Invertibility of Sparse Complex Gaussian Matrices

Edward Zeng
UC Berkeley
edwardzeng@berkeley.edu

June 10, 2022

Abstract
Let $\sigma_n(\cdot)$ denote the least singular value of a $n \times n$ matrix. It is well-known that
$$\mathbb{P}[\sigma_n(A) \leq \varepsilon] \leq \varepsilon n$$
if $A$ is drawn from the real Ginibre ensemble of $n \times n$ matrices and
$$\mathbb{P}[\sigma_n(A) \leq \varepsilon] \leq \varepsilon^2 n^2$$
if $A$ is drawn from the complex Ginibre ensemble. In this paper, we will show a similar phenomenon occurs for sparse random matrices.

1 Introduction
Let $\sigma_n(\cdot)$ denote the least singular value of a $n \times n$ matrix. It is well-known that
$$\mathbb{P}[\sigma_n(A) \leq \varepsilon] \leq \varepsilon n$$
if $A$ is drawn from the real Ginibre ensemble of $n \times n$ matrices and
$$\mathbb{P}[\sigma_n(A) \leq \varepsilon] \leq \varepsilon^2 n^2$$
if $A$ is drawn from the complex Ginibre ensemble. (For proofs, see [Ede88].) In this paper, we will show a similar phenomenon occurs for sparse random matrices.

Let $\delta_p$ denote the Bernoulli random variable
$$\mathbb{P}[\delta_p = 1] = p, \quad \mathbb{P}[\delta_p = 0] = 1 - p.$$
In [BR17] Basak and Rudelson prove the following result:

Theorem 1.1. Let $A$ be a $n \times n$ random matrix with i.i.d. entries $A_{ij} = \xi_{ij}\delta_{ij}$ where
1. $\xi_{ij}$ and $\delta_{ij}$ are all independent.

2. $\xi_{ij}$ are i.i.d. sub-Gaussian random variables with

\[ E\xi_{ij} = 0, \quad E\xi_{ij}^2 = 1, \quad \xi_{ij} \in \mathbb{R}. \]

3. $\delta_{ij}$ are i.i.d. copies of $\delta_p$.

Then there exist constants $c, C, C' > 0$ depending on the sub-Gaussian norm of $\xi_{ij}$ such that for $p \geq C' \log n$ we have

\[ P[\sigma_n(A) \leq \varepsilon] \leq e^{-cnp} + C \varepsilon \exp \left( c \log \left( \frac{1}{np} \right) \right) \cdot \sqrt{\frac{n}{p}}, \quad \varepsilon \geq 0. \]

Let us consider a special case of Theorem 1.1 where we fix $\delta \in (0, 1)$ and set

\[ p = n^{\delta-1} = \frac{n^{\delta}}{n}. \]

Then there exist constants $c, C > 0$ depending on the sub-Gaussian norm of $\xi_{ij}$ such that for sufficiently large $n$ we have

\[ P[\sigma_n(A) \leq \varepsilon] \leq e^{-cn^\delta} + C \varepsilon \exp \left( c \frac{1 - \delta}{\delta} \right) \cdot n^{\frac{\delta n}{2}} \]

\[ = e^{-cn^\delta} + O(\varepsilon n^{\frac{\delta n}{2}}). \]

In this paper, we will show that the dependence on $\varepsilon$ can be upgraded to $\varepsilon^2$ if we move from real sub-Gaussian entries to complex Gaussian entries. More formally, our main result is:

**Theorem 1.2.** Fix $\delta \in (0, 1)$ and let $A$ be the $n \times n$ random matrix in Theorem 1.1 but with

$\xi_{ij} \sim \mathcal{N}_\mathbb{C}(0, 1), \quad \delta_{ij} \overset{d}{=} \delta_n \delta^{1-1}$.

Then there exists a constant $c > 0$ such that for sufficiently large $n$ we have

\[ P[\sigma_n(A) \leq \varepsilon] \leq e^{-cn^\delta} + O(\varepsilon^2 n^{2-\delta}), \quad \varepsilon \geq 0. \]

Here and throughout this paper, $\mathcal{N}_\mathbb{C}(0, 1)$ will denote the complex normal distribution with mean 0 and variance 1.

## 2 Proof Overview

From now on, $A$ will refer to the random matrix as defined Theorem 1.2.

The proof of Theorem 1.2 will require careful analysis of nets on the unit sphere. Because such arguments need upper bound estimates on $\|A\|_{op}$, we will break up the proof of Theorem 1.2 into two parts:
Theorem 2.1. For any $K > 0$, there exists a constant $c > 0$ such that for sufficiently large $n$ we have
\[ P[\sigma_n(A) \leq \varepsilon \text{ and } \|A\|_{op} \leq Kn^{\frac{3}{2}}] \leq e^{-cn^d} + O(\varepsilon^2 n^{2-d}), \quad 0 \leq \varepsilon \leq 1. \]

Theorem 2.2. There exists constants $K, c > 0$ such that for sufficiently large $n$ we have
\[ P[\|A\|_{op} \geq Kn^{\frac{3}{2}}] \leq e^{-cn^d}. \]

Combining Theorems 2.1 and 2.2 with a union bound gives Theorem 1.2.

Remark. Careful readers may have noticed that Theorem 2.1 restricts $\varepsilon$ to $[0, 1]$. It is very easy to extend the statement to all $\varepsilon \geq 0$; the added restriction will just make the proof a little simpler.

2.1 Theorem 2.1 Proof Overview

As the proof of Theorem 2.1 comprises the bulk of this paper, we outline it here. Let $Y_i$ denote the $i$-th column of $A$ and let $W_i$ denote the span of the $Y_j$’s for $j \neq i$. Using standard arguments, we show it suffices to prove that it is unlikely for $\text{dist}(Y_i, W_1)$ to be small. Observe that
\[ \text{dist}(Y_1, W_1) \text{ small } \iff \|Y_1, \eta_1\| \text{ small for any } \eta_1 \in W_1^\perp, \|\eta_1\|_2 = 1. \quad (1) \]

Notice that $\eta_1 \in W_1^\perp \iff \eta_1 \in \ker(B)$, where $A = [Y_1 B^*]$. Writing $S^{n-1} \subset \mathbb{C}^n$ as the unit sphere, we can rewrite (1) as
\[ \text{dist}(Y_1, W_1) \text{ small } \iff \|Y_1, \eta_1\| \text{ small for any } \eta_1 \in \ker(B) \cap S^{n-1}. \]

We show that for generic vectors $\eta \in S^{n-1}$ we have good control on the lower tail of $\|Y_1, \eta\|$. In the rare case that $\eta$ is structured and does not provide a good lower tail bound on $\|Y_1, \eta\|$, we show that it is unlikely $\eta \in \ker(B)$. Combining these arguments will then provide us a proof of Theorem 2.1.

This general strategy of reducing the least singular problem into a statement of distance between columns and then doing casework was developed by Rudelson in [Rud08] and has since become the standard technique in attacking such problems. [BR17] and this paper are no exceptions to the rule. However, the technical difficulties that arise when working with sparse matrices are markedly different than in other settings. We will draw some inspiration from [BR17] because of this, but our casework and analysis will differ from [BR17] significantly.

2.2 Generic and Structured Vectors

Let us now motivate our definitions of generic and structured vectors by considering the distribution of $|\langle Y_1, \eta \rangle|$ for $\eta \in S^{n-1}$. 

Let $\hat{\eta}, G \in \mathbb{C}^n$ denote the random vectors

$$\hat{\eta}_i = \delta_{i1} \eta_i, \quad G_i = \xi_{i1}. $$

(Notice that $G$ is a Gaussian random vector.) Then we can rewrite

$$|\langle Y_1, \eta \rangle| = |\langle G, \hat{\eta} \rangle|. $$

Observe if we condition on the value of $\|\hat{\eta}\|_2$, we obtain

$$\langle G, \hat{\eta} \rangle \sim \mathcal{N}(0, \|\hat{\eta}\|_2). $$

Consequently, if we want a good lower tail bound on $|\langle Y_1, \eta \rangle|$, we want it to be unlikely for $\|\hat{\eta}\|_2$ to be small.

How do we make it unlikely for $\|\hat{\eta}\|_2$ to be small? One way is to require $\eta$ to have many coordinates of a certain size so that it is likely some coordinates of this size will not be zeroed out. This leads us to our definition.

Partition the unit sphere $S^{n-1}$ into the sets

$$S_{IC} = \left\{ x \in S^{n-1} : \sum_{|x_i|^2 \leq \frac{1}{c_1^2}} |x_i|^2 \geq \varepsilon_1 \right\}$$

$$S_{MC} = \left\{ x \in S^{n-1} : \sum_{|x_i|^2 \leq \frac{1}{c_2^2}} |x_i|^2 < \varepsilon_1, \quad \sum_{\frac{1}{c_2^2} < |x_i|^2 \leq \frac{1}{c_2^2} - \varepsilon_2} |x_i|^2 \geq \varepsilon_2 \right\}$$

$$S_{HC} = \left\{ x \in S^{n-1} : \sum_{|x_i|^2 \leq \frac{1}{c_3^2} - \varepsilon_3} |x_i|^2 < \varepsilon_1 + \varepsilon_2 \right\}$$

where $c_1, c_2, \varepsilon_1, \varepsilon_2$ are constants that will be determined later. We call $S_{IC}$ the set of incompressible (or generic) vectors and call $S^{n-1} - S_{IC}$ the set of compressible (or structured) vectors. In particular, we will call $S_{MC}$ the set of moderately compressible vectors, and $S_{HC}$ the set of highly compressible vectors. In the following sections, we first show that when $\eta$ is compressible, it is likely that $\eta \notin \ker(B)$, or equivalently,

$$\|B\eta\|_2 > 0.$$ 

For presentation purposes (and to avoid proof duplication), we will do this by showing it is likely that

$$\|A\eta\|_2$$

cannot be too small for compressible $\eta$. The argument will work in exactly the same way when $A$ is replaced by $B$. Because of technical issues introduced by sparsity, we will treat the case $\eta \in S_{IC}$ (Section 5) differently from the case $\eta \in S_{MC}$ (Section 6).
On the other hand, in Section 7 we show when \( \eta \) is incompressible, it is likely that 
\[
|\langle Y_1, \eta \rangle|
\]
cannot be too small. We then put together our arguments in Section 8 to prove Theorem 2.1.

3 Notation

For the reader’s convenience, we collect some notation that will continue to appear throughout the paper.

| Symbol | Meaning |
|--------|---------|
| \( \delta_p \) | As defined in Section 1. |
| \( n, A, \delta_{ij}, \xi_{ij}, \delta \) | As defined in Theorem 1.2. |
| \( Y_i \) | \( i \)-th column of \( A \). |
| \( W_i \) | Span of the \( Y_i \)'s for \( j \neq i \). |
| \( B \) | Matrix such that \( A = [Y_1 \ B^\ast] \). |
| \( S^{n-1} \) | Complex unit sphere \( S^{n-1} \subset \mathbb{C}^n \). |
| \( S_{\{HC,MC,IC\}} \) | As defined in Section 2.2. |
| \( c_1, c_2, \xi_1, \xi_2 \) | As defined in Section 2.2. |
| \( \mathcal{N}_C(\mu, \sigma^2) \) | Complex Gaussian distribution. |
| \( \mathcal{N}_R(\mu, \sigma^2) \) | Real Gaussian distribution. |

4 Preliminaries

In this section we will investigate the properties of the set
\[
V = \left\{ x \in S^{n-1} : \sum_{|x_i|^2 \leq \frac{1}{2}} |x_i|^2 < d_1, \sum_{\frac{1}{2} < |x_i|^2 \leq \frac{1}{2}} |x_i|^2 \geq d_2 \right\}.
\]
We will assume \( 0 < d_1 \leq 0.1 \) and \( d_1 \ll d_2 \), which hold in all our use cases.

Define
\[
\text{Sparse}(m) = \{ x \in \mathbb{C}^n : |\text{supp}(x)| \leq m \}.
\]
Let \( \mathcal{N} \) be a maximal \( \sqrt{d_1} \)-net of the \( \text{Sparse}(a) \) vectors on \( S^{n-1} \). Let us approximate any \( x \in V \) with an element in \( \mathcal{N} \) in the following manner:

1. Set \( x^{(0)} = x \in V \).
2. Define \( x^{(1)} \in \mathbb{C}^n \) such that
\[
x_i^{(1)} = \begin{cases} x_i^{(0)} & \frac{1}{\delta} < |x_i^{(0)}|^2 \\ 0 & \text{otherwise}. \end{cases}
\]
3. Set \( x^{(2)} = x^{(1)}/\|x^{(1)}\|_2 \).
4. Let $x^{(3)} \in \mathcal{N}$ be $\sqrt{d_1}$-approximation of $x^{(2)}$.

Notice that

$$\|x - x^{(3)}\|_2 \leq \sum_{i=0}^{2} \|x^{(i)} - x^{(i+1)}\|_2 \leq 3\sqrt{d_1}.$$  

so every point in $V$ is $3\sqrt{d_1}$-close to a point in $\mathcal{N}$.

However, we are not satisfied with $x^{(3)}$ just being a Sparse$(a)$ approximation of $x$; we would like $x^{(3)}$ to inherit some properties of $x$ as well. In fact, we claim

$$\sum_{\frac{1}{b} < |x_i|^2 \leq \frac{3}{b}} |x_i^{(3)}|^2 \geq d_2 - 5\sqrt{d_1}. \quad (2)$$

To see this, first note

$$\sum_{\frac{1}{b} < |x_i|^2 \leq \frac{3}{b}} |x_i^{(3)}|^2 \geq d_2 \implies \sum_{\frac{1}{b} < |x_i^{(1)}|^2 \leq \frac{1}{a}} |x_i^{(1)}|^2 \geq d_2 \implies \sum_{\frac{1}{b} < |x_i^{(2)}|^2 \leq \frac{1}{a - \sqrt{2}b}} |x_i^{(2)}|^2 \geq d_2 \implies \sum_{\frac{1}{b} < |x_i^{(3)}|^2 \leq \frac{3}{b}} |x_i^{(3)}|^2 \geq d_2.$$

Define $LB, LA, NL \subset \{1, \ldots, n\}$ to be the sets of indices

- $i \in LB \implies \frac{1}{a} < |x_i^{(2)}|^2 \leq \frac{3}{b'}$  \quad $|x_i^{(3)}|^2 \leq \frac{1}{2a}$
- $i \in LA \implies \frac{1}{a} < |x_i^{(2)}|^2 \leq \frac{3}{b'}$  \quad $|x_i^{(3)}|^2 > \frac{4}{b}$
- $i \in NL \implies \frac{1}{a} < |x_i^{(2)}|^2 \leq \frac{3}{b'}$  \quad $\frac{1}{2a} < |x_i^{(3)}|^2 \leq \frac{4}{b}$.  

(Think of $LB$ standing for “loss from below,” $LA$ standing for “loss from above,” and $NL$ standing for “no loss.”) Observe we have

\[
d_1 \geq \sum_{i \in LB} \left( |x_i^{(2)}| - |x_i^{(3)}| \right)^2 \\
\geq \sum_{i \in LB} \left( |x_i^{(2)}| - \frac{|x_i^{(2)}|}{\sqrt{2}} \right)^2 \\
\geq \left( 1 - \frac{1}{\sqrt{2}} \right)^2 \sum_{i \in LB} |x_i^{(2)}|^2
\]
and

\[ d_1 \geq \sum_{i \in \mathcal{L}} (|x_i^{(3)}| - |x_i^{(2)}|)^2 \]
\[ \geq \sum_{i \in \mathcal{L}} \left( \sqrt{\frac{4}{5}} - \sqrt{\frac{3}{5}} \right)^2 \]
\[ \geq \left( \sqrt{\frac{4}{3}} - 1 \right)^2 \sum_{i \in \mathcal{L}} \frac{3}{b} \]
\[ \geq \left( \sqrt{\frac{4}{3}} - 1 \right)^2 \sum_{i \in \mathcal{L}} |x_i^{(2)}|^2. \]

Thus we obtain

\[ \sum_{i \in \mathcal{L}} |x_i^{(2)}|^2 \leq 12d_1 + 42d_1 = 54d_1. \] (3)

Furthermore, we have

\[ \sum_{i \in \mathcal{N}_L} |x_i^{(2)}|^2 - \sum_{i \in \mathcal{N}_L} |x_i^{(3)}|^2 = \sum_{i \in \mathcal{N}_L} |x_i^{(2)}|^2 - |x_i^{(3)}|^2 \]
\[ \leq \sum_{i \in \mathcal{N}_L} |x_i^{(2)} - x_i^{(3)}|^2 + 2|x_i^{(2)}| \cdot |x_i^{(2)} - x_i^{(3)}| \]
\[ \leq \|x^{(2)} - x^{(3)}\|_2^2 + 2\|x^{(3)}\|_2 \|x^{(2)} - x^{(3)}\|_2 \]
\[ \leq d_1 + 2\sqrt{d_1}. \] (4)

From (3) and (4) we get

\[ \sum_{\frac{1}{3} < \sqrt{d_1}^2 \leq \frac{1}{2}} |x_i^{(2)}|^2 - \sum_{\frac{1}{3} < \sqrt{d_1}^2 \leq \frac{1}{2}} |x_i^{(3)}|^2 \leq 54d_1 + d_1 + 2\sqrt{d_1} \leq 57\sqrt{d_1}, \]

proving (2).

Consequently, every point in \( V \) is \( 3\sqrt{d_1} \)-close to a point in

\[ \Sigma_V := \mathcal{N} \cap \left\{ x \in \mathbb{S}^{n-1} : \sum_{\frac{1}{3} < \sqrt{|x|^2} \leq \frac{1}{2}} |x_i|^2 \geq d_2 - 57\sqrt{d_1} \right\}. \]

It is also well-known that

\[ |\mathcal{N}| < \left( \frac{3}{\sqrt{d_1}} \right)^{2a} \left( \frac{n}{a} \right) < \left( \frac{3}{\sqrt{d_1}} \right)^{2a} \left( \frac{ne}{a} \right)^{a}. \]

(For a proof, see Lemma 5.2 in [Ver12]). Hence we have shown
Lemma 4.1. Let $V$ be defined as above with $0 < 57 \sqrt{d_1} < d_2 < 1$. Then every point in $V$ is $3 \sqrt{d_1}$-close to a point in $\Sigma_V$, where

$$|\Sigma_V| < \left( \frac{3}{\sqrt{d_1}} \right)^{2a} \left( \frac{ne}{a} \right)^a$$

and

$$x \in \Sigma_V \implies \left\{ \begin{array}{l} x \in \text{Sparse}(a) \\ \sum_{\frac{1}{30} < |x_i|^2 \leq \frac{1}{3}} |x_i|^2 \geq d_2 - 57 \sqrt{d_1} \end{array} \right.$$

Sometimes, instead of studying $\|Ax\|$ for $x \in \Sigma_V$ directly, we would like to first zero out some entries of $x$. Hence we will need:

Lemma 4.2. Suppose $x, y \in \mathbb{C}^n$ such that $|x_i| \geq |y_i|$ for all $1 \leq i \leq n$. Then

$$P[\|Ax\|_2 \leq t] \leq P[\|Ay\|_2 \leq t], \quad t \geq 0.$$

Proof. The key is to exploit the fact that the nonzero entries of $A$ are drawn from the Gaussian distribution. Note we can rewrite

$$P[\|Ax\|_2 \leq t] = P[\|Ax\|_2^2 \leq t^2] = P\left[ \sum_{i=1}^{n} |R_i \cdot x|^2 \leq t^2 \right],$$

where $\sum |R_i \cdot x|^2$ is a sum of independent random variables.

Let us first consider $|R_1 \cdot x|$. If we condition on the values of $\delta_{11}, \ldots, \delta_{1n}$, then

$$|R_1 \cdot x| \sim \mathcal{N}_C \left( 0, \sum_{i=1}^{n} \delta_{1i} |x_i|^2 \right), \quad |R_1 \cdot y| \sim \mathcal{N}_C \left( 0, \sum_{i=1}^{n} \delta_{1i} |y_i|^2 \right).$$

As

$$\sum_{i=1}^{n} \delta_{1i} |x_i|^2 \geq \sum_{i=1}^{n} \delta_{1i} |y_i|^2,$$

we conclude

$$P \left[ |R_1 \cdot x| \leq t' \left| \delta_{1i}, \ldots, \delta_{1n} \right. \right] \leq P \left[ |R_1 \cdot y| \leq t' \left| \delta_{1i}, \ldots, \delta_{1n} \right. \right], \quad t' \geq 0,$$

which implies

$$P[|R_1 \cdot x| \leq t'] \leq P[|R_1 \cdot y| \leq t'], \quad t' \geq 0.$$

Hence

$$P \left[ \sum_{i=1}^{n} |R_i \cdot x|^2 \leq t^2 \right] \leq P \left[ |R_1 \cdot y|^2 + \sum_{i=2}^{n} |R_i \cdot x|^2 \leq t^2 \right].$$

The lemma follows by repeating this argument for $i = 2, \ldots, n$. \hfill \Box

We are now ready to tackle Theorem 2.1.
5 The Highly Compressible Case

In this section, we would like to show

**Theorem 5.1.** There exists a constant $c > 0$ such that for sufficiently large $n$ we have

$$
\mathbb{P} \left[ \inf_{x \in S_{nc}} \|Ax\|_2 \leq \varepsilon \text{ and } \|A\|_{op} \leq Kn^{\frac{\gamma}{2}} \right] \leq e^{-cn^\delta}, \quad 0 \leq \varepsilon \leq 1.
$$

Let $R_i$ denote the $i$-th row of $A$. When $x$ is highly compressible, most of the mass of $x$ is concentrated on $O(n^{1-\delta})$ coordinates. Consequently, as each entry of $A$ is equal to 0 with probability $n^{\delta-1}$, it is difficult to obtain a good lower tail bound on $\|Ax\|_2^2 = \sum_i |R_i \cdot x|^2$

using standard arguments like Paley-Zygmund or Chernoff bounds. Instead, we employ an alternative strategy by exploiting the fact that there is very little cancellation in the sum

$$R_i \cdot x = \sum_j A_{ij} x_j.$$

To do this, we introduce the following lemma.

**Lemma 5.2 (Row Bound).** For $y \in \mathbb{C}^n$ and $J \subset \{1, \ldots, n\}$, define $I_y(J)$ to be the set

$$\{i \in [1, n] : |A_{i,j^*}| \geq 1 \text{ for some } j^* \in J, A_{i,j} = 0 \text{ for } j \in \text{supp}(y) \setminus j^* \}.$$

Then

$$\mathbb{P} \left[ |I_y(J)| \leq \frac{3}{20} |J| n^{\delta} (1 - n^{\delta-1})^{m-1} \right] \leq \exp \left( -\frac{3}{80} |J| n^{\delta} (1 - n^{\delta-1})^{m-1} \right),$$

where $m = |\text{supp}(y)|$.

**Proof.** Denote $c_y = \mathbb{P}[|X| \geq 1], \quad X \sim \mathcal{N}(0, 1)$

and observe

$$\mathbb{P}[i \in I_y(J)] = c_y |J| n^{\delta-1} (1 - n^{\delta-1})^{m-1} > \frac{3}{10} |J| n^{\delta-1} (1 - n^{\delta-1})^{m-1}.$$

The result then follows by applying the multiplicative Chernoff bound. \qed

**Remark.** Notice that Lemma 5.2 is strong for small $m$, i.e. when $y$ is sparse.

As an illustration of our argument, let us suppose $x \in S^{n-1}$ is a $O(n^{1-\delta})$-sparse vector. Without loss of generality suppose the entries of $x$ are in non-increasing order of magnitude. Let us denote

$$x_{[a:b]} = [0, \ldots, 0, x_a, x_{a+1}, \ldots, x_b, 0, \ldots, 0]^T.$$
Then for any step size $s$, we note
\[ x = \sum_{i=0}^{n-1} x_{[is+1,(i+1)s]}. \]

Consequently, we have
\[ \|Ax\|_2^2 \geq \sum_{i=0}^{n-1} |I_x([is+1, (i+1)s])| \|x_{(i+1)s}\|^2. \]

The sparsity of $x$ implies that
\[ \left(1 - n^{\delta-1}\right)^{|\text{supp}(x)|-1} \]

is bounded below by a constant, so from Lemma 5.2 followed by a union bound we conclude that with probability
\[ \left( \frac{n}{s} \right) e^{-O(n^\delta s)} \]
we have
\[ \|Ax\|_2^2 \leq C n^\delta \sum_{i=0}^{n-1} s|x_{(i+1)s}|^2 \]
for some constant $C > 0$. Pictorially, the sum $\sum_{i=0}^{n-1} s|x_{(i+1)s}|^2$ can be drawn as follows:

Notice this sum is maximized when the step size $s = 1$; however, the resulting probability
\[ n e^{-O(n^\delta)} \]
is not good enough to overcome the entropy of the net of $O(n^{1-\delta})$-sparse vectors. Consequently, we will disjointize $S_{H^C}$ into many subsets, using smaller values of $s$ when we can afford to and using larger values when we cannot. In the following subsections, we will make this argument rigorous.
Remark. In our illustration above, we have implicitly assumed that $n$ is divisible by the step size $s$, and so $\frac{n}{s}$ is an integer. This need not be the case; if $n$ is not divisible by $s$, we will have to consider $\lfloor \frac{n}{s} \rfloor$ instead (along with all the necessary modifications). The argument, however, will be exactly the same. Consequently, for notational convenience, we will carry out all our computations as if $n$ is divisible by $s$ — but this need not be the case.

5.1
In this subsection, we will bound

$$
P\left[ \inf_{x \in V} \|Ax\|_2 \leq \varepsilon \text{ and } \|A\|_{op} \leq Kn^{\frac{1}{2}} \right], \quad 0 \leq \varepsilon \leq 1.
$$

where

$$
V = \left\{ x \in S_{n-1} : \sum_{|x_i|^2 \leq \frac{1}{t}} |x_i|^2 < d_1, \quad \sum_{\frac{1}{t} < |x_i|^2 \leq t} |x_i|^2 \geq d_2 \right\}.
$$

with $a \leq c_2 n^{1-\delta}$ and $d_2/\sqrt{d_1}$ sufficiently large.

From Lemma 4.1, we have

$$
P\left[ \inf_{x \in V} \|Ax\|_2 \leq \varepsilon \text{ and } \|A\|_{op} \leq Kn^{\frac{1}{2}} \right]
\leq P\left[ \inf_{x \in \Sigma_V} \|Ax\|_2 \leq \varepsilon + 3K\sqrt{d_1} n^{\frac{1}{2}} \text{ and } \|A\|_{op} \leq Kn^\frac{1}{2} \right]
\leq \sum_{x \in \Sigma_V} P\left[ \|Ax\|_2 \leq \varepsilon + 3K\sqrt{d_1} n^{\frac{1}{2}} \text{ and } \|A\|_{op} \leq Kn^\frac{1}{2} \right]
\leq \sum_{x \in \Sigma_V} P\left[ \|Ax\|_2 \leq \varepsilon + 3K\sqrt{d_1} n^{\frac{1}{2}} \right],
$$

where

$$
x \in \Sigma_V \implies \sum_{\frac{1}{t} < |x_i|^2 \leq t} |x_i|^2 \geq d_2 - 57\sqrt{d_1}.
$$

As $\varepsilon$ is bounded, for sufficiently large $n$ we have

$$
P\left[ \inf_{x \in V} \|Ax\|_2 \leq \varepsilon \text{ and } \|A\|_{op} \leq Kn^\frac{1}{2} \right] \leq \sum_{x \in \Sigma_V} P\left[ \|Ax\|_2 \leq 4K\sqrt{d_1} n^{\frac{1}{2}} \right].
$$

Fix an arbitrary $x \in \Sigma_V$ and let $y$ denote the vector $x$ but with the entries $|x_i|^2 > \frac{1}{t}$ set to 0. By Lemma 4.2, we have

$$
P\left[ \|Ax\|_2 \leq 4K\sqrt{d_1} n^{\frac{1}{2}} \right] \leq P\left[ \|Ay\|_2 \leq 4K\sqrt{d_1} n^{\frac{1}{2}} \right].
so it suffices to bound the right hand side.

Without loss of generality assume the entries of $y$ are in non-increasing order of magnitude. For sufficiently large $n$, observe Lemma 5.2 implies

$$
\mathbb{P} \left[ |I_y(j)| \leq \frac{1}{10e^{c_2}} |j| n^\delta \right] \leq \exp \left( -\frac{1}{40e^{c_2}} |j| n^\delta \right).
$$

Then with probability

$$
\geq 1 - \frac{n}{s} \exp \left( -\frac{1}{40e^{c_2}} s n^\delta \right)
$$

we have

$$
\|Ay\|_2^2 \geq \frac{1}{10e^{c_2}} s n^\delta \sum_{i=0}^{\delta-1} |y_{(i+1)b}|^2
$$

$$
\geq \frac{1}{10e^{c_2}} n^\delta \|y_{b:n}\|_2^2
$$

$$
\geq \frac{1}{10e^{c_2}} n^\delta \left( d_2 - 57\sqrt{d_1} - \frac{4(s-1)}{b} \right).
$$

Setting

$$
s = 1 + \left\lfloor \frac{(d_2 - 57\sqrt{d_1})b}{8} \right\rfloor
$$

we obtain

$$
\mathbb{P} \left[ \|Ay\|_2^2 \leq \frac{1}{20e^{c_2}} n^\delta (d_2 - 57\sqrt{d_1}) \right] \leq n \exp \left( -\frac{1}{40e^{c_2}} \left( 1 + \left\lfloor \frac{(d_2 - 57\sqrt{d_1})b}{8} \right\rfloor \right) n^\delta \right)
$$

for sufficiently large $n$. Thus we get

**Lemma 5.3.** Let $V$ be defined as before with $a \leq c_2 n^{1-\delta}$ and

$$
0 < 57\sqrt{d_1} + 320e^{c_2} K^2 d_1 < d_2 < 1.
$$

Then for $0 \leq \varepsilon \leq 1$ and sufficiently large $n$ we have

$$
\mathbb{P} \left[ \inf_{x \in V} \|Ax\|_2 \leq \varepsilon \text{ and } \|A\|_{op} \leq Kn^{\frac{\delta}{2}} \right]
$$

$$
\leq n \exp \left( -\frac{1}{40e^{c_2}} \left( 1 + \left\lfloor \frac{(d_2 - 57\sqrt{d_1})b}{8} \right\rfloor \right) n^\delta \right) \left( \frac{3}{\sqrt{d_1}} \right)^{2a} \left( \frac{ne}{a} \right)^a.
$$

Lemma 5.3 will allow us to prove Theorem 5.1.
5.2 Proof of Theorem 5.1

Proof. Let \( m \) be the largest integer such that

\[
(m - 1)\frac{\delta}{2} < 1 - \delta.
\]

Write \( S_{HC} \) as the finite union

\[
S_{HC} = \bigcup_{k=1}^{m} S_{HC}^k,
\]

where

\[
S_{HC}^k = \left\{ x \in S^{n-1} : \sum_{|x_i|^2 \leq \frac{1}{n}} |x_i|^2 < d_4^k, \quad \sum_{\frac{1}{n} < |x_i|^2 \leq \frac{1}{n^2}} |x_i|^2 \geq d_2^k \right\}
\]

and \( a_k, b_k, d_{1k}, d_{2k} > 0 \) are finite sequences such that

\[
a_k = n^{\frac{\delta}{2}} \text{ for } k < m, \quad a_m = c_2 n^{1-\delta}, \quad b_k = a_{k-1}, \quad b_1 = 1,
\]

\[
d_{21} + d_{11} = 1, \quad d_{1m} = \varepsilon_1 + \varepsilon_2, \quad d_{1k} = d_{1(k+1)} + d_{2(k+1)},
\]

\[
d_{2k} > 57 \sqrt{d_{1k} + 320 e^{c_2} K^2 d_{1k}}.
\]

(Note the \( d_{1k}, d_{2k} \)'s will only exist if \( \varepsilon_1 + \varepsilon_2 \) is very small, which we can enforce with our choice of \( \varepsilon_1 \) and \( \varepsilon_2 \) in Section 6.) Then by Lemma 5.3, we get

\[
P\left[ \inf_{x \in S_{HC}} \|Ax\|_2 \leq \varepsilon \text{ and } \|A\|_{op} \leq Kn^{\frac{\delta}{2}} \right] \leq \sum_k P\left[ \inf_{x \in S_{HC}} \|Ax\|_2 \leq \varepsilon \text{ and } \|A\|_{op} \leq Kn^{\frac{\delta}{2}} \right] \leq e^{-cn^k}
\]

for some \( c > 0 \) and sufficiently large \( n \), as desired. \[\Box\]

6 The Moderately Compressible Case

In this section, we would like to show

**Theorem 6.1.** There exists a constant \( c > 0 \) such that for sufficiently large \( n \) we have

\[
P\left[ \inf_{x \in S_{MC}} \|Ax\|_2 \leq \varepsilon \text{ and } \|A\|_{op} \leq Kn^{\frac{\delta}{2}} \right] \leq e^{-cn}, \quad 0 \leq \varepsilon \leq 1.
\]
Moderately compressible vectors are generic enough for us to use a more traditional argument.

From Lemma 4.1 there exists a net $\Sigma_{MC} := \Sigma_{S_{MC}}$ such that every point in $S_{MC}$ is $3\sqrt{\varepsilon_1}$-close to a point in $\Sigma_{MC}$ and

$$x \in \Sigma_{MC} \implies \sum_{1 \leq |i| \leq \varepsilon_2 - \delta_1 \sqrt{\varepsilon_1}} x_i^2 \geq \varepsilon_2 - 5\sqrt{\varepsilon_1}.$$ 

Thus we have

$$\mathbb{P} \left[ \inf_{x \in \Sigma_{MC}} \|Ax\|_2 \leq \varepsilon \text{ and } \|A\| \leq Kn^{\frac{1}{2}} \right] \leq \sum_{x \in \Sigma_{MC}} \mathbb{P} \left[ \|Ax\|_2 \leq 4K \sqrt{\varepsilon_1} n^{\frac{1}{2}} \right]$$

$$= \sum_{x \in \Sigma_{MC}} \mathbb{P} \left[ \sum_{i=1}^{n} |R_i \cdot x|^2 \leq (4K \sqrt{\varepsilon_1} n^{\frac{1}{2}})^2 \right], \ R_i \text{ i-th row}$$

$$\leq \sum_{x \in \Sigma_{MC}} \mathbb{P} \left[ |R_i \cdot x|^2 \leq \frac{(4K \sqrt{\varepsilon_1})^2}{n(1-\delta')^2} \text{ for } (1-\delta')n \text{ values of } i \right], \ \delta' > 0$$

$$\leq \sum_{x \in \Sigma_{MC}} \left( \frac{n}{1-\delta'} \right) \mathbb{P} \left[ |R_1 \cdot x|^2 \leq \frac{(4K \sqrt{\varepsilon_1})^2}{n(1-\delta')^2} \right] (1-\delta')n \quad \delta' > 0$$

$$= \sum_{x \in \Sigma_{MC}} \left( \frac{n}{\delta' n} \right) \mathbb{P} \left[ |R_1 \cdot x| \leq \frac{4K \sqrt{\varepsilon_1}}{n^{\frac{1}{2}} \delta'} \right] (1-\delta')n \quad \delta' > 0.$$

To bound this probability, we will need a few intermediate results.

**Lemma 6.2.** Let $X \sim \mathcal{N}(0, \sigma^2)$. Then

$$\mathbb{P}[|X| \leq \varepsilon] \leq \frac{\varepsilon^2}{\sigma^2}.$$ 

**Proof.** Let $X_1 = \Re(X)$ and $X_2 = \Im(X)$. Note $X_1, X_2 \sim \mathcal{N}(0, \frac{\sigma^2}{2})$ and are independent. Furthermore, we have

$$\frac{2}{\sigma^2} |X|^2 = \frac{2}{\sigma^2} X_1^2 + \frac{2}{\sigma^2} X_2^2,$$

so $\frac{2}{\sigma^2} |X|^2$ is the $\chi^2$-distribution with 2 degrees of freedom. Consequently, we
have
\[
P \left( \frac{2}{a^2} |X|^2 \leq t \right) = 1 - e^{-\frac{t}{2}} \\
= e^{-\frac{t}{2}} (e^{\frac{t}{2}} - 1) \\
= e^{-\frac{t}{2}} \left( \sum_{i=1}^{\infty} \left( \frac{t}{2} \right)^i i! \right) \\
= e^{-\frac{t}{2}} \left( \sum_{i=1}^{\infty} \frac{1}{i} \left( \frac{t}{2} \right)^{i-1} \frac{1}{(i-1)!} \right) \\
\leq e^{-\frac{t}{2}} \frac{t}{2} e^{\frac{t}{2}} \\
= \frac{t}{2}. 
\]
Thus we obtain
\[
P(|X| \leq \varepsilon) = P(|X|^2 \leq \varepsilon^2) = P \left( \frac{2}{a^2} |X|^2 \leq \frac{2\varepsilon^2}{a^2} \right) \leq \frac{\varepsilon^2}{a^2}.
\]

Lemma 6.3. Suppose
\[
\sum_{i=1}^{n} a_i = a, \quad a_i \geq 0.
\]
Let the $b_i$'s be independent random variables such that $b_i \overset{d}{=} a_i \delta_{n^{\delta-1}}$. Then
\[
P \left( \sum_{i=1}^{n} b_i \leq n^{\delta-1} ta \right) \leq 1 - (1 - t)^2 \frac{a}{a + (n^{1-\delta} - 1) \max a_i}
\]
for $0 \leq t \leq 1$.

Proof. From the Paley-Zygmund inequality, we get
\[
P \left( \sum_{i=1}^{n} b_i \leq n^{\delta-1} ta \right) \leq 1 - (1 - t)^2 \frac{n^{2(\delta-1)}a^2}{n^{2(\delta-1)}a^2 + \text{Var} \left( \sum b_i \right)} \\
\leq 1 - (1 - t)^2 \frac{n^{2(\delta-1)}a^2}{n^{2(\delta-1)}a^2 + \sum \left( n^{\delta-1}a_j^2 - n^{2(\delta-1)}a_j^2 \right)} \\
\leq 1 - (1 - t)^2 \frac{a^2}{a^2 + (n^{1-\delta} - 1) \sum a_j^2}
\]
from which the result follows. \hfill \Box
Lemma 6.4. For any $x \in \mathbb{C}^n$, we have

$$P \left[ |R_1 \cdot x| \leq \varepsilon \right] \leq 1 - (1 - t)^2 \frac{\|x\|_2^2}{\|x\|_2^2 + (n^{1-\delta} - 1) \sup |x_i|^2 + \varepsilon^2 t n^{1-\delta} n^2}$$

for any $0 < t < 1$.

Proof. Define the random variable $\sigma^2 = \sum_{i: R_i \neq 0} |x_i|^2$. By Lemma 6.3, we note that

$$P \left[ \sigma^2 \leq n^{1-\delta} t \|x\|_2^2 \right] \leq 1 - (1 - t)^2 \frac{\|x\|_2^2}{\|x\|_2^2 + (n^{1-\delta} - 1) \sup |x_i|^2}$$

Now assume that $\sigma^2 > n^{1-\delta} t \|x\|_2^2$. Conditioning on the value of $\sigma^2$, we note that $R_1 \cdot x \sim N_C(0, \sigma^2)$. Thus Lemma 6.2 implies

$$P[|R_1 \cdot x| \leq \varepsilon] \leq \frac{\varepsilon^2}{\sigma^2} \leq \frac{\varepsilon^2}{n^{1-\delta} t \|x\|_2^2}$$

Lemma 6.4 then follows from the union bound. \(\square\)

Let us fix an arbitrary $x \in \Sigma_{MC}$ and let $y$ denote the vector $x$ with all entries $|x_i|^2 \geq \frac{4}{c_2 n^{1-\delta}}$ set to 0. Applying Lemma 4.2 and Lemma 6.4 then allows us to conclude that

$$P \left[ |R_1 \cdot x| \leq \frac{4K \sqrt{\varepsilon_1}}{n^{1-\delta} \sqrt{\delta'}} \right] \leq P \left[ |R_1 \cdot y| \leq \frac{4K \sqrt{\varepsilon_1}}{n^{1-\delta} \sqrt{\delta'}} \right]$$

$$\leq 1 - (1 - t)^2 \frac{\|y\|_2^2}{1 + (n^{1-\delta} - 1) \sup |y_i|^2 + \frac{16K^2 \varepsilon_1}{t \|y\|_2^2 \delta'}}$$

$$\leq 1 - (1 - t)^2 \frac{\varepsilon_2 - 57 \sqrt{\varepsilon_1}}{1 + \frac{4}{c_2} + \frac{4}{c_2 n^{1-\delta}} + \frac{16K^2 \varepsilon_1}{t (\varepsilon_2 - 57 \sqrt{\varepsilon_1}) \delta'}}$$

As

$$|\Sigma_{MC}| \leq \left( \frac{3}{\sqrt{\varepsilon_1}} \right)^{2c_1 n} \left( \frac{\varepsilon}{c_1} \right)^{c_1 n} = \left( \frac{3}{\sqrt{\varepsilon_1}} \right)^{2c_1 n} \left( \exp(c_1 (1 - \ln c_1)) \right)^n$$

and

$$\left( \frac{n}{\delta' n} \right) \leq \left( \frac{\varepsilon}{\delta'} \right)^{\delta' n} = \left( \exp(\delta'(1 - \ln \delta')) \right)^n,$$

to prove Theorem 6.1 it suffices to choose appropriate values for $t, \varepsilon_1, \varepsilon_2, c_1, c_2$ and $\delta'$ such that

$$\left( \frac{3}{\sqrt{\varepsilon_1}} \right)^{2c_1 n} \left( \exp(c_1 (1 - \ln c_1)) \right)^n \left( \exp(\delta'(1 - \ln \delta')) \right)^n \cdot \left( 1 - (1 - t)^2 \frac{\varepsilon_2 - 57 \sqrt{\varepsilon_1}}{1 + \frac{4}{c_2} + \frac{4}{c_2 n^{1-\delta}} + \frac{16K^2 \varepsilon_1}{t (\varepsilon_2 - 57 \sqrt{\varepsilon_1}) \delta'}} \right)^{(1-\delta') n} \leq e^{-cn}$$

16
for some $c > 0$ and sufficiently large $n$.

We will make our choices in the following (chronological) manner:

1. Pick $c_2 > 0$ and pick $t \in (0, 1)$ such that $(1 - t)^2 = 0.5$.
2. Pick $\varepsilon_2 > 0$ small enough such that if $\varepsilon_1$ is chosen to be $< \varepsilon_2$, then $\varepsilon_1 + \varepsilon_2$ is small enough for the proof of Theorem 5.1 to be valid.
3. Pick $\delta' > 0$ small enough such that
   \[
   \exp\left(\frac{\delta'(1 - \ln \delta')}{1 - \delta'}\right) \cdot \left(1 - 0.4 \frac{\varepsilon_2}{1 + \frac{4}{c_2 - \frac{4}{c_2 n^{1-\delta}}}}\right) < 1 - 0.3 \frac{\varepsilon_2}{1 + \frac{4}{c_2 - \frac{4}{c_2 n^{1-\delta}}}}.
   \]
   
   Note this is possible as $\lim_{\delta' \to 0} \delta'(1 - \ln \delta') = 0$.
4. Pick $\varepsilon_1 > 0$ small enough such that
   \[
   1 - 0.5 \frac{\varepsilon_2 - 57\sqrt{\varepsilon_1}}{1 + \frac{4}{c_2 - \frac{4}{c_2 n^{1-\delta}}}} + \frac{16K^2 \varepsilon_1}{t(\varepsilon_2 - 57\sqrt{\varepsilon_1})\delta'} < 1 - 0.4 \frac{\varepsilon_2}{1 + \frac{4}{c_2 - \frac{4}{c_2 n^{1-\delta}}}}.
   \]
5. Finally, pick $c_1 > 0$ small enough such that
   \[
   \left(\frac{3}{\sqrt{\varepsilon_1}}\right)^{2c_1 n} \left(1 - 0.3 \frac{\varepsilon_2}{1 + \frac{4}{c_2 - \frac{4}{c_2 n^{1-\delta}}}}\right)^{(1-\delta')n} < \left(1 - 0.2 \frac{\varepsilon_2}{1 + \frac{4}{c_2 - \frac{4}{c_2 n^{1-\delta}}}}\right)^{(1-\delta')n}
   \]
   and
   \[
   \exp\left(\frac{c_1 (1 - \ln c_1)}{1 - \delta'}\right) \cdot \left(1 - 0.2 \frac{\varepsilon_2}{1 + \frac{4}{c_2 - \frac{4}{c_2 n^{1-\delta}}}}\right) < 1 - 0.1 \frac{\varepsilon_2}{1 + \frac{4}{c_2 - \frac{4}{c_2 n^{1-\delta}}}}.
   \]

This proves Theorem 6.1.

## 7 Incompressible Vectors

Finally, we turn our attention to incompressible vectors. Let $Y_1$ denote the first column of $A$.

**Theorem 7.1.** There exists a constant $c > 0$ such that for sufficiently large $n$ and any $\eta \in \mathcal{S}_{IC}$ we have

\[
\mathbb{P}\left(|\langle Y_1, \eta \rangle| < t \right) \leq e^{-cn^c} + O(t^2 n^{1-\delta}).
\]

**Proof.** Notice there exists at least $\lambda_0 n$ values of $i$ such that

\[
|\eta_i|^2 \geq \frac{1}{\lambda_1 n}
\]
for some constants $\lambda_0, \lambda_1 > 0$ depending on $c_1, \varepsilon_1$. Let $S$ denote the set of such indices $i$. From the multiplicative Chernoff bound, we obtain

$$\Pr \left[ |\{i \in S : Y_i \neq 0\}| \leq \frac{1}{2} \lambda_0 n^\delta \right] \leq \exp \left( - \frac{1}{8} \lambda_0 n^\delta \right).$$

Let us now condition on the event that

$$|\{i \in S : Y_i \neq 0\}| > \frac{1}{2} \lambda_0 n^\delta.$$

Then by Lemma 6.2 we conclude

$$\Pr \left[ |\langle Y_i, \eta \rangle| < t \right] \leq \frac{2\lambda_1}{\lambda_0} t^2 n^{1-\delta}.$$

Theorem 7.1 then follows by a union bound.

\[\square\]

\section{8 Proof of Theorem 2.1}

We are now ready to give a proof of Theorem 2.1.

\textit{Proof.} Note we can rewrite

$$\Pr \left[ \sigma_i(A) \leq \varepsilon \text{ and } \|A\| \leq Kn^{\frac{\delta}{2}} \right] = \Pr \left[ \inf_{x \in S^{n-1}} \|Ax\|_2 \leq \varepsilon \text{ and } \|A\| \leq Kn^{\frac{\delta}{2}} \right] \leq \sum_{i \in \{HC,MC,IC\}} \Pr \left[ \inf_{x \in S^i} \|Ax\|_2 \leq \varepsilon \text{ and } \|A\| \leq Kn^{\frac{\delta}{2}} \right].$$

After applying Theorems 5.1 and 6.1, to prove Theorem 2.1 it suffices to show

$$\Pr \left[ \inf_{x \in S_{IC}} \|Ax\|_2 \leq \varepsilon \text{ and } \|A\| \leq Kn^{\frac{\delta}{2}} \right] \leq e^{-cn^\delta} + O(\varepsilon^2 n^{2-\delta})$$

for some $c > 0$ and sufficiently large $n$. Let $E$ denote the event that there exists $x \in S_{IC}$ such that

$$\|Ax\|_2 \leq \varepsilon, \quad \|A\| \leq Kn^{\frac{\delta}{2}}.$$

Let $Y_i$ denote the $i$-th column of $A$ and let $W_i$ denote the span of the $Y_j$'s for $j \neq i$. Observe that

$$|x_i| \text{ dist}(Y_i, W_i) \leq \varepsilon.$$

As $x \in S_{IC}$, there exists at least $\lambda_0 n$ values of $i$ such that

$$|x_i|^2 \geq \frac{1}{\lambda_1 n}.$$
for some constants $\lambda_0, \lambda_1 > 0$ depending on $\epsilon_1, \epsilon_2$. Notice such condition implies that $\text{dist}(Y_i, W_i) \leq \epsilon \sqrt{\lambda_1 n}$. Let $E_i$ denote the event that both this and $E$ occur, and note that

$$\mathbb{P}(E) \leq \frac{1}{\lambda_0 n} \sum_i \mathbb{P}[E_i] = \frac{1}{\lambda_0} \mathbb{P}[E_1]$$

by double counting.

As $|\langle Y_i \cdot \eta_1 \rangle| \leq \text{dist}(Y_i, W_i)$ for $\eta_1 \perp W_1, \eta_1 \in S^{n-1}$, we conclude that

$$\mathbb{P}(E_1) \leq \mathbb{P}\left[\forall \eta_1 \in S^{n-1} \text{ s.t. } B\eta_1 = 0 \text{ we have } |\langle Y_i \cdot \eta_1 \rangle| \leq \epsilon \sqrt{\lambda_1 n} \right.$$  

and $\|B\|_{op} \leq Kn^{\frac{1}{2}}].$

By applying Theorems 5.1 and 6.1 but with $A$ replaced with $B$, we get

$$\mathbb{P}\left[\inf_{\eta \in S^{n-1} \sim S_{IC}} \|B\eta\|_2 = 0 \text{ and } \|B\|_{op} \leq Kn^{\frac{1}{2}} \right] \leq e^{-cn^b}.$$

for some $c > 0$ and sufficiently large $n$. (The same arguments work for $(n-1)\times n$ matrices.) Therefore, it suffices to bound

$$\mathbb{P}\left[\forall \eta \in S_{IC} \text{ we have } |\langle Y_i \cdot \eta \rangle| \leq \epsilon \sqrt{\lambda_1 n} \right].$$

However, if $\eta \in S_{IC}$, then Theorem 7.1 implies

$$\mathbb{P}\left[|\langle Y_i \cdot \eta \rangle| \leq \epsilon \sqrt{\lambda_1 n} \right] \leq e^{-cn^b} + O(\epsilon^2 n^{2-b})$$

for some $c > 0$ and sufficiently large $n$. Theorem 2.1 follows. □

### 9 Operator Norm Bound

To prove Theorem 1.2, it remains to prove Theorem 2.2. To do this, we will use a result that follows from Theorem 1.7 in [BR17]:

**Lemma 9.1.** Let $A'$ be the $n \times n$ random matrix in Theorem 1.2 but with

$$\xi_{ij} \sim \mathcal{N}_\mathbb{R}(0, 1), \quad \delta_{ij} \overset{d}{=} \delta_{n^{b-1}}.$$

Then there exists constants $K', c' > 0$ such that for sufficiently large $n$ we have

$$\mathbb{P}[\|A'\|_{op} \geq K' n^{\frac{1}{2}}] \leq e^{-cn^b}.$$

Write

$$A = \Re(A) + i\Im(A)$$
where $\Re(A)$ denotes the real part of $A$ and $\Im(A)$ denotes the imaginary part of $A$. Observe $\Re(A), \Im(A)$ are i.i.d. random matrices. Furthermore, from Lemma 9.1 there exists $K', \epsilon > 0$ such that

$$\Pr \left[ \|\Re(A)\|_{op} \geq \frac{K'}{\sqrt{2}} n^{\frac{3}{2}} \right] \leq e^{-\epsilon n^b},$$

$$\Pr \left[ \|\Im(A)\|_{op} \geq \frac{K'}{\sqrt{2}} n^{\frac{3}{2}} \right] \leq e^{-\epsilon n^b}.$$

As $\|A\|_{op} \leq \|\Re(A)\|_{op} + \|\Im(A)\|_{op}$, the union bound gives us

$$\Pr \left[ \|A\|_{op} \geq \sqrt{2} K' n^{\frac{3}{2}} \right] \leq 2e^{-\epsilon n^b},$$

proving Theorem 2.2 and thus Theorem 1.2.

## 10 Discussion On Sparse Regularization

When a matrix has undesirable properties (such as a small least singular value), it is often useful to add a random regularizing perturbation. That is, given a matrix $A \in \mathbb{C}^{n \times n}$ one often studies $A + \lambda A$, where $A$ is a random matrix with independent entries and $\lambda > 0$ is a small constant. (For an example of this technique in action, see [ST04].)

In many cases, $A$ is chosen from the complex Ginibre ensemble. The original motivation of this paper was the following question:

What happens if the perturbation $A$ is a random sparse matrix?

Theorem 1.2 states that for a certain type of random sparse matrix $A$, we have

$$\Pr \left[ \sigma_n(0 + \lambda A) \leq \epsilon \right] \leq e^{-\epsilon n^b} + O_n \left( \frac{\epsilon^2}{\lambda^2} \right).$$

Intuitively, as $M = 0$ is “as singular as a matrix can be,” one may hope that $M = 0$ is a worst-case scenario and be able to conclude that

$$\Pr \left[ \sigma_n(M + \lambda A) \leq \epsilon \right] \leq e^{-\epsilon n^b} + O_n \left( \frac{\epsilon^2}{\lambda^2} \right), \quad \forall M \in \mathbb{C}^{n \times n}.$$

Indeed, such shift-independence is true for $A$ drawn from the Ginibre ensemble. However, in this section, we will show that Theorem 1.2 is not shift-independent.

In fact, we prove a stronger statement:

**Theorem 10.1.** Suppose $A$ is a $n \times n$ random matrix with independent entries where each entry has variance $< K$. If $\Pr(a_{ij} = 0) = P > 0$ for any $a_{ij}$, then there exists a matrix $M$ with arbitrarily large $t = \|M\|_{op}$ such that

$$\Pr \left[ \sigma_n(M + \lambda A) \leq \frac{C \lambda^2 n^{\frac{3}{2}}}{t} \right] \geq 0.99 P$$
where \( C > 0 \) depends only on \( P \) and \( K \).

**Remark.** In Theorem 10.1, we do not require \( A \) to have identically distributed entries. Also the power of \( n \) has not been optimized.

Observe Theorem 10.1 applies if *any* entry of \( A \) has a nonzero chance of being 0. This includes having a random matrix with i.i.d. \( N(0, 1) \) entries, but with just one entry randomly set to 0.

**Proof.** Recall that

\[
\sigma_n(M + \lambda A) = \inf_{x \in S^{n-1}} \| (M + \lambda A)x \|_2
\]

where \( S^{n-1} \subset \mathbb{C}^n \) denotes the unit sphere. For any deterministic unitary matrices \( U \) and \( V \), observe that

\[
\inf_{x \in S^{n-1}} \| (M + \lambda A)x \|_2 = \inf_{x \in S^{n-1}} \| V^*(M + \lambda A)Ux \|_2
\]

\[
= \inf_{x \in S^{n-1}} \| V^*MU + \lambda V^*AUx \|_2.
\]

By an appropriate choice of \( U \) and \( V \) to permute the rows and columns of \( A \), we can set

\[(V^*AU)_{nn} = a_{ij}.
\]

Consequently, it suffices to prove Theorem 1.2 with \( i = j = n \).

Let us define

\[
M = \begin{pmatrix}
t & 0 & 0 & \ldots & 0 & 0 \\
0 & t & 0 & \ldots & 0 & 0 \\
0 & 0 & t & \ldots & 0 & 0 \\
0 & 0 & 0 & \ddots & t & 0 \\
0 & 0 & 0 & \ldots & 0 & 0
\end{pmatrix}.
\]

In other words, \( M = t \cdot \text{Id} \) but with the lower right corner entry zeroed out. Let us also define

\[
x = \begin{pmatrix}
-t \frac{a_{1n}}{t} \\
-t \frac{a_{2n}}{t} \\
\vdots \\
-t \frac{a_{(n-1)n}}{t} \\
1
\end{pmatrix}.
\]

Observe now that

\[
Mx = \begin{pmatrix}
t & 0 & 0 & \ldots & 0 & 0 \\
0 & t & 0 & \ldots & 0 & 0 \\
0 & 0 & t & \ldots & 0 & 0 \\
0 & 0 & 0 & \ddots & t & 0 \\
0 & 0 & 0 & \ldots & 0 & 0
\end{pmatrix} \begin{pmatrix}
-t \frac{a_{1n}}{t} \\
-t \frac{a_{2n}}{t} \\
\vdots \\
-t \frac{a_{(n-1)n}}{t} \\
1
\end{pmatrix} = -\lambda \begin{pmatrix}
a_{1n} \\
a_{2n} \\
\vdots \\
a_{(n-1)n} \\
0
\end{pmatrix}.
\]
In the event that \( a_{nn} = 0 \) we have

\[
\| (M + \lambda A)x \|_2 = \| Mx + \lambda Ax \|_2 = \lambda \left\| \begin{pmatrix} a_{11} & \ldots & a_{1(n-1)} \\ a_{21} & \ldots & a_{2(n-1)} \\ \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{n(n-1)} \end{pmatrix} \begin{pmatrix} -\frac{\lambda}{2} a_{1n} \\ -\frac{\lambda}{2} a_{2n} \\ \vdots \\ -\frac{\lambda}{2} a_{(n-1)n} \end{pmatrix} \right\|_2 \\
= \frac{\lambda^2}{t} \left\| \begin{pmatrix} a_{11} & \ldots & a_{1(n-1)} \\ a_{21} & \ldots & a_{2(n-1)} \\ \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{n(n-1)} \end{pmatrix} \right\|_{\text{HS}}.
\]

A straightforward application of Markov’s inequality yields

\[
\left\| \begin{pmatrix} a_{11} & \ldots & a_{1(n-1)} \\ a_{21} & \ldots & a_{2(n-1)} \\ \vdots & \ddots & \vdots \\ a_{n1} & \ldots & a_{n(n-1)} \end{pmatrix} \right\|_{\text{op}} \leq C' n,
\]

with probability \( 1 - 0.01 P \) for a sufficiently large \( C' \) depending only on \( K \) and \( P \). Thus with probability \( 0.99 P \) we have

\[
\sigma_n(M + \lambda A) \leq \frac{\| (M + \lambda A)x \|_2}{\| x \|_2} \leq C' n^{\frac{3}{2}} \frac{\lambda^2}{t},
\]

as desired. \( \square \)

Consequently, a correct generalization of the lower tail estimate in Theorem 1.2 to the shifted random matrix \( M + \lambda A \) will depend on \( M \) or some property of \( M \) (perhaps, say, the norm \( \| M \|_{\text{op}} \)). Thus a natural follow-up question is to ask:

How does \( \mathbb{P}[\sigma_n(M + \lambda A) < \varepsilon] \) vary with \( M \)?

11 Acknowledgements

This work was done as part of my senior thesis at UC Berkeley. First and foremost, I would like to thank my wonderful advisor, Professor Nikhil Srivastava, who suggested this problem and whose guidance and encouragement made this all possible. I would also like to thank Professor Mark Rudelson for providing thoughtful feedback on an earlier version of this thesis.
References

[BR17] Anirban Basak and Mark Rudelson. Invertibility of sparse non-hermitian matrices. *Advances in Mathematics*, 320:426–483, 2017.

[Ede88] Alan Edelman. Eigenvalues and condition numbers of random matrices. *SIAM Journal on Matrix Analysis and Applications*, 9(4):543–560, 1988.

[Rud08] Mark Rudelson. Invertibility of random matrices: norm of the inverse. *Ann. of Math*, 168(2):575–600, 2008.

[ST04] Daniel A. Spielman and Shang-Hua Teng. Smoothed analysis of algorithms: Why the simplex algorithm usually takes polynomial time. *J. ACM*, 51(3):385–463, 2004.

[Ver12] Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Yonina C. Eldar and Kutyniok, editors, *Compressed Sensing: Theory and Applications*, page 210–268. Cambridge University Press, 2012.