An Unconstrained ℓ_q Minimization with $0 < q \leq 1$ for Sparse Solution of Under-determined Linear Systems

Ming-Jun Lai *and Jingyue Wang
Department of Mathematics
The University of Georgia
Athens, GA 30602.

November 9, 2010

Abstract

We study an unconstrained version of the ℓ_q minimization for the sparse solution of under-determined linear systems for $0 < q \leq 1$. Although the minimization is non-convex when $q < 1$, we introduce a regularization and develop an iterative algorithm. We show that the iterative algorithm converges and the iterative solutions converge to the sparse solution under some additional assumptions on under-determined linear systems. Numerical experiments are presented to demonstrate the effectiveness of our approach.

Key Words and Phrases: Compressed Sensing, Sparse Solution, ℓ_q Minimization
Mathematics Subject Classification 2000: 15A09, 41A15, 65F10, 90C26
Short Title: An Unconstrained ℓ_q Minimization with $0 < q \leq 1$

1 Introduction

We are interested in computing the sparse solution of under-determined linear systems in the following sense: letting A be a matrix of size $m \times N$ with $m << N$ and $\mathbf{b}$ be a vector which is compressible, i.e., there exists a vector $\mathbf{x}^*$ with $\|\mathbf{x}^*\|_0 < m$ such that $\mathbf{b} = A\mathbf{x}^*$, we would like to find the solution of the following minimization

$$\min_{\mathbf{x} \in \mathbf{R}^N} \{\|\mathbf{x}\|_0, \quad A\mathbf{x} = \mathbf{b}\}, \tag{1}$$

*mjlai@math.uga.edu. This author is partly supported by the National Science Foundation under grant DMS-0713807

where $\|\mathbf{x}\|_0$ denotes the number of nonzero components of $\mathbf{x}$. The solution is called the sparse solution of $A\mathbf{x} = \mathbf{b}$. This is one of critical problems in compressed sensing research. This problem is motivated by data compression, error correction decodes, n-term approximation, and etc.. (See, e.g. [28]). It is known that problem (1) needs non-polynomial time to solve (cf. [30]). It is crucial to recognize that one natural approach to tackle (1) is to solve the following convex minimization problem:

$$\min_{\mathbf{x}\in\mathbf{R}^N} \{\|\mathbf{x}\|_1, \quad A\mathbf{x} = \mathbf{b}\}, \tag{2}$$

where $\|\mathbf{x}\|_1 = \sum_{j=1}^N |x_j|$ is the standard ℓ_1 norm. Note that it is equivalent to a linear programming problem which can be solved by the standard simplex method or interior point method.

The study of this problem (2) was pioneered by Donoho, Candès and their collaborators. Many researchers have made a lot of contributions related to the existence, uniqueness, and other properties of the sparse solution as well as computational algorithms and their convergence analysis to tackle problem (1). See survey papers in [1], [3], and [2].

To motivate our study, let us outline some research results related to numerical algorithms for the computation of sparse solutions of (1). First of all, the ℓ_1 minimization (2) by Candès and his collaborators (cf. [5]) is a successful approach to find sparse solutions of (1) if the sparsity $s = \|\mathbf{x}\|_0$ is not very large. A MATLAB program based on a linear programming method for the sparse solution is available on-line at the Candès webpage. The performance of the ℓ_1 method is further improved based on the ideas of repeating reweighted iteration (cf. [8]). Another approach is based on orthogonal greedy algorithm (OGA). See [35] and [36] for some theoretic study and [33] for an efficient numerical algorithm. The performance of the OGA in [33] is much improved based on the greedy ℓ_1 algorithm proposed recently in [27]. Another approach for the computation of the sparse solutions is based on ℓ_q minimization with $0 < q < 1$. That is, we consider the following

$$\min_{\mathbf{x}\in\mathbf{R}^N} \{\|\mathbf{x}\|_q^q, \quad A\mathbf{x} = \mathbf{b}\}, \tag{3}$$

where $\|\mathbf{x}\|_q^q = \sum_{j=1}^N |x_j|^q$ for $0 < q \le 1$. This minimization is motivated by the following fact:

$$\lim_{q\to 0_+} \|\mathbf{x}\|_q^q = \|\mathbf{x}\|_0.$$

This approach was initiated by [24] and many researchers have worked on this direction. Even though it is NP hard (cf. [23]), there are at least three advantages of using this approach to the best of the authors' knowledge. One is the result in [10]: for a Gaussian random matrix A, the restricted q-isometry property of order s holds if s is almost proportional to m when $q \to 0_+$. Another advantage demonstrated in [34], [9] and [20] is when $\delta_{2s} < 1$ (or $\delta_{2s+1} < 1$, $\delta_{2s+2} < 1$), the solution of the ℓ_q minimization is a sparse solution when $q > 0$ small enough, where δ_{2s} is the restricted isometry constants of matrix

A (similar for $\delta_{2s+1}, \delta_{2s+2}$). Note that for the ℓ_1 minimization, one needs $\delta_{2s} < \sqrt{2} - 1$ or $\delta_{2s} < 2/(3+\sqrt{2})$ or $4/(6+\sqrt{6})$ as shown in [4], [20] and [19]. The third advantage is that the ℓ_q minimization can be applied to a wider class of random matrices A, e.g., when A is a random matrix whose entries are iid copies of a pre-Gaussian random variable. See [21]. In addition, there are many other approaches, e.g., optimal basis pursuit(OMB) method (for problem (2)), soft-thresholding iterations, standard and damped Landweber iterations ([12]) for problem (2), iterative reweighted least squares (IRLS) method (cf. [14]) (for problems (2) and (3)) and etc..

In this paper we shall consider another version of ℓ_q minimization:

$$\min_{\mathbf{x} \in \mathbf{R}^N} \|\mathbf{x}\|_q^q + \frac{1}{2\lambda}\|A\mathbf{x} - \mathbf{b}\|_2^2, \tag{4}$$

which $\|\mathbf{x}\|_2^2 = \sum_{j=1}^N x_j^2$ and $\lambda > 0$ is a parameter which is sufficiently small, e.g., $\lambda = 10^{-8}$. See Theorem 2.7 for an upper bound for λ. Clearly, this is a standard unconstrained version of the original ℓ_q minimization (3). Due to the singularity of the gradient of the associated functional above because of the sparsity of the solution $\mathbf{x}$, we introduce the following regularized version of the unconstrained ℓ_q minimization:

$$\min_{\mathbf{x} \in \mathbf{R}^N} \|\mathbf{x}\|_{q,\epsilon} + \frac{1}{2\lambda}\|A\mathbf{x} - \mathbf{b}\|_2^2, \tag{5}$$

where

$$\|\mathbf{x}\|_{q,\epsilon} = \sum_{j=1}^N (\epsilon + x_j^2)^{q/2}$$

and $\epsilon > 0$ is another parameter which will go to zero in order to approximate $\|\mathbf{x}\|_q^q$. This is the main minimization problem we study in this paper. There are many unconstrained versions of problem (2) have been studied in the literature (cf. [37] and references therein). In addition, there are several studies on the unconstrained ℓ_q minimization (4), e.g., [11] and [32]. The researchers in their papers [32] and [11] used several formats to regularize (4). These regularized minimizations are different from the one in (5). They obtained several interesting results on the lower bound for nonzero entries in the local minimizers of (4). Their study inspires us to consider the lower bound for nonzero entries of our local minimizers. We find ϵ is a good indicator if an entry of local minimizer is zero or not zero.

Clearly, the above problem (5) has a solution for any $q \in (0, 1]$ and $\epsilon > 0$. We shall derive an iterative algorithm to compute a critical point $\mathbf{x}^{\epsilon,q}$ of (5). We prove that the iterative algorithm is convergent for any starting point. Although $\mathbf{x}^{\epsilon,q}$ is a critical point of the nonconvex minimization problem (5), we shall show that it is a global minimizer under some additional assumptions on matrix A, the existence of a sparse solution, and the sparsity of critical points. We shall introduce a concept called matrices of completely full rank and recall the standard notion of the restricted isometry property (RIP). See [25] for a verifiable sufficient condition for sparse solution. A matrix which is of completely full rank can be renormalized to be a matrix with a RIP. Under the assumption that $\mathbf{x}^{\epsilon,q}$ is a local minimizer for each ϵ and λ dependent on ϵ, we shall show that $\mathbf{x}^{\epsilon,q}, \epsilon \to 0_+$ has a convergent subsequence which converges to a minimizer $\mathbf{y}^q$ of the constrained ℓ_q minimization problem (3). Under the sparsity conditions which can be verified, we can show that the limit $\mathbf{y}^q$ is the sparse solution of (1). This convergence requires that λ be dependent on ϵ and $\lambda \to 0$ when $\epsilon \to 0$. Furthermore, we discuss the convergence of the minimizer $\mathbf{z}^q$ of unconstrained ℓ_q minimization problem (4) for sufficiently small λ, but $\lambda \not\to 0$. We shall use the notion of Γ-convergence to show that the minimizer $\mathbf{z}^q$ converges to solution of our original under-determined linear system (1) when $q \to 0_+$. These form our main results in this paper.

However, it is easy to see that there is a lot of computation for various ϵ, q and λ to perform. Namely we have to compute $\mathbf{x}^{\epsilon,q}$ for many different small values ϵ for a fixed q to get a limit $\mathbf{y}^q$ of a convergent subsequence. Each $\mathbf{x}^{\epsilon,q}$ is computed by using an iterative algorithm which is proved to be convergent. Then we look for a convergent subsequence of $\{\mathbf{y}^q, q > 0\}$. Thus, the computation is expensive. However, if we know the information about the sparsity, say s of the sparse solution of (1), we can determine immediately if $\mathbf{x}^{\epsilon,q}$ is already a solution or not by checking if $\|\mathbf{x}^{\epsilon,q}\|_0 = s$ or not and if $A\mathbf{x}^{\epsilon,q} - \mathbf{b} = 0$ or not. Thus the iteration can be quickly stopped. Suppose we do not know the sparsity. Letting $\mathbf{y}^*$ be the limit of a subsequence of $\{\mathbf{y}^q, q > 0\}$, if the sparsity of $\mathbf{y}^*$ is $\le m/2$, the chance of $\mathbf{y}^*$ to be the sparse solution is good as we shall explain it in the next section. Reduce ϵ and use $\mathbf{y}^*$ as an initial solution to solve the minimization problem (5) again. Otherwise, if the sparsity of $\mathbf{y}^*$ is $>> m/2$ or closed to m, then the chance of $\mathbf{y}^*$ to be a solution of the following minimization (6) is slim. It is better to start with a completely new initial guess and construct a new subsequence $\mathbf{y}^q$. This gives us reasons of the better performance of our algorithm than the other schemes in our numerical experiments. Note that when $q \to 0_+$, the unconstrained ℓ_q minimization converges to

$$\min_{\mathbf{x}\in\mathbf{R}^N} \|\mathbf{x}\|_0 + \frac{1}{2\lambda}\|A\mathbf{x} - \mathbf{b}\|_2^2, \tag{6}$$

which is the unconstrained version of our main problem (1). With λ small enough, the solution can be viewed as a good numerical approximation of the sparse solution. Extensive numerical experiments have been performed to compare with many other methods as explained above. Our unconstrained ℓ_q minimization does indeed perform much better. In particular, our approach performs the best for under-determined linear systems $A\mathbf{x} = \mathbf{b}$ with uniform random matrices A, i.e., the entries of A are random variables of uniform distribution.

The paper consists of three sections, in addition to this introductory section: our analysis of the unconstrained ℓ_q minimization when $q < 1$ in §2, some additional properties of the unconstrained ℓ_q minimization when $q = 1$ in §3, and finally in §4 numerical results to demonstrate how well our unconstrained ℓ_q minimization can find the sparse solutions.

2 Analysis of Unconstrained ℓ_q Minimization with $0 < q < 1$

We begin with some elementary properties of the minimization problem (5). Let $L_q(\epsilon, \mathbf{x})$ be the function associated with (5). It is easy to see that the problem has a solution. We use $\mathbf{x}^{\epsilon,q}$ to denote a critical point of (5).

Consider the following one variable function of α

$$L_q(\epsilon, \mathbf{x}^{\epsilon,q} + \alpha\mathbf{y}) = \|\mathbf{x}^{\epsilon,q} + \alpha\mathbf{y}\|_{q,\epsilon} + \frac{1}{2\lambda}\|A\mathbf{x}^{\epsilon,q} + \alpha A\mathbf{y} - \mathbf{b}\|^2.$$

A critical point $\mathbf{x}^{\epsilon,q}$ satisfies the following gradient equations

$$\left[\frac{qx_j^{\epsilon,q}}{(\epsilon + (x_j^{\epsilon,q})^2)^{1-q/2}}\right]_{1\le j\le N} + \frac{1}{\lambda}A^T(A\mathbf{x}^{\epsilon,q} - \mathbf{b}) = 0. \tag{7}$$

This is a necessary condition for minimizers. We now derive an iterative method to solve the above equations due to their nonlinearity. Starting with any initial $\mathbf{x}^{(1)}$, we solve the following system of linear equations for $\mathbf{x}^{(k+1)}$:

$$\left[\frac{qx_j^{(k+1)}}{(\epsilon + (x_j^{(k)})^2)^{1-q/2}}\right]_{1\le j\le N} + \frac{1}{\lambda}A^T(A\mathbf{x}^{(k+1)} - \mathbf{b}) = 0 \tag{8}$$

or

$$\left(A^TA + \text{ diag }\left[\frac{q\lambda}{(\epsilon + |x_j^{(k)}|^2)^{1-q/2}}, j = 1, \cdots, N\right]\right)\mathbf{x}^{(k+1)} = A^T\mathbf{b} \tag{9}$$

for $k = 1, 2, 3, \cdots,$. It is easy to see that the above linear system is invertible for any $\mathbf{x}^{(k)}$ as long as $\epsilon > 0$. Thus, the iterative method is well defined. We now show that these $\mathbf{x}^{(k)}$ converge to a critical point of the minimization problem (5). We begin with

Lemma 2.1 *Fix any $\epsilon > 0$. Let $\mathbf{x}^{(k+1)}$ be the solution of (9) for $k = 1, 2, 3, \cdots$. Then*

$$\|A\mathbf{x}^{(k)} - A\mathbf{x}^{(k+1)}\|^2 \le 2\lambda(L_q(\epsilon, \mathbf{x}^{(k)}) - L_q(\epsilon, \mathbf{x}^{(k+1)})). \tag{10}$$

Proof. Mainly we need the following inequality

$$(\epsilon + |x|^2)^{q/2} - (\epsilon + |y|^2)^{q/2} - \frac{qy(x-y)}{(\epsilon + |x|^2)^{1-q/2}} \ge 0, \tag{11}$$

where $0 < q < 1$. This inequality can be verified by a direct computation. We now compute

$$L_q(\epsilon, \mathbf{x}^{(k)}) - L_q(\epsilon, \mathbf{x}^{(k+1)}) \;=\; \sum_{j=1}^N(\epsilon + |x_j^{(k)}|^2)^{q/2} - \sum_{j=1}^N(\epsilon + |x_j^{(k+1)}|^2)^{q/2}$$

$$
\begin{aligned}
& +\frac{1}{2\lambda}\left(\|A\mathbf{x}^{(k)}-\mathbf{b}\|^2-\|A\mathbf{x}^{(k+1)}-\mathbf{b}\|^2\right) \\
= \; & \sum_{j=1}^{N}(\epsilon+|x_j^{(k)}|^2)^{q/2}-(\epsilon+|x_j^{(k+1)}|^2)^{q/2}+\frac{1}{2\lambda}\|A\mathbf{x}^{(k)}-A\mathbf{x}^{(k+1)}\|^2 \\
& +\frac{1}{\lambda}(A\mathbf{x}^{(k+1)}-\mathbf{b})^T(A\mathbf{x}^{(k)}-A\mathbf{x}^{(k+1)}).
\end{aligned}
$$

The last term can be simplified to be

$$
-\sum_{j=1}^{N}\frac{qx_j^{(k+1)}(x_j^{(k)}-x_j^{(k+1)})}{(\epsilon+|x_j^{(k)}|^2)^{1-q/2}}
$$

by using (8) based on a dot-product with $\mathbf{x}^{(k)}-\mathbf{x}^{(k+1)}$. With this we have

$$
\begin{aligned}
& L_q(\epsilon,\mathbf{x}^{(k)})-L_q(\epsilon,\mathbf{x}^{(k+1)}) \\
= \; & \sum_{j=1}^{N}\left((\epsilon+|\mathbf{x}_j^{(k)}|^2)^{q/2}-(\epsilon+|\mathbf{x}_j^{(k+1)}|^2)^{q/2}-\frac{x_j^{(k+1)}(x_j^{(k)}-x_j^{(k+1)})}{(\epsilon+|x_j^{(k)}|^2)^{1-q/2}}\right) \\
& +\frac{1}{2\lambda}\|A\mathbf{x}^{(k)}-A\mathbf{x}^{(k+1)}\|^2 \\
\geq \; & \frac{1}{2\lambda}\|A\mathbf{x}^{(k)}-A\mathbf{x}^{(k+1)}\|^2 \geq 0
\end{aligned}
$$

by using (11). The result (10) follows immediately. ■

We are now ready to prove the convergence of our iterative algorithm from any starting point.

Theorem 2.1 *Fix an $\epsilon > 0$ and parameter $\lambda > 0$. There exists $\mathbf{x}^* \in \mathbf{R}^n$ such that the iterative solutions $\mathbf{x}^{(k)}$ in (9) converges to $\mathbf{x}^*$, i.e.,*

$$
\lim_{k\to\infty}\mathbf{x}^{(k)}=\mathbf{x}^*
$$

and $\mathbf{x}^*$ is a critical point of the problem (5).

Proof. By Lemma 2.1, we have

$$
L_q(\epsilon,\mathbf{x}^{(k+1)}) \leq L_q(\epsilon,\mathbf{x}^{(k)}). \tag{12}
$$

That is, $L_q(\epsilon,\mathbf{x}^{(k)})$ is decreasing. Let $\lim_{k\to\infty} L_q(\epsilon,\mathbf{x}^{(k)}) = M$. It is clear that $\|\mathbf{x}^{(k)}\|_q$ is bounded due to the fact that

$$
\|\mathbf{x}^{(k)}\|_q \leq \|\mathbf{x}^{(k)}\|_{q,\epsilon} \leq L_q(\epsilon,\mathbf{x}^{(k+1)}) \leq M+1
$$

for k sufficiently large. Hence, there exist a vector $\mathbf{x} \in \mathbf{R}^n$ and a convergent subsequence $\mathbf{x}^{(k_j)}$ such that $\mathbf{x}^{(k_j)} \to \mathbf{x}$. Note that $\mathbf{x}^{(k_j+1)}$ solves (8). By (9), $\mathbf{x}^{(k_j+1)}$ is also a convergent subsequence. Let us say $\mathbf{x}^{(k_j+1)} \to \mathbf{y}$. Note that $L_q(\epsilon, \mathbf{x}) = M = L_q(\epsilon, \mathbf{y})$. Then by (10),

$$\|A\mathbf{x} - A\mathbf{y}\|^2 \leq 2\lambda(L_q(\epsilon, \mathbf{x}) - L_q(\epsilon, \mathbf{y})) = 0.$$

Thus we have $A\mathbf{y} = A\mathbf{x}$. Then it follows that $\|\mathbf{x}\|_{q,\epsilon} = \|\mathbf{y}\|_{q,\epsilon}$.

We now claim that $\mathbf{x} = \mathbf{y}$. Indeed, using dot product in (8) with $\mathbf{x} - \mathbf{y}$ for $k = k_j$ and letting $j \to \infty$, we have

$$\sum_{j=1}^N \frac{q y_j(x_j - y_j)}{(\epsilon + |x_j|^2)^{1-q/2}} + \frac{1}{\lambda}(A\mathbf{x} - A\mathbf{y})^T(A\mathbf{y} - \mathbf{b}) = 0$$

As we have proved that $A\mathbf{x} = A\mathbf{y}$ above, we have

$$\sum_{j=1}^N \frac{y_j(x_j - y_j)}{(\epsilon + |x_j|^2)^{1-q/2}} = 0.$$

Combining the above equation with the fact $\|\mathbf{x}\|_{q,\epsilon} = \|\mathbf{y}\|_{q,\epsilon}$ we just proved above, we end with

$$\|\mathbf{x}\|_{q,\epsilon} - \|\mathbf{y}\|_{q,\epsilon} - \sum_{j=1}^N \frac{q y_j(x_j - y_j)}{(\epsilon + |x_j|^2)^{1-q/2}} = 0.$$

In other words,

$$\sum_{j=1}^N \left((\epsilon + x_j^2)^{q/2} - (\epsilon + y_j^2)^{q/2} - \frac{q y_j(x_j - y_j)}{(\epsilon + x_j^2)^{1-q/2}} \right) = 0$$

By inequality (11), each summation term is nonnegative and hence has to be zero term by term. Furthermore, each term can be rewritten as

$$\begin{aligned} 0 &= (\epsilon + x_j^2)^{q/2} - (\epsilon + y_j^2)^{q/2} - \frac{y_j(x_j - y_j)}{(\epsilon + x_j^2)^{1-q/2}} \\ &= \frac{q x_j^2 - 2q x_j y_j + q y_j^2}{2(\epsilon + x_j^2)^{1-q/2}} + \frac{2\epsilon + (2-q)x_j^2 + q y_j^2 - 2(\epsilon + x_j^2)^{1-q/2}(\epsilon + y_j^2)^{q/2}}{2(\epsilon + x_j^2)^{1-q/2}}. \end{aligned}$$

Since both of the above two terms are nonnegative, it follows from the first term above that $x_j = y_j$ for all j and hence, $\mathbf{x} = \mathbf{y}$. Let us denote $\mathbf{x}$ by $\mathbf{x}^*$. Therefore, from (8) for $k = k_j$ with $j \to \infty$, we know $\mathbf{x}^*$ satisfies (7) with $\mathbf{x}^*$ in place of $\mathbf{x}^{\epsilon,q}$. Thus $\mathbf{x}^*$ is a critical point. ∎

We remark that $A\mathbf{x}^{\epsilon,q} \neq \mathbf{b}$ if $\mathbf{b} \neq 0$. Otherwise, we would have $\mathbf{x}^{\epsilon,q} = 0$ by (7) and hence, a contradiction $A\mathbf{x}^{\epsilon,q} = 0$. When $q = 1$, we can see that the functional $L_1(\epsilon, \mathbf{x})$

is strictly convex. The uniqueness of the minimizer which satisfies the gradient equation implies that the limit $\mathbf{x}^*$ of our iterative solution is $\mathbf{x}^{\epsilon,1}$, the minimizer of (5) with $q = 1$. Although we only proved that $\mathbf{x}^*$ is a critical point when $q < 1$, numerical experiments show that the $\mathbf{x}^*$ achieves the minimum for some ϵ small enough. Indeed, we can show the limit of iterative solution is the global minimizer for (5) under some additional assumptions on the sensing matrix A and sparsity of the limit solution.

To this end, let us introduce the concept of completely full rank. We say a matrix A of size $m \times n$ with $m < n$ is of *completely full rank* if any sub-matrices of A of size $m \times m$ are full rank. For example, $A = [(x_j)^{i-1}]_{1\le i\le m, 1\le j\le n}$ with x_j distinct is a matrix of completely full rank. For another example, let A with

$$A^T = [1, \cos(x_j), \sin(x_j), \cdots, \cos(mx_j), \sin(mx_j)]_{j=1,\cdots,n}$$

for all $x_j \in [0, 2\pi), j = 1, \cdots, n$ be a matrix of size $(2m+1) \times n$. Then A is of completely full rank since A is a Tchebysheff system (cf. [26]).

Lemma 2.2 *Suppose that A is of completely full rank. Let*

$$\mathcal{A} = \begin{bmatrix} A & 0_m \\ I_n & R_m \end{bmatrix}.$$

where 0_m is a zero block matrix of size $m \times m$, I_n is the identity matrix of size $n \times n$, and R_m is a zero matrix except for $R(r(i), i) = 1$ for $i = 1, \cdots, m$ with $r(1), \cdots, r(m)$ being the first m entries of a random permutation of integers $1, 2, \cdots, n$. Then $\mathcal{A}$ is invertible and $\|\mathcal{A}^{-1}\|_2$ is bounded above by a constant C which is dependent on A for any random permutation.

Proof. Without loss of generality, we may assume that

$$R_m = \begin{bmatrix} I_m \\ 0_{n-m,m} \end{bmatrix}$$

with $0_{n-m,m}$ being a zero block matrix of size $(n-m) \times m$. Since A is completely full rank, we use the rows from $m+1$ to $2m$ of $\mathcal{A}$ to make $\mathcal{A}(1 : m, 1 : m)$ to be zero. Then we use the rows $2m+1$ to $m+n$ to make $\mathcal{A}(1 : m, m+1 : n)$ to zero. Note that $\mathcal{A}(1 : m, n+1 : n+m)$ is $-A(1 : m, 1 : m)$. Clearly, the resulting matrix of $\mathcal{A}$ is invertible and the norm of the inverse of the resulting matrix is dependent on the norm of the inverse of $A(1 : m, 1 : m)$. Similar for other random matrix R_m. ■

Lemma 2.3 *Suppose that A is of completely full rank. If $\mathbf{x}^{\epsilon,q} \in \mathbf{R}^n$ has a sparsity $\|\mathbf{x}^{\epsilon,q}\|_0 \le m/2$. Then for any $\mathbf{y} \in R^n$ with sparsity $\|\mathbf{y}\|_0 \le m/2$,*

$$\|\mathbf{x}^{\epsilon,q} - \mathbf{y}\| \le C\|A\mathbf{x}^{\epsilon,q} - A\mathbf{y}\|, \tag{13}$$

where C is a constant that depends on A.

Proof. Consider $\mathbf{x}^{\epsilon,q}-\mathbf{y}$ whose sparsity is at most m. Without loss of generality, we may assume that the first m entries of $\mathbf{x}^{\epsilon,q}-\mathbf{y}$ are nonzero. Let $\mathbf{z}$ be a vector of size $(n+m)\times 1$ whose first m entries are the first m entries of $\mathbf{x}^{\epsilon,q}-\mathbf{y}$, whose entries with indices $n+1,\cdots,n+m$ are the negative of the first m components of $\mathbf{x}^{\epsilon,q}-\mathbf{y}$ and whose remaining entries are zero. It is easy to see that

$$\mathcal{A}\mathbf{z}=\left[\begin{array}{c} A\mathbf{x}^{\epsilon,q}-A\mathbf{y} \\ 0_{m,1}\end{array}\right].$$

Then

$$\|\mathbf{z}\|_2^2=2\|\mathbf{x}^{\epsilon,q}-\mathbf{y}\|_2^2\le\|\mathcal{A}^{-1}\|_2^2\|A\mathbf{x}^{\epsilon,q}-A\mathbf{y}\|_2^2.$$

This completes the proof. ∎

We are now ready to prove that all critical points will be the global minimizers under the sparsity assumptions described in the following theorem.

Theorem 2.2 *Suppose that A is of completely full rank. For any ϵ,q, there is a $\lambda_{\epsilon,q}$ dependent on ϵ,q so that for any $\lambda<\lambda_{\epsilon,q}$, if $\mathbf{x}^{\epsilon,q}$ is a critical point for problem (5) and if the sparsity $\|\mathbf{x}^{\epsilon,q}\|_0\le m/2$, then for any $\mathbf{y}$ with sparsity $\|\mathbf{y}\|_0\le m/2$,*

$$\|\mathbf{y}-\mathbf{x}^{\epsilon,q}\|^2\le C\lambda(L_q(\epsilon,\mathbf{y})-L_q(\epsilon,\mathbf{x}^{\epsilon,q})) \tag{14}$$

for a positive constant $C>0$.

Proof. We first calculate $L_q(\epsilon,\mathbf{y})-L_q(\epsilon,\mathbf{x}^{\epsilon,q})$ and show for an appropriately chosen λ, (14) is true.

$$\begin{aligned}
L_q(\epsilon,\mathbf{y})-L_q(\epsilon,\mathbf{x}^{\epsilon,q}) &= \sum_{j=1}^{N}(\epsilon+|\mathbf{y}_j|^2)^{q/2}-\sum_{j=1}^{N}(\epsilon+|\mathbf{x}_j^{\epsilon,q}|^2)^{q/2} \\
&\quad +\frac{1}{2\lambda}\left(\|A\mathbf{y}-\mathbf{b}\|^2-\|A\mathbf{x}^{\epsilon,q}-\mathbf{b}\|^2\right) \\
&= \sum_{j=1}^{N}(\epsilon+|\mathbf{y}_j|^2)^{q/2}-(\epsilon+|\mathbf{x}_j^{\epsilon,q}|^2)^{q/2} \\
&\quad +\frac{1}{2\lambda}\|A\mathbf{y}-A\mathbf{x}^{\epsilon,q}\|^2+\frac{1}{\lambda}(A\mathbf{x}^{\epsilon,q}-\mathbf{b})^T(A\mathbf{y}-A\mathbf{x}^{\epsilon,q}).
\end{aligned}$$

The last term can be simplified to be

$$-\sum_{j=1}^{N}\frac{q\mathbf{x}_j^{\epsilon,q}(\mathbf{y}_j-\mathbf{x}_j^{\epsilon,q})}{(\epsilon+(\mathbf{x}_j^{\epsilon,q})^2)^{1-q/2}}$$

by using the Euler-Lagrange equation of (5). With this we have

$$L_q(\epsilon,\mathbf{y})-L_q(\epsilon,\mathbf{x}^{\epsilon,q})$$

$$
\begin{aligned}
&= \sum_{j=1}^{N}\left((\epsilon+\mathbf{y}_j^2)^{q/2}-(\epsilon+(\mathbf{x}_j^{\epsilon,q})^2)^{q/2}-\frac{q\mathbf{x}_j^{\epsilon,q}(\mathbf{y}_j-\mathbf{x}_j^{\epsilon,q})}{(\epsilon+(\mathbf{x}_j^{\epsilon,q})^2)^{1-q/2}}\right)\\
&\quad +\frac{1}{2\lambda}\|A\mathbf{y}-A\mathbf{x}^{\epsilon,q}\|^2\\
&\geq \sum_{j=1}^{N}\left((\epsilon+\mathbf{y}_j^2)^{q/2}-(\epsilon+(\mathbf{x}_j^{\epsilon,q})^2)^{q/2}-\frac{q\mathbf{x}_j^{\epsilon,q}(\mathbf{y}_j-\mathbf{x}_j^{\epsilon,q})}{(\epsilon+(\mathbf{x}_j^{\epsilon,q})^2)^{1-q/2}}\right)\\
&\quad +\frac{C}{2\lambda}\|\mathbf{y}-\mathbf{x}^{\epsilon,q}\|^2 \qquad \text{(by Lemma 2.3)}\\
&= \sum_{j=1}^{N}\left((\epsilon+\mathbf{y}_j^2)^{q/2}-(\epsilon+(\mathbf{x}_j^{\epsilon,q})^2)^{q/2}-\frac{q\mathbf{x}_j^{\epsilon,q}(\mathbf{y}_j-\mathbf{x}_j^{\epsilon,q})}{(\epsilon+(\mathbf{x}_j^{\epsilon,q})^2)^{1-q/2}}+\frac{C}{2\lambda}(\mathbf{y}_j-\mathbf{x}_j^{\epsilon,q})^2\right)
\end{aligned}
$$

We show that if λ is small enough, each term in the sum is nonnegative. It is reduced to prove that for a fixed a, function

$$f(x)=(\epsilon+x^2)^{q/2}-(\epsilon+a^2)^{q/2}-\frac{qa(x-a)}{(\epsilon+a^2)^{1-q/2}}+\frac{C}{2\lambda}(x-a)^2$$

has a global minimum at a. Simple calculus shows

$$f''(x)=\frac{C}{\lambda}+\frac{q(\epsilon-(1-q)x^2)}{(\epsilon+x^2)^{2-q/2}}.$$

The second term is smooth and tends to zero as x goes to infinity, thus is bounded. The bound depends on ϵ and q. Therefore there exists a $\lambda_{\epsilon,q}>0$ so that if $\lambda<\lambda_{\epsilon,q}$, $f''(x)>0$, $f(x)$ is convex. It's easy to verify the only minimum occurs at $x=a$ where $f(a)=0$ and $f(x)\geq 0$ for any x. We choose $\lambda=\lambda_{\epsilon,q}/2$. Thus

$$
\begin{aligned}
&L_q(\epsilon,\mathbf{y})-L_q(\epsilon,\mathbf{x}^{\epsilon,q})\\
&= \sum_{j=1}^{N}\left((\epsilon+\mathbf{y}_j^2)^{q/2}-(\epsilon+(\mathbf{x}_j^{\epsilon,q})^2)^{q/2}-\frac{q\mathbf{x}_j^{\epsilon,q}(\mathbf{y}_j-\mathbf{x}_j^{\epsilon,q})}{(\epsilon+(\mathbf{x}_j^{\epsilon,q})^2)^{1-q/2}}+\frac{C}{2\lambda_{\epsilon,q}}(\mathbf{y}_j-\mathbf{x}_j^{\epsilon,q})^2\right)\\
&\qquad +\sum_{j=1}^{N}\left(\frac{C}{2\lambda_{\epsilon,q}}(\mathbf{y}_j-\mathbf{x}_j^{\epsilon,q})^2\right)\\
&\geq \sum_{j=1}^{N}\left(\frac{C}{2\lambda_{\epsilon,q}}(\mathbf{y}_j-\mathbf{x}_j^{\epsilon,q})^2\right).
\end{aligned}
$$

Note that C depends on A, the result follows. ■

If $\mathbf{y}$ is a global minimizer of (5) which is sparse in the sense of $\|\mathbf{y}\|_0\leq m/2$, then (14) would be negative unless $\mathbf{x}^{\epsilon,q}=\mathbf{y}$. That is, the limit $\mathbf{x}^{\epsilon,q}$ of our iterative solution is a global minimizer.

The discussion above can be generalized. To this end, we need the concept of restricted isometry property (RIP). We say matrix A possesses the s^{th} RIP if there exists a nonzero number $\delta_s < 1$ such that

$$(1-\delta_s)\|\mathbf{x}_T\|_2^2 \leq \|A\mathbf{x}_T\|_2^2 \leq (1+\delta_s)\|\mathbf{x}_T\|_2^2, \tag{15}$$

where $\mathbf{x}_T$ is a vector in $\mathbf{R}^n$ whose nonzero entries are those with indices in T for all $T \subset \{1, 2, \cdots, n\}$ with $\#(T) \leq s$. The concept was introduced in [6] and [7] which generated a great deal of interest. Many random matrices such as Gaussian, sub-Gaussian, and pre-Gaussian random matrices are shown to have RIP with overwhelming probability. See [6], [10] and [21] and the references therein. For a non-random matrix to have a RIP, we can show

Theorem 2.3 *Any matrix A of completely full rank can be normalized to have the m^{th} RIP. On the other hand, if A has the m^{th} RIP, then A is of completely full rank.*

Proof. Let $I = \{T \subset \{1, \cdots, n\}, \#(T) = m\}$ be the collection of all index sets each of which is a subset of $\{1, \cdots, n\}$ with cardinality m, i.e., $T = \{i_1, \cdots, i_m\}$ with $1 \leq i_1 < i_2 < \cdots < i_m \leq n$ for all choices of $i_1, \cdots, i_m$. Let α_T and β_T be the smallest and largest singular values of A_T which consists of columns with indices in T. Since A is of completely full rank, there exist α and β such that $\alpha_T \geq \alpha > 0$ and $\beta_T \leq \beta < \infty$ for all $T \in I$. If we normalize A by $\hat{A} := A/\beta$, then we have

$$\frac{\alpha^2}{\beta^2}\|\mathbf{x}_T\|_2^2 \leq \|\hat{A}_T\mathbf{x}_T\|_2^2 \leq \|\mathbf{x}\|_2^2, \quad \forall T \in I.$$

By letting $1-\delta_m = \dfrac{\alpha^2}{\beta^2}$, we know that $\hat{A}$ possesses m^{th} RIP.

When A has the m^{th} RIP, i.e. $\delta_m < 1$, then the definition $\|A_T\mathbf{x}_T\|^2 \geq (1-\delta_m)\|\mathbf{x}_T\|^2$ for all T with $\#(T) = m$ implies that A_T is of full rank for all T with $\#(T) = m$. That is, A is of completely full rank. ■

If A is of completely full rank, so is αA for any scalar α. This gives us a method to check if a matrix A has the m^{th} RIP or can be scaled to have the m^{th} RIP. The above Theorem 2.3 provides a method to construct a matrix A with the s^{th} RIP with $0 < s < m$. Indeed, take a matrix $\tilde{A}$ of size $s \times n$ which is of completely full rank and add $m-s$ rows of any real numbers of length n to form a matrix B of size $m \times n$. Then by using the method in the proof above, B can be rescaled to A so that A possesses the s^{th} RIP.

With the concept of RIP, we are able to prove a similar result to Theorem 2.2.

Theorem 2.4 *Suppose that A possesses the s^{th} RIP. There exists a small enough λ depending on ϵ, q, and δ_s so that for any global minimizer $\mathbf{y}^q$ and any critical point $\mathbf{x}^{\epsilon,q}$ of problem (5), if the sparsity $\|\mathbf{y}^q - \mathbf{x}^{\epsilon,q}\|_0 \leq s$, then $\mathbf{x}^{\epsilon,q}$ is a minimizer of problem (5).*

Proof. Note that we also have the following inequality as in (14) for this case.

$$\|\mathbf{y} - \mathbf{x}^{\epsilon,q}\|^2 \leq C\lambda_{\epsilon,q}(L_q(\epsilon, \mathbf{y}) - L_q(\epsilon, \mathbf{x}^{\epsilon,q})). \tag{16}$$

The proof of (16) is almost identical to the proof of Theorem 2.2. The only adaption is now we use the s^{th} RIP property to bound $\|A\mathbf{y}^q - A\mathbf{x}^{\epsilon,q}\|^2$ from below:

$$\|A\mathbf{y}^q - A\mathbf{x}^{\epsilon,q}\|^2 \geq (1 - \delta_s)\|\mathbf{y}^q - \mathbf{x}^{\epsilon,q}\|^2$$

and the rest of the proof keeps the same. Thus since the right side of (16) is also non-positive, we have $L_q(\epsilon, \mathbf{y}^q) = L_q(\epsilon, \mathbf{x}^{\epsilon,q})$ and $\mathbf{x}^{\epsilon,q}$ is a minimizer of problem (5). ■

These explain that if a global minimizer of (5) is sparse and a critical point is sparse, then the critical point is a global minimizer. That is, the sparse minimizer is unique.

Now letting ϵ and correspondingly λ tend 0, we show that these minimizers for all $\epsilon, \lambda > 0$ with $\epsilon, \lambda \to 0$ converge to a solution of the constrained ℓ_q minimization problem (3).

Theorem 2.5 *Fix $q > 0$. Let $\mathbf{x}^k$ be a critical point of (5) associated with ϵ_k, λ_k for $k = 1, 2, 3, \cdots$ with ϵ_k, λ_k decreasing to 0, then there exists a subsequence from $\mathbf{x}^k$ which converges to $\mathbf{y}^q$ and $\mathbf{y}^q$ is a minimizer of (3).*

Proof. We first show that $\mathbf{x}^k$ is bounded in $\mathbf{R}^n$. In fact for a fixed $\mathbf{z}$ satisfying $A\mathbf{z} = \mathbf{b}$

$$\|\mathbf{x}^k\|_q^q \leq \sum_{j=1}^{N}(\epsilon_k + (\mathbf{x}_j^k)^2)^{q/2} \leq L_q(\epsilon_k, \mathbf{x}^k) \leq L_q(\epsilon_k, \mathbf{z}) \leq \|\mathbf{z}\|_q^q + N\epsilon_k^{q/2}. \tag{17}$$

Therefore there exists a subsequence that converges to $\mathbf{y}^q$ in $\mathbf{R}^n$. For simplicity we assume $\mathbf{x}^k$ converges to $\mathbf{y}^q$ in $\mathbf{R}^n$. Taking $k \to \infty$ in (17), we obtain

$$\|\mathbf{y}^q\|_q^q \leq \|\mathbf{z}\|_q^q. \tag{18}$$

Meanwhile since $L_q(\epsilon_k, \mathbf{x}^k) \leq \|\mathbf{z}\|_q^q + N\epsilon_k^{q/2} \leq M$ for some $M > 0$ independent of k, we have

$$\frac{1}{2\lambda_k}\|A\mathbf{x}^k - \mathbf{b}\|^2 \leq M \text{ or } \|A\mathbf{x}^k - \mathbf{b}\|^2 \leq 2\lambda_k M \tag{19}$$

By passing to the limit in (19), we have

$$A\mathbf{y}^q = \mathbf{b} \tag{20}$$

Combining (18) and (20), we conclude that $\mathbf{y}^q$ is a minimizer of problem (3). ■

We use the following theorem to help us find the solution of the constrained ℓ_q problem (3).

Theorem 2.6 *Suppose $q \in (0,1)$ is a fixed positive number and A possesses the m^{th} RIP. Assume $\mathbf{y}^q$ is the solution of problem (3) with sparsity $\|\mathbf{y}^q\|_0 \le m/2$. Let $\mathbf{x}^k$ be a critical point of (5) associated with ϵ_k, λ_k for $k = 1, 2, 3, \cdots$ which are decreasing to zero. If the sparsity $\|\mathbf{x}^k\|_0 \le m/2$ for k sufficiently large, then there exists a subsequence from $\mathbf{x}^k$ which converges to $\mathbf{y}^q$.*

Proof. Let $\mathbf{z}^k$ be the solution of problem (5). We truncate $\mathbf{z}^k$ by setting its $N - [m/2]$ smallest entries to zero and denote the resulting vector by $\widetilde{\mathbf{z}}^k$. By one of the assumptions, $\mathbf{y}^q$ has a sparsity no greater than $m/2$ and by Theorem 2.5, $\{\mathbf{z}^k\}$ has a subsequence that tend to $\mathbf{y}^q$ as k tends to infinity. For simplicity we assume $\lim_{k\to\infty} \mathbf{z}^k = \mathbf{y}^q$. One concludes that $\widetilde{\mathbf{z}}^k$ also tends to $\mathbf{y}^q$ as k tends to infinity and

$$\lim_{k\to\infty} A\widetilde{\mathbf{z}}^k = \lim_{k\to\infty} A\mathbf{z}^k = b \tag{21}$$

Applying Theorem 2.2 with $\mathbf{y} = \widetilde{\mathbf{z}}^k$, $\mathbf{x}^{\epsilon,q} = \mathbf{x}^k$, we have

$$\begin{aligned}\|\widetilde{\mathbf{z}}^k - \mathbf{x}^k\|^2 &\le C\lambda_k(L_q(\epsilon_k, \widetilde{\mathbf{z}}^k) - L_q(\epsilon_k, \mathbf{x}^k)) \\ &\le C\lambda_k(L_q(\epsilon_k, \mathbf{z}^k) - L_q(\epsilon_k, \mathbf{x}^k)) + C\lambda_k(L_q(\epsilon_k, \widetilde{\mathbf{z}}^k) - L_q(\epsilon_k, \mathbf{z}^k)).\end{aligned}$$

By definition, the first term is not greater than zero and hence,

$$\begin{aligned}\|\widetilde{\mathbf{z}}^k - \mathbf{x}^k\|^2 &\le C\lambda_k(L_q(\epsilon_k, \widetilde{\mathbf{z}}^k) - L_q(\epsilon_k, \mathbf{z}^k)) \\ &= C\lambda_k(\|\widetilde{\mathbf{z}}^k\|_{\epsilon_k,q} - \|\mathbf{z}^k\|_{\epsilon_k,q}) + (\|A\widetilde{\mathbf{z}}^k - b\|^2 - \|A\mathbf{z}^k - b\|^2).\end{aligned}$$

It is straightforward to see the right side tends to zero as k tends to infinity. Hence,

$$\lim_{k\to\infty} \mathbf{x}^k = \lim_{k\to\infty} \widetilde{\mathbf{z}}^k = \mathbf{y}^q.$$

That is, critical points $\mathbf{x}^k$ converge to the solution of the unconstrained minimization (3). ■

We now use Theorem 2.6 to explain our computational procedure. According to [34] (or [20], [9]), the minimizer $\mathbf{y}^q$ of the constrained ℓ_q minimization (3) is the sparse solution if $q > 0$ is small enough when the sparsity of the sparse solution is $\le m/2$. Let us assume such q. In this case, $\|\mathbf{y}^q\|_0 \le m/2$. Let $\{\mathbf{x}^k, k \ge 1\}$ be the sequence of critical points in Theorem 2.6. By the discussion above we use the ϵ to truncate the limit of our iterative solution $\mathbf{x}^k$. If the number of nonzero entries of the truncated $\mathbf{x}^k$ is bigger than $m/2$, this $\mathbf{x}^k$ is not good. A new initial guess is necessary to generate another iterative sequence to have a new limit $\mathbf{x}^k$. If the number of nonzero entries of the truncated $\mathbf{x}^k$ is less than or equal to $m/2$, we keep this limit vector and continue the procedure to find a good $\mathbf{x}^{k+1}$ with $\epsilon_{k+1} < \epsilon_k$. We have several methods to generate a new initial guess. If we exhaust all initial guess generators, we deem the our unconstrained ℓ_q approach fails. These justify our numerical approach to be given in the last section.

In the above, we have discussed the minimizers of problem (5) with $\lambda \to 0$. We now devote the remaining part of this section to the study of the minimizer of unconstrained ℓ_q minimization (4) without letting $\lambda \to 0$, but keeping λ small enough (we will specify how small is). Let $\mathbf{z}^q$ be a minimizer of (4) for each $q \in (0,1)$. We show that $\mathbf{z}^q$ converges to the sparse solution of the original problem (1) as $q \to 0_+$. We shall use the concept of Γ-convergence which was introduced by E. De Giorgi and T. Franzoni in 1975 (cf. [22]). We first give the definition for the Γ-convergence.

Definition 2.1 *Let (X, d) be a metric space with metric d. We say that a sequence of functionals $E_k : X \to [-\infty, \infty]$ is Γ-convergent to a functional $E : X \to [-\infty, \infty]$ as $k \to \infty$ if for all $u \in X$ we have*

(i) for every sequence $\{u_k \in X\}$ converging to u

$$E(u) \le \liminf_k E_k(u_k)$$

(ii) there exists a sequence $\{u_k \in X\}$ converging to u such that

$$E(u) \ge \limsup_k E_k(u_k),$$

or equivalently

$$E(u) = \lim_k E_k(u_k).$$

Next we prove that if the minimizers of E_k have a cluster point, it is a minimizer of E under the assumption of the Γ-convergence of E_k to E. We start with the following

Lemma 2.4 *If a sequence of functionals E_k is $\Gamma-$convergent to a functional E on X as $k \to \infty$, for any subsequence $\{E_{k_j}\}$ of $\{E_k\}$,*

$$\limsup_{k_j \to \infty} \inf_{u \in X} E_{k_j}(u) \le \inf_{v \in X} E(v).$$

Proof. For any vector $v \in X$, by the definition of $\Gamma-$convergence, there exists $\{u_k\}$ converging to v such that,

$$\limsup_{k \to \infty} E_k(u_k) \le E(v).$$

Note that $\inf_{u \in X} E_{k_j}(u) \le E_{k_j}(u_{k_j})$,

$$\begin{aligned}\limsup_{k_j \to \infty} \inf_{u \in X} E_{k_j}(u) &\le \limsup_{k_j \to \infty} E_{k_j}(u_{k_j}) \\ &\le \limsup_{k \to \infty} E_k(u_k) \\ &\le E(v).\end{aligned}$$

Since v is arbitrarily chosen, we have

$$\limsup_{k_j\to\infty} \inf_{u\in X} E_{k_j}(u) \le \inf_{v\in X} E(v).$$

■

One important consequence of a Γ-convergent sequence of functionals is the following standard result (cf. [29])

Lemma 2.5 *Suppose that a sequence of functionals E_k is $\Gamma-$convergent to a functional E on X as $k \to \infty$. Letting E_{k_j} be a subsequence and u_{k_j} be the minimizer of E_{k_j}, if u_{k_j} converges to u in X, then u is a minimizer of E.*

Proof. By the definition of $\Gamma-$convergence,

$$\begin{aligned} E(u) &\le \liminf_{k_j\to\infty} E_{k_j}(u_{k_j}) \\ &\le \limsup_{k_j\to\infty} E_{k_j}(u_{k_j}) \\ &= \limsup_{k_j\to\infty} \inf_{v\in X} E_{k_j}(v) \\ &\le \inf_{v\in X} E(v). \end{aligned}$$

The first line follows from the definition of $\Gamma-$convergence and the last line is the result of Lemma 2.4. ■

More details on Γ convergence can be found in [29]. The preparatory results above are enough for our current purpose.

Consider $L_q(0,\mathbf{x}), q \in (0,1)$ to be a sequence of functionals. Let $L_0(0,\mathbf{x})$ be another functional associated with the minimization in (6). We claim that $L_q(0,\mathbf{x}), q \in (0,1)$ are Γ-convergent to $L_0(0,\mathbf{x})$. Indeed, for any sequence $\mathbf{x}^q \in \mathbf{R}^n, q \in (0,1)$ which converges to $\mathbf{x}$ as $q \to 0_+$, we can see $\|\mathbf{b}-A\mathbf{x}^q\|_2^2$ converge to $\|\mathbf{b}-A\mathbf{x}\|_2^2$ easily. Writing $\mathbf{x} = (x_1, \cdots, x_n)^T$, let $\delta = \min\{|x_i| > 0\}$. Since $\mathbf{x}^q \to \mathbf{x}$, we have

$$\begin{aligned} L_q(0,\mathbf{x}^q) &\ge \sum_{|x_i|>0} |x^{q,i}|^q + \frac{1}{2\lambda}\|\mathbf{b} - A\mathbf{x}^q\|_2^2 \\ &\ge \sum_{|x_i|>0} |\frac{1}{2}\delta|^q + \frac{1}{2\lambda}\|\mathbf{b} - A\mathbf{x}^q\|_2^2 \end{aligned}$$

for q sufficiently small. It follows that

$$\liminf_{q\to 0_+} L_q(0,\mathbf{x}^q) \ge \sum_{|x_i|>0} 1 + \frac{1}{2\lambda}\|\mathbf{b} - A\mathbf{x}\|_2^2 = L_0(0,\mathbf{x}).$$

On the other hand, for any $\mathbf{x}$, we choose a particular sequence $\mathbf{x}^q = \mathbf{x}$ for all $q \in (0,1)$. Then we have

$$\limsup_{q\to 0} L_q(0,\mathbf{x}^q) = L_0(0,\mathbf{x}).$$

These show that $L_q(0,\mathbf{x}), q \in (0,1]$ are Γ-convergent to $L_0(0,\mathbf{x})$.

Assuming that $\widehat{\mathbf{x}}$ is the sparse solution of our original problem (1) with sparsity $s = \|\widehat{\mathbf{x}}\|_0$, let

$$D = \min_{\|\mathbf{x}\|_0 \le s-1} \|A\mathbf{x} - \mathbf{b}\|_2^2. \tag{22}$$

It is easy to see that $D > 0$. Then by Theorem 2.5, we conclude the following

Theorem 2.7 *Let $\mathbf{z}^q$ be a minimizer of the unconstrained ℓ_q minimization problem (4) with λ satisfying $\dfrac{D}{\lambda} > 2s$, where s is the sparsity of our main problem (1). Then $\{\mathbf{z}^q, q \in (0,1)\}$ contains at least one convergent subsequence and the limit of any subsequence from $\{\mathbf{z}^q, q \in (0,1)\}$ is a sparse solution of (1).*

Proof. Since $\mathbf{z}^q, q \in (0,1)$ are bounded, $\mathbf{z}^q$ contains a convergent subsequence. We use Lemma 2.5 to conclude that the limit, say $\mathbf{x}^0$ of the convergent subsequence is a minimizer of $L_0(0,\mathbf{x})$. Since the under-determined linear system has a sparse solution $\widehat{\mathbf{x}}$, $L_0(0,\widehat{\mathbf{x}}) = \|\widehat{\mathbf{x}}\|_0$ which is the minimal value for any λ.

Since $\mathbf{x}^0$ is a minimizer of $L_0(0,\mathbf{x})$, we have $L_0(0,\mathbf{x}^0) = \|\widehat{\mathbf{x}}\|_0$ by Lemma 2.5. That is, $\mathbf{x}^0$ has to be a vector such that $A\mathbf{x}^0 = \mathbf{b}$ and $\|\mathbf{x}^0\|_0 = \|\widehat{\mathbf{x}}\|_0$. Otherwise, if $A\mathbf{x}^0 \ne \mathbf{b}$, then $\|\mathbf{x}^0\|_0 \le s - 1$ and

$$L_0(0,\mathbf{x}^0) = \|\mathbf{x}^0\|_0 + \frac{1}{2\lambda}\|A\mathbf{x}^0 - \mathbf{b}\|_2^2 \ge \|\mathbf{x}^0\|_0 + \frac{1}{2\lambda}D > 1 + s$$

which contradicts to the fact $L_0(0,\mathbf{x}^0) = \|\widehat{\mathbf{x}}\|_0 = s$. This completes the proof. ■

As the sparse solution $\widehat{\mathbf{x}}$ may not be unique, the sequence $\mathbf{z}^q, q \in (0,1)$ does not converge in general. The result in Theorem 2.7 above shows that the limit of any subsequence is a sparse solution of (1).

3 Some Additional Properties of Unconstrained ℓ_1 Minimization

In this section we exhibit more properties of the unconstrained ℓ_q minimization when $q = 1$. We have the following stability property of the unconstrained ℓ_1 minimization.

Proposition 3.1 (Stability) *Suppose that $q = 1$. Let $\mathbf{x_b}$ be a minimizer for input data $\mathbf{b}$ in problem (5). Similarly, for an input data $\mathbf{c}$, let $\mathbf{x_c}$ be a minimizer of (5) with $\mathbf{b}$ replaced by $\mathbf{c}$. Then*

$$\|A\mathbf{x_b} - A\mathbf{x_c}\|_2 \le \|\mathbf{b} - \mathbf{c}\|_2 .$$

In particular, the above property holds for a minimizer $\mathbf{x_b}$ of (2) and a minimizer $\mathbf{x_c}$ of the minimization problem (2) with $\mathbf{b}$ replaced by $\mathbf{c}$. In addition, if A is completely full rank and $\|\mathbf{x_b}\|_0 \le m/2$ as well $\|\mathbf{x_c}\|_0 \le m/2$, then there exists a positive constant C d

$$\|\mathbf{x_c} - \mathbf{x_b}\| \le C\|\mathbf{b} - \mathbf{c}\|_2.$$

Proof. We mainly use the following inequality:

$$\left(\frac{x}{\sqrt{\epsilon+|x|^2}}-\frac{y}{\sqrt{\epsilon+|y|^2}}\right)(x-y)\geq 0. \tag{23}$$

which can be verified easily. Fix an $\epsilon>0$. Let $\mathbf{x_b}$ be the minimizer satisfying the equation (7) with $q=1$ associated with $\mathbf{b}$. Similarly, let $\mathbf{x_c}$ satisfy (7) associated with $\mathbf{c}$ replacing $\mathbf{b}$. For convenience, let us write $\mathbf{x}:=\mathbf{x_b}$ and $\mathbf{y}:=\mathbf{x_c}$. Multiplying $\mathbf{x}-\mathbf{y}$ to both sides of (7), we have

$$\sum_{j=1}^N \frac{x_j(x_j-y_j)}{\sqrt{\epsilon+(x_j)^2}}+\frac{1}{\lambda}(A(\mathbf{x}-\mathbf{y}))^T(A\mathbf{x}-\mathbf{b})=0. \tag{24}$$

Similarly, we have

$$\sum_{j=1}^N \frac{y_j(x_j-y_j)}{\sqrt{\epsilon+|x_j|^2}}+\frac{1}{\lambda}(A(\mathbf{x}-\mathbf{y}))^T(A\mathbf{y}-\mathbf{c})=0. \tag{25}$$

The subtractions of the above equations yields

$$\frac{1}{\lambda}(A(\mathbf{x}-\mathbf{y}))^T(A\mathbf{x}-A\mathbf{y}-\mathbf{b}+\mathbf{c})=-\sum_{j=1}^N\left(\frac{x_j}{\sqrt{\epsilon+|x_j|^2}}-\frac{y_j}{\sqrt{\epsilon+|y_j|^2}}\right)(x_j-y_j)$$

which is less than or equal to zero by (23). It follows that

$$(A\mathbf{x}-A\mathbf{y})^T(A\mathbf{x}-A\mathbf{y}-\mathbf{b}+\mathbf{c})\leq 0$$

or

$$\|A\mathbf{x}-A\mathbf{y}\|_2^2\leq(A\mathbf{x}-A\mathbf{c})^T(\mathbf{b}-\mathbf{c}).$$

An application of Cauchy-Schwarz inequality and let $\epsilon\to 0_+$ together with Theorem 2.5 yield the proof of. ■

A simple corollary is the following

Corollary 3.1 *Let $N(A)$ be the null space of A. For any two minimizers $\mathbf{x}^*$ and $\mathbf{x}$ of the unconstrained ℓ_1 minimization (2), $\mathbf{x}-\mathbf{x}^*\in Null(A)$. Similar for any two minimizers of (3).*

Next we prove the following

Proposition 3.2 (Extremal Value) *Suppose that $q=1$. Let $\mathbf{x}_{\mathbf{b},\epsilon}$ be the minimizers for input data $\mathbf{b}$ in problem (5) with $q=1$. Then*

$$\min_{\mathbf{x}\in\mathbf{R}^N} L_1(\epsilon,\mathbf{x})=\frac{1}{2\lambda}\left(\|\mathbf{b}\|_2^2-\|A\mathbf{x}_{\mathbf{b},\epsilon}\|_2^2\right)+\sum_{j=1}^N\frac{\epsilon}{\sqrt{\epsilon+|(\mathbf{x}_{\mathbf{b},\epsilon})_j|^2}}. \tag{26}$$

Consequently, for a minimizer $\mathbf{x_b}$ *of problem (4), we have*

$$\min_{\mathbf{x}\in\mathbf{R}^N}\{\|\mathbf{x}\|_1+\frac{1}{2\lambda}\|A\mathbf{x}-\mathbf{b}\|_2^2\}=\frac{1}{2\lambda}\left(\|\mathbf{b}\|_2^2-\|A\mathbf{x_b}\|_2^2\right). \tag{27}$$

Proof. For convenience, let us write $\mathbf{x}:=\mathbf{x}_{\mathbf{b},\epsilon}$. Multiplying $\mathbf{x}$ to the both sides of equation (7), we have

$$\sum_{j=1}^N\frac{x_j^2}{\sqrt{\epsilon+|x_j|^2}}+\frac{1}{\lambda}(A\mathbf{x})^T(A\mathbf{x}-\mathbf{b})=0.$$

The first term can be rewritten to

$$\sum_{j=1}^N\frac{x_j^2}{\sqrt{\epsilon+|x_j|^2}}=\|\mathbf{x}\|_\epsilon-\sum_{j=1}^N\frac{\epsilon}{\sqrt{\epsilon+|x_j|^2}}.$$

The second term can be rewritten to

$$\frac{1}{\lambda}(A\mathbf{x})^T(A\mathbf{x}-\mathbf{b})=\frac{1}{2\lambda}\left(\|A\mathbf{x}-\mathbf{b}\|_2^2+(A\mathbf{x}+\mathbf{b})^T(A\mathbf{x}-\mathbf{b})\right).$$

Combining these two equations together, we have

$$L_1(\epsilon,\mathbf{x})=\sum_{j=1}^N\frac{\epsilon}{\sqrt{\epsilon+|x_j|^2}}-\frac{1}{2\lambda}\left((A\mathbf{x}+\mathbf{b})^T(A\mathbf{x}-\mathbf{b})\right).$$

which yields (26) in this proposition. By letting ϵ go to zero in (26), we get (27). ■

It follows from (27) that for a minimizer $\mathbf{x_b}$ of problem (4), we have the following Pythagorean inequality:

$$\|A\mathbf{x_b}-\mathbf{b}\|_2^2+\|\mathbf{Ax_b}\|_2^2\le\|\mathbf{b}\|_2^2.$$

Next we generalize the proof of Lemma 2.1 to have

Proposition 3.3 *Suppose that* $q=1$. *Let* $\mathbf{x}$ *be the minimizer for problem (4) with* $q=1$. *Then for any* $\mathbf{y}$

$$\|A\mathbf{y}-A\mathbf{x}\|^2\le 2\lambda(L_1(0,\mathbf{y})-L_1(0,\mathbf{x})). \tag{28}$$

Proof. We follow the idea of proving Lemma 2.1. First we consider the problem (5) with $q=1$. It is easy to verify

$$\sqrt{\epsilon+y^2}-\sqrt{\epsilon+x^2}\ge\frac{(y-x)x}{\sqrt{\epsilon+|x|^2}} \tag{29}$$

Then

$$L_1(\epsilon,\mathbf{y})-L_1(\epsilon,\mathbf{x}) \;=\; \sum_{j=1}^N\sqrt{\epsilon+|y_j|^2}-\sum_{j=1}^N\sqrt{\epsilon+|x_j|^2}$$

$$
\begin{aligned}
& +\frac{1}{2\lambda}\left(\|A\mathbf{y}-\mathbf{b}\|^2-\|A\mathbf{x}-\mathbf{b}\|^2\right) \\
= \; & \sum_{j=1}^{N}\sqrt{\epsilon+|y_j|^2}-\sqrt{\epsilon+|x_j|^2}+\frac{1}{2\lambda}\|A\mathbf{y}-A\mathbf{x}\|^2 \\
& +\frac{1}{\lambda}(A\mathbf{x}-\mathbf{b})^T(A\mathbf{y}-A\mathbf{x})
\end{aligned}
$$

The last term can be reduced to

$$
-\sum_{j=1}^{N}\frac{(y_i-x_i)x_i}{\sqrt{\epsilon+|x|^2}}
$$

Then

$$
\begin{aligned}
& L_1(\epsilon,\mathbf{y})-L_1(\epsilon,\mathbf{x}) \\
= \; & \sum_{j=1}^{N}\left(\sqrt{\epsilon+|y_j|^2}-\sqrt{\epsilon+|x_j|^2}-\frac{x_j(y_j-x_j)}{\sqrt{\epsilon+|x_j|^2}}\right)+\frac{1}{2\lambda}\|A\mathbf{y}-A\mathbf{x}\|^2 \\
\geq \; & \frac{1}{2\lambda}\|A\mathbf{y}-A\mathbf{x}\|^2
\end{aligned}
$$

by using (29). Next we let $\epsilon\to 0$ to conclude the result in this proposition. ■

4 Computation of Sparse Solutions of Under-determined Linear Systems

In this section, we first explain that ϵ may be used for determining nonzero entries of critical points of (5) in practice when ϵ is very small, e.g., $\epsilon=10^{-5}$. For any fixed $\epsilon\in(0,1)$, let $\mathbf{x}^{\epsilon,q}$ be a critical point of (5) from Theorem 2.1 which satisfies (7). We observe from (7) that writing $A=[\mathbf{a}^{(1)},\cdots,\mathbf{a}^{(N)}]$, for each j with $1\leq j\leq N$,

$$
\begin{aligned}
\frac{q(x_j^{\epsilon,q})^2}{(\epsilon+(x^{\epsilon,q})^2)^{2-q}} &\leq\frac{1}{\lambda^2}\|\mathbf{a}^{(j)}\|_2^2\|A\mathbf{x}^{\epsilon,q}-\mathbf{b}\|_2^2 \\
&\leq\frac{2}{\lambda}\|\mathbf{a}^j\|_2^2L_q(\epsilon,\mathbf{x}^{\epsilon,q})\leq\frac{2}{\lambda}\|\mathbf{a}^j\|_2^2L_q(\epsilon,\mathbf{x}^{(1)}),
\end{aligned}
\tag{30}
$$

where $\mathbf{x}^{(1)}$ is the initial vector generating the sequence $\{\mathbf{x}^{(1)},\cdots,\mathbf{x}^{(k)},\cdots\}$ which contains a subsequence convergent to $\mathbf{x}^{\epsilon,q}$. For any entry $x_j^{\epsilon,q}$, we have two cases: $|x_j^{\epsilon,q}|^2>\epsilon$ and $|x_j^{\epsilon,q}|^2\leq\epsilon$. Let us discuss these two cases. Suppose $|x_j^{\epsilon,q}|^2>\epsilon$. Letting $L=$

$\frac{2}{\lambda}\max_j \|\mathbf{a}^{(j)}\|_2^2 L_q(1,\mathbf{x}^{(1)})$ be a constant bigger than the right-hand side of the above inequality (30), the above inequality (30) implies

$$\frac{q}{2^{2-q}}\frac{1}{|x_j^{\epsilon,q}|^{2(1-q)}} = \frac{q|x_j^{\epsilon,q}|^2}{(2|x_j^{\epsilon,q}|^2)^{2-q}} < \frac{q|x_j^{\epsilon,q}|^2}{(\epsilon+|x_j^{\epsilon,q}|^2)^{2-q}} < L.$$

It follows that $|x_j^{\epsilon,q}|^{2(1-q)} > \frac{q}{2^{2-q}L}$ or $|x_j^{\epsilon,q}|^2 > \left(\frac{q}{2^{2-q}L}\right)^{1/(1-q)}$.

On the other hand, if $|x_j^{\epsilon,q}|^2 \le \epsilon$, we have

$$\frac{q|x_j^{\epsilon,q}|^2}{(2\epsilon)^{2-q}} \le \frac{q|x_j^{\epsilon,q}|^2}{(\epsilon+|x_j^{\epsilon,q}|^2)^{2-q}} < L.$$

That is, $q|x_j^{\epsilon,q}|^2 < 2^{2-q}L\epsilon^{2-q}$ or $|x_j^{\epsilon,q}|^2 < \left(\frac{2^{2-q}L}{q}\right)\epsilon^{2-q}$. For convenience, let

$$B = \left(\frac{2^{2-q}L}{q}\right)^{1/(1-q)} < \infty.$$

We summarize the above discussion to have

Theorem 4.1 *Let $\mathbf{x}^{\epsilon,q}$ be a critical point of (5) from Theorem 2.1 for a fixed $\lambda > 0$. If ϵ is so small that $B\epsilon < 1$, then for each entry $x_j^{\epsilon,q}$ of $\mathbf{x}^{\epsilon,q}$,*

$$\textit{either } |x_j^{\epsilon,q}|^2 > (1/B) \textit{ or } |x_j^{\epsilon,q}|^2 \le (B\epsilon)^{1-q}\epsilon.$$

From the above discussion, we can see that if the square of an entry of a critical point is less than ϵ in magnitude, then it is much less than ϵ since $B\epsilon < 1$ is very small. Intuitively, such an entry should be treated as a zero. Similarly, if the square of an entry is larger than ϵ, it will be much bigger than ϵ since $|x_j^{\epsilon,q}|^2 > (1/B) = \frac{1}{B\epsilon}\epsilon$. Heuristically ϵ is a good indicator for entries of a critical point whether they are zero entries or nonzero entries. Let us threshold $\mathbf{x}^{\epsilon,q}$ by ϵ, that is, set all entries to be zero if $|x_i^{\epsilon,q}|^2 \le \epsilon$. Let $\widehat{\mathbf{x}}^{\epsilon,q}$ be the thresholded vector of $\mathbf{x}^{\epsilon,q}$. If ϵ is very small, we can view $\widehat{\mathbf{x}}^{\epsilon,q}$ as an approximate critical point. Since nonzero entries are bigger than $1/\sqrt{B}$ in magnitude, we have

$$\left(\frac{1}{B}\right)^{q/2}\|\widehat{\mathbf{x}}^{\epsilon,q}\|_0 \le \|\widehat{\mathbf{x}}^{\epsilon,q}\|_q^q \le \|\mathbf{x}^{\epsilon,q}\|_q^q \le L_q(\epsilon,\mathbf{x}^{(1)}) \tag{31}$$

by (12). It follows $\|\widehat{\mathbf{x}}^{\epsilon,q}\|_0 \le B^{q/2}L_q(\epsilon,\mathbf{x}^{(1)})$. One heuristic condition to ensure a critical point $\mathbf{x}^k$ with $\|\mathbf{x}^k\|_0 \le m/2$ is to choose a good initial guess such that $L_q(\epsilon,\mathbf{x}^{(1)})$ is as close to $m/2$ as possible since $B^{q/2}$ is close to 1 if q is very small. If we choose $\mathbf{x}^{(1)}$ to be a solution of $A\mathbf{x}^{(1)} = \mathbf{b}$ with small value $\|\mathbf{x}^{(1)}\|_q^q$, e.g., $< m/2$. then $L_q(\epsilon,\mathbf{x}^{(1)}) \le \|\mathbf{x}^{(1)}\|_q^q + N\epsilon^q \le m/2$ if ϵ is small enough. Then $\widehat{\mathbf{x}}^{\epsilon,q}$ may be a good approximate global minimizer.

Next we compare the performance of our unconstrained ℓ_q minimization described in Section 2 with six other existing algorithms, namely the orthogonal greedy algorithm (OGA)(cf. [33]), the ℓ_1 greedy algorithm (L1G) (cf. [27]), the standard ℓ_1 (L1) (cf. [5]) and the reweighted ℓ_1(RWL1) algorithms (cf. [8]) which can be obtained on-line from the Candès webpage, the regularized orthogonal matching pursuit (ROMP) (cf. [31]), in addition to the ℓ_q (Lq) algorithm developed in [20]. In our unconstrained ℓ_q (nLq) minimization, we choose $\lambda = 10^{-8}$ and run our iterative algorithm explained in Section 2 for many small $\epsilon > 0$ and $q = i/30, i = 1, \cdots, 30$. We have to admit that it takes very long time to get the sparse solution. However, if the sparsity is known, our algorithm finds the sparse solution quickly since an intermediate iterative solution $\mathbf{x}^{\epsilon,q}$ may have already been a solution. We use this additional assumption for our algorithm during this numerical experiment. In this comparison, we used 500 random pairs $(A, \mathbf{x})$ with matrices A of size $m \times N$ with $N = 250$ and $m = 50$ as in [14] and vectors $\mathbf{x} \in \mathbf{R}^N$ for sparsity $\|\mathbf{x}\|_0 = s$ for $s = 1, 2, \cdots, 30$, where matrices A are Gaussian random matrices whose entries are iid of $N(0, \sigma)$ with $\sigma = 1/50$. Thus $\mathbf{b} = A\mathbf{x}$ are known given vectors. These random matrices and vectors are generated by using MATLAB command **randn**. It is a classic result that these matrices have an s-RIP with overwhelming probability for an integer $s < m$. Although they may not be of completely full rank, m^{th} RIP condition is just a sufficient condition for our theory. Also in practice, the given matrix A may not be of completely full rank. We thus only use Gaussian random matrices for our experiments.

We use all 7 methods to solve $A\mathbf{x} = \mathbf{b}$. Several algorithms require some initial guesses. We have three schemes to generate initial guesses: 1) the pseudo-inverse method, 2) a solution from the ℓ_1 minimization, and 3) a solution of the linear system with submatrix, m columns randomly chosen from A.

For each s and each pair A and $\mathbf{b}$, we run each of the 7 algorithms to obtain vectors $\widetilde{\mathbf{x}}$, and we considered the recovery a success if $\|\mathbf{x} - \widetilde{\mathbf{x}}\|_\infty < 10^{-5}$. That is, our tolerance is 10^{-5}. We plot the percentages of successfully finding the sparse solutions for each method in Fig. 1. Here nL_q denotes our unconstrained ℓ_q minimization method. From the graph, we can see that our method is very close to the best performer: the ℓ_1 greedy algorithm. The interested reader may use our graph to compare with other methods which are not listed in this paper, e.g., the one in [14].

We have to point out that our algorithm is slower than the ℓ_1 greedy (L1G) algorithm since we have to do many iterative solutions for various ϵ and q. Nevertheless, our algorithm does offer some advantage when the matrix A is an uniform random matrix. That is, we also use these 7 algorithms above to test the recovery of the sparse solutions of under-determined linear systems for uniform random matrices A using MATLAB command **rand**, i.e. $A = rand(50, 250)$ and vector $\mathbf{x}$. All the procedures are exactly the same as above. Our method is clearly better as shown in Fig. 2.

Acknowledgement 4.1 *The authors would like to thank both anonymous reviewers for their comments and corrections which improve the presentation of the paper. In particular,*

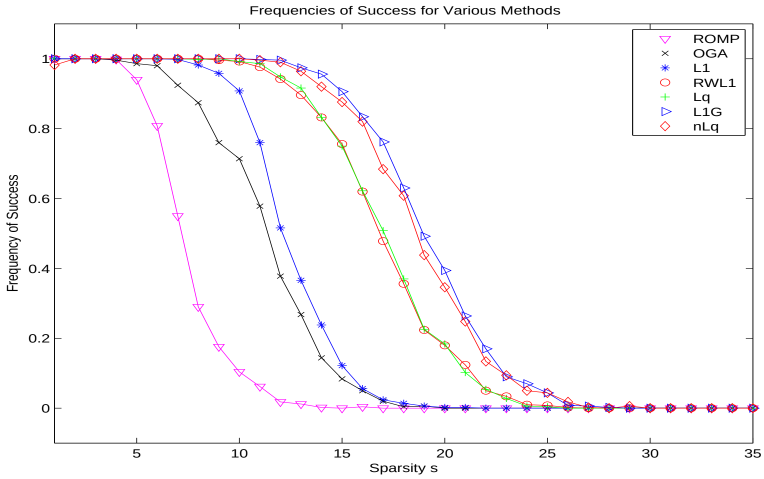

Figure 1: Comparison of the 7 algorithms: the regularized orthogonal matching pursuit (ROMP), the orthogonal greedy algorithm (OGA), the ℓ_1 greedy algorithm (L1G), the standard ℓ_1 (L1), the reweighted ℓ_1(RWL1) algorithms, the constrained ℓ_q (Lq) and unconstrained ℓ_q (nLq) for sparse solution of under-determined linear system associated with Gaussian random matrices

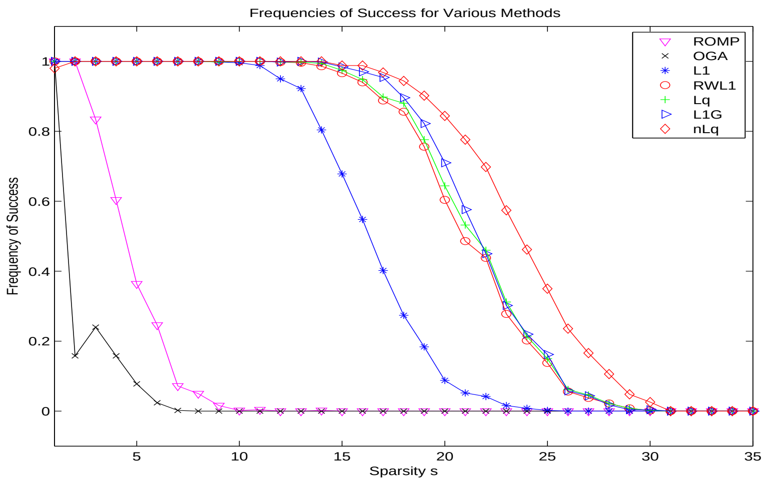

Figure 2: Comparison of the 7 algorithms for sparse solution of under-determined linear system associated with uniform random matrices

the authors sincerely thank one of them whose provides the reference [11] which is closely related to the study in this paper.

References

[1] R. Baraniuk, Compressive sensing, IEEE Signal Processing Magazine, (2007), vol. 24, 118–121.

[2] A. M. Bruckstein, D. L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Review, (2009) vol. 51, No . 1, pp. 34–81.

[3] E. J. Candès, Compressive sampling, International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich, 2006, pp. 1433–1452.

[4] E. J. Candès, The restricted isometry property and its implications for compressed sensing, Comptes Rendus de l'Académie des Sciences, Série I, 346 (2008), 589–592.

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

[6] E. J. Candès and T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory 51 (2005), no. 12, 4203–4215.

[7] E. J. Candès and T. Tao, Near-optimal signal recovery from random projections: universal encoding strategies, IEEE Trans. Inform. Theory 52 (2006), no. 12, 5406–5425.

[8] E. J. Candès, M. Watkin, and S. Boyd, Enhancing Sparsity by Reweighted l_1 Minimization, To appear in J. Fourier Anal. Appl., 2009.

[9] R. Chartrand, Exact reconstruction of sparse signals via nonconvex minimization, IEEE Signal Process. Lett., 14 (2007), 707–710.

[10] R. Chartrand and V. Staneva, Restricted isometry properties and nonconvex compressive sensing, Inverse Problem, 24(2008), 1–14.

[11] X. Chen, F. Xu and Y. Ye, Lower Bound Theory of Nonzero Entries in Solutions of l_2-l_p Minimization, http://www.stanford.edu/~yyye/

[12] I. Daubechies, M. Defreise, and C. De Mol, An iterative thresholding algorithm for linear inverse problem with a sparsity constraint, Comm. Pure and Applied Math., (2004) vol LVII, 1413–1457.

[13] I. Daubechies, M. Fornasier, I. Loris, Accelerated projected gradient method for linear inverse problems with sparsity constraints, J. Fourier Anal. Appl. 14 (2008), no. 5-6, 764–792.

[14] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntuk, Iteratively re-weighted least squares minimization for sparse recovery, Commun. Pure Appl. Math. 2009, to appear.

[15] D. L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (2006), no. 4, 1289–1306.

[16] D. L. Donoho and M. Elad, Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization, Proc. Natl. Acad. Sci. USA 100 (2003), no. 5, 2197–2202.

[17] D. L. Donoho, M. Elad, and V. N. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise. IEEE Trans. Inform. Theory, 52 (2006), 6–18.

[18] D. L. Donoho and B. F. Logan, Signal recovery and the large sieve, SIAM J. Appl. Math. 52 (1992), no. 2, 557–591.

[19] S. Foucart, A note on guaranteed sparse recovery via ℓ_1-minimization, Appl. Comput. Harmonic Anal., 29(2010), 97–103.

[20] S. Foucart and M. J. Lai, Sparsest Solutions of Under-determined Linear Systems via ℓ_q minimization for $0 < q \leq 1$, Applied Comput. Harmonic Analysis, 26(2009) 395–407.

[21] S. Foucart and M. J. Lai, Sparse Recovery with Pre-Gaussian Random Matrices, Studia Math. 200 (2010), 91–102.

[22] E. de Giorgi and T. Franzoni, Su un tipo di convergenza variazionale; Atti Accad. Naz. Lincei Rend. Cl. Sci. Mat. (8) 58(1975), 842–850.

[23] D. Ge, X. Jiang Y. Ye, A note on complexity of L_p minimization, Mathematical Programming manuscript, 2010.

[24] R. Gribnoval and M. Nielsen, Sparse decompositions in unions of bases, IEEE Trans. Info. Theory, 49(2003), 3320–3325.

[25] A. Juditsky and A. Nemirovskiy, On verifiable sufficient conditions for sparse signal recovery via ℓ_1 Minimization, manuscript, 2008.

[26] S. Karlin and W. J. Studden, Tchebycheff systems: with applications in Analysis and Statistics, Interscience, New York, 1966.

[27] I. Kozlov and A. Petukhov, Sparse Solutions of Underdetermined Linear Systems, in Handbook of Geomathematics, W. Freeden, M.Z. Nashed, T. Sonar (Eds.), Springer, 2010, 1243–1259.

[28] M. J. Lai, On sparse solution of underdetermined linear systems, in Journal of Concrete and Applicable Mathematics, 8(2010), 296–327.

[29] G. dal Maso, An introduction to Γ-convergence, Birkhäuser, Basel 1993.

[30] B. K. Natarajan, Sparse approximate solutions to linear systems, SIAM J. Comput., vol. 24, pp. 227–234, 1995.

[31] D. Needell, R. Vershynin, Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit, Found. Comput. Math., in Press, 2009.

[32] M. Nikolova, M. Ng, S. Zhang, and W.-K. Ching, Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization, SIAM J. Imaging Sci. 1 (2008), 2–25.

[33] A. Petukhov, Fast implementation of orthogonal greedy algorithm for tight wavelet frames, Signal Processing, 86 (2006), 471–479.

[34] Q. Sun, Recovery of sparsest signals via ℓ_q minimization, manuscript, 2010.

[35] V. N. Temlyakov, Nonlinear methods of approximation, Foundations of Comp. Math., 3 (2003), 33–107.

[36] J. A. Tropp, Greed is good: algorithmic results for sparse approximation, IEEE Trans. Inf. Theory, 50 (2004), 2231–2242.

[37] W. Yin, S. Osher, D. Goldfarb, and J. Darbon, Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing, SIAM J. Imaging Sciences, 1(2008), 143–168.