On the Scaling Law for Compressive Sensing and its Applications

Weiyu Xu
Cornell University
Ithaca NY 14853, USA
Email: wx42@cornell.edu

Ao Tang
Cornell University
Ithaca NY 14853, USA
Email: atang@ece.cornell.edu

***Abstract*—ℓ_1 minimization can be used to recover sufficiently sparse unknown signals from compressed linear measurements. In fact, exact thresholds on the sparsity (the size of the support set), under which with high probability a sparse signal can be recovered from i.i.d. Gaussian measurements, have been computed and are referred to as "weak thresholds" [4]. It was also known that there is a tradeoff between the sparsity and the ℓ_1 minimization recovery stability. In this paper, we give a *closed-form* characterization for this tradeoff which we call the scaling law for compressive sensing recovery stability. In a nutshell, we are able to show that as the sparsity backs off ϖ $(0 < \varpi < 1)$ from the weak threshold of ℓ_1 recovery, the parameter for the recovery stability will scale as $\frac{1}{\sqrt{1-\varpi}}$. Our result is based on a careful analysis through the Grassmann angle framework for the Gaussian measurement matrix. We will further discuss how this scaling law helps in analyzing the iterative reweighted ℓ_1 minimization algorithms. If the nonzero elements over the signal support follow a amplitude probability density function (pdf) $f(\cdot)$ whose t-th derivative $f^t(0) \neq 0$ for some integer $t \geq 0$, then a certain iterative reweighted ℓ_1 minimization algorithm can be analytically shown to lift the phase transition thresholds (weak thresholds) of the plain ℓ_1 minimization algorithm.**

I. Introduction

Compressive sensing addresses the problem of recovering sparse signals from under-determined systems of linear equations [18]. In particular, if $\mathbf{x}$ is an $n \times 1$ real-numbered vector that is known to have at most k nonzero elements where $k < n$, and $\mathbf{A}$ is an $m \times n$ measurement matrix with $k < m < n$, then for appropriate values of k, m and n, it is possible to efficiently recover $\mathbf{x}$ from $\mathbf{y} = \mathbf{Ax}$ [1], [2], [3], [5]. The most well recognized powerful recovery algorithm is ℓ_1 minimization which can be formulated as follows:

$$\min_{\mathbf{Az}=\mathbf{Ax}} \|\mathbf{z}\|_1 \tag{1}$$

The first result that established the fundamental phase transitions of signal recovery using ℓ_1 minimization is due to Donoho and Tanner [2], [4], where it was shown that if the measurement matrix is i.i.d. Gaussian, for a given ratio of $\delta = \frac{m}{n}$, ℓ_1 minimization can successfully recover *every* k-sparse signal, provided that $\mu = \frac{k}{n}$ is smaller that a certain threshold. This statement is true asymptotically as $n \to \infty$ and with high probability. This threshold guarantees the recovery of *all* sufficiently sparse signals and is therefore referred to as a "strong" threshold. It therefore does not depend on the actual distribution of the nonzero entries of the sparse signal and thus is a universal result.

Another notion introduced and computed in [2], [4] is that of a *weak* threshold $\mu_W(\delta)$ under which signal recovery is guaranteed for *almost all* support sets and *almost all* sign patterns of the sparse signal, with high probability as $n \to \infty$. The weak threshold is the one that can be observed in simulations of ℓ_1 minimization and allows for signal recovery beyond the strong threshold. It is also universal in the sense that it applies to any amplitude that the nonzero signal entries take.

When the sparsity of the signal $\mathbf{x}$ is larger than the weak threshold $\mu_W(\delta)n$, a common stability result for the ℓ_1 minimization is that, for a set $K \subseteq \{1, 2, ..., n\}$ with cardinality $|K|$ small enough for A to satisfy the restrict isometry condition [3] or the null space robustness property [13] [14], the decoding error is bounded by,

$$\|\mathbf{x} - \hat{\mathbf{x}}\|_1 \leq D\|\mathbf{x}_{\overline{K}}\|_1, \tag{2}$$

where $\hat{\mathbf{x}}$ is any minimizer to ℓ_1 minimization, D is a constant, $\overline{K}$ is the complement of the set K and $\mathbf{x}_{\overline{K}}$ is the part of $\mathbf{x}$ over the set $\overline{K}$.

To date, known bounds on $|K|/n$, for the restricted isometry condition to hold with overwhelming probability, are small compared with the weak threshold $\mu_W(\delta)$ [3]. [9] [14] used the Grassmann angle approach to characterize sharp bounds on the stability of ℓ_1 minimization and showed that, for an arbitrarily small ϵ_0, as long as $|K|/n = (1 - \epsilon_0)\mu_W(\delta)n$, with overwhelming probability as $n \to \infty$, (2) holds for some constant D (D of course depends on $|K|/n$). However, no closed-form formula for D were given.

In this paper, we give a *closed-form* characterization for this tradeoff which we call the scaling law for compressive sensing recovery stability. Namely, we will give a closed-form bound for D as a function of $|K|/n$. It is the first result of such kind. This result is obtained from close analysis through the Grassmann angle framework for the Gaussian measurement matrix. We will further discuss how this scaling law helps in analyzing the iterative reweighted ℓ_1 minimization algorithm.

Using this scaling law results for the stability and the Grassmann angle framework for the weighted ℓ_1 minimization, we prove that a certain *iterative reweighted* ℓ_1 algorithm indeed has better weak recovery guarantees for particular classes of sparse signals, including sparse Gaussian signals. We previously introduced these algorithms in [16], and had proven that for a very restricted class of sparse signals they outperform standard ℓ_1 minimization. In this paper, we are able to extend this result to a much wider and more reasonable class of sparse signals. The key to our result is the fact that for these signals, ℓ_1 minimization has an *approximate support recovery* property which can be exploited via a reweighted ℓ_1 algorithm, to obtain a provably superior weak threshold. More specifically, if the nonzero elements over the signal support follow a probability density function (pdf) $f(\cdot)$ whose t-th derivative $f^t(0) \neq 0$ for some $t \geq 0$, then a certain iterative reweighted ℓ_1 minimization algorithm can be analytically shown to lift the phase transition thresholds (weak thresholds) of the plain ℓ_1 minimization algorithm through using the scaling law for the sparse recovery stability. This extends our earlier results of weak threshold improvements for sparse vectors with nonzero elements following the Gaussian distribution, whose pdf is itself nonzero at the origin (namely its 0-th derivative is nonzero).

It is worth noting that different variations of reweighted ℓ_1 algorithms have been recently introduced in the literature and, have shown experimental improvement over ordinary ℓ_1 minimization [15], [7]. In [7] approximately sparse signals have been considered, where perfect recovery is never possible. However, it has been shown that the recovery noise can be reduced using an iterative scheme. In [15], a similar algorithm is suggested and is empirically shown to outperform ℓ_1 minimization for exactly sparse signals with non-flat distributions. Unfortunately, [15] provides no theoretical performance guarantee.

This paper is organized as follows. In Section II and III, we introduce the basic concepts and system model. In Section IV, we introduce and derive the main result of this paper: the scaling law for the compressive sensing recovery stability. In the following sections, we will use the scaling law to give new analysis results about the iterative reweighted ℓ_1 minimization algorithms.

II. Basic Definitions

A sparse signal with exactly k nonzero entries is called k-sparse. For a vector $\mathbf{x}$, $\|\mathbf{x}\|_1$ denotes the ℓ_1 norm. The support (set) of $\mathbf{x}$, denoted by $supp(\mathbf{x})$, is the index set of its nonzero coordinates. For a vector $\mathbf{x}$ that is not exactly k-sparse, we define the k-support of $\mathbf{x}$ to be the index set of the largest k entries of $\mathbf{x}$ in amplitude, and denote it by $supp_k(\mathbf{x})$. For a subset K of the entries of $\mathbf{x}$, $\mathbf{x}_K$ means the vector formed by those entries of $\mathbf{x}$ indexed in K. Finally, $\max|\mathbf{x}|$ and $\min|\mathbf{x}|$ mean the absolute value of the maximum and minimum entry of $\mathbf{x}$ in magnitude, respectively.

III. Signal Model and Problem Description

We consider sparse random signals with i.i.d. nonzero entries. In other words we assume that the unknown sparse signal is an $n\times 1$ vector $\mathbf{x}$ with exactly k nonzero entries, where each nonzero entry is independently sampled from a well defined distribution. The measurement matrix $\mathbf{A}$ is a $m \times n$ matrix with i.i.d. Gaussian entries with a ratio of dimensions $\delta = \frac{m}{n}$. Compressed sensing theory guarantees that if $\mu = \frac{k}{n}$ is smaller than a certain threshold, then every k-sparse signal can be recovered using ℓ_1 minimization. The relationship between δ and the maximum threshold of μ for which such a guarantee exists is called the *strong sparsity threshold*, and is denoted by $\mu_S(\delta)$. A more practical performance guarantee is the so-called *weak sparsity threshold*, denoted by $\mu_W(\delta)$, and has the following interpretation. For a fixed value of $\delta = \frac{m}{n}$ and i.i.d. Gaussian matrix $\mathbf{A}$ of size $m \times n$, a random k-sparse vector $\mathbf{x}$ of size $n\times 1$ with a randomly chosen support set and a random sign pattern can be recovered from $\mathbf{Ax}$ using ℓ_1 minimization with high probability, if $\frac{k}{n} < \mu_W(\delta)$. Similar recovery thresholds can be obtained by imposing more or less restrictions. For example, strong and weak thresholds for nonnegative signals have been evaluated in [6].

We assume that the support size of $\mathbf{x}$, namely k, is slightly larger than the weak threshold of ℓ_1 minimization. In other words, $k = (1+\epsilon_0)\mu_W(\delta)$ for some $\epsilon_0 > 0$. This means that if we use ℓ_1 minimization, a randomly chosen $\mu_W(\delta)n$-sparse signal will be recovered perfectly with very high probability, whereas a randomly selected k-sparse signal will not. We would like to show that for a strictly positive ϵ_0, the iterative reweighted ℓ_1 algorithm of Section V can indeed recover a randomly selected k-sparse signal with high probability, which means that it has an improved weak threshold.

IV. The Scaling Law for the Compressive Sensing Stability

In this section, we will derive the scaling of the ℓ_1 recovery stability as a function of the signal sparsity. More specifically, we are interested in characterizing a closed-form relationship between C and the sparsity $|K|$ in the following theorem.

Theorem 1. *Let A be a general $m \times n$ measurement matrix, $\mathbf{x}$ be an n-element vector and $\mathbf{y} = A\mathbf{x}$. Denote K as a subset of $\{1,2,\ldots,n\}$ such that its cardinality $|K| = k$ and further denote $\overline{K} = \{1,2,\ldots,n\}\setminus K$. Let $\mathbf{w}$ denote an $n\times 1$ vector. Let $C > 1$ be a fixed number.*

Given a specific set K and suppose that the part of $\mathbf{x}$ on K, namely $\mathbf{x}_K$ is fixed. $\forall \mathbf{x}_{\overline{K}}$, any solution $\hat{\mathbf{x}}$ produced by the ℓ_1 minimization satisfies

$$\|\mathbf{x}_K\|_1 - \|\hat{\mathbf{x}}_K\|_1 \leq \frac{2}{C-1}\|\mathbf{x}_{\overline{K}}\|_1$$

and

$$\|(\mathbf{x}-\hat{\mathbf{x}})_{\overline{K}}\|_1 \leq \frac{2C}{C-1}\|\mathbf{x}_{\overline{K}}\|_1,$$

if and only if $\forall \mathbf{w} \in \mathbb{R}^n$ such that $A\mathbf{w} = 0$, we have

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\overline{K}}}{C}\|_1 \geq \|\mathbf{x}_K\|_1. \quad (3)$$

In fact, if (3) is satisfied, we will have the stability result

$$\|(\mathbf{x}-\hat{\mathbf{x}})_{\overline{K}}\|_1 \leq \frac{2C}{C-1}\|\mathbf{x}_{\overline{K}}\|_1.$$

In [9], it was established that when the matrix A is sampled from an i.i.d. Gaussian ensemble, $C = 1$, considering a single index set K, there exists a constant ratio $0 < \mu_W < 1$ such that if $\frac{|K|}{n} \leq \mu_W$, then with overwhelming probability as $n \to \infty$, the condition (3) holds for all $\mathbf{w} \in \mathbb{R}^n$ satisfying $A\mathbf{w} = 0$. Now if we take a single index set K with cardinality $\frac{|K|}{n} = (1 - \varpi)\mu_W$, we would like to derive a characterization of C, as a function of $\frac{|K|}{n} = (1 - \varpi)\mu_W$, such that the condition (3) holds for all $\mathbf{w} \in \mathbb{R}^n$ satisfying $A\mathbf{w} = 0$. The main result of this paper is stated in the following theorem.

Theorem 2. *Assume the $m \times n$ measurement matrix A is sampled from an i.i.d. Gaussian ensemble, let K be a single index set with $\frac{|K|}{n} = (1 - \varpi)\mu_W$, where μ_W is the weak threshold for ideally sparse signals and ϖ is any real number between 0 and 1. We also let $\mathbf{x}$ be an n-dimensional signal vector with $\mathbf{x}_K$ being an arbitrary but fixed signal component. Then with overwhelming probability, the condition (3) holds for all $\mathbf{w} \in \mathbb{R}^n$ satisfying $A\mathbf{w} = 0$, with respect to the parameter $C = \frac{1}{\sqrt{1-\varpi}}$.*

Proof: When the measurement matrix A is sampled from an i.i.d. Gaussian ensemble, it is known that the probability that the condition (3) holds for all $\mathbf{w} \in \mathbb{R}^n$ satisfying $A\mathbf{w} = 0$ is the *Grassmann angle*, namely the probability that an $(n-m)$-dimensional uniformly distributed subspace intersects a polyhedral cone trivially (intersecting only at the apex of the cone). The complementary probability that the condition (3) does not hold for all $\mathbf{w} \in \mathbb{R}^n$ satisfying $A\mathbf{w} = 0$ is the *complementary Grassmann angle*. In our problem, without loss of generality, we scale $\mathbf{x}_K$ (extended to an n-dimensional vector supported on K) to a point in the relative interior of a $(k-1)$-dimensional face F of the weighted ℓ_1 ball,

$$\text{SP} = \{\mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\overline{K}}}{C}\|_1 \leq 1\}. \tag{4}$$

The polyhedral cone we are interested in for the complementary Grassmann angle is the cone $\text{SP} - \mathbf{x}_K$, namely the cone obtained by setting $\mathbf{x}_K$ as the apex, and observing SP from this apex.

Building on the works by Santalö [11] and McMullen [12] in high dimensional integral geometry and convex polytopes, the complementary Grassmann angle for the $(k-1)$-dimensional face F can be explicitly expressed as the sum of products of internal angles and external angles [10]:

$$P = 2 \times \sum_{s \geq 0} \sum_{G \in \Im_{m+1+2s}(\text{SP})} \beta(F, G)\gamma(G, \text{SP}), \tag{5}$$

where s is any nonnegative integer, G is any $(m+1+2s)$-dimensional face of the SP ($\Im_{m+1+2s}(\text{SP})$ is the set of all such faces), $\beta(\cdot,\cdot)$ stands for the internal angle and $\gamma(\cdot,\cdot)$ stands for the external angle.

The internal angles and external angles are basically defined as follows [10][12]:

- An internal angle $\beta(F_1, F_2)$ is the fraction of the hypersphere S covered by the cone obtained by observing the face F_2 from the face F_1. [1] The internal angle $\beta(F_1, F_2)$ is defined to be zero when $F_1 \nsubseteq F_2$ and is defined to be one if $F_1 = F_2$.
- An external angle $\gamma(F_3, F_4)$ is the fraction of the hypersphere S covered by the cone of outward normals to the hyperplanes supporting the face F_4 at the face F_3. The external angle $\gamma(F_3, F_4)$ is defined to be zero when $F_3 \nsubseteq F_4$ and is defined to be one if $F_3 = F_4$.

When $C = 1$, we denote the probability P in (5) as P_1. By definition, the weak threshold μ_W is the supremum of $\frac{|K|}{n} \leq \mu_W$ such that the probability P_1 in (5) goes to 0 as $n \to \infty$. We need to show for $\frac{|K|}{n} = (1 - \varpi)\mu_W$ and $C = \frac{1}{\sqrt{1-\varpi}}$, (5) also goes to 0 as $n \to \infty$. To that end, we only need to show the probability P' that, there exists an $\mathbf{w}$ from the null space of A such that

$$\|\mathbf{x}_K + \mathbf{w}_K\|_1 + \|\frac{\mathbf{w}_{\overline{K_1}}}{C_\infty}\|_1 + \|\frac{\mathbf{w}_{\overline{K_2}}}{C}\|_1 < \|\mathbf{x}_K\|_1 \tag{6}$$

goes to 0 as $n \to \infty$, where C_∞ is a large number which we may take as ∞ at the end, $\overline{K_1}$, $\overline{K_2}$ and K are disjoint sets such that $|\overline{K_1} \bigcup K| = \mu_W n$ and $\overline{K_1} \bigcup \overline{K_2} = \overline{K}$.

Then the probability P' will be equal to the probability that an $(n-m)$-dimensional uniformly distributed subspace intersects the polyhedral cone $\text{WSP} - \mathbf{x}_K$ nontrivially (intersecting at some other points besides the apex of the cone), where WSP is the polytope

$$\text{WSP} = \{\mathbf{y} \in \mathbb{R}^n \mid \|\mathbf{y}_K\|_1 + \|\frac{\mathbf{y}_{\overline{K_1}}}{C_\infty}\|_1 + \|\frac{\mathbf{y}_{\overline{K_2}}}{C}\|_1 \leq 1\}. \tag{7}$$

Then P' is also a complementary Grassmann angle, which can be expressed by [10]:

$$P' = 2 \times \sum_{s \geq 0} \sum_{G \in \Im_{m+1+2s}(\text{WSP})} \beta(F, G)\gamma(G, \text{WSP}). \tag{8}$$

Now we only need to show $P' \leq P_1$. If we denote $l = (m+1+2s)+1$ and $k = (1-\varpi)\mu_W n$, in the polytope WSP, then there are in total $\binom{n-k}{l-k}2^{l-k}$ faces G of dimension $(l-1)$ such that $F \subseteq G$ and $\beta(F, G) \neq 0$.

However, we argue that when C_∞ is very large, only $\binom{n-k_1}{l-k_1}2^{l-k}$ such faces G of dimension $(l-1)$ will contribute nonzero terms to P' in (8), where $k_1 = \mu_W n$. In fact, a certain $(l-1)$-dimensional face G supported on the index set L is the convex hull of $C_i e_i$, where $i \in L$, C_i is the corresponding weighting for index i (which is 1 for the set K, C_∞ for the set $\overline{K_1}$ and C for the set $\overline{K_2}$), and e_i is the standard unit coordinate vector. Now we show that if $\overline{K_1} \nsubseteq L$, the corresponding term in (8) for the face G will be 0 when C_∞ is very large.

Lemma 1. *Suppose that F is a $(k-1)$-dimensional face of WSP supported on the subset K with $|K| = k$. Then the external angle $\gamma(G, \text{WSP})$ between an $(l-1)$-dimensional face*

[1] Note the dimension of the hypersphere S here matches the dimension of the corresponding cone discussed. Also, the center of the hypersphere is the apex of the corresponding cone. All these defaults also apply to the definition of the external angles.

G supported on the set $L (F \subseteq G)$ and the polytope WSP is 0 *when $\overline{K_1} \nsubseteq L$ and C_∞ is large.*

Proof: Without loss of generality, assume $K = \{n-k+1, \cdots, n\}$. Consider the $(l-1)$-dimensional face

$$G = \text{conv}\{C_{n-l+1} \times e^{n-l+1}, ..., C_{n-k} \times e^{n-k}, e^{n-k+1}, ..., e^n\}$$

of WSP. The 2^{n-l} outward normal vectors of the supporting hyperplanes of the facets containing G are given by

$$\{\sum_{p=1}^{n-l} j_p e_p / C_p + \sum_{p=n-l+1}^{n-k} e_p / C_p + \sum_{p=n-k+1}^{n} e_p, j_p \in \{-1, 1\}\}.$$

Then the outward normal cone $c(G, \text{WSP})$ at the face G is the positive hull of these normal vectors. When $\overline{K_1} \nsubseteq L$, the fraction of the surface of the $(n-l-1)$-dimensional sphere taken by the cone $c(G, \text{WSP})$ is 0 since the corresponding C_p is very large. ∎

Now let us look at the internal angle $\beta(F, G)$ between the $(k-1)$-dimensional face F and an $(l-1)$-dimensional face G, where $\overline{K_1}$ is a subset of the support set of G. Notice that the only interesting case is when $F \subseteq G$ since $\beta(F, G) \neq 0$ only if $F \subseteq G$. We will see if $F \subseteq G$, the cone $c(F, G)$ formed by observing G from F is the direct sum of a $(k-1)$-dimensional linear subspace and the positive hull of $(l-k)$ vectors. These $(l-k)$ vectors are in the form

$$v_i = (-\frac{1}{k}, ..., -\frac{1}{k}, 0, ..., C_i, 0, ...0), i \in L \setminus K.$$

For those vectors v_i with $i \in \overline{K_1}$, $C_i = C_\infty$. When C_∞ is very large, the considered cone takes half of the space at each i-th coordinate with $i \in \overline{K_1}$.

So by the definition of the internal angle, the internal angle $\beta(F, G)$ is equal to $\frac{1}{2^{k_1-k}} \times \beta(F, G_1)$, where G_1 is supported only on the set $L \setminus \overline{K_1}$. It is known that this internal angle $\beta(F, G_1)$ is equal to the fraction of an $(l-k_1-1)$-dimensional sphere taken by a polyhedral cone formed by $(l-k_1)$ unit vectors with inner product $\frac{1}{1+C^2 k}$ between each other. In this case, the internal angle is given by

$$\beta(F, G) = \frac{1}{2^{k_1-k}} \frac{V_{l-k_1-1}(\frac{1}{1+C^2 k}, l-k_1-1)}{V_{l-k_1-1}(S^{l-k_1-1})}, \quad (9)$$

where $V_i(S^i)$ denotes the i-th dimensional surface measure on the unit sphere S^i, while $V_i(\alpha', i)$ denotes the surface measure for regular spherical simplex with $(i+1)$ vertices on the unit sphere S^i and with inner product as α' between these $(i+1)$ vertices. Thus (9) is equal to $B(\frac{1}{1+C^2 k}, l-k_1)$, where

$$B(\alpha', m') = \theta^{\frac{m'-1}{2}} \sqrt{(m'-1)\alpha' + 1} \pi^{-m'/2} \alpha'^{-1/2} J(m', \theta), \quad (10)$$

with $\theta = (1-\alpha')/\alpha'$ and

$$J(m', \theta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} (\int_0^{\infty} e^{-\theta v^2 + 2iv\lambda} dv)^{m'} e^{-\lambda^2} d\lambda. \quad (11)$$

If we take $C = \frac{1}{\sqrt{1-\varpi}}$, then

$$\frac{1}{1+C^2 k} = \frac{1}{1+k_1}.$$

By comparison, $\beta(F, G) = \frac{1}{2^{k_1-k}} \times \beta(F, G)$ is exactly the $\frac{1}{2^{k_1-k}} \beta(F_1, G_1)$ term appearing in the expression for the Grassmann angle P between the face F_1 supported on the set K_1 and the polytope SP, where G_1 is an $(l-1)$-dimensional face of SP supported on the set L.

Similar to the derivation for the internal angle, we can show that the external angle $\gamma(G, \text{WSP})$ is also exactly equal to $\gamma(G_1, \text{SP})$ term appearing in the expression for the Grassmann angle P between the face F_1 supported on the set K_1 and the polytope SP, where G_1 an $(l-1)$-dimensional face of SP supported on the set L.

Since there are in total only $\binom{n-k_1}{l-k_1} 2^{l-k}$ such faces G of dimension $(l-1)$ will contribute nonzero terms to P' in (8), substituting the results for the internal and external angles, we have $P = P'$. Thus for $\frac{|K|}{n} = (1-\varpi)\mu_W$ and $C = \frac{1}{\sqrt{1-\varpi}}$, with high probability, the condition the condition (3) holds for all $\mathbf{w} \in \mathbb{R}^n$ satisfying $A\mathbf{w} = 0$.

∎

V. Iterative Weighted ℓ_1 Algorithm

Beginning from this section, we will see how the stability result is used in analyzing the iterative reweighted ℓ_1 minimization algorithms. We focus on the following algorithm from [16], [17], consisting of two ℓ_1 minimization steps: a standard one and a weighted one. The input to the algorithm is the vector $\mathbf{y} = \mathbf{A}\mathbf{x}$, where $\mathbf{x}$ is a k-sparse signal with $k = (1+\epsilon_0)\mu_W(\delta)n$, and the output is an approximation $\mathbf{x}^*$ to the unknown vector $\mathbf{x}$. We assume that k, or an upper bound on it, is known. Also $\omega > 1$ is a predetermined weight.

Algorithm 1.

1) *Solve the ℓ_1 minimization problem:*

$$\hat{\mathbf{x}} = \arg\min \|\mathbf{z}\|_1 \ \textit{subject to} \ \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}. \quad (12)$$

2) *Obtain an approximation for the support set of $\mathbf{x}$: find the index set $L \subset \{1, 2, ..., n\}$ which corresponds to the largest k elements of $\hat{\mathbf{x}}$ in magnitude.*

3) *Solve the following weighted ℓ_1 minimization problem and declare the solution as output:*

$$\mathbf{x}^* = \arg\min \|\mathbf{z}_L\|_1 + \omega \|\mathbf{z}_{\overline{L}}\|_1 \ \textit{subject to} \ \mathbf{A}\mathbf{z} = \mathbf{A}\mathbf{x}. \quad (13)$$

The intuition behind the algorithm is as follows. In the first step we perform a standard ℓ_1 minimization. If the sparsity of the signal is beyond the weak threshold $\mu_W(\delta)n$, then ℓ_1 minimization is not capable of recovering the signal. However, we use the output of the ℓ_1 minimization to identify an index set, L, which hopefully contains most of the nonzero entries of $\mathbf{x}$. We finally perform a weighted ℓ_1 minimization by penalizing those entries of $\mathbf{x}$ that are not in L because they have a lower chance of being nonzero elements.

In the next sections we formally prove that the above intuition is correct and that, for certain classes of signals, Algorithm 1 has a recovery threshold beyond that of standard ℓ_1 minimization. The idea of the proof is as follows. In Section VI, we prove that there is a large overlap between the index

978-1-4244-9848-2/11$26.00©2011 IEEE

set L, found in Step 2 of the algorithm, and the support set of the unknown signal $\mathbf{x}$ (denoted by K)—see Theorem 3. Then in Section VII, we show that the large overlap between K and L can result in perfect recovery of $\mathbf{x}$, beyond the standard weak threshold, when a weighted ℓ_1 minimization is used in Step 3.

VI. Approximate Support Recovery, Steps 1 and 2 of the Algorithm

In this section, we carefully study the first two steps of Algorithm 1. The unknown signal $\mathbf{x}$ is assumed to be a Gaussian k-sparse vector with support set K, where $k = |K| = (1+\epsilon_0)\mu_W(\delta)n$, for some $\epsilon_0 > 0$. The set L, as defined in the algorithm, is in fact the k-support set of $\hat{\mathbf{x}}$. We show that for small enough ϵ_0, the intersection of L and K is very large with high probability, so that L can be counted as a good approximation to K.

In order to lower bound $|L \cap K|$, we separate our work in two steps. First, we state a general lemma that bounds $|K \cap L|$ as a function of $\|\mathbf{x}-\hat{\mathbf{x}}\|_1$ [17]. Then, we mention an intrinsic property of ℓ_1 minimization called *weak robustness* that provides an upper bound on the quantity $\|\mathbf{x}-\hat{\mathbf{x}}\|_1$.

Definition 1. *For a k-sparse signal* $\mathbf{x}$*, we define* $W(\mathbf{x},\lambda)$ *to be the size of the largest subset of nonzero entries of* $\mathbf{x}$ *that has a* ℓ_1 *norm less than or equal to* λ*.*

$$W(\mathbf{x},\lambda) := \max\{|S| \mid S \subseteq \mathit{supp}(\mathbf{x}),\ \|\mathbf{x}_S\|_1 \le \lambda\}$$

Note that $W(\mathbf{x},\lambda)$ is increasing in λ.

Lemma 2. *[17] Let* $\mathbf{x}$ *be a k-sparse vector and* $\hat{\mathbf{x}}$ *be another vector. Also, let* K *be the support set of* $\mathbf{x}$ *and* L *be the k-support set of* $\hat{\mathbf{x}}$*. Then*

$$|K \cap L| \ge k - W(\mathbf{x}, \|\mathbf{x}-\hat{\mathbf{x}}\|_1) \tag{14}$$

We now review the notion of weak robustness, which allows us to bound $\|\mathbf{x}-\hat{\mathbf{x}}\|_1$, and has the following formal definition [9].

Definition 2. *Let the set* $S \subset \{1,2,\cdots,n\}$ *and the subvector* $\mathbf{x}_S$ *be fixed. A solution* $\hat{\mathbf{x}}$ *is called weakly robust if, for some* $C > 1$ *called the robustness factor, and all* $\mathbf{x}_{\overline{S}}$*, it holds that*

$$\|(\mathbf{x}-\hat{\mathbf{x}})_{\overline{S}}\|_1 \le \frac{2C}{C-1}\|\mathbf{x}_{\overline{S}}\|_1 \tag{15}$$

and

$$\|\mathbf{x}_S\| - \|\hat{\mathbf{x}}_S\| \le \frac{2}{C-1}\|\mathbf{x}_{\overline{S}}\|_1 \tag{16}$$

The weak robustness notion allows us to bound the error in $\|\mathbf{x}-\hat{\mathbf{x}}\|_1$ in the following way. If the matrix $\mathbf{A}_S$, obtained by retaining only those columns of $\mathbf{A}$ that are indexed by S, has full column rank, then the quantity

$$\kappa = \max_{\mathbf{A}\mathbf{w}=0,\mathbf{w}\neq 0} \frac{\|\mathbf{w}_S\|_1}{\|\mathbf{w}_{\overline{S}}\|_1}$$

must be finite, and one can write

$$\|\mathbf{x}-\hat{\mathbf{x}}\|_1 \le \frac{2C(1+\kappa)}{C-1}\|\mathbf{x}_{\overline{S}}\|_1 \tag{17}$$

From [9] and the scaling law discovered in this paper, we know that for Gaussian i.i.d. measurement matrices $\mathbf{A}$, ℓ_1 minimization is weakly robust, i.e., there exists a robustness factor $C > 1$ as a function of $\frac{|S|}{n} < \mu_W(\delta)$ for which (15) and (16) hold. Now let $k_1 = (1-\epsilon_1)\mu_W(\delta)n$ for some small $\epsilon_1 > 0$, and K_1 be the k_1-support set of $\mathbf{x}$, namely, the set of the largest k_1 entries of $\mathbf{x}$ in magnitude. Based on equation (17) we may write

$$\|\mathbf{x}-\hat{\mathbf{x}}\|_1 \le \frac{2C(1+\kappa)}{C-1}\|\mathbf{x}_{\overline{K_1}}\|_1 \tag{18}$$

For a fixed value of δ, C in (18) is a function of ϵ_1 following the scaling law discovered in this paper, and becomes arbitrarily close to 1 as $\epsilon_1 \to 0$. κ is also a bounded function of ϵ_1 and therefore we may replace it with an upper bound κ^*. We now have a bound on $\|\mathbf{x}-\hat{\mathbf{x}}\|_1$. To explore this inequality and understand its asymptotic behavior, we apply a third result, which is a certain concentration bound on the order statistics of the random variables following certain amplitude distributions.

Lemma 3. *Suppose* $X_1, X_2, \cdots, X_N$ *are* N *i.i.d. random variables whose amplitudes, with a mean value of* $E(|X|)$*, follow the probability density function* $f(x)$ *for* $x \ge 0$*. Let* $S_N = \sum_{i=1}^{N}|X_i|$ *and let* S_M *be the sum of the smallest* M *numbers among the* $|X_i|$*, for each* $1 \le M \le N$*. Then for every* $\epsilon > 0$*, as* $N \to \infty$*, we have*

$$\mathbb{P}(|\frac{S_N}{N} - E(|X|)| > \epsilon) \to 0,$$
$$\mathbb{P}(|\frac{S_M}{S_N} - \frac{1}{E(|X|)}\int_0^{F^{-1}(\frac{M}{N})} xf(x)dx| > \epsilon) \to 0,$$

where $F(x)$ *is the corresponding cumulative distribution function for the considered random variable amplitude* $|X|$*.*

Without loss of generality, we assume $E(|X|) = 1$. As a direct consequence of Lemma 3 we can write:

$$\mathbb{P}(|\frac{\|\mathbf{x}_{\overline{K_1}}\|_1}{\|\mathbf{x}\|_1} - \int_0^{F^{-1}(\frac{\epsilon_0+\epsilon_1}{1+\epsilon_0})} xf(x)dx| > \epsilon) \to 0 \tag{19}$$

for all $\epsilon > 0$ as $n \to \infty$. Define

$$\zeta(\epsilon_0) := \inf_{\epsilon_1 > 0} \frac{2C(1+\kappa^*)}{C-1} \int_0^{F^{-1}(\frac{\epsilon_0+\epsilon_1}{1+\epsilon_0})} xf(x)dx > \epsilon$$

Combining (18) with (19) we can get

$$\mathbb{P}(\frac{\|\mathbf{x}-\hat{\mathbf{x}}\|_1}{\|\mathbf{x}\|_1} - \zeta(\epsilon_0) < \epsilon) \to 1 \tag{20}$$

for all $\epsilon > 0$ as $n \to \infty$. In summary, we have showed that $|K \cap L| \ge k - W(\mathbf{x}, \|\mathbf{x}-\hat{\mathbf{x}}\|_1)$, and then "weak robustness" of ℓ_1 minimization then guarantee that for large n with high probability $\|\mathbf{x}-\hat{\mathbf{x}}\|_1 \le \zeta(\epsilon_0)\|\mathbf{x}\|_1$. These results will further lead to the main claim on the support recovery.

Theorem 3 (Support Recovery). *Let* $\mathbf{A}$ *be an i.i.d. Gaussian* $m \times n$ *measurement matrix with* $\frac{m}{n} = \delta$. *Let* $k = (1+\epsilon_0)\mu_W(\delta)$ *and* $\mathbf{x}$ *be an* $n \times 1$ *random* k*-sparse vector whose nonzero element amplitude follows the distribution of* $f(x)$. *Suppose that* $\hat{\mathbf{x}}$ *is the approximation to* $\mathbf{x}$ *given by the* ℓ_1 *minimization, namely* $\hat{\mathbf{x}} = argmin_{\mathbf{Az}=\mathbf{Ax}}\|\mathbf{z}\|_1$. *Then, for any* $\epsilon_0 > 0$ *and for all* $\epsilon > 0$, *as* $n \to \infty$,

$$\mathbb{P}(\frac{|supp(\mathbf{x}) \cap supp_k(\hat{\mathbf{x}})|}{k} - (1 - F(y^*)) > -\epsilon) \to 1, \qquad (21)$$

where y^* *is the solution to* y *in the equation* $\int_0^y xf(x)dx = \zeta(\epsilon_0)$.

Moreover, if the integer $t \geq 0$ *is the smallest integer for which the amplitude distribution* $f(x)$ *has a nonzero t-th order derive at the origin, namely* $f^{(t)}(0) \neq 0$, *then as* $\epsilon_0 \to 0$, *with high probability,*

$$\frac{|supp(\mathbf{x}) \cap supp_k(\hat{\mathbf{x}})|}{k} = 1 - O(\epsilon_0^{\frac{1}{t+2}}). \qquad (22)$$

The proof of Theorem 3 relies on the scaling law for recovery stability in this paper and concentration Lemma 3. Note that if $\epsilon_0 \to 0$, then Theorem 3 implies that $\frac{|K \cap L|}{k}$ becomes arbitrarily close to 1. We can also see that the support recovery is better when the probability distribution function of $f(x)$ has a lower order of nonzero derivative. This is consistent with the better recovery performance observed for such distributions in simulations of the iterative reweighted ℓ_1 minimization algorithms.

VII. Perfect Recovery, Step 3 of the Algorithm

In Section VI we showed that. if ϵ_0 is small, the k-support of $\hat{\mathbf{x}}$, namely $L = supp_k(\hat{\mathbf{x}})$, has a significant overlap with the true support of $\mathbf{x}$. The scaling law gives a quantitative lower bound on the size of this overlap in Theorem 3. In Step 3 of Algorithm 1, weighted ℓ_1 minimization is used, where the entries in $\overline{L}$ are assigned a higher weight than those in L. In [8], we have been able to analyze the performance of such weighted ℓ_1 minimization algorithms. The idea is that if a sparse vector $\mathbf{x}$ can be partitioned into two sets L and $\overline{L}$, where in one set the fraction of non-zeros is much larger than in the other set, then (13) can potentially increase the recovery threshold of ℓ_1 minimization.

Theorem 4. *[8] Let* $L \subset \{1, 2, \cdots, n\}$, $\omega > 1$ *and the fractions* $f_1, f_2 \in [0, 1]$ *be given. Let* $\gamma_1 = \frac{|L|}{n}$ *and* $\gamma_2 = 1 - \gamma_1$. *There exists a threshold* $\delta_c(\gamma_1, \gamma_2, f_1, f_2, \omega)$ *such that with high probability, almost all random sparse vectors* $\mathbf{x}$ *with* at least $f_1\gamma_1 n$ *nonzero entries over the set* L, *and* at most $f_2\gamma_2 n$ *nonzero entries over the set* $\overline{L}$ *can be perfectly recovered using* $\min_{\mathbf{Az}=\mathbf{Ax}} \|\mathbf{z}_L\|_1 + \omega\|\mathbf{z}_{\overline{L}}\|_1$, *where* $\mathbf{A}$ *is a* $\delta_c n \times n$ *matrix with i.i.d. Gaussian entries. Furthermore, for appropriate* ω,

$$\mu_W(\delta_c(\gamma_1, \gamma_2, f_1, f_2, \omega)) < f_1\gamma_1 + f_2\gamma_2,$$

i.e., standard ℓ_1 *minimization using a* $\delta_c n \times n$ *measurement matrix with i.i.d. Gaussian entries cannot recover such* x.

A software package for computing such thresholds can also be found in [19]. We then summarize the threshold improvement result in the following theorem, with the detailed proofs omitted due to limited space.

Theorem 5 (Perfect Recovery). *Let* $\mathbf{A}$ *be an* $m \times n$ *i.i.d. Gaussian matrix with* $\frac{m}{n} = \delta$. *If* $\delta_c(\mu_W(\delta), 1 - \mu_W(\delta), 1, 0, \omega) < \delta$, *then there exist* $\epsilon_0 > 0$ *and* $\omega > 0$ *such that, with high probability as* n *grows to infinity, Algorithm 1 perfectly recovers a random* $(1 + \epsilon_0)\mu_W(\delta)n$*-sparse vector with i.i.d. nonzero entries following an amplitude distribution whose pdf has a nonzero derive of some finite order at the origin.*

VIII. Acknowledgement

The research is supported by NSF under CCF-0835706.

References

[1] D. Donoho," Compressed sensing", *IEEE Trans. on Information Theory*, 52(4), pp. 1289-1306, April 2006.

[2] D. Donoho, "High-Dimensional Centrally Symmetric Polytopes with Neighborliness Proportional to Dimension ", *Discrete and Computational Geometry* , 102(27), pp. 617-652, 2006, Springer .

[3] E. Candès and T. Tao, "Decoding by linear programming", *IEEE Trans. on Information Theory*, 51(12), pp. 4203 - 4215, December 2005.

[4] D. Donoho and J. Tanner, "Thresholds for the Recovery of Sparse Solutions via L1 Minimization", *Proceedings of the Conference on Information Sciences and Systems*, March 2006.

[5] R. G. Baraniuk and M. B. Wakin "Random Projections of Smooth Manifolds ", *Journal of Foundations of Computational Mathematics*, Volume 9, No.1, Feb. 2009,

[6] D. Donoho and J. Tanner, "Sparse nonnegative solutions of underdetermined linear equations by linear programming" Proc. National Academy of Sciences, 102(27), pp.9446-9451, 2005.

[7] D. Needell, "Noisy signal recovery via iterative reweighted L1-minimization" Proc. Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA Nov. 2009.

[8] A. Khajehnejad, W. Xu, A. Avestimehr, Babak Hassibi, "Weighted ℓ_1 minimization for Sparse Recovery with Prior Information", Proceedings of ISIT 2009.

[9] W. Xu and B. Hassibi, "On Sharp Performance Bounds for Robust Sparse Signal Recoveries", *Proceedings of the International Symposium on Information Theory 2009.*

[10] Branko Grünbaum. *Convex polytopes*, volume 221 of *Graduate Texts in Mathematics.* Springer-Verlag, New York, Second Edition, 2003. Prepared and with a preface by Volker Kaibel, Victor Klee and Gnterü M. Ziegler.

[11] L. A. Santaló. Geometría integral enespacios de curvatura constante. *Rep.Argetina Publ.Com.Nac.Energí Atómica, Ser.Mat 1, No.1*, 1952.

[12] Peter McMullen. Non-linear angle-sum relations for polyhedral cones and polytopes. *Math. Proc. Cambridge Philos. Soc.*, 78(2):247–261, 1975.

[13] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *J. Amer. Math. Soc. 22 (2009), 211-231*, 2008.

[14] Weiyu Xu and Babak Hassibi. Compressed sensing over the grassmann manifold: A unified analytical framework. *Proceedings of the Forty-Sixth Annual Allerton Conference on Communication, Control, and Computing*, 2008.

[15] E. J. Candès, M. B. Wakin, and S. Boyd, "Enhancing Sparsity by Reweighted l1 Minimization", *Journal of Fourier Analysis and Applications*, 14(5), pp. 877-905, special issue on sparsity, December 2008.

[16] W. Xu, A. Khajehnejad, S. Avestimehr, B. Hassibi, "Breaking through the Thresholds: an Analysis for Iterative Reweighted ℓ_1 Minimization via the Grassmann Angle Framework", in Proceedings of ICASSP 2010.

[17] A. Khajehnejad, W.Xu, S. Avestimehr, B. Hassibi, "Improved Sparse Recovery Thresholds with Two Step Reweighted L1 Minimization", in Proceedings of ISIT, 2010.

[18] http://www.dsp.ece.rice.edu/cs

[19] http://www.its.caltech.edu/~amin/weighted_l1_codes/