Advances in Pure Mathematics, 2016, 6, 498-506
Published Online June 2016 in SciRes. http://www.scirp.org/journal/apm
http://dx.doi.org/10.4236/apm.2016.67036

An Efficient Adaptive Iteratively Reweighted ℓ_1 Algorithm for Elastic ℓ_q Regularization

Yong Zhang, Wanzhou Ye

Department of Mathematics, College of Science, Shanghai University, Shanghai, China
Email: 13820161@shu.edu.cn, wzhy@shu.edu.cn

Received 12 May 2016; accepted 13 June 2016; published 16 June 2016

Abstract

In this paper, we propose an efficient adaptive iteratively reweighted ℓ_1 algorithm (A-IRL1 algorithm) for solving the elastic ℓ_q regularization problem. We prove that the sequence generated by the A-IRL1 algorithm is convergent for any rational $q \in (0,1)$ and the limit is a critical point of the elastic ℓ_q regularization problem. Under certain conditions, we present an error bound for the limit point of convergent sequence.

Keywords

Compressed Sensing, Elastic ℓ_q Minimization, Nonconvex Optimization, Convergence, Critical Point

1. Introduction

Compressed sensing (CS) has been emerging as very active research field and brings about great changes in the fields of signal processing in recent years [1] [2]. The main task of CS focuses on the recovery of sparse signal from a small number of linear measurement data. It can be mathematically modeled as following optimization problem,

$$\min_{x \in R^N} \|x\|_0 \text{ subject to } Ax = b, \tag{1}$$

where $b \in R^m$, $A \in R^{m \times N}$ (commonly $m < N$) is a measurement matrix and $\|x\|_0$, formally called the ℓ_0 quasi-norm, denotes the number of nonzero components of $x = (x_1, x_2, \cdots, x_N)^T \in R^N$. In general, it is difficult to tackle problem (1) due to its nonsmooth and nonconvex nature. In recent years, some researchers have

How to cite this paper: Zhang, Y. and Ye, W.Z. (2016) An Efficient Adaptive Iteratively Reweighted ℓ_1 Algorithm for Elastic ℓ_q Regularization. *Advances in Pure Mathematics*, **6**, 498-506. http://dx.doi.org/10.4236/apm.2016.67036

proposed the ℓ_q norm regularization problem [3]-[5] with $0<q\le 1$, that is, to consider the following ℓ_q regularization problem

$$\min_{x\in R^N} \|x\|_q^q \text{ subject to } Ax=b, \tag{2}$$

or the unconstrained ℓ_q regularization problem

$$\min_{x\in R^N} \left\{\frac{1}{2}\|Ax-b\|_2^2+\lambda\|x\|_q^q\right\}, \tag{3}$$

where $\|x\|_q=\left(\sum_{i=1}^N |x_i|^q\right)^{1/q}$ for $0<q\le 1$ and $\lambda>0$ is a regularization parameter.

When $q=1$, it is well known that the problems (2) and (3) are both convex optimization problems, and therefore, can be solved efficiently [6] [7]. On the other hand, when $0<q<1$, the above problems (2) and (3) lead to nonconvex, nonsmooth and even non-Lipschitz optimization problem. It is difficult to solve them fastly and efficiently. Iterative reweighted algorithms, which include iteratively reweighted ℓ_1 algorithm [8] and iteratively reweighted least squares [9], are very effective for solving the nonconvex ℓ_q regularization problem.

In this paper, we consider the following elastic ℓ_q regularization problem,

$$\min_{x\in R^N} \left\{\frac{1}{2}\|Ax-b\|_2^2+\lambda_1\|x\|_q^q+\lambda_2\|x\|_2^2\right\}, \tag{4}$$

where $\lambda_1,\lambda_2>0$ are two parameters. When $q=1$, the above problem (4) reduces to the well-known elastic-net regularization proposed by Zou and Hastie [10], which is an effective method for variable selection. In [10], Zou *et al*. showed that this method outperformed Lasso [11] in terms of prediction accuracy for both simulation studies and real-data applications on variable selection. For further statistical properties of the elastic-net regularization in detail, we refer to references [12] [13]. When $0<q<1$, problem (4) is an extension of elastic net regularization from ℓ_1 penalty to ℓ_q penalty. In statistics, elastic ℓ_q regularization is usually very effective for group variable selection.

Obviously, for $0<q<1$, the ℓ_q norm term in (4) is not differentiable at zero. Therefore, in this paper, we study the following relaxed elastic ℓ_q minimization problem with $0<q<1$

$$\min_{x\in R^N} L(x,\varepsilon,\lambda_1,\lambda_2)=\frac{1}{2}\|Ax-b\|_2^2+\lambda_1\sum_{i=1}^N\left(|x_i|+\varepsilon\right)^q+\lambda_2\|x\|_2^2, \tag{5}$$

The model (5) can be considered as an approximation to the model (4) as $\varepsilon\to 0$. In order to solve the above problem (5), we propose the following adaptive iteratively reweighted ℓ_1 minimization algorithm (A-IRL1 algorithm),

$$x^{k+1}\in\arg\min_{x\in R^N}\left\{L_k(x,\varepsilon_k,\lambda_1,\lambda_2)=\frac{1}{2}\|Ax-b\|_2^2+\lambda_1\|W^k x\|_1+\lambda_2\|x\|_2^2\right\}, \tag{6}$$

where the weight $W^k=diag\left(w^k\right)$ is defined by the previous iterates and updated in each iteration as

$$w_i^k=\frac{q}{\left(|x_i^k|+\varepsilon_k\right)^{1-q}},\ i=1,\cdots,N.$$

The adaptive iteratively update of ε_k in the proposed algorithm is the same as the one in [9], which is also adopted in [14]. The A-IRL1 algorithm (6) solves a convex $\ell_2-\ell_1$ minimization problem, which can be solved by many efficient algorithms [6] [7] [15].

The relaxed elastic ℓ_q regularization problem (5) can be solved by A-IRL1 algorithm (6). In this paper, we prove that any sequence generated by the A-IRL1 algorithm (6) is convergent itself for any rational $q\in(0,1)$ as the case $\varepsilon_k\to\varepsilon_*>0$. Moreover, we present an error bound between the limit point and the sparse solution of problem (1).

The rest of this paper is organized as follows: In Section 2, we summarize the A-IRL1 algorithm for solving elastic ℓ_q regularization problem (5). In Section 3, we present a detail convergence analysis for the A-IRL1 algorithm (6). We prove that the A-IRL1 algorithm is convergent for any rational $q \in (0,1)$ based on an algebraic method with $\varepsilon_* > 0$. Furthermore, under certain conditions, we present an error bound between the limit point and the sparse solution of problem (1). Finally, a conclusion is given in Section 4.

2. A-IRL1 Algorithm for Solving Elastic ℓ_q Regularization

We give a detailed implementation of A-IRL1 algorithm (6) for solving elastic ℓ_q regularization problem (5). The algorithm is summarized as **Algorithm 1**.

Algorithm 1. Adaptive Iteratively reweighted ℓ_1 algorithm (6) for solving elastic ℓ_q regularization (5).

Input: matrix $A \in R^{m\times N}$, vector $b \in R^m$ and estimated sparsity level s;
Choose approximate parameters $\lambda_1 > 0, \lambda_2 > 0$, $q \in (0,1)$.
Initialize $x^0 \in R^N$ such that $Ax^0 = b$, $k = 0$, $\varepsilon_0 = 1$, $w_i^0 = \frac{q}{(|x_i^0|+\varepsilon_0)^{1-q}}$, $i = 1, \cdots, N$.
For $k = 0, 1, 2, \cdots$
  $x^{k+1} \in \arg\min_{x\in R^N} \left\{\frac{1}{2}\|Ax-b\|_2^2 + \lambda_1\|W^k x\|_1 + \lambda_2\|x\|_2^2\right\}$,
  Update $\varepsilon_{k+1} = \min\{\varepsilon_k, \alpha \cdot r(x^{k+1})_{s+1}\}$, where $\alpha \in (0,1)$ is a constant.
  Update w_i^{k+1} by $w_i^{k+1} = \frac{q}{(|x_i^{k+1}|+\varepsilon_{k+1})^{1-q}}$, $i = 1, \cdots, N$.
End For

In **Algorithm 1**, $r(x)$ is the rearrangement of the absolute values of $x \in R^N$ in decreasing order. If $\varepsilon_{k+1} = 0$, we choose x^{k+1} to be the approximate sparse solution and stop iteration. Otherwise, we stop the algorithm within a reasonable time and return the last x^{k+1}.

It is clear from Algorithm 1 that $\{\varepsilon_k\}$ is a nonincreasing sequence which is convergent to some nonnegative number ε_*. In the next section, we prove that the sequence $\{x^k\}$ is convergent when $\varepsilon_* > 0$, and the limit is a critical point of problem (5) with $\varepsilon = \varepsilon_*$. Furthermore, we also present an error bound for the limit point.

3. Convergence of Algorithm 1

In this section, we first prove that the the sequence $\{x^k\}$ generated by Algorithm 1 is bounded and asymptotically regular. Then, based on an algebraic method, we prove that Algorithm 1 is convergent for any rational $q \in (0,1)$ with $\varepsilon_* > 0$. Next, we begin with the following inequality.

Lemma 1. *Given* $0 < q < 1$ *and* $\varepsilon_k \ge \varepsilon_{k+1} \ge 0$, *then the inequality*

$$(\varepsilon_k + |\alpha|)^q - (\varepsilon_{k+1} + |\beta|)^q - \frac{q(|\alpha| - |\beta|)}{(\varepsilon_k + |\alpha|)^{1-q}} \ge 0, \tag{7}$$

holds for any $\alpha, \beta \in R$.

Proof. We first define $f(t) = t^q$ $(t > 0)$. For any $t_1, t_2 > 0$, by the mean value theorem, we have

$$f(t_1) - f(t_2) = f'(\xi)(t_1 - t_2) \quad \text{where } \xi \text{ lies between } t_1 \text{ and } t_2. \tag{8}$$

The following inequality is always hold for any $t_1 > t_2$, $t_1 < t_2$ or $t_1 = t_2$,

$$f(t_1) - f(t_2) \ge f'(t_1)(t_1 - t_2).$$

Let $t_1 = \varepsilon_k + |\alpha|$ and $t_2 = \varepsilon_{k+1} + |\beta|$, we thus have

$$(\varepsilon_k + |\alpha|)^q - (\varepsilon_{k+1} + |\beta|)^q \ge \frac{q((\varepsilon_k - \varepsilon_{k+1}) + (|\alpha| - |\beta|))}{(\varepsilon_k + |\alpha|)^{1-q}} \ge \frac{q(|\alpha| - |\beta|)}{(\varepsilon_k + |\alpha|)^{1-q}}. \tag{9}$$

After rewriting the terms of (9), we thus get the desired inequality (7).

Our next result shows the monotonicity of $L(x^k, \varepsilon_k, \lambda_1, \lambda_2)$ along the sequence $\{x^k\}$ and this sequence is also asymptotically regular.

Lemma 2. *Let* $\{x^k\}$ *be the sequence generated by Algorithm 1. Then we have*

$$\|Ax^k - Ax^{k+1}\|_2^2 \le 2\left(L(x^k, \varepsilon_k, \lambda_1, \lambda_2) - L(x^{k+1}, \varepsilon_{k+1}, \lambda_1, \lambda_2)\right). \tag{10}$$

Furthermore,

$$\left\|x^k - x^{k+1}\right\|_2^2 \le \lambda_2^{-1}\left(L\left(x^k, \varepsilon_k, \lambda_1, \lambda_2\right) - L\left(x^{k+1}, \varepsilon_{k+1}, \lambda_1, \lambda_2\right)\right). \tag{11}$$

Proof. Since x^{k+1} is a solution of problem (6), we thus have,

$$0 \in \partial L_k\left(x^{k+1}, \varepsilon_k, \lambda_1, \lambda_2\right).$$

Besides, we can get the subgradient of $L_k\left(x, \varepsilon_k, \lambda_1, \lambda_2\right)$ as follows,

$$\partial L_k\left(x, \varepsilon_k, \lambda_1, \lambda_2\right) = \lambda_1 \left[\frac{q\partial\left|x_i\right|}{\left(\left|x_i^k\right| + \varepsilon_k\right)^{1-q}}\right]_{1\le i\le N} + A^{\mathrm{T}}\left(Ax - b\right) + 2\lambda_2 x. \tag{12}$$

Hence, we find

$$0 \in \partial L_k\left(x^{k+1}, \varepsilon_k, \lambda_1, \lambda_2\right) = \lambda_1 \left[\frac{q\partial\left|x_i^{k+1}\right|}{\left(\left|x_i^k\right| + \varepsilon_k\right)^{1-q}}\right]_{1\le i\le N} + A^{\mathrm{T}}\left(Ax^{k+1} - b\right) + 2\lambda_2 x^{k+1}, \tag{13}$$

which means that there exists $c_i^{k+1} \in \partial\left|x_i^{k+1}\right|, i = 1, \cdots, N$ such that

$$\lambda_1 \left[\frac{qc_i^{k+1}}{\left(\left|x_i^k\right| + \varepsilon_k\right)^{1-q}}\right]_{1\le i\le N} + A^{\mathrm{T}}\left(Ax^{k+1} - b\right) + 2\lambda_2 x^{k+1} = 0, \tag{14}$$

where $c_i^{k+1} = \begin{cases} 1, & \text{if } x_i^{k+1} > 0, \\ -1, & \text{if } x_i^{k+1} < 0, \\ \alpha, & \text{if } x_i^{k+1} = 0, \alpha \in [-1,1]. \end{cases}$

We compute

$$\begin{aligned} & L\left(x^k, \varepsilon_k, \lambda_1, \lambda_2\right) - L\left(x^{k+1}, \varepsilon_{k+1}, \lambda_1, \lambda_2\right) \\ & = \lambda_1 \sum_{i=1}^N \left(\left(\left|x_i^k\right| + \varepsilon_k\right)^q - \left(\left|x_i^{k+1}\right| + \varepsilon_{k+1}\right)^q\right) + \frac{1}{2}\left(\left\|Ax^k - b\right\|_2^2 - \left\|Ax^{k+1} - b\right\|_2^2\right) + \lambda_2\left(\left\|x^k\right\|_2^2 - \left\|x^{k+1}\right\|_2^2\right) \\ & = \lambda_1 \sum_{i=1}^N \left(\left(\left|x_i^k\right| + \varepsilon_k\right)^q - \left(\left|x_i^{k+1}\right| + \varepsilon_{k+1}\right)^q\right) + \frac{1}{2}\left\|Ax^k - Ax^{k+1}\right\|_2^2 + \lambda_2\left\|x^k - x^{k+1}\right\|_2^2 \\ & \quad + \left(Ax^k - Ax^{k+1}\right)^{\mathrm{T}}\left(Ax^{k+1} - b\right) + 2\lambda_2\left(x^k - x^{k+1}\right)^{\mathrm{T}} x^{k+1}. \end{aligned} \tag{15}$$

Using (14), we have

$$\left(Ax^k - Ax^{k+1}\right)^{\mathrm{T}}\left(Ax^{k+1} - b\right) + 2\lambda_2\left(x^k - x^{k+1}\right)^{\mathrm{T}} x^{k+1} = \lambda_1 \sum_{i=1}^N \frac{qc_i^{k+1}\left(x_i^{k+1} - x_i^k\right)}{\left(\left|x_i^k\right| + \varepsilon_k\right)^{1-q}}. \tag{16}$$

Substituting (16) into (15) and using Lemma 1 yields

$$\begin{aligned} & L\left(x^k, \varepsilon_k, \lambda_1, \lambda_2\right) - L\left(x^{k+1}, \varepsilon_{k+1}, \lambda_1, \lambda_2\right) \\ & = \lambda_1 \sum_{i=1}^N \left(\left(\left|x_i^k\right| + \varepsilon_k\right)^q - \left(\left|x_i^{k+1}\right| + \varepsilon_{k+1}\right)^q + \frac{qc_i^{k+1}\left(x_i^{k+1} - x_i^k\right)}{\left(\left|x_i^k\right| + \varepsilon_k\right)^{1-q}}\right) + \frac{1}{2}\left\|Ax^k - Ax^{k+1}\right\|_2^2 + \lambda_2\left\|x^k - x^{k+1}\right\|_2^2 \\ & \ge \lambda_1 \sum_{i=1}^N \left(\left(\left|x_i^k\right| + \varepsilon_k\right)^q - \left(\left|x_i^{k+1}\right| + \varepsilon_{k+1}\right)^q + \frac{q\left(\left|x_i^{k+1}\right| - \left|x_i^k\right|\right)}{\left(\left|x_i^k\right| + \varepsilon_k\right)^{1-q}}\right) + \frac{1}{2}\left\|Ax^k - Ax^{k+1}\right\|_2^2 + \lambda_2\left\|x^k - x^{k+1}\right\|_2^2 \\ & \ge \frac{1}{2}\left\|Ax^k - Ax^{k+1}\right\|_2^2 + \lambda_2\left\|x^k - x^{k+1}\right\|_2^2, \end{aligned} \tag{17}$$

where the first inequality uses $c_i^{k+1}x_i^{k+1}=\left|x_i^{k+1}\right|$ and $\left|c_i^{k+1}\right|\le 1$, and the last inequality uses Lemma 1. Therefore, from (17) we get the desired results (10) and (11).

From Lemma 2 (10), we know that $L\left(x^k,\varepsilon_k,\lambda_1,\lambda_2\right)$ is monotonically decreasing and bounded. Otherwise, $L\left(x,\varepsilon,\lambda_1,\lambda_2\right)\to\infty$ as $\|x\|_2\to\infty$. Therefore, $\left\{x^k\right\}$ is also bounded. On the other hand, from (11) we obtain that the sequence $\left\{x^k\right\}$ is asymptotically regular.

In order to prove that the whole sequence generated by Algorithm 1 is convergent, we need the following lemma, which plays an important role in the proof of convergence. The following lemma mainly states that for almost every system of n polynomial equations in n complex variables, if its corresponding highest ordered system of equations have only trivial solution, then there is a finite number of solutions to the n polynomial equations. For detailed proof refer to Theorem 3.1 in [16].

Lemma 3. ([16]) *Let n polynomial equations in n complex variables $P\left(z,\overline{w}\right)=0$ be given, and let $Q\left(z,\overline{a},\overline{c}\right)=0$ be its corresponding highest ordered system of equations. If $Q\left(z,\overline{a},\overline{c}\right)=0$ has only the trivial solution $z=0$, then $P\left(z,\overline{w}\right)=0$ has $\Gamma=\prod_{i=1}^{n}q_i$ solutions, where q_i is the degree of P_i.*

With above lemmas, we are now in a position to present the convergence of Algorithm 1 for any rational number $q\in\left(0,1\right)$ with $\varepsilon_*>0$.

Theorem 1 *For any $\lambda_1,\lambda_2>0$, if the limit of $\left\{\varepsilon_k\right\}$ is $\varepsilon_*>0$, then the sequence $\left\{x^k\right\}$ generated by Algorithm* 1 *is convergent. Denoting the limit by x^*, i.e., $\lim_{k\to\infty}x^k=x^*$. Moreover, the limit x^* is a critical point of problem* (5) *with $\varepsilon=\varepsilon_*$.*

Proof. From (10), we know that the sequence $\left\{L\left(x^k,\varepsilon_k,\lambda_1,\lambda_2\right)\right\}$ is monotonically decreasing and bounded below. Thus, we can infer that the sequence $\left\{x^k\right\}$ is also bounded. The boundedness of $\left\{x^k\right\}$ implies that there exists at least one convergent subsequence. We assume that $\left\{x^{k_j}\right\}$ is any one of the convergent subsequences of $\left\{x^k\right\}$ with limit x^*. By (11), we know that the sequence $\left\{x^{k_j+1}\right\}$ also converges to x^*. Now replacing x^k, x^{k+1}, ε_k, c_i^{k+1} with x^{k_j}, x^{k_j+1}, ε_{k_j}, $c_i^{k_j+1}$ in (14) respectively, and letting $j\to\infty$ yields

$$\left[\frac{\lambda_1 q c_i^*}{\left(\left|x_i^*\right|+\varepsilon_*\right)^{1-q}}\right]_{1\le i\le N}+A^{\mathrm{T}}\left(Ax^*-b\right)+2\lambda_2 x^*=0, \tag{18}$$

where $c_i^*\in\partial\left|x_i^*\right|, i=1,\cdots,N$.

The above Equation (18) demonstrates that the limit of any convergent subsequence of $\left\{x^k\right\}$ is a stationary point of problem (5) with $\varepsilon=\varepsilon_*>0$. In order to prove the convergence of the whole sequence $\left\{x^k\right\}$, one first needs to prove that the limit point set, denoted by $Y_{\left\{x^k\right\}}$, which contains all the limit points of convergent subsequence of $\left\{x^k\right\}$, is a finite set.

A classification is made for the limit point set $Y_{\left\{x^k\right\}}$ with different sparsity s, $1\le s\le N$. That is the set

$$\Omega=\left\{x^*\in Y_{\left\{x^k\right\}},\left|supp\left(x^*\right)\right|=s\right\},$$

which contains all the limit points with each sparsity s. If we prove the set Ω is a finite set, then we obtain that the limit point set $Y_{\left\{x^k\right\}}$ is also a finite set. Without loss of generality, we define a set

$$C^s=\left\{x^*=\left(x_1^*,x_2^*,\cdots,x_s^*,0,\cdots,0\right)^{\mathrm{T}}\in\Omega\right\}. \tag{19}$$

Furthermore, for any given $\eta=\left(\eta_1,\cdots,\eta_s\right)^{\mathrm{T}}$ with $\eta_i=\pm1, i=1,\cdots,s$, we define another set

$$C^s\left(\eta_1,\cdots,\eta_s\right)=\left\{x^*\in C^s, sign\left(x_i^*\right)=\eta_i, 1\le i\le s\right\}. \tag{20}$$

From (19) and (20), we have $C^s=\bigcup_{\eta_i=\pm1,1\le i\le s}C^s\left(\eta_1,\cdots,\eta_s\right)$. If we prove that the set $C^s\left(\eta_1,\cdots,\eta_s\right)$ is a finite set, then it implies that the set C^s is also finite, and we further conclude that the limit point set $Y_{\left\{x^k\right\}}$ is a finite set.

For any $x^* = \left(x_1^*, x_2^*, \cdots, x_s^*, 0, \cdots, 0\right)^{\mathrm{T}} \in C^s\left(\eta_1, \cdots, \eta_s\right)$ with $\eta_i = \pm 1, i = 1, \cdots, s$, let $S = supp\left(x^*\right)$ denotes the support set of x^* with $|S| = s$. By (18), we know that x^* satisfies the following equation

$$\left[\frac{\lambda_1 q \eta_i}{\left(x_i^* \eta_i + \varepsilon_*\right)^{1-q}}\right]_{1\le i\le s} + A_S^{\mathrm{T}}\left(A_S x_S^* - b\right) + 2\lambda_2 x_S^* = 0, \tag{21}$$

where x_S^* denotes the subvector of x^* with components restricted to S and A_S denotes the submatrix of A with columns restricted to S. Next, if we prove the Equation (21) has finite solutions, then we can obtain the set $C^s\left(\eta_1, \cdots, \eta_s\right)$ as a finite set.

It is clear that (21) can be rewritten as follows:

$$\left[\frac{\lambda_1 q \eta_i}{\left(x_i^* \eta_i + \varepsilon_*\right)^{1-q}}\right]_{1\le i\le s} + \left(A_S^{\mathrm{T}} A_S + 2\lambda_2 I_S\right) x_S^* - A_S^{\mathrm{T}} b = 0, \tag{22}$$

where $A_S^{\mathrm{T}} A_S + 2\lambda_2 I_S$ is an $s \times s$ positive-definite matrix, $A_S^{\mathrm{T}} b \in R^s$ and I_S is the $s \times s$ identity matrix. We observe that (22) can further be rewritten as follows:

$$\lambda_1 q \eta + \mathbb{W}\left(\left(A_S^{\mathrm{T}} A_S + 2\lambda_2 I_S\right) x_S^* - A_S^{\mathrm{T}} b\right) = 0, \tag{23}$$

where $\mathbb{W}$ is an $s \times s$ diagonal matrix with the diagonal entries $\mathbb{W}_{ii} = \left(x_i^* \eta_i + \varepsilon_*\right)^{1-q}$ for $i = 1, \cdots, s$. Without loss of generality, we denote $A_S^{\mathrm{T}} A_S + 2\lambda_2 I_S = \left(a_{ij}\right)_{s\times s}$ and $A_S^{\mathrm{T}} b = \left(p_1, p_2, \cdots, p_s\right)^{\mathrm{T}}$. Let $1 - q = \frac{\gamma}{p}$, where γ, p are two positive integers. By using simple calculation for Equation (23), we get the following system:

$$\begin{cases} \left(a_{11}x_1^* + a_{12}x_2^* + \cdots + a_{1s}x_s^* - p_1\right)^{2p}\left(\left(x_1^*\right)^2 + 2\varepsilon_*\eta_1 x_1^* + \varepsilon_*^2\right)^{\gamma} - \left(q\lambda_1\right)^{2p} = 0, \\ \left(a_{21}x_1^* + a_{22}x_2^* + \cdots + a_{2s}x_s^* - p_2\right)^{2p}\left(\left(x_2^*\right)^2 + 2\varepsilon_*\eta_2 x_2^* + \varepsilon_*^2\right)^{\gamma} - \left(q\lambda_1\right)^{2p} = 0, \\ \qquad \vdots \\ \left(a_{s1}x_1^* + a_{s2}x_2^* + \cdots + a_{ss}x_s^* - p_s\right)^{2p}\left(\left(x_s^*\right)^2 + 2\varepsilon_*\eta_s x_s^* + \varepsilon_*^2\right)^{\gamma} - \left(q\lambda_1\right)^{2p} = 0. \end{cases} \tag{24}$$

Since all the solutions of Equation (21) satisfy (24), we can thus show that Equation (21) has finite solutions as long as we can prove that (24) has finite solutions. To do that, we show that the following system has finite solutions:

$$\begin{cases} \left(a_{11}u_1 + a_{12}u_2 + \cdots + a_{1s}u_s - p_1\right)^{2p}\left(u_1^2 + 2\varepsilon_*\eta_1 u_1 + \varepsilon_*^2\right)^{\gamma} - \left(q\lambda_1\right)^{2p} = 0, \\ \left(a_{21}u_1 + a_{22}u_2 + \cdots + a_{2s}u_s - p_2\right)^{2p}\left(u_2^2 + 2\varepsilon_*\eta_2 u_2 + \varepsilon_*^2\right)^{\gamma} - \left(q\lambda_1\right)^{2p} = 0, \\ \qquad \vdots \\ \left(a_{s1}u_1 + a_{s2}u_2 + \cdots + a_{ss}u_s - p_s\right)^{2p}\left(u_s^2 + 2\varepsilon_*\eta_s u_s + \varepsilon_*^2\right)^{\gamma} - \left(q\lambda_1\right)^{2p} = 0, \end{cases} \tag{25}$$

where $u = \left(u_1, u_2, \cdots, u_s\right)^{\mathrm{T}} \in R^s$. Now, we extract the highest order terms from system (25) to get the following system:

$$\begin{cases} \left(a_{11}u_1 + a_{12}u_2 + \cdots + a_{1s}u_s\right)^{2p} u_1^{2\gamma} = 0, \\ \left(a_{21}u_1 + a_{22}u_2 + \cdots + a_{2s}u_s\right)^{2p} u_2^{2\gamma} = 0, \\ \qquad \vdots \\ \left(a_{s1}u_1 + a_{s2}u_2 + \cdots + a_{ss}u_s\right)^{2p} u_s^{2\gamma} = 0. \end{cases} \tag{26}$$

To prove that system (26) has only trivial solution, we use the method of proof by contradiction. Without loss of generality, we assume $u=(u_1,u_2,\cdots,u_t,0,0,\cdots,0)\in R^s$ is a nonzero solution of (26), $u_i\neq 0$ for $i=1,2,\cdots,t$, and $1\le t\le s$. By the assumption $u_i\neq 0$ for $i=1,2,\cdots,t$, and from (26) we can get the following equation:

$$Bu^t=0 \tag{27}$$

where $B=(a_{ij})_{t\times t}$ is the $t\times t$ leading principal submatrix of matrix $A_S^{\mathrm{T}}A_S+2\lambda_2 I_S$ and $u^t=(u_1,u_2,\cdots,u_t)^{\mathrm{T}}\in R^t$ for $1\le t\le s$. Because the matrix $A_S^{\mathrm{T}}A_S+2\lambda_2 I_S$ is positive definite; implies that the matrix B is also positive definite, and thus we have $u_i=0$ for $i=1,2,\cdots,t$. This contradicts the assumption that $u_i\neq 0$, $i=1,2,\cdots,t$. Therefore, we get that the system (26) has only trivial solutions. According to Lemma 3, we deduce that the system (25) has finite solutions, which further implies that the Equation (21) has also finite solutions, that is, the set $C^s(\eta_1,\cdots,\eta_s)$ is a finite set. Therefore, we get that the limit point set $Y_{\{x^k\}}$ is a finite set.

Combining with $\|x^k-x^{k+1}\|_2\to 0$ as $k\to\infty$, we thus obtain that the sequence $\{x^k\}$ is convergent. By the convergence of sequence $\{x^k\}$ and (18), we obtain that the limit x^* is a critical point of problem (5) with $\varepsilon=\varepsilon_*$.

Theorem 1 gives a detailed convergence proof of Algorithm 1 based on an algebraic approach. In the next, we will present an error bound between the convergent limit and the sparse solution of problem (1).

Under the Restricted Isometry Property (RIP) on the matrix *A*, we present an error bound between the convergent limit and the sparse solution of problem (1). First of all, we give a definition of RIP on the matrix *A* as follows.

Definition: For every integer $1\le s\le N$, we define δ_s as the *s*-restricted isometry constant of *A* as the smallest positive quantity such that

$$(1-\delta_s)\|x\|_2^2\le\|Ax\|_2^2\le(1+\delta_s)\|x\|_2^2 \tag{28}$$

for all subsets $T\subset\{1,2,\cdots,N\}$ of cardinality at most *s* and vectors *x* supported on *T*.

Under the RIP assumption, we can ensure that the limit x^* is a reasonable approximation of the sparse solution if x^* has a very small tail in the sense that

$$\sigma_s(x^*)_p=\inf_{\|y\|_0\le s}\|x^*-y\|_p$$

for $p\ge 1$, which is the error term of the best *s*-term approximation of x^* in the ℓ_p-norm.

With the concept of RIP, we are able to prove the result of following theorem.

Theorem 2. *Suppose that x is an s-sparse solution of* (1) *satisfying* $Ax=b$. *Assume that A satisfies the RIP of order* 2*s with* $\delta_{2s}<1$ *and* $\varepsilon_k\to\varepsilon_*$. *For any fixed* $\lambda_1,\lambda_2>0$.

(1) When $\varepsilon_*>0$, the limit x^* of convergent sequence $\{x^k\}$ is a critical point of problem (5) with $\varepsilon=\varepsilon_*$, and it satisfies

$$\|x^*-x\|_2\le C_1\sqrt{\lambda_1}+C_2\sigma_s(x^*)_2, \tag{29}$$

(2) When $\varepsilon_*=0$, there must exist a subsequence from $\{x^k\}$ converging to an *s*-sparse point x^{ε_*} which satisfies

$$\|x^{\varepsilon_*}-x\|_2\le C\sqrt{\lambda_1}. \tag{30}$$

Here C_1, C_2 and *C* are positive constants dependent on $\delta_{2s},\lambda_1,\lambda_2$ and the initial point x^0.

Proof. (1). In Theorem 1, we have proved that the limit x^* of convergent sequence $\{x^k\}$ is a critical point of problem (5) with $\varepsilon=\varepsilon_*>0$.

We use Lemma 2 to get

$$\begin{aligned}L(x^*,\varepsilon_*,\lambda_1,\lambda_2)&\le L(x^k,\varepsilon_k,\lambda_1,\lambda_2)\le L(x^0,\varepsilon_0,\lambda_1,\lambda_2)\\&\le\lambda_1\left(\|x^0\|_q^q+N\varepsilon_0^q\right)+\lambda_2\|x^0\|_2^2=\lambda_1\left(\|x^0\|_q^q+N\right)+\lambda_2\|x^0\|_2^2,\end{aligned} \tag{31}$$

where we use the assumption that the initial value x^0 satisfies $Ax^0 = b$ and $\varepsilon_0 = 1$ in Algorithm 1. By (31), we have

$$\left\|Ax^* - b\right\|_2 \le \sqrt{2L\left(x^*, \varepsilon_*, \lambda_1, \lambda_2\right)} \le \sqrt{2\lambda_1\left(\left\|x^0\right\|_q^q + N + \frac{\lambda_2}{\lambda_1}\left\|x^0\right\|_2^2\right)}.$$

et S be the index set of the s-sparse solution x, and let S^* be the index set of s largest entries in the absolute value of x^*. Since $\|x\|_0 \le s$, we have

$$\begin{aligned}
\left\|x^* - x\right\|_2 &\le \left\|\left(x^* - x\right)_{S\cup S^*}\right\|_2 + \left\|x^*_{\left(S\cup S^*\right)^c}\right\|_2 \\
&\le \frac{1}{\sqrt{1-\delta_{2s}}}\left\|A\left(x^* - x\right)_{S\cup S^*}\right\|_2 + \left\|x^*_{\left(S\cup S^*\right)^c}\right\|_2 \\
&\le \frac{1}{\sqrt{1-\delta_{2s}}}\left\|Ax^* - b\right\|_2 + \left(\frac{\|A\|_2}{\sqrt{1-\delta_{2s}}} + 1\right)\left\|x^*_{\left(S\cup S^*\right)^c}\right\|_2 \\
&\le \frac{1}{\sqrt{1-\delta_{2s}}}\sqrt{2\lambda_1\left(\left\|x^0\right\|_q^q + N + \frac{\lambda_2}{\lambda_1}\left\|x^0\right\|_2^2\right)} + \left(\frac{\|A\|_2}{\sqrt{1-\delta_{2s}}} + 1\right)\sigma_s\left(x^*\right)_2.
\end{aligned}$$

Letting $C_1 = \sqrt{2\left(\left\|x^0\right\|_q^q + N + \frac{\lambda_2}{\lambda_1}\left\|x^0\right\|_2^2\right)} \Big/ \sqrt{1-\delta_{2s}}$ and $C_2 = \frac{\|A\|_2}{\sqrt{1-\delta_{2s}}} + 1$.

(3) If $\varepsilon_* = 0$, then $\varepsilon_k = 0$ for some k or $\varepsilon_k = \alpha \cdot r\left(x^{m_k}\right)_{s+1}$ holds for sufficiently large k and some integer $m_k \le k$. In the former case, x^k is an s-sparse vector, and we denote $x^{\varepsilon_*} = x^k$. In the latter case, by the boundedness of $\left\{x^{m_k}\right\}$, we have $x^{\varepsilon_*} = \lim_{k\to\infty} x^{m_k}$. Then $r\left(x^{\varepsilon_*}\right)_{s+1} = \lim_{k\to\infty} r\left(x^{m_k}\right)_{s+1} = \lim_{k\to\infty}\frac{\varepsilon_k}{\alpha} = 0$. That is, x^{ε_*} is an s-sparse vector. Therefore, in both cases, we have an s-sparse limit point x^{ε_*}. Without loss of generality, we assume $x^k \to x^{\varepsilon_*}$. Using RIP of A, we get

$$\begin{aligned}
\left\|x^{\varepsilon_*} - x\right\|_2 &\le \frac{1}{\sqrt{1-\delta_{2s}}}\left\|A\left(x^{\varepsilon_*} - x\right)\right\|_2 = \frac{1}{\sqrt{1-\delta_{2s}}}\lim_{k\to\infty}\left\|Ax^k - b\right\|_2 \\
&\le \frac{1}{\sqrt{1-\delta_{2s}}}\lim_{k\to\infty}\sqrt{2L\left(x^k, \varepsilon_k, \lambda_1, \lambda_2\right)} = \frac{1}{\sqrt{1-\delta_{2s}}}\sqrt{2L\left(x^*, \varepsilon_*, \lambda_1, \lambda_2\right)} \\
&\le \frac{1}{\sqrt{1-\delta_{2s}}}\sqrt{2L\left(x^0, \varepsilon_*, \lambda_1, \lambda_2\right)} = \frac{1}{\sqrt{1-\delta_{2s}}}\sqrt{2\lambda_1\left(\left\|x^0\right\|_q^q + \frac{\lambda_2}{\lambda_1}\left\|x^0\right\|_2^2\right)},
\end{aligned}$$

where the third inequality uses (31) with $\varepsilon_k = \varepsilon_*$, the last equality uses the assumption $Ax^0 = b$ and $\varepsilon_* = 0$. Denoting $C = \sqrt{2\left(\left\|x^0\right\|_q^q + \frac{\lambda_2}{\lambda_1}\left\|x^0\right\|_2^2\right)} \Big/ \sqrt{1-\delta_{2s}}$. This completes the proof.

Under the condition of RIP on the matrix A, when $\varepsilon_* > 0$, Theorem 2 provide an error bound between the convergent limit and the sparse solution of problem (1). While $\varepsilon_* = 0$, we present an error bound for the limit point of any convergent subsequence. In this case, the limit point of any convergent subsequence is an s-sparse vector.

4. Conclusion

The iteratively reweighted ℓ_1 algorithm has been widely used for solving nonconvex optimization problem. In this paper, we propose an efficient adaptive iteratively reweighted ℓ_1 algorithm (6) for solving the elastic ℓ_q regularization (5) and we prove the convergence of the algorithm. In particular, we first prove that the sequence generated by Algorithm 1 is bounded and the sequence is asymptotically regular. When $\varepsilon_* > 0$, based on an algebraic method, we prove that the sequence generated by Algorithm 1 is convergent for any rational $q \in (0,1)$ and the limit is a critical point of problem (5) with $\varepsilon = \varepsilon_*$. Furthermore, under the condition of the RIP on the matrix *A*, when $\varepsilon_* > 0$, we present an error bound between the convergent limit and the sparse solution of problem (1). While $\varepsilon_* = 0$, we present an error bound for the limit point of any convergent subsequence. Our established convergence results provide a theoretical guarantee for a wide range of applications of adaptive iteratively reweighted ℓ_1 algorithm.

References

[1] Donoho, D.L. (2006) Compressed Sensing. *IEEE Transactions on Information Theory*, **52**, 1289-1306. http://dx.doi.org/10.1109/TIT.2006.871582

[2] Candès, E., Romberg, J. and Tao, T. (2006) Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information. *IEEE Transactions on Information Theory*, **52**, 489-509. http://dx.doi.org/10.1109/TIT.2005.862083

[3] Chartrand, R. (2007) Exact Reconstruction of Sparse Signals via Nonconvex Minimization. *IEEE Signal Processing Letters*, **14**, 707-710. http://dx.doi.org/10.1109/LSP.2007.898300

[4] Xu, Z.B., Chang, X.Y., Xu, F.M. and Zhang, H. (2012) $L_{1/2}$ Regularization: A Thresholding Representation Theory and a Fast Solver. *IEEE Transactions on Neural Networks and Learning Systems*, **23**, 1013-1027. http://dx.doi.org/10.1109/TNNLS.2012.2197412

[5] Foucart, S. and Lai, M. (2009) Sparsest Solutions of Underdetermined Linear Systems via ℓ_q Minimization for $0 < q \le 1$. *Applied and Computational Harmonic Analysis*, **26**, 395-407. http://dx.doi.org/10.1016/j.acha.2008.09.001

[6] Chen, S.S., Donoho, D.L. and Saunders, M.A. (1998) Atomic Decomposition by Basis Pursuit. *SIAM Journal on Scientific Computing*, **20**, 33-61. http://dx.doi.org/10.1137/S1064827596304010

[7] Daubechies, I., Defrise, M. and Christine, D.M. (2004) An Iterative Thresholding Algorithm for Linear Inverse Problems with a Sparsity Constraint. *Communications on Pure and Applied Mathematics*, **57**, 1413-1457. http://dx.doi.org/10.1002/cpa.20042

[8] Candes, E.J., Wakin, M.B. and Boyd, S.P. (2008) Enhancing Sparity by Reweighted ℓ_1 Minimization. *Journal of Fourier Analysis and Applications*, **14**, 877-905. http://dx.doi.org/10.1007/s00041-008-9045-x

[9] Daubechies, I., Devore, R., Fornasier, M. and Güntük, C.S. (2010) Iteratively Reweighted Least Suqares Minimization for Sparse Recovery. *Communications on Pure and Applied Mathematics*, **63**, 1-38. http://dx.doi.org/10.1002/cpa.20303

[10] Zou, H. and Hastie, T. (2005) Regularization and Variable Selection via the Elastic Net. *Journal of the Royal Statistical Society*: *Series B*, **67**, 301-320. http://dx.doi.org/10.1111/j.1467-9868.2005.00503.x

[11] Tibshirani, R. (1996) Regression Shrinkage and Selection via The Lasso. *Journal of the Royal Statistical Society*: *Series B*, **58**, 267-288.

[12] Jin, B.T., Lorenz, D.A. and Schiffler, S. (2009) Elastic-Net Regularization: Error Estimates and Active Set Methods. *Inverse Problems*, **25**, Article ID: 115022. http://dx.doi.org/10.1088/0266-5611/25/11/115022

[13] De Mola, C., De Vitob, E. and Rosasco, L. (2009) Elastic-Net Regularization in Learning Theory. *Journal of Complexity*, **25**, 201-230. http://dx.doi.org/10.1016/j.jco.2009.01.002

[14] Lai, M.J., Xu, Y.Y. and Yin, W.T. (2013) Improved Iteratively Reweighted Least Squares for Unconstrained Smoothed ℓ_q Minimization. *SIAM Journal on Numerical Analysis*, **51**, 927-957. http://dx.doi.org/10.1137/110840364

[15] Donoho, D.L. and Tsaig, Y. (2008) Fast Solution of ℓ_1 Norm Minimization Problems When the Solution May Be Sparse. *IEEE Transactions on Information Theory*, **54**, 4789-4812. http://dx.doi.org/10.1109/TIT.2008.929958

[16] Garcia, C.B. and Li, T.Y. (1980) On the Number of Solutions to Polynomial Systems of Equations. *SIAM Journal on Numerical Analysis*, **17**, 540-546. http://dx.doi.org/10.1137/0717046