On the Performance of Sparse Recovery via ℓ_p-minimization $(0 \leq p \leq 1)$

Meng Wang, *Student Member, IEEE,* Weiyu Xu, and Ao Tang, *Member, IEEE*

***Abstract*—It is known that a high-dimensional sparse vector $\mathbf{x}^*$ in $\mathcal{R}^n$ can be recovered from low-dimensional measurements $\mathbf{y} = A\mathbf{x}^*$ where $A^{m\times n}(m < n)$ is the measurement matrix. In this paper, with A being a random Gaussian matrix, we investigate the recovering ability of ℓ_p-minimization $(0 \leq p \leq 1)$ as p varies, where ℓ_p-minimization returns a vector with the least ℓ_p quasi-norm among all the vectors $\mathbf{x}$ satisfying $A\mathbf{x} = \mathbf{y}$. Besides analyzing the performance of strong recovery where ℓ_p-minimization is required to recover all the sparse vectors up to certain sparsity, we also for the first time analyze the performance of "weak" recovery of ℓ_p-minimization $(0 \leq p < 1)$ where the aim is to recover all the sparse vectors on one support with a fixed sign pattern. When $\alpha(:= \frac{m}{n}) \to 1$, we provide sharp thresholds of the sparsity ratio (i.e. percentage of nonzero entries of a vector) that differentiates the success and failure via ℓ_p-minimization. For strong recovery, the threshold strictly decreases from 0.5 to 0.239 as p increases from 0 to 1. Surprisingly, for weak recovery, the threshold is $2/3$ for all p in $[0, 1)$, while the threshold is 1 for ℓ_1-minimization. We also explicitly demonstrate that ℓ_p-minimization $(p < 1)$ can return a denser solution than ℓ_1-minimization. For any $\alpha \in (0, 1)$, we provide bounds of the sparsity ratio for strong recovery and weak recovery respectively below which ℓ_p-minimization succeeds. Our bound of strong recovery improves on the existing bounds when α is large. In particular, regarding the recovery threshold, this paper argues that ℓ_p-minimization has a higher threshold with smaller p for strong recovery; the threshold is the same for all p for sectional recovery; and ℓ_1-minimization can outperform ℓ_p-minimization for weak recovery. These are in contrast to traditional wisdom that ℓ_p-minimization, though computationally more expensive, always has better sparse recovery ability than ℓ_1-minimization since it is closer to ℓ_0-minimization. Finally, we provide an intuitive explanation to our findings. Numerical examples are also used to unambiguously confirm and illustrate the theoretical predictions.**

***Index Terms*—Compressed sensing, Sparse recovery, ℓ_p-minimization, recovery threshold, phase transition.**

I. Introduction

We consider recovering a vector $\mathbf{x}$ in $\mathcal{R}^n$ from an m-dimensional measurement $\mathbf{y} = A\mathbf{x}$, where $A^{m\times n}(m < n)$ is the measurement matrix. Obviously, given $\mathbf{y}$ and A, $A\mathbf{x} = \mathbf{y}$ is an underdetermined linear system and admits an infinite number of solutions. However, if $\mathbf{x}$ is sparse, i.e. it only has a small number of nonzero entries compared with its dimension, one can actually recover $\mathbf{x}$ from $\mathbf{y}$ under certain conditions. This topic is known as *compressed sensing* and draws much attention recently, for example, [7], [8], [18], [20].

Given $\mathbf{x} \in \mathcal{R}^n$, its support T is defined as $T = \{i \in \{1, ..., n\} : x_i \neq 0\}$. The cardinality $|T|$ of set T is the sparsity of $\mathbf{x}$, which also equals to the ℓ_0 norm $\|\mathbf{x}\|_0 := |\{i : x_i \neq 0\}|$. We say $\mathbf{x}$ is ρn-sparse if $|T| = \rho n$ for some $\rho < 1$. Given the measurement $\mathbf{y}$ and the measurement matrix A, together with the assumption that $\mathbf{x}$ is sparse, one natural estimate of $\mathbf{x}$ is the vector with the least ℓ_0 norm that can produce the measurement $\mathbf{y}$. Mathematically, to recover $\mathbf{x}$, we solve the following ℓ_0-minimization problem:

$$\min_{\mathbf{x}\in\mathcal{R}^n} \|\mathbf{x}\|_0 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{y}. \tag{1}$$

However, (1) is combinatorial and computationally intractable except for small problems, and one commonly used approach is to solve a closely related ℓ_1-minimization problem:

$$\min_{\mathbf{x}\in\mathcal{R}^n} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{y}, \tag{2}$$

where $\|\mathbf{x}\|_1 := \sum_i |x_i|$. (2) is a convex problem and can be recast as a linear program, thus can be solved efficiently. Conditions under which (2) can successfully recover $\mathbf{x}$ have been extensively studied in the literature of compressed sensing. For example, the Restricted Isometry Property (RIP) conditions [6]–[8] can guarantee that (2) accurately recovers the sparse vector.

Among the explosion of research on compressed sensing ([1]–[3], [5], [14], [28], [34], [35]) recently, there has been great research interest in recovering $\mathbf{x}$ by ℓ_p-minimization for $0 < p < 1$ ([9], [10], [12], [13], [15], [24], [31]) as follows,

$$\min_{\mathbf{x}\in\mathcal{R}^n} \|\mathbf{x}\|_p \quad \text{s.t.} \quad A\mathbf{x} = \mathbf{y}. \tag{3}$$

Recall that $\|\mathbf{x}\|_p^p := (\sum_i |x_i|^p)$ for $p > 0$. Though $\|\cdot\|_p$ is a quasi-norm when $p < 1$ as it violates the triangular inequality, $\|\cdot\|_p^p$ follows the triangular inequality. We say $\mathbf{x}$ can be recovered by ℓ_p-minimization if and only if it is the unique solution to (3). (3) is non-convex, and finding the global minimum is in general computationally hard. [9], [10], [12] employ heuristic algorithms to compute a local minimum of (3) and show numerically that these heuristics can indeed recover sparse vectors, and the support size of these vectors can be larger than that of the vectors recoverable from ℓ_1-minimization. Then the question is what is the relationship between the sparsity of a vector and the successful recovery with ℓ_p-minimization $(p < 1)$? How sparse should a vector be so that ℓ_p-minimization can recover it? What is the threshold of sparsity that differentiates the success and failure of recovering by ℓ_p-minimization? [26] shows the sparsity up to which ℓ_p-minimization can successfully recover all the sparse vectors at least does not decrease as p decreases. [31] provides a sufficient condition for successful recovery via ℓ_p-minimization based on Restricted Isometry Constants

M. Wang, W. Xu and A. Tang are with School of ECE, Cornell University, Ithaca, NY, 14853. Email: {mw467, wx42}@cornell.edu, atang@ece.cornell.edu

and provides a lower bound of the support size up to which ℓ_p-minimization can recover all such sparse vectors. A later paper [13] improves this result by considering a modified Restricted p-Isometry Constant. [24] provides a lower bound of recovery threshold by considering a generalized version of RIP condition, and [4] numerically calculates this bound.

Here are the main contributions of this paper. For strong recovery where ℓ_p-minimization needs to recover all the vectors up to a certain sparsity, we provide a sharp threshold $\rho^*(p)$ of the ratio of the support size to the dimension which differentiates the success and the failure of ℓ_p-minimization when $\alpha(=\frac{m}{n}) \to 1$. This is an exact threshold compared with a lower bound of successful recovery in previous results. When p increases from 0 to 1, $\rho^*(p)$ decreases from 0.5 to 0.239. This coincides with the intuition that the performance of ℓ_p-minimization is improved when p decreases. When $\alpha \in (0,1)$ is fixed, we provide a positive bound $\rho^*(\alpha, p)$ for all $\alpha \in (0,1)$ and all $p \in (0,1]$ of strong recovery such that with a Gaussian measurement matrix $A^{m\times n}$, ℓ_p-minimization can recover all the $\rho^*(\alpha,p)n$-sparse vectors with overwhelming probability. $\rho^*(\alpha,p)$ improves on the existing bounds in large α region.

We also analyze the performance of ℓ_p-minimization for *weak* recovery where we need to recover all the sparse vectors on one support with one sign pattern. To the best of our knowledge, there is no existing result in this regard for $p<1$. We characterize the successful weak recovery through a necessary and sufficient condition regarding the null space of the measurement matrix. When $\alpha \to 1$, we provide a sharp threshold $\rho_w^*(p)$ of the ratio of the support size to the dimension which differentiates the success and the failure of ℓ_p-minimization. The weak threshold indicates that if we would like to recover every vector over one support with size less than $\rho_w^*(p)n$ and with one sign pattern, (though the support and sign patterns are not known a priori), and we generate a random Gaussian measurement matrix independently of the vectors, then with overwhelmingly high probability, ℓ_p-minimization will recover all such vectors regardless of the amplitudes of the entries of a vector. For ℓ_1-minimization, given a vector, if we randomly generate a Gaussian matrix and apply ℓ_1-minimization, then its recovering ability observed in simulation exactly captures the weak recovery threshold, see [17], [18]. Interestingly, when $\alpha \to 1$ and n is large enough, we prove that the weak threshold $\rho_w^*(p)$ is 2/3 for all $p \in [0,1)$, and is lower than the weak threshold of ℓ_1-minimization, which is 1. In this region, ℓ_1-minimization outperforms ℓ_p-minimization for all $p \in [0,1)$ if we only need to recover sparse vectors on one support with one sign pattern. We also explicitly show that ℓ_p-minimization ($p \in (0,1)$) can return a vector denser than the original sparse vector while ℓ_1-minimization successfully recovers the sparse vector. Finally, for every $\alpha \in (0,1)$, we provide a positive bound $\rho_w^*(\alpha,p)$ such that ℓ_p-minimization successfully recovers all the $\rho_w^*(\alpha,p)n$-sparse vectors on one support with one sign pattern.

The rest of the paper is organized as follows. We introduce the null space condition of successful ℓ_p-minimization in Section II. We especially define the successful weak recovery for $p<1$ and provide a necessary and sufficient condition. We use an example to illustrate that the solution of ℓ_1-minimization can be sparser than that of ℓ_p-minimization ($p \in (0,1)$). Section III provides thresholds of the sparsity ratio of the successful recovery via ℓ_p-minimization for all $p \in [0,1]$ both in strong recovery and in weak recovery when the measurement matrix is random Gaussian matrix and $\alpha \to 1$. For $\alpha \in (0,1)$, Section IV provides bounds of sparsity ratio below which ℓ_p-minimization is successful in the strong sense and in the weak sense respectively. We compare the performance of ℓ_p-minimization ($p<1$) and the performance of ℓ_1-minimization in Section V and provide numerical results in Section VI. Section VII concludes the paper. We only state the results in the main text and please refer to the Appendix for the proofs.

II. Successful Recovery of ℓ_p-minimization

We first introduce the null space characterization of the measurement matrix A to capture the successful recovery via ℓ_p-minimization ($p \in [0,1]$). Besides the strong recovery that has been studied in [4], [14], [24]–[26], [31], [33], we especially provide a necessary and sufficient condition for the success of *weak* recovery in the sense that ℓ_p-minimization only needs to recover all the sparse vectors on one support with one sign pattern. For example, in practice, given an unknown vector to recover, we randomly generate a measurement matrix and solve the ℓ_1-minimization problem, the simulation result of recovery performance with respect to the sparsity of the vector indeed represents the performance of weak recovery.

Given a measurement matrix $A^{m\times n}$, let $B^{n\times(n-m)}$ denote a matrix whose columns form a basis of the null space of A, then we have $AB = \mathbf{0}$. Let B_i ($i \in \{1,...,n\}$) denote the i^{th} row of B. Let B_T denote the submatrix of B with $T \subseteq \{1,...,n\}$ as the set of row indices. Let $T^c \subseteq \{1,...,n\}$ be the complimentary set of T. In this paper, we will study the sparse recovery property of ℓ_p-minimization by analyzing the null space of A.

We first state the null space condition for the success of strong recovery via ℓ_p-minimization [23][26] in the sense that ℓ_p-minimization should recover all the sparse vectors up to a certain sparsity.

Theorem 1 ([23][26])**.** *$\mathbf{x}$ is the unique solution to ℓ_p-minimization problem ($0 \le p \le 1$) for every vector $\mathbf{x}$ up to ρn-sparse if and only if*

$$\|B_T\mathbf{z}\|_p^p < \|B_{T^c}\mathbf{z}\|_p^p \tag{4}$$

for every non-zero $\mathbf{z} \in \mathcal{R}^{n-m}$, and every support T with $|T| \le \rho n$.

One important property is that if the condition (4) is satisfied for some $0 < p \le 1$, then it is also satisfied for all $q \in [0,p]$ [15][27]. Therefore, if ℓ_p-minimization could recover all the ρn-sparse vectors $\mathbf{x}$, then ℓ_q-minimization ($0 \le q \le p$) could also recover all the ρn-sparse vectors. Intuitively, the strong recovery performance of ℓ_q-minimization should be at least as good as that of ℓ_p-minimization when $0 \le q < p \le 1$.

A. Weak recovery for ℓ_p-minimization

Though ℓ_p-minimization ($p<1$) should be at least as good as ℓ_1-minimization for strong recovery, the argument may not be true for weak recovery. For weak recovery, we would like to recover all the vectors on some support T with some sign pattern σ, and $\sigma_i \in \{1,-1\}$ for every i in T. $\sigma_i = 1$ if a vector is positive on index i, and $\sigma_i = -1$ if a vector is negative on index i. Given any non-zero vector $\mathbf{z} \in \mathcal{R}^{n-m}$, we define $T^- := \{i \in T : B_i\mathbf{z}\sigma_i < 0\}$, $T^+ := \{i \in T : B_i\mathbf{z}\sigma_i > 0\}$, and $T^0 := \{i \in T : B_i\mathbf{z} = 0\}$. Note that when B is given, T^-, T^+ and T^0 depend on $\mathbf{z}$, and they can be empty. In this paper for weak recovery, we consider recovering nonnegative vectors on some support T for notational simplicity. In this case, T^- and T^+ are simplified to be $T^- = \{i \in T : B_i\mathbf{z} < 0\}$ and $T^+ = \{i \in T : B_i\mathbf{z} > 0\}$. However, all the results also hold for any specific support and any sign pattern.

We first state the null space condition for successful weak recovery via ℓ_1-minimization as follows, please see [21], [26], [32], [36], [38] for this result.

Theorem 2. *For every nonnegative $\mathbf{x} \in \mathcal{R}^n$ on some support T, $\mathbf{x}$ is the unique solution to ℓ_1-minimization problem (2) if and only if*

$$\|B_{T^-}\mathbf{z}\|_1 < \|B_{T^c}\mathbf{z}\|_1 + \|B_{T^+}\mathbf{z}\|_1 \tag{5}$$

holds for all non-zero $\mathbf{z} \in \mathcal{R}^{n-m}$.

Note that for every nonnegative vector $\mathbf{x}$ on a fixed support T, the condition to successfully recover it via ℓ_1-minimization is the same, as stated in Theorem 2. Therefore if one vector $\mathbf{x}$ can be successfully recovered, all the other nonnegative sparse vectors on T can also be recovered. Conversely, if some vector $\mathbf{x}$ cannot be successfully recovered, then every other nonnegative vector on T cannot be recovered either. However, the condition of successful recovery via ℓ_p-minimization ($0 \leq p < 1$) varies for different nonnegative sparse vectors even if they have the same support. In other words, the recovery condition depends on the amplitudes of the entries of the vector. Here we consider the worst case scenario for weak recovery in the sense that the recovery via ℓ_p-minimization is defined to be "successful" if it can recover *all* the nonnegative vectors on a fixed support. The null space condition for weak recovery in this definition via ℓ_1-minimization is still the same as that in Theorem 2. We characterize the ℓ_p-minimization ($p \in (0,1)$) case in Theorem 3 and the ℓ_0-minimization case in Theorem 4.

Theorem 3. *Given any $p \in (0,1)$, ℓ_p-minimization (3) can successfully recover all the nonnegative vectors $\mathbf{x} \in \mathcal{R}^n$ on some support T if and only if the following condition holds for every non-zero $\mathbf{z} \in \mathcal{R}^{n-m}$:*
if T^+ is not empty, then

$$\|B_{T^-}\mathbf{z}\|_p^p \leq \|B_{T^c}\mathbf{z}\|_p^p; \tag{6}$$

and if T^+ is empty, then

$$\|B_{T^-}\mathbf{z}\|_p^p < \|B_{T^c}\mathbf{z}\|_p^p.$$

Similarly, the null space condition for the weak recovery of ℓ_0-minimization is as follows, we skip its proof as it is similar to that of Theorem 3.

Theorem 4. *ℓ_0-minimization problem (1) can successfully recover all the nonnegative vectors $\mathbf{x} \in \mathcal{R}^n$ on support T, if and only if*

$$\|B_{T^-}\mathbf{z}\|_0 < \|B_{T^c}\mathbf{z}\|_0 \tag{7}$$

for all non-zero $\mathbf{z} \in \mathcal{R}^{n-m}$.

For the strong recovery, the null space conditions of ℓ_1-minimization and ℓ_p-minimization ($0 \leq p < 1$) share the same form (4), and if (4) holds for some $p \leq 1$, it also holds for all $q \in [0,p]$. However, for recovery of sparse vectors on one support with one sign pattern, from Theorem 2, 3 and 4, we know that although the conditions of ℓ_p-minimization ($0 < p < 1$) and ℓ_0-minimization share a similar form in (6) and (7), the condition of ℓ_1-minimization has a very different form in (5). Moreover, if (6) holds for some $p \in (0,1)$, it does not necessarily hold for all $q \in (0,p)$. Therefore the way that the performance of weak recovery changes over p may be quite different from the way that the performance of strong recovery changes over p. Moreover, the performance of weak recovery of ℓ_1 may be significantly different from that of ℓ_p-minimization for $p \in (0,1)$. We will further discuss this issue.

B. The solution of ℓ_1-minimization can be sparser than that of ℓ_p-minimization ($p \in (0,1)$)

ℓ_p-minimization ($p \in (0,1)$) may not perform as well as ℓ_1-minimization in some cases, for example in the weak recovery which we will discuss in Section III and Section IV. Here we employ a numerical example to illustrate that in certain cases ℓ_1-minimization can recover the sparse vector while ℓ_p-minimization ($p \in (0,1)$) cannot, and the solution of ℓ_p-minimization is denser than the original sparse vector.

Example 1. ℓ_p-minimization returns a denser solution than ℓ_1-minimization.

Let the measurement matrix A be a $(6k-1) \times 6k$ matrix with $\boldsymbol{\beta} \in \mathcal{R}^{6k}$ as a basis of its null space, and $\beta_i = 1$ for all $i \in \{1, ..., k\}$, $\beta_i = -1$ for all $i \in \{k+1, ..., 2k\}$, and $\beta_i = 1/64$ for all $i \in \{2k+1, ..., 6k\}$. Then every vector in the null space can be represented as $h\beta$, for some $h \in \mathcal{R}$. Note that $\|h\beta\|_1/2 = \frac{33k|h|}{32}$, and $\|h\beta_T\|_1 \leq (\lceil\frac{33}{32}k\rceil - 1)|h| < \|h\beta\|_1/2$ for all $T \subset \{1, ..., 6k\}$ with $|T| \leq (\lceil\frac{33}{32}k\rceil - 1)$ and for all $h \in \mathcal{R}$, and $\|h\beta_{\hat{T}}\|_1 = \lceil\frac{33}{32}k\rceil|h| \geq \|h\beta\|_1/2$ for all h if $\hat{T} = \{1, ..., \lceil\frac{33}{32}k\rceil\}$. Then according to Theorem 1, ℓ_1-minimization can recover all the $(\lceil\frac{33}{32}k\rceil - 1)$-sparse vectors in $\mathcal{R}^{6k}$, but fails to recover some $\lceil\frac{33}{32}k\rceil$-sparse vector. Similarly, $\|h\beta\|_{0.5}^{0.5}/2 = \frac{5k|h|}{4}$, and $\|h\beta_T\|_{0.5}^{0.5} \leq (\lceil\frac{5}{4}k\rceil - 1)|h| < \|h\beta\|_{0.5}^{0.5}/2$ for all $T \subset \{1, ..., 6k\}$ with $|T| \leq (\lceil\frac{5}{4}k\rceil - 1)$ and for all $h \in \mathcal{R}$, and $\|h\beta_{\hat{T}}\|_{0.5}^{0.5} = \lceil\frac{5}{4}k\rceil|h| \geq \|h\beta\|_{0.5}^{0.5}/2$ for all h if $\hat{T} = \{1, ..., \lceil\frac{5}{4}k\rceil\}$. Therefore by Theorem 1, $\ell_{0.5}$-minimization can recover all the $(\lceil\frac{5}{4}k\rceil - 1)$-sparse vectors in $\mathcal{R}^{6k}$, but fails to recover some $\lceil\frac{5}{4}k\rceil$-sparse vector. Therefore, in terms of strong recovery, $\ell_{0.5}$-minimization has a better performance than ℓ_1-minimization as it can recover all the vectors up to a higher sparsity.

Before discussing the weak recovery performance, we should first point out that when the null space is only one-dimensional, the ℓ_p-minimization problem for all $p \in (0,1]$ can be easily solved. Let $\mathbf{x}^*$ denote the sparse vector we would like to recover, and let $\tilde{\mathbf{x}}$ denote a vector that can produce the same measurements as $\mathbf{x}^*$, and mathematically, $A\tilde{\mathbf{x}} = A\mathbf{x}^*$. Then every vector $\mathbf{x}$ such that $A\mathbf{x} = A\mathbf{x}^*$ holds should satisfy $\mathbf{x} = \tilde{\mathbf{x}} + h\boldsymbol{\beta}$ for some $h \in \mathcal{R}$. Then the ℓ_p-minimization problem ($p \in (0, 1]$) is equivalent to

$$\min_{h\in\mathcal{R}} \|\tilde{\mathbf{x}} + h\boldsymbol{\beta}\|_p^p. \tag{8}$$

Given $\tilde{\mathbf{x}}$ and $\boldsymbol{\beta}$, $\|\tilde{\mathbf{x}} + h\boldsymbol{\beta}\|_p^p$ is a function of h. Define set $S = \{-\frac{\tilde{x}_i}{\beta_i} \mid \beta_i \neq 0\}$, let q denote the number of different elements in S, and let s_i ($i = 1, ..., q$) denote the ordered elements in S, and $s_i < s_j$ if $i < j$. Let I_0 denote the interval $(-\infty, s_1]$, let I_i denote the interval $[s_i, s_{i+1}]$ ($i = 1, ..., q-1$), and let I_q denote the interval $[s_q, +\infty)$. Note that for each interval I_i ($0 \leq i \leq q$), $\|\tilde{\mathbf{x}} + h\boldsymbol{\beta}\|_p^p$ is concave on I_i for every $p \in (0, 1)$, and $\|\tilde{\mathbf{x}} + h\boldsymbol{\beta}\|_1$ is linear on I_i. Therefore the minimum value of $\|\tilde{\mathbf{x}}+h\boldsymbol{\beta}\|_p^p$ ($p \in (0, 1]$) on I_i ($1 \leq i \leq q-1$) should be achieved at one of the endpoints of I_i, either s_i or s_{i+1}. Since when h goes to $-\infty$ or $+\infty$, $\|\tilde{\mathbf{x}} + h\boldsymbol{\beta}\|_p^p$ goes to $+\infty$, then the minimum value of $\|\tilde{\mathbf{x}} + h\boldsymbol{\beta}\|_p^p$ ($p \in (0, 1]$) on I_1 should be achieved at s_1, and the minimum value on I_{q+1} should be achieved at s_q. Thus, let $\mathbf{x}^i = \tilde{\mathbf{x}} + s_i\boldsymbol{\beta}$ for every $i = 1, ..., q$, and let $i^* := \arg\min_{1\leq i\leq q} \|\mathbf{x}^i\|_p^p$, then $\mathbf{x}^{i^*}$ is the solution to ℓ_p-minimization problem. We call $\mathbf{x}^i$'s as "singular vectors". Therefore, to solve (8), we only need to find all the singular vectors, and the one with the least ℓ_p quasi-norm (or ℓ_1 norm) is the solution to ℓ_p-minimization (or ℓ_1-minimization). If $\mathbf{x}^{i^*} = \mathbf{x}^*$, then we say $\mathbf{x}^*$ can be successfully recovered.

Now consider the "weak" recovery as to recover all the nonnegative vectors on support $T = \{1, ..., 2k\}$. According to Theorem 2 and Theorem 3, one can check that ℓ_1-minimization can indeed recover all the nonnegative vectors on support T, however, $\ell_{0.5}$-minimization fails to recover some vectors in this case. For example, consider a $2k$-sparse vector $\mathbf{x}^*$ with $x_i^* = 9$ for all $i \in \{1, ..., k\}$, $x_i^* = 1$ for all $i \in \{k+1, ..., 2k\}$, and $x_i^* = 0$ for all $i \in \{2k + 1, ..., 6k\}$. There are three singular vectors in this case: $\mathbf{x}^1 = \mathbf{x}^*$, $\mathbf{x}^2 = \mathbf{x}^* + \beta$ and $\mathbf{x}^3 = \mathbf{x}^* - 9\beta$. Since $\|\mathbf{x}^1\|_1 = 10k$, $\|\mathbf{x}^2\|_1 = 10k + k/16$, and $\|\mathbf{x}^3\|_1 = 10k + 9k/16$, then $\mathbf{x}^1$ is the solution of ℓ_1-minimization, and $\mathbf{x}^*$ is successfully recovered. Now consider $\ell_{0.5}$-minimization, since $\|\mathbf{x}^1\|_{0.5}^{0.5} = 4k$, $\|\mathbf{x}^2\|_{0.5}^{0.5} = (\sqrt{10} + 0.5)k$, and $\|\mathbf{x}^3\|_{0.5}^{0.5} = (\sqrt{10} + 1.5)k$, then $\mathbf{x}^2$ is the solution of $\ell_{0.5}$-minimization, and it is $5k$-sparse. Thus, the solution of $\ell_{0.5}$-minimization is a $5k$-sparse vector although the original vector $\mathbf{x}^*$ is only $2k$-sparse. Therefore $\ell_{0.5}$-minimization fails to recover some nonnegative $2k$-sparse vector $\mathbf{x}^*$ while $\mathbf{x}^*$ is the solution to ℓ_1-minimization, and the solution of $\ell_{0.5}$-minimization is denser than the original vector $\mathbf{x}^*$.

III. Recovery thresholds when $\frac{m}{n} \to 1$

In this paper we focus on the case that the measurement matrix A has i.i.d. standard Gaussian $\mathcal{N}(0, 1)$ entries. Then for a matrix $B^{n\times(n-m)}$ with i.i.d. $\mathcal{N}(0, 1)$ entries, the column space of B is equivalent in distribution to the null space of A, please refer to [8][37] for details. Then in later analysis, we will use B to represent a basis of the null space of A.

We first focus on the case that $\alpha = \frac{m}{n} \to 1$ and provide recovery thresholds of ℓ_p-minimization for every $p \in [0, 1]$. We consider two types of thresholds: one in the *strong* sense as we require ℓ_p-minimization to recover *all* ρn-sparse vectors (Section III-A), one in the *weak* sense as we only require ℓ_p-minimization to recover *all the vectors on a certain support with a certain sign pattern* (Section III-B). Since in our setup the measurement matrix A has i.i.d. $\mathcal{N}(0, 1)$ entries, the weak recovery performance does not depend on the specific choice of the support and the sign pattern. We call it a threshold as for any sparsity below that threshold, ℓ_p-minimization can recover all the sparse vectors either in the strong sense or the weak sense when α is close enough to 1 and n is large enough, and for any sparsity above that threshold, ℓ_p-minimization fails to recover some sparse vector no matter how large α and n are. These thresholds can be viewed as the limiting behavior of ℓ_p-minimization, since for any constant $\alpha \in (0, 1)$, the recovery thresholds of ℓ_p-minimization would be no greater than the ones provided here.

A. Strong Recovery

In this section, for given p, we shall provide a threshold $\rho^*(p)$ of *strong recovery* such that for any $\rho < \rho^*(p)$, ℓ_p-minimization (3) can recover *all* ρn-sparse vectors $\mathbf{x}$ with overwhelming probability when α is close enough to 1. Our technique here stems from [22], which only focuses on the strong recovery of ℓ_1-minimization.

We have already discussed in Section II that the performance of ℓ_q-minimization should be no worse than ℓ_p-minimization for strong recovery when $0 \leq q < p \leq 1$. Although there are results about bound of the sparsity below which ℓ_p-minimization can recover all the sparse vectors, no existing result has explicitly calculated the recovery threshold of ℓ_p-minimization for $p < 1$ which differentiates the success and failure of ℓ_p-minimization. To this end, we will first define $\rho^*(p)$ in the following lemma, and then prove that $\rho^*(p)$ is indeed the threshold of strong recovery in later part.

Lemma 1. *Let X_1, X_2,...,X_n be i.i.d. $\mathcal{N}(0, 1)$ random variables and let Y_1, Y_2,...,Y_n be the sorted ordering (in non-increasing order) of $|X_1|^p$, $|X_2|^p$,...,$|X_n|^p$ for some $p \in (0, 1]$. For given $\rho > 0$, define S_ρ as $\sum_{i=1}^{\lceil\rho n\rceil} Y_i$. Let S denote $E[S_1]$, the expected value of S_1. Then there exists a constant $\rho^*(p)$ such that $\lim_{n\to\infty} \frac{E[S_{\rho^*}]}{S} = \frac{1}{2}$.*

ρ^* is a function of p, and in fact is strictly decreasing as stated in Proposition 1.

Proposition 1. *The function $\rho^*(p)$ is strictly decreasing in p on $(0, 1]$.*

Note that $\rho^*(p)$ goes to $\frac{1}{2}$ as p tends to zero from (13) and (14). We plot ρ^* against p numerically in Fig. 1. We also obtain that $\rho^*(1) = 0.239...$, which coincides with the result in [22].

Now we proceed to prove that ρ^* is the threshold of successful recovery with ℓ_p-minimization for p in $(0, 1]$. First

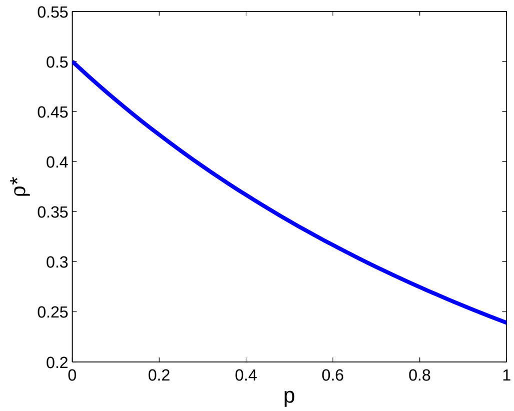


Fig. 1. Threshold ρ^* of successful recovery with ℓ_p-minimization

we state the concentration property of S_ρ in the following lemma.

Lemma 2. *For any $p \in (0,1]$, let X_1,...,X_n, Y_1,...,Y_n, S_ρ and S be as in Lemma 1. For any $\rho > 0$ and any $\delta > 0$, there exists a constant $c_1 > 0$ such that when n is large enough, with probability at least $1 - 2e^{-c_1 n}$, $|S_\rho - E[S_\rho]| \leq \delta S$.*

Roughly speaking, Lemma 2 states that S_ρ is concentrated around its expectation $E[S_\rho]$ for every ρ. For our purpose in this paper, the following two corollaries of Lemma 2 are important for the later proof.

Corollary 1. *For any $\rho < \rho^*$, there exists a $\delta > 0$ and a constant $c_2 > 0$ such that when n is large enough, with probability at least $1 - 2e^{-c_2 n}$, $S_\rho \leq (\frac{1}{2} - \delta)S$.*

Corollary 2. *For any $\epsilon > 0$, there exists a constant $c_3 > 0$ such that when n is large enough, with probability at least $1 - 2e^{-c_3 n}$, it holds that $(1-\epsilon)S \leq S_1 \leq (1+\epsilon)S$.*

From the above two corollaries and applying the union bound, one can easily show that with overwhelming probability the sum of the largest $\lceil \rho n \rceil$ terms of Y_i's is less than half of the total sum S_1 if $\rho < \rho^*$. The following lemma extends the result to all the vectors $B\mathbf{z}$ simultaneously where matrix $B^{n\times(n-m)}$ has i.i.d. Gaussian entries and $\mathbf{z}$ is any non-zero vector in $\mathcal{R}^{n-m}$.

Lemma 3. *For any $0 < p \leq 1$, given any $\rho < \rho^*(p)$, there exist constants $0 < c_4 < 1$, $c_5 > 0$, $\delta > 0$ such that when $\alpha = \frac{m}{n} > c_4$ and n is large enough, with probability at least $1 - e^{-c_5 n}$, an $n \times (n-m)$ matrix B with i.i.d. $\mathcal{N}(0,1)$ entries has the following property: for every non-zero $\mathbf{z} \in \mathcal{R}^{n-m}$ and every subset $T \subseteq \{1,...,n\}$ with $|T| \leq \lceil \rho n \rceil$, $\|B_{T^c}\mathbf{z}\|_p^p - \|B_T\mathbf{z}\|_p^p \geq \delta S \|\mathbf{z}\|_2^p$.*

We remark here that in Lemma 3 and all the following results in this paper, when we say "with probability at least $1 - e^{-cn}$ for some constant $c > 0$", by "constant" we mean c does not depend on the measurement matrix A, but c could depend on other parameters in various occasions.

Lemma 3 indicates that when $\alpha > c_4$ and n is large enough, with overwhelming probability $\sum_{i\in T^c}|(B\mathbf{z})_i|^p - \sum_{i\in T}|(B\mathbf{z})_i|^p \geq \delta S \|\mathbf{z}\|_2^p > 0$ holds for every non-zero $\mathbf{z}$ and every set T with $|T| \leq \lceil \rho n \rceil$, then from Theorem 1, in this case every $\lceil \rho n \rceil$-sparse vector $\mathbf{x}$ is the unique solution to the ℓ_p-minimization problem (3) with overwhelming probability. We can now establish one main result regarding the threshold of successful recovery via ℓ_p-minimization.

Theorem 5. *For any $0 < p \leq 1$, given any $\rho < \rho^*(p)$, there exist constants $0 < c_4 < 1$, $c_5 > 0$ such that when $\alpha > c_4$ and n is large enough, with probability at least $1 - e^{-c_5 n}$, an $m \times n$ matrix A with i.i.d. $\mathcal{N}(0,1)$ entries has the following property: for every $\mathbf{x} \in \mathcal{R}^n$ with its support T satisfying $|T| \leq \lceil \rho n \rceil$, $\mathbf{x}$ is the unique solution to the ℓ_p-minimization problem (3).*

We remark here that $\rho^*(p)$ is a sharp bound for successful recovery. For any $\rho > \rho^*(p)$, from Lemma 1 and Lemma 2, for any $\mathbf{z}$ in $\mathcal{R}^{n-m}$, with overwhelming probability the sum of the largest $\lceil \rho n \rceil$ terms of $|B_i\mathbf{z}|^p$'s is more than the half of the total sum S_1, i.e. the null space condition stated in Theorem 1 for successful recovery via ℓ_p-minimization fails with overwhelming probability. Therefore, ℓ_p-minimization fails to recover some ρn-sparse vector with overwhelming probability if $\rho > \rho^*(p)$. Proposition 1 implies that the threshold strictly decreases as p increases. The performance of ℓ_{p_1}-minimization is better than that of ℓ_{p_2}-minimization for $0 < p_1 < p_2 \leq 1$ in strong recovery as ℓ_{p_1}-minimization can recover vectors up to a higher sparsity.

B. Weak Recovery

We have demonstrated in Section III-A that the threshold for strong recovery strictly decreases as p increases from 0 to 1. Here we provide a weak recovery threshold for all $p \in [0,1)$ when $\alpha \to 1$. As we shall see, for weak recovery, the threshold of ℓ_p-minimization is the same for all $p \in [0,1)$, and is lower than the threshold of ℓ_1-minimization.

Recall that for successful weak recovery, ℓ_p-minimization should recover all the vectors on some fixed support with a fixed sign pattern, and the equivalent null space characterization is stated in Theorem 3 and Theorem 4.

Note that to simply the notation, for the remaining part of the paper, we will say a vector is ρn-sparse or the size of the support is ρn instead of using the notation $\lceil \rho n \rceil$. However, the support size should always be an integer.

We define $x^0 = 1$ for all $x \neq 0$, and $0^0 = 0$. To characterize the recovery threshold of ℓ_p-minimization in this case, we first state the following lemma.

Lemma 4. *Let X_1, X_2,...,X_n be i.i.d. $\mathcal{N}(0,1)$ random variables and T be a set of indices with size $|T| = \rho n$ for some $\rho > 0$. For every $p \in [0,1)$, for every $\epsilon > 0$, when n is large enough, with probability at least $1 - e^{-c_6 n}$ for some constant $c_6 > 0$, the following two properties hold simultaneously:*

- $\frac{1}{2}\rho n(\mu - \epsilon) < \sum_{i\in T: X_i<0}|X_i|^p < \frac{1}{2}\rho n(\mu + \epsilon)$
- $(1-\rho)n(\mu - \epsilon) < \sum_{i\in T^c}|X_i|^p < (1-\rho)n(\mu + \epsilon)$.

where $\mu = E[|X|^p]$, $X \sim \mathcal{N}(0,1)$.

The proof of Lemma 4 is based on concentration of measure, and the arguments are similar to those in the proof of Lemma 2. Lemma 4 implies that $\sum_{i\in T: X_i<0}|X_i|^p <$

$\sum_{i\in T^c}|X_i|^p$ holds with high probability when $|T| = \rho n < \frac{2}{3}n$. Applying the net arguments similar to those in the proof of Lemma 3, we can also show that with overwhelming probability the statement holds for all vectors $B\mathbf{z}$ simultaneously where matrix $B^{n\times(n-m)}$ has i.i.d. Gaussian entries and $\mathbf{z}$ is any non-zero vector in $\mathcal{R}^{n-m}$. Then we can establish the main result regarding the threshold of successful recovery with ℓ_p-minimization from vectors on one support with the same sign pattern.

Theorem 6. *For any $p \in [0, 1)$, given any $\rho < \rho_w^* := \frac{2}{3}$, there exist constants $c_7 \in (0, 1)$, $c_8 > 0$ such that when $\alpha > c_7$ and n is large enough, with probability at least $1-e^{-c_8 n}$, an $m\times n$ matrix A with i.i.d. $\mathcal{N}(0, 1)$ entries has the following property: for every nonnegative vector $\mathbf{x}$ on some support T satisfying $|T| \leq \rho n$, $\mathbf{x}$ is the unique solution to the ℓ_p-minimization problem.*

We remark here that ρ_w^* is a sharp bound for successful recovery in this setup. For any $\rho > \rho_w^*$, from Lemma 4, with overwhelming probability that $\sum_{i\in T:B_i\mathbf{z}<0}|B_i\mathbf{z}|^p > \sum_{i\in T^c}|B_i\mathbf{z}|^p$, then Theorem 3 and Theorem 4 indicate that the ℓ_p-minimization ($p \in [0, 1)$) fails to recover some nonnegative ρn-sparse vector $\mathbf{x}$ on T in this case. Note that for a random Gaussian measurement matrix, from symmetry one can check that this result does not depend on the specific choice of the support and the sign pattern. In fact, ρ_w^* in Theorem 6 is the weak recovery threshold for any fixed support and any fixed sign pattern.

Surprisingly, the successful recovery threshold ρ_w^* when we only consider recovering vectors on one support with one sign pattern is $\frac{2}{3}$ for all p in $[0, 1)$ and is strictly less than the threshold for $p = 1$, which is 1 [17]. Thus in this case, ℓ_1-minimization has better recovery performance than ℓ_p-minimization ($p \in [0, 1)$) in terms of the sparsity requirement for the sparse vector. Although the strong recovery performance can be improved if we apply ℓ_p-minimization with a smaller p, ℓ_1-minimization can indeed outperform ℓ_p-minimization for all $p \in [0, 1)$ in weak recovery if α is close to 1 and n is large enough.

It might be counterintuitive at first sight to see that the weak threshold of ℓ_0-minimization is less than that of ℓ_1-minimization, so let us take a moment to consider what the result means. We choose recovering all nonnegative vectors on some support T ($|T| = \rho n$) for the weak recovery, the argument follows for all the other supports and all the other sign patterns. The results about weak recovery threshold indicate that for any $\rho \in (2/3, 1)$, when n is sufficiently large and α is close enough to 1, for a random Gaussian measurement matrix A, ℓ_1-minimization would recover all the nonnegative vectors on support T with overwhelming probability, while ℓ_0-minimization would fail to recover some nonnegative vector on T with overwhelming probability. The failure of ℓ_0-minimization indicates that there exists a nonnegative vector $\mathbf{x}$ on support T and a vector $\mathbf{x}'$ on support T' such that $|T'| \leq |T|$, and $A\mathbf{x} = A\mathbf{x}'$. Note that $\mathbf{x}'$ could have negative entries, or T' may not be a subset of T. Therefore, if $\mathbf{x}$ is the sparse vector we would like to recover from $A\mathbf{x}$, ℓ_0-minimization would fail since $\|\mathbf{x}'\|_0 \leq \|\mathbf{x}\|_0$. However, $\|\mathbf{x}\|_1 < \|\mathbf{x}'\|_1$ should hold since ℓ_1-minimization can successfully return $\mathbf{x}$ as its solution. Of course when $\mathbf{x}'$ is the sparse vector we would like to recover, ℓ_1-minimization would return $\mathbf{x}$ and fail to recover $\mathbf{x}'$. However, since ℓ_1-minimization would recover all the nonnegative vectors on T, then either $T' \nsubseteq T$ holds or $\mathbf{x}'$ has negative entries. Therefore when we consider recovering nonnegative vectors on T for the weak recovery, $\mathbf{x}'$ is not taken into account, and ℓ_1-minimization works better than ℓ_0-minimization. Thus, although the performance of ℓ_1-minimization is not as good as that of ℓ_p-minimization ($p \in [0, 1)$) in the strong recovery which requires to recover all the vectors up to certain sparsity, ℓ_1-minimization can recover all the ρn-sparse ($\rho > 2/3$) vectors on some support with some sign pattern, while for ℓ_p-minimization ($p \in [0, 1)$), the size of the largest support on which it can recover all the vectors with one sign pattern is no greater than $2n/3$. In a word, when we aim to recover all the vectors up to certain sparsity, ℓ_p-minimization is better for smaller p, however, when we aim to recover all the vectors on one support with one sign pattern, ℓ_1-minimization may have a better performance.

IV. Recovery Bounds for Fixed $\frac{m}{n}$

We considered the limiting case that $\alpha \to 1$ in Section III and provided the limiting thresholds of sparsity ratio for successful recovery via ℓ_p-minimization both in the strong sense and in the weak sense. Here we focus on the case that α is fixed ($0 < \alpha < 1$). For any α and p, we will provide a bound $\rho^*(\alpha, p)$ for strong recovery and a bound $\rho_w^*(\alpha, p)$ for weak recovery such that ℓ_p-minimization can recover all the $\rho^*(\alpha, p)n$-sparse vectors with overwhelming probability, and recover all the $\rho_w^*(\alpha, p)n$-sparse vectors on one support with one sign pattern with overwhelming probability. Note that the thresholds we provided in Section III is tight in the sense that for any $\rho > \rho^*$ in the strong recovery or any $\rho > \rho_w^*$ in the weak recovery, with overwhelming probability ℓ_p-minimization would fail to recover some ρn-sparse vector. However, $\rho^*(\alpha, p)$ and $\rho_w^*(\alpha, p)$ we provide in this section are lower bounds for the thresholds of strong recovery and weak recovery respectively, and might not be tight in general.

A. Strong Recovery

From Theorem 1 we know that in order to successfully recover all the ρn-sparse vectors via ℓ_p-minimization, $\|B_T\mathbf{z}\|_p^p < \frac{1}{2}\|B\mathbf{z}\|_p^p$ should hold for every non-zero vector $\mathbf{z} \in \mathcal{R}^{n-m}$, and every set $T \subset \{1, ..., n\}$ with $|T| \leq \rho n$. The key idea to obtain a lower bound $\rho^*(\alpha, p)$ is as follows. We first calculate a lower bound of $\|B\mathbf{z}\|_p^p$ for all $\mathbf{z}$ in $\mathcal{S}$, where $\mathcal{S}$ is the unit sphere in $\mathcal{R}^{n-m}$. Then for any ρ, we calculate an upper bound of $\|B_T\mathbf{z}\|_p^p$ for all T with $|T| = \rho n$ and all $\mathbf{z}$ in $\mathcal{S}$. Then we define $\rho^*(\alpha, p)$ to be the largest ρ such that the aforementioned upper bound is less than half of the lower bound. According to Theorem 1, ℓ_p-minimization is now guaranteed to recover all the $\rho^*(\alpha, p)n$-sparse vectors. The problem regarding characterizing the lower bound and the upper bound here is that B has i.i.d. $\mathcal{N}(0, 1)$ entries, and therefore for any $\mathbf{z} \in \mathcal{S}$ and any T and for any constant $c > 0$, there always exist a positive probability that $B\mathbf{z}$ is less than c, and similarly a positive probability that $B_T\mathbf{z}$ is greater than c. Thus, strictly speaking, no finite value would be a lower bound of $\|B\mathbf{z}\|_p^p$, nor an upper bound of $\|B_T\mathbf{z}\|_p^p$. To address this issue, we will look for a "lower bound" of $\|B\mathbf{z}\|_p^p$ for all $\mathbf{z}$ in $\mathcal{S}$ in Lemma 5 in the sense that the violation probability decays to zero exponentially, and likewise an "upper bound" of $\|B_T\mathbf{z}\|_p^p$ for all T with $|T| = \rho n$ and all $\mathbf{z}$ in $\mathcal{S}$ in Lemma 6 such that the probability it is exceeded decays exponentially to zero. We want the "lower bound (upper bound)" to be as large (small) as possible as long as its violation probability has exponential decay to zero, and we do not focus on the decay rate here. We still define $\rho^*(\alpha, p)$ to be the largest ρ such that the "upper bound" is less than half of the "lower bound". We then show in Theorem 7 that ℓ_p-minimization can recover all the $\rho^*(\alpha, p)n$-sparse vectors with overwhelming probability.

Lemma 5. *For any α and p, there exists a constant $\lambda_{\min}(\alpha, p) > 0$ and some constant $c_9 > 0$ such that with probability at least $1 - e^{-c_9 n}$, for every $\mathbf{z} \in \mathcal{S}$, $\|B\mathbf{z}\|_p^p > \lambda_{\min}(\alpha, p)n$.*

Lemma 6. *Given any α, p and corresponding $\lambda_{\min}(\alpha, p) > 0$, there exists a constant $\rho^*(\alpha, p) > 0$ and some constant $c_{10} > 0$ such that with probability at least $1 - e^{-c_{10} n}$, for every $\mathbf{z} \in \mathcal{S}$ and for every set $T \subset \{1, 2, ..., n\}$ with $|T| \le \rho^*(\alpha, p)n$, $\|B_T\mathbf{z}\|_p^p < \frac{1}{2}\lambda_{\min}(\alpha, p)n$.*

Together with Lemma 5 and Lemma 6, we are ready to present our result on bounds for strong recovery of ℓ_p-minimization with given $\alpha \in (0, 1)$.

Theorem 7. *For any $0 < p \le 1$, any $0 < \alpha < 1$, for matrix $A^{m \times n}$ ($\alpha = \frac{m}{n}$) with i.i.d. $\mathcal{N}(0, 1)$ entries, there exists a constant $c_{11} > 0$ such that with probability at least $1 - e^{-c_{11} n}$, $\mathbf{x}$ is the unique solution to the ℓ_p-minimization problem (3) for every vector $\mathbf{x}$ up to $\rho^*(\alpha, p)n$-sparse.*

Theorems 7 implies that for every $\alpha \in (0, 1)$ and every $p \in (0, 1]$, there exists a positive constant $\rho^*(\alpha, p)$ such that ℓ_p-minimization can recover all the $\rho^* n$-sparse vectors with overwhelming probability. Since $\rho^*(\alpha, p)$ is a lower bound of the threshold of the strong recovery, we would like the lower bound to be as high as possible. Clearly, the value of $\rho^*(\alpha, p)$ depends on the "lower bound" of $\|B\mathbf{z}\|_p^p$ and the "upper bound" of $\|B_T\mathbf{z}\|_p^p$ with $|T| = \rho n$ for a given ρ. In order to improve $\rho^*(\alpha, p)$, we need to improve the "lower bound" of $\|B\mathbf{z}\|_p^p$ and the "upper bound" of $\|B_T\mathbf{z}\|_p^p$. Therefore, besides establishing the existence of "lower (upper) bound", we make some efforts to increase (decrease) the "lower (upper) bound" while making sure that the probability of violating these bounds has exponential decay to zero. To be more specific, we first calculate $\lambda_{\min}(\alpha, p)$ in Lemma 5 as a "lower bound" of $\|B\mathbf{z}\|_p^p$. The key idea is as follows. Given any constant $b > 0$, we characterize the probability that $\|B\mathbf{z}\|_p^p \le bn$ holds for some $\mathbf{z} \in \mathcal{S}$ by techniques like γ-net arguments, the Chernoff bound and the union bound. Then $\lambda_{\min}(\alpha, p)$ is chosen to be the largest value b such that the probability still maintains exponential decay to zero. With the obtained $\lambda_{\min}(\alpha, p)$, we next calculate $\rho^*(\alpha, p)$ in Lemma 6. The idea is similar to that in calculating $\lambda_{\min}(\alpha, p)$. For any given $\rho > 0$, we calculate an upper bound of the probability that there exists some $\mathbf{z} \in \mathcal{S}$ and some support T with $|T| = \rho n$ such that $\|B_T\mathbf{z}\|_p^p \ge \lambda_{\min}(\alpha, p)n/2$. Then $\rho^*(\alpha, p)$ is chosen to be the largest ρ such that the probability still has exponential decay to zero. Please refer to Appendix-J and Appendix-K for the detailed calculation of $\lambda_{\min}(\alpha, p)$ and $\rho^*(\alpha, p)$.

We numerically compute $\rho^*(\alpha, p)$ by calculating first $\lambda_{\max}(\alpha, p)$ in Lemma 9 from (43), and then $\lambda_{\min}(\alpha, p)$ in Lemma 5 from (53), and finally $\rho^*(\alpha, p)$ in Lemma 6 from (58). Fig. 2 shows the curve of $\rho^*(\alpha, p)$ against α for different p, and Fig. 3 shows the curve of $\rho^*(\alpha, p)$ against p for different α. Note that for any p, $\lim_{\alpha \to 1} \rho^*(\alpha, p)$ is slightly smaller than the limiting threshold of strong recovery we obtained in Section III-A. For example, when $p = 0.5$, the threshold $\rho^*(0.5)$ we obtained in Section III-A is 0.3406, and the bound $\rho^*(\alpha, 0.5)$ we obtained here is approximately 0.268 when α goes to 1. This is because in Section III-A we employed a finer technique to characterize the sum of the largest ρn terms of n i.i.d. random variables directly, while in Section IV-A introducing the union bound causes some slackness.

Compared with the bound obtained in [4] through restricted isometry condition, our bound $\rho^*(\alpha, p)$ is tighter when α is relatively large. For example, when $p = 1$, the bound in [4] (Fig.3.2(a)) is in the order of 10^{-3} for all $\alpha \in (0, 1)$ and upper bounded by 0.0035, while $\rho^*(\alpha, 1)$ is greater than 0.0039 for all $\alpha \ge 0.8$ and increases to 0.1308 as $\alpha \to 1$. When $p = 0.5$, the bound in [4] (Fig.3.2(c)) is in the order of 10^{-3} for all $\alpha \in (0, 1)$ and upper bounded by 0.01, while here $\rho^*(\alpha, 0.5)$ is greater than 0.011 for all $\alpha \ge 0.65$ and increases to 0.268 as $\alpha \to 1$. Therefore, although [4] provides a better bound than ours when α is small, our bound ρ^* improves over that in [4] when α is relatively large.

[13] provides a lower bound of strong recovery threshold for every α and very p. For example, it shows that when n is large enough, ℓ_0-minimization can recover all the $\frac{\alpha n}{119}$-sparse vectors for given α. Their result is better than ours when α is small. However, our bound is higher than that in [13] when α is large. For example, when $\alpha = 0.5$, [13] indicates that a lower bound of recovery threshold in terms of the ratio of sparsity to the dimension n is $0.5/119 \approx 0.004$ for ℓ_0-minimization. Our result shows that $\rho^*(0.5, 0.7)$ is already 0.004, and $\rho^*(0.5, 0.1)$ is as high as 0.0379, which is approximately ten times the bound $0.5/119$ in [13].

[17] applies geometric face counting technique to the strong bound of successful recovery of ℓ_1-minimization (Fig.1.1). Since if the necessary and sufficient condition (4) is satisfied for $p = 1$, then it is also satisfied for all $p < 1$, therefore the bound in [19] can serve as the bound of successful recovery for all $0 < p < 1$. Our bound $\rho^*(\alpha, p)$ in Section IV is higher than that in [17] when α is relatively large.

B. Weak Recovery

Theorem 3 provides a sufficient condition for successful recovery of every nonnegative ρn-sparse vector $\mathbf{x}$ on one support T, which requires $\|B_{T^-}\mathbf{z}\|_p^p < \|B_{T^c}\mathbf{z}\|_p^p$ to hold for all non-zero $\mathbf{z} \in \mathcal{R}^{n-m}$, where given $\mathbf{z}$, $T^- = \{i : B_i\mathbf{z} < 0\}$.

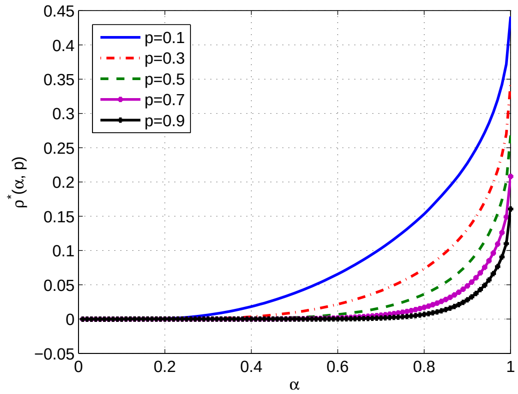


Fig. 2. $\rho^*(\alpha, p)$ against α for different p

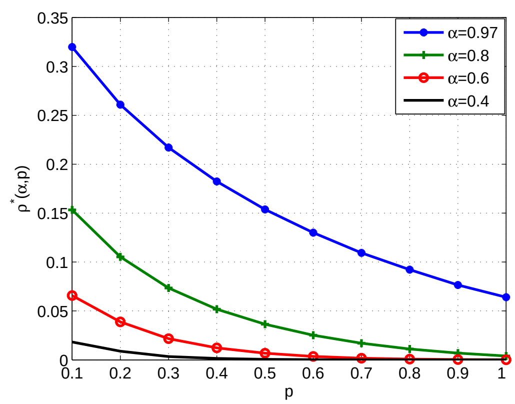


Fig. 3. $\rho^*(\alpha, p)$ against p for different α

We will use arguments similar to those in Section IV-A to obtain a lower bound $\rho_w^*(\alpha, p)$ of the weak recovery threshold. Given α, p and $\rho \in (0, 1)$, we will establish a "lower bound" of $\|B_{T^c}\mathbf{z}\|_p^p$ for all $\mathbf{z} \in \mathcal{S}$ in Lemma 7 in the sense that the violation probability of this "lower bound" decays exponentially to zero, and likewise establish an "upper bound" of $\|B_{T^-}\mathbf{z}\|_p^p$ in Lemma 8. If there exists $\rho_w^*(\alpha, p) > 0$ such that the corresponding "lower bound" of $\|B_{T^c}\mathbf{z}\|_p^p$ is greater than the "upper bound" of $\|B_{T^-}\mathbf{z}\|_p^p$, then $\rho_w^*(\alpha, p)$ serves as a lower bound of recovery threshold of ℓ_p-minimization for vectors on a fixed support with a fixed sign pattern.

The techniques used to establish the "lower bound" of $\|B_{T^c}\mathbf{z}\|_p^p$ for all $\mathbf{z} \in \mathcal{S}$ is the same as that in Lemma 5. We state the result in Lemma 7, please refer to Appendix-M for its proof.

Lemma 7. *Given α, p and set $T \subset \{1, ..., n\}$ with $|T| = \rho n$, with probability at least $1 - e^{-c_{12}n}$ for some $c_{12} > 0$, for all $\mathbf{z} \in \mathcal{S}$, $\|B_{T^c}\mathbf{z}\|_p^p < (1-\rho)\lambda_{\max}(\frac{\alpha-\rho}{1-\rho}, p)n$, and with probability at least $1 - e^{-c_{13}n}$ for some $c_{13} > 0$, for all $\mathbf{z} \in \mathcal{S}$, $\|B_{T^c}\mathbf{z}\|_p^p > (1-\rho)\lambda_{\min}(\frac{\alpha-\rho}{1-\rho}, p)n$, where $\lambda_{\max}(\alpha, p)$ and $\lambda_{\min}(\alpha, p)$ are defined in (43) and (53) respectively.*

Given T with $|T| = \rho n$, Lemma 7 provides a "lower bound" of $\|B_{T^c}\mathbf{z}\|_p^p$ which holds with overwhelming probability for all $\mathbf{z} \in \mathcal{S}$. Next we will provide an "upper bound" of $\|B_{T^-}\mathbf{z}\|_p^p$ for all $\mathbf{z} \in \mathcal{S}$ in Lemma 8. One should be cautious that the set T^- varies for different $\mathbf{z}$.

Lemma 8. *Given α, p and set $T \subset \{1, ..., n\}$ with $|T| = \rho n$, with probability at least $1 - e^{-c_{14}n}$ for some $c_{14} > 0$, for every $\mathbf{z} \in \mathcal{S}$, $\|B_{T^-}\mathbf{z}\|_p^p < \rho\tilde{\lambda}_{\max}(\alpha, p, \rho)n$, for some $\tilde{\lambda}_{\max}(\alpha, p, \rho) > 0$.*

To improve the lower bound of weak recovery threshold, given ρ, we want $\tilde{\lambda}_{\max}(\alpha, p, \rho)$ in Lemma 8 to be as small as possible while at the same time the probability that $\|B_{T^-}\mathbf{z}\|_p^p \geq \rho\tilde{\lambda}_{\max}(\alpha, p, \rho)n$ for some T with $|T| = \rho n$ and some $\mathbf{z}$ in $\mathcal{S}$ still has exponential decay to zero. Efforts are made in Appendix-N to improve $\tilde{\lambda}_{\max}(\alpha, p, \rho)$, which can be computed from (70).

With the help of Lemma 7 and Lemma 8, we are ready to present the result regarding the lower bound of recovery threshold via ℓ_p-minimization in the weak sense for given α.

Theorem 8. *For any $0 < p \leq 1$, any $0 < \alpha < 1$, for matrix $A^{m\times n}$ ($m = \alpha n$) with i.i.d. $\mathcal{N}(0, 1)$ entries, there exist constants $\rho_w^*(\alpha, p) > 0$ and $c_{15} > 0$ such that with probability at least $1 - e^{-c_{15}n}$, $\mathbf{x}$ is the unique solution to the ℓ_p-minimization problem (3) for every nonnegative $\rho_w^*(\alpha, p)n$-sparse vector $\mathbf{x}$ on fixed support T.*

Theorem 8 establishes the existence of a positive bound $\rho_w^*(\alpha, p)$ of weak recovery threshold. To obtain $\rho_w^*(\alpha, p)$, for every ρ we first calculate $\lambda_{\min}(\frac{\alpha-\rho}{1-\rho}, p)$ in Lemma 7 from (53) to obtain a "lower bound" of $\|B_{T^c}\mathbf{z}\|_p^p$ for all $\mathbf{z}$ in $\mathcal{S}$ and calculate $\tilde{\lambda}_{\max}(\alpha, p, \rho)$ in Lemma 8 from (70) to obtain an "upper bound" of $\|B_{T^-}\mathbf{z}\|_p^p$ for all $\mathbf{z}$ in $\mathcal{S}$. We then find the largest $\rho_w^*(\alpha, p)$ such that the "lower bound" of $\|B_{T^c}\mathbf{z}\|_p^p$ is larger than the "upper bound" of $\|B_{T^-}\mathbf{z}\|_p^p$, or mathematically, (71) holds. We numerically calculate this bound and illustrate the results in Fig. 4 and Fig. 5. Fig. 4 shows the curve of $\rho_w^*(\alpha, p)$ against α for different p, and Fig. 5 shows the curve of $\rho_w^*(\alpha, p)$ against p for different α. When $\alpha \to 1$, $\rho_w^*(\alpha, p)$ goes to $2/3$ for all $p \in (0, 1)$, which coincides with the limiting threshold discussed in Section III-B. As indicated in Fig. 1.2 of [20], the weak recovery threshold of ℓ_1-minimization is greater than 2/3 for all α that is greater than 0.9, since the weak recovery threshold of ℓ_p-minimization ($p \in [0, 1)$) when $\alpha \to 1$ is all 2/3, therefore for all $\alpha > 0.9$, the weak recovery threshold of ℓ_1-minimization is greater than that of ℓ_p-minimization for all $p \in [0, 1)$.

V. ℓ_1-MINIMIZATION CAN PERFORM BETTER THAN ℓ_p-MINIMIZATION ($p \in [0, 1)$) FOR SPARSE RECOVERY

For strong recovery, if ℓ_1-minimization can recover all the k-sparse vectors, then ℓ_p-minimization is also guaranteed to recover all the k-sparse vectors for all $p \in [0, 1)$. However, for weak recovery, the performance of ℓ_1-minimization is better than that of ℓ_p-minimization for all $p \in [0, 1)$ in at least the large α region ($\alpha > 0.9$), and the same result holds for all choices of supports and sign patterns. Then one may naturally ask why ℓ_1-minimization outperforms ℓ_p-minimization ($p < 1$) in recovering vectors on every specific support with every specific sign pattern, but is not as good as ℓ_p-minimization

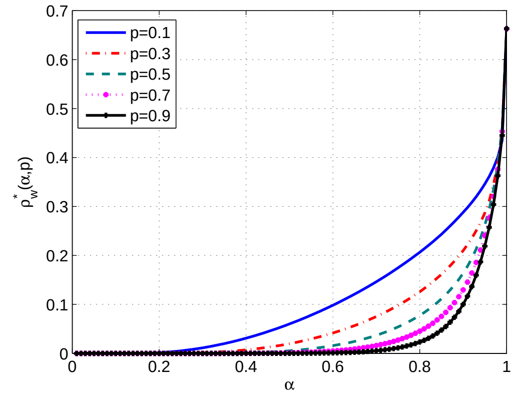


Fig. 4. $\rho_w^*(\alpha, p)$ against α for different p

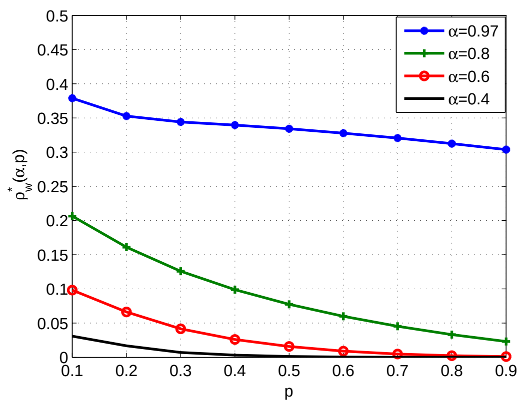


Fig. 5. $\rho_w^*(\alpha, p)$ against p for different α

in recovering vectors on all the supports with all the sign patterns? We next provide an intuitive explanation.

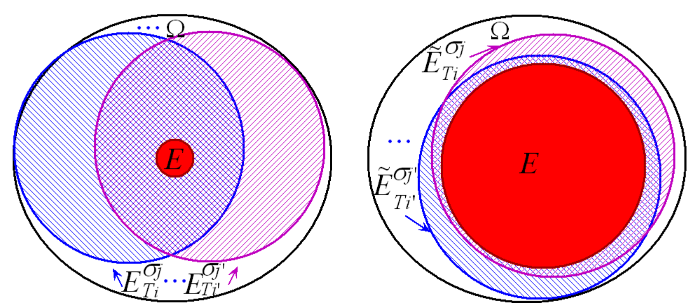
Let $\alpha < 1$ be very close to 1, let n be large enough and let A be a random Gaussian matrix. Then with overwhelming probability ℓ_1-minimization can recover all the vectors up to $\rho_1^s n$-sparse and ℓ_p-minimization with some $p \in [0, 1)$ can recover all the vectors up to $\rho_p^s n$-sparse, and we know $\rho_1^s < \rho_p^s$ from our discussion on strong bound. Note that since the limiting threshold of strong recovery via ℓ_p-minimization increases to 0.5 as p decreases to 0, then we have $\rho_1^s < \rho_p^s \leq 0.5$. However, if we only consider the ability to recover all the vectors on one support with one sign pattern, with overwhelming probability ℓ_1-minimization can recover vectors up to $\rho_1^w n$-sparse, while ℓ_p-minimization can recover vectors up to $\rho_p^w n$-sparse. From previous discussion about weak recovery threshold, we know that when α is very close to 1, $\rho_1^w > \frac{2}{3} > \rho_p^w > \frac{1}{2}$. And this result holds for any specific choice of the support and the sign pattern. Therefore we have $\rho_1^w > \rho_p^w > \rho_p^s > \rho_1^s$. We illustrate the difference of ℓ_1 and ℓ_p-minimization in Fig. 6 and Fig. 7. Let Ω be the set of all $m \times n$ matrices with entries drawn from standard Gaussian distribution, and the probability measure $P(\Omega) = 1$. We pick $\rho \in (\rho_1^s, \rho_p^s)$ in Fig. 6. Since $\rho < \rho_1^w$, for any fixed support T_i with $|T_i| = \rho n$ and any fixed sign pattern σ_j, with high probability ℓ_1-minimization can recover all the ρn-sparse vectors on T_i with sign pattern σ_j. Let $E_{T_i}^{\sigma_j}$ denote the event that ℓ_1-minimization can recover all the ρn-sparse vectors on support T_i with sign patter σ_j. There are $\binom{n}{\rho n}$ different supports, and for each support, there are $2^{\rho n}$ different sign patterns. Then $P(E_{T_i}^{\sigma_j})$ is very close to 1 for every T_i and σ_j as shown in Fig. 6(a). Since we also have $\rho > \rho_1^s$, then with high probability strong recovery of ℓ_1-minimization fails, in other words, ℓ_1-minimization would fail to recover at least one vector with at most ρn non-zero entries. Let E denote the event that ℓ_1-minimization can recover all the ρn-sparse vectors, then we have

$$E = \bigcap_{i \in \{1,\ldots,\binom{n}{\rho n}\}, j \in \{1,\ldots,2^{\rho n}\}} E_{T_i}^{\sigma_j}.$$

Then although $P(E_{T_i}^{\sigma_j})$ is the same for all T_i and σ_j and is very close to 1, $P(E)$ is close to 0, as indicated in Fig. 6(a). For ℓ_p-minimization, since $\rho < \rho_p^s$, then with high probability, ℓ_p-minimization can recover all the ρn-sparse vectors. In Fig. 6(b), $\tilde{E}$ denotes the event that ℓ_p-minimization can recover all the ρn-sparse vectors, then

$$\tilde{E} = \bigcap_{i \in \{1,\ldots,\binom{n}{\rho n}\}, j \in \{1,\ldots,2^{\rho n}\}} \tilde{E}_{T_i}^{\sigma_j},$$

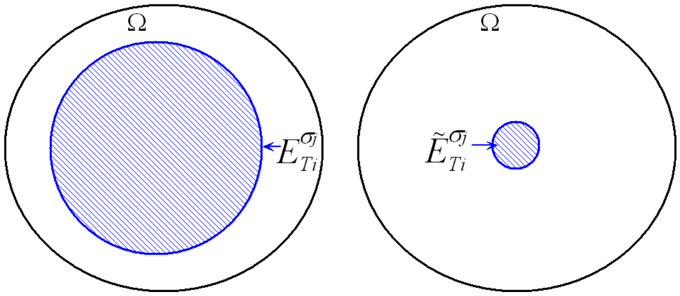
where $\tilde{E}_{T_i}^{\sigma_j}$ denotes the event that ℓ_p-minimization recovers all the vectors on support T_i with sign pattern σ_j. In this case, $P(\tilde{E})$ is close to 1 as indicated in Fig. 6(b). In Fig. 7, we pick $\rho \in (\rho_p^w, \rho_1^w)$. Then given any support T_i and any sign pattern σ_j, ℓ_1-minimization can recover all the vectors on T_i with sign pattern σ_j with high probability, while ℓ_p-minimization fails to recover at least one vector on T_i with sign pattern σ_j with high probability. Therefore $P(E_{T_i}^{\sigma_j})$ is close to 1, while $P(\tilde{E}_{T_i}^{\sigma_j})$ is close to 0 for any given T_i and σ_j. Therefore, if the sparse vectors we would like to recover are on one same support and share the same sign pattern, ℓ_1-minimization can be a better choice than ℓ_p-minimization for all $p \in [0, 1)$ regardless of the amplitudes of the entries of a vector.



(a) ℓ_1-minimization (b) ℓ_p-minimization

Fig. 6. Comparison of ℓ_1 and ℓ_p-minimization for $\rho \in (\rho_1^s, \rho_p^s)$.

To better understand how the recovery performance changes from strong recovery to weak recovery, let us consider another type of recovery: sectional recovery, which measures the ability of recovering all the vectors on one support T. Therefore, the requirement for successful sectional recovery is stricter than that of weak recovery, but is looser than that of strong recovery. The necessary and sufficient condition of successful sectional recovery can be stated as:



(a) ℓ_1-minimization (b) ℓ_p-minimization

Fig. 7. Comparison of ℓ_1 and ℓ_p-minimization for $\rho \in (\rho_p^w, \rho_1^w)$.

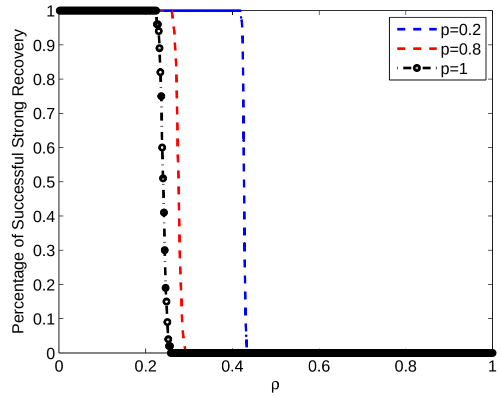


Fig. 8. Strong recovery threshold with 499×500 Gaussian matrix

Theorem 9. *ℓ_p-minimization problem ($p \in [0,1]$) can recover all the ρn-sparse vectors $\mathbf{x}$ on some support T if and only if*

$$\|B_T\mathbf{z}\|_p^p < \|B_{T^c}\mathbf{z}\|_p^p \tag{9}$$

for all non-zero $\mathbf{z} \in \mathcal{R}^{n-m}$.

The difference of the null space condition for strong recovery and sectional recovery is that (9) should hold for every support T for strong recovery, but only needs to hold for one specific support T for sectional recovery. Though for strong recovery, if the null space condition holds for $p \in [0,1]$, it also holds for all $q \in [0,p]$, this argument is not true for sectional recovery. Consider a simple example that the basis B of null space of A contains only one vector in $\mathcal{R}^4$ and $T = \{1,2\}$. If $B = [16,16,1,36]$, then one can check that $\|B_T\|_1 = 32 < 37 = \|B_{T^c}\|_1$, but $\|B_T\|_{0.5}^{0.5} = 8 > 7 = \|B_{T^c}\|_{0.5}^{0.5}$. If $B = [1,4,1,9]$, then $\|B_T\|_1 < \|B_{T^c}\|_1$, and $\|B_T\|_{0.5}^{0.5} < \|B_{T^c}\|_{0.5}^{0.5}$. Therefore the null space condition of successful sectional recovery holds for p does not necessarily imply that it holds for another $q \neq p$.

Using the techniques as in Section III-B, one can show that when $\alpha \to 1$ and n is large enough, the recovery threshold of sectional recovery is 1/2 for all $p \in [0,1]$. We skip the proof here as it follows the lines in Section III-B. To summarize, regarding the recovery threshold when $\alpha \to 1$, ℓ_p-minimization ($p \in [0,1]$) has a higher threshold for smaller p for strong recovery; the threshold is 1/2 for all $p \in [0,1]$ for sectional recovery; and the threshold is 2/3 for all $p \in [0,1)$ and is 1 for $p = 1$ for weak recovery. We can see how recovery performance changes when the requirement for successful recovery changes from strong to weak.

VI. Numerical Experiments

We present the results of numerical experiments to explore the performance of ℓ_p-minimization. First we consider the special case that the null space of the measurement matrix is only one dimensional. In this case, we can in fact compute the recovery threshold easily.

Experiment 1. Recovery thresholds when measurement matrices have one-dimensional null space

The null space of the measurement matrix A is only one-dimensional, and let vector β denote the basis of the null space of A. Then $\lambda\beta$ is in the null space of A for every $\lambda \in \mathcal{R}$, and every vector in the null space of A can be represented as $\lambda\beta$ for some $\lambda \in \mathcal{R}$. Thus, the strong recovery threshold and the weak recovery threshold of ℓ_1-minimization and ℓ_p-minimization can be directly computed by Theorem 1, Theorem 2 and Theorem 3, since we only need to check whether or not the null space condition holds for both β and $-\beta$. From Theorem 1, the strong recovery threshold of ℓ_p-minimization ($p \in (0,1]$) is the integer k such that the sum of the largest k terms of $|\beta_i|^p$ ($i \in \{1,...,n\}$) is less than $\|\boldsymbol{\beta}\|_p^p/2$ and the sum of the largest $k+1$ terms of $|\beta_i|^p$ ($i \in \{1,...,n\}$) is greater than or equal to $\|\boldsymbol{\beta}\|_p^p/2$. For weak recovery, we consider recovering all the nonnegative k-sparse vectors on support $T = \{1,...,k\}$. From Theorem 2, the weak recovery threshold of ℓ_1-minimization is the largest integer k such that both $\|\boldsymbol{\beta}_{T^-}\|_1 < \|\boldsymbol{\beta}_{T^c}\|_1 + \|\boldsymbol{\beta}_{T^+}\|_1$ and $\|\boldsymbol{\beta}_{T^+}\|_1 < \|\boldsymbol{\beta}_{T^c}\|_1 + \|\boldsymbol{\beta}_{T^-}\|_1$ hold. From Theorem 3, the weak recovery threshold of ℓ_p-minimization is the largest integer k such that both $\|\boldsymbol{\beta}_{T^-}\|_p^p < \|\boldsymbol{\beta}_{T^c}\|_p^p$ and $\|\boldsymbol{\beta}_{T^+}\|_p^p < \|\boldsymbol{\beta}_{T^c}\|_p^p$ hold.

We generate one hundred random Gaussian matrices $A^{499\times500}$, and for each random matrix A, we compute its corresponding strong (and weak) recovery threshold of ℓ_1 (and ℓ_p)-minimization. For each ρ between 0 and 1, we count the percentage of random matrices with which ℓ_1 (and ℓ_p)-minimization can recover all the ρn-sparse vectors in the strong sense (and in the weak sense). Fig. 8 shows the strong recovery thresholds for different p and Fig. 9 shows the weak recovery thresholds. We can see that the strong recovery threshold strictly decreases as p increases. However, the weak recovery threshold of ℓ_1-minimization is close to 0.9, which is greater than the weak recovery threshold of ℓ_p-minimization for every $p < 1$.

Except for special cases like Experiment 1, (3) is indeed non-convex and it is hard to compute its global minimum. In following experiments we employ the iteratively reweighted least squares algorithm [11][12] to compute the local minimum of (3), please refer to [12] about the details of the algorithm.

Experiment 2. ℓ_p-minimization using IRLS [12]

We fix $n = 200$ and $m = 100$, and increase ρ from 0.01 to 0.5. For each ρ, we repeat the following procedure one hundred times. We first generate an n-dimensional vector $\mathbf{x}$ with ρn non-zero entries. The location of the non-zero entries are chosen randomly, and each non-zero value follows from

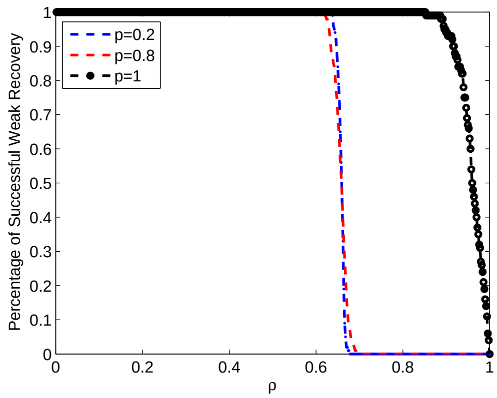


Fig. 9. Weak recovery threshold with 499×500 Gaussian matrix

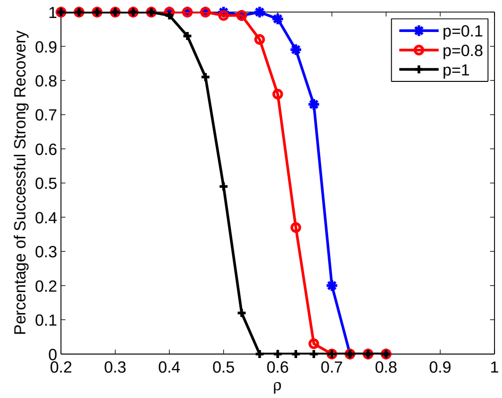


Fig. 11. Successful strong recovery of ρn-sparse vectors

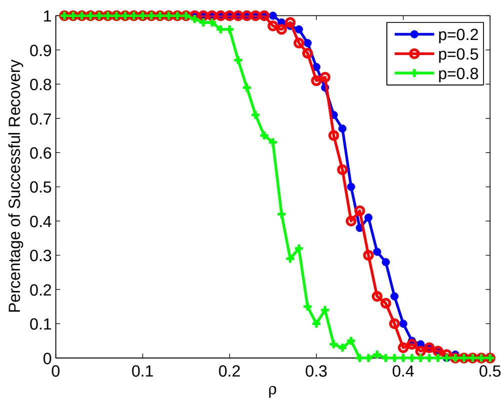


Fig. 10. Successful recovery of ρn-sparse vectors via ℓ_p-minimization

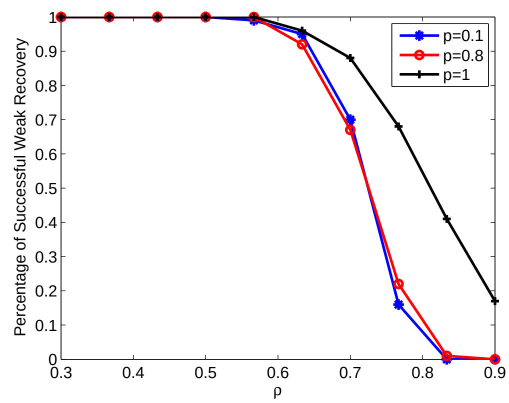


Fig. 12. Successful weak recovery of ρn-sparse vectors

standard Gaussian distribution. We then generate an $m \times n$ matrix A with i.i.d. $\mathcal{N}(0, 1)$ entries. We let $\mathbf{y} = A\mathbf{x}$ and run the iteratively reweighted least squares algorithm to search for a local minimum of (3) with p chosen to be 0.2, 0.5, and 0.8 respectively. Let $\mathbf{x}^*$ be the output of the algorithm, if $\|\mathbf{x}^* - \mathbf{x}\|_2 \leq 10^{-4}$, we say the recovery of $\mathbf{x}$ is the successful. Fig. 10 records the percentage of times that the recovery is successful for different sparsity ρn. Note that the iteratively reweighted least squares algorithm is designed to obtain a local minimum of the ℓ_p-minimization problem (3), and is not guaranteed to obtain the global minimum. However, as shown in Figure 10, it indeed recovers the sparse vectors up to certain sparsity. For $\ell_{0.2}$, $\ell_{0.5}$ and $\ell_{0.8}$-minimization computed by the heuristic, the sparsity ratios of successful recovery are 0.25, 0.24, and 0.15 respectively.

Experiment 3. Strong recovery vs. weak recovery

We also compare the performance of ℓ_p-minimization and ℓ_1-minimization both for strong recovery in Fig. 11 and for weak recovery in Fig. 12 when α is large. We employ the trial version of `MOSEK` [30] to solve ℓ_1-minimization and still employ the iteratively reweighted least squares algorithm to compute a local minimum of ℓ_p-minimization. We fix $n = 60$ and $m = 58$ and independently generate one hundred random matrices $A^{m \times n}$ with i.i.d. $\mathcal{N}(0, 1)$ entries and evaluate the performance of strong recovery and weak recovery. For each matrix, we increase ρ from 0.2 to 1. In weak recovery, we consider recovering nonnegative vectors on a fixed support $T = \{1, ..., \rho n\}$. For a given ρ, we generate one thousand vectors and claim the weak recovery of ρn-sparse vectors to be successful if and only if all the vectors are successfully recovered. For each vector $\mathbf{x}$, $x_i = |z_i|$ ($i \in T$), and z_i is generated from $\mathcal{N}(0, 1)$ with probability 0.5, and $\mathcal{N}(1000, 1)$ with probability 0.5. As discussed in Section II, the condition for successful weak recovery via ℓ_1-minimization is the same for every nonnegative vector on T, therefore for a fixed matrix A, if ℓ_1-minimization recovers all the vectors we generated, it should also recover all the nonnegative vectors on T. ℓ_p-minimization ($p \in [0, 1)$), on the other hand, can recover some nonnegative vectors on T while at the same time fails to recover some other nonnegative vectors on T. Therefore, since we could not check every nonnegative $\mathbf{x}$ on T, ℓ_p-minimization ($p < 1$) can still fail to recover some other nonnegative vector on T even if we declare the weak recovery to be "successful". In strong recovery, for each ρ, we generate one thousand vectors and claim the strong recovery to be successful if and only if all these vectors are correctly recovered. For each such random ρn-sparse vector $\mathbf{x}$, we first randomly pick a support T with $|T| = \rho n$, and then for each x_i ($i \in T$), x_i is generated from $\mathcal{N}(0, 1)$ with probability 0.5, from $\mathcal{N}(1000, 1)$ with probability 0.25, and from $\mathcal{N}(-1000, 1)$ with probability 0.25. The average performance of one hundred random matrices for strong recovery is plotted in Fig. 11, and the average performance of weak recovery is plotted in Fig. 12. Note that we only apply iteratively reweighted least squares algorithm to approximate the performance of ℓ_p-minimization, therefore the solution returned by the algorithm may not always be the solution of ℓ_p-minimization. Simulation results indicate that for strong recovery, the recovery threshold increases as p decreases, while for the weak recovery, interestingly, the recovery threshold of ℓ_1-minimization is higher than any other ℓ_p-minimization for $p < 1$.

VII. Conclusion

This paper analyzes the ability of ℓ_p-minimization ($0 \leq p \leq 1$) to recover high-dimensional sparse vectors from low-dimensional linear measurements where the measurement matrix $A^{m\times n}$ has i.i.d. standard Gaussian entries. When $\alpha = m/n \to 1$, we provide a tight threshold $\rho^*(p)$ of the sparsity ratio separating the success and failure of strong recovery which requires to recover all the sparse vectors. $\rho^*(p)$ strictly decreases from 0.5 to 0.239 as p increases from 0 to 1. For weak recovery which only needs to recover sparse vectors on some support with some sign pattern, we first provide an equivalent null space characterization of successful weak recovery, then prove that the threshold of sparsity ratio separating the success and failure of ℓ_p-minimization is 2/3 for all $p < 1$, compared with the threshold 1 for ℓ_1-minimization. For any $\alpha < 1$, we provide a bound $\rho^*(\alpha, p)$ of sparsity ratio below which strong recovery via ℓ_p-minimization succeeds with overwhelming probability, and our bound $\rho^*(\alpha, p)$ improves on the existing bounds in the large α region. We also provide a bound $\rho^*_w(\alpha, p)$ of sparsity ratio below which weak recovery succeeds with overwhelming probability.

Throughout the paper, we assume that the measurements $\mathbf{y} = A\mathbf{x}$ are exact, and it would be interesting to consider the case that the measurements are noisy, i.e. $\mathbf{y} = A\mathbf{x}+\mathbf{e}$ where $\mathbf{e}$ is the vector of noise. Moreover, we assume that $\mathbf{x}$ is exactly sparse, i.e. most of its entries are exactly zero. The extension of results to approximately sparse vectors whose coefficients (if ordered) decay rapidly is also worth pursuit.

Acknowledgments: The authors thank J. Tanner of University of Edinburgh, Edinburgh, UK, for helpful suggestions, and anonymous reviewers for their efforts in helping improving this paper. The research is supported by NSF under CCF-0835706.

References

[1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” *Constructive Approximation*, vol. 28, pp. 253–263, 2008.
[2] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss., “Combining geometry and combinatorics: a unified approach to sparse signal recovery,” 2008. [Online]. Available: http://arxiv.org/abs/0804.4666
[3] R. Berinde and P. Indyk, “Sparse recovery using sparse random matrices,” *MIT-CSAIL Technical Report*, 2008.
[4] J. Blanchard, C. Cartis, and J. Tanner, “The restricted isometry property and ℓ^q regularization: phase transitions for sparse approximation,” *Preprint*, 2009.
[5] A. Bruckstein, M. Elad, and M. Zibulevsky, “On the uniqueness of nonnegative sparse solutions to underdetermined systems of equations,” *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4813–4820, Nov. 2008.
[6] E. Candès, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489 – 509, Feb. 2006.
[7] E. Candès and T. Tao, “Decoding by linear programming,” *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[8] ——, “Near-optimal signal recovery from random projections: Universal encoding strategies?” *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
[9] R. Chartrand, “Exact reconstruction of sparse signals via nonconvex minimization,” *Signal Process. Lett.*, vol. 14, no. 10, pp. 707–710, 2007.
[10] ——, “Nonconvex compressed sensing and error correction,” in *Proc. ICASSP*, 2007.
[11] R. Chartrand and W. Yin, http://www.caam.rice.edu/~wy1/paperfiles/TR08-01/IRLS_CS.rar.
[12] ——, “Iteratively reweighted algorithms for compressive sensing,” in *Proc. IEEE ICASSP 2008.*, Apr. 2008, pp. 3869 –3872.
[13] R. Chartrand and V. Staneva, “Restricted isometry properties and nonconvex compressive sensing,” *Inverse Problems*, vol. 24, pp. 35 020–35 033(14), 2008.
[14] A. Cohen, W. Dahmen, and R. DeVore, “Compressed sensing and best k-term approximation,” *Journal of the American Mathematical Society*, vol. 22, pp. 211–231, 2009.
[15] M. E. Davies and R. Gribonval, “Restricted isometry constants where l_p sparse recovery can fail for $0 < p \leq 1$,” *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2203–2214, 2009.
[16] A. Dembo, O. Zeitouni, A. Dembo, and O. Zeitouni, in *Large Deviations Techniques and Applications*, ser. Stochastic Modelling and Applied Probability, B. Rozovskii and G. Grimmett, Eds., 2010, vol. 38, pp. 11–70.
[17] D. Donoho, “High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension,” *Discrete Comput. Geom.*, 2006.
[18] D. L. Donoho and J. Tanner, “Sparse nonnegative solution of underdetermined linear equations by linear programming,” in *Proc. Natl. Acad. Sci. U.S.A.*, vol. 102, no. 27, 2005, pp. 9446–9451.
[19] ——, “Counting the faces of randomly-projected hypercubes and orthants, with applications,” *Journal of the American Mathematical Society*, vol. 22, no. 1, pp. 1–53, 2009.
[20] D. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
[21] D. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition,” *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845 – 2862, Nov. 2001.
[22] C. Dwork, F. McSherry, and K. Talwar, “The price of privacy and the limits of lp decoding,” in *Proc. STOC*, 2007, pp. 85–94.
[23] M. Elad and A. Bruckstein, “A generalized uncertainty principle and sparse representation in pairs of bases,” *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2558 – 2567, Sep. 2002.
[24] S. Foucart and M.-J. Lai, “Sparsest solutions of underdetermined linear systems via l_q-minimization for $0 < q \leq 1$,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 395 – 407, 2009.
[25] J.-J. Fuchs, “On sparse representations in arbitrary redundant bases,” *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341 – 1344, Jun. 2004.
[26] R. Gribonval and M. Nielsen, “Sparse representations in unions of bases,” *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320 – 3325, Dec. 2003.
[27] ——, “Highly sparse representations from dictionaries are unique and independent of the sparseness measure,” *Applied and Computational Harmonic Analysis*, vol. 22, no. 3, pp. 335 – 355, 2007.
[28] J. Haupt and R. Nowak, “Signal reconstruction from noisy random projections,” *IEEE Trans. Inf. Theory*, vol. 52, no. 9, pp. 4036 –4048, Sep. 2006.
[29] M. Ledoux, Ed., *The Concentration of Measure Phenomenon*. American Mathematical Society, 2001.
[30] Mosek Optimization Software, http://www.mosek.com/.
[31] R. Saab, R. Chartrand, and O. Yilmaz, “Stable sparse approximations via nonconvex optimization,” in *Proc. ICASSP*, 2008.
[32] M. Stojnic, “A simple performance analysis of ℓ_1 optimization in compressed sensing,” in *Proc. IEEE ICASSP 2009.*, Apr. 2009, pp. 3021 –3024.
[33] M. Stojnic, W. Xu, and B. Hassibi, “Compressed sensing - probabilistic analysis of a null-space characterization,” in *Proc. ICASSP*, 2008, pp. 3377–3380.

[34] J. Wright and Y. Ma, "Dense error correction via l_1-minimization," in *Proc. ICASSP 2009.*, Apr. 2009, pp. 3033 –3036.
[35] W. Xu and B. Hassibi, "Efficient compressive sensing with deterministic guarantees using expander graphs," in *Information Theory Workshop, 2007. ITW '07. IEEE*, Sept. 2007, pp. 414–419.
[36] ——, "Compressed sensing over the Grassmann manifold: A unified analytical framework," in *Proc. Allerton 2008*, Sep. 2008, pp. 562 – 567.
[37] ——, "Compressive sensing over the Grassmann manifold: a unified geometric framework," *Preprint*, 2010.
[38] Y. Zhang, "When is missing data recoverable," Tech. Rep., 2006.

APPENDIX

A. Proof of Theorem 3

Proof: Necessary part. Suppose the condition fails for some $\mathbf{z}$, then there are two cases: (1) T^+ is empty, and (2) T^+ is not empty for that particular $\mathbf{z}$.

First consider the case T^+ is empty, then we have $\|B_{T^-}\mathbf{z}\|_p^p \geq \|B_{T^c}\mathbf{z}\|_p^p$ since we assume the condition in Theorem 3 fails for $\mathbf{z}$. Define a vector $\mathbf{x}$ as follows. Let $x_i = 0$ for every i in T^c, let $x_i = -B_i\mathbf{z}$ for every i in T^-. Let x_i be any positive value for every i in T^0. Then according to the definition of $\mathbf{x}$, we have

$$\begin{aligned} & \|\mathbf{x}+B\mathbf{z}\|_p^p \\ = \; & \|\mathbf{x}_{T^-}+B_{T^-}\mathbf{z}\|_p^p + \|\mathbf{x}_{T^0}+B_{T^0}\mathbf{z}\|_p^p + \|B_{T^c}\mathbf{z}\|_p^p \\ = \; & 0 + \|\mathbf{x}_{T^0}\|_p^p + \|B_{T^c}\mathbf{z}\|_p^p \\ = \; & \|\mathbf{x}\|_p^p - \|\mathbf{x}_{T^-}\|_p^p + \|B_{T^c}\mathbf{z}\|_p^p \\ = \; & \|\mathbf{x}\|_p^p - \|B_{T^-}\mathbf{z}\|_p^p + \|B_{T^c}\mathbf{z}\|_p^p \\ \leq \; & \|\mathbf{x}\|_p^p. \end{aligned}$$

Since $\|\mathbf{x}+B\mathbf{z}\|_p^p \leq \|\mathbf{x}\|_p^p$, (3) cannot successfully recover $\mathbf{x}$, which is a contradiction.

Secondly, consider the case that T^+ is not empty. Then $\|B_{T^-}\mathbf{z}\|_p^p > \|B_{T^c}\mathbf{z}\|_p^p$ since we assume the condition in Theorem 3 fails for $\mathbf{z}$. Let $\delta = \|B_{T^-}\mathbf{z}\|_p^p - \|B_{T^c}\mathbf{z}\|_p^p > 0$. Define a vector $\mathbf{x}$ as follows. Let $x_i = 0$ for every i in T^c, let $x_i = -B_i\mathbf{z}$ for every i in T^-, and let x_i be any positive value for every i in T^0. For every i in T^+, since $p \in (0,1)$, we can pick $x_i > 0$ large enough such that $\|\mathbf{x}_{T^+}+B_{T^+}\mathbf{z}\|_p^p - \|\mathbf{x}_{T^+}\|_p^p < \frac{\delta}{2}$. Then

$$\begin{aligned} \|\mathbf{x}+B\mathbf{z}\|_p^p &= 0 + \|\mathbf{x}_{T^+}+B_{T^+}\mathbf{z}\|_p^p + \|\mathbf{x}_{T^0}\|_p^p + \|B_{T^c}\mathbf{z}\|_p^p \\ &< \|\mathbf{x}_{T^+}\|_p^p + \frac{\delta}{2} + \|\mathbf{x}_{T^0}\|_p^p + \|B_{T^c}\mathbf{z}\|_p^p \\ &= \|\mathbf{x}_{T^+}\|_p^p + \frac{\delta}{2} + \|\mathbf{x}_{T^0}\|_p^p + \|B_{T^-}\mathbf{z}\|_p^p - \delta \\ &= \|\mathbf{x}\|_p^p - \frac{\delta}{2}. \end{aligned}$$

Thus $\|\mathbf{x}+B\mathbf{z}\|_p^p < \|\mathbf{x}\|_p^p$, $\mathbf{x}$ is not a solution to (3), which is also a contradiction.

Sufficient part. Assume the null space condition holds, then for any nonnegative $\mathbf{x}$ on support T, and any non-zero $\mathbf{z} \in \mathcal{R}^{n-m}$, we have

$$\begin{aligned} & \|\mathbf{x}+B\mathbf{z}\|_p^p \\ = \; & \|\mathbf{x}_{T^+}+B_{T^+}\mathbf{z}\|_p^p + \|\mathbf{x}_{T^-}+B_{T^-}\mathbf{z}\|_p^p \\ & + \|\mathbf{x}_{T^0}\|_p^p + \|B_{T^c}\mathbf{z}\|_p^p \\ \geq \; & \|\mathbf{x}_{T^+}+B_{T^+}\mathbf{z}\|_p^p + \|\mathbf{x}_{T^-}\|_p^p - \|B_{T^-}\mathbf{z}\|_p^p \\ & + \|\mathbf{x}_{T^0}\|_p^p + \|B_{T^c}\mathbf{z}\|_p^p, \end{aligned} \tag{10}$$

where the inequality follows from the triangular property that $|\mathbf{x}_i + B_i\mathbf{z}|^p \geq |\mathbf{x}_i|^p - |B_i\mathbf{z}|^p$ holds for all i and all $p \in (0,1)$.

If T^+ is not empty, then $\|\mathbf{x}_{T^+}+B_{T^+}\mathbf{z}\|_p^p > \|\mathbf{x}_{T^+}\|_p^p$ since $B_i\mathbf{z} > 0$ for every i in T^+, and $B_i\mathbf{z}$ and x_i have the same sign. Since we also have $\|B_{T^-}\mathbf{z}\|_p^p \leq \|B_{T^c}\mathbf{z}\|_p^p$ from assumption, therefore by (10) we have $\|\mathbf{x}+B\mathbf{z}\|_p^p > \|\mathbf{x}\|_p^p$. If T^+ is empty, then $\|B_{T^-}\mathbf{z}\|_p^p < \|B_{T^c}\mathbf{z}\|_p^p$ from assumption, therefore by (10) we also have $\|\mathbf{x}+B\mathbf{z}\|_p^p > \|\mathbf{x}\|_p^p$. Thus, $\|\mathbf{x}+B\mathbf{z}\|_p^p > \|\mathbf{x}\|_p^p$ for all non-zero $\mathbf{z} \in \mathcal{R}^{n-m}$, then $\mathbf{x}$ is the solution to (3). ∎

B. Proof of Lemma 1

Proof: Let $X \sim \mathcal{N}(0,1)$ and let $Z = |X|$. Let $f(z)$ and $F(z)$ denote the p.d.f. and c.d.f. of Z respectively. Then

$$f(z) = \begin{cases} \sqrt{2/\pi} e^{-\frac{1}{2}z^2} & \text{if } z \geq 0, \\ 0 & \text{if } z < 0. \end{cases} \tag{11}$$

$$F(z) = \begin{cases} \operatorname{erf}(z/\sqrt{2}) = \int_0^z \sqrt{2/\pi} e^{-\frac{1}{2}x^2} dx, & \text{if } z \geq 0 \\ 0 & \text{if } z < 0. \end{cases} \tag{12}$$

Define $g(t) = \int_t^\infty z^p f(z) dz$. g is continuous and decreasing in $[0,\infty]$, and $g(0) = E[Z^p] = \frac{S}{n}$, $\lim_{t\to\infty} g(t) = 0$. Then there exists z^* such that $g(z^*) = \frac{g(0)}{2}$, i.e.

$$\int_0^{z^*} x^p f(x) dx - \int_{z^*}^\infty x^p f(x) dx = 0. \tag{13}$$

Define

$$\rho^* = 1 - F(z^*). \tag{14}$$

We claim ρ^* has the desired property.

Let

$$T_{z^*} = \sum_{i: Y_i \geq z^{*p}} Y_i = \sum_{i=1}^n Y_i \mathbf{1}_{\{Y_i \geq z^{*p}\}},$$

where $\mathbf{1}$ is the indicator function. Then

$$\begin{aligned} E[T_{z^*}] &= E[\sum_{i=1}^n Y_i \mathbf{1}_{\{Y_i \geq z^{*p}\}}] = E[\sum_{i=1}^n |X_i|^p \mathbf{1}_{\{|X_i|^p \geq z^{*p}\}}] \\ &= nE[Z^p \mathbf{1}_{\{Z \geq z^*\}}] = n \int_{z^*}^\infty z^p f(z) dz = ng(z^*). \end{aligned}$$

Let h be the smallest integer such that $Y_h \geq z^{*p}$ and $Y_{h+1} < z^{*p}$, then $T_{z^*} = \sum_{i=1}^h Y_i$. We also have that $h = \sum_{i=1}^n \mathbf{1}_{\{|X_i|^p \geq z^{*p}\}} = \sum_{i=1}^n \mathbf{1}_{\{|X_i| \geq z^*\}}$. Note that $P(|X_i| \geq z^*) = 1 - F(z^*) = \rho^*$, thus h follows the Binomial distribution $B(n, \rho^*)$. Then its expectation $E[h] = \rho^* n$, and the variance $E[(h - \rho^* n)^2] = n\rho^*(1-\rho^*)$.

We claim that

$$|T_{z^*} - S_{\rho^*}| \leq \frac{|h - \lceil \rho^* n \rceil| S_{\rho^*}}{\lceil \rho^* n \rceil}. \tag{15}$$

To see this, consider three different cases, $h = \lceil \rho^* n \rceil$, $h > \lceil \rho^* n \rceil$ and $h < \lceil \rho^* n \rceil$. If $h = \lceil \rho^* n \rceil$, then $T_{z^*} = S_{\rho^*}$, and (15) holds trivially. If $h > \lceil \rho^* n \rceil$, then $|T_{z^*} - S_{\rho^*}| = \sum_{i=\lceil \rho^* n \rceil + 1}^h Y_i$. Note that for every $i > \lceil \rho^* n \rceil$, $Y_i \leq Y_{\lceil \rho^* n \rceil} \leq S_{\rho^*}/\lceil \rho^* n \rceil$, then (15) follows. If $h < \lceil \rho^* n \rceil$, then $|T_{z^*} - S_{\rho^*}| = \sum_{i=h+1}^{\lceil \rho^* n \rceil} Y_i$. Since $Y_i \geq Y_j$ for all $i \leq j$, then $\sum_{i=h+1}^{\lceil \rho^* n \rceil} Y_i / (\lceil \rho^* n \rceil - h) \leq \sum_{i=1}^h Y_i / h$, which leads to

$\sum_{i=h+1}^{\lceil\rho^*n\rceil} Y_i/(\lceil\rho^*n\rceil - h) \leq \sum_{i=1}^{\lceil\rho^*n\rceil} Y_i/\lceil\rho^*n\rceil = S_{\rho^*}/\lceil\rho^*n\rceil$, and (15) follows. Combining three cases, we conclude that (15) always holds. Then

$$\begin{aligned} E[|T_{z^*} - S_{\rho^*}|] &\leq \frac{E[|h - \lceil\rho^*n\rceil| S_{\rho^*}]}{\lceil\rho^*n\rceil} \\ &\leq \frac{\sqrt{E[(h - \lceil\rho^*n\rceil)^2]E[S_{\rho^*}^2]}}{\lceil\rho^*n\rceil}, \quad (16) \end{aligned}$$

where the second inequality follows from the Cauchy-Schwarz inequality. We have

$$\begin{aligned} & E[(h - \lceil\rho^*n\rceil)^2] \\ = \ & E[(h - \rho^*n)^2] + 2(\rho^*n - \lceil\rho^*n\rceil)E[h - \rho^*n] \\ & +(\rho^*n - \lceil\rho^*n\rceil)^2 \\ \leq \ & n\rho^*(1-\rho^*) + 1. \end{aligned}$$

Besides,

$$\begin{aligned} E[S_{\rho^*}^2] &\leq E[S_1^2] = E[(\sum_{i=1}^{n} |X_i|^p)^2] \\ &= E[\sum_{i=1}^{n} |X_i|^{2p} + \sum_{i,j:i\neq j} |X_i|^p|X_j|^p] \\ &= nE[|X|^{2p}] + n(n-1)(E[|X|^p])^2, \end{aligned}$$

where the third equality follows since X_1, X_2,...,X_n are i.i.d. $\mathcal{N}(0,1)$ random variables. Then from (16) we have

$$\begin{aligned} & E[|T_{z^*} - S_{\rho^*}|] \\ \leq \ & \frac{\sqrt{(n\rho^*(1-\rho^*)+1)(nE[|X|^{2p}] + n(n-1)(E[|X|^p])^2)}}{\lceil\rho^*n\rceil} \\ = \ & O(\sqrt{n}). \end{aligned}$$

Since $E[|T_{z^*} - S_{\rho^*}|]$ is upper bounded by $O(\sqrt{n})$, $E[T_{z^*}] = ng(z^*)$, and $S = ng(0)$, we have

$$\begin{aligned} \lim_{n\to\infty} \frac{E[S_{\rho^*}]}{S} &= \lim_{n\to\infty} \frac{E[T_{z^*}]}{S} + \lim_{n\to\infty} \frac{E[S_{\rho^*} - T_{z^*}]}{S} \\ &= \frac{g(z^*)}{g(0)} + 0 = \frac{1}{2}. \end{aligned}$$

■

C. Proof of Proposition 1

Proof: From the definition of z^* in (13), we have

$$H(z^*, p) := \int_0^{z^*} x^p f(x)dx - \int_{z^*}^{\infty} x^p f(x)dx = 0, \quad (17)$$

where $f(\cdot)$ and $F(\cdot)$ are defined in (11) and (12). From the Implicit Function Theorem,

$$\frac{dz^*}{dp} = -\frac{\frac{\partial H}{\partial p}}{\frac{\partial H}{\partial z^*}} = -\frac{\int_0^{z^*} x^p(\ln x)f(x)dx - \int_{z^*}^{\infty} x^p(\ln x)f(x)dx}{2z^{*p}f(z^*)}.$$

From (14), we have $\frac{d\rho^*}{dz^*} = -f(z^*)$. From the chain rule, we know $\frac{d\rho^*}{dp} = \frac{d\rho^*}{dz^*}\frac{dz^*}{dp}$, thus

$$\frac{d\rho^*}{dp} = \frac{\int_0^{z^*} x^p(\ln x)f(x)dx - \int_{z^*}^{\infty} x^p(\ln x)f(x)dx}{2z^{*p}} \quad (18)$$

Note that

$$\begin{aligned} \int_0^{z^*} x^p(\ln x)f(x)dx &< \int_0^{z^*} x^p(\ln z^*)f(x)dx \\ &= \int_{z^*}^{\infty} x^p(\ln z^*)f(x)dx \\ &< \int_{z^*}^{\infty} x^p(\ln x)f(x)dx, \quad (19) \end{aligned}$$

where the equality follows from (17). Then the numerator of (18) is less than 0 from (19), thus $\frac{d\rho^*}{dp} < 0$. ■

D. Proof of Lemma 2

Proof: Let $\mathbf{X} = [X_1, ..., X_n]^T$. If two vectors $\mathbf{X}$ and $\mathbf{X}'$ only differ in co-ordinate i, then for any p, $|S_\rho(\mathbf{X}) - S_\rho(\mathbf{X}')| \leq ||X_i|^p - |X_i'|^p|$. Thus for any $\mathbf{X}$ and $\mathbf{X}'$,

$$|S_\rho(\mathbf{X}) - S_\rho(\mathbf{X}')| \leq \sum_{i:X_i\neq X_i'} \big||X_i|^p - |X_i'|^p\big|.$$

Since $\big||X_i|^p - |X_i'|^p\big| \leq |X_i - X_i'|^p$ for all $p \in (0,1]$,

$$|S_\rho(\mathbf{X}) - S_\rho(\mathbf{X}')| \leq \sum_i |X_i - X_i'|^p. \quad (20)$$

From the isoperimetric inequality for the Gaussian measure [29], for any set $A \in \mathcal{R}^n$ with measure at least a half, the set $A_t = \{\mathbf{x} \in \mathcal{R}^n : d(\mathbf{x}, A) \leq t\}$ has measure at least $1 - e^{-t^2/2}$, where $d(\mathbf{x}, A) = \inf_{\mathbf{y}\in A} \|\mathbf{x} - \mathbf{y}\|_2$. Let M_ρ be the median value of $S_\rho = S_\rho(\boldsymbol{X})$. Define set $A = \{\mathbf{x} \in \mathcal{R}^n : S_\rho(\mathbf{x}) \leq M_\rho\}$, then

$$P(d(\mathbf{x}, A) \leq t) \geq 1 - e^{-t^2/2}.$$

We claim that $d(\mathbf{x}, A) \leq t$ implies that $S_\rho(\mathbf{x}) \leq M_\rho + n^{(1-p/2)}t^p$. If $\mathbf{x} \in A$, then $S_\rho(\mathbf{x}) \leq M_\rho$, thus the claim holds as $n^{1-p/2}t^p$ is nonnegative. If $\mathbf{x} \notin A$, then there exists $\mathbf{x}' \in A$ such that $\|\mathbf{x} - \mathbf{x}'\|_2 \leq t$. Let $u_i = 1$ for all i and let $v_i = |x_i - x_i'|^p$. From Hölder's inequality,

$$\begin{aligned} \sum_i |x_i - x_i'|^p &\leq \left(\sum_i |u_i|^{2/(2-p)}\right)^{1-p/2} \left(\sum_i |v_i|^{2/p}\right)^{p/2} \\ &\leq n^{(1-p/2)}(t^2)^{p/2} = n^{(1-p/2)}t^p \quad (21) \end{aligned}$$

From (20) and (21), $|S_\rho(\mathbf{x}) - S_\rho(\mathbf{x}')| \leq n^{(1-p/2)}t^p$. Since $\mathbf{x} \notin A$ and $\mathbf{x}' \in A$, then $S_\rho(\mathbf{x}) > M_\rho \geq S_\rho(\mathbf{x}')$. Thus $S_\rho(\mathbf{x}) \leq M_\rho + n^{(1-p/2)}t^p$, which verifies our claim. Then

$$P(S_\rho(\mathbf{x}) \leq M_\rho + n^{(1-p/2)}t^p) \geq P(d(\mathbf{x}, A) \leq t) \geq 1 - e^{-t^2/2}. \quad (22)$$

Similarly,

$$P(S_\rho(\mathbf{x}) \geq M_\rho - n^{(1-p/2)}t^p) \geq 1 - e^{-t^2/2}. \quad (23)$$

Combining (22) and (23),

$$P(|S_\rho(x) - M_\rho| \geq n^{(1-p/2)}t^p) \leq 2e^{-t^2/2}. \quad (24)$$

The difference of $E[S_\rho]$ and M_ρ can be bounded as follows,

$$\begin{aligned}|E[S_\rho]-M_\rho| &\leq E[|S_\rho-M_\rho|]\\ &= \int_0^\infty P(|S_\rho(x)-M_\rho|\geq y)dy\\ &\leq \int_0^\infty 2e^{-\frac{1}{2}y^{\frac{2}{p}}n^{(1-\frac{2}{p})}}dy\\ &= n^{(1-\frac{p}{2})}\int_0^\infty 2e^{-\frac{1}{2}s^{\frac{2}{p}}}ds\end{aligned}$$

Note that $c:=\int_0^\infty 2e^{-\frac{1}{2}s^{(2/p)}}ds$ is a finite constant for all $p\in(0,1]$. As $p>0$ and $S=nE[|x_i|^p]$, thus for any $\delta>0$, $cn^{(1-\frac{p}{2})}<\frac{\delta}{2}S$ when n is large enough.

Let $t=\left(\frac{1}{2}\delta Sn^{(\frac{p}{2}-1)}\right)^{\frac{1}{p}}=(\frac{1}{2}\delta E[|x_i|^p])^{\frac{1}{p}}\sqrt{n}$, from (24) with probability at least $1-2e^{-\frac{1}{2}(\frac{1}{2}\delta E[|x_i|^p])^{\frac{2}{p}}n}$, $|S_\rho-M_\rho|<\frac{1}{2}\delta S$. Thus $|S_\rho-E[S_\rho]|\leq|S_\rho-M_\rho|+|M_\rho-E[S_\rho]|<\delta S$ with probability at least $1-2e^{-c_1n}$ for some constant c_1. ■

E. Proof of Corollary 1

Proof: From Lemma 1 we know that for every $\epsilon>0$, there exists M large enough such that

$$E[S_{\rho^*}]\leq(\frac{1}{2}+\epsilon)S \tag{25}$$

for all $n\geq M$ where $S=E[S_1]$. Then $E[\sum_{i=\lceil\rho^*n\rceil+1}^n Y_i]=S-E[S_{\rho^*}]\geq(\frac{1}{2}-\epsilon)S$. Since $E[\sum_{i=\lceil\rho^*n\rceil+1}^n Y_i]$ is a summation of $n-\lceil\rho^*n\rceil$ terms, and $E[Y_{\lceil\rho^*n\rceil}]\geq E[Y_i]$ for all $i\geq\lceil\rho^*n\rceil$, then we have

$$E[Y_{\lceil\rho^*n\rceil}]\geq\frac{E[\sum_{i=\lceil\rho^*n\rceil+1}^n Y_i]}{n-\lceil\rho^*n\rceil}\geq\frac{(\frac{1}{2}-\epsilon)S}{n-\lceil\rho^*n\rceil}\geq\frac{(\frac{1}{2}-\epsilon)S}{n}. \tag{26}$$

Then for any $\rho<\rho^*$, for every $\epsilon>0$, when n is large enough,

$$\begin{aligned}E[S_\rho] &= E[S_{\rho^*}]-\sum_{i=\lceil\rho n\rceil+1}^{\lceil\rho^*n\rceil}E[Y_i]\\ &\leq E[S_{\rho^*}]-(\lceil\rho^*n\rceil-\lceil\rho n\rceil)E[Y_{\lceil\rho^*n\rceil}]\\ &\leq (\frac{1}{2}+\epsilon)S-(\lceil\rho^*n\rceil-\lceil\rho n\rceil)\frac{(\frac{1}{2}-\epsilon)S}{n}\\ &\leq (\frac{1}{2}+\epsilon)S-(\rho^*-\rho-\frac{1}{n})(\frac{1}{2}-\epsilon)S\end{aligned}$$

where the first inequality holds since each Y_i with $i\leq\lceil\rho^*n\rceil$ has expectation at least as large as $E[Y_{\lceil\rho^*n\rceil}]$, and the second inequality follows from (25) and (26). Then for any $\rho<\rho^*$, we can pick $\epsilon>0$ small enough such that $E[S_\rho]/S\leq(\frac{1}{2}+\epsilon)-(\rho^*-\rho-\frac{1}{n})(\frac{1}{2}-\epsilon)\leq\frac{1}{2}-2\delta$ for a suitable $\delta>0$ when n is large enough. The result follows by combining the above with Lemma 2. ■

F. Proof of Lemma 3

Proof: For any given $\gamma>0$, there exists a γ-net Σ in $\mathcal{R}^{n-m}$ of cardinality less than $(1+\frac{2}{\gamma})^{n-m}$([29]). A γ-net Σ is a set of points in $\mathcal{R}^{n-m}$ such that $\|\mathbf{v}^k\|_2=1$ for all $\mathbf{v}^k$ in Σ and for any $\mathbf{z}\in\mathcal{R}^{n-m}$ with $\|\mathbf{z}\|_2=1$, there exists some $\mathbf{v}^k$ such that $\|\mathbf{z}-\mathbf{v}^k\|_2\leq\gamma$.

Since B has i.i.d $\mathcal{N}(0,1)$ entries, then $B\mathbf{v}^k$ has n i.i.d. $\mathcal{N}(0,1)$ entries for every $\mathbf{v}^k$. From Corollary 1 and 2, we know that given any $\rho<\rho^*$, for some $\delta>0$ and for every $\epsilon>0$, there exists $c_2>0$ and c_3 such that with probability at least $1-2e^{-c_2n}-2e^{-c_3n}$, we have

$$S_\rho(B\mathbf{v}^k)\leq(\frac{1}{2}-\delta)S \tag{27}$$

and

$$(1-\epsilon)S\leq S_1(B\mathbf{v}^k)\leq(1+\epsilon)S \tag{28}$$

both hold for *one* vector $\mathbf{v}^k$ in Σ. Then applying union bound, we know that (27) and (28) hold for *all* vectors in Σ with probability at least

$$1-(1+2/\gamma)^{n-m}(2e^{-c_2n}+2e^{-c_3n}). \tag{29}$$

Let $\alpha=m/n$, then as long as α is large enough, say greater than $c_4:=1-\frac{\min(c_2,c_3)}{2\ln(1+2/\gamma)}$, then (29) is greater than $1-e^{-c_5n}$ for some constant $c_5>0$.

For any $\mathbf{z}$ such that $\|\mathbf{z}\|_2=1$, there exists $\mathbf{v}_0$ in Σ such that $\|\mathbf{z}-\mathbf{v}_0\|_2\triangleq\gamma_1\leq\gamma$. Let $\mathbf{z}_1$ denote $\mathbf{z}-\mathbf{v}_0$, then $\|\mathbf{z}_1-\gamma_1\mathbf{v}_1\|_2\triangleq\gamma_2\leq\gamma_1\gamma\leq\gamma^2$ for some $\mathbf{v}_1$ in Σ. Repeating this process, we have

$$\mathbf{z}=\sum_{j\geq0}\gamma_j\mathbf{v}_j \tag{30}$$

where $\gamma_0=1$, $\gamma_j\leq\gamma^j$ and $\mathbf{v}_j\in\Sigma$. Thus for any $\mathbf{z}\in\mathcal{R}^{n-m}$, we have $\mathbf{z}=\|\mathbf{z}\|_2\sum_{j\geq0}\gamma_j\mathbf{v}_j$.

For any index set T with $|T|\leq\lceil\rho n\rceil$,

$$\begin{aligned}\|B_T\mathbf{z}\|_p^p &= \|\mathbf{z}\|_2^p\|\sum_{j\geq0}\gamma_jB_T\mathbf{v}_j\|_p^p\\ &\leq \|\mathbf{z}\|_2^p\sum_{j\geq0}\gamma^{jp}\|B_T\mathbf{v}_j\|_p^p\\ &\leq S\|\mathbf{z}\|_2^p\frac{1-2\delta}{2(1-\gamma^p)},\end{aligned}$$

where the first inequality holds from the triangular inequality and the fact that $\gamma_j\leq\gamma^j$. The second inequality holds with overwhelming probability by (27) and (29).

$$\begin{aligned}\|B\mathbf{z}\|_p^p &= \|\mathbf{z}\|_2^p\|\sum_{j\geq0}\gamma_jB\mathbf{v}_j\|_p^p\\ &\geq \|\mathbf{z}\|_2^p(\|Bv_0\|_p^p-\sum_{j\geq1}\gamma_j^p\|B\mathbf{v}_j\|_p^p)\\ &\geq \|\mathbf{z}\|_2^p(\|B\mathbf{v}_0\|_p^p-\sum_{j\geq1}\gamma^{jp}\|B\mathbf{v}_j\|_p^p)\\ &\geq \|\mathbf{z}\|_2^p((1-\epsilon)S-\sum_{j\geq1}\gamma^{jp}(1+\epsilon)S)\\ &\geq S\|\mathbf{z}\|_2^p\frac{1-2\gamma^p-\epsilon}{1-\gamma^p},\end{aligned}$$

where the first inequality holds from the triangular inequality and the third inequality holds with overwhelming probability by (28) and (29). Thus $\|B_{T^c}\mathbf{z}\|_p^p-\|B_T\mathbf{z}\|_p^p\geq S\|\mathbf{z}\|_2^p\frac{2\delta-2\gamma^p-\epsilon}{1-\gamma^p}$ holds with probability at least $1-e^{-c_5n}$. For the given δ from Corollary 1, we can pick γ and ϵ small enough such that $\|B_{T^c}\mathbf{z}\|_p^p-\|B_T\mathbf{z}\|_p^p\geq\delta S\|\mathbf{z}\|_2^p$. ■

G. Proof of Lemma 4

Proof: We first consider the case that $p = 0$. Now $\mu = E[|X|^0] = 1$, where $X \sim \mathcal{N}(0,1)$. We have $\sum_{i\in T: X_i<0} |X_i|^p = \sum_{i\in T} \mathbf{1}_{\{X_i<0\}}$. Since $P(X_i < 0) = 0.5$ independently for all i in T, then $E[\sum_{i\in T} \mathbf{1}_{\{X_i<0\}}] = \rho n/2$, and from the Chernoff bound, we have

$$P(\sum_{i\in T} \mathbf{1}_{\{X_i<0\}} \geq (1+\epsilon)\rho n/2) \leq e^{-\epsilon^2 \rho n/2},$$

and

$$P(\sum_{i\in T} \mathbf{1}_{\{X_i<0\}} \geq (1-\epsilon)\rho n/2) \leq e^{-\epsilon^2 \rho n/2}.$$

It is easy to see that with probability one $\sum_{i\in T^c} |X_i|^p = \sum_{i\in T^c} \mathbf{1}_{\{X_i\neq 0\}} = (1-\rho)n$ holds. Therefore Lemma 4 follows for $p = 0$.

Now we consider the case that $p \in (0,1)$. Let $\mathbf{X} = [X_1, ..., X_n]^T$. Let $S_{T^-}(\mathbf{X}) = \sum_{i\in T: X_i<0} |X_i|^p$. For any $\mathbf{X}$ and $\mathbf{X}'$,

$$\begin{aligned}
& |S_{T^-}(\mathbf{X}) - S_{T^-}(\mathbf{X}')| \\
= \; & |\sum_{i\in T} |X_i|^p \mathbf{1}_{\{X_i<0\}} - \sum_{i\in T} |X_i'|^p \mathbf{1}_{\{X_i'<0\}}| \\
\leq \; & \sum_{i\in T} \big| |X_i|^p \mathbf{1}_{\{X_i<0\}} - |X_i'|^p \mathbf{1}_{\{X_i'<0\}} \big| \\
\leq \; & \sum_{i\in T} |X_i - X_i'|^p, \qquad (31)
\end{aligned}$$

where the first inequality follows from the triangular inequality. To see why the second inequality holds, we consider three different cases. If both $X_i < 0$ and $X_i' < 0$ hold, then $\big||X_i|^p \mathbf{1}_{\{X_i<0\}} - |X_i'|^p \mathbf{1}_{\{X_i'<0\}}\big| = \big||X_i|^p - |X_i'|^p\big| \leq |X_i - X_i'|^p$ where the inequality holds since $p \in (0,1)$. If both X_i and X_i' are nonnegative, then clearly $\big||X_i|^p \mathbf{1}_{\{X_i<0\}} - |X_i'|^p \mathbf{1}_{\{X_i'<0\}}\big| = 0 \leq |X_i - X_i'|^p$. If only one of X_i and X_i' is negative, we assume $X_i < 0$ without loss of generality, then $\big||X_i|^p \mathbf{1}_{\{X_i<0\}} - |X_i'|^p \mathbf{1}_{\{X_i'<0\}}\big| = |X_i|^p \leq |X_i - X_i'|^p$, where the inequality holds since $X_i < 0$ and $X_i' \geq 0$. Combining the three cases, we know that $\big||X_i|^p \mathbf{1}_{\{X_i<0\}} - |X_i'|^p \mathbf{1}_{\{X_i'<0\}}\big| \leq |X_i - X_i'|^p$ always holds, thus the second inequality in (31) holds.

From the isoperimetric inequality for the Gaussian measure [29], for any set $A \in \mathcal{R}^n$ with measure at least a half, the set $A_t = \{\mathbf{x} \in \mathcal{R}^n : d(\mathbf{x}, A) \leq t\}$ has measure at least $1 - e^{-t^2/2}$, where $d(\mathbf{x}, A) = \inf_{\mathbf{y}\in A} \|\mathbf{x} - \mathbf{y}\|_2$. Let M_{T^-} be the median value of $S_{T^-} = S_{T^-}(\mathbf{X})$. Define set $A = \{\mathbf{x} \in \mathcal{R}^n : S_{T^-}(\mathbf{x}) \leq M_{T^-}\}$, then

$$P(d(\mathbf{x}, A) \leq t) \geq 1 - e^{-t^2/2}.$$

We claim that $d(\mathbf{x}, A) \leq t$ implies that $S_{T^-}(\mathbf{x}) \leq M_{T^-} + (\rho n)^{(1-p/2)} t^p$. If $\mathbf{x} \in A$, then $S_{T^-}(\mathbf{x}) \leq M_{T^-}$, thus the claim holds as $(\rho n)^{1-p/2} t^p$ is nonnegative. If $\mathbf{x} \notin A$, then there exists $\mathbf{x}' \in A$ such that $\|\mathbf{x} - \mathbf{x}'\|_2 \leq t$. For i in T, let $u_i = 1$ and let $v_i = |x_i - x_i'|^p$. From Hölder's inequality,

$$\begin{aligned}
\sum_{i\in T} |x_i - x_i'|^p \;&\leq\; \left(\sum_{i\in T} |u_i|^{2/(2-p)}\right)^{1-p/2} \left(\sum_{i\in T} |v_i|^{2/p}\right)^{p/2} \\
&\leq\; (\rho n)^{(1-p/2)} (t^2)^{p/2} = (\rho n)^{(1-p/2)} t^p \qquad (32)
\end{aligned}$$

From (31) and (32), $|S_{T^-}(\mathbf{x}) - S_{T^-}(\mathbf{x}')| \leq (\rho n)^{(1-p/2)} t^p$. Since $\mathbf{x} \notin A$ and $\mathbf{x}' \in A$, then $S_{T^-}(\mathbf{x}) > M_{T^-} \geq S_{T^-}(\mathbf{x}')$. Thus $S_{T^-}(\mathbf{x}) \leq M_{T^-} + (\rho n)^{(1-p/2)} t^p$, which verifies our claim. Then

$$\begin{aligned}
P(S_{T^-}(\mathbf{x}) \;&\leq\; M_{T^-} + (\rho n)^{(1-p/2)} t^p) \geq P(d(\mathbf{x}, A) \leq t) \\
&\geq\; 1 - e^{-t^2/2}. \qquad (33)
\end{aligned}$$

Similarly,

$$P(S_{T^-}(\mathbf{x}) \geq M_{T^-} - (\rho n)^{(1-p/2)} t^p) \geq 1 - e^{-t^2/2}. \qquad (34)$$

Combining (33) and (34),

$$P(|S_{T^-}(x) - M_{T^-}| \geq (\rho n)^{(1-p/2)} t^p) \leq 2e^{-t^2/2}. \qquad (35)$$

The difference of $E[S_{T^-}]$ and M_{T^-} can be bounded as follows,

$$\begin{aligned}
|E[S_{T^-}] - M_{T^-}| \;&\leq\; E[|S_{T^-} - M_{T^-}|] \\
&=\; \int_0^\infty P(|S_{T^-}(\mathbf{x}) - M_{T^-}| \geq y) dy \\
&\leq\; \int_0^\infty 2e^{-\frac{1}{2} y^{\frac{2}{p}} (\rho n)^{(1-\frac{2}{p})}} dy \\
&=\; (\rho n)^{(1-\frac{p}{2})} \int_0^\infty 2e^{-\frac{1}{2} s^{\frac{2}{p}}} ds
\end{aligned}$$

Note that $c := \int_0^\infty 2e^{-\frac{1}{2}s^{(2/p)}} ds$ is a finite constant for all $p \in (0,1]$. Since $p > 0$, for any $\epsilon > 0$, $c(\rho n)^{(1-\frac{p}{2})} < \epsilon \rho n/4$ when n is large enough.

Let $t = (\frac{\epsilon}{4})^{\frac{1}{p}} \sqrt{\rho n}$, from (35) with probability at least $1 - 2e^{-\frac{1}{2}(\frac{\epsilon}{4})^{\frac{2}{p}} \rho n}$, $|S_{T^-} - M_{T^-}| < \epsilon \rho n/4$. Thus $|S_{T^-} - E[S_{T^-}]| \leq |S_{T^-} - M_{T^-}| + |M_{T^-} - E[S_{T^-}]| < \epsilon \rho n/2$ holds with probability at least $1 - 2e^{-d_1 n}$ for some constant d_1. Since $E[S_{T^-}] = \mu \rho n/2$, where $\mu = E[|X|^p]$, then

$$\frac{1}{2}\rho n(\mu - \epsilon) < \sum_{i\in T: X_i<0} |X_i|^p < \frac{1}{2}\rho n(\mu + \epsilon)$$

holds with probability at least $1 - 2e^{-d_1 n}$.

Similarly we can prove that with probability at least $1 - 2e^{-d_2 n}$ for some $d_2 > 0$,

$$(1-\rho)n(\mu - \epsilon) < \sum_{i\in T^c} |X_i|^p < (1-\rho)n(\mu + \epsilon)$$

holds. Then by a simple union bound, the above two statements hold at the same time with probability at least $1 - 2e^{-d_2 n} - 2e^{-d_2 n}$, thus Lemma 4 follows. ■

H. Proof of Theorem 6

Proof: From Lemma 4, applying similar arguments in the proof of Lemma 3, we get that when $\alpha > c_7$ for some $0 < c_7 < 1$ and n is large enough, with probability $1 - e^{-c_8 n}$ for some $c_8 > 0$,

- $\frac{1}{2}\rho n(\mu - \epsilon) < \sum_{i\in T: B_i \mathbf{v}<0} |B_i \mathbf{v}|^p < \frac{1}{2}\rho n(\mu + \epsilon)$
- $(1-\rho)n(\mu - \epsilon) < \sum_{i\in T^c} |B_i \mathbf{v}|^p < (1-\rho)n(\mu + \epsilon)$

hold for all the vectors $\mathbf{v}$ in a γ-net Σ at the same time. Let $\mathcal{S}$ be the unit sphere in $\mathcal{R}^{n-m}$. Pick any $\mathbf{z} \in \mathcal{S}$, from (30) we have $\mathbf{z} = \sum_{j\geq 0} \gamma_j \mathbf{v}_j$, where $\gamma_0 = 1$, $\mathbf{v}_j \in \Sigma$ for all j and $\gamma_j \leq \gamma^j$.

Given $\mathbf{z}$, let $T^- = \{i \in T : B_i\mathbf{z} < 0\}$. For any i in T^-,

$$\begin{aligned}
|B_i\mathbf{z}|^p &= \big|\sum_{j\geq 0}\gamma_j B_i\mathbf{v}_j\big|^p \\
&= \big|\sum_{j:B_i\mathbf{v}_j<0}\gamma_j B_i\mathbf{v}_j + \sum_{j:B_i\mathbf{v}_j\geq 0}\gamma_j B_i\mathbf{v}_j\big|^p \\
&\leq \big|\sum_{j:B_i\mathbf{v}_j<0}\gamma_j B_i\mathbf{v}_j\big|^p \\
&\leq \sum_{j:B_i\mathbf{v}_j<0}\gamma^{jp}|B_i\mathbf{v}_j|^p
\end{aligned}$$

where the first inequality holds as $B_i\mathbf{z} < 0$. Then

$$\begin{aligned}
\|B_{T^-}\mathbf{z}\|_p^p &\leq \sum_{i\in T^-}\sum_{j:B_i\mathbf{v}_j<0}\gamma^{jp}|B_i\mathbf{v}_j|^p \\
&\leq \sum_{i\in T}\sum_{j:B_i\mathbf{v}_j<0}\gamma^{jp}|B_i\mathbf{v}_j|^p \\
&= \sum_{j\geq 0}\gamma^{jp}\sum_{i\in T:B_i\mathbf{v}_j<0}|B_i\mathbf{v}_j|^p \qquad (36)\\
&< \frac{1}{2(1-\gamma^p)}\rho n(\mu+\epsilon), \qquad (37)
\end{aligned}$$

where the last inequality holds with overwhelming probability.

We also have

$$\begin{aligned}
&\|B_{T^c}\mathbf{z}\|_p^p = \|(\sum_{j\geq 0}\gamma_j B_{T^c}\mathbf{v}_j)\|_p^p \\
\geq\ & \|B_{T^c}\mathbf{v}_0\|_p^p - \sum_{j\geq 1}\gamma^{jp}\|B_{T^c}\mathbf{v}_j\|_p^p \\
>\ & (1-\rho)n(\mu-\epsilon) - \sum_{j\geq 1}\gamma^{jp}(1-\rho)n(\mu+\epsilon) \\
\geq\ & (1-\rho)n\frac{\mu-2\mu\gamma^p-\epsilon}{1-\gamma^p}, \qquad (38)
\end{aligned}$$

where the second inequality holds with overwhelming probability.

Combining (37) and (38), we have for every $\mathbf{z} \in \mathcal{S}$, $\|B_{T^c}\mathbf{z}\|_p^p - \|B_{T^-}\mathbf{z}\|_p^p > \frac{n\mu}{1-\gamma^p}\big(1-\frac{3}{2}\rho-2\gamma^p(1-\rho)-\frac{\epsilon}{\mu}(1-\frac{\rho}{2})\big)$ holds at the same time with overwhelming probability. Then with overwhelming probability, for every non-zero $\mathbf{z} \in \mathcal{R}^{n-m}$, we have $\|B_{T^c}\mathbf{z}\|_p^p - \|B_{T^-}\mathbf{z}\|_p^p > \|\mathbf{z}\|_2^p\frac{n\mu}{1-\gamma^p}\big(1-\frac{3}{2}\rho-2\gamma^p(1-\rho)-\frac{\epsilon}{\mu}(1-\frac{\rho}{2})\big)$. For any $\rho < \frac{2}{3}$, we can pick γ and ϵ small enough such that the righthand side is positive. The result follows by applying Theorem 3 and Theorem 4. ■

I.

Lemma 9. *Given any α and p, there exists a constant $\lambda_{\max}(\alpha,p) > 0$ and some constant $c_{16} > 0$ such that with probability at least $1-e^{-c_{16}n}$, for every $\mathbf{z} \in \mathcal{S}$, $\|B\mathbf{z}\|_p^p < \lambda_{\max}(\alpha,p)n$.*

To help improve the lower bound of the recovery threshold, we would like $\lambda_{\max}(\alpha,p)$ to be as small as possible, while at the same time, the probability that $\|B\mathbf{z}\|_p^p$ exceeds $\lambda_{\max}(\alpha,p)n$ for some $\mathbf{z}$ in $\mathcal{S}$ still has exponential decay to zero. Therefore, in the following proof, besides establishing the existence of $\lambda_{\max}(\alpha,p)$, we make some efforts to reduce the value of $\lambda_{\max}(\alpha,p)$, and $\lambda_{\max}(\alpha,p)$ can be computed following the lines and finally through (43).

Proof: Define $c_{\max} = \frac{1}{n}\max_{\mathbf{z}\in\mathcal{S}}\|B\mathbf{z}\|_p^p$, then for any non-zero vector $\mathbf{z}$, $\|B\mathbf{z}\|_p^p \leq \|\mathbf{z}\|_p^p c_{\max}n$. Let Σ_1 be a γ-net of $\mathcal{S}$ with cardinality at most $(1+2/\gamma)^{n-m}$ [29] and $\gamma > 0$ to be chosen later, and define

$$\eta = \frac{1}{n}\max_{\mathbf{z}\in\Sigma_1}\|B\mathbf{z}\|_p^p.$$

Then from the definition of γ-net, for every $\mathbf{z} \in \mathcal{S}$, there exists $\mathbf{z}' \in \Sigma_1$ such that $\|\mathbf{z}-\mathbf{z}'\|_2 \leq \gamma$. Note that for every $\mathbf{z} \in \mathcal{S}$, $\|B\mathbf{z}\|_p^p \leq \|B\mathbf{z}'\|_p^p + \|B(\mathbf{z}-\mathbf{z}')\|_p^p = \|B\mathbf{z}'\|_p^p + \|\mathbf{z}-\mathbf{z}'\|_2^p\|B\frac{\mathbf{z}-\mathbf{z}'}{\|\mathbf{z}-\mathbf{z}'\|_2}\|_p^p \leq \eta n + \gamma^p c_{\max}n$, where the first inequality follows from the triangular inequality and the second inequality follows from the definition of η and $c_{\max}$. Then $c_{\max}n \leq \eta n + \gamma^p c_{\max}n$, which leads to

$$c_{\max} \leq \eta/(1-\gamma^p). \qquad (39)$$

To characterize $c_{\max}$, we first characterize η. For any $a > E[|X|^p]$ where $X \sim \mathcal{N}(0,1)$, we calculate the probability that $\|B\mathbf{z}\|_p^p \geq an$ for some $\mathbf{z}$ in Σ_1. Note that $\forall \mathbf{z} \in \mathcal{S}$, $B_i\mathbf{z}$ ($i = 1,...,n$) are i.i.d. $\mathcal{N}(0,1)$ random variables where B_i is the i^{th} row of B. Then

$$\begin{aligned}
&P(\eta \geq a) = P(\exists \mathbf{z} \in \Sigma_1 \text{s.t. } \|B\mathbf{z}\|_p^p \geq an) \\
\leq\ & \sum_{\mathbf{z}\in\Sigma_1}P(\|B\mathbf{z}\|_p^p \geq an) \\
\leq\ & (1+2/\gamma)^{n-m}\min_{t>0}e^{-tan}E[e^{t\sum_i|B_i\mathbf{z}|^p}] \\
=\ & (1+2/\gamma)^{(1-\alpha)n}\min_{t>0}e^{-tan}(E[e^{t|X|^p}])^n \\
=\ & e^{((1-\alpha)\log(1+\frac{2}{\gamma})+\min_{t>0}(\log(E[e^{t|X|^p}])-at))n}, \qquad (40)
\end{aligned}$$

where $X \sim \mathcal{N}(0,1)$, the first inequality follows from the union bound, and the second inequality follows from the Chernoff bound.

To obtain a good upper bound of η, we would like to find the smallest a such that the upper bound of $P(\eta \geq a)$ in (40) still exponentially decays to zero, note that we do not care about the decay rate here. To solve the minimization problem in the righthand side of (40), note that $\log(E[e^{t|X|^p}])$ is the cumulant generating function and is known to be convex [16] with respect to t, then $\log(E[e^{t|X|^p}])-at$ is also convex, and its minimum is achieved where its first derivative with respect to t is 0. Define $t^* := \arg\min_t[\log(E[e^{t|X|^p}])-at]$, then we have

$$\begin{aligned}
0 &= \frac{d[\log(E[e^{t|X|^p}])-at]}{dt}\Big|_{t=t^*} \\
&= \frac{E[|X|^p e^{t^*|X|^p}]}{E[e^{t^*|X|^p}]} - a. \qquad (41)
\end{aligned}$$

(41) determines t^* given a. The derivative of t^* with respect to a is

$$\begin{aligned}
\frac{dt^*}{da} &= \big(\frac{da}{dt^*}\big)^{-1} \\
&= \frac{(E[e^{t^*|X|^p}])^2}{E[e^{t^*|X|^p}]E[|X|^{2p}e^{t^*|X|^p}]-(E[|X|^p e^{t^*|X|^p}])^2},
\end{aligned}$$

Note that $(E[|X|^p e^{t^*|X|^p}])^2 = (E[e^{t^*|X|^p/2} \cdot (|X|^p e^{t^*|X|^p/2})])^2 < E[e^{t^*|X|^p}]E[|X|^{2p}e^{t^*|X|^p}]$, where the inequality follows from Cauchy-Schwarz inequality and the fact that the functions $e^{t^*|X|^p}$ and $|X|^{2p}e^{t^*|X|^p}$ are not linearly dependent. Thus, $\frac{dt^*}{da} > 0$. Since when $a = E[|X|^p]$, we have $t^* = 0$ from (41), then when $a > E[|X|^p]$ we have $t^* > 0$. Thus when $a > E[|X|^p]$, it holds that $t^* = \arg\min_{t>0}(\log(E[e^{t|X|^p}]) - at)$. Given a, we can numerically compute t^* by (41) and plug it into (40) to obtain an upper bound of $P(\eta \geq a)$. Then the question is how small can a be while the exponent on the righthand side of (40) is still negative. Note that given γ, the exponent on the righthand side of (40) is negative when a is large enough. To see this, if we let $t = 2(1-\alpha)\log(1+2/\gamma)/a$, then $\log(E[e^{t|X|^p}]) - at$ goes to $-2(1-\alpha)\log(1+2/\gamma)$ as a goes to infinity. Thus, when a is sufficiently large, $\min_{t>0}\log(E[e^{t|X|^p}]) - at < -(1-\alpha)\log(1+2/\gamma) < 0$, in other words, the exponent on the righthand side of (40) is negative. Pick $\hat{a}(\alpha,p,\gamma)$ such that the exponent on the righthand side of (40) is negative for all $a \geq \hat{a}(\alpha,p,\gamma)$, and positive for all $a \leq \hat{a}(\alpha,p,\gamma) - \epsilon$ for a very small $\epsilon > 0$. Therefore

$$(1-\alpha)\log(1+\frac{2}{\gamma}) + \min_{t>0}(\log(E[e^{t|X|^p}]) - \hat{a}(\alpha,p,\gamma)t) < 0. \tag{42}$$

Then there exists some constant $c_{16} > 0$ such that

$$\begin{aligned} & P(\eta \geq \hat{a}(\alpha,p,\gamma)) \\ \leq\ & e^{((1-\alpha)\log(1+\frac{2}{\gamma})+\min_{t>0}(\log(E[e^{t|X|^p}])-\hat{a}(\alpha,p,\gamma)t))n} \\ =\ & e^{-c_{16}n}. \end{aligned}$$

Then the probability that $\|B\mathbf{z}\|_p^p \geq \frac{\hat{a}(\alpha,p,\gamma)n}{1-\gamma^p}$ holds for some $\mathbf{z} \in \mathcal{S}$ is

$$\begin{aligned} P(\max_{\mathbf{z}\in\mathcal{S}}\|B\mathbf{z}\|_p^p \geq \frac{\hat{a}(\alpha,p,\gamma)n}{1-\gamma^p}) &= P(c_{\max} \geq \frac{\hat{a}(\alpha,p,\gamma)}{1-\gamma^p}) \\ &\leq P(\frac{\eta}{1-\gamma^p} \geq \frac{\hat{a}(\alpha,p,\gamma)}{1-\gamma^p}) \\ &\leq e^{-c_{16}n}, \end{aligned}$$

where the first inequality follows from (39). Thus, for all $\gamma \in (0,1)$, $\hat{a}(\alpha,p,\gamma)n/(1-\gamma^p)$ can be viewed as an upper bound of $\|B\mathbf{z}\|$ for all $\mathbf{z} \in \mathcal{S}$ in the sense that the probability that $\|B\mathbf{z}\|_p^p \geq \hat{a}(\alpha,p,\gamma)n/(1-\gamma^p)$ for some $\mathbf{z} \in \mathcal{S}$ decays exponentially to zero for every γ in $(0,1)$. Since we would like such an upper bound to be as small as possible, we let

$$\lambda_{\max}(\alpha,p) = \min_{\gamma\in(0,1)} \hat{a}(\alpha,p,\gamma)/(1-\gamma^p), \tag{43}$$

then with probability at least $1 - e^{-c_{16}n}$ for some $c_{16}(\alpha,p,\lambda_{\max}) > 0$, for every $\mathbf{z} \in \mathcal{S}$, $\|B\mathbf{z}\|_p^p < \lambda_{\max}n$ holds. Thus, the statement follows. ∎

J. *Calculation of $\lambda_{\min}(\alpha,p)$ in Lemma 5*

Given α and p, define

$$c_{\max} = \frac{1}{n}\sup_{\mathbf{z}\in\mathcal{S}}\|B\mathbf{z}\|_p^p = \frac{1}{n}\max_{\mathbf{z}\in\mathcal{S}}\|B\mathbf{z}\|_p^p,$$

where the second equality holds by compactness. Thus, for any vector $\mathbf{z}$, $\|B\mathbf{z}\|_p^p \leq \|\mathbf{z}\|_p^p c_{\max} n$. Define

$$c_{\min} = \frac{1}{n}\min_{\mathbf{z}\in\mathcal{S}}\|B\mathbf{z}\|_p^p.$$

Pick a γ-net Σ_2 of $\mathcal{S}$ with cardinality at most $(1+2/\gamma)^{n-m}$ [29] and $\gamma > 0$ to be chosen later, we define

$$\theta = \frac{1}{n}\min_{\mathbf{z}\in\Sigma_2}\|B\mathbf{z}\|_p^p.$$

Then for every $\mathbf{z} \in \mathcal{S}$, there exists $\mathbf{z}' \in \Sigma_2$ such that $\|\mathbf{z} - \mathbf{z}'\|_2 \leq \gamma$. We have

$$\|B\mathbf{z}\|_p^p \geq \|B\mathbf{z}'\|_p^p - \|B(\mathbf{z}-\mathbf{z}')\|_p^p \geq \theta n - \gamma^p c_{\max} n, \tag{44}$$

where the first inequality follows from triangular inequality and the second inequality follows from the definition of $c_{\max}$. Since (44) holds for every $\mathbf{z}$ in $\mathcal{S}$, we have

$$c_{\min} \geq \theta - \gamma^p c_{\max}. \tag{45}$$

We aim to find a value $\lambda_{\min}(\alpha,p)$ as large as possible such that $c_{\min} > \lambda_{\min}(\alpha,p)$ still holds with overwhelming probability. We will calculate a "lower bound" of θ and an "upper bound" of $c_{\max}$, and then obtain a "lower bound" of $c_{\min}$ by (45).

We first consider the lower bound of θ. For any constant $b > 0$, we will calculate the probability that θ is less than b. We want to obtain a value b large enough but this probability still decays exponentially to 0. And we treat such a value as the lower bound of θ. Given any constant $b > 0$,

$$\begin{aligned} & P(\theta \leq b) = P(\exists \mathbf{z}\in\Sigma_2 \text{ s.t. } \|B\mathbf{z}\|_p^p \leq bn) \\ \leq\ & \sum_{\mathbf{z}\in\Sigma_2} P(\|B\mathbf{z}\|_p^p \leq bn) \\ \leq\ & (1+2/\gamma)^{n-m}e^{tbn}E[e^{-t\sum_i|B_i\mathbf{z}|^p}], \quad \forall t>0 \\ =\ & (1+2/\gamma)^{(1-\alpha)n}e^{tbn}E[e^{-t|X|^p}]^n, \quad \forall t>0 \\ =\ & e^{((1-\alpha)\log(1+2/\gamma)+\log(E[e^{-t|X|^p}])+bt)n}, \quad \forall t>0, \end{aligned} \tag{46}$$

where $X \sim \mathcal{N}(0,1)$, and the first inequality follows from the union bound. The second inequality follows from the Chernoff bound and the fact that $P(\|B\mathbf{z}\|_p^p \leq bn)$ is the same for all $\mathbf{z} \in \Sigma_2$ since B has i.i.d. $\mathcal{N}(0,1)$ entries. Note that

$$\begin{aligned} E[e^{-t|X|^p}] &= \sqrt{2/\pi}\int_0^\infty e^{-tx^p}e^{-\frac{1}{2}x^2}dx \\ &= t^{-\frac{1}{p}}\sqrt{2/\pi}\int_0^\infty e^{-y^p}e^{-\frac{1}{2}(t^{-\frac{1}{p}}y)^2}dy. \quad (47) \\ &\leq t^{-\frac{1}{p}}\sqrt{2/\pi}\int_0^\infty e^{-y^p}dy \\ &= t^{-\frac{1}{p}}\sqrt{2/\pi}\,\Gamma(1/p)/p, \quad (48) \end{aligned}$$

where (47) holds from changing variables using $x = t^{-\frac{1}{p}}y$, and the inequality follows from the fact that $e^{-\frac{1}{2}(t^{-\frac{1}{p}}y)^2} \leq 1$ for all $y \geq 0$. When $t > 1$, then $t^{-\frac{1}{p}} < 1$, then from (47) we have

$$E[e^{-t|X|^p}] \geq t^{-\frac{1}{p}}\sqrt{2/\pi}\int_0^\infty e^{-y^p-\frac{1}{2}y^2}dy.$$

Since $\int_0^\infty e^{-y^p-\frac{1}{2}y^2}dy$ exists and is positive, then combining (48) and (49), we have when $t>1$,

$$E[e^{-t|X|^p}] = \Theta(t^{-\frac{1}{p}}). \tag{49}$$

Since (46) holds for all $t>0$, we let $t=\gamma^{-p(1-\alpha+\epsilon)}>1$ for any ϵ such that $0<\epsilon\leq\alpha$ and let $b=1/t$, then from (46) we have

$$P(\theta \leq \gamma^{p(1-\alpha+\epsilon)}) \leq e^{((1-\alpha)\log(1+2/\gamma)+\log(\Theta(\gamma^{1-\alpha+\epsilon}))+1)n}.$$

Note that since $\epsilon>0$, when γ is sufficiently small, we have

$$(1-\alpha)\log(1+\frac{2}{\gamma})+\log(\Theta(\gamma^{1-\alpha+\epsilon}))+1<0. \tag{50}$$

Therefore when $\gamma\leq\xi$ for some small enough $\xi>0$, there exists $\kappa>0$ (depending on γ and ϵ) such that

$$P(\theta \leq \gamma^{p(1-\alpha+\epsilon)}) \leq e^{-\kappa n}. \tag{51}$$

Thus, for every $\epsilon\in(0,\alpha)$ and for all $\gamma\leq\xi$ with some ξ depending on ϵ, the probability that $\theta\leq\gamma^{p(1-\alpha+\epsilon)}$ decays exponentially to zero, though the decaying rate depends on ϵ and γ.

Lemma 9 indicates that there exists $\lambda_{\max}(\alpha,p)$ and $c_{16}>0$ such that

$$P(c_{\max}<\lambda_{\max}(\alpha,p)) \geq 1-e^{-c_{16}n}. \tag{52}$$

Then after characterizing θ and $c_{\max}$ separately, we are ready to characterize $c_{\min}$. We have

$$\begin{aligned}
& P(c_{\min}\leq\gamma^{p(1-\alpha+\epsilon)}-\gamma^p\lambda_{\max}(\alpha,p)) \\
\leq\ & P(\theta-\gamma^p c_{\max}\leq\gamma^{p(1-\alpha+\epsilon)}-\gamma^p\lambda_{\max}(\alpha,p)) \\
\leq\ & P(\theta\leq\gamma^{p(1-\alpha+\epsilon)})+P(c_{\max}\geq\lambda_{\max}(\alpha,p)) \\
\leq\ & e^{-\kappa n}+e^{-c_{12}n},
\end{aligned}$$

where the first inequality follows from (45), and the last inequality follows from (51) and (52). Then for every $\epsilon\in(0,\alpha)$, for all $\gamma\leq\xi(\epsilon)$, there exists constant $c_9>0$ (depending on ϵ and γ) such that $P(c_{\min}\leq\gamma^{p(1-\alpha+\epsilon)}-\gamma^p\lambda_{\max}(\alpha,p))\leq e^{-c_9 n}$. Given $\lambda_{\max}(\alpha,p)$, let

$$\lambda_{\min}(\alpha,p)=\max_{0<\epsilon<\alpha,0<\gamma\leq\xi(\epsilon)}\gamma^{p(1-\alpha+\epsilon)}-\gamma^p\lambda_{\max}(\alpha,p). \tag{53}$$

Note that since $1-\alpha+\epsilon<1$, $\gamma^{p(1-\alpha+\epsilon)}-\gamma^p\lambda_{\max}>0$ when γ is sufficiently small, therefore $\lambda_{\min}>0$, and Lemma 5 follows.

K. Calculation of $\rho^*(\alpha,p)$ in Lemma 6

For any given set $T\subset\{1,2,...,n\}$ with $|T|=\rho n$ ($0<\rho<1$), define

$$d_{\max}=\frac{1}{n}\max_{\mathbf{z}\in\mathcal{S}}\|B_T\mathbf{z}\|_p^p.$$

Since B has i.i.d. Gaussian entries, then the distribution of $d_{\max}$ is the same for any T with $|T|=\rho n$. Given a γ-net Σ_3 of $\mathcal{S}$ with cardinality at most $(1+2/\gamma)^{n-m}$ and $\gamma>0$ to be chosen later, define

$$\tau=\frac{1}{n}\max_{\mathbf{z}\in\Sigma_3}\|B_T\mathbf{z}\|_p^p.$$

Then for every $\mathbf{z}\in\mathcal{S}$, there exists $\mathbf{z}'\in\Sigma_3$ such that $\|\mathbf{z}-\mathbf{z}'\|_2\leq\gamma$. Then for every $\mathbf{z}\in\mathcal{S}$, we have $\|B_T\mathbf{z}\|_p^p\leq\|B_T\mathbf{z}'\|_p^p+\|B_T(\mathbf{z}-\mathbf{z}')\|_p^p\leq\tau n+\gamma^p d_{\max}n$. That means $d_{\max}n\leq\tau n+\gamma^p d_{\max}n$, which implies

$$d_{\max}\leq\tau/(1-\gamma^p). \tag{54}$$

Given $\lambda_{\min}(\alpha,p)$ (denoted by $\lambda_{\min}$ here for simplicity), in order to obtain $\rho^*(\alpha,p)$ in Lemma 6, we essentially need to find the largest ρ such that the probability that $d_{\max}\geq\lambda_{\min}/2$ holds for some support T with $|T|=\rho n$ can still decay exponentially to 0. From (54), we first consider the probability that $\tau\geq\lambda_{\min}(1-\gamma^p)/2$ holds for a given set T.

$$\begin{aligned}
& P(\tau\geq\lambda_{\min}(1-\gamma^p)/2,\ \text{given } T) \\
=\ & P(\exists\mathbf{z}\in\Sigma_3 \text{ s.t. } \|B_T\mathbf{z}\|_p^p\geq\lambda_{\min}(1-\gamma^p)n/2) \\
\leq\ & \sum_{\mathbf{z}\in\Sigma_3}P(\|B_T\mathbf{z}\|_p^p\geq\frac{\lambda_{\min}(1-\gamma^p)n}{2}) \\
=\ & \sum_{\mathbf{z}\in\Sigma_3}P(\sum_{i\in T}|B_i\mathbf{z}|^p\geq\frac{\lambda_{\min}(1-\gamma^p)n}{2}) \\
\leq\ & (1+2/\gamma)^{n-m}\min_{t>0}e^{-t\lambda_{\min}(1-\gamma^p)n/2}E[e^{t\sum_{i\in T}|B_i\mathbf{z}|^p}] \\
=\ & (1+2/\gamma)^{(1-\alpha)n}\min_{t>0}e^{-t\lambda_{\min}(1-\gamma^p)n/2}(E[e^{t|X|^p}])^{\rho n} \\
=\ & e^{((1-\alpha)\log(1+\frac{2}{\gamma})+\min_{t>0}(\rho\log(E[e^{t|X|^p}])-t\lambda_{\min}(1-\gamma^p)/2))n}
\end{aligned} \tag{55}$$

where $X\sim\mathcal{N}(0,1)$, the first inequality follows from the union bound and the second inequality follows from the Chernoff bound. Note that since B has i.i.d. $\mathcal{N}(0,1)$ entries, (55) holds for any T as long as $|T|=\rho n$.

Now consider the probability that $\|B_T\mathbf{z}\|_p^p\geq\frac{1}{2}\lambda_{\min}n$ for some $\mathbf{z}\in\mathcal{S}$ and T with $|T|=\rho n$.

$$\begin{aligned}
& P(\exists\mathbf{z}\in\mathcal{S},\exists T,\ \text{s.t. } |T|=\rho n,\|B_T\mathbf{z}\|_p^p\geq\lambda_{\min}n/2) \\
\leq\ & \binom{n}{\rho n}P(\exists\mathbf{z}\in\mathcal{S} \text{ s.t. } \|B_T\mathbf{z}\|_p^p\geq\lambda_{\min}n/2, \\
& \text{for given } T\subset\{1,2,...,n\} \text{ and } |T|=\rho n) \\
=\ & \binom{n}{\rho n}P(d_{\max}\geq\lambda_{\min}/2,\ \text{given } T) \\
\leq\ & \binom{n}{\rho n}P(\tau/(1-\gamma^p)\geq\lambda_{\min}/2,\ \text{given } T) \\
=\ & \binom{n}{\rho n}P(\tau\geq\lambda_{\min}(1-\gamma^p)/2,\ \text{given } T) \\
\leq\ & 2^{nH(\rho)}e^{((1-\alpha)\log(1+2/\gamma)+\min_{t>0}(\rho\log(E[e^{t|X|^p}])-t\lambda_{\min}(1-\gamma^p)/2))n} \\
=\ & e^{(H(\rho)\log2+(1-\alpha)\log(1+2/\gamma)+\min_{t>0}(\rho\log(E[e^{t|X|^p}])-t\lambda_{\min}(1-\gamma^p)/2))n}
\end{aligned} \tag{56}$$

where the first inequality follows from the union bound, the second inequality follows from (54), and the third inequality follows from (55) and the fact that $\binom{n}{\rho n}\leq2^{nH(\rho)}$, where $H(\rho)=-\rho\log(\rho)-(1-\rho)\log(1-\rho)$.

To obtain a good upper bound of $P(\exists\mathbf{z}\in\mathcal{S},\exists T,\text{s.t. }|T|\leq\rho n,\|B_T\mathbf{z}\|_p^p\geq\lambda_{\min}n/2)$, we first would like to solve the minimization problem on the righthand side of (56). Note that $\log(E[e^{t|X|^p}])$ is the cumulant generating function and is known to be convex [16] with respect to t. Then $\rho\log(E[e^{t|X|^p}])-t\lambda_{\min}(1-\gamma^p)/2$ is also convex, then its minimum is achieved where its first derivative with respect to t

is 0. Define $t^* := \arg\min_t [\rho \log(E[e^{t|X|^p}]) - t\lambda_{\min}(1-\gamma^p)/2]$. We have

$$\begin{aligned} 0 &= \frac{d[\rho \log(E[e^{t|X|^p}]) - t\lambda_{\min}(1-\gamma^p)/2]}{dt}\Big|_{t=t^*} \\ &= \frac{\rho E[|X|^p e^{t^*|X|^p}]}{E[e^{t^*|X|^p}]} - \lambda_{\min}(1-\gamma^p)/2, \end{aligned}$$

which is equivalent to

$$\rho = \frac{\lambda_{\min}(1-\gamma^p)E[e^{t^*|X|^p}]}{2E[|X|^p e^{t^*|X|^p}]}. \tag{57}$$

(57) determines t^* given ρ, $\lambda_{\min}$ and γ. The derivative of t^* with respect to ρ is

$$\frac{dt^*}{d\rho} = \left(\frac{d\rho}{dt^*}\right)^{-1} = \frac{2(E[|X|^p e^{t^*|X|^p}])^2}{\lambda_{\min}(1-\gamma^p)\left((E[|X|^p e^{t^*|X|^p}])^2 - E[e^{t^*|X|^p}]E[|X|^{2p}e^{t^*|X|^p}]\right)}.$$

Note that $(E[|X|^p e^{t^*|X|^p}])^2 = (E[e^{t^*|X|^p/2} \cdot (|X|^p e^{t^*|X|^p/2})])^2 < E[e^{t^*|X|^p}]E[|X|^{2p}e^{t^*|X|^p}]$, where the inequality follows from Cauchy-Schwarz inequality and the fact that functions $e^{t^*|X|^p}$ and $|X|^{2p}e^{t^*|X|^p}$ are not linearly dependent. Therefore from (58) we know $\frac{dt^*}{d\rho} < 0$. Since when $\rho = \lambda_{\min}(1-\gamma^p)/(2E[|X|^p])$, one can obtain from (57) that $t^* = 0$, therefore when $\rho < \lambda_{\min}(1-\gamma^p)/(2E[|X|^p])$, the corresponding t^* is always positive. Thus, when $\rho < \lambda_{\min}(1-\gamma^p)/(2E[|X|^p])$, t^* defined in (57) is the solution to $\min_{t>0}(\rho \log(E[e^{t|X|^p}]) - t\lambda_{\min}(1-\gamma^p)/2)$. Given ρ, γ, and α, we can numerically compute t^* by (57) and plug it into (56) to obtain an upper bound of $P(\exists \mathbf{z} \in \mathcal{S}, \exists T, \text{ s.t. } |T| \le \rho n, \|B_T\mathbf{z}\|_p^p \ge \lambda_{\min} n/2)$.

Now that given α and $\lambda_{\min}$, for any ρ, (56) provides an upper bound of the probability that there exists some $\mathbf{z} \in \mathcal{S}$ and some T with $|T| = \rho n$ such that $\|B_T\mathbf{z}\|_p^p \ge \lambda_{\min} n/2$ holds. The next question is how large ρ could be such that this upper bound still decays exponentially to zero. The largest ρ is indeed the $\rho^*(\alpha, p)$ we would like to calculate.

Note that given α, p, and $\lambda_{\min}$, for every γ, as ρ goes to 0, $H(\rho)$ goes to 0, and $\min_{t>0}(\rho\log(E[e^{t|X|^p}]) - t\lambda_{\min}(1-\gamma^p)/2$ goes to $-\infty$, thus, there exists $\hat{\rho}(\alpha, p, \gamma) > 0$ such that the exponent on the righthand side of (56) is negative for all $\rho \le \hat{\rho}(\alpha, p, \gamma)$, and is positive for all $\rho > \hat{\rho}(\alpha, p, \gamma) + \epsilon$ for some very small $\epsilon > 0$. In other words, for each γ, $P(\exists \mathbf{z} \in \mathcal{S}, \exists T, \text{ s.t. } |T| = \hat{\rho}(\alpha, p, \gamma)n, \|B_T\mathbf{z}\|_p^p \ge \lambda_{\min} n/2) \le e^{-cn}$ for some positive c depending on γ. We then optimize $\hat{\rho}(\alpha, p, \gamma)$ over $\gamma \in (0,1)$, and let

$$\rho^*(\alpha, p) = \max_{\gamma \in (0,1)} \hat{\rho}(\alpha, p, \gamma),$$

then with probability at least $1 - e^{-c_{10}n}$ for some $c_{10} > 0$, for every $\mathbf{z} \in \mathcal{S}$ and for every set $T \subset \{1, 2, ..., n\}$ with $|T| \le \rho^*(\alpha, p)n$, $\|B_T\mathbf{z}\|_p^p < \lambda_{\min} n/2$ holds simultaneously. Then Lemma 6 follows.

L. Proof of Theorem 7

Proof: Let $\mathcal{S}$ be the unit sphere in $\mathcal{R}^{n-m}$. Then

$$\begin{aligned} & P(\text{Strong recovery succeeds to recover} \\ & \quad \text{vectors up to } \rho^*(\alpha,p)n\text{-sparse}) \\ = \; & P(\forall \text{ non-zero } \mathbf{z} \in \mathcal{R}^{n-m}, \forall T \text{ with } |T| = \rho^*(\alpha,p)n, \\ & \|B_T\mathbf{z}\|_p^p < \frac{1}{2}\|B\mathbf{z}\|_p^p) \\ = \; & P(\forall \mathbf{z} \in \mathcal{S}, \forall T \text{ with } |T| = \rho^*(\alpha,p)n, \\ & \|B_T\mathbf{z}\|_p^p < \frac{1}{2}\|B\mathbf{z}\|_p^p) \\ \ge \; & P(\forall \mathbf{z} \in \mathcal{S}, \forall T \text{ with } |T| = \rho^*(\alpha,p)n, \\ & \|B_T\mathbf{z}\|_p^p < \frac{1}{2}\lambda_{\min}(\alpha,p)n, \text{ and } \|B\mathbf{z}\|_p^p > \lambda_{\min}(\alpha,p)n) \\ \ge \; & 1 - P(\exists \mathbf{z} \in \mathcal{S}, \text{ s.t. } \|B\mathbf{z}\|_p^p \le \lambda_{\min}(\alpha,p)n) \\ & -P(\exists \mathbf{z} \in \mathcal{S}, \exists T \text{ with } |T| = \rho^*(\alpha,p)n \text{ s.t.} \\ & \|B_T\mathbf{z}\|_p^p \ge \lambda_{\min}(\alpha,p)n/2) \\ = \; & 1 - e^{-c_9 n} - e^{-c_{10}n}, \end{aligned} \tag{58}$$

where the first equality follows from Theorem 1, the second equality holds since for any non-zero $\mathbf{z} \in \mathcal{R}^{n-m}$, $\mathbf{z}/\|\mathbf{z}\|_2 \in \mathcal{S}$. From Lemma 5 we know there exists $c_9 > 0$ such that $P(\exists \mathbf{z} \in \mathcal{S}, \text{ s.t. } \|B\mathbf{z}\|_p^p \le \lambda_{\min}(\alpha,p)n) \le e^{-c_9 n}$, and from Lemma 6 we know there exists $c_{10} > 0$ such that $P(\exists \mathbf{z} \in \mathcal{S}, \exists T \text{ s.t. } \|B_T\mathbf{z}\|_p^p \ge \frac{1}{2}\lambda_{\min}(\alpha,p)n) \le e^{-c_{10}n}$, then there exists $c_{11} > 0$ which depends on α, p and $\lambda_{\min}$ such that the righthand side of (58) is greater than $1 - e^{-c_{11}n}$. Therefore, ℓ_p-minimization can recover all the $\rho^*(\alpha,p)n$-sparse vectors with probability at least $1 - e^{-c_{11}n}$. ■

M. Proof of Lemma 7

Proof: Let $\alpha' = \frac{\alpha - \rho}{1-\rho}$. Define $c'_{\max} = \frac{1}{(1-\rho)n}\max_{\mathbf{z}\in\mathcal{S}} \|B_{T^c}\mathbf{z}\|_p^p$. Let Σ_4 be a γ-net of $\mathcal{S}$ with cardinality at most $(1 + 2/\gamma)^{n-m}$ and γ be the value where $\lambda_{\max}(\alpha', p)$ is achieved in (43). We use $\lambda_{\max}$ to denote $\lambda_{\max}(\alpha', p)$ for simplicity here in the proof. Then from (43) we have

$$\lambda_{\max} = \hat{a}(\alpha', p, \gamma)/(1-\gamma^p), \tag{59}$$

where according to (42), $\hat{a}(\alpha', p, \gamma)$ has the property that

$$(1-\alpha')\log(1 + \frac{2}{\gamma}) + \min_{t>0}(\log(E[e^{t|X|^p}]) - \hat{a}(\alpha', p, \gamma)t) < 0. \tag{60}$$

Combining (59) and (60), we have

$$(1-\alpha')\log(1 + \frac{2}{\gamma}) + \min_{t>0}(\log(E[e^{t|X|^p}]) - \lambda_{\max}(1-\gamma^p)t) < 0. \tag{61}$$

Define

$$\eta' = \frac{1}{(1-\rho)n}\max_{\mathbf{z}\in\Sigma_4} \|B_{T^c}\mathbf{z}\|_p^p.$$

Then by arguments similar to those that lead to (39), we have

$$c'_{\max} \le \eta'/(1-\gamma^p).$$

We first show that with overwhelming probability, $\|B_{T^c}\mathbf{z}\|_p^p < (1-\rho)\lambda_{\max} n$ for all $\mathbf{z}$ in $\mathcal{S}$, or equivalently

$c'_{\max} < \lambda_{\max}$. Note that

$$\begin{aligned}
& P(c'_{\max} \geq \lambda_{\max}) \\
\leq\; & P(\eta'/(1-\gamma^p) \geq \lambda_{\max}) \\
=\; & P(\exists \mathbf{z} \in \Sigma_4 \text{ s.t. } \|B_{T^c}\mathbf{z}\|_p^p \geq (1-\rho)\lambda_{\max}(1-\gamma^p)n) \\
\leq\; & \sum_{\mathbf{z}\in\Sigma_4} P(\|B_{T^c}\mathbf{z}\|_p^p \geq (1-\rho)\lambda_{\max}(1-\gamma^p)n) \\
\leq\; & (1+\frac{2}{\gamma})^{n-m} \min_{t>0} \frac{E[e^{t\sum_{i\in T_c}|B_i\mathbf{z}|^p}]}{e^{t(1-\rho)\lambda_{\max}(1-\gamma^p)n}} \\
=\; & (1+\frac{2}{\gamma})^{(1-\alpha)n} \min_{t>0} \frac{(E[e^{t|X|^p}])^{(1-\rho)n}}{e^{t(1-\rho)\lambda_{\max}(1-\gamma^p)n}} \\
=\; & e^{(1-\rho)n(\frac{1-\alpha}{1-\rho}\log(1+\frac{2}{\gamma})+\min_{t>0}(\log(E[e^{t|X|^p}])-\lambda_{\max}(1-\gamma^p)t))} \\
=\; & e^{(1-\rho)n((1-\alpha')\log(1+\frac{2}{\gamma})+\min_{t>0}(\log(E[e^{t|X|^p}])-\lambda_{\max}(1-\gamma^p)t))} \quad (62)
\end{aligned}$$

where $X \sim \mathcal{N}(0,1)$. Combining (61) and (62), we conclude that there exists $c_{12} > 0$ such that $P(c'_{\max} \geq \lambda_{\max}) \leq e^{-c_{12}n}$. Therefore with probability at least $1 - e^{-c_{12}n}$, for all $\mathbf{z} \in \mathcal{S}$, $\|B_{T^c}\mathbf{z}\|_p^p < (1-\rho)\lambda_{\max}(\alpha', p)n$ holds.

Similarly, define $c'_{\min} = \frac{1}{(1-\rho)n}\min_{\mathbf{z}\in\mathcal{S}} \|B\mathbf{z}\|_p^p$. Let Σ_5 be a $\hat{\gamma}$-net of $\mathcal{S}$ with cardinality at most $(1+2/\hat{\gamma})^{n-m}$ and $\hat{\gamma}$ be the value where $\lambda_{\min}(\alpha', p)$ is achieved, note that from (53) we have

$$\lambda_{\min}(\alpha', p) = \hat{\gamma}^{p(1-\alpha'+\epsilon)} - \hat{\gamma}^p\lambda_{\max}(\alpha', p)$$

for some $\epsilon \in (0, \alpha')$. From (50) we also have that

$$(1-\alpha')\log(1+2/\hat{\gamma}) + \log(\Theta(\hat{\gamma}^{1-\alpha'+\epsilon})) + 1 < 0. \quad (63)$$

We use $\lambda_{\min}$ and $\lambda_{\max}$ to denote $\lambda_{\min}(\alpha', p)$ and $\lambda_{\max}(\alpha', p)$ for simplicity. We define

$$\theta' = \frac{1}{(1-\rho)n}\min_{\mathbf{z}\in\Sigma_5}\|B_{T^c}\mathbf{z}\|_p^p.$$

Using the same arguments as those for (45), we have

$$c'_{\min} \geq \theta' - \gamma^p c'_{\max}.$$

We next show that with overwhelming probability, $\|B_{T^c}\mathbf{z}\|_p^p > (1-\rho)\lambda_{\min}n$ for all $\mathbf{z}$ in $\mathcal{S}$, or equivalently $c'_{\min} > \lambda_{\min}$. Note that the probability that $c'_{\min} \leq \lambda_{\min}$ is

$$\begin{aligned}
& P(c'_{\min} \leq \lambda_{\min}) \\
=\; & P(c'_{\min} \leq \hat{\gamma}^{p(1-\alpha'+\epsilon)} - \hat{\gamma}^p\lambda_{\max}) \\
\leq\; & P(\theta' - \hat{\gamma}^p c'_{\max} \leq \hat{\gamma}^{p(1-\alpha'+\epsilon)} - \hat{\gamma}^p\lambda_{\max}) \\
\leq\; & P(\theta' \leq \hat{\gamma}^{p(1-\alpha'+\epsilon)}) + P(c'_{\max} \geq \lambda_{\max}) \\
\leq\; & P(\theta' \leq \hat{\gamma}^{p(1-\alpha'+\epsilon)}) + e^{-c_{13}n}, \quad (64)
\end{aligned}$$

where the last inequality follows from $P(c'_{\max} \geq \lambda_{\max}) \leq e^{-c_{12}n}$. To calculate $P(\theta' \leq \hat{\gamma}^{p(1-\alpha'+\epsilon)})$, note that

$$\begin{aligned}
& P(\theta' \leq \hat{\gamma}^{p(1-\alpha'+\epsilon)}) \\
=\; & P(\exists \mathbf{z} \in \Sigma_5 \text{ s.t. } \|B_{T^c}\mathbf{z}\|_p^p \leq (1-\rho)\hat{\gamma}^{p(1-\alpha'+\epsilon)}n) \\
\leq\; & \sum_{\mathbf{z}\in\Sigma_5} P(\sum_{i\in T^c}|B_i\mathbf{z}|^p \leq (1-\rho)\hat{\gamma}^{p(1-\alpha'+\epsilon)}n) \\
\leq\; & (1+\frac{2}{\hat{\gamma}})^{(1-\alpha)n} e^{(1-\rho)n}(E[e^{-\hat{\gamma}^{-p(1-\alpha'+\epsilon)}|X|^p}])^{(1-\rho)n} \\
=\; & e^{(1-\rho)n((1-\alpha')\log(1+\frac{2}{\hat{\gamma}})+\log(E[e^{-\hat{\gamma}^{-p(1-\alpha'+\epsilon)}|X|^p}])+1)} \\
=\; & e^{(1-\rho)n((1-\alpha')\log(1+\frac{2}{\hat{\gamma}})+\log(\Theta(\hat{\gamma}^{1-\alpha'+\epsilon}))+1)}, \quad (65)
\end{aligned}$$

where $X \sim \mathcal{N}(0,1)$, the first inequality follows from the union bound, the second inequality follows from the Chernoff bound, and the last equality follows from (49). Combining (63) and (65), we have

$$P(\theta' \leq \hat{\gamma}^{p(1-\alpha'+\epsilon)}) \leq e^{-\kappa n}, \quad (66)$$

for some positive $\kappa > 0$. Thus, from (64) and (66) we have

$$P(c'_{\min} \leq \lambda_{\min}) \leq e^{-\kappa n} + e^{-c_{12}n} \leq e^{-c_{13}n},$$

for some $c_{13} > 0$. Then, with probability at least $1 - e^{-c_{13}n}$, for all $\mathbf{z} \in \mathcal{S}$, $\|B_{T^c}\mathbf{z}\|_p^p > (1-\rho)\lambda_{\min}(\alpha', p)n$. ■

N. Calculation of $\tilde{\lambda}_{\max}(\alpha, p, \rho)$ in Lemma 8

Proof: Define $\tilde{c}_{\max} = \frac{1}{\rho n}\max_{\mathbf{z}\in\mathcal{S}}\|B_{T^-}\mathbf{z}\|_p^p$. Let Σ_6 be a γ-net of $\mathcal{S}$ with cardinality at most $(1+2/\gamma)^{n-m}$ and $\gamma > 0$ to be chosen later, and define $\tilde{\eta} = \frac{1}{\rho n}\max_{\mathbf{z}\in\Sigma_4}\|B_{T^-}\mathbf{z}\|_p^p$. Then from (30), for any $\mathbf{z} \in \mathcal{S}$, $\mathbf{z} = \sum_{j\geq 0}\gamma_j\mathbf{v}_j$ holds, where $\gamma_0 = 1$, $\gamma_j \leq \gamma^j$ and $\mathbf{v}_j \in \Sigma_6$. From (36) we have

$$\begin{aligned}
\|B_{T^-}\mathbf{z}\|_p^p &\leq \sum_{j\geq 0}\gamma^{jp}\sum_{i\in T: B_i\mathbf{v}_j<0}|B_i\mathbf{v}_j|^p \\
&\leq \sum_{j\geq 0}\gamma^{jp}\tilde{\eta}\rho n \\
&= \tilde{\eta}\rho n/(1-\gamma^p), \quad (67)
\end{aligned}$$

where the second inequality follows from the definition of $\tilde{\eta}$. Since (67) holds for every $\mathbf{z} \in \mathcal{S}$, then $\tilde{c}_{\max}\rho n \leq \tilde{\eta}\rho n/(1-\gamma^p)$, which leads to $\tilde{c}_{\max} \leq \tilde{\eta}/(1-\gamma^p)$. For any given $\mathbf{z} \in \mathcal{S}$, define a random variable S_i for each i in T, and S_i is equal to 1 if $B_i\mathbf{z} < 0$ and equal to 0 otherwise. Then $\|B_{T^-}\mathbf{z}\|_p^p = \sum_{i\in T}|B_i\mathbf{z}|^p S_i$.

Given γ, for any $\tilde{a}$, we will characterize the probability that $\tilde{c}_{\max}$ is greater than $\tilde{a}/(1-\gamma^p)$. We will find the smallest value of $\tilde{a}$ such that this probability still exponentially decays to zero, and take the corresponding $\tilde{a}/(1-\gamma^p)$ as an upper bound of $\tilde{c}_{\max}$. Note that

$$\begin{aligned}
& P(\tilde{c}_{\max} \geq \frac{\tilde{a}}{1-\gamma^p}) \leq P(\frac{\tilde{\eta}}{1-\gamma^p} \geq \frac{\tilde{a}}{1-\gamma^p}) \\
=\; & P(\tilde{\eta} \geq \tilde{a}) = P(\exists \mathbf{z} \in \Sigma_6 \text{ s.t. } \|B_{T^-}\mathbf{z}\|_p^p \geq \tilde{a}\rho n) \\
\leq\; & \sum_{\mathbf{z}\in\Sigma_6} P(\|B_{T^-}\mathbf{z}\|_p^p \geq \tilde{a}\rho n) \\
=\; & (1+\frac{2}{\gamma})^{n-m} P(\sum_{i\in T}|B_i\mathbf{z}|^p S_i \geq \tilde{a}\rho n) \\
\leq\; & (1+\frac{2}{\gamma})^{(1-\alpha)n} \min_{t>0} \frac{(E[e^{t|X|^p S}])^{\rho n}}{e^{t\tilde{a}\rho n}} \\
=\; & e^{((1-\alpha)\log(1+\frac{2}{\gamma})+\rho\min_{t>0}(\log(E[e^{t|X|^p S}])-\tilde{a}t))n}, \quad (68)
\end{aligned}$$

where $X \sim \mathcal{N}(0,1)$, $S = 1$ if $X < 0$ and $S = 0$ otherwise.

To solve the minimization problem in the righthand side of (68), note that $\log(E[e^{t|X|^p S}])$ is the cumulant generating function and is convex [16] with respect to t, then $\log(E[e^{t|X|^p S}]) - \tilde{a}t$ is also convex, and its minimum is achieved where its first derivative with respect to t is 0. Define

$t^* := \arg\min_t [\log(E[e^{t|X|^p S}]) - \tilde{a}t]$, then we have

$$\begin{aligned} 0 &= \frac{d[\log(E[e^{t|X|^p S}]) - \tilde{a}t]}{dt}|_{t=t^*} \\ &= \frac{E[|X|^p S e^{t^*|X|^p S}]}{E[e^{t^*|X|^p S}]} - \tilde{a}. \end{aligned} \tag{69}$$

(69) determines t^* given $\tilde{a}$. The derivative of t^* with respect to $\tilde{a}$ is

$$\frac{dt^*}{d\tilde{a}} = \left(\frac{d\tilde{a}}{dt^*}\right)^{-1} = \frac{(E[e^{t^*|X|^p S}])^2}{E[e^{t^*|X|^p S}]E[|X|^{2p}S^2 e^{t^*|X|^p S}] - (E[|X|^p S e^{t^*|X|^p S}])^2} > 0,$$

where the inequality follows from Cauchy-Schwarz inequality. Since when $\tilde{a} = E[|X|^p S]$, $t^* = 0$ from (69), then when $\tilde{a} > E[|X|^p S]$, we have $t^* > 0$. Thus $t^* = \min_{t>0}(\log(E[e^{t|X|^p}]) - at)$ when $\tilde{a} > E[|X|^p S]$. Given $\tilde{a}$, we can numerically compute t^* by (69) and plug it into (68) to obtain an upper bound of $P(\tilde{c}_{\max} \geq \frac{\tilde{a}}{1-\gamma^p})$.

Then the question is how small can $\tilde{a}$ be while the exponent on the righthand side of (68) is still negative. Given γ, the exponent on the righthand side of (68) is negative when $\tilde{a}$ is large enough. To see this, note that if $t = 2(1-\alpha)\log(1+2/\gamma)/(\tilde{a}\rho)$, then $\log(E[e^{t|X|^p S}]) - \tilde{a}t$ goes to $-2(1-\alpha)\log(1+2/\gamma)/\rho$ as $\tilde{a}$ goes to infinity. Thus, when $\tilde{a}$ is sufficiently large, $\rho\min_{t>0}(\log(E[e^{t|X|^p S}]) - \tilde{a}t) < -(1-\alpha)\log(1+2/\gamma)$. Therefore, the exponent on the righthand side of (68) is negative when $\tilde{a}$ is large enough. Thus, we can pick $\bar{a}(\alpha,p,\rho,\gamma)$ such that the exponent on the righthand side of (68) is negative for all $a \geq \bar{a}(\alpha,p,\rho,\gamma)$, and positive for all $a \leq \bar{a}(\alpha,p,\rho,\gamma) - \epsilon$ for some small enough $\epsilon > 0$. Therefore

$$(1-\alpha)\log(1+\frac{2}{\gamma}) + \rho\min_{t>0}(\log(E[e^{t|X|^p S}]) - \bar{a}(\alpha,p,\rho,\gamma)t) < 0.$$

Then there exists some constant $c_{14} > 0$ such that

$$\begin{aligned} & P(\tilde{c}_{\max} \geq \frac{\bar{a}(\alpha,p,\rho,\gamma)}{1-\gamma^p}) \\ \leq\ & e^{((1-\alpha)\log(1+\frac{2}{\gamma}) + \rho\min_{t>0}(\log(E[e^{t|X|^p S}]) - \bar{a}(\alpha,p,\rho,\gamma)t))n} \\ =\ & e^{-c_{14}n}. \end{aligned}$$

Thus, for all $\gamma \in (0,1)$, $\hat{a}(\alpha,p,\rho,\gamma)\rho n/(1-\gamma^p)$ can be viewed as an upper bound of $\|B_{T^-}\mathbf{z}\|$ for all $\mathbf{z} \in \mathcal{S}$ in the sense that the probability that $\|B_{T^-}\mathbf{z}\|_p^p \geq \hat{a}(\alpha,p,\rho,\gamma)\rho n/(1-\gamma^p)$ for some $\mathbf{z} \in \mathcal{S}$ decays exponentially to zero. Since we would like such an upper bound to be as small as possible, we let

$$\tilde{\lambda}_{\max}(\alpha,p,\rho) = \min_{\gamma\in(0,1)} \frac{\bar{a}(\alpha,p,\rho,\gamma)}{1-\gamma^p}, \tag{70}$$

then with overwhelming probability, $\tilde{c}_{\max} < \tilde{\lambda}_{\max}(\alpha,p,\rho)$, or equivalently, for every $\mathbf{z} \in \mathcal{S}$, $\|B_{T^-}\mathbf{z}\|_p^p < (1-\rho)\tilde{\lambda}_{\max}(\alpha,p,\rho)n$. Thus, Lemma 8 follows. ■

O. Proof of Theorem 8

Proof: We first consider the case that there exists some $\rho_w^*(\alpha,p)$ (denoted by ρ_w^* for simplicity here in this proof) such that $\rho_w^* > \rho^*(\alpha,p)$, where $\rho^*(\alpha,p)$ is the lower bound of strong threshold in Theorem 7, and the following inequality holds,

$$\rho_w^*\tilde{\lambda}_{\max}(\alpha,p,\rho_w^*) \leq (1-\rho_w^*)\lambda_{\min}(\frac{\alpha-\rho_w^*}{1-\rho_w^*},p). \tag{71}$$

We will show that such ρ_w^* indeed has the property that Theorem 8 states, i.e. it is a lower bound of weak recovery threshold.

Now consider the probability that ℓ_p-minimization can recover all the $\rho_w^* n$-sparse $\mathbf{x}$ on one fixed support T with one fixed sign pattern. From Theorem 3 we know that $\|B_{T^-}\mathbf{z}\|_p^p < \|B_{T^c}\mathbf{z}\|_p^p$ for all non-zero $\mathbf{z} \in \mathcal{R}^{n-m}$ is a sufficient condition for the success of weak recovery, thus

$$\begin{aligned} & P(\text{Weak recovery succeeds up to } \rho_w^* n\text{-sparse}) \\ \geq\ & P(\forall \text{ non-zero } \mathbf{z} \in \mathcal{R}^{n-m}, \|B_{T^-}\mathbf{z}\|_p^p < \|B_{T^c}\mathbf{z}\|_p^p) \\ =\ & P(\forall \mathbf{z} \in \mathcal{S}, \|B_{T^-}\mathbf{z}\|_p^p < \|B_{T^c}\mathbf{z}\|_p^p) \\ \geq\ & P(\forall \mathbf{z} \in \mathcal{S}, \|B_{T^-}\mathbf{z}\|_p^p < \rho_w^*\tilde{\lambda}_{\max}(\alpha,p,\rho_w^*), \text{ and} \\ & \|B_{T^c}\mathbf{z}\|_p^p > (1-\rho_w^*)\lambda_{\min}(\frac{\alpha-\rho_w^*}{1-\rho_w^*},p)) \\ \geq\ & 1 - e^{-c_{14}n} - e^{-c_{13}n}, \end{aligned} \tag{72}$$

where the first equality holds since for any non-zero $\mathbf{z} \in \mathcal{R}^{n-m}$, $\mathbf{z}/\|\mathbf{z}\|_2 \in \mathcal{S}$, and the second inequality follows from (71). From Lemma 7 we know there exists $c_{13} > 0$ such that $P(\forall \mathbf{z} \in \mathcal{S}, \|B_{T^c}\mathbf{z}\|_p^p > (1-\rho_w^*)\lambda_{\min}(1-\frac{1-\alpha}{1-\rho_w^*},p)) \geq 1 - e^{-c_{13}n}$, and from Lemma 8 we know there exists $c_{14} > 0$ such that $P(\forall \mathbf{z} \in \mathcal{S}, \|B_{T^-}\mathbf{z}\|_p^p < \rho_w^*\tilde{\lambda}_{\max}(\alpha,p,\rho_w^*)) \geq 1 - e^{-c_{14}n}$, then the third inequality of (72) holds from the union bound. Thus, there exists $c_{15} > 0$ such that with probability at least $1 - e^{-c_{15}n}$, ℓ_p-minimization problem can recover all $\rho_w^* n$-sparse vectors on fixed support T with fixed sign pattern, then Theorem 8 holds.

Now consider the case that there is no $\rho_w^* > \rho^*(\alpha,p)$ satisfying (71), where $\rho^*(\alpha,p) > 0$ is the lower bound of strong threshold in Theorem 7, then we can simply define $\rho_w^*(\alpha,p) := \rho^*(\alpha,p)$. Since $\rho_w^*(\alpha,p)$ is a lower bound of strong threshold and then a lower bound of weak threshold, thus Theorem 8 follows. ■

Meng Wang (S'06) received the B.E. (Hon.) and M.S. (Hon.) in Electrical Engineering from Tsinghua University, Beijing, China, in 2005 and 2007, respectively. She is currently working towards the Ph.D degree in the School of Electrical and Computer Engineering at Cornell University. Her research interests include communication networks, signal processing, and nonconvex optimization and its applications.

Weiyu Xu received his B.E. degree in Information Engineering from Beijing University of Posts and Telecommunications in 2002, and a M.S. degree in Electronic Engineering from Tsinghua University in 2005. He received his M.S. degree in 2006 and Ph.D. degree in 2009, both in Electrical Engineering, from California Institute of Technology, with a minor in Applied and Computational Mathematics. He was honored as an outstanding graduate of Tsinghua University in 2005, a recipient of the Information Science and Technology Fellowship at California Institute of Technology, a finalist for the Best Student Paper Award at the 2008 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), and the recipient of the Charles Wilts doctoral thesis award in 2010. His research interests are in signal processing, compressive sensing, communication networks, information and coding theory.

Ao Tang (S01-M07) received the B.E. (Hon.) in electronics engineering from Tsinghua University, Beijing, China, and the Ph.D. in electrical engineering with a minor in applied and computational mathematics from the California Institute of Technology (Caltech), Pasadena, CA, in 1999 and 2006, respectively.

He is currently an Assistant Professor in the School of Electrical and Computer Engineering at the Cornell University, where his research interests center around various networks including communication networks, power networks and on-chip networks.

Dr. Tang was the recipient of the 2006 George B. Dantzig Best Dissertation Award, the 2007 Charles Wilts best dissertation Prize, and a 2009 IBM Faculty Award.