On the Performance of Sparse Recovery via \(\ell_p\)-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 \(x^*\) in \(\mathbb{R}^n\) can be recovered from low-dimensional measurements \(y = Ax^*\) 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 \(\ell_p\)-minimization (0 \leq p \leq 1) as \(p\) varies, where \(\ell_p\)-minimization returns a vector with the least \(\ell_p\) quasi-norm among all the vectors \(x\) satisfying \(Ax = y\). Besides analyzing the performance of strong recovery where \(\ell_p\)-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 \(\ell_p\)-minimization (0 \leq p < 1) where the aim is to recover all the sparse vectors on one support with a fixed sign pattern. When \(\alpha (\equiv \frac{m}{n}) \to 1\), we provide sharp thresholds of the sparsity ratio (i.e., percentage of nonzero entries of a vector) that differentiates the success and failure via \(\ell_p\)-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 \(\ell_1\)-minimization. We also explicitly demonstrate that \(\ell_0\)-minimization (\(p < 1\)) can return a denser solution than \(\ell_1\)-minimization. For any \(\alpha \in (0, 1)\), we provide bounds of the sparsity ratio for strong recovery and weak recovery respectively below which \(\ell_p\)-minimization succeeds. Our bound of strong recovery improves on the existing bounds when \(\alpha\) is large. In particular, regarding the recovery threshold, this paper argues that \(\ell_p\)-minimization has a higher threshold with smaller \(p\) for strong recovery; the threshold is the same for all \(p\) for sectional recovery; and \(\ell_1\)-minimization can outperform \(\ell_p\)-minimization for weak recovery. These are in contrast to traditional wisdom that \(\ell_p\)-minimization, though computationally more expensive, always has better sparse recovery ability than \(\ell_1\)-minimization since it is closer to \(\ell_0\)-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, \(\ell_p\)-minimization, recovery threshold, phase transition.

I. INTRODUCTION

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

Given \(x \in \mathbb{R}^n\), 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 \(x\), which also equals to the \(\ell_0\) norm \(\|x\|_0 := |\{i : x_i \neq 0\}|\). We say \(x\) is \(\rho m\)-sparse if \(|T| = \rho m\) for some \(\rho < 1\). Given the measurement \(y\) and the measurement matrix \(A\), together with the assumption that \(x\) is sparse, one natural estimate of \(x\) is the vector with the least \(\ell_0\) norm that can produce the measurement \(y\). Mathematically, to recover \(x\), we solve the following \(\ell_0\)-minimization problem:

\[
\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{s.t.} \quad Ax = y.
\]

However, (1) is combinatorial and computationally intractable except for small problems, and one commonly used approach is to solve a closely related \(\ell_1\)-minimization problem:

\[
\min_{x \in \mathbb{R}^n} \|x\|_1 \quad \text{s.t.} \quad Ax = y,
\]

where \(\|x\|_1 := \sum_i |x_i|\). (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 \(x\) 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 ([11]–[3], [5], [14], [28], [34], [35]) recently, there has been great research interest in recovering \(x\) by \(\ell_p\)-minimization for 0 < \(p < 1\) ([9], [10], [12], [13], [15], [24], [31]) as follows,

\[
\min_{x \in \mathbb{R}^n} \|x\|_p \quad \text{s.t.} \quad Ax = y.
\]

Recall that \(\|x\|_p = (\sum_i |x_i|^p)^{1/p}\) for \(p > 0\). Though \(\| \cdot \|_p\) is a quasi-norm when \(p < 1\) as it violates the triangular inequality, \(\| \cdot \|_p\) follows the triangular inequality. We say \(x\) can be recovered by \(\ell_p\)-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 \(\ell_1\)-minimization. Then the question is what is the relationship between the sparsity of a vector and the successful recovery with \(\ell_p\)-minimization (\(p < 1\))? How sparse should a vector be so that \(\ell_p\)-minimization can recover it? What is the threshold of sparsity that differentiates the success and failure of recovering by \(\ell_p\)-minimization? [26] shows the sparsity up to which \(\ell_p\)-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 \(\ell_p\)-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 \( \ell_p \)-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 \( \ell_p \)-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 \( \ell_p \)-minimization when \( \alpha \left( = \frac{m}{n} \right) \rightarrow \infty \). 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 to 0.5. This coincides with the intuition that the performance of \( \ell_p \)-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 \( A^{m \times n} \), \( \ell_p \)-minimization can recover all the \( \rho^*(\alpha, p)n \)-sparse vectors with overwhelming probability. \( \rho^*(\alpha, p) \) improves on the existing bounds in large \( \alpha \) region.

We also analyze the performance of \( \ell_p \)-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 \( \alpha \rightarrow 1 \), we provide a sharp threshold \( \rho_{\alpha, p}(p) \) of the ratio of the support size to the dimension which differentiates the success and the failure of \( \ell_p \)-minimization. The weak threshold indicates that if we would like to recover every vector over one support with size less than \( \rho_{\alpha, p}(p)n \) 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, \( \ell_p \)-minimization will recover all such vectors regardless of the amplitudes of the entries of a vector. For \( \ell_1 \)-minimization, given a vector, if we randomly generate a Gaussian matrix and apply \( \ell_1 \)-minimization, then its recovering ability observed in simulation exactly captures the weak recovery threshold, see [17], [18]. Interestingly, when \( \alpha \rightarrow 1 \) and \( n \) is large enough, we prove that the weak threshold \( \rho_{\alpha, p}(p) \) is \( 2/3 \) for all \( p \in [0, 1] \), and is lower than the weak threshold of \( \ell_1 \)-minimization, which is 1. In this region, \( \ell_1 \)-minimization outperforms \( \ell_p \)-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 \( \ell_p \)-minimization (\( p \in (0, 1) \)) can return a vector denser than the original sparse vector while \( \ell_1 \)-minimization successfully recovers the sparse vector. Finally, for every \( \alpha \in (0, 1) \), we provide a positive bound \( \rho_{\alpha, p}(\alpha, p) \) such that \( \ell_p \)-minimization successfully recovers all the \( \rho_{\alpha, p}(\alpha, p)n \)-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 \( \ell_p \)-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 \( \ell_1 \)-minimization can be sparser than that of \( \ell_p \)-minimization (\( p \in (0, 1) \)). Section III provides thresholds of the sparsity ratio of the successful recovery via \( \ell_p \)-minimization for all \( p \in [0, 1] \) both in strong recovery and in weak recovery when the measurement matrix is random Gaussian matrix and \( \alpha \rightarrow 1 \). For \( \alpha \in (0, 1) \), Section IV provides bounds of sparsity ratio below which \( \ell_p \)-minimization is successful in the strong sense and in the weak sense respectively. We compare the performance of \( \ell_p \)-minimization (\( p < 1 \)) and the performance of \( \ell_1 \)-minimization in Section V and provide numerical results in Section VI. Section VII concludes the paper.
be true for weak recovery. For weak recovery, we would like to recover all the vectors on some support $T$ with some sign pattern $\sigma$, 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 $z \in \mathbb{R}^{n-m}$, we define $T^- := \{i \in T : B_z\sigma_i < 0\}$, $T^+ := \{i \in T : B_z\sigma_i > 0\}$, and $T^0 := \{i \in T : B_z\sigma_i = 0\}$. Note that when $B$ is given, $T^-$, $T^+$ and $T^0$ depend on $z$, 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 $T^- = \{i \in T : B_z\sigma_i < 0\}$ and $T^+ = \{i \in T : B_z\sigma_i > 0\}$. 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 $\ell_1$-minimization as follows, please see [21], [26], [32], [36], [38] for this result.

**Theorem 2.** For every nonnegative $x \in \mathbb{R}^n$ on some support $T$, $x$ is the unique solution to $\ell_1$-minimization problem (2) if and only if

$$
\|B_T - z\|_1 < \|B_T^* z\|_1 + \|B_T z\|_1
$$

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

Note that for every nonnegative vector $x$ on a fixed support $T$, the condition to successfully recover it via $\ell_1$-minimization is the same, as stated in Theorem 2. Therefore if one vector $x$ can be successfully recovered, all the other nonnegative sparse vectors on $T$ can also be recovered. Conversely, if some vector $x$ cannot be successfully recovered, then every other nonnegative vector on $T$ cannot be recovered either. However, the condition of successful recovery via $\ell_p$-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.

We further consider the worst case scenario for weak recovery in the sense that the recovery via $\ell_p$-minimization is defined to be “successful” if it can recover all the nonnegative vectors on a specific support. The null space condition for weak recovery in this definition via $\ell_1$-minimization is still the same as that in Theorem 2. We characterize the $\ell_p$-minimization ($p \in (0, 1)$) case in Theorem 3 and the $\ell_p$-minimization case in Theorem 4.

**Theorem 3.** Given any $p \in (0, 1)$, $\ell_p$-minimization (3) can successfully recover all the nonnegative vectors $x \in \mathbb{R}^n$ on some support $T$ if and only if the following condition holds for every non-zero $z \in \mathbb{R}^{n-m}$:

if $T^+$ is not empty, then

$$
\|B_T - z\|_p < \|B_T^* z\|_p;
$$

and if $T^+$ is empty, then

$$
\|B_T - z\|_p < \|B_T z\|_p.
$$

Similarly, the null space condition for the weak recovery of $\ell_0$-minimization is as follows, we skip its proof as it is similar to that of Theorem 3.

**Theorem 4.** $\ell_0$-minimization problem (1) can successfully recover all the nonnegative vectors $x \in \mathbb{R}^n$ on support $T$, if and only if

$$
\|B_T - z\|_0 < \|B_T^* z\|_0
$$

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

For the strong recovery, the null space conditions of $\ell_1$-minimization and $\ell_p$-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 $\ell_p$-minimization ($0 < p < 1$) and $\ell_0$-minimization share a similar form in (6) and (7), the condition of $\ell_1$-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 $\ell_1$ may be significantly different from that of $\ell_p$-minimization for $p \in (0, 1)$. We will further discuss this issue.

B. The solution of $\ell_1$-minimization can be sparser than that of $\ell_p$-minimization ($p \in (0, 1)$)

$\ell_p$-minimization ($p \in (0, 1)$) may not perform as well as $\ell_1$-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 $\ell_1$-minimization can recover the sparse vector while $\ell_p$-minimization ($p \in (0, 1)$) cannot, and the solution of $\ell_p$-minimization is denser than the original sparse vector.

**Example 1.** $\ell_p$-minimization returns a denser solution than $\ell_1$-minimization.

Let the measurement matrix $A$ be a $(6k - 1) \times 6k$ matrix with $\beta \in \mathbb{R}^{6k}$ as a basis of its null space, and $\beta_i = 1$ for all $i \in \{1, \ldots, k\}$, $\beta_i = -1$ for all $i \in \{k + 1, \ldots, 2k\}$, and $\beta_i = 1/64$ for all $i \in \{2k+1, \ldots, 6k\}$. Then every vector in the null space can be represented as $h\beta$, for some $h \in \mathbb{R}$. Note that $\|h\beta\|_1/2 = \left(\frac{33}{33}\right)k|\beta|$, and $\|h\beta\|_1 \leq \left(\frac{33}{33}\right)k|\beta| < \|h\beta\|_1/2$ for all $T \subset \{1, \ldots, 6k\}$ with $|T| \leq \left(\frac{33}{33}\right)k - 1$ and for all $h \in \mathbb{R}$, and $\|h\beta_T\|_1 = \left(\frac{33}{33}\right)k|\beta| \geq \|h\beta\|_1/2$ for all $h$ if $T = \{1, \ldots, \left(\frac{33}{33}\right)k\}$. Then according to Theorem 1, $\ell_1$-minimization can recover all the $(\left(\frac{33}{33}\right)k - 1)$-sparse vectors in $\mathbb{R}^{6k}$, but fails to recover some $(\left(\frac{33}{33}\right)k)$-sparse vector. Similarly, $\|h\beta_T\|_0/2 = \left(\frac{5}{5}\right)k|\beta|$, and $\|h\beta_T\|_0 < \left(\frac{5}{5}\right)k|\beta| < \|h\beta_T\|_0/2$ for all $T \subset \{1, \ldots, 6k\}$ with $|T| \leq \left(\frac{5}{5}\right)k - 1$ and for all $h \in \mathbb{R}$, and $\|h\beta_T\|_0 = \left(\frac{5}{5}\right)k|\beta| \geq \|h\beta\|_0/2$ for all $h$ if $T = \{1, \ldots, \left(\frac{5}{5}\right)k\}$. Therefore by Theorem 1, $\ell_0$-minimization can recover all the $(\left(\frac{5}{5}\right)k - 1)$-sparse vectors in $\mathbb{R}^{6k}$, but fails to recover some $(\left(\frac{5}{5}\right)k)$-sparse vector. Therefore, in terms of strong recovery, $\ell_0$-minimization has a better performance than $\ell_1$-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 $\ell_p$-minimization problem for all $p \in (0, 1]$ can be easily solved. Let $x^*$ denote the sparse vector we would
like to recover, and let \( \mathbf{x} \) denote a vector that can produce the same measurements as \( \mathbf{x}^* \). Then every vector \( \mathbf{x} \) such that \( A\mathbf{x} = A\mathbf{x}^* \) holds should satisfy \( \mathbf{x} = \mathbf{x}^* + h\beta \) for some \( h \in \mathbb{R} \). Then the \( \ell_p \)-minimization problem \((p \in (0, 1])\) is equivalent to

\[
\min_{h \in \mathbb{R}} \| \mathbf{x} + h\beta \|^p_p.
\]

(8)

Given \( \mathbf{x} \) and \( \beta \), \( \| \mathbf{x} + h\beta \|^p_p \) is a function of \( h \). Define set \( S = \{ -\frac{1}{\| \beta \|_1} \mid \beta \neq 0 \} \), 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 \( I_1 \) 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 \( I_i \) \((0 \leq i \leq q)\), \( \| \mathbf{x} + h\beta \|^p_p \) is concave on \( I_i \) for every \( p \in (0, 1) \), and \( \| \mathbf{x} + h\beta \|^p_p \) is linear on \( I_i \). Therefore the minimum value of \( \| \mathbf{x} + h\beta \|^p_p \) \((p \in (0, 1))\) 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 \(+\infty \), \( \| \mathbf{x} + h\beta \|^p_p \) goes to \(+\infty \), then the minimum value of \( \| \mathbf{x} + h\beta \|^p_p \) \((p \in (0, 1))\) 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 \( x^i = x + s_i\beta \) for every \( i = 1, ..., q \), and let \( x^* := \arg \min_{1 \leq i \leq q} \| x^i \|^p_p \), then \( x^* \) is the solution to \( \ell_p \)-minimization problem. We call \( x^i \)’s as “singular vectors”. Therefore, to solve (8), we only need to find all the singular vectors, and the one with the least \( \ell_p \)-norm (or \( \ell_1 \) norm) is the solution to \( \ell_p \)-minimization (or \( \ell_1 \)-minimization). If \( x^* = x^* \), then we say \( x^* \) can be successfully recovered.

Now consider the “weak” recovery as to recover all the nonnegative vectors on support \( T = \{1, \ldots, 2k\} \). According to Theorem 2 and Theorem 3, one can check that \( \ell_1 \)-minimization can indeed recover all the nonnegative vectors on support \( T \), however, \( \ell_0 \)-minimization fails to recover some vectors in this case. For example, consider a 2\( k \)-sparse vector \( x^* \) with \( x^*_i = 9 \) for all \( i \in \{1, \ldots, k\} \), \( x^*_i = 1 \) for all \( i \in \{k+1, \ldots, 2k\} \), and \( x^*_i = 0 \) for all \( i \in \{2k+1, \ldots, 6k\} \). There are three singular vectors in this case: \( x^1 = x^* \), \( x^2 = x^* + \beta \), and \( x^3 = x^* - 9\beta \). Since \( \| x^1 \|_1 = 10^k \), \( \| x^2 \|_1 = 10^k + k/16 \), and \( \| x^3 \|_1 = 10^k + 9^k/16 \), then \( x^3 \) is the solution to \( \ell_1 \)-minimization, and \( x^* \) is successfully recovered. Now consider \( \ell_0 \)-minimization, since \( \| x^1 \|^2_{2,5} = 4k \), \( \| x^2 \|^2_{2,5} = (\sqrt{10^k} + 0.5)^2 \), and \( \| x^3 \|^2_{2,5} = (\sqrt{10^k} + 1.5)^2 \), then \( x^3 \) is the solution to \( \ell_0 \)-minimization, and it is 5\( k \)-sparse. Thus, the solution of \( \ell_0 \)-minimization is a 5\( k \)-sparse vector although the original \( x^* \) is only 2\( k \)-sparse. Therefore \( \ell_0 \)-minimization fails to recover some nonnegative 2\( k \)-sparse vector \( x^* \) while \( x^* \) is the solution to \( \ell_1 \)-minimization, and the solution of \( \ell_0 \)-minimization is denser than the original vector \( x^* \).

III. RECOVERY THRESHOLDS WHEN \( \frac{m}{n} \rightarrow 1 \)

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 \( \alpha = \frac{m}{n} \rightarrow 1 \) and provide recovery thresholds of \( \ell_p \)-minimization for every \( p \in [0, 1] \).

We consider two types of thresholds: one in the strong sense as we require \( \ell_p \)-minimization to recover all \( pn \)-sparse vectors (Section III-A), one in the weak sense as we only require \( \ell_p \)-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 any sparsity below that threshold, \( \ell_p \)-minimization can recover all the sparse vectors either in the strong sense or the weak sense when \( \alpha \) is close enough to 1 and \( n \) is large enough, and for any sparsity above that threshold, \( \ell_p \)-minimization fails to recover some sparse vector no matter how large \( \alpha \) and \( n \) are. These thresholds can be viewed as the limiting behavior of \( \ell_p \)-minimization, since for any constant \( \alpha \in (0, 1) \), the recovery thresholds of \( \ell_p \)-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 \( \rho < \rho^*(p) \), \( \ell_p \)-minimization (3) can recover all \( pn \)-sparse vectors \( \mathbf{x} \) with overwhelming probability when \( \alpha \) is close enough to 1. Our technique here stems from [22], which only focuses on the strong recovery of \( \ell_1 \)-minimization.

We have already discussed in Section II that the performance of \( \ell_q \)-minimization should be no worse than \( \ell_p \)-minimization for strong recovery when \( 0 \leq q < p \leq 1 \). Although there are results about bound of the sparsity below which \( \ell_p \)-minimization can recover all the sparse vectors, no existing result has explicitly calculated the recovery threshold of \( \ell_p \)-minimization for \( p < 1 \) which differentiates the success and failure of \( \ell_p \)-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 \( Y_1, Y_2, ..., Y_m \) 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 \( S_{\rho} \) as \( \sum_{i=1}^{[pn]} Y_i \). Let \( S \) denote \( E[S_1] \), the expected value of \( S_1 \). Then there exists a constant \( \rho^*(p) \) such that \( \lim_{n \rightarrow \infty} \frac{E[S_{\rho^*(p)}]}{S} = \frac{1}{2} \).

\( \rho^* \) 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 \( \rho^* \) 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 \( \rho^* \) is the threshold of successful recovery with \( \ell_p \)-minimization for \( p \in (0, 1] \). First
we state the concentration property of $S_\rho$ in the following lemma.

**Lemma 2.** For any $p \in (0, 1]$, let $X_1, \ldots, X_n, Y_1, \ldots, Y_n, S_\rho$ 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 $1 - 2e^{-c_1n}$, $|S_\rho - E[S_\rho]| \leq \delta S$.

Roughly speaking, Lemma 2 states that $S_\rho$ is concentrated around its expectation $E[S_\rho]$ for every $\rho$. 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 $1 - 2e^{-c_2n}$, $S_\rho \leq (\frac{1}{2} - \delta)S$.

**Corollary 2.** For any $\epsilon > 0$, there exists a constant $c_3 > 0$ such that when $n$ is large enough, with probability at least $1 - 2e^{-c_3n}$, it holds that $(1 - \epsilon)S \leq S_\rho \leq (1 + \epsilon)S$.

From the above two corollaries and applying the union bound, one can easily show that with overwhelming probability the sum of the largest $\lfloor \rho n \rfloor$ 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 $Bz$ where matrix $B^{n \times (n-m)}$ has i.i.d. Gaussian entries and $z$ 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 $\alpha = \frac{p}{n} > c_4$ and $n$ is large enough, with probability at least $1 - e^{-c_5n}$, an $n \times (n-m)$ matrix $B$ with i.i.d. $\mathcal{N}(0, 1)$ entries has the following property: for every non-zero $z \in \mathcal{R}^{n-m}$ and every subset $T \subseteq \{1, \ldots, n\}$ with $|T| \leq \lfloor \rho n \rfloor$, $\|B_Tz\|^p - \|B_Tz\|^p \geq \delta S\|z\|^p$.

We remark here that in Lemma 3 and all the following results in this paper, when we say “with probability at least $1 - e^{-c_5n}$ 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 $\sum_{i \in T^c} |(Bz)_i|^p - \sum_{i \in T} |(Bz)_i|^p \geq \delta S\|z\|^p > 0$ holds for every non-zero $z$ and every set $T$ with $|T| \leq \lfloor \rho n \rfloor$, then from Theorem 1, in this case every $\lfloor \rho n \rfloor$-sparse vector $x$ is the unique solution to the $\ell_p$-minimization problem (3) with overwhelming probability. We can now establish one main result regarding the threshold of successful recovery via $\ell_p$-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 $1 - e^{-c_5n}$, an $m \times n$ matrix $A$ with i.i.d. $\mathcal{N}(0, 1)$ entries has the following property: for every $x \in \mathcal{R}^n$ with its support $T$ satisfying $|T| \leq \lfloor \rho n \rfloor$, $x$ is the unique solution to the $\ell_p$-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 $z$ in $\mathcal{R}^{n-m}$, with overwhelming probability the sum of the largest $\lfloor \rho n \rfloor$ terms of $|Bz|^p$'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 $\ell_p$-minimization fails with overwhelming probability. Therefore, $\ell_p$-minimization fails to recover some $\rho n$-sparse vector with overwhelming probability if $\rho < \rho^*(p)$. Proposition 1 implies that the threshold strictly decreases as $p$ increases. The performance of $\ell_p$-minimization is better than that of $\ell_p$-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 $\alpha \rightarrow 1$. As we shall see, for weak recovery, the threshold of $\ell_p$-minimization is the same for all $p \in [0, 1)$, and is lower than the threshold of $\ell_1$-minimization.

Recall that for successful weak recovery, $\ell_p$-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 $\rho n$-sparse or the size of the support is $\rho n$ instead of using the notation $\lfloor \rho n \rfloor$. 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 $\ell_p$-minimization in this case, we first state the following lemma.

**Lemma 4.** Let $X_1, X_2, \ldots, X_n$ 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 $1 - e^{-c_5n}$ for some constant $c_6 > 0$, the following two properties hold simultaneously:

- $\frac{1}{2}pn(\mu - \epsilon) < \sum_{i \in T \cap X_i < 0} |X_i|^p < \frac{3}{2}pn(\mu + \epsilon)$
- $(1 - \rho)n(\mu - \epsilon) < \sum_{i \in T \cap X_i > 0} |X_i|^p < (1 - \rho)n(\mu + \epsilon)$.

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 $\sum_{i \in T \cap X_i < 0} |X_i|^p < \frac{1}{2}pn(\mu - \epsilon)$ and $\sum_{i \in T \cap X_i > 0} |X_i|^p > \frac{1}{2}pn(\mu + \epsilon)$.
\[ \sum_{i \in T^c} |X_i|^p \] holds with high probability when \(|T| = \rho n < \frac{4}{3} n\). 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 \(Bz\) simultaneously where matrix \(B^{n \times (n-m)}\) has i.i.d. Gaussian entries and \(z\) is any non-zero vector in \(R^{n-m}\). Then we can establish the main result regarding the threshold of successful recovery with \(\ell_p\)-minimization from vectors on one support with the same sign pattern.

**Theorem 6.** For any \(p \in [0, 1]\), given any \(0 < \rho^*_w := \frac{2}{3}\), there exist constants \(c_1 \in (0, 1)\), \(c_2 > 0\) such that when \(\alpha > c_1\) and \(n\) is large enough, with probability at least \(1 - e^{-\alpha n}\), an \(m \times n\) matrix \(A\) with i.i.d. \(N(0, 1)\) entries has the following property: for every nonnegative vector \(x\) on some support \(T\) satisfying \(|T| \leq \rho n\), \(x\) is the unique solution to the \(\ell_p\)-minimization problem.

We remark here that \(\rho^*_w\) is a sharp bound for successful recovery in this setup. For any \(\rho > \rho^*_w\), from Lemma 4, with overwhelming probability that \(\sum_{i \in T' : B_i < 0} |B_i|^{p'} > \sum_{i \in T^c} |B_i|^{p'}\), then Theorem 3 and Theorem 4 indicate that the \(\ell_{p'}\)-minimization \((p \in [0, 1])\) fails to recover some nonnegative \(\rho n\)-sparse vector \(x\) 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, \(\rho^*_w\) in Theorem 6 is the weak recovery threshold for any fixed support and any fixed sign pattern.

Surprisingly, the successful recovery threshold \(\rho^*_w\) 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, \(\ell_1\)-minimization has better recovery performance than \(\ell_{p'}\)-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 \(\ell_{p'}\)-minimization with a smaller \(p\), \(\ell_1\)-minimization can indeed outperform \(\ell_{p'}\)-minimization for all \(p \in [0, 1]\) in weak recovery if \(\alpha\) is close to 1 and \(n\) is large enough.

It might be counterintuitive at first sight to see that the weak threshold of \(\ell_0\)-minimization is less than that of \(\ell_1\)-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 \(\alpha\) is close enough to 1, for a random Gaussian measurement matrix \(A\), \(\ell_1\)-minimization would recover all the nonnegative vectors on support \(T\) with overwhelming probability, while \(\ell_0\)-minimization would fail to recover some nonnegative vector on \(T\) with overwhelming probability. The failure of \(\ell_0\)-minimization indicates that there exists a nonnegative vector \(x\) on support \(T\) and a vector \(x'\) on support \(T'\) such that \(|T'| \leq |T|\), and \(Ax = Ax'\). Note that \(x'\) could have negative entries, or \(T'\) may not be a subset of \(T\). Therefore, if \(x\) is the sparse vector we would like to recover from \(Ax\), \(\ell_0\)-minimization would fail since \(\|x'\|_0 \leq \|x\|_0\).

However, \(\|x\|_1 < \|x'\|_1\) should hold since \(\ell_1\)-minimization can successfully return \(x\) as its solution. Of course when \(x'\) is the sparse vector we would like to recover, \(\ell_1\)-minimization would return \(x\) and fail to recover \(x'\). However, since \(\ell_1\)-minimization would recover all the nonnegative vectors on \(T\), then either \(T' \not\subset T\) holds or \(x'\) has negative entries. Therefore when we consider recovering nonnegative vectors on \(T\) for the weak recovery, \(x'\) is not taken into account, and \(\ell_1\)-minimization works better than \(\ell_0\)-minimization. Thus, although the performance of \(\ell_1\)-minimization is not as good as that of \(\ell_{p'}\)-minimization \((p \in [0, 1])\) in the strong recovery which requires to recover all the vectors up to certain sparsity, \(\ell_1\)-minimization can recover all the \(\rho n\)-sparse vectors on some support with some sign pattern, while for \(\ell_{p'}\)-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, \(\ell_{p'}\)-minimization is better for smaller \(p\), however, when we aim to recover all the vectors on one support with one sign pattern, \(\ell_1\)-minimization may have a better performance.

**IV. Recovery Bounds for Fixed \(\frac{a}{n}\)**

We considered the limiting case that \(\alpha \to 1\) in Section III and provided the limiting thresholds of sparsity ratio for successful recovery via \(\ell_{p'}\)-minimization both in the strong sense and in the weak sense. Here we focus on the case that \(\alpha\) is fixed \((0 < \alpha < 1)\). For any \(\alpha\) 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 \(\ell_{p'}\)-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 \(\ell_{p'}\)-minimization would fail to recover some \(\rho n\)-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 \(\rho n\)-sparse vectors via \(\ell_{p'}\)-minimization, \(\|B_Tz\|_{p'} < \frac{1}{2}\|Bz\|_{p'}\) should hold for every non-zero vector \(z \in \mathbb{R}^{n-m}\), and every set \(T \subset \{1, \ldots, n\}\) 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 \(\|Bz\|_{p'}^\alpha\) for all \(z \in S\), where \(S\) is the unit sphere in \(\mathbb{R}^{n-m}\). Then for any \(\rho\), we calculate an upper bound of \(\|B_Tz\|_{p'}\) for all \(T\) with \(|T| = \rho n\) and all \(z \in S\). Then we define \(\rho^*(\alpha, p)\) to be the largest \(\rho\) such that the aforementioned upper bound is less than half of the lower bound. According to Theorem 1, \(\ell_{p'}\)-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. \(N(0, 1)\) entries, and therefore for any \(z \in S\) 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 $S$ 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 $S$ 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 $\rho$ such that the "upper bound" is less than half of the "lower bound". We then show in Theorem 7 that $\ell_p$-minimization can recover all the $\rho^*(\alpha, p)n$-sparse vectors with overwhelming probability.

**Lemma 5.** For any $\alpha$ 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 $\mathbf{z} \in S$, $\|B\mathbf{z}\|_p^p > \lambda_{\min}(\alpha, p)n$.

**Lemma 6.** Given any $\alpha$, $p$ 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 $\mathbf{z} \in S$ and for every set $T \subset \{1, 2, ..., n\}$ with $|T| \leq \rho^*(\alpha, p)n$, $\|B_T\mathbf{z}\|_p^p < \frac{1}{2} \lambda_{\min}(\alpha, p)n$.

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

**Theorem 7.** For any $0 < p \leq 1$, any $0 < \alpha < 1$, for matrix $A^{m \times n}$ ($\alpha = \frac{m}{n}$) with i.i.d. $N(0, 1)$ 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 $\ell_p$-minimization problem (3) for every vector $\mathbf{x}$ up to $\rho^*(\alpha, p)n$-sparse.

Theorem 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 $\ell_p$-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 $\rho$. 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 $\|B\mathbf{z}\|_p^p \leq bn$ holds for some $\mathbf{z} \in S$ by techniques like $\gamma$-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 $\mathbf{z} \in S$ and some support $T$ with $|T| = \rho n$ such that $\|B_T\mathbf{z}\|_p^p \geq \lambda_{\min}(\alpha, p)n/2$. Then $\rho^*(\alpha, p)$ is chosen to be the largest $\rho$ 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 $\alpha$ for different $p$, and Fig. 3 shows the curve of $\rho^*(\alpha, p)$ against $p$ for different $\alpha$. Note that for any $p$, $\lim_{\alpha \rightarrow 1} \rho^*(\alpha, p)$ 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 $\rho^*(0.5)$ we obtained here is approximately 0.268 when $\alpha$ goes to 1. This is because in Section III-A we employed a finer technique to characterize the sum of the largest $\rho n$ 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 $\alpha$ 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 $\alpha \geq 0.8$ and increases to 0.1308 as $\alpha \rightarrow 1$. 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 $\alpha \geq 0.65$ and increases to 0.268 as $\alpha \rightarrow 1$. Therefore, although [4] provides a better bound than ours when $\alpha$ is small, our bound $\rho^*$ improves over that in [4] when $\alpha$ is relatively large.

[13] provides a lower bound of strong recovery threshold for every $\alpha$ and very $p$. For example, it shows that when $n$ is large enough, $\ell_0$-minimization can recover all the $\frac{m}{n}$-sparse vectors for given $\alpha$. Their result is better than ours when $\alpha$ is small. However, our bound is higher than that in [13] when $\alpha$ 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 $\ell_0$-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 $\ell_1$-minimization (Fig.1.1). Since 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 $\alpha$ is relatively large.

**B. Weak Recovery**

Theorem 3 provides a sufficient condition for successful recovery of every nonnegative $pn$-sparse vector $\mathbf{x}$ on one support $T$, which requires $\|B_T\mathbf{z}\|_p^p \leq \|B_T\mathbf{z}\|_2^2$ to hold for all non-zero $\mathbf{z} \in \mathbb{R}^{n-m}$, where given $\mathbf{z}$, $T^- = \{i : B_i\mathbf{z} < 0\}$. 
vectors on a fixed support with a fixed sign pattern. A lower bound of recovery threshold of exponentially to zero, and likewise establish an "upper bound" that the violation probability of this "lower bound" decays its proof.

Lemma 7. \( \rho c = \alpha, p \) and \( z = \rho \bar{\lambda}_{\text{max}}(\alpha, p, \rho) \), for all \( \alpha, p, \rho \) and set \( z = \rho \bar{\lambda}_{\text{max}}(\alpha, p, \rho) \) for all \( \alpha, p \), and \( \rho \) as small as possible while at the same time the probability that \( ||B_T^{-}\bar{z}||^p = \rho \bar{\lambda}_{\text{max}}(\alpha, p, \rho) n \) for some \( T \) with \( |T| = \rho n \) and some \( z \in S \) still has exponential decay to zero. Efforts are made in Appendix-N to improve \( \bar{\lambda}_{\text{max}}(\alpha, p, \rho) \), 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 \( \ell_p \)-minimization in the weak sense for given \( \alpha \).

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 \( 1 - e^{-c_{15}n} \), \( x \) is the unique solution to the \( \ell_p \)-minimization problem (3) for every nonnegative \( \rho_w(\alpha, p) \)-sparse vector \( x \) 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 \( p \) we first calculate \( \bar{\lambda}_{\text{min}}(\alpha, p, \rho) \) in Lemma 7 from (53) to obtain a "lower bound" of \( ||B_T^{-}\bar{z}||^p \) for all \( z \in S \) and calculate \( \bar{\lambda}_{\text{max}}(\alpha, p, \rho) \) in Lemma 8 from (70) to obtain an "upper bound" of \( ||B_T^{-}\bar{z}||^p \) for all \( z \in S \). We then find the largest \( \rho_w(\alpha, p) \) such that the "lower bound" of \( ||B_T^{-}\bar{z}||^p \) is larger than the "upper bound" of \( ||B_T^{-}\bar{z}||^p \), 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 \( \alpha \) for different \( p \) and Fig. 5 shows the curve of \( \rho_w(\alpha, p) \) against \( \alpha \) for different \( p \). When \( \alpha \to 1 \) and \( \rho_w(\alpha, p) \) 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 \( \ell_1 \)-minimization is greater than 2/3 for all \( \alpha \) that is greater than 0.9, since the weak recovery threshold of \( \ell_p \)-minimization \( (p \in [0, 1]) \) when \( \alpha \to 1 \) is all 2/3, therefore for all \( \alpha > 0.9 \), the weak recovery threshold of \( \ell_1 \)-minimization is greater than that of \( \ell_p \)-minimization for all \( p \in (0, 1) \).

V. \( \ell_1 \)-MINIMIZATION CAN PERFORM BETTER THAN \( \ell_p \)-MINIMIZATION \( (p \in [0, 1]) \) FOR SPARSE RECOVERY

For strong recovery, if \( \ell_1 \)-minimization can recover all the \( k \)-sparse vectors, then \( \ell_p \)-minimization is also guaranteed to recover all the \( k \)-sparse vectors for all \( p \in (0, 1) \). However, for weak recovery, the performance of \( \ell_1 \)-minimization is better than that of \( \ell_p \)-minimization for all \( p \in (0, 1) \) in at least the large \( \alpha \) region \( (\alpha > 0.9) \), and the same result holds for all choices of supports and sign patterns. Then one may naturally ask why \( \ell_1 \)-minimization outperforms \( \ell_p \)-minimization \( (p < 1) \) in recovering vectors on every specific support with every specific sign pattern, but is not as good as \( \ell_p \)-minimization for all \( z \in S \) in Lemma 8. One should be cautious that the set \( T^- \) varies for different \( z \).

Lemma 8. Given \( \alpha, p \) and set \( T \subset \{1, \ldots, n\} \) with \( |T| = \rho n \), for every \( z \in S \), \( ||B_T^{-}\bar{z}||^p < \rho \bar{\lambda}_{\text{max}}(\alpha, p, \rho) n \), for some \( \bar{\lambda}_{\text{max}}(\alpha, p, \rho) > 0 \).
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 $\ell_1$-minimization can recover all the vectors up to $\rho_1^{n}$-sparse and $\ell_p$-minimization with some $p \in [0, 1)$ can recover all the vectors up to $\rho_p^{n}$-sparse, and we know $\rho_1^{n} < \rho_p^{n}$ from our discussion on strong bound. Note that since the limiting threshold of strong recovery via $\ell_p$-minimization increases to 0.5 as $p$ decreases to 0, then we have $\rho_1^{n} < \rho_p^{n} \leq 0.5$. However, if we only consider the ability to recover the vectors on one support with one sign pattern, with overwhelming probability $\ell_1$-minimization can recover vectors up to $\rho_1^{n}$-sparse, while $\ell_p$-minimization can recover vectors up to $\rho_p^{n}$-sparse. From previous discussion about weak recovery threshold, we know that when $\alpha$ is very close to 1, $\rho_1^{n} > \frac{\alpha}{2} > \rho_p^{n} > \frac{\alpha}{2}$. And this result holds for any specific choice of the support and the sign pattern. Therefore we have $\rho_1^{n} > \rho_p^{n} > \rho_s^{n} > \rho_1^{n}$. We illustrate the difference of $\ell_1$ and $\ell_p$-minimization in Fig. 6 and Fig. 7. Let $\Omega$ 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 $\rho \in (\rho_1^{n}, \rho_p^{n})$ in Fig. 6. Since $\rho < \rho_1^{n}$, for any fixed support $T_i$ with $|T_i| = mn$ and any fixed sign pattern $\sigma_j$, with high probability $\ell_1$-minimization can recover all the $\rho n$-sparse vectors on $T_i$ with sign pattern $\sigma_j$. Let $E_{T_i}^{\sigma_j}$ denote the event that $\ell_1$-minimization can recover all the $\rho n$-sparse vectors on support $T_i$ with sign pattern $\sigma_j$. There are $\binom{n}{mn}$ different supports, and for each support, there are $2^{mn}$ different sign patterns. Then $P(E_{T_i}^{\sigma_j})$ is very close to 1 for every $T_i$ and $\sigma_j$ as shown in Fig. 6(a). Since we also have $\rho > \rho_1^{n}$, then with high probability strong recovery of $\ell_1$-minimization fails, in other words, $\ell_1$-minimization would fail to recover at least one vector with at most $\rho n$ non-zero entries. Let $E$ denote the event that $\ell_1$-minimization can recover all the $\rho n$-sparse vectors, then we have

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

Then although $P(E_{T_i}^{\sigma_j})$ is the same for all $T_i$ and $\sigma_j$ and is very close to 1, $P(E)$ is close to 0, as indicated in Fig. 6(a). For $\ell_p$-minimization, since $\rho < \rho_1^{n}$, then with high probability, $\ell_p$-minimization can recover all the $\rho n$-sparse vectors. In Fig. 6(b), $\tilde{E}$ denotes the event that $\ell_p$-minimization can recover all the $\rho n$-sparse vectors, then

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

where $\tilde{E}_{T_i}^{\sigma_j}$ denotes the event that $\ell_p$-minimization recovers all the vectors on support $T_i$ with sign pattern $\sigma_j$. In this case, $P(\tilde{E})$ is close to 1 as indicated in Fig. 6(b). In Fig. 7, we pick $\rho \in (\rho_1^{n}, \rho_p^{n})$. Then given any support $T_i$ and any sign pattern $\sigma_j$, $\ell_1$-minimization can recover the vectors on $T_i$ with sign pattern $\sigma_j$ with high probability, while $\ell_p$-minimization fails to recover at least one vector on $T_i$ with sign pattern $\sigma_j$ with high probability. Therefore $P(\tilde{E}_{T_i}^{\sigma_j})$ is close to 1, while $P(\tilde{E}_{T_i}^{\sigma_j})$ is close to 0 for any given $T_i$ and $\sigma_j$. Therefore, if the sparse vectors we would like to recover are on one same support and share the same sign pattern, $\ell_1$-minimization can be a better choice than $\ell_p$-minimization for all $p \in [0, 1)$ regardless of the amplitudes of the entries of a vector.

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:
Experiment 1. Recovery thresholds when measurement the recovery threshold easily. In this case, we can in fact compute special case that the null space of the measurement matrix $\ell$-recovery changes from strong to weak. Performance changes when the requirement for successful and is 1 for $p$. We can see how recovery threshold strictly decreases as $p$ increases. However, the weak recovery threshold of $\ell_1$-minimization is close to 0.9, which is greater than the weak recovery threshold of $\ell_p$-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. 

Theorem 9. $\ell_p$-minimization problem ($p \in [0,1]$) can recover all the $\rho n$-sparse vectors $x$ on some support $T$ if and only if
\[
\|B_Tz\|_p^p < \|B_Tz\|_p^p
\]
for all non-zero $z \in \mathbb{R}^{n-m}$.

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 $\mathbb{R}^4$ and $T = \{1,2\}$. If $B = [16,16,1,36]$, then one can check that $\|B_T\|_{1} = 32 < 37 = \|B_T\|_{1}$, but $\|B_T\|_{0.5} = 8 > 7 = \|B_T\|_{0.5}$. If $B = [1,4,1,9]$, then $\|B_T\|_{1} < \|B_T\|_{1}$, and $\|B_T\|_{0.5} < \|B_T\|_{0.5}$. 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 $\alpha \rightarrow 1$ 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 $\alpha \rightarrow 1$, $\ell_p$-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 $\ell_p$-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 $\beta$ denote the basis of the null space of $A$. Then $\lambda \beta$ is in the null space of $A$ for every $\lambda \in \mathbb{R}$, and every vector in the null space of $A$ can be represented as $\lambda \beta$ for some $\lambda \in \mathbb{R}$. Thus, the strong recovery threshold and the weak recovery threshold of $\ell_1$-minimization and $\ell_p$-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 $\beta$ and $-\beta$. From Theorem 1, the strong recovery threshold of $\ell_p$-minimization ($p \in (0,1]$) is the integer $k$ such that the sum of the largest $k$ terms of $|\beta_i|^p$ ($i \in \{1,\ldots,n\}$) is less than $\|\beta\|_p^p/2$ and the sum of the largest $k+1$ terms of $|\beta_i|^p$ ($i \in \{1,\ldots,n\}$) is greater than or equal to $\|\beta\|_p^p/2$. For weak recovery, we consider recovering all the nonnegative $k$-sparse vectors on support $T = \{1,\ldots,k\}$. From Theorem 2, the weak recovery threshold of $\ell_1$-minimization is the largest integer $k$ such that both $\|\beta_{T^-}\|_1 < \|\beta_{T^-}\|_1 + \|\beta_{T^+}\|_1$ and $\|\beta_{T^-}\|_1 + \|\beta_{T^+}\|_1 < \|\beta_{T^-}\|_1 + \|\beta_{T^+}\|_1$ hold. From Theorem 3, the weak recovery threshold of $\ell_p$-minimization is the largest integer $k$ such that both $\|\beta_{T^-}\|_1 < \|\beta_{T^-}\|_1$ and $\|\beta_{T^+}\|_1 < \|\beta_{T^+}\|_1$ hold.

We generate one hundred random Gaussian matrices $A_{499 \times 500}$, and for each random matrix $A$, we compute its corresponding strong (and weak) recovery threshold of $\ell_1$ and $\ell_p$-minimization. For each $p$ between 0 and 1, we count the percentage of random matrices with which $\ell_1$ and $\ell_p$-minimization can recover all the $\rho n$-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 $\ell_1$-minimization is close to 0.9, which is greater than the weak recovery threshold of $\ell_p$-minimization for every $p < 1$.
standard Gaussian distribution. We then generate an $m \times n$ matrix $A$ with i.i.d. $\mathcal{N}(0, 1)$ entries. We let $y = Ax$ 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 $x^*$ be the output of the algorithm, if $\|x^* - x\|_2 \leq 10^{-4}$, we say the recovery of $x$ is the successful. Fig. 10 records the percentage of times that the recovery is successful for different sparsity $\rho n$. Note that the iteratively reweighted least squares algorithm is designed to obtain a local minimum of the $\ell_p$-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 $\ell_p$-minimization and $\ell_1$-minimization both for strong recovery in Fig. 11 and for weak recovery in Fig. 12 when $\alpha$ is large. We employ the trial version of MOSEK [30] to solve $\ell_1$-minimization and still employ the iteratively reweighted least squares algorithm to compute a local minimum of $\ell_p$-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 $\rho$ from 0.2 to 1. In weak recovery, we consider recovering nonnegative vectors on a fixed support $T = \{1, ..., \rho n\}$. For a given $\rho$, we generate one thousand vectors and claim the weak recovery of $\rho n$-sparse vectors to be successful if and only if all the vectors are successfully recovered. For each vector $x$, $x_i = |z_i|$ ($i \in T$), 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 $\ell_1$-minimization is the same for every nonnegative vector on $T$, therefore for a fixed matrix $A$, if $\ell_1$-minimization recovers all the vectors we generated, it should also recover all the nonnegative vectors on $T$. $\ell_p$-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 $x$ on $T$, $\ell_p$-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 $\rho$, 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 $\rho n$-sparse vector $x$, 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 $\ell_p$-minimization, therefore the solution returned by the algorithm may not always be the solution of $\ell_p$-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 $\ell_1$-minimization is higher than any other $\ell_p$-minimization for $p < 1$.

VII. CONCLUSION

This paper analyzes the ability of $\ell_p$-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 $\ell_p$-minimization is $2/3$ for all $p < 1$, compared with the threshold $1$ for $\ell_1$-minimization. For any $\alpha < 1$, we provide a bound $\rho^*(\alpha, p)$ of sparsity ratio below which strong recovery via $\ell_p$-minimization succeeds with overwhelming probability, and our bound $\rho^*(\alpha, p)$ improves on the existing bounds in the large $\alpha$ region. We also provide a bound $\rho^*_\alpha(\alpha, p)$ of sparsity ratio below which weak recovery succeeds with overwhelming probability.

Throughout the paper, we assume that the measurements $y = Ax$ are exact, and it would be interesting to consider the case that the measurements are noisy, i.e. $y = Ax + e$ where $e$ is the vector of noise. Moreover, we assume that $x$ 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] I. Blanchard, C. Curtis, and J. Tanner, “The restricted isometry property and $\ell^2$ 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] R. Chartrand and W. Yin, http://www.caam.rice.edu/~wy1/paperfiles/TR08-01/IRLS_CS.rar.
[11] ——, “Iteratively reweighted algorithms for compressive sensing,” in Proc. ICASSP, 2007.
[12] R. Chartrand and V. Staneva, “Restricted isometry properties and non-convex compressive sensing,” Inverse Problems, vol. 24, pp. 35020–35033(14), 2008.
[13] 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.
[14] 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.
[15] 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.
[16] D. Donoho, “High-dimensional centrally symmetric polytopes with neighborhood proportional to dimension,” Discrete Comput. Geom., 2006.
[17] 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.
[18] ——, “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.
[19] D. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, April 2006.
[20] D. Donoho and X. Huo, “Uncertainty principles and ideal atomic decomposition,” IEEE Trans. Inf. Theory, vol. 47, no. 7, pp. 2845 – 2862, Nov. 2001.
[21] C. Dwork, F. McSherry, and K. Talwar, “The price of privacy and the limits of lp decoding,” in Proc. STOC, 2007, pp. 85–94.
[22] 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.
[23] 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.
[24] J.-F.uchs, “On sparse representations in arbitrary redundant bases,” IEEE Trans. Inf. Theory, vol. 50, no. 6, pp. 1341 – 1344, Jun. 2004.
[25] R. Gribonval and M. Nielsen, “Sparse representations in unions of bases,” IEEE Trans. Inf. Theory, vol. 49, no. 12, pp. 3320 – 3325, Dec. 2003.
[26] ——, “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.
[27] J. Haupt and R. Nowak, “Signal reconstruction from noisy random projections,” IEEE Trans. Inf. Theory, vol. 52, no. 9, pp. 4036 – 4048, Sep. 2006.
[28] M. Ledoux, Ed., The Concentration of Measure Phenomenon. American Mathematical Society, 2001.
[29] Mosek Optimization Software, http://www.mosek.com/.
[30] R. Saab, R. Chartrand, and O. Yilmaz, “Stable sparse approximations via nonconvex optimization,” in Proc. ICASSP, 2008.
[31] M. Stojnic, “A simple performance analysis of $l_1$ optimization in compressed sensing,” in Proc. IEEE ICASSP 2009, Apr. 2009, pp. 3021 – 3024.
[32] M. Stojnic, W. Xu, and B. Hassibi, “Compressed sensing - probabilistic analysis of a null-space characterization,” in Proc. ICASSP, 2008, pp. 3377–3380.
A. Proof of Theorem 3

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

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

\[
\sum_i \|x + B_{T^-}z\|_p^p = \sum_i \|x + B_{T^-}z\|_p^p + \sum_i \|x + B_{T^+}z\|_p^p = 0 + \sum_i \|x + B_{T^-}z\|_p^p + \sum_i \|x + B_{T^+}z\|_p^p = 0 + \sum_i \|x + B_{T^-}z\|_p^p + \sum_i \|x + B_{T^+}z\|_p^p = \|x + B_{T^-}z\|_p^p + \|x + B_{T^+}z\|_p^p \leq \|x\|_p^p.
\]

Thus \( \|x + B_{T^-}z\|_p^p < \|x\|_p^p \) is not a solution to (3), which is a contradiction.

Secondly, consider the case \( T^+ \) is not empty. Then \( \|B_{T^-}-z\|_p^p > \|B_{T^+}-z\|_p^p \) since we assume the condition in Theorem 3 fails for \( z \). Define a vector \( x \) as follows. Let \( x_i = 0 \) for every \( i \in T^+ \), let \( x_i = -B_{T^-}z \) 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 \( x_i \geq 0 \) large enough such that \( \|x + B_{T^-}z\|_p^p - \|x + B_{T^+}z\|_p^p < \frac{\delta}{2} \). Then

\[
\|x + B_{T^-}z\|_p^p = 0 + \sum_i \|x + B_{T^-}z\|_p^p = \sum_i \|x + B_{T^-}z\|_p^p = \sum_i \|x + B_{T^-}z\|_p^p - \delta \leq \|x\|_p^p.
\]

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

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

\[
\|x + B_{T^-}z\|_p^p = \sum_i \|x + B_{T^-}z\|_p^p + \sum_i \|x + B_{T^+}z\|_p^p = \sum_i \|x + B_{T^-}z\|_p^p + \sum_i \|x + B_{T^+}z\|_p^p \geq \sum_i \|x + B_{T^-}z\|_p^p + \sum_i \|x + B_{T^+}z\|_p^p = \sum_i \|x + B_{T^-}z\|_p^p + \sum_i \|x + B_{T^+}z\|_p^p.
\]

where the inequality follows from the triangular property that \( \|x_i + B_{T^-}z\|_p^p \geq \|x_i\|_p^p - \|B_{T^-}z\|_p^p \) holds for all \( i \) and all \( p \in (0, 1) \).

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

B. Proof of Lemma 1

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

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

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

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

\[
\int_{z^*}^\infty z^p f(z) dz = \int_0^{z^*} z^p f(z) dz = 0.
\]

Define

\[
\rho^* = 1 - F(z^*).
\]

We claim \( \rho^* \) has the desired property. Let

\[
T_{z^*} = \sum_{i : |x_i| > z^*} Y_i = \sum_{i = 1}^n Y_i 1_{\{X_i \geq z^*\}},
\]

where \( 1 \) is the indicator function. Then

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

Let \( h \) be the smallest integer such that \( Y_h \geq z^{*p} \) and \( Y_{h+1} < z^{*p} \), then \( T_h = \sum_{i=1}^h Y_i \). We also have that \( h = \sum_{i=1}^n 1_{\{|X_i|^p \geq z^{*p}\}} = \sum_{i=1}^n 1_{\{|X_i| \geq z^*\}} \). 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

\[
|T_{z^*} - S_{\rho^*}| \leq \frac{h - \rho^* n}{\rho^* n} S_{\rho^*}.
\]

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

and (15) follows. Combining three cases, we conclude that (15) always holds. Then

\[ E[|T_{z^*} - S_{p^*}|] \leq E[|h - \lceil \rho n \rceil| S_{p^*} / \rho] \leq \sqrt{E[(h - \lceil \rho n \rceil)^2] E[S_{p^*}^2] / \rho}, \]

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

\[ E[(h - \lceil \rho n \rceil)^2] = E[(h - \rho n)^2 + 2(\rho n - \lceil \rho n \rceil)E[h - \rho n] + (\rho n - \lceil \rho n \rceil)^2] \leq n\rho(1 - \rho^*) + 1. \]

Besides,

\[ E[S_{p^*}^2] \leq E[S_{p^*}^2] = E(\sum_{i=1}^{n} |X_i|^p)^2 \leq E(\sum_{i=1}^{n} |X_i|^{2p} + \sum_{i \neq j} |X_i|^p |X_j|^p) = nE[X|X|^p] + n(n)E[X|^p]_2, \]

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

\[ E[|T_{z^*} - S_{p^*}|] \leq \sqrt{(n\rho^*(1 - \rho^*) + 1)(nE[X|X|^p] + n(n - 1)E[X|^p]_2)^2 / \rho}, \]

or \( O(\sqrt{n}) \).

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

\[ \lim_{n \to \infty} \frac{E[S_{p^*}]}{S} = \lim_{n \to \infty} \frac{E[T_{z^*}]}{S} + \lim_{n \to \infty} \frac{E[S_{p^*} - T_{z^*}]}{S} = \frac{f(z^*)}{g(0)} + 0 = \frac{1}{2}. \]

C. Proof of Proposition 1

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

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

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

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

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

\[ \frac{dp^*}{dz^*} = \frac{dp^*}{dp} \left(\frac{dz^*}{dp}\right), \]

thus

\[ \frac{dp^*}{dp} = \frac{f(z^*)}{\int_0^{z^*} x^p f(x) dx - \int_{z^*}^{\infty} x^p f(x) dx}{2z^*}. \]

Note that

\[ \int_0^{z^*} x^p \ln x f(x) dx - \int_0^{\infty} x^p \ln x f(x) dx = \int_{z^*}^{\infty} x^p \ln x f(x) dx > 0 \]

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

D. Proof of Lemma 2

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

\[ |S_{p}(X) - S_{p}(X')| \leq \sum |X_i^p - X_i^p|p. \]

Since \( \|X_i^p - X_i^p| \leq |X_i - X_i|^p \) for all \( p \in (0, 1) \),

\[ |S_{p}(X) - S_{p}(X')| \leq \sum |X_i - X_i|^p. \]

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

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

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

\[ \sum_i |x_i - x_i'|^p \leq \left( \sum_i |u_i|^{p/(2-p)} \right)^{1-p/2} \left( \sum_i |v_i|^{p/2} \right)^{p/2} \leq \left( n^{1-p}/(2-p) \right)^{1-p/2} \left( n^{1-p}/2 \right)^{p/2} \]

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

\[ P(S_{p}(x) \leq M_{p} + n^2/(p/2) \) \( t^2 p \) \geq P(d(x, A) \leq t) \geq 1 - e^{-t^2/2}. \]

Similarly,

\[ P(S_{p}(x) \geq M_{p} - n^2/(p/2) \) \( t^2 p \rangle \geq 1 - e^{-t^2/2}. \]

Combining (22) and (23),

\[ P(|S_{p}(x) - M_{p}| \geq n^2/(p/2) \) \( t^2 p \leq 2e^{-t^2/2}. \]
The difference of $E[S_\rho]$ and $M_\rho$ can be bounded as follows,
\[
|E[S_\rho] - M_\rho| \leq E[|S_\rho - M_\rho|] = \int_0^\infty P(|S_\rho(x) - M_\rho| \geq y)dy \\
\leq \int_0^\infty 2e^{-\frac{1}{4}y^2\frac{1}{n}(\frac{1}{2})}dy \\
= n^{(\frac{1}{2} - \frac{1}{2})}\int_0^\infty 2e^{-\frac{1}{4}y^2\frac{1}{n}}ds
\]
Note that $c := \int_0^\infty 2e^{-\frac{1}{4}y^{(s/\rho)}}ds$ is a finite constant for all $p \in (0, 1]$. As $p > 0$ and $S = nE[|x_i|p]$, thus for any $\delta > 0$, $cn(\frac{1}{2} - \frac{1}{2}) < \frac{1}{2}$ when $n$ is large enough.

Let $t = (\frac{1}{2} \delta sn^{\frac{1}{2} - 1})^\frac{1}{2} = (\frac{1}{2} \delta E[|x_i|^p])^\frac{1}{2},$ from (24) with probability at least $1 - 2e^{-\frac{1}{2}(\frac{1}{2} \delta E[|x_i|^p])}\frac{1}{n}$, $|S_\rho - M_\rho| < \frac{1}{2}\delta S$. Thus $|S_\rho - E[S_\rho]| \leq |S_\rho - M_\rho| + |M_\rho - E[S_\rho]| < \delta S$ with probability at least $1 - 2e^{-\epsilon c_n^2}$ 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
\[
E[S_{\rho^*}] \leq \frac{1}{2} + \epsilon)S
\]
for all $n \geq M$ where $S = E[S_i]$. Then $E[\sum_{l=\rho n^i} \gamma_{\rho n}] = S - E[S_{\rho^*}] \geq (\frac{1}{2} - \epsilon)S$. Since $E[\sum_{l=\rho n^i} \gamma_{\rho n}]$ is a summation of $n - \rho n^i$ terms, and $E[Y_{\rho n^i}] \geq E[Y_i]$ for all $i \geq \rho n$, then we have
\[
E[Y_{\rho n^i}] \geq \frac{E[\sum_{l=\rho n^i} \gamma_{\rho n}]}{n - \rho n^i} \geq \frac{\left(\frac{1}{2} - \epsilon\right)S}{n - \rho n^i} \geq \frac{\left(\frac{1}{2} - \epsilon\right)S}{n}.
\]
Then for any $\rho < \rho^*$, for every $\epsilon > 0$, when $n$ is large enough,
\[
E[S_\rho] = E[S_\rho^*] - \sum_{i = \rho n^i+1}^{\gamma_{\rho n}} E[Y_i] \\
\leq E[S_\rho^*] - \left[(\rho n^i - \rho n)\right] E[Y_{\rho n^i}] \\
\leq \frac{1}{2} + \epsilon)S - (\rho n^i - \rho n)\left(\frac{1}{2} - \epsilon\right)S/n \\
\leq \frac{1}{2} + \epsilon)S - (\rho n^i - \rho n)\left(\frac{1}{2} - \epsilon\right)S/n
\]
where the first inequality holds since each $Y_i$ with $i \leq \rho n^i$ has expectation at least as large as $E[Y_{\rho n^i}]$, and the second inequality follows from (25) and (26). Then for any $\rho < \rho^*$, we can pick $\epsilon > 0$ small enough such that $E[S_\rho^*]/S \leq (\frac{1}{2} + \epsilon) - (\rho n^i - \rho n)\left(\frac{1}{2} - \epsilon\right) \leq \frac{1}{2} - 2\delta$ 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 $\gamma$-net $\Sigma$ in $\mathcal{R}^{n-m}$ of cardinality less than $(1 + \frac{2}{\gamma})^{n-m}$ (29). A $\gamma$-net $\Sigma$ is a set of points in $\mathcal{R}^{n-m}$ such that $\|v^k\|_2 = 1$ for all $v^k$ in $\Sigma$ and for any $z \in \mathcal{R}^{n-m}$ with $\|z\|_2 = 1$, there exists some $v^k$ such that $\|z - v^k\|_2 \leq \gamma$.

Since $B$ has i.i.d. $N(0, 1)$ entries, then $Bu_k$ has $n$ i.i.d. $N(0, 1)$ entries for every $u_k$. 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 $1 - 2e^{-c_2 n^2} = 2e^{-c_3 n}$, we have
\[
S_{\rho}(Bu_k) \leq \left(\frac{1}{2} - \delta\right)S
\]
(27)
and
\[
(1 - \epsilon)S \leq S_{\rho}(Bu_k) \leq (1 + \epsilon)S
\]
(28)
both hold for one vector $u_k$ in $\Sigma$. Then applying union bound, we know that (27) and (28) hold for all vectors in $\Sigma$ with probability at least
\[
1 - (1 + 2/\gamma)^{n-m}(2e^{-c_2 n^2} + 2e^{-c_3 n}).
\]
(29)
Let $\alpha = m/n$, then as long as $\alpha$ is large enough, say greater than $c_4 := 1 - \min(c_2, c_3)$, then (29) is greater than $1 - e^{-c_5 n}$ for some constant $c_5 > 0$.

For any $z$ such that $\|z\|_2 = 1$, there exists $v_0$ in $\Sigma$ such that $\|z - v_0\|_2 \leq c_1 \leq \gamma$. Let $z_1$ denote $z - v_0$, then $\|z_1 - v_1\|_2 \leq \gamma_2 \leq \gamma_1 \leq \gamma^2$ for some $v_1$ in $\Sigma$. Repeating this process, we have
\[
z = \sum_{j=0}^{\gamma_j} v_j
\]
(30)
where $\gamma_0 = 1, \gamma_1 \leq \gamma^j$ and $v_j \in \Sigma$. Thus for any $z \in \mathcal{R}^{n-m}$, we have $z = \|z\|_2 \sum_{j=0}^{\gamma_j} v_j$.

For any index set $\mathcal{T}$ with $|\mathcal{T}| \leq |\rho n|$, we have
\[
\|Bz\|_p^p = \|z\|_p^p \sum_{j=0}^{\gamma_j} Bv_j \|_p^p \\
\leq \|z\|_p^p \sum_{j=0}^{\gamma_j} \|Bv_j\|_p^p \\
\leq S\|z\|_p^p \left(\frac{1}{2} - \frac{\delta}{1 - \gamma^p}\right)
\]
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).

\[
\|Bz\|_p^p = \|z\|_p^p \sum_{j=0}^{\gamma_j} Bv_j \|_p^p \\
\geq \|z\|_p^p (\|Bv_0\|_p^p - \sum_{j=1}^{\gamma_j} \|Bv_j\|_p^p) \\
\geq \|z\|_p^p ((1 - \epsilon)S - \sum_{j=1}^{\gamma_j} \|Bv_j\|_p^p) \\
\geq S\|z\|_p^p \left(\frac{1}{2} - \frac{\gamma^p}{1 - \gamma^p}\right)
\]
where the first inequality holds from the triangular inequality and the third inequality holds with overwhelming probability by (28) and (29). Thus $\|B_Tz\|_p^p \leq \|B_Tz\|_p^p \leq S\|z\|_p^p \left(\frac{2\delta - 2\gamma^p}{1 - \gamma^p}\right)$ holds with probability at least $1 - e^{-c_6 n}$. For the given $\delta$ from Corollary 1, we can pick $\gamma$ and $\epsilon$ small enough such that $\|B_Tz\|_p^p \leq \|B_Tz\|_p^p \leq \delta S\|z\|_p^p$.
**G. Proof of Lemma 4**

**Proof:** We first consider the case that $p = 0$. Now $E[X_0^p] = 1$, where $X \sim \mathcal{N}(0,1)$. We have $\sum_{i \in T: X_i < 0} |X_i|^p = \sum_{i \in T} 1_{\{X_i < 0\}}$. Since $P(X_i < 0) = 0.5$ independently for all $i \in T$, then $E[\sum_{i \in T} 1_{\{X_i < 0\}}] = \rho n / 2$, and from the Chernoff bound, we have

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

and

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

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

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

$$|S_{T^-}(X) - S_{T^-}(X')| = \left| \sum_{i \in T} |X_i|^p 1_{\{X_i < 0\}} - \sum_{i \in T} |X_i'|^p 1_{\{X_i' < 0\}} \right| \leq \sum_{i \in T} \left| |X_i|^p 1_{\{X_i < 0\}} - |X_i'|^p 1_{\{X_i' < 0\}} \right| \leq \sum_{i \in T} |X_i - X_i'|^p, \quad (31)$$

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

From the isoperimetric inequality for the Gaussian measure [29], for any set $A \subset \mathbb{R}^n$ with measure at least a half, the set $A_1 = \{x \in \mathbb{R}^n : d(x, A) \leq t\}$ has measure at least $1 - e^{-t^2 / 2}$, where $d(x, A) = \inf_{y \in A} \|x - y\|_2$. Let $M_{T^-}$ be the median value of $S_{T^-} = S_{T^-}(X)$. Define set $A = \{x \in \mathbb{R}^n : S_{T^-}(x) \leq M_{T^-}\}$, then

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

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

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

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

$$P(S_{T^-}(x) \leq M_{T^-} + (\rho n)^{(1-p/2)}t^p) \geq P(d(x, A) \leq t) \geq 1 - e^{-t^2 / 2}. \quad (33)$$

Similarly,

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

Combining (33) and (34),

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

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

$$|E[S_{T^-}] - M_{T^-}| \leq E[|S_{T^-} - M_{T^-}|] = \int_0^\infty P(|S_{T^-}(x) - M_{T^-}| \geq y)dy \leq \int_0^\infty 2e^{-\frac{y^2}{2}} dy = (\rho n)^{(1-\frac{2}{p})} \int_0^\infty 2e^{-\frac{y^2}{2}} dy = (\rho n)^{(1-\frac{2}{p})}.$$}

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

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

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

holds with probability at least $1 - 2e^{-d_2 n}$.

Similarly we can prove that with probability at least $1 - 2e^{-d_3 n}$ for some $d_3 > 0$,

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

holds. Then by a simple union bound, the above two statements hold at the same time with probability at least $1 - 2e^{-d_2 n} - 2e^{-d_3 n}$, thus Lemma 4 follows. \hfill \blacksquare

**H. Proof of Theorem 6**

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

- $\frac{1}{2}\rho n (\mu - \epsilon) \leq \sum_{i \in T: B_i \in S} |B_i|^p \leq \frac{1}{2} \rho n (\mu + \epsilon)$
- $\frac{1}{2}(1 - \rho)n (\mu - \epsilon) \leq \sum_{i \in T: B_i \in S} |B_i|^p \leq (1 - \rho)n (\mu + \epsilon)$

hold for all the vectors $v$ in a $\gamma$-net $\Sigma$ at the same time. Let $S$ be the unit sphere in $\mathbb{R}^{n-m}$. Pick any $x \in S$, from (30) we have $z = \sum_{j \geq 0} \gamma_j v_j$, where $\gamma_0 = 1$, $v_j \in \Sigma$ for all $j$ and $\gamma_j \leq \gamma^3$. \hfill \blacksquare
Given $z$, let $T^- = \{i \in T : B_i z < 0\}$. For any $i$ in $T^-$,
\[ |B_i z|^p \leq \sum_{j \geq 0} |\gamma_j B_i v_j|^p \]
\[ = \sum_{j : B_i v_j < 0} |\gamma_j B_i v_j|^p + \sum_{j : B_i v_j \geq 0} |\gamma_j B_i v_j|^p \]
\[ \leq \sum_{j : B_i v_j < 0} |\gamma_j B_i v_j|^p \]
where the first inequality holds as $B_i z < 0$. Then
\[ |B_{T^-} z|^p \leq \sum_{i \in T^-} \sum_{j : B_i v_j < 0} \gamma_j |B_i v_j|^p \]
\[ = \sum_{j \geq 0} \sum_{i \in T^-} \gamma_j |B_i v_j|^p \]
\[ \leq \frac{1}{2(1 - \gamma p)} \rho n (\mu + \epsilon), \]
(37) where the last inequality holds with overwhelming probability.

We also have
\[ |B_{T^-} z|^p = \left( \sum_{j \geq 0} |\gamma_j B_{T^-} v_j|^p \right) \]
\[ \geq \left( 1 - \rho \right) n (\mu - \epsilon) - \sum_{j \geq 1} \gamma_j |1 - \rho| n (\mu + \epsilon) \]
\[ \geq \left( 1 - \rho \right) n \frac{2 \gamma \rho - \epsilon}{1 - \gamma p}, \]
(38) where the second inequality holds with overwhelming probability.

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

I.

Lemma 9. Given any $\alpha$ 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 $z \in S$, $|B z|^p < \lambda_{max}(\alpha, p) n$.

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 $|B z|^p$ exceeds $\lambda_{max}(\alpha, p) n$ for some $z$ in $S$ 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 $\xi_{max} = \frac{1}{n} \min_{z \in S} |B z|^p$, then for any non-zero vector $z$, $|B z|^p \leq \xi_{max} n$. Let $\Sigma_i$ be a $\gamma$-net of $S$ with cardinality at most $(1 + 2/\gamma)^{n-m}$ [29] and $\gamma > 0$ to be chosen later, and define
\[ \eta = \frac{1}{n} \max_{z \in \Sigma_i} |B z|^p. \]
Then from the definition of $\gamma$-net, for every $z \in S$, there exists $z' \in \Sigma_i$ such that $\|z - z'\|_2 \leq \gamma$. Note that for every $z \in S$, $|B z|^p \leq |B z'|^p + |B(z - z')|^p = |B z'|^p + |z - z'|^p$, $|B z|^p \leq \eta n + \gamma p c_{max} n$, where the first inequality follows from the triangle inequality and the second inequality follows from the definition of $\eta$ and $c_{max}$. Then $c_{max} n \leq \eta n + \gamma p c_{max} n$, which leads to
\[ c_{max} \leq \eta / (1 - \gamma p). \]
(39)

To characterize $c_{max}$, we first characterize $\eta$. For any $a > E|X|^p$ where $X \sim \mathcal{N}(0, 1)$, we calculate the probability that $|B z|^p \geq an$ for some $z$ in $\Sigma_i$. Note that for all $z \in S$, $B z$ ($i = 1, \ldots, n$) are i.i.d. $\mathcal{N}(0, 1)$ random variables where $B_i$ is the $i$th row of $B$. Then
\[ P(\eta \geq a) = P(\exists z \in \Sigma_i: \|B z\|^p \geq an) \]
\[ \leq \sum_{z \in \Sigma_i} P(\|B z\|^p \geq an) \]
\[ \leq (1 + 2/\gamma)^{n-m} \min_{t > 0} e^{-t \gamma \sum_{i \in S_i} |B_i|^p} \]
\[ = (1 + 2/\gamma)^{(1 - \alpha) n} \min_{t > 0} e^{-t \gamma} E(e^{t \gamma |X|^p})^n \]
\[ = e^{((1 - \alpha) n \log(1 + 2/\gamma) + \min_{t > 0} \log(E|e^{t \gamma |X|^p}|) - \alpha n)}, \]
(40)
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 $\eta$, 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 right-hand side of (40), note that $\log(E[e^{t \gamma |X|^p}])$ is the cumulant generating function and is known to be convex [16] with respect to $t$, then $\log(E[e^{t \gamma |X|^p}]) - at$ is also convex, and its minimum is achieved where its first derivative with respect to $t$ is 0. Define $t^* := \arg \min_t (\log(E[e^{t \gamma |X|^p}]) - at)$, then we have
\[ 0 = \frac{d \log(E[e^{t \gamma |X|^p}]) - at}{dt} \bigg|_{t = t^*} \]
\[ = \frac{E[|X|^p e^{t \gamma |X|^p}]}{E[e^{t \gamma |X|^p}]} - a. \]
(41)
(41) determines $t^*$ given $a$. The derivative of $t^*$ with respect to $a$ is
\[ \frac{dt^*}{da} = \left( \frac{da}{dt^*} \right)^{-1} \]
\[ = \frac{(E[e^{t^* \gamma |X|^p}])^2}{E[e^{t^* \gamma |X|^p} E[|X|^p e^{t^* \gamma |X|^p}]} - (E[|X|^p e^{t^* \gamma |X|^p}])^2, \]
\[ (E[|X|^p e^{t^* \gamma |X|^p}])^2 \]
Note that $(E[|X|^p e^{t|X|^p}]^2 = (E[e^{t|X|^p}]^2 - (E[|X|^p e^{t|X|^p/2}]^2 < E[e^{t|X|^p}] E[|X|^{2p} e^{t|X|^p}]$, 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, $\frac{dt}{d\lambda} > 0$. 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 $\gamma$, 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 $\hat{a}(a,p,\gamma)$ such that the exponent on the righthand side of (40) is negative for all $a \geq \hat{a}(a,p,\gamma)$, and positive for all $a \leq \hat{a}(a,p,\gamma) - \epsilon$ for a very small $\epsilon > 0$. Therefore

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

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

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

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

\[P(\max_{z \in S} \|Bz\|^p_p \geq \frac{\hat{a}(a,p,\gamma)n}{1-\gamma^p}) = P(c_{\max} \geq \frac{\hat{a}(a,p,\gamma)}{1-\gamma^p}) \leq P(\eta \geq \frac{\hat{a}(a,p,\gamma)}{1-\gamma^p}) \leq e^{-c_{16}n},\]

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

\[\lambda_{\max}(a,p) = \min_{\gamma \in (0,1)} \hat{a}(a,p,\gamma)/(1-\gamma^p),\]

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

\\

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

Given $\alpha$ and $p$, define

\[c_{\max} = \frac{1}{n} \sup_{z \in S} \|Bz\|^p_p = \frac{1}{n} \max_{z \in S} \|Bz\|^p_p,\]

where the second equality holds by compactness. Thus, for any vector $z$, $\|Bz\|^p_p \leq \|z\|^p_{\max}n$. Define

\[c_{\min} = \frac{1}{n} \min_{z \in S} \|Bz\|^p_p.\]

Pick a $\gamma$-net $\Sigma$ of $S$ with cardinality at most $(1+2/\gamma)^{n-m}$ \cite{[9]} and $\gamma > 0$ to be chosen later, we define

\[\theta = \frac{1}{n} \min_{z \in \Sigma} \|Bz\|^p_p.\]

Then for every $z \in S$, there exists $z' \in \Sigma$ such that $\|z - z'\|_2 \leq \gamma$. We have

\[\|Bz\|^p_p \geq \|Bz'\|^p_p - \|B(z-z')\|^p_p \geq \theta n - \gamma^p \lambda_{\max}n,\]

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

\[c_{\min} \geq \theta - \gamma^p \lambda_{\max}.\]

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

We first consider the lower bound of $\theta$. For any constant $b > 0$, we will calculate the probability that $\theta$ 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 $\theta$. Given any constant $b > 0$,

\[P(\theta \leq b) = P(\exists z \in \Sigma \text{ s.t. } \|Bz\|^p_p \leq bn) \leq \sum_{z \in \Sigma} P(\|Bz\|^p_p \leq bn) \leq (1+2/\gamma)^{n-m} \min_{t>0} E[e^{-t\Sigma}]} \|Bz\|^p_p, \forall t > 0 \leq (1+2/\gamma)^{(1-\alpha)n} \min_{t>0} E[e^{-t|X|^p}], \forall t > 0 = e^{(1-\alpha)\log(1+2/\gamma)} + \log(E[|X|^p]) + bt) \leq 0, (46)\]

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 $P(\|Bz\|^p_p \leq bn)$ is the same for all $z \in \Sigma$ since $B$ has i.i.d. $\mathcal{N}(0,1)$ entries. Note that

\[E[e^{-t|X|^p}] = \sqrt{2/\pi} \int_0^\infty e^{-tx} e^{-\frac{1}{2}x^2} dx = t^{-\frac{1}{2}} \sqrt{2/\pi} \int_0^\infty e^{-y^2} e^{-\frac{1}{2}(t-\frac{1}{2})y^2} dy. (47)\]

\[\leq t^{-\frac{1}{2}} \sqrt{2/\pi} \int_0^\infty e^{-y^2} dy = t^{-\frac{1}{2}} \sqrt{2/\pi} \Gamma(1/p)/p, (48)\]

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

\[E[e^{-t|X|^p}] \geq t^{-\frac{1}{2}} \sqrt{2/\pi} \int_0^\infty e^{-y^2 - \frac{1}{2}y^2} dy.\]
Since \( \int_0^\infty e^{-y^2 - \frac{1}{2}y^2} dy \) exists and is positive, then combining (48) and (49), we have when \( t > 1 \),
\[
E[e^{\gamma |X|}] = \Theta(t^{-\gamma}).
\] (49)

Since (46) holds for all \( t > 0 \), we let \( t = \gamma^{-p(1-\alpha+\epsilon)} \) for any \( \epsilon \) 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 \( \gamma \) is sufficiently small, we have
\[
(1-\alpha)\log(1+\frac{2}{\gamma})+\log(\Theta(\gamma^{1-\alpha+\epsilon})) + 1 < 0.
\] (50)

Therefore when \( \gamma \leq \xi \) for some small enough \( \xi > 0 \), there exists \( \epsilon > 0 \) (depending on \( \gamma \) and \( \epsilon \)) such that
\[
P(\theta \leq \gamma^{p(1-\alpha+\epsilon)}) \leq e^{-\epsilon\gamma n}. \] (51)

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

Lemma 9 indicates that there exists \( \lambda_{\max}(\alpha, p) \) and \( c_{10} > 0 \) such that
\[
P(c_{\max} < \lambda_{\max}(\alpha, p)) \geq 1 - e^{-c_{10}n} \] (52)

Then after characterizing \( \theta \) and \( c_{\max} \) separately, we are ready to characterize \( c_{\min} \). We have
\[
P(c_{\min} < \gamma^{p(1-\alpha+\epsilon)} - \gamma^p\lambda_{\max}(\alpha, p)) \leq P(\theta - \gamma^{p}c_{\max} \leq \gamma^{p(1-\alpha+\epsilon)} - \gamma^p\lambda_{\max}(\alpha, p)) \leq P(\theta \leq \gamma^{p(1-\alpha+\epsilon)}) + P(c_{\max} \geq \lambda_{\max}(\alpha, p)) \leq e^{-\epsilon\gamma n} + e^{-c_{10}n},
\]

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 \( \epsilon \) and \( \gamma \)) such that \( P(c_{\min} \leq \gamma^{p(1-\alpha+\epsilon)} - \gamma^p\lambda_{\max}(\alpha, p)) \leq e^{-c_9\gamma n}. \) Given \( \lambda_{\max}(\alpha, p) \), let
\[
\lambda_{\min}(\alpha, p) = \max_{0<\epsilon<\alpha,0<\gamma<\xi(\epsilon)} \gamma^{p(1-\alpha+\epsilon)} - \gamma^p\lambda_{\max}(\alpha, p).
\] (53)

Note that since \( 1 - \alpha + \epsilon < 1 \), \( \gamma^{p(1-\alpha+\epsilon)} - \gamma^p\lambda_{\max}(\alpha, p) > 0 \) when \( \gamma \) 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 |\( T \)\| = \( mn \) \( (0 < \rho < 1) \), define
\[
d_{\max} = \frac{1}{n} \max_{z \in S} \|B_T z\|_p.
\]

Since \( B \) has i.i.d. Gaussian entries, then the distribution of \( d_{\max} \) is the same for any \( T \) with |\( T \)\| = \( mn \). Given a \( \gamma \)-net \( \Sigma_3 \) of \( S \) with cardinality at most \( (1+2/\gamma)^{n-m} \) and \( \gamma > 0 \) to be chosen later, define
\[
\tau = \frac{1}{n} \max_{z \in \Sigma_3} \|B_T z\|_p,
\]

Then for every \( z \in S \), there exists \( z' \in \Sigma_3 \) such that \( \|z - z'|_2 \leq \gamma \). Then for every \( z \in S \), we have \( \|B_T z\|_p \leq \|B_T z'\|_p + \|B_T (z - z')\|_p \leq \tau n + \gamma \tau d_{\max}. \) That means \( d_{\max} \leq \tau/(1-\gamma^2). \) (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 \( \rho \) such that the probability that \( c_{\max} \geq \lambda_{\min}/2 \) holds for some support \( T \) with |\( T \)\| = \( mn \) can still decay exponentially to zero. From (54), we first consider the probability that \( \tau \geq \lambda_{\min}(1-\gamma^2)/2 \) holds for a given set \( T \).

\[
P(\tau \geq \lambda_{\min}(1-\gamma^2)/2, \text{ given } T) = P(\exists z \in \Sigma_3 \text{ s.t. } \|B_T z\|_p \geq \lambda_{\min}(1-\gamma^2)n/2) \leq \sum_{z \in \Sigma_3} P(\|B_T z\|_p \geq \lambda_{\min}(1-\gamma^2)n/2) \leq \sum_{z \in \Sigma_3} P(\|z\|_p \geq \lambda_{\min}(1-\gamma^2)n/2) \leq (1+2/\gamma)^{n-m} \min_{t>0} e^{-t\lambda_{\min}(1-\gamma^2)n/2}\epsilon^{\epsilon t^{\epsilon}1/n} \leq (1+2/\gamma)^{(1-\alpha)n} \min_{t>0} e^{-t\lambda_{\min}(1-\gamma^2)n/2}\epsilon^{\epsilon t^{\epsilon}1/n} \leq (1-\alpha)(1-\gamma^2)) \min_{t>0} (\epsilon^{\epsilon t^{\epsilon}1/n}1/2) \] (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 |\( T \)\| = \( mn \).

Now consider the probability that \( \|B_T z\|_p \geq \frac{1}{2}\lambda_{\min} \) for some \( z \in S \) and \( T \) with |\( T \)\| = \( mn \).

\[
P(\exists z \in S, \exists T, \text{ s.t. } |T| = mn, \|B_T z\|_p \geq \lambda_{\min}/2) \leq \left( \frac{n}{\log(1+2/\gamma)} \right)^{n-m} \min_{t>0} e^{t\lambda_{\min}(1-\gamma^2)n/2}\epsilon^{\epsilon t^{\epsilon}1/n} \leq \left( \frac{n}{\log(1+2/\gamma)} \right)^{n-m} \min_{t>0} e^{t\lambda_{\min}(1-\gamma^2)n/2}\epsilon^{\epsilon t^{\epsilon}1/n} \leq \left( \frac{n}{\log(1+2/\gamma)} \right)^{n-m} \min_{t>0} (\epsilon^{\epsilon t^{\epsilon}1/n}1/2) \] (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 \( \left( \frac{n}{\log(1+2/\gamma)} \right)^{n-m} \leq \frac{2^{\alpha H(\rho)}}{H(\rho)} \), where \( H(\rho) = -\rho \log(\rho) - (1-\rho) \log(1-\rho) \).

To obtain a good upper bound of \( P(\exists z \in S, \exists T, \text{ s.t. } |T| = mn, \|B_T z\|_p \geq \lambda_{\min}/2) \), we first would like to solve the minimization problem on the right-hand side of (56). Note that log\( (E[e^{1/1+n}]) \) is the cumulant generating function and is known to be convex [16] with respect to \( t \). Then \( \rho \log(\epsilon^{-1/1+n}) - t\lambda_{\min}(1-\gamma^2)/2 \) is also convex, then its minimum is achieved when its first derivative with respect to \( t \)
\( \lambda \) is 0. Define \( t^* := \arg\min_t [\rho \log(E[e^t|X|^p]) - \lambda\min(1 - \gamma^p)/2] \).

We have

\[
0 = \frac{d[\rho \log(E[e^{\lambda |X|^p}]) - \lambda\min(1 - \gamma^p)/2]_t}{dt} \bigg|_{t = t^*} = \frac{\rho E[|X|^p e^{\lambda |X|^p}]}{E[e^{\lambda |X|^p}]} - \lambda\min(1 - \gamma^p)/2,
\]

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 \( \rho, \lambda\min \) and \( \gamma \). The derivative of \( t^* \) with respect to \( \rho \) is

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

Note that \( (E[|X|^p e^{t^* |X|^p}]^2 - E[e^{t^* |X|^p}]E[|X|^p e^{t^* |X|^p}]) \) is 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 \( \min_{t > 0} (\rho \log(E[e^{t |X|^p}]) - \lambda\min(1 - \gamma^p)/2) \). Given \( \rho, \gamma \), and \( \alpha \), we can numerically compute \( t^* \) by (57) and plug it into (56) to obtain an upper bound of \( P(\exists z \in S, T \mid |T| = \rho n, \|B_T z\|^p_P \geq \lambda\min n/2) \).

Note that given \( \alpha, \rho, \lambda\min \), (56) provides an upper bound of the probability that there exists some \( z \in S \) and some \( T \) with \( |T| = \rho n \) such that \( \|B_T z\|^p_P \geq \lambda\min n/2 \). The next question is how large \( \rho \) can be such that this upper bound still decays exponentially to zero. The largest \( \rho \) is indeed the \( \rho^*(\alpha, p) \) we would like to calculate.

Note that given \( \alpha, p, \) and \( \lambda\min \), for every \( \gamma \), as \( \rho \) goes to 0, \( H(\rho) \) goes to 0, and \( \min_{t > 0} (\rho \log(E[e^{t |X|^p}]) - \lambda\min(1 - \gamma^p)/2 \) goes to \(-\infty \), thus, there exists \( \hat{\rho}(\alpha, p, \gamma) > 0 \) such that the exponent on the right-hand side of (56) is negative for all \( \rho \leq \hat{\rho}(\alpha, p, \gamma) \), and is positive for all \( \rho > \hat{\rho}(\alpha, p, \gamma) + \epsilon \) for some very small \( \epsilon > 0 \). In other words, for each \( \gamma \), \( P(\exists z \in S, T \mid |T| = \rho n, \|B_T z\|^p_P \geq \lambda\min n/2) \leq e^{-c\rho} \) for some positive \( c \) depending on \( \gamma \). We then optimize \( \hat{\rho}(\alpha, p, \gamma) \) over \( \gamma \in (0, 1) \), and let

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

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

L. Proof of Theorem 7

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

\[
P(\forall \text{ non-zero } z \in \mathcal{R}^{n-m}, \forall T \text{ with } |T| = \rho^*(\alpha, p, n), \|B_T z\|^p_P < 1/2 \|B_T z\|^p_P) \leq P(\forall z \in S, \forall T \mid |T| = \rho^*(\alpha, p, n), \|B_T z\|^p_P < 1/2 \|B_T z\|^p_P) \leq P(\forall z \in S, \forall T \mid |T| = \rho^*(\alpha, p, n), \|B_T z\|^p_P \leq \lambda\min(\alpha, p) n, \|B_T z\|^p_P > \lambda\min(\alpha, p) n) \leq 1 - P(\exists z \in S, \forall T \mid |T| = \rho^*(\alpha, p, n) - \|B_T z\|^p_P \leq \lambda\min(\alpha, p) n) \leq 1 - e^{-c_{11}n} \|B_T z\|^p_P \geq \lambda\min(\alpha, p) n/2) \leq 1 - e^{-c_{11}n}.
\]

where the first equality follows from Theorem 1, the second equality holds since for any non-zero \( z \in \mathcal{R}^{n-m}, \|z\|_2 \leq 1 \). From Lemma 5 we know there exists \( c_9 > 0 \) such that \( P(\exists z \in S, \forall T \mid |T| = \rho n, \|B_T z\|^p_P \leq \lambda\min(\alpha, p) n) \leq e^{-c_{10}n} \), and from Lemma 6 we know there exists \( c_{10} > 0 \) such that \( P(\exists z \in S, \forall T \mid |T| = \rho n, \|B_T z\|^p_P \geq \lambda\min(\alpha, p) n) \leq e^{-c_{11}n} \), then there exists \( c_{11} > 0 \) which depends on \( \alpha, p \) and \( \lambda\min \) such that the right-hand side of (58) is greater than \( 1 - e^{-c_{11}n} \). Therefore, \( \ell_p \)-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 \( \alpha' = \frac{\alpha}{\rho} \). Define \( \alpha'_{\max} = \max_{\rho \in (1-n)\rho_{\max} \leq \rho} \|B_T z\|^p_P \). Let \( \Sigma_4 \) be a \( \gamma \)-net of \( S \) with cardinality at most \( (1 + 2/\gamma)^{n-m} \) and be the value where \( \lambda\max(\alpha', p) \) is achieved in (43). We use \( \lambda\max \) to denote \( \lambda\max(\alpha', p) \) simplicity here in the proof. From (43) we have

\[
\lambda\max = \tilde{a}(\alpha', p, \gamma)/(1 - \gamma^p), \tag{59}
\]

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

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

(60)

Combining (59) and (60), we have

\[
(1 - \gamma^p) \log(1 + 2/\gamma) + \min_{t > 0} (\log(E[e^{1/2 |X|^p}]) - \lambda\max(1 - \gamma^p)t) < 0.
\]

(61)

Define

\[
\eta' = \frac{1}{(1 - \gamma^p)\rho_{\max} \max_{z \in \Sigma_4} \|B_T z\|^p_P}.
\]

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

\[
\eta' \leq \lambda\max/(1 - \gamma^p).
\]

We first show that with overwhelming probability, \( \|B_T z\|^p_P < (1 - \rho)\lambda\max n \) for all \( z \in S \), or equivalently

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

then with probability at least \( 1 - e^{-c_{10}n} \) for some \( c_{10} > 0 \), for every \( z \in S \) and for every set \( T \subset \{1, 2, \ldots, n\} \) with \( |T| \leq \rho^*(\alpha, p) n, \|B_T z\|^p_P < \lambda\min n/2 \) holds simultaneously. Then Lemma 6 follows.
$c'_{\text{max}} < \lambda_{\text{max}}$. Note that
\[
P(c'_{\text{max}} \geq \lambda_{\text{max}}) \leq P(\eta'/(1 - \gamma^p) \geq \lambda_{\text{max}})
\]
\[
= P(\exists z \in \Sigma_{T} \text{ s.t. } ||B_T z||_p^p \geq (1 - \rho)\lambda_{\text{max}}(1 - \gamma^p)n)
\]
\[
\leq \sum_{z \in \Sigma_{T}} P(||B_T z||_p^p \geq (1 - \rho)\lambda_{\text{max}}(1 - \gamma^p)n)
\]
\[
\leq (1 + \frac{2}{\gamma})\min_{t > 0} \frac{E[e^{\gamma E}[\|X\|^p]]}{\gamma} \max_{t > 0} \min_{\rho}(1 - \rho)(1 - \gamma^p)n
\]
\[
= e^{(1 - \rho)n(1 - \alpha') \log(1 + \frac{2}{\gamma}) + \min_{t > 0}(\log(E[e^{\gamma E}[\|X\|^p]]) - \lambda_{\text{max}}(1 - \gamma^p)n))}
\]
\[
= e^{(1 - \rho)n(1 - \alpha') \log(1 + \frac{2}{\gamma}) + \min_{t > 0}(\log(E[e^{\gamma E}[\|X\|^p]]) - \lambda_{\text{max}}(1 - \gamma^p)n))}
\]
\[
\leq \frac{(1 + \frac{2}{\gamma})(1 - \alpha') n}{\min_{t > 0} \frac{E[e^{\gamma E}[\|X\|^p]]}{\gamma} \max_{t > 0} \min_{\rho}(1 - \rho)(1 - \gamma^p)n}
\]
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
\[
P(\theta' \leq \gamma p(1 - \alpha' + \epsilon)) \leq e^{-Kn},
\]
for some positive $K > 0$. Thus, from (64) and (66) we have
\[
P(\alpha_{\text{max}}^* \leq \lambda_{\text{max}}) \leq e^{-Kn} + e^{-c_{12}n} \leq e^{-c_{12}n},
\]
for some $c_{13} > 0$. Then, with probability at least $1 - e^{-c_{13}n}$, for all $z \in S$, $||B_T z||_p^p > (1 - \rho)\lambda_{\text{max}}(\alpha', p)n$.

\section{Calculation of $\lambda_{\text{max}}(\alpha, p, \rho)$ in Lemma 8}

\textbf{Proof:} Define $\tilde{c}_{\text{max}} = \frac{1}{p_{\text{max}}} \max_{z \in S} ||B_T z||_p^p$. Let $\Sigma_{T}$ be a $\gamma$-net of $S$ with cardinality at most $(1 + 2/\gamma)^n$ and $\gamma > 0$ to be chosen later, and define $\tilde{\eta} = \frac{1}{p_{\text{max}}} \max_{z \in S} ||B_T z||_p^p$. Then from (30), for any $z \in S$, $z = \sum_{j > 0} \gamma v_j$, holds where $\gamma_j = 1$, $\gamma_j \leq \gamma^j$ and $v_j \in \Sigma_{T}$. From (36) we have
\[
||B_T z||_p^p \leq \sum_{j > 0} \gamma^j \sum_{i \in T : B_i v_j < 0} |B_i v_j|_p^p
\]
\[
\leq \sum_{j > 0} \gamma^j \tilde{\eta} \rho m
\]
\[
= \tilde{\eta} \rho m/(1 - \gamma^p),
\]
where the second inequality follows from the definition of $\tilde{\eta}$. Since (67) holds for every $z \in S$, then $\tilde{c}_{\text{max}} \rho m \leq \tilde{\eta} \rho m/(1 - \gamma^p)$, which leads to $\tilde{c}_{\text{max}} \leq \tilde{\eta} \rho m/(1 - \gamma^p)$. For any given $z \in S$, define a random variable $S_i$ for each $i \in T$, and $S_i$ is equal to 1 if $B_i z < 0$ and equal to 0 otherwise. Then $||B_T z||_p^p = \sum_{i \in T} |B_i z|_p^p S_i$.

Given $\gamma$, for any $\tilde{a}$, we will characterize the probability that $\tilde{c}_{\text{max}}$ 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 $c_{\text{max}}$. Note that
\[
P(c_{\text{max}} \geq \tilde{a}/(1 - \gamma^p)) \leq P(\tilde{\eta} \rho m/(1 - \gamma^p) \geq \frac{\tilde{a}}{1 - \gamma^p})
\]
\[
= P(\tilde{\eta} \geq \tilde{a}) = P(\exists z \in \Sigma_{T} \text{ s.t. } ||B_T z||_p^p \geq \tilde{a} \rho m)
\]
\[
\leq \sum_{z \in \Sigma_{T}} P(||B_T z||_p^p \geq \tilde{a} \rho m)
\]
\[
= (1 + \frac{2}{\gamma})^{-n \rho m} \sum_{i \in T} |B_i z|_p^p S_i \geq \tilde{a} \rho m
\]
\[
\leq (1 + \frac{2}{\gamma})^{-n \rho m} \min_{t > 0} \frac{E[e^{\gamma E}[\|X\|^p]]}{\gamma} \max_{t > 0} \min_{\rho}(1 - \rho)(1 - \gamma^p)n
\]
\[
= e^{(1 - \rho)n(1 - \alpha') \log(1 + \frac{2}{\gamma}) + \rho \min_{t > 0}(\log(E[e^{\gamma E}[\|X\|^p]]) - \tilde{a} \rho m))},
\]
where $X \sim \mathcal{N}(0, 1)$, $S = 1$ if $X < 0$ and $S = 0$ otherwise.

To solve the minimization problem in the right-hand side of (68), note that log$(E[e^{\gamma E}[\|X\|^p]])$ is the cumulant generating function and is convex [16] with respect to $t$, then log$(E[e^{\gamma E}[\|X\|^p]]) - \tilde{a} \rho m$ is also convex, and its minimum is achieved where its first derivative with respect to $t$ is 0. Define
\[
\lambda_{\text{max}}(\alpha, p, \rho)
\]
\[ t^* := \arg\min_t \log \mathbb{E}[e^{t|X|^p}] - \tilde{a}t, \]
then we have
\[ 0 = \left. \frac{d}{dt} \log \mathbb{E}[e^{t|X|^p}] \right|_{t=t^*} = \frac{E[X|p] e^{t^*|X|^p}}{E[e^{t^*|X|^p}]} - \tilde{a}. \]  
(69)
determines \( t^* \) given \( \tilde{a} \). The derivative of \( t^* \) with respect to \( \tilde{a} \) is
\[ \frac{dt^*}{d\tilde{a}} = \left( \frac{d\tilde{a}}{dt^*} \right)^{-1} = \frac{(E[e^{t^*|X|^p}])^2}{E[e^{t^*|X|^p}] E[|X|^p] e^{t^*|X|^p} - (E[|X|^p] e^{t^*|X|^p})^2} > 0, \]
where the inequality follows from Cauchy-Schwarz inequality. Since when \( \tilde{a} = E[|X|^p] \), \( t^* = 0 \) from (69), then when \( \tilde{a} > E[|X|^p] \), we have \( t^* > 0 \). Thus \( t^* = \min_{t>0} (\log (E[e^{t|X|^p}]) - \tilde{a}t) \) when \( \tilde{a} > E[|X|^p] \). Given \( \tilde{a} \), we can numerically compute \( t^* \) by (69) and plug it into (68) to obtain an upper bound of \( \mathbb{P}(\lambda_{\max} \geq \frac{\tilde{a}}{1-\gamma_0}) \).

Then the question is how small can \( \tilde{a} \) be while the exponent on the righthand side of (68) is still negative. Given \( \gamma \), 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 \( \log(E[e^{t|X|^p}]) - \tilde{a}t \) goes to \(-2(1-\alpha) \log(1+2/\gamma)/\rho \) as \( \tilde{a} \) goes to infinity. Thus, when \( \tilde{a} \) is sufficiently large, \( \rho \min_{t>0} (\log (E[e^{t|X|^p}]) - \tilde{a}t) < -2(1-\alpha) \log(1+2/\gamma) \).

Therefore, the exponent on the righthand side of (68) is negative when \( \tilde{a} \) is large enough. Thus, we can pick \( \tilde{a}(\alpha, p, \rho, \gamma) \) such that the exponent on the righthand side of (68) is negative for all \( \alpha \geq \tilde{a}(\alpha, p, \rho, \gamma) \), and positive for all \( a \leq \tilde{a}(\alpha, p, \rho, \gamma) - \epsilon \) for some small enough \( \epsilon \). Therefore
\[ (1-\alpha) \log(1+2/\gamma) + \rho \min_{t>0} (\log (E[e^{t|X|^p}]) - \tilde{a}(\alpha, p, \rho, \gamma)t) < 0. \]

Then there exists some constant \( c_{14} > 0 \) such that
\[ P(\tilde{c}_{\max} \geq \frac{\tilde{a}(\alpha, p, \rho, \gamma)}{1-\gamma_0}) \leq e^{\left((1-\alpha) \log(1+2/\gamma) + \rho \min_{t>0} (\log (E[e^{t|X|^p}]) - \tilde{a}(\alpha, p, \rho, \gamma)t)\right)n} = e^{-c_{14}n}. \]

Thus, for all \( \gamma \in (0, 1) \), \( \tilde{a}(\alpha, p, \rho, \gamma) \rho n/(1-\gamma_0) \) can be viewed as an upper bound of \( \|B_T-z\|_p^\gamma \) for all \( z \in S \) in the sense that the probability that \( \|B_T-z\|_p^\gamma \geq \tilde{a}(\alpha, p, \rho, \gamma) \rho n/(1-\gamma_0) \) for some \( z \in S \) decays exponentially to zero. Since we would like such an upper bound to be as small as possible, we let
\[ \tilde{\lambda}_{\max}(\alpha, p, \rho) = \min_{\gamma \in (0, 1)} \frac{\tilde{a}(\alpha, p, \rho, \gamma)}{1-\gamma_0}, \]  
(70)
then with overwhelming probability, \( \tilde{c}_{\max} < \tilde{\lambda}_{\max}(\alpha, p, \rho) \), or equivalently, for every \( z \in S \), \( \|B_T-z\|_p^\gamma < (1-\rho)\tilde{\lambda}_{\max}(\alpha, p, \rho)n \). 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 \( \rho_w^* \) 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,
\[ \rho_w^* \tilde{\lambda}_{\max}(\alpha, p, \rho_w^*) \leq (1-\rho_w^*) \lambda_{\min}(\alpha-\rho_w^*, p). \]  
(71)

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

Now consider the probability that \( \ell_p \)-minimization can recover all the \( \rho_w^* \)-sparse \( x \) on one fixed support \( T \) with one fixed sign pattern. From Theorem 3 we know that \( \|B_T-z\|_p^p \geq \|B_T-z\|_p^\gamma \) for all non-zero \( z \in \mathbb{R}^{n-m} \) is a sufficient condition for the success of weak recovery, thus
\[ P(\text{Weak recovery succeeds up to } \rho_w^* \text{-sparse}) \geq P(\forall \text{ non-zero } z \in \mathbb{R}^{n-m}, \|B_T-z\|_p^p < \|B_T-z\|_p^\gamma) \]
\[ = P(\forall z \in S, \|B_T-z\|_p^p < \|B_T-z\|_p^\gamma) \]
\[ \geq P(\forall z \in S, \|B_T-z\|_p^p < \rho_w^* \tilde{\lambda}_{\max}(\alpha, p, \rho_w^*)) \text{ and }\]
\[ \|B_T-z\|_p^p > (1-\rho_w^*) \lambda_{\min}(\alpha-\rho_w^*, p)) \]
\[ \geq 1 - e^{-c_{14}n} - e^{-c_{14}n}, \]  
(72)
where the first equality holds since for any non-zero \( z \in \mathbb{R}^{n-m} \), \( z/\|z\|_2 \in S \), and the second inequality follows from (71). From Lemma 7 we know there exists \( c_{14} > 0 \) such that \( P(\forall z \in S, \|B_T-z\|_p^p > (1-\rho_w^*) \lambda_{\min}(\alpha-\rho_w^*, p)) \) \( \geq 1 - e^{-c_{14}n} \), and from Lemma 8 we know there exists \( c_{15} > 0 \) such that \( P(\forall z \in S, \|B_T-z\|_p^p < \rho_w^* \tilde{\lambda}_{\max}(\alpha, p, \rho_w^*)) \geq 1 - e^{-c_{14}n} \), then the third inequality of (72) holds from the union bound.

Thus, there exists \( c_{15} > 0 \) such that with probability at least \( 1 - e^{-c_{14}n} \), \( \ell_p \)-minimization problem can recover all \( \rho_w^* \)-sparse vectors on fixed support \( T \) with fixed sign pattern, then Theorem 8 holds.

Now consider the case that there is no \( \rho_w^* \) satisfying (71), where \( \rho^*(\alpha, p) \) 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 (’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.