Singularity of random Bernoulli matrices

Konstantin Tikhomirov
Georgia Institute of Technology, Atlanta GA

Sept 18th, 2020
Problem

Let $X_1, X_2, \ldots, X_n$ be independent random vectors uniformly distributed on vertices of the $n$–dimensional cube $[-1, 1]^n$. What is the probability that the vectors are linearly dependent?
Let $X_1, X_2, \ldots, X_n$ be independent random vectors uniformly distributed on vertices of the $n$–dimensional cube $[-1,1]^n$. What is the probability that the vectors are linearly dependent?

The question can be restated in terms of random matrices. Let $B_n$ be an $n \times n$ random matrix with i.i.d $\pm 1$ entries. What is the probability that the matrix is singular? It is not elementary even to show that this probability goes to zero as $n$ grows to infinity.
Let $X_1, X_2, \ldots, X_n$ be independent random vectors uniformly distributed on vertices of the $n$–dimensional cube $[-1,1]^n$. What is the probability that the vectors are linearly dependent?

The question can be restated in terms of random matrices. Let $B_n$ be an $n \times n$ random matrix with i.i.d $\pm 1$ entries. What is the probability that the matrix is singular? It is not elementary even to show that this probability goes to zero as $n$ grows to infinity.

Let $s_{\text{min}}(B_n) = \inf_{x \in S^{n-1}} \|B_n x\|_2$ — smallest singular value of $B$ (i.e. smallest eigenvalue of positive semidefinite matrix $(B_n B_n^\top)^{1/2}$). Columns/rows of $B_n$ are linearly dependent if and only if $s_{\text{min}}(B_n) = 0$. Then the above question is

$$\mathbb{P}\{s_{\text{min}}(B_n) = 0\} \leq \ ?$$
Let $A$ be an $n \times n$ matrix, $s_{\text{max}}(A) = \|A\| = \sup_{x \in S^{n-1}} \|Ax\|_2$ — the largest singular value of $A$. The condition number

$$\kappa(A) = \frac{s_{\text{max}}(A)}{s_{\text{min}}(A)} = \|A\| \|A^{-1}\|.$$ 

The condition number serves as measure of loss of precision when solving systems of linear equations.
Let $A$ be an $n \times n$ matrix, $s_{\text{max}}(A) = \|A\| = \sup_{x \in S^{n-1}} \|Ax\|_2$ — the largest singular value of $A$. The condition number

$$\kappa(A) = \frac{s_{\text{max}}(A)}{s_{\text{min}}(A)} = \|A\| \|A^{-1}\|.$$

The condition number serves as measure of loss of precision when solving systems of linear equations. Assume we look for solution of a system

$$Ax = b,$$

but the coefficient vector $b$ is given with an error $\delta b$. Thus, we are solving the system

$$Ay = b + \delta b, \text{ where } y = x + \delta x.$$
Let $\mathbf{A}$ be an $n \times n$ matrix, $s_{\text{max}}(\mathbf{A}) = \|\mathbf{A}\| = \sup_{\mathbf{x} \in \mathbb{S}^{n-1}} \|\mathbf{A}\mathbf{x}\|_2$ — the largest singular value of $\mathbf{A}$. The condition number

$$\kappa(\mathbf{A}) = \frac{s_{\text{max}}(\mathbf{A})}{s_{\text{min}}(\mathbf{A})} = \|\mathbf{A}\| \|\mathbf{A}^{-1}\|.$$  

The condition number serves as measure of loss of precision when solving systems of linear equations. Assume we look for solution of a system

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

but the coefficient vector $\mathbf{b}$ is given with an error $\delta\mathbf{b}$. Thus, we are solving the system

$$\mathbf{A}\mathbf{y} = \mathbf{b} + \delta\mathbf{b}, \text{ where } \mathbf{y} = \mathbf{x} + \delta\mathbf{x}.$$  

We clearly have

$$\frac{\|\delta\mathbf{x}\|_2}{\|\mathbf{x}\|_2} = \frac{\|\mathbf{A}^{-1}\delta\mathbf{b}\|_2}{\|\mathbf{A}^{-1}\mathbf{b}\|_2} = \frac{\|\mathbf{A}^{-1}\delta\mathbf{b}\|_2}{\|\delta\mathbf{b}\|_2} \frac{\|\delta\mathbf{b}\|_2}{\|\mathbf{b}\|_2} \frac{\|\mathbf{b}\|_2}{\|\mathbf{A}^{-1}\mathbf{b}\|_2} \leq \kappa(\mathbf{A}) \frac{\|\delta\mathbf{b}\|_2}{\|\mathbf{b}\|_2}.$$
In the 1940-es–1950-es, the condition number of random matrices was studied by von Neumann and Goldstine using numerical simulations. In particular, they conjectured that, for an $n \times n$ random matrix $G_n$ with i.i.d standard normal entries, the condition number $\kappa(G_n) = O(n)$ with probability close to one.
History of the question

In the 1940-es–1950-es, the condition number of random matrices was studied by von Neumann and Goldstine using numerical simulations. In particular, they conjectured that, for an $n \times n$ random matrix $G_n$ with i.i.d standard normal entries, the condition number $\kappa(G_n) = O(n)$ with probability close to one.

The limiting distribution of the condition number, and the smallest singular value, of Gaussian random matrices was only computed on 1980-es by Edelman, using a formula for the joint distribution of its singular values. Edelman proved that

$$
\mathbb{P}\{s_{\min}(G_n) \leq tn^{-1/2}\} = 1 - \exp \left( - \frac{t^2}{2} - t \right) + o(1), \quad t > 0.
$$

So, typically $s_{\min}(G_n)$ is of order $n^{-1/2}$. There are various arguments which show that $s_{\max}(G_n) = (2 + o(1))\sqrt{n}$ with high probability. Thus, the condition number of the standard $n \times n$ Gaussian matrix $G_n$ is typically of order $n$. Corresponding results for non-Gaussian random matrices were obtained much later.
History of the question

In the 1940-es–1950-es, the condition number of random matrices was studied by von Neumann and Goldstine using numerical simulations. In particular, they conjectured that, for an \( n \times n \) random matrix \( G_n \) with i.i.d standard normal entries, the condition number \( \kappa(G_n) = O(n) \) with probability close to one.

The limiting distribution of the condition number, and the smallest singular value, of Gaussian random matrices was only computed on 1980-es by Edelman, using a formula for the joint distribution of its singular values. Edelman proved that

\[
\mathbb{P}\{s_{\min}(G_n) \leq tn^{-1/2}\} = 1 - \exp\left(-t^2/2 - t\right) + o(1), \quad t > 0.
\]

So, typically \( s_{\min}(G_n) \) is of order \( n^{-1/2} \). There are various arguments which show that \( s_{\max}(G_n) = (2 + o(1))\sqrt{n} \) with high probability. Thus, the condition number of the standard \( n \times n \) Gaussian matrix \( G_n \) is typically of order \( n \).
In the 1940-es–1950-es, the condition number of random matrices was studied by von Neumann and Goldstine using numerical simulations. In particular, they conjectured that, for an $n \times n$ random matrix $G_n$ with i.i.d standard normal entries, the condition number $\kappa(G_n) = O(n)$ with probability close to one.

The limiting distribution of the condition number, and the smallest singular value, of Gaussian random matrices was only computed on 1980-es by Edelman, using a formula for the joint distribution of its singular values. Edelman proved that

$$\mathbb{P}\{s_{\min}(G_n) \leq tn^{-1/2}\} = 1 - \exp\left(-\frac{t^2}{2} - t\right) + o(1), \quad t > 0.$$ 

So, typically $s_{\min}(G_n)$ is of order $n^{-1/2}$. There are various arguments which show that $s_{\max}(G_n) = (2 + o(1))\sqrt{n}$ with high probability. Thus, the condition number of the standard $n \times n$ Gaussian matrix $G_n$ is typically of order $n$.

*Corresponding results for non-Gaussian random matrices were obtained much later.*
In 1960-es, Komlós showed that for $n \times n$ random matrix $B_n$ with i.i.d $\pm 1$ entries, 

$$\mathbb{P}\{B_n \text{ is singular}\} = o(1).$$
In 1960-es, Komlós showed that for $n \times n$ random matrix $B_n$ with i.i.d $\pm 1$ entries,

$$\mathbb{P}\{B_n \text{ is singular}\} = o(1).$$

The estimate was greatly improved about 30 years later by Kahn, Komlós and Szemerédi (1995), who showed that

$$\mathbb{P}\{B_n \text{ is singular}\} \leq 0.999^n,$$

i.e. the singularity probability is 	extbf{exponentially small in dimension}. 
The trivial bound

$$\mathbb{P}\{B_n \text{ is singular}\} \geq \mathbb{P}\{\text{Two rows/columns of } B_n \text{ are equal up to a sign}\}$$

implies that

$$\mathbb{P}\{B_n \text{ is singular}\} \geq (1 - o(1))n^22^{1-n}.$$ 

It is natural to expect that equality of two rows or columns is the main contribution to singularity which leads to
The trivial bound

\[ \Pr\{B_n \text{ is singular}\} \geq \Pr\{\text{Two rows/columns of } B_n \text{ are equal up to a sign}\} \]

implies that

\[ \Pr\{B_n \text{ is singular}\} \geq (1 - o(1))n^22^{1-n}. \]

It is natural to expect that equality of two rows or columns is the main contribution to singularity which leads to

**Strong conjecture**

\[ \Pr\{B_n \text{ is singular}\} = (1 + o(1))n^22^{1-n}. \]

**Weak conjecture**

\[ \Pr\{B_n \text{ is singular}\} = \left(\frac{1}{2} + o(1)\right)^n. \]

Both conjectures are folklore and have been restated many times in the literature.
Existing approaches

- The argument of Kahn–Komlós–Szemerédi and its development due to Tao–Vu and Bourgain–Vu–Wood: based on replacing the original distribution of the entries with a “lazy” one (a large weight is assigned to 0). Produced optimal singularity estimates for certain models of random discrete matrices.
Existing approaches

- The argument of Kahn–Komlós–Szemerédi and its development due to Tao–Vu and Bourgain–Vu–Wood: based on replacing the original distribution of the entries with a “lazy” one (a large weight is assigned to 0). Produced optimal singularity estimates for certain models of random discrete matrices.

- Approach based on the Littlewood–Offord theory (Tao–Vu, Rudelson–Vershynin):

  - \( s_{\min}(B_n) = 0 \iff \langle \text{col}_i(B_n), Y_i \rangle = 0 \) for some \( i \leq n \), where \( Y_i \) is a vector orthogonal to \( \text{col}_j(B_n), j \neq i \).
Existing approaches

- The argument of Kahn–Komlós–Szemerédi and its development due to Tao–Vu and Bourgain–Vu–Wood: based on replacing the original distribution of the entries with a “lazy” one (a large weight is assigned to 0). Produced optimal singularity estimates for certain models of random discrete matrices.

- Approach based on the Littlewood–Offord theory (Tao–Vu, Rudelson–Vershynin):
  - $s_{\min}(B_n) = 0 \iff \langle \text{col}_i(B_n), Y_i \rangle = 0$ for some $i \leq n$, where $Y_i$ is a vector orthogonal to $\text{col}_j(B_n), j \neq i$.
  - Upper bounds of the form $\mathbb{P}\{\langle \text{col}_n(B_n), Y_n \rangle \leq \varepsilon\} \leq ?$
Existing approaches

- The argument of Kahn–Komlós–Szemerédi and its development due to Tao–Vu and Bourgain–Vu–Wood: based on replacing the original distribution of the entries with a “lazy” one (a large weight is assigned to 0). Produced optimal singularity estimates for certain models of random discrete matrices.

- Approach based on the Littlewood–Offord theory (Tao–Vu, Rudelson–Vershynin):
  - \( s_{\min}(B_n) = 0 \iff \langle \text{col}_i(B_n), Y_i \rangle = 0 \) for some \( i \leq n \), where \( Y_i \) is a vector orthogonal to \( \text{col}_j(B_n) \), \( j \neq i \).
  - Upper bounds of the form \( \mathbb{P}\{\langle \text{col}_n(B_n), Y_n \rangle \leq \varepsilon \} \leq \) are obtained by conditioning on a typical realization of \( Y_n \) and using the randomness of \( \text{col}_n(B_n) \).
Existing approaches

- The argument of Kahn–Komlós–Szemerédi and its development due to Tao–Vu and Bourgain–Vu–Wood: based on replacing the original distribution of the entries with a “lazy” one (a large weight is assigned to 0). Produced optimal singularity estimates for certain models of random discrete matrices.

- Approach based on the Littlewood–Offord theory (Tao–Vu, Rudelson–Vershynin):
  - \( s_{\min}(B_n) = 0 \iff \langle \text{col}_i(B_n), Y_i \rangle = 0 \) for some \( i \leq n \), where \( Y_i \) is a vector orthogonal to \( \text{col}_j(B_n) \), \( j \neq i \).
  - Upper bounds of the form \( \mathbb{P}\{\langle \text{col}_n(B_n), Y_n \rangle \leq \varepsilon \} \leq ? \) are obtained by conditioning on a typical realization of \( Y_n \) and using the randomness of \( \text{col}_n(B_n) \).
  - Strategy: (certain special conditions on \( Y_n \)) \( \implies \) \( (\mathbb{P}\{\langle \text{col}_n(B_n), Y_n \rangle = 0 \} \) is small).

Classical inequalities of this type are due to Erdős–Littlewood–Offord, Levy–Kolmogorov–Rogozin, Esseen, Kesten...
The original approach of Kahn–Komlós–Szemerédi was later improved by Tao and Vu who showed that

\[ \mathbb{P}\{B_n \text{ is singular}\} \leq \left(\frac{3}{4} + o(1)\right)^n. \]
The original approach of Kahn–Komlós–Szemerédi was later improved by Tao and Vu who showed that

$$\mathbb{P}\{B_n \text{ is singular}\} \leq \left(\frac{3}{4} + o(1)\right)^n.$$  

Further improvement was obtained by Bourgain–Vu–Wood. They showed that

$$\mathbb{P}\{B_n \text{ is singular}\} \leq \left(\frac{1}{\sqrt{2}} + o(1)\right)^n,$$

and also obtained asymptotically optimal estimates for some models of discrete random matrices. In particular, they showed that for \(n \times n\) matrix \(M_n\) with i.i.d three-valued entries taking values \(+1\) and \(-1\) with probability \(1/4\) and zero with probability a half,

$$\mathbb{P}\{M_n \text{ is singular}\} = \left(\frac{1}{2} + o(1)\right)^n.$$
For a random variable $X$, its *Levy concentration function* is defined as
\[
\mathcal{L}(X, t) := \sup_{\lambda \in \mathbb{R}} \mathbb{P}\{|X - \lambda| \leq t\}, \quad t \geq 0.
\]
For a random variable $X$, its *Levy concentration function* is defined as

$$\mathcal{L}(X, t) := \sup_{\lambda \in \mathbb{R}} \mathbb{P}\{ |X - \lambda| \leq t \}, \quad t \geq 0.$$ 

**Erdős–Littlewood–Offord**

Let $b_1, b_2, \ldots, b_n$ be i.i.d ±1 random variables, and let $a_1, a_2, \ldots, a_n$ be non-zero real numbers. Then

$$\mathcal{L}\left( \sum_{i=1}^{n} a_i b_i, \min_{i \leq n} |a_i| \right) \leq \frac{C}{\sqrt{n}},$$

for a universal constant $C > 0$. 

**Esseen**

Let $X$ be a real random variable. Then

$$\mathcal{L}(X, t) \leq C \int_{-1}^{1} |E e^{2\pi i s X / t}| \, ds,$$

where $C > 0$ is a universal constant.
Erdős–Littlewood–Offord’s and Esseen’s inequalities

For a random variable \( X \), its Levy concentration function is defined as

\[
\mathcal{L}(X, t) := \sup_{\lambda \in \mathbb{R}} \mathbb{P}\{|X - \lambda| \leq t\}, \quad t \geq 0.
\]

**Erdős–Littlewood–Offord**

Let \( b_1, b_2, \ldots, b_n \) be i.i.d ±1 random variables, and let \( a_1, a_2, \ldots, a_n \) be non-zero real numbers. Then

\[
\mathcal{L}\left( \sum_{i=1}^{n} a_i b_i, \min_{i \leq n} |a_i| \right) \leq \frac{C}{\sqrt{n}},
\]

for a universal constant \( C > 0 \).

**Esseen**

Let \( X \) be a real random variable. Then

\[
\mathcal{L}(X, t) \leq C \int_{-1}^{1} \left| \mathbb{E} e^{2\pi is X/t} \right| ds,
\]

where \( C > 0 \) is a universal constant.
It is hard to extract quantitative information about the smallest singular value of square random matrices from the theorem of Kahn–Komlós–Szemerédi and its refinements. An important step in this direction was made independently by Tao–Vu and Rudelson.
It is hard to extract quantitative information about the smallest singular value of square random matrices from the theorem of Kahn–Komlós–Szemerédi and its refinements. An important step in this direction was made independently by Tao–Vu and Rudelson.

Tao and Vu (2007) estimated the smallest singular value of discrete random matrices by studying arithmetic structure of “potential” almost null vectors of the matrix. In particular, their result implies that for any $K > 0$ there is $L > 0$ depending only on $K$ such that for all sufficiently large $n$

$$\mathbb{P}\{s_{\text{min}}(B_n) \leq n^{-L}\} \leq n^{-K}.$$
Singularity of $B_n$: quantitative argument of Tao–Vu

It is hard to extract quantitative information about the smallest singular value of square random matrices from the theorem of Kahn–Komlós–Szemerédi and its refinements. An important step in this direction was made independently by Tao–Vu and Rudelson.

Tao and Vu (2007) estimated the smallest singular value of discrete random matrices by studying arithmetic structure of “potential” almost null vectors of the matrix. In particular, their result implies that for any $K > 0$ there is $L > 0$ depending only on $K$ such that for all sufficiently large $n$

$$\mathbb{P}\{s_{\min}(B_n) \leq n^{-L}\} \leq n^{-K}.$$ 

The proof is based on a theorem which asserts that any fixed integer vector $\nu = (\nu_1, \ldots, \nu_n)$ with $\sup_{r \in \mathbb{R}} \mathbb{P}\{\sum_i b_i \nu_i = r\} \geq n^{-R}$, almost all coordinates of $\nu$ are contained in a generalized arithmetic progression with some special properties. This paper advanced development of the Littlewood–Offord theory, dealing with anti-concentration inequalities for random linear combinations under various structural assumptions on the coefficients.
The result of Rudelson–Vershynin (2008) strengthens the theorem of Kahn–Komlós–Szemerédi simultaneously in two directions: by providing strong quantitative estimates on $s_{\text{min}}$ and by taking a broader class of distributions.
The result of Rudelson–Vershynin (2008) strengthens the theorem of Kahn–Komlós–Szemerédi simultaneously in two directions: by providing strong quantitative estimates on $s_{\text{min}}$ and by taking a broader class of distributions.

**Rudelson–Vershynin, 2008**

Let $A_n$ be $n \times n$ random matrix with i.i.d entries of zero mean, unit variance and with bounded subgaussian moment. Then for any $t > 0$

$$\mathbb{P}\{s_{\text{min}}(A_n) \leq t/\sqrt{n}\} \leq Ct + c^n,$$

where $C > 0$ and $c \in (0, 1)$ may only depend on the subgaussian moment.
The result of Rudelson–Vershynin (2008) strengthens the theorem of Kahn–Komlós–Szemerédi simultaneously in two directions: by providing strong quantitative estimates on $s_{\min}$ and by taking a broader class of distributions.

**Rudelson–Vershynin, 2008**

Let $A_n$ be $n \times n$ random matrix with i.i.d entries of zero mean, unit variance and with bounded subgaussian moment. Then for any $t > 0$

$$
\mathbb{P}\{s_{\min}(A_n) \leq t/\sqrt{n}\} \leq Ct + c^n,
$$

where $C > 0$ and $c \in (0, 1)$ may only depend on the subgaussian moment.

The argument of Rudelson and Vershynin is based on three key components:

- Compressible and incompressible vectors;
- Reduction to structural properties of random normals;
- The notion of the *Least Common Denominator*. 
A vector $x \in S^{n-1}$ is called $(\delta, \rho)$–compressible (for some parameters $\delta, \rho \in (0, 1)$ if the Euclidean distance from that vector to the set of $\delta n$–sparse vectors is at most $\rho$. 

A vector $x \in S^{n-1}$ is called $(\delta, \rho)\text{-compressible}$ (for some parameters $\delta, \rho \in (0, 1)$) if the Euclidean distance from that vector to the set of $\delta n$–sparse vectors is at most $\rho$.

For example, the unit vector $(0.5, 0.8, 0.2, 0.2, 0.1, -0.1, -0.1) \in S^6$ is $(4/7, \sqrt{0.03})\text{-compressible}$. 
A vector $x \in S^{n-1}$ is called $(\delta, \rho)$–compressible (for some parameters $\delta, \rho \in (0, 1)$ if the Euclidean distance from that vector to the set of $\delta n$–sparse vectors is at most $\rho$.

For example, the unit vector $(0.5, 0.8, 0.2, 0.2, 0.1, -0.1, -0.1) \in S^6$ is $(4/7, \sqrt{0.03})$–compressible.

The remaining (not $(\delta, \rho)$–compressible) unit vectors are called $(\delta, \rho)$–incompressible.

The set of unit $(\delta, \rho)$–compressible vectors will be denoted by $\text{Comp}_n(\delta, \rho)$, and incompressible — by $\text{Incomp}_n(\delta, \rho)$.
A vector $x \in S^{n-1}$ is called $(\delta, \rho)$–compressible (for some parameters $\delta, \rho \in (0, 1)$ if the Euclidean distance from that vector to the set of $\delta n$–sparse vectors is at most $\rho$.

For example, the unit vector $(0.5, 0.8, 0.2, 0.2, 0.1, -0.1, -0.1) \in S^6$ is $(4/7, \sqrt{0.03})$–compressible.

The remaining (not $(\delta, \rho)$–compressible) unit vectors are called $(\delta, \rho)$–incompressible.

The set of unit $(\delta, \rho)$–compressible vectors will be denoted by $\text{Comp}_n(\delta, \rho)$, and incompressible — by $\text{Incomp}_n(\delta, \rho)$. Thus,

$$S^{n-1} = \text{Comp}_n(\delta, \rho) \sqcup \text{Incomp}_n(\delta, \rho).$$

The compressible and incompressible vectors are handled differently.
When the parameters \( \delta \) and \( \rho \) are small, the set \( \text{Comp}_n(\delta, \rho) \) is small as well (in the sense that the covering numbers can be efficiently bounded). This becomes useful when estimating probability of an event of the type

\[
\{ \|A_n x\|_2 \text{ is small for some } x \in \text{Comp}_n(\delta, \rho) \}.
\]

**Lemma (Litvak–Pajor–Rudelson–Tomczak-Jaegermann, 2005)**

Let \( \mathcal{N} \) be a smallest Euclidean \( \varepsilon \)-net in \( \text{Comp}_n(\delta, \rho) \), i.e. the smallest subset of \( \text{Comp}_n(\delta, \rho) \) such that \( \text{dist}(x, \mathcal{N}) \leq \varepsilon \) for every \( x \in \text{Comp}_n(\delta, \rho) \), where \( \rho \leq \varepsilon / 2 \). Then

\[
|\mathcal{N}| \leq \left( \frac{C}{\delta \varepsilon} \right)^{\delta n}.
\]
When the parameters $\delta$ and $\rho$ are small, the set $\text{Comp}_n(\delta, \rho)$ is small as well (in the sense that the covering numbers can be efficiently bounded). This becomes useful when estimating probability of an event of the type

$$\{\|A_n x\|_2 \text{ is small for some } x \in \text{Comp}_n(\delta, \rho)\}.$$

**Lemma (Litvak–Pajor–Rudelson–Tomczak-Jaegermann, 2005)**

Let $\mathcal{N}$ be a smallest Euclidean $\varepsilon$–net in $\text{Comp}_n(\delta, \rho)$, i.e. the smallest subset of $\text{Comp}_n(\delta, \rho)$ such that $\text{dist}(x, \mathcal{N}) \leq \varepsilon$ for every $x \in \text{Comp}_n(\delta, \rho)$, where $\rho \leq \varepsilon / 2$. Then

$$|\mathcal{N}| \leq \left( \frac{C}{\delta \varepsilon} \right)^{\delta n}.$$

**Proposition (Litvak–Pajor–Rudelson–Tomczak-Jaegermann, 2005)**

Let $A_n$ be an $n \times n$ matrix with i.i.d. centered subgaussian entries of unit variance. Then there are $c, \delta, \rho$ depending only on the subgaussian moment such that

$$\mathbb{P}\{\|A_n x\|_2 \leq c \sqrt{n} \text{ for some } x \in \text{Comp}_n(\delta, \rho)\} \leq 2 \exp(-cn).$$
Proof of the proposition: Let $\mathcal{N}$ be a smallest Euclidean $\varepsilon$–net in the set of compressible vectors. Apply the relation

$$\mathbb{P}\{\|A_nx\|_2 \leq c\sqrt{n} \text{ for some } x \in \text{Comp}_n(\delta, \rho)\}$$

$$\leq |\mathcal{N}| \sup_{y \in \mathcal{N}} \mathbb{P}\{\|A_ny\|_2 \leq c\sqrt{n} + \varepsilon L\sqrt{n}\} + \mathbb{P}\{\|A_n\| \geq L\sqrt{n}\},$$

valid for any $\varepsilon, L > 0$. 
Proof of the proposition: Let $\mathcal{N}$ be a smallest Euclidean $\varepsilon$–net in the set of compressible vectors. Apply the relation

$$\mathbb{P}\{\|A_n x\|_2 \leq c \sqrt{n} \text{ for some } x \in \text{Comp}_n(\delta, \rho)\}$$

$$\leq |\mathcal{N}| \sup_{y \in \mathcal{N}} \mathbb{P}\{\|A_n y\|_2 \leq c \sqrt{n} + \varepsilon L \sqrt{n}\} + \mathbb{P}\{\|A_n\| \geq L \sqrt{n}\},$$

valid for any $\varepsilon, L > 0$.

- The cardinality of $\mathcal{N}$ is then bounded above by $\left(\frac{C}{\delta \varepsilon}\right)^{\delta n}$. 


Proof of the proposition: Let \( \mathcal{N} \) be a smallest Euclidean \( \varepsilon \)-net in the set of compressible vectors. Apply the relation

\[
P\{\|A_n x\|_2 \leq c \sqrt{n} \text{ for some } x \in \text{Comp}_n(\delta, \rho)\} \\
\leq |\mathcal{N}| \sup_{y \in \mathcal{N}} P\{\|A_n y\|_2 \leq c \sqrt{n} + \varepsilon L \sqrt{n}\} + P\{\|A_n\| \geq L \sqrt{n}\},
\]

valid for any \( \varepsilon, L > 0 \).

- The cardinality of \( \mathcal{N} \) is then bounded above by \( \left( \frac{C}{\delta \varepsilon} \right)^{\delta n} \).
- Further, for large enough \( L \), \( P\{\|A_n\| \geq L \sqrt{n}\} \leq e^{-n} \) (subgaussian assumption).
Proof of the proposition: Let $\mathcal{N}$ be a smallest Euclidean $\varepsilon$–net in the set of compressible vectors. Apply the relation

$$\mathbb{P}\{\|A_n x\|_2 \leq c\sqrt{n} \text{ for some } x \in \text{Comp}_n(\delta, \rho)\}$$

$$\leq |\mathcal{N}| \sup_{y \in \mathcal{N}} \mathbb{P}\{\|A_n y\|_2 \leq c\sqrt{n} + \varepsilon L\sqrt{n}\} + \mathbb{P}\{\|A_n\| \geq L\sqrt{n}\},$$

valid for any $\varepsilon, L > 0$.

- The cardinality of $\mathcal{N}$ is then bounded above by $(\frac{C}{\delta \varepsilon})^\delta n$.
- Further, for large enough $L$, $\mathbb{P}\{\|A_n\| \geq L\sqrt{n}\} \leq e^{-n}$ (subgaussian assumption).
- For each unit vector $y$, the probability of $\{\|A_n y\|_2 \leq c\sqrt{n} + \varepsilon L\sqrt{n}\}$ can be bounded by $e^{-c' n}$, assuming that $c$ and $\varepsilon L$ are small enough.
**Proof of the proposition:** Let $\mathcal{N}$ be a smallest Euclidean $\varepsilon$–net in the set of compressible vectors. Apply the relation

$$
\mathbb{P}\left\{ \|A_n x\|_2 \leq c \sqrt{n} \text{ for some } x \in \text{Comp}_n(\delta, \rho) \right\}
\leq |\mathcal{N}| \sup_{y \in \mathcal{N}} \mathbb{P}\left\{ \|A_n y\|_2 \leq c \sqrt{n} + \varepsilon L \sqrt{n} \right\} + \mathbb{P}\left\{ \|A_n\| \geq L \sqrt{n} \right\},
$$

valid for any $\varepsilon, L > 0$.

- The cardinality of $\mathcal{N}$ is then bounded above by $\left( \frac{C}{\delta \varepsilon} \right)^{\delta n}$.
- Further, for large enough $L$, $\mathbb{P}\{\|A_n\| \geq L \sqrt{n}\} \leq e^{-n}$ (subgaussian assumption).
- For each unit vector $y$, the probability of $\{\|A_n y\|_2 \leq c \sqrt{n} + \varepsilon L \sqrt{n}\}$ can be bounded by $e^{-c'n}$, assuming that $c$ and $\varepsilon L$ are small enough.

Thus,

$$
\mathbb{P}\left\{ \|A_n x\|_2 \leq c \sqrt{n} \text{ for some } x \in \text{Comp}_n(\delta, \rho) \right\}
\leq \left( \frac{C}{\delta \varepsilon} \right)^{\delta n} e^{-c'n} + e^{-n} \leq 2e^{-c''n},
$$

for a right choice of the parameters.
\[ \text{Incomp}_n(\delta, \rho) = \{x \in S^{n-1} : \text{dist}(x, \text{The set of } \delta n\text{-sparse vectors}) > \rho \}. \]
Incomp\(_n(\delta, \rho) = \{x \in S^{n-1} : \text{dist}(x, \text{The set of } \delta n\text{-sparse vectors}) > \rho\}.$

The incompressible vectors occupy almost entire sphere. On the other hand, a very useful property of any incompressible vector is that it is flat in the sense that a proportional to \(n\) number of its components are of absolute value \(\approx 1/\sqrt{n}\).
The incompressible vectors occupy almost entire sphere. On the other hand, a very useful property of any incompressible vector is that it is flat in the sense that a proportional to $n$ number of its components are of absolute value $\approx 1/\sqrt{n}$.

An application of Markov's inequality to random indicators gives

**Proposition (Rudelson–Vershynin, 2008)**

Let $A_n$ be an $n \times n$ random matrix. Then

$$\mathbb{P}\{\|A_n x\|_2 \leq t n^{-1/2} \text{ for some } x \in \text{Incomp}_n(\delta, \rho)\}$$

$$\leq \frac{C}{\delta n} \sum_{i=1}^{n} \mathbb{P}\{\text{dist}(\text{col}_i(A_n), H_i(A_n)) \leq t/\rho\},$$

where $H_i(A_n)$ is the random linear subspace spanned by columns of $A_n$ except for the $i$–th.
Summary of the above:

Using that the smallest singular value $s_{\text{min}}(A_n) = \min_{x \in S_n^{-1}} \|A_n x\|_2$, and the partition $S_n^{-1} = \text{Comp}_n(\delta,\rho) \sqcup \text{Incomp}_n(\delta,\rho)$, we get

$$P\{s_{\text{min}}(A_n) \leq t/\sqrt{n}\} \leq P\{\|A_n x\|_2 \leq t/\sqrt{n} \text{ for some } x \in \text{Comp}_n(\delta,\rho)\} + P\{\|A_n x\|_2 \leq t/\sqrt{n} \text{ for some } x \in \text{Incomp}_n(\delta,\rho)\} \leq 2 \exp(-cn) + C\delta P\{\text{dist}(\text{col}_n(A_n), H_n(A_n)) \leq t/\rho\} = 2 \exp(-cn) + C\delta P\{|\langle\text{col}_n(A_n), Y_n\rangle| \leq t/\rho\},$$

where $Y_n$ is a random unit vector orthogonal to the first $n-1$ columns of $A_n$ (and independent from $\text{col}_n(A_n)$). Thus, the problem of estimating $s_{\text{min}}$ is reduced to studying anti-concentration properties of the random variable $\langle\text{col}_n(A_n), Y_n\rangle$. 
Summary of the above:

Using that the smallest singular value $s_{\min}(A_n) = \min_{x \in S^{n-1}} \|A_n x\|_2$, and the partition $S^{n-1} = \text{Comp}_n(\delta, \rho) \sqcup \text{Incomp}_n(\delta, \rho)$, we get

$$\mathbb{P}\{s_{\min}(A_n) \leq t/\sqrt{n}\} \leq \mathbb{P}\{\|A_n x\|_2 \leq t/\sqrt{n} \text{ for some } x \in \text{Comp}_n(\delta, \rho)\}$$

$$+ \mathbb{P}\{\|A_n x\|_2 \leq t/\sqrt{n} \text{ for some } x \in \text{Incomp}_n(\delta, \rho)\}$$

$$\leq 2 \exp(-cn) + \frac{C}{\delta} \mathbb{P}\{\text{dist}(\text{col}_n(A_n), H_n(A_n)) \leq t/\rho\}$$

$$= 2 \exp(-cn) + \frac{C}{\delta} \mathbb{P}\{|\langle \text{col}_n(A_n), Y_n\rangle| \leq t/\rho\},$$

where $Y_n$ is a random unit vector orthogonal to the first $n - 1$ columns of $A_n$ (and independent from $\text{col}_n(A_n)$).
Summary of the above:

Using that the smallest singular value $s_{\min}(A_n) = \min_{x \in S^{n-1}} \|A_n x\|_2$, and the partition $S^{n-1} = \text{Comp}_n(\delta, \rho) \sqcup \text{Incomp}_n(\delta, \rho)$, we get

\[
P\{s_{\min}(A_n) \leq t/\sqrt{n}\} \\
\leq P\{\|A_n x\|_2 \leq t/\sqrt{n} \text{ for some } x \in \text{Comp}_n(\delta, \rho)\} \\
+ P\{\|A_n x\|_2 \leq t/\sqrt{n} \text{ for some } x \in \text{Incomp}_n(\delta, \rho)\} \\
\leq 2 \exp(-cn) + \frac{C}{\delta} P\{\text{dist}(\text{col}_n(A_n), H_n(A_n)) \leq t/\rho\} \\
= 2 \exp(-cn) + \frac{C}{\delta} P\{|\langle \text{col}_n(A_n), Y_n\rangle| \leq t/\rho\},
\]

where $Y_n$ is a random unit vector orthogonal to the first $n-1$ columns of $A_n$ (and independent from $\text{col}_n(A_n)$).

Thus, the problem of estimating $s_{\min}$ is reduced to studying anti-concentration properties of the random variable $\langle \text{col}_n(A_n), Y_n\rangle$. 

The Lévy concentration function of a random variable $\xi$ is defined as

$$
\mathcal{L}(\xi, s) := \sup_{r \in \mathbb{R}} \mathbb{P}\{|\xi - r| \leq s\}, \quad s \geq 0.
$$

In view of the above, the theorem of Rudelson–Vershynin would be implied by the estimate

$$
\mathcal{L}\left(\langle \text{col}_n(A_n), Y_n \rangle, t \right) \leq Ct + Ce^{-c'n}, \quad t > 0.
$$
The Lévy concentration function of a random variable $\xi$ is defined as

$$\mathcal{L}(\xi, s) := \sup_{r \in \mathbb{R}} \mathbb{P}\{|\xi - r| \leq s\}, \quad s \geq 0.$$ 

In view of the above, the theorem of Rudelson–Vershynin would be implied by the estimate

$$\mathcal{L}(\langle \text{col}_n(A_n), Y_n \rangle, t) \leq Ct + Ce^{-c'n}, \quad t > 0.$$ 

Let $\mathcal{T}$ be the subset of $S^{n-1}$ consisting of all unit vectors $y$ such that

$$\mathcal{L}(\langle \text{col}_n(A_n), y \rangle, t) \leq Ct + Ce^{-c'n}, \quad t > 0.$$ 

Then, using the independence of $Y_n$ and $\text{col}_n(A_n)$, the above estimate (and the main result) would follow if

$$\mathbb{P}\{Y_n \in \mathcal{T}\} \geq 1 - 2 \exp(-c'n).$$
The least common denominator of a unit vector $y$ is defined as

$$\text{LCD}(y) := \inf \{ \theta > 0 : \text{dist}(\theta y, \mathbb{Z}^n) < \min(c_1 \|\theta y\|_2, c_2 \sqrt{n}) \},$$

where $c_1, c_2 > 0$ are two small constants. The least common denominator characterizes structural properties of the vector.
The least common denominator of a unit vector $y$ is defined as
\[
\text{LCD}(y) := \inf \{ \theta > 0 : \text{dist}(\theta y, \mathbb{Z}^n) < \min(c_1 \|\theta y\|_2, c_2 \sqrt{n}) \},
\]
where $c_1, c_2 > 0$ are two small constants. The least common denominator characterizes structural properties of the vector. The key element of the Rudelson–Vershynin argument is the relation between the least common denominator and properties of the random sums.

Rudelson–Vershynin, 2008

If $a_1, \ldots, a_n$ are i.i.d random variables of unit variance then
\[
\mathcal{L} \left( \sum_{i=1}^{n} a_i y_i, t \right) \leq \tilde{C} t + \tilde{C} e^{-c'n} + \frac{\tilde{C}}{\text{LCD}(y)} \quad \text{for all } t > 0.
\]
The least common denominator of a unit vector $y$ is defined as

$$\text{LCD}(y) := \inf \{ \theta > 0 : \text{dist}(\theta y, \mathbb{Z}^n) < \min(c_1\|\theta y\|_2, c_2\sqrt{n}) \},$$

where $c_1, c_2 > 0$ are two small constants. The least common denominator characterizes structural properties of the vector. The key element of the Rudelson–Vershynin argument is the relation between the least common denominator and properties of the random sums.

Rudelson–Vershynin, 2008

If $a_1, \ldots, a_n$ are i.i.d random variables of unit variance then

$$\mathcal{L}\left(\sum_{i=1}^{n} a_i y_i, t\right) \leq \tilde{C} t + \tilde{C} e^{-c'n} + \frac{\tilde{C}}{\text{LCD}(y)} \quad \text{for all } t > 0.$$

Recall that the set $\mathcal{T}$ was defined as

$$\mathcal{T} = \{ y \in S^{n-1} : \mathcal{L}(\langle \text{col}_n(A_n), y \rangle, t) \leq Ct + Ce^{-c'n}, \quad t > 0 \}.$$
The least common denominator of a unit vector $y$ is defined as

$$\text{LCD}(y) := \inf \{ \theta > 0 : \text{dist}(\theta y, \mathbb{Z}^n) < \min(c_1\|\theta y\|_2, c_2\sqrt{n}) \},$$

where $c_1, c_2 > 0$ are two small constants. The least common denominator characterizes structural properties of the vector. The key element of the Rudelson–Vershynin argument is the relation between the least common denominator and properties of the random sums.

Rudelson–Vershynin, 2008

If $a_1, \ldots, a_n$ are i.i.d random variables of unit variance then

$$\mathcal{L}\left(\sum_{i=1}^{n} a_i y_i, t\right) \leq \tilde{C} t + \tilde{C} e^{-c'n} + \frac{\tilde{C}}{\text{LCD}(y)} \quad \text{for all } t > 0.$$

Recall that the set $\mathcal{T}$ was defined as

$$\mathcal{T} = \{ y \in S^{n-1} : \mathcal{L}(\langle \text{col}_n(A_n), y \rangle, t) \leq Ct + Ce^{-c'n}, \quad t > 0 \}.$$  

Thus, with the appropriate choice of constants, the above estimate gives

$$\{ y \in S^{n-1} : \text{LCD}(y) \geq e^{c'n} \} \subset \mathcal{T}.$$
Hence, the main theorem follows as long as

\[ \mathbb{P}\{ \text{LCD}(Y_n) \geq e^{c'n} \} \geq 1 - e^{-c''n}. \]

This last bound is proved using a special covering argument:
Hence, the main theorem follows as long as
\[ \mathbb{P}\{ \text{LCD}(Y_n) \geq e^{c'n} \} \geq 1 - e^{-c''n}. \]

This last bound is proved using a special covering argument:

- First, the set of compressible vectors is ruled out:
  \[ \mathbb{P}\{ Y_n \text{ is } (\delta, \rho)\text{-compressible} \} \leq 2e^{-\tilde{c}n}. \]

It can also be shown that every incompressible vector has LCD of order at least \( \sqrt{n} \).
Hence, the main theorem follows as long as
\[ P\{\text{LCD}(Y_n) \geq e^{c'n}\} \geq 1 - e^{-c''n}. \]

This last bound is proved using a special covering argument:

- First, the set of compressible vectors is ruled out:
  \[ P\{Y_n \text{ is } (\delta, \rho)-\text{compressible}\} \leq 2e^{-\tilde{c}n}. \]

It can also be shown that every incompressible vector has LCD of order at least \( \sqrt{n} \).

- The set of incompressible vectors with “subexponential” LCD is partitioned into \( cn \) subsets, each of the form \( S_D := \{y \in \text{Incomp}_n(\delta, \rho) : \text{LCD}(y) \in [D, 2D)\} \), for some \( D \in [\sqrt{n}, e^{cn}] \).
Hence, the main theorem follows as long as
\[
P\{ \text{LCD}(Y_n) \geq e^{c'n} \} \geq 1 - e^{-c''n}.
\]
This last bound is proved using a special covering argument:
- First, the set of compressible vectors is ruled out:
  \[
P\{ Y_n \text{ is } (\delta, \rho) \text{-compressible} \} \leq 2e^{-\tilde{c}n}.
  \]
- It can also be shown that every incompressible vector has LCD of order at least \( \sqrt{n} \).
- The set of incompressible vectors with “subexponential” LCD is partitioned into \( cn \) subsets, each of the form \( S_D := \{ y \in \text{Incomp}_n(\delta, \rho) : \text{LCD}(y) \in [D, 2D) \} \), for some \( D \in [\sqrt{n}, e^{cn}] \). It is then shown, using a “standard” \( \varepsilon \)-net on \( S_D \), that \( P\{ Y_n \in S_D \} \) is exponentially small in dimension \( n \).
Hence, the main theorem follows as long as
\[ \mathbb{P}\{ \text{LCD}(Y_n) \geq e^{c'n} \} \geq 1 - e^{-c''n}. \]

This last bound is proved using a special covering argument:

- First, the set of compressible vectors is ruled out:
  \[ \mathbb{P}\{ Y_n \text{ is } (\delta, \rho)-\text{compressible} \} \leq 2e^{-\tilde{c}n}. \]

It can also be shown that every incompressible vector has LCD of order at least \( \sqrt{n} \).

- The set of incompressible vectors with “subexponential” LCD is partitioned into \( cn \) subsets, each of the form \( S_D := \{ y \in \text{Incomp}_n(\delta, \rho) : \text{LCD}(y) \in [D, 2D) \} \), for some \( D \in [\sqrt{n}, e^{cn}] \). It is then shown, using a “standard” \( \varepsilon \)-net on \( S_D \), that \( \mathbb{P}\{ Y_n \in S_D \} \) is exponentially small in dimension \( n \).

In this argument, the Least Common Denominator acts as a “proxy” in measuring anti-concentration properties of the inner products:

\[ Y_n \in \{ y \in S^{n-1} : \text{LCD}(y) \geq e^{c'n} \} \subset T \text{ with very high probability.} \]
Let $B_n$ be the $n \times n$ Bernoulli random matrix with i.i.d $\pm 1$ entries.

- The result of Rudelson–Vershynin implies the estimate

\[ \mathbb{P}\{s_{\text{min}}(B_n) \leq t/\sqrt{n}\} \leq Ct + c^n, \quad t > 0, \]

for some constants $C > 0, c \in (0, 1)$. In particular, $B_n$ is non-singular with probability at least $1 - c^n$. 

The argument of Kahn–Komlós–Szemerédi and its development by Tao–Vu and Bourgain–Vu–Wood provides stronger singularity probability estimates:

\[ \mathbb{P}\{B_n \text{ is singular}\} \leq (1/\sqrt{2} + o(1))n, \]

although the method does not seem to imply strong small ball probability estimates for $s_{\text{min}}(B_n)$.

The folklore conjecture in the field is that $\mathbb{P}\{B_n \text{ is singular}\} = (1/2 + o(1))n$, that is, considerable contribution to the matrix singularity comes from the event that two rows or columns of the matrix are equal.
Summary of earlier results

Let $B_n$ be the $n \times n$ Bernoulli random matrix with i.i.d $\pm 1$ entries.

- The result of Rudelson–Vershynin implies the estimate

$$\mathbb{P}\{s_{\text{min}}(B_n) \leq t/\sqrt{n}\} \leq Ct + c^n, \quad t > 0,$$

for some constants $C > 0$, $c \in (0, 1)$. In particular, $B_n$ is non-singular with probability at least $1 - c^n$.

- The argument of Kahn–Komlós–Szemerédi and its development by Tao–Vu and Bourgain–Vu–Wood provides stronger singularity probability estimates:

$$\mathbb{P}\{B_n \text{ is singular}\} \leq \left(\frac{1}{\sqrt{2}} + o(1)\right)^n,$$

although the method does not seem to imply strong small ball probability estimates for $s_{\text{min}}(B_n)$.
Let $B_n$ be the $n \times n$ Bernoulli random matrix with i.i.d $\pm 1$ entries.

- The result of Rudelson–Vershynin implies the estimate
  \[ \Pr\{s_{\min}(B_n) \leq t/\sqrt{n}\} \leq Ct + c^n, \quad t > 0, \]
  for some constants $C > 0, c \in (0, 1)$. In particular, $B_n$ is non-singular with probability at least $1 - c^n$.

- The argument of Kahn–Komlós–Szemerédi and its development by Tao–Vu and Bourgain–Vu–Wood provides stronger singularity probability estimates:
  \[ \Pr\{B_n \text{ is singular}\} \leq \left(\frac{1}{\sqrt{2}} + o(1)\right)^n, \]
  although the method does not seem to imply strong small ball probability estimates for $s_{\min}(B_n)$.

- The folklore conjecture in the field is that
  \[ \Pr\{B_n \text{ is singular}\} = \left(\frac{1}{2} + o(1)\right)^n, \]
  that is, considerable contribution to the matrix singularity comes from the event that two rows or columns of the matrix are equal.
Theorem (T.'18)

Let $B_n$ be an $n \times n$ random matrix with i.i.d $\pm 1$ entries. Then for any $\varepsilon > 0$ there is $C > 0$ depending only on $\varepsilon$ such that

$$\mathbb{P}\{s_{\min}(B_n) \leq t/\sqrt{n}\} \leq Ct + C(1/2 + \varepsilon)^n, \quad t > 0.$$ 

In particular,

$$\mathbb{P}\{B_n \text{ is singular}\} = \left(\frac{1}{2} + o_n(1)\right)^n.$$

Moreover, for any fixed $p \in (0, 1/2]$, and the sequence of random matrices $M_n$ with i.i.d. Bernoulli($p$) entries,

$$\mathbb{P}\{M_n \text{ is singular}\} = \left(1 - p + o_n(1)\right)^n.$$
Lemma
For any \( \varepsilon \in (0, 1] \) there are \( n_0 \in \mathbb{N}, \gamma > 0 \) and \( \delta, \rho \in (0, 1) \) depending only on \( \varepsilon \) such that for \( n \geq n_0 \),

\[
\mathbb{P}\left\{ \| B_n x \|_2 \leq \gamma \sqrt{n} \text{ for some } x \in \text{Comp}_n(\delta, \rho) \right\} \leq \left( \frac{1}{2} + \varepsilon \right)^n.
\]

Proof: \( \varepsilon \)-net argument.
Proof: preliminary reductions

Lemma

For any $\varepsilon \in (0, 1]$ there are $n_0 \in \mathbb{N}$, $\gamma > 0$ and $\delta, \rho \in (0, 1)$ depending only on $\varepsilon$ such that for $n \geq n_0$,

$$
\mathbb{P}\{\|B_n x\|_2 \leq \gamma \sqrt{n} \text{ for some } x \in \text{Comp}_n(\delta, \rho)\} \leq (1/2 + \varepsilon)^n.
$$

Proof: $\varepsilon$–net argument.

Thus,

$$
\mathbb{P}\{s_{\min}(B_n) \leq t/\sqrt{n}\} \\
\leq \mathbb{P}\{\|B_n x\|_2 \leq t/\sqrt{n} \text{ for some } x \in \text{Comp}_n(\delta, \rho)\} \\
+ \frac{1}{\delta} \mathbb{P}\{|\langle \text{col}_n(B_n), Y_n \rangle| \leq t/\rho\} \\
\leq \left(\frac{1}{2} + \varepsilon\right)^n + C \mathbb{P}\{|\langle \text{col}_n(B_n), Y_n \rangle| \leq Ct\},
$$

where $Y_n$ is the random unit normal vector to first $n - 1$ column of $B_n$. 
We need to show that the normal vector $Y_n$ is typically “very unstructured”, so that its inner product with the column $\text{col}_n(B_n)$ behaves almost as a random variable with a bounded density up to the scale $(\frac{1}{2} + \varepsilon)^n$. 
Proof: preliminary reductions

We need to show that the normal vector $Y_n$ is typically “very unstructured”, so that its inner product with the column $\text{col}_n(B_n)$ behaves almost as a random variable with a bounded density up to the scale $(\frac{1}{2} + \varepsilon)^n$.

Fix a large constant $L > 0$, and for any $N \leq (2 - \varepsilon)^n$ ($N$ – an integer power of 2), define a set of unit vectors

$$Q_N := \left\{ x \in \text{Incomp}_n(\delta, \rho) : \sup \left\{ t \in [0, 1] : \mathcal{L}(\langle \text{col}_n(B_n), x \rangle, t) > Lt \right\} \in \left[ \frac{1}{2N}, \frac{1}{N} \right] \right\},$$

where

$$\mathcal{L}(\xi, t) := \sup_{r} \mathbb{P}\{|\xi - r| \leq t\}$$

is the Lévy concentration function.
Proof: preliminary reductions

We need to show that the normal vector $Y_n$ is typically “very unstructured”, so that its inner product with the column $\text{col}_n(B_n)$ behaves almost as a random variable with a bounded density up to the scale $(\frac{1}{2} + \varepsilon)^n$.

Fix a large constant $L > 0$, and for any $N \leq (2 - \varepsilon)^n$ ($N$ – an integer power of 2), define a set of unit vectors

$$Q_N := \left\{ x \in \text{Incomp}_n(\delta, \rho) : \right.$$ 

$$\sup \left\{ t \in [0, 1] : \mathcal{L}(t \langle \text{col}_n(B_n), x \rangle, t) > Lt \right\} \in \left[ \frac{1}{2N}, \frac{1}{N} \right] \} ,$$

where

$$\mathcal{L}(\xi, t) := \sup_r \mathbb{P}\{|\xi - r| \leq t\}$$

is the Lévy concentration function.

Conditioning on any realization of $Y_n$ inside $Q_N$, we have

$$\mathbb{P}_{\text{col}_n(B_n)} \left\{ |\langle \text{col}_n(B_n), Y_n \rangle| \leq Ct \right\} \lesssim t + 1/N .$$

Thus, to prove the theorem, it is enough to show that for each admissible $N \leq (2 - \varepsilon)^n$,

$$\mathbb{P}\{ Y_n \in Q_N \} \leq (1/2 + \varepsilon)^n .$$
Proof: discretization

Let $B_{1..n-1}$ be the matrix obtained by removing $n$–th column from $B_n$. It can be shown that conditioned on any realization of $B_{1..n-1}$ such that $Y_n \in Q_N$, there exists a vector $Y \in \left(\frac{1}{N}\mathbb{Z}\right)^n$ such that

1. (distance to $Y_n$) $\|Y_n - Y\|_\infty \leq \frac{1}{N}$.
Proof: discretization

Let $B_{1..n-1}$ be the matrix obtained by removing $n$–th column from $B_n$. It can be shown that conditioned on any realization of $B_{1..n-1}$ such that $Y_n \in Q_N$, there exists a vector $\mathbf{Y} \in \left(\frac{1}{N}\mathbb{Z}\right)^n$ such that

- (distance to $Y_n$) $\|Y_n - \mathbf{Y}\|_\infty \leq \frac{1}{N}$;
- (anti-concentration) $\mathbb{P}_b\{\left|\sum_{i=1}^n b_i Y_i\right| \leq t\} \leq C Lt$ for all $t \geq 1/N$;
Proof: discretization

Let $B_{1..n-1}$ be the matrix obtained by removing $n$–th column from $B_n$. It can be shown that conditioned on any realization of $B_{1..n-1}$ such that $Y_n \in Q_N$, there exists a vector $Y \in \left( \frac{1}{N} \mathbb{Z} \right)^n$ such that

- (distance to $Y_n$) $\| Y_n - Y \|_\infty \leq \frac{1}{N}$;
- (anti-concentration) $\mathbb{P}_b \left\{ \left| \sum_{i=1}^n b_i Y_i \right| \leq t \right\} \leq C L t$ for all $t \geq 1/N$;
- (concentration) $L_b \left( \sum_{i=1}^n b_i Y_i, 1/N \right) \geq c L_b \left( \sum_{i=1}^n b_i (Y_n)_i, 1/N \right)$;
Proof: discretization

Let $B_{1..n-1}$ be the matrix obtained by removing $n$-th column from $B_n$. It can be shown that conditioned on any realization of $B_{1..n-1}$ such that $Y_n \in Q_N$, there exists a vector $Y \in \left(\frac{1}{N} \mathbb{Z}\right)^n$ such that

- (distance to $Y_n$) $\|Y_n - Y\|_\infty \leq \frac{1}{N}$;
- (anti-concentration) $\mathbb{P}_b\{\left| \sum_{i=1}^{n} b_i Y_i \right| \leq t\} \leq C L t$ for all $t \geq 1/N$;
- (concentration) $\mathcal{L}_b \left( \sum_{i=1}^{n} b_i Y_i, 1/N \right) \geq c \mathcal{L}_b \left( \sum_{i=1}^{n} b_i (Y_n)_i, 1/N \right)$;
- (distance to the column span) $\|B_{1..n-1}^T Y\|_2 \leq Cn/N$.

Here, $b_1, \ldots, b_n$ are $\pm 1$ variables mutually independent with $Y_n, Y_n$. 
Proof: discretization

Let $B_{1..n-1}$ be the matrix obtained by removing $n$–th column from $B_n$. It can be shown that conditioned on any realization of $B_{1..n-1}$ such that $Y_n \in Q_N$, there exists a vector $\mathbf{Y} \in (\frac{1}{N}\mathbb{Z})^n$ such that

- (distance to $Y_n$) $\|Y_n - \mathbf{Y}\|_\infty \leq \frac{1}{N}$;
- (anti-concentration) $\mathbb{P}_b\{\left| \sum_{i=1}^n b_i Y_i \right| \leq t \} \leq C Lt$ for all $t \geq 1/N$;
- (concentration) $\mathcal{L}_b\left(\sum_{i=1}^n b_i Y_i, 1/N\right) \geq c \mathcal{L}_b\left(\sum_{i=1}^n b_i(Y_n)_i, 1/N\right)$;
- (distance to the column span) $\|B_{1..n-1}^\top \mathbf{Y}\|_2 \leq Cn/N$.

Here, $b_1, \ldots, b_n$ are ±1 variables mutually independent with $Y_n, Y_n$.

The procedure is called “random rounding” in computer science literature. In the random matrix context, it was first used by Livshyts (2018).

To construct the approximation $\mathbf{Y}$ of the vector $Y_n$, satisfying the conditions mentioned above, we replace each component $(Y_n)_i$ with a random variable $Y_i$ distributed on the set $\{\lfloor N(Y_n)_i \rfloor/N, \lfloor N(Y_n)_i \rfloor/N + 1/N\}$, and such that $\mathbb{E}_i Y_i = (Y_n)_i$. Then with high probability $\mathbf{Y}$ satisfies the needed properties.
Proof: preliminary reductions

Denote by $\mathcal{N}_N$ a set of all realizations of $\mathbf{Y}$ for all possible realizations of $Y_n \in Q_N$. 

...
Denote by $\mathcal{N}_N$ a set of all realizations of $Y$ for all possible realizations of $Y_n \in Q_N$.

That is, $\mathcal{N}_N$ is the set of all incompressible vectors $y$ in $\frac{1}{N} \mathbb{Z}^n$ such that $\|y\|_2 \in [1/2, 2]$ and

- $\mathbb{P}_b\{|\sum_{i=1}^n b_i y_i| \leq t\} \leq C L t$ for all $t \geq 1/N$;
- $\mathcal{L}_b(\sum_{i=1}^n b_i y_i, 1/N) \geq c L/N$,

so that, in view of the properties of the approximations, the event $\{Y_n \in Q_N\}$ is contained inside the event

$$\{\|B_{1..n-1}^T y\|_2 \leq C n/N \text{ for some } y \in \mathcal{N}_N\}.$$
Proof: preliminary reductions

Denote by $\mathcal{N}_N$ a set of all realizations of $Y$ for all possible realizations of $Y_n \in Q_N$.

That is, $\mathcal{N}_N$ is the set of all incompressible vectors $y$ in $\frac{1}{N} \mathbb{Z}^n$ such that $\|y\|_2 \in [1/2, 2]$ and

- $\mathbb{P}_b\{ \left| \sum_{i=1}^n b_i y_i \right| \leq t \} \leq C L t$ for all $t \geq 1/N$;
- $\mathcal{L}_b\left( \sum_{i=1}^n b_i y_i, 1/N \right) \geq c L/N$,

so that, in view of the properties of the approximations, the event $\{ Y_n \in Q_N \}$ is contained inside the event

$$\left\{ \|B_{1..n-1} y\|_2 \leq C n / N \text{ for some } y \in \mathcal{N}_N \right\}.$$

Then we get

$$\mathbb{P}\{ Y_n \in Q_N \} \leq |\mathcal{N}_N| \sup_{y \in \mathcal{N}} \mathbb{P}\{ \|B_{1..n-1} y\|_2 \leq C n / N \} \leq |\mathcal{N}_N| (\tilde{C} \sqrt{n} / N)^n.$$
Proof: preliminary reductions

Denote by \( \mathcal{N}_N \) a set of all realizations of \( \mathbf{Y} \) for all possible realizations of \( Y_n \in Q_N \).

That is, \( \mathcal{N}_N \) is the set of all *incompressible* vectors \( y \) in \( \frac{1}{N} \mathbb{Z}^n \) such that \( \|y\|_2 \in [1/2, 2] \) and

\[
\begin{align*}
\mathbb{P}_b \{ \left| \sum_{i=1}^n b_i y_i \right| \leq t \} & \leq C L t \text{ for all } t \geq 1/N; \\
\mathcal{L}_b \left( \sum_{i=1}^n b_i y_i, 1/N \right) & \geq c L/N,
\end{align*}
\]

so that, in view of the properties of the approximations, the event \( \{ Y_n \in Q_N \} \) is contained inside the event

\[
\{ \|B_{1..n-1} y\|_2 \leq Cn/N \text{ for some } y \in \mathcal{N}_N \}.
\]

Then we get

\[
\mathbb{P}\{ Y_n \in Q_N \} \leq |\mathcal{N}_N| \sup_{y \in \mathcal{N}} \mathbb{P}\{ \|B_{1..n-1} y\|_2 \leq Cn/N \} \leq |\mathcal{N}_N| \left( \tilde{C} \sqrt{n}/N \right)^n.
\]

The estimate is satisfactory as long as it is possible to show that \( |\mathcal{N}_N| \leq (1/2 + \varepsilon)^n (\tilde{C} \sqrt{n}/N)^{-n} \) for a *large* constant \( \tilde{C} \).
Denote by $\mathcal{N}_N$ a set of all realizations of $Y$ for all possible realizations of $Y_n \in Q_N$.

That is, $\mathcal{N}_N$ is the set of all \textit{incompressible} vectors $y$ in $\frac{1}{N}\mathbb{Z}^n$ such that $\|y\|_2 \in [1/2, 2]$ and

\begin{itemize}
  \item $\mathbb{P}_b\{\left| \sum_{i=1}^{n} b_i y_i \right| \leq t \} \leq C L t$ for all $t \geq 1/N$;
  \item $\mathcal{L}_b(\sum_{i=1}^{n} b_i y_i, 1/N) \geq c L/N$,
\end{itemize}

so that, in view of the properties of the approximations, the event $\{Y_n \in Q_N\}$ is contained inside the event

$$\left\{ \|B_{1..n-1} y\|_2 \leq C n/N \text{ for some } y \in \mathcal{N}_N \right\}.$$

Then we get

$$\mathbb{P}\{Y_n \in Q_N\} \leq |\mathcal{N}_N| \sup_{y \in \mathcal{N}} \mathbb{P}\{\|B_{1..n-1} y\|_2 \leq C n/N\} \leq |\mathcal{N}_N| (\tilde{C} \sqrt{n}/N)^n.$$

The estimate is satisfactory as long as it is possible to show that $|\mathcal{N}_N| \leq (1/2 + \varepsilon)^n (\tilde{C} \sqrt{n}/N)^{-n}$ for a \textbf{large} constant $\tilde{C}$.

The trivial upper bound $|\mathcal{N}_N| \leq |\frac{1}{N}\mathbb{Z}^n \cap 2B_2^n|$ is not sufficient.
The set $\mathcal{N}_N$ is not convenient in a sense that it does not embed into a cartesian product of satisfactory cardinality. We can use the following elementary fact:

**Lemma**

Let $B_D := \frac{1}{N} \mathbb{Z}^n \cap 2B_2^n$. Then there exist $2^n$ subsets of $\frac{1}{N} \mathbb{Z}^n$ — $W_1, \ldots, W_{2^n}$ — such that $B_D \subset \bigcup_{i=1}^{2^n} W_i$.
The set $\mathcal{N}_N$ is not convenient in a sense that it does not embed into a cartesian product of satisfactory cardinality. We can use the following elementary fact:

**Lemma**

Let $B_D := \frac{1}{N} \mathbb{Z}^n \cap 2B_2^n$. Then there exist $2^n$ subsets of $\frac{1}{N} \mathbb{Z}^n$ — $W_1, \ldots, W_{2^n}$ — such that $B_D \subset \bigcup_{i=1}^{2^n} W_i$, each $W_i$ is a Cartesian product of symmetric integer intervals,
The set $\mathcal{N}_N$ is not convenient in a sense that it does not embed into a cartesian product of satisfactory cardinality. We can use the following elementary fact:

**Lemma**

Let $B_D := \frac{1}{N} \mathbb{Z}^n \cap 2B_2^n$. Then there exist $2^n$ subsets of $\frac{1}{N} \mathbb{Z}^n$ — $W_1, \ldots, W_{2^n}$ — such that $B_D \subset \bigcup_{i=1}^{2^n} W_i$, each $W_i$ is a Cartesian product of symmetric integer intervals, and $\sum_{i=1}^{2^n} |W_i| \leq C^n |B_D|$.

It will be more useful to study sets $\mathcal{N}_N \cap W_i$ instead of $\mathcal{N}_N$ itself.
The set $\mathcal{N}_N$ is not convenient in a sense that it does not embed into a cartesian product of satisfactory cardinality. We can use the following elementary fact:

**Lemma**

Let $B_D := \frac{1}{N} \mathbb{Z}^n \cap 2B_2^n$. Then there exist $2^n$ subsets of $\frac{1}{N} \mathbb{Z}^n$ — $W_1, \ldots, W_{2^n}$ — such that $B_D \subset \bigcup_{i=1}^{2^n} W_i$, each $W_i$ is a Cartesian product of symmetric integer intervals, and $\sum_{i=1}^{2^n} |W_i| \leq C^n |B_D|$. It will be more useful to study sets $\mathcal{N}_N \cap W_i$ instead of $\mathcal{N}_N$ insel.

**Theorem (Net cardinality)**

Let $M \geq 1$. Take a large $n \geq n_0(M, \varepsilon)$, and $N \leq (2 - \varepsilon)^n$. Then

$$|\mathcal{N}_N \cap W_i| \leq e^{-M n} (\sqrt{n}/N)^{-n}.$$  

The above theorem implies the main result.
The same theorem can be restated in probabilistic form:

**Theorem (Net cardinality — probabilistic form)**

Let $\delta \in (0, 1]$, $M \geq 1$. There exist $L_B = L_B(\delta) > 0$ depending **only** on $\delta$ (and not on $M$) with the following property. Take a large $n \geq n_0(\delta, M)$, and $N \leq (2 - \varepsilon)^n$, and let

$$A := \{-2N, \ldots, -N - 1, N + 1, \ldots, 2N\}^{\delta n} \times \{-N, -N + 1, \ldots, N\}^{n - \delta n}.$$

Further, assume that a random vector $(\xi_1, \ldots, \xi_n)$ is uniform on $A$. Then

$$\mathbb{P}\left\{2^{-n} \sup_{\lambda \in \mathbb{R}} \sum_{(v_j)_{j=1}^n \in \{-1,1\}^n} 1_{[-\sqrt{n}, \sqrt{n}]}(\lambda + v_1 \xi_1 + \cdots + v_n \xi_n) > \frac{L_B N}{N}\right\} \leq e^{-M n}.$$
The same theorem can be restated in probabilistic form:

**Theorem (Net cardinality — probabilistic form)**

Let \( \delta \in (0, 1] \), \( M \geq 1 \). There exist \( L_B = L_B(\delta) > 0 \) depending **only** on \( \delta \) (and not on \( M \)) with the following property. Take a large \( n \geq n_0(\delta, M) \), and \( N \leq (2 - \varepsilon)^n \), and let

\[
A := \{-2N, \ldots, -N - 1, N + 1, \ldots, 2N\}^{\delta n} \times \{-N, -N + 1, \ldots, N\}^{n - \delta n}.
\]

Further, assume that a random vector \((\xi_1, \ldots, \xi_n)\) is uniform on \(A\). Then

\[
\mathbb{P}\left\{ 2^{-n} \sup_{\lambda \in \mathbb{R}} \sum_{(v_j)_{j=1}^{n} \in \{-1, 1\}^n} 1_{[-\sqrt{n}, \sqrt{n}]}(\lambda + v_1\xi_1 + \cdots + v_n\xi_n) > \frac{L_B}{N} \right\} \leq e^{-M n}.
\]

The crucial point of the statement is that \( L_B \) does not depend on \( M \), so one can make \( M \) arbitrarily small as long as \( n \) is large. This statement (in fact, a little more technical version) is translated into the cardinality estimates for the net \( \mathcal{N}_N \) discussed in the previous slide.
Let
\[ A := \{ -2N, \ldots, -N - 1, N + 1, \ldots, 2N \}^{\delta n} \times \{ -N, -N + 1, \ldots, N \}^{n - \delta n}. \]

Further, assume that a random vector \((\xi_1, \ldots, \xi_n)\) is uniform on \(A\).

For each \(f \in \ell_1(\mathbb{Z})\) and \(\ell \geq 1\), define
\[
f_{A,\ell}(t) := 2^{-\ell} \sum_{(v_j)_{j=1}^{\ell} \in \{-1,1\}^\ell} f\left(t + v_1\xi_1 + \cdots + v_\ell\xi_\ell\right).
\]

Thus, \(f_{A,\ell}\) can be viewed as an average of \(2^\ell\) (random) translations of \(f\).
Let
\[ A := \{-2N, \ldots, -N - 1, N + 1, \ldots, 2N\}^{\delta n} \times \{-N, -N + 1, \ldots, N\}^{n - \delta n}. \]
Further, assume that a random vector \((\xi_1, \ldots, \xi_n)\) is uniform on \(A\).
For each \(f \in \ell_1(\mathbb{Z})\) and \(\ell \geq 1\), define
\[
    f_{A,\ell}(t) := 2^{-\ell} \sum_{(v_j)_{j=1}^{\ell} \in \{-1,1\}^\ell} f(t + v_1\xi_1 + \cdots + v_\ell\xi_\ell).
\]
Thus, \(f_{A,\ell}\) can be viewed as an average of \(2^\ell\) (random) translations of \(f\).

Randomized averaging for log-Lipschitz functions

For any \(\delta \in (0, 1]\), \(M \geq 1\) there are \(n_0 = n_0(\delta, M) \geq 1\), \(\eta_0 = \eta_0(\delta, M) \in (0, 1]\) depending on \(\delta, M\) and \(L_B = L_B(\delta) > 0\) depending only on \(\delta\) (and not on \(M\)) with the following property. Take \(n \geq n_0\), \(N \leq (2 - \varepsilon)^n\), and let \((f(t))_{t \in \mathbb{Z}}\) be a sequence of positive reals with \(\|f\|_1 = 1\) and such that \(\log f\) is \(\eta_0\)-Lipschitz on \(\mathbb{Z}\). Then
\[
    \mathbb{P}\left\{\|f_{A,n}\|_\infty > L_B(N\sqrt{n})^{-1}\right\} \leq \exp(-Mn).
\]
It can be shown that the above theorem implies the main result.
Let
\[ A := \{-2N, \ldots, -N - 1, N + 1, \ldots, 2N\}^{\delta n} \times \{-N, -N + 1, \ldots, N\}^{n-\delta n}. \]
Assume \((\xi_1, \ldots, \xi_n)\) is uniform on \(A\). For each \(f \in \ell_1(\mathbb{Z})\) and \(\ell \geq 1\),
\[ f_{A,\ell}(t) := 2^{-\ell} \sum_{(v_j)_{j=1}^{\ell} \in \{-1,1\}^\ell} f(t + v_1 \xi_1 + \cdots + v_\ell \xi_\ell). \]

Preprocessing step
For any \(M > 0, \delta \in (0, 1)\) there are \(L = L(M, \delta) > 0\) and \(n_0 = n_0(M, \delta) \in \mathbb{N}\) with the following property. Let \(f \in \ell_1(\mathbb{Z})\) with \(\|f\|_1 = 1\) and \(f(t) > 0\), let \(n \geq n_0, n/2 \leq \ell \leq n\), and let \(A\) be as above for some \(N \leq 2^n\). Then
\[ \mathbb{P}\{\|f_{A,\ell}\|_\infty > \max(L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty)\} \leq \exp(-Mn). \]
Let
\[ \mathcal{A} := \{-2N, \ldots, -N - 1, N + 1, \ldots, 2N\}^{\delta n} \times \{-N, -N + 1, \ldots, N\}^{n-\delta n}. \]

Assume \((\xi_1, \ldots, \xi_n)\) is uniform on \(\mathcal{A}\). For each \(f \in \ell_1(\mathbb{Z})\) and \(\ell \geq 1\),
\[ f_{A,\ell}(t) := 2^{-\ell} \sum_{(v_j)_{j=1}^\ell \in \{-1,1\}^\ell} f(t + v_1\xi_1 + \cdots + v_\ell\xi_\ell). \]

**Preprocessing step**

For any \(M > 0, \delta \in (0, 1)\) there are \(L = L(M, \delta) > 0\) and \(n_0 = n_0(M, \delta) \in \mathbb{N}\) with the following property. Let \(f \in \ell_1(\mathbb{Z})\) with \(\|f\|_1 = 1\) and \(f(t) > 0\), let \(n \geq n_0, n/2 \leq \ell \leq n\), and let \(\mathcal{A}\) be as above for some \(N \leq 2^n\). Then
\[ \mathbb{P}\{\|f_{A,\ell}\|_\infty > \max \left(L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty \right) \} \leq \exp(-Mn). \]

The main difference with the previous statement is that \(L\) is allowed to depend on \(M\). The log-Lipschitzness is not needed here.
For each $f \in \ell_1(\mathbb{Z})$ and $\ell \geq 1$,

$$f_{A,\ell}(t) := 2^{-\ell} \sum_{(v_j)_{j=1}^\ell \in \{-1,1\}^\ell} f(t + v_1 \xi_1 + \cdots + v_\ell \xi_\ell),$$

so that $f_{A,\ell}(t) = \frac{1}{2} (f_{A,\ell-1}(t + \xi_\ell) + f_{A,\ell-1}(t - \xi_\ell))$. 
Randomized averaging: preprocessing step

For each \( f \in \ell_1(\mathbb{Z}) \) and \( \ell \geq 1 \),

\[
    f_{A,\ell}(t) := 2^{-\ell} \sum_{(v_j)_{j=1}^\ell \in \{-1,1\}^\ell} f(t + v_1 \xi_1 + \cdots + v_\ell \xi_\ell),
\]

so that \( f_{A,\ell}(t) = \frac{1}{2} (f_{A,\ell-1}(t + \xi_\ell) + f_{A,\ell-1}(t - \xi_\ell)) \).

Let \( R > 0 \) be a parameter and let \( m \in \{1, 2, \ldots, \ell\} \). We say that a point \( t \in \mathbb{Z} \) decays at time \( m \) if

\[
    f_{A,m-1}(t + 2\xi_m) \leq \frac{R}{N\sqrt{n}} \quad \text{and} \quad f_{A,m-1}(t - 2\xi_m) \leq \frac{R}{N\sqrt{n}}.
\]

Further, given any \( t \in \mathbb{Z} \) and a sequence \( (v_i)_{i=1}^\ell \in \{-1, 1\}^\ell \), the descendant sequence for \( t \) with respect to \( (v_i)_{i=1}^\ell \) is a random sequence \( (t_i)_{i=0}^\ell \), where

\[
    t_i = t - \sum_{j=1}^i v_j \xi_j, \ 1 \leq i \leq \ell \quad \text{(where } t_0 := t). \]

The preprocessing step is accomplished by proving the following: the event that \( \ell_\infty \)-norm of \( f_{A,\ell} \) is "large" is contained within the event that there exists a descendant sequence such that a proportional number of its elements do not decay.
Randomized averaging: preprocessing step

\( R \) is a parameter. A point \( t \in \mathbb{Z} \) decays at time \( m \) if
\[
f_{A,m-1}(t + 2\xi_m) \leq \frac{R}{N \sqrt{n}} \text{ and } f_{A,m-1}(t - 2\xi_m) \leq \frac{R}{N \sqrt{n}}.
\]

**Lemma**

Let \( L > 0 \), and set \( R := \varepsilon L \). Define event \( \mathcal{E} \) as the subset the probability space such that there exists a sequence \( (v_i)_{i=1}^\ell \in \{-1, 1\}^\ell \) and a point \( t \in \mathbb{Z} \) so that the descendant sequence \( (t_i)_{i=0}^\ell \) for \( t \) with respect to \( (v_i)_{i=1}^\ell \) (i.e. \( t_i = t - \sum_{j=1}^i v_j \xi_j \), \( 1 \leq i \leq \ell \), and \( t_0 := t \)) satisfies
\[
\left| \{1 \leq i \leq \ell : \text{ } t_{i-1} \text{ does not decay at time } i\} \right| \geq -n \log \left( \frac{(1/2 + \varepsilon)/(1/2 + \varepsilon/2)}{2 \log (1/2 + \varepsilon/2)} \right).
\]

Then \( \mathcal{E} \supset \{\|f_{A,\ell}\|_\infty > \max(L(N \sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty)\} \).
Randomized averaging: preprocessing step

$R$ is a parameter. A point $t \in \mathbb{Z}$ decays at time $m$ if $f_{A,m-1}(t + 2\xi_m) \leq \frac{R}{N\sqrt{n}}$ and $f_{A,m-1}(t - 2\xi_m) \leq \frac{R}{N\sqrt{n}}$.

**Lemma**

Let $L > 0$, and set $R := \varepsilon L$. Define event $\mathcal{E}$ as the subset the probability space such that there exists a sequence $(v_i)_{i=1}^{\ell} \in \{-1, 1\}^\ell$ and a point $t \in \mathbb{Z}$ so that the descendant sequence $(t_i)_{i=0}^{\ell}$ for $t$ with respect to $(v_i)_{i=1}^{\ell}$ (i.e. $t_i = t - \sum_{j=1}^{i} v_j \xi_j$, $1 \leq i \leq \ell$, and $t_0 := t$) satisfies

$$\left| \left\{ 1 \leq i \leq \ell : t_i - 1 \text{ does not decay at time } i \right\} \right| \geq -\frac{n \log \left( (1/2 + \varepsilon)/(1/2 + \varepsilon/2) \right)}{2 \log (1/2 + \varepsilon/2)}.$$

Then $\mathcal{E} \supset \{ \|f_{A,\ell}\|_\infty > \max \left( L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty \right) \}$.

Less formally, the lemma asserts that the event interesting for us (the $\ell_\infty$ norm of $f_{A,\ell}$ is large) is contained in the event that there is a descendant sequence with a proportion of $n$ elements not decaying.
Proof of the Lemma:

Fix a realization of $\xi_1, \ldots, \xi_\ell$ such that $\|f_{A,\ell}\|_\infty > \max(L(N\sqrt{n}) - 1, (1/2 + \varepsilon)\ell\|f\|_\infty)$.

Construct a sequence of integers $(t_i)_{\ell=0}^\infty$ inductively in inverse order as follows. Take $t_\ell$ to be any integer such that $f_{A,\ell}(t_\ell) > \max(L(N\sqrt{n}) - 1, (1/2 + \varepsilon)\ell\|f\|_\infty)$.

At $(\ell - i + 1)$–st step ($1 \leq i \leq \ell$) we assume that $t_i$ has been defined. It follows from the definition of $f_{A,i}$ that $f_{A,i-1}(t_i + v_i \xi_i) \geq f_{A,i}(t_i)$ for some $v_i \in \{-1, 1\}$.

Then we set $t_i := t_i + v_i \xi_i$.

The sequence $(t_i)_{\ell=0}^\infty$ constructed above is the descendant sequence for $t_0$ with respect to $(v_i)_{\ell=0}^\infty$, which satisfies the conditions:

(a) $f_{A,i-1}(t_i - 1) \geq f_{A,i}(t_i)$ for all $1 \leq i \leq \ell$;

(b) $f_{A,\ell}(t_\ell) > \max(L(N\sqrt{n}) - 1, (1/2 + \varepsilon)\ell\|f\|_\infty)$.

It remains to show that the sequence $(t_i)_{\ell=0}^\infty$ satisfies the condition $|\{1 \leq i \leq \ell : t_i - 1 \text{ does not decay at time } i\}| \geq -n \log((1/2 + \varepsilon)/(1/2 + \varepsilon/2))^2 \log((1/2 + \varepsilon/2))$. 

Proof of the Lemma:

- Fix a realization of $\xi_1, \ldots, \xi_\ell$ such that
  \[ \| f_{A, \ell} \|_\infty > \max \left( L (N \sqrt{n})^{-1}, (\frac{1}{2} + \varepsilon) \ell \| f \|_\infty \right). \]
Proof of the Lemma:

1. Fix a realization of $\xi_1, \ldots, \xi_\ell$ such that
   $$\|f_{A,\ell}\|_\infty > \max \left( L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty \right).$$

2. Construct a sequence of integers $(t_i)_{i=0}^\ell$ inductively in inverse order as follows. Take $t_\ell$ to be any integer such that $f_{A,\ell}(t_\ell) > \max \left( L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty \right)$. 

It remains to show that the sequence $(t_i)_{i=0}^\ell$ satisfies the condition
$$\left| \left\{ 1 \leq i \leq \ell : t_i - 1 \text{ does not decay at time } i \right\} \right| \geq -n \log \left( \frac{1}{2 + \varepsilon} \right)^2 \log \left( \frac{1}{2 + \varepsilon/2} \right).$$
Proof of the Lemma:

- Fix a realization of $\xi_1, \ldots, \xi_\ell$ such that 
  \[ \| f_{A, \ell} \|_\infty > \max \left( L \left( N \sqrt{n} \right)^{-1}, (1/2 + \varepsilon)^\ell \| f \|_\infty \right). \]

- Construct a sequence of integers $(t_i)_{i=0}^\ell$ inductively in inverse order as follows. Take $t_\ell$ to be any integer such that 
  \[ f_{A, \ell}(t_\ell) > \max \left( L \left( N \sqrt{n} \right)^{-1}, (1/2 + \varepsilon)^\ell \| f \|_\infty \right). \]

  At $(\ell - i + 1)$–st step $(1 \leq i \leq \ell)$ we assume that $t_i$ has been defined. It follows from the definition of $f_{A, i}$, that 
  $f_{A, i-1}(t_i + v_i \xi_i) \geq f_{A, i}(t_i)$ for some $v_i \in \{-1, 1\}$. Then we set $t_{i-1} := t_i + v_i \xi_i$. 
Proof of the Lemma:

- Fix a realization of $\xi_1, \ldots, \xi_\ell$ such that
  $$\|f_{A,\ell}\|_\infty > \max \left(L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty\right).$$

- Construct a sequence of integers $(t_i)_{i=0}^\ell$ inductively in inverse order as follows. Take $t_\ell$ to be any integer such that
  $$f_{A,\ell}(t_\ell) > \max \left(L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty\right).$$

  At $(\ell - i + 1)$–st step $(1 \leq i \leq \ell)$ we assume that $t_i$ has been defined.

  It follows from the definition of $f_{A,i}$, that $f_{A,i}(t_i + v_i\xi_i) \geq f_{A,i}(t_i)$ for some $v_i \in \{-1, 1\}$. Then we set $t_{i-1} := t_i + v_i\xi_i$.

- The sequence $(t_i)_{i=0}^\ell$ constructed above is the descendant sequence for $t_0$ with respect to $(v_i)_{i=1}^\ell$, which satisfies the conditions
  
  (a) $f_{A,i-1}(t_{i-1}) \geq f_{A,i}(t_i)$ for all $1 \leq i \leq \ell$;

  (b) $f_{A,\ell}(t_\ell) > \max \left(L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty\right)$.
Proof of the Lemma:

- Fix a realization of $\xi_1, \ldots, \xi_\ell$ such that $\|f_{A,\ell}\|_\infty > \max \left( L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty \right)$. 

- Construct a sequence of integers $(t_i)_{i=0}^\ell$ inductively in inverse order as follows. Take $t_\ell$ to be any integer such that $f_{A,\ell}(t_\ell) > \max \left( L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty \right)$.

  At $(\ell - i + 1)$-st step $(1 \leq i \leq \ell)$ we assume that $t_i$ has been defined. It follows from the definition of $f_{A,i}$, that $f_{A,i-1}(t_i + v_i \xi_i) \geq f_{A,i}(t_i)$ for some $v_i \in \{-1, 1\}$. Then we set $t_{i-1} := t_i + v_i \xi_i$.

- The sequence $(t_i)_{i=0}^\ell$ constructed above is the descendant sequence for $t_0$ with respect to $(v_i)_{i=1}^\ell$, which satisfies the conditions

  (a) $f_{A,i-1}(t_{i-1}) \geq f_{A,i}(t_i)$ for all $1 \leq i \leq \ell$;

  (b) $f_{A,\ell}(t_\ell) > \max \left( L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty \right)$.

- It remains to show that the sequence $(t_i)_{i=0}^\ell$ satisfies the condition

  $$|\{1 \leq i \leq \ell : t_{i-1} \text{ does not decay at time } i\}| \geq -\frac{n \log \left( ((1/2 + \varepsilon)/(1/2 + \varepsilon/2))^\ell \right)}{2 \log (1/2 + \varepsilon/2)}.$$
Proof of the Lemma (continued):

We are given a sequence \((t_i)_{i=0}^\ell\) such that

\[ t_i - t_{i-1} = t_i + v_i \xi_i, \quad i \geq 1, \]

and

\[(a) f_{A,i-1}(t_{i-1}) \geq f_{A,i}(t_i) \quad \text{for all} \quad 1 \leq i \leq \ell; \]

\[(b) f_{A,\ell}(t_{\ell}) > \max(L(N\sqrt{n}) - 1, (1/2 + \epsilon) \ell \|f\|_\infty). \]

The goal: show \(|\{1 \leq i \leq \ell : t_i - t_{i-1} \text{ does not decay at time } i\}|\) is large.

Assume that \(1 \leq i \leq \ell\) is such that \(t_i - t_{i-1}\) decays at time \(i\). We have

\[ f_{A,i}(t_i) = \frac{1}{2} (f_{A,i-1}(t_i - 2v_i \xi_i) + f_{A,i-1}(t_i + v_i \xi_i)). \]

By definition of decay at time \(i\), both \(f_{A,i-1}(t_i - 2v_i \xi_i)\) and \(f_{A,i-1}(t_i - 2v_i \xi_i)\) are less than \(\epsilon L_N \sqrt{n}\), hence less than \(\epsilon f_{A,i-1}(t_i - 1)\). Thus,

\[ f_{A,i}(t_i) \leq \left(\frac{1}{2} + \epsilon/2\right) f_{A,i-1}(t_i - 1). \]

Applying the relation for all \(i\) when \(t_i - t_{i-1}\) decays, we get for \(u = |\{1 \leq i \leq \ell : t_i - t_{i-1} \text{ decays at time } i\}|\):

\[ \left(\frac{1}{2} + \epsilon/2\right) u \|f\|_\infty < f_{A,\ell}(t_{\ell}) \leq \left(\frac{1}{2} + \epsilon/2\right) u \|f\|_\infty. \]

This implies a required estimate.
Proof of the Lemma (continued):

We are given a sequence \((t_i)_{i=0}^{\ell}\) such that \(t_{i-1} = t_i + \nu_i \xi_i, \ i \geq 1,\) and

(a) \(f_{A,i-1}(t_{i-1}) \geq f_{A,i}(t_i)\) for all \(1 \leq i \leq \ell;\)

(b) \(f_{A,\ell}(t_\ell) > \max (L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty).\)

The goal: show \(|\{1 \leq i \leq \ell : \ t_{i-1} \text{ does not decay at time } i\}|\) is large.
Proof of the Lemma (continued):

We are given a sequence \((t_i)_{i=0}^\ell\) such that \(t_{i-1} = t_i + \nu_i \xi_i, \ i \geq 1\), and

(a) \(f_{A,i-1}(t_{i-1}) \geq f_{A,i}(t_i)\) for all \(1 \leq i \leq \ell\);

(b) \(f_{A,\ell}(t_\ell) > \max (L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty)\).

The goal: show \(\left| \{1 \leq i \leq \ell : t_{i-1} \text{ does not decay at time } i \} \right|\) is large.

Assume that \(1 \leq i \leq \ell\) is such that \(t_{i-1}\) decays at time \(i\). We have

\[
f_{A,i}(t_i) = \frac{1}{2} \left( f_{A,i-1}(t_i - \nu_i \xi_i) + f_{A,i-1}(t_i + \nu_i \xi_i) \right)
= \frac{1}{2} \left( f_{A,i-1}(t_{i-1} - 2\nu_i \xi_i) + f_{A,i-1}(t_{i-1}) \right).
\]
Proof of the Lemma (continued):

We are given a sequence \((t_i)_{i=0}^{\ell}\) such that \(t_{i-1} = t_i + v_i \xi_i, \ i \geq 1\), and

(a) \(f_{A,i-1}(t_{i-1}) \geq f_{A,i}(t_i)\) for all \(1 \leq i \leq \ell\);

(b) \(f_{A,\ell}(t_\ell) > \max \left(L(N \sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_{\infty}\right)\).

The goal: show \(\left|\{1 \leq i \leq \ell : \ t_{i-1} \text{ does not decay at time } i\}\right|\) is large.

Assume that \(1 \leq i \leq \ell\) is such that \(t_{i-1}\) decays at time \(i\). We have

\[
f_{A,i}(t_i) = \frac{1}{2} \left(f_{A,i-1}(t_i - v_i \xi_i) + f_{A,i-1}(t_i + v_i \xi_i)\right)
= \frac{1}{2} \left(f_{A,i-1}(t_{i-1} - 2v_i \xi_i) + f_{A,i-1}(t_{i-1})\right).
\]

By definition of decay at time \(i\), both \(f_{A,i-1}(t_{i-1} + 2\xi_i)\) and \(f_{A,i-1}(t_{i-1} - 2\xi_i)\) are less than \(\frac{\varepsilon L}{N \sqrt{n}}\), hence less than \(\varepsilon f_{A,i-1}(t_{i-1})\). Thus,

\[
f_{A,i}(t_i) \leq (1/2 + \varepsilon/2) f_{A,i-1}(t_{i-1}).
\]
Randomized averaging: preprocessing step

Proof of the Lemma (continued):

We are given a sequence \((t_i)_{i=0}^\ell\) such that \(t_{i-1} = t_i + \nu_i \xi_i, \ i \geq 1\), and

(a) \(f_{A,i-1}(t_{i-1}) \geq f_{A,i}(t_i)\) for all \(1 \leq i \leq \ell\);

(b) \(f_{A,\ell}(t_\ell) > \max \left( L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty \right)\).

The goal: show \(|\{1 \leq i \leq \ell : t_{i-1} \text{ does not decay at time } i\}|\) is large.

Assume that \(1 \leq i \leq \ell\) is such that \(t_{i-1}\) decays at time \(i\). We have

\[
f_{A,i}(t_i) = \frac{1}{2} \left( f_{A,i-1}(t_i - \nu_i \xi_i) + f_{A,i-1}(t_i + \nu_i \xi_i) \right)
   = \frac{1}{2} \left( f_{A,i-1}(t_{i-1} - 2\nu_i \xi_i) + f_{A,i-1}(t_{i-1}) \right).
\]

By definition of decay at time \(i\), both \(f_{A,i-1}(t_{i-1} + 2\xi_i)\) and \(f_{A,i-1}(t_{i-1} - 2\xi_i)\) are less than \(\frac{\varepsilon L}{N\sqrt{n}}\), hence less than \(\varepsilon f_{A,i-1}(t_{i-1})\). Thus,

\[
f_{A,i}(t_i) \leq (1/2 + \varepsilon/2) f_{A,i-1}(t_{i-1}).
\]

Applying the relation for all \(i\) when \(t_{i-1}\) decays, we get for \(u = |\{1 \leq i \leq \ell : t_{i-1} \text{ decays at time } i\}|\):

\[
(1/2 + \varepsilon)^\ell \|f\|_\infty < f_{A,\ell}(t_\ell) \leq (1/2 + \varepsilon/2)^u \|f\|_\infty.
\]

This implies a required estimate.
Preprocessing step (to be completed)

For any $M > 0$, $\delta \in (0, 1)$ there are $L = L(M, \delta) > 0$ and $n_0 = n_0(M, \delta) \in \mathbb{N}$ with the following property. Let $f \in \ell_1(\mathbb{Z})$ with $\|f\|_1 = 1$ and $f(t) > 0$, let $n \geq n_0$, $n/2 \leq \ell \leq n$, and let $A$ be as above for some $N \leq 2^n$. Then

$$\mathbb{P}\{\|f_{A,\ell}\|_\infty > \max (L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)\ell \|f\|_\infty) \} \leq \exp(-Mn).$$
Preprocessing step (to be completed)

For any $M > 0$, $\delta \in (0, 1)$ there are $L = L(M, \delta) > 0$ and $n_0 = n_0(M, \delta) \in \mathbb{N}$ with the following property. Let $f \in \ell_1(\mathbb{Z})$ with $\|f\|_1 = 1$ and $f(t) > 0$, let $n \geq n_0$, $n/2 \leq \ell \leq n$, and let $A$ be as above for some $N \leq 2^n$. Then

$$\mathbb{P}\{\|f_{A,\ell}\|_\infty > \max\left(L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty\right)\} \leq \exp(-Mn).$$

Lemma proved earlier

Define event $\mathcal{E}$ as the subset the probability space such that there exists a sequence $(v_i)_{i=1}^\ell \in \{-1, 1\}^\ell$ and a point $t \in \mathbb{Z}$ so that the descendant sequence $(t_i)_{i=0}^\ell$ for $t$ with respect to $(v_i)_{i=1}^\ell$ (i.e. $t_i = t - \sum_{j=1}^i v_j \xi_j$, $1 \leq i \leq \ell$, and $t_0 := t$) satisfies

$$\left|\left\{1 \leq i \leq \ell : t_{i-1} \text{ does not decay at time } i\right\}\right| \geq -\frac{n \log \frac{1/2+\varepsilon}{1/2+\varepsilon/2}}{2 \log \left(1/2+\varepsilon/2\right)}.$$

Then $\mathcal{E} \supset \{\|f_{A,\ell}\|_\infty > \max\left(L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)^\ell \|f\|_\infty\right)\}$. 

34
For any $M > 0$, $\delta \in (0, 1)$ there are $L = L(M, \delta) > 0$ and $n_0 = n_0(M, \delta) \in \mathbb{N}$ with the following property. Let $f \in \ell_1(\mathbb{Z})$ with $\|f\|_1 = 1$ and $f(t) > 0$, let $n \geq n_0$, $n/2 \leq \ell \leq n$, and let $A$ be as above for some $N \leq 2^n$. Then

$$\mathbb{P}\{\|f_{A,\ell}\|_\infty > \max(L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)\ell \|f\|_\infty)\} \leq \exp(-Mn).$$

Define event $E$ as the subset the probability space such that there exists a sequence $(v_i)_{i=1}^\ell \in \{-1, 1\}^\ell$ and a point $t \in \mathbb{Z}$ so that the descendant sequence $(t_i)_{i=0}^\ell$ for $t$ with respect to $(v_i)_{i=1}^\ell$ (i.e. $t_i = t - \sum_{j=1}^i v_j \xi_j$, $1 \leq i \leq \ell$, and $t_0 := t$) satisfies

$$\left|\{1 \leq i \leq \ell : t_{i-1} \text{ does not decay at time } i\}\right| \geq -\frac{n \log 1/2 + \varepsilon/2}{2 \log (1/2 + \varepsilon/2)}.$$

Then $E \supset \{\|f_{A,\ell}\|_\infty > \max(L(N\sqrt{n})^{-1}, (1/2 + \varepsilon)\ell \|f\|_\infty)\}$.

Hence, to complete the preprocessing step, it is enough to show that $\mathbb{P}(E) \leq \exp(-Mn)$. 
Proof of the preprocessing step (completion):
Recall that \((\xi_1, \ldots, \xi_n)\) is uniformly distributed on
\[
\mathcal{A} := \{-2N, \ldots, -N - 1, N + 1, \ldots, 2N\}^{\delta n} \times \{-N, -N + 1, \ldots, N\}^{n-\delta n}.
\]
Proof of the preprocessing step (completion):
Recall that \((\xi_1, \ldots, \xi_n)\) is uniformly distributed on
\[
A := \{-2N, \ldots, -N - 1, N + 1, \ldots, 2N\}^{\delta n} \times \{-N, -N + 1, \ldots, N\}^{n-\delta n}.
\]

It can be shown that for any point \(t \in \mathbb{Z}\) such that the last element of a descendant sequence \((t_i)_{i=0}^\ell\) (with respect to some sequence in \(\{-1, 1\}^\ell\) and with \(t_0 = t\)) satisfies \(f_{A, \ell}(t_\ell) > (N\sqrt{n})^{-1}\), we have \(t \in \{s \in \mathbb{Z} : f(s) > (N\sqrt{n})^{-1}\} + Cn \{-N, -N + 1, \ldots, N - 1, N\} \).
Proof of the preprocessing step (completion):
Recall that $(\xi_1, \ldots, \xi_n)$ is uniformly distributed on

$$\mathcal{A} := \{-2N, \ldots, -N - 1, N + 1, \ldots, 2N\}^{\delta_n} \times \{-N, -N + 1, \ldots, N\}^{n-\delta_n}.$$ 

It can be shown that for any point $t \in \mathbb{Z}$ such that the last element of a descendant sequence $(t_i)_{i=0}^{\ell}$ (with respect to some sequence in $\{-1, 1\}^{\ell}$ and with $t_0 = t$) satisfies $f_{A, \ell}(t_\ell) > (N\sqrt{n})^{-1}$, we have $t \in \{s \in \mathbb{Z} : f(s) > (N\sqrt{n})^{-1}\} + Cn \{-N, -N + 1, \ldots, N - 1, N\}$.

Then, setting $D$ to be the above set,

$$\mathbb{P}(\mathcal{E}) \leq 2^\ell |D| \sup_{t \in D, (v_i)_{i=1}^{\ell} \in \{-1, 1\}^{\ell}} \mathbb{P}\{\text{The descendant sequence } (t_i)_{i=0}^{\ell}\text{ for } t \text{ w.r.t } (v_i)_{i=1}^{\ell}\text{ satisfies } |\{1 \leq i \leq \ell : t_{i-1} \text{ does not decay at } i\}| \gtrsim n\}.$$
Proof of the preprocessing step (completion):
Recall that \((\xi_1, \ldots, \xi_n)\) is uniformly distributed on

\[ A := \{-2N, \ldots, -N-1, N+1, \ldots, 2N\}^{\delta n} \times \{-N, -N+1, \ldots, N\}^{n-\delta n}. \]

- It can be shown that for any point \(t \in \mathbb{Z}\) such that the last element of a descendant sequence \((t_i)_{i=0}^{\ell}\) (with respect to some sequence in \(\{-1, 1\}^\ell\) and with \(t_0 = t\)) satisfies \(f_{A,\ell}(t_\ell) > (N\sqrt{n})^{-1}\), we have \(t \in \{s \in \mathbb{Z} : f(s) > (N\sqrt{n})^{-1}\} + Cn \{-N, -N+1, \ldots, N-1, N\}\).
- Then, setting \(D\) to be the above set,

\[
\mathbb{P}(\mathcal{E}) \leq 2^\ell |D| \sup_{t \in D, (v_i)_{i=1}^{\ell} \in \{-1, 1\}^\ell} \mathbb{P}\{\text{The descendant sequence } (t_i)_{i=0}^{\ell} \text{ for } t \text{ w.r.t } (v_i)_{i=1}^{\ell} \text{ satisfies } |\{1 \leq i \leq \ell : t_{i-1} \text{ does not decay at } i\}| \gtrsim n\}.
\]
- The proof of the estimate for \(\mathbb{P}(\mathcal{E})\) is then completed by taking the union bound.
The rest of the proof of the weak Bernoulli singularity conjecture
Recall that the strong Bernoulli singularity conjecture (still open) asserts that \( \mathbb{P}\{B_n \text{ is singular}\} = (1 + o(1))n^22^{-n} \), where \( B_n \) is the \( n \times n \) random matrix with i.i.d \( \pm 1 \) entries. The conjecture can be modified to the setting of matrices with i.i.d Bernoulli(\( p \)) (0/1) entries:
Recall that the strong Bernoulli singularity conjecture (still open) asserts that $\mathbb{P}\{B_n \text{ is singular}\} = (1 + o(1))n^2 2^{1-n}$, where $B_n$ is the $n \times n$ random matrix with i.i.d $\pm 1$ entries. The conjecture can be modified to the setting of matrices with i.i.d Bernoulli($p$) (0/1) entries:

Singularity of Bernoulli($p$) random matrices

Let $(p_n)$ be a sequence of numbers in $(0, 1)$, and for each $n$, let $M_n$ be an $n \times n$ random matrix with i.i.d Bernoulli($p_n$) entries. Is it true that

$\mathbb{P}\{M_n \text{ is singular}\} = (1 + o_n(1))\mathbb{P}\{\text{There is a pair of linearly dependent rows or columns in } M_n\}$?
The strong Bernoulli conjecture for Bernoulli($p$) matrices

Recall that the strong Bernoulli singularity conjecture (still open) asserts that $\mathbb{P}\{B_n \text{ is singular}\} = (1 + o(1))n^22^{1-n}$, where $B_n$ is the $n \times n$ random matrix with i.i.d $\pm 1$ entries. The conjecture can be modified to the setting of matrices with i.i.d Bernoulli($p$) (0/1) entries:

**Singularity of Bernoulli($p$) random matrices**

Let $(p_n)$ be a sequence of numbers in $(0, 1)$, and for each $n$, let $M_n$ be an $n \times n$ random matrix with i.i.d Bernoulli($p_n$) entries. Is it true that

$$\mathbb{P}\{M_n \text{ is singular}\} = (1 + o_n(1))\mathbb{P}\{\text{There is a pair of linearly dependent rows or columns in } M_n\}?$$

In the regime $p_n < 1/2$, the conjecture can be rewritten more explicitly:

**Singularity of Bernoulli($p$) random matrices for $p < 1/2$**

Let $\varepsilon > 0$, let $(p_n)$ be a sequence of numbers in $(0, 1/2 - \varepsilon)$, and for each $n$, let $M_n$ be an $n \times n$ random matrix with i.i.d Bernoulli($p_n$) entries. Is it true that

$$\mathbb{P}\{M_n \text{ is singular}\} = \mathbb{P}\{\text{There is a zero row or zero column of } M_n\}?$$
The strong Bernoulli conjecture for Bernoulli($p$) matrices

Basak and Rudelson (2018) considered the setting of Bernoulli($p_n$) matrices, and resolved the conjecture in the regime $n p_n \approx \log n$.

**Theorem (Basak–Rudelson)**

Assume that $np_n \leq \ln n + O(\ln \ln n)$. Let $M_n$ be an $n \times n$ random matrix with i.i.d Bernoulli($p_n$) entries. Then

$$
P\{M_n \text{ is singular}\} = (1 + o_n(1))P\{\text{There is a zero row or column in } M_n\}.
$$
Basak and Rudelson (2018) considered the setting of Bernoulli($p_n$) matrices, and resolved the conjecture in the regime $n p_n \approx \log n$.

**Theorem (Basak–Rudelson)**

Assume that $np_n \leq \ln n + O(\ln \ln n)$. Let $M_n$ be an $n \times n$ random matrix with i.i.d Bernoulli($p_n$) entries. Then

$$\mathbb{P}\{M_n \text{ is singular}\} = (1 + o_n(1))\mathbb{P}\{\text{There is a zero row or column in } M_n\}.$$

Further, a recent result in collaboration with A.Litvak resolved the strong Bernoulli($p$) conjecture under assumption $\log n \ll np_n \ll n$:

**Theorem (A.E.Litvak–T., 20+)**

There are universal constants $C, c > 0$ with the following property. For each $n$, let $p_n$ be a number in $(0,1)$, and assume that $C \leq \liminf \frac{np_n}{\log n}$ and $\limsup p_n \leq c$. For each $n$, let $M_n$ be the $n \times n$ matrix with i.i.d. Bernoulli($p_n$) entries. Then

$$\mathbb{P}\{M_n \text{ is singular}\} = (1 + o_n(1))\mathbb{P}\{M_n \text{ has a zero row or zero column}\}$$

$$= (2 + o_n(1)) n (1 - p_n)^n.$$
Singularity of sparse Bernoulli($p$) matrices: improved anti-concentration for dependent variables

The proof of the joint result with A. Litvak starts with the following simple observation.
The proof of the joint result with A. Litvak starts with the following simple observation.

First, assume that \( b_1, \ldots, b_n \) are i.i.d Bernoulli(\( p \)) random variables (\( p \) is small), and let \( x = (x_1, x_2, \ldots, x_n) \) be any fixed vector in \( \mathbb{R}^n \). Then, regardless of the structure of \( x \), we have

\[
\max_{\lambda \in \mathbb{R}} \mathbb{P}\left\{ \sum_{i=1}^{n} b_i x_i = \lambda \right\} \geq \mathbb{P}\{ \text{all } b_i \text{ are zero} \} = (1 - p)^n \approx e^{-pn}.
\]
The proof of the joint result with A. Litvak starts with the following simple observation.

First, assume that $b_1, \ldots, b_n$ are i.i.d Bernoulli($p$) random variables ($p$ is small), and let $x = (x_1, x_2, \ldots, x_n)$ be any fixed vector in $\mathbb{R}^n$. Then, regardless of the structure of $x$, we have

$$\max_{\lambda \in \mathbb{R}} \mathbb{P}\left\{ \sum_{i=1}^{n} b_i x_i = \lambda \right\} \geq \mathbb{P}\{\text{all } b_i \text{ are zero}\} = (1 - p)^n \approx e^{-pn}.$$ 

Now, let $b_1, \ldots, b_n$ are Bernoulli($p$) variables such that $(b_1, \ldots, b_n)$ is uniformly distributed on the set of 0/1 vectors with $pn$ ones. Let $x = (x_1, x_2, \ldots, x_n)$ be a vector in $\mathbb{R}^n$ in “general” position (so that no two sign combinations of its coordinates are equal). Then

$$\max_{\lambda \in \mathbb{R}} \mathbb{P}\left\{ \sum_{i=1}^{n} b_i x_i = \lambda \right\} = \left(\frac{n}{pn}\right)^{-1} \approx p^{pn}.$$
The proof of the joint result with A. Litvak starts with the following simple observation.

First, assume that \( b_1, \ldots, b_n \) are i.i.d Bernoulli\((p)\) random variables (\( p \) is small), and let \( x = (x_1, x_2, \ldots, x_n) \) be any fixed vector in \( \mathbb{R}^n \). Then, regardless of the structure of \( x \), we have

\[
\max_{\lambda \in \mathbb{R}} \mathbb{P}\left\{ \sum_{i=1}^{n} b_i x_i = \lambda \right\} \geq \mathbb{P}\{ \text{all } b_i \text{ are zero} \} = (1 - p)^n \approx e^{-pn}.
\]

Now, let \( b_1, \ldots, b_n \) are Bernoulli\((p)\) variables such that \((b_1, \ldots, b_n)\) is uniformly distributed on the set of 0/1 vectors with \( pn \) ones. Let \( x = (x_1, x_2, \ldots, x_n) \) be a vector in \( \mathbb{R}^n \) in “general” position (so that no two sign combinations of its coordinates are equal). Then

\[
\max_{\lambda \in \mathbb{R}} \mathbb{P}\left\{ \sum_{i=1}^{n} b_i x_i = \lambda \right\} = \binom{n}{pn}^{-1} \approx p^{pn}.
\]

When \( p \) is small, conditioning on the sum of \( b_i \)'s allows to prove much stronger anti-concentration inequalities compared to the i.i.d setting.
Singularity of sparse Bernoulli($p$) matrices: elements of the proof

The proof of the estimate $\mathbb{P}\{M_n \text{ is singular} \} = (2 + o_n(1)) n (1 - p_n)^n$ in the regime $\log n \ll np_n \ll n$ is based on

- Special covering arguments to treat invertibility of $M_n$ over “close to sparse” unit vectors;
- Conditioning on column-sums of $M_n$;
- A Littlewood–Offord–type inequality for linear combinations of dependent Bernoulli variables;
- A double counting procedure for estimating cardinalities of certain discrete subsets of $\mathbb{R}^n$. 
Singularity of sparse Bernoulli($p$) matrices: elements of the proof

Given $n \in \mathbb{N}$, $1 \leq m \leq n/2$, a vector $y \in \mathbb{R}^n$ and parameters $K \geq 1$, we define the degree of unstructuredness of vector $y$ as

$$
\text{UD}_{n}(y, m, K) := \sup \left\{ t > 0 : \left( \left\lfloor \frac{n}{m} \right\rfloor ! \right)^m \left( n - m \lfloor n/m \rfloor \right)! \frac{n!}{n!} \cdot \sum_{S_1, \ldots, S_m} \int_{-t}^{t} \prod_{i=1}^{m} \psi\left( \left| \mathbb{E} \exp \left( 2\pi i y_{\eta[S_i]} m^{-1/2} s \right) \right| \right) ds \leq K \right\},
$$

where the sum is taken over all sequences $(S_i)_{i=1}^{m}$ of disjoint subsets $S_1, \ldots, S_m \subset [n]$, each of cardinality $\lfloor n/m \rfloor$. Here $\eta[S_i], i \leq m$, denote mutually independent integer random variables uniformly distributed on respective $S_i$’s. The function $\psi$ in the definition acts as a smoothing of $\max(c, t)$ for certain small positive constant $c$. 
Singularity of sparse Bernoulli($p$) matrices: elements of the proof

\[
\text{UD}_n(y, m, K) := \sup \left\{ t > 0 : \frac{((\lfloor n/m \rfloor)!)^m (n - m\lfloor n/m \rfloor)!}{n!} \cdot \sum_{S_1, \ldots, S_m} \int_{-t}^{t} \prod_{i=1}^{m} \psi(\|\mathbb{E} \exp(2\pi i y_{\eta[S_i] \cdot m^{-1/2} s})\|) \, ds \leq K \right\}.
\]

The functional $\text{UD}_n(y, m, K)$ can be interpreted as follows. The expression inside the supremum is the average value of the integral

\[
\int_{-t}^{t} \prod_{i=1}^{m} \psi(\|\mathbb{E} \exp(2\pi i y_{\eta[S_i] \cdot m^{-1/2} s})\|) \, ds,
\]

with the average taken over all choices of sequences $(S_i)_{i=1}^{m}$. The function under the integral, disregarding the smoothing $\psi$, is the absolute value of the characteristic function of the random variable $\langle y, Z \rangle$, where $Z$ is a random 0/1–vector with exactly $m$ ones, and with the $i$-th “1” distributed uniformly on $S_i$. 

42
Singularity of sparse Bernoulli($p$) matrices: elements of the proof

$$\textbf{UD}_n(y, m, K) := \sup \left\{ t > 0 : \frac{((\lfloor n/m \rfloor)!)^m (n - m\lfloor n/m \rfloor)!}{n!} \cdot \sum_{S_1, \ldots, S_m} \int_t^{-t} \prod_{i=1}^m \psi \left( |E \exp \left( 2\pi i y_{\eta[S_i]} m^{-1/2} s \right) | \right) ds \leq K \right\}.$$  

The definition of $\textbf{UD}_n(\cdot)$ and a lemma of Esseen imply

A Littlewood–Offord-type inequality in terms of the unstructuredness

Let $m, n$ be positive integers with $m \leq n/2$, and let $K \geq 1$. Further, let $\nu \in \mathbb{R}^n$, and let $X = (X_1, \ldots, X_n)$ be a random $0/1$–vector in $\mathbb{R}^n$ uniformly distributed on the set of vectors with $m$ ones and $n - m$ zeros. Then for some $C > 0$ depending only on $K$,

$$\mathcal{L} \left( \sum_{i=1}^n \nu_i X_i, \sqrt{m} \tau \right) \leq C \left( \tau + \textbf{UD}_n(\nu, m, K)^{-1} \right)$$  

for all $\tau > 0$. 

43
Singularity of sparse Bernoulli($p$) matrices: comparison of $\textbf{UD}_n$ and the Least Common Denominator

The Least Common Denominator of Rudelson–Vershynin was defined as

$$\text{LCD}(y) := \inf \{ \theta > 0 : \text{dist}(\theta y, \mathbb{Z}^n) < \min(c_1\|\theta y\|_2, c_2\sqrt{n}) \},$$

where $c_1, c_2 > 0$ are two small constants. The bigger LCD is, the larger is the interval of the values of $\theta$ on which the rescaled vector $\theta y$ is far from the integer lattice.
The Least Common Denominator of Rudelson–Vershynin was defined as

$$\text{LCD}(y) := \inf \{ \theta > 0 : \text{dist}(\theta y, \mathbb{Z}^n) < \min(c_1\|\theta y\|_2, c_2\sqrt{n}) \},$$

where $c_1, c_2 > 0$ are two small constants. The bigger LCD is, the larger is the interval of the values of $\theta$ on which the rescaled vector $\theta y$ is far from the integer lattice.

The definition of $\text{UD}_n$ is “raw”; it does not refer directly to the structure of $y$ but instead is defined in terms of the integral of the characteristic function of some auxiliary random variable. It seems reasonable to expect that the degree of unstructuredness $\text{UD}_n$ can be replaced with a nicer function of distances of rescaled coordinate projections of $y$ to respective integer lattices.
Singularity of sparse Bernoulli$(p)$ matrices: comparison of $\textbf{UD}_n$ and the Least Common Denominator

The Least Common Denominator of Rudelson–Vershynin was defined as

$$\text{LCD}(y) := \inf \{ \theta > 0 : \text{dist}(\theta y, \mathbb{Z}^n) < \min(c_1\|\theta y\|_2, c_2\sqrt{n}) \},$$

where $c_1, c_2 > 0$ are two small constants. The bigger LCD is, the larger is the interval of the values of $\theta$ on which the rescaled vector $\theta y$ is far from the integer lattice.

The definition of $\textbf{UD}_n$ is “raw”; it does not refer directly to the structure of $y$ but instead is defined in terms of the integral of the characteristic function of some auxiliary random variable. It seems reasonable to expect that the degree of unstructuredness $\textbf{UD}_n$ can be replaced with a nicer function of distances of rescaled coordinate projections of $y$ to respective integer lattices.

While “stability” of LCD with respect to small perturbations of the vector is easy to establish, proving the same property for $\textbf{UD}_n$ requires additional work. The smoothing function $\psi$ was introduced for that purpose. The stability is crucial when applying approximation ($\varepsilon$–net) arguments.
Singularity of sparse Bernoulli($p$) matrices: anti-concentration on a lattice in terms of $\text{UD}_n$

**Theorem (Anti-concentration on a lattice in terms of $\text{UD}_n$)**

Let $M \geq 1$, $\rho, \delta \in (0, 1/4]$. There exist $K = K(\delta, \rho) \geq 1$, $n_0 = n_0(M, \delta, \rho)$, and $C = C(M, \delta, \rho) \in \mathbb{N}$ with the following property.
Singularity of sparse Bernoulli($p$) matrices: anti-concentration on a lattice in terms of $\text{UD}_n$

**Theorem (Anti-concentration on a lattice in terms of $\text{UD}_n$)**

Let $M \geq 1$, $\rho, \delta \in (0, 1/4]$. There exist $K = K(\delta, \rho) \geq 1$, $n_0 = n_0(M, \delta, \rho)$, and $C = C(M, \delta, \rho) \in \mathbb{N}$ with the following property.

Let $\sigma$ be a permutation of $[n]$, $h \in \mathbb{R}$, and let $Q_1, Q_2 \subset [n]$ be disjoint subsets such that $|Q_1|, |Q_2| = \lceil \delta n \rceil$. Then

$$\text{P}\{\text{UD}_n(X, m, K) < \frac{km}{2^C}\} \leq e^{-M n^{0.45}}.$$
Singularity of sparse Bernoulli(\( p \)) matrices: anti-concentration on a lattice in terms of \( \text{UD}_n \)

**Theorem (Anti-concentration on a lattice in terms of \( \text{UD}_n \))**

Let \( M \geq 1, \rho, \delta \in (0, 1/4] \). There exist \( K = K(\delta, \rho) \geq 1, n_0 = n_0(M, \delta, \rho) \), and \( C = C(M, \delta, \rho) \in \mathbb{N} \) with the following property.

Let \( \sigma \) be a permutation of \([n] \), \( h \in \mathbb{R} \), and let \( Q_1, Q_2 \subset [n] \) be disjoint subsets such that \(|Q_1|, |Q_2| = \lceil \delta n \rceil \).

Let \( n \geq n_0, m \geq C \) with \( n/m \geq C, k \geq 1 \).
Singularity of sparse Bernoulli($p$) matrices: anti-concentration on a lattice in terms of $\text{UD}_n$

**Theorem (Anti-concentration on a lattice in terms of $\text{UD}_n$)**

Let $M \geq 1$, $\rho, \delta \in (0, 1/4]$. There exist $K = K(\delta, \rho) \geq 1$, $n_0 = n_0(M, \delta, \rho)$, and $C = C(M, \delta, \rho) \in \mathbb{N}$ with the following property.

Let $\sigma$ be a permutation of $[n]$, $h \in \mathbb{R}$, and let $Q_1, Q_2 \subset [n]$ be disjoint subsets such that $|Q_1|, |Q_2| = \lceil \delta n \rceil$.

Let $n \geq n_0$, $m \geq C$ with $n/m \geq C$, $k \geq 1$.

Let $X = (X_1, \ldots, X_n)$ be a random vector uniformly distributed on the set

$$\Lambda_n := \left\{ x \in \frac{1}{k} \mathbb{Z}^n : |x_{\sigma(i)}| \leq (2n/i)^{100} \text{ for all } i \leq n; \right.\]

$$\left. \min_{i \in Q_1} x_i \geq h, \text{ and } \max_{i \in Q_2} x_i \leq h - \rho \right\}.$$
Singularity of sparse Bernoulli($p$) matrices: anti-concentration on a lattice in terms of $\text{UD}_n$

**Theorem (Anti-concentration on a lattice in terms of $\text{UD}_n$)**

Let $M \geq 1$, $\rho, \delta \in (0, 1/4]$. There exist $K = K(\delta, \rho) \geq 1$, $n_0 = n_0(M, \delta, \rho)$, and $C = C(M, \delta, \rho) \in \mathbb{N}$ with the following property.

Let $\sigma$ be a permutation of $[n]$, $h \in \mathbb{R}$, and let $Q_1, Q_2 \subset [n]$ be disjoint subsets such that $|Q_1|, |Q_2| = \lceil \delta n \rceil$.

Let $n \geq n_0$, $m \geq C$ with $n/m \geq C$, $k \geq 1$.

Let $X = (X_1, \ldots, X_n)$ be a random vector uniformly distributed on the set

$$
\Lambda_n := \left\{ x \in \frac{1}{k} \mathbb{Z}^n : |x_{\sigma(i)}| \leq (2n/i)^{100} \text{ for all } i \leq n;
\right. \\
\left. \min_{i \in Q_1} x_i \geq h, \text{ and } \max_{i \in Q_2} x_i \leq h - \rho \right\}.
$$

Then

$$
\mathbb{P}\left\{ \text{UD}_n(X, m, K) < km^{1/2} / C \right\} \leq e^{-Mn}.
$$
Let, as before, $M_n$ denote the $n \times n$ matrix with i.i.d Bernoulli($p_n$) entries. In the regime $n p_n = \Theta(\log n)$, the relation $\mathbb{P}\{M_n \text{ is singular}\} = (1 + o_n(1)) \mathbb{P}\{M_n \text{ has a zero row or column}\}$ has very recently been confirmed by H.Huang. Together with the result of A.Litvak–T., this resolves the strong Bernoulli singularity conjecture for sufficiently sparse matrices.
Let, as before, $M_n$ denote the $n \times n$ matrix with i.i.d Bernoulli($p_n$) entries. In the regime $np_n = \Theta(\log n)$, the relation $\mathbb{P}\{ M_n \text{ is singular} \} = (1 + o_n(1)) \mathbb{P}\{ M_n \text{ has a zero row or column} \}$ has very recently been confirmed by H.Huang. Together with the result of A.Litvak–T., this resolves the strong Bernoulli singularity conjecture for sufficiently sparse matrices.

Singularity of dense Bernoulli($p$) matrices below 1/2 threshold

Assume that $p_n = \Theta(1)$, and $\limsup p_n < 1/2$. Is it true that

$$\mathbb{P}\{ M_n \text{ is singular} \} = (1 + o_n(1)) \mathbb{P}\{ M_n \text{ has a zero row or column} \}
= (2 + o(1)) n (1 - p_n)^n?$$
Further directions

Let, as before, $M_n$ denote the $n \times n$ matrix with i.i.d Bernoulli($p_n$) entries. In the regime $np_n = \Theta(\log n)$, the relation $\mathbb{P}\{M_n \text{ is singular}\} = (1 + o_n(1))\mathbb{P}\{M_n \text{ has a zero row or column}\}$ has very recently been confirmed by H.Huang. Together with the result of A.Litvak–T., this resolves the strong Bernoulli singularity conjecture for sufficiently sparse matrices.

Singularity of dense Bernoulli($p$) matrices below 1/2 threshold

Assume that $p_n = \Theta(1)$, and $\lim \sup p_n < 1/2$. Is it true that

$$\mathbb{P}\{M_n \text{ is singular}\} = (1 + o_n(1))\mathbb{P}\{M_n \text{ has a zero row or column}\}
= (2 + o(1))n (1 - p_n)^n?$$

Singularity of random matrices is “local”

Let $\xi$ be a random variable (perhaps with some moment growth conditions). For each $n$, let $A_n$ be a random matrix with i.i.d. entries equidistributed with $\xi$. Is it true that $\mathbb{P}\{A_n \text{ is singular}\} =$

$$(1 + o_n(1))\mathbb{P}\{\text{There are 2 columns or 2 rows of } A_n \text{ which are linearly dependent}\}?$$
Thank you!