Non-asymptotic theory of random matrices: extreme singular values

Mark Rudelson and Roman Vershynin

University of Michigan

International Congress of Mathematicians
Hyderabad, India, 2010

Random matrices in functional analysis

The need to understand spectra of random matrices appears in various areas of science and engineering:

- Statistics: Principal Component Analysis [Wishart'20...]
- Quantum mechanics: excitation spectra of nuclei [Wigner, Dyson 50-60...]
- Numerical analysis: average analysis of matrix algorithms [von Neumann'50, Smale'80...]
- Functional analysis: probabilistic constructions of normed spaces and linear operators [Milman'70, Gluskin'70...]:

Random matrices in functional analysis

- Randomized constructions of good linear operators in finite dimensional normed spaces $(\mathbb{R}^n, \|\cdot\|)$. This is a functional analytic version of the probabilistic method of P. Erdös in combinatorics (random graphs).
- Example: Kashin's theorem on Euclidean subspaces of L_1^N with the norm $\|f\|_{L_1} = \frac{1}{N}\sum_{i=1}^N |f(i)|$.

Euclidean subspaces of L_1 [Kashin'77]

For all $N = (1+\delta)n$, there exist subspace E of L_1^N of dimension n which is uniformly isomorphic to L_2^n:

$$\|f\|_{L_2} \lesssim_\delta \|f\|_{L_1} \le \|f\|_{L_2} \quad \text{for all } f \in E.$$

- Only randomized constructions of E are known: kernel or image of random Gaussian matrix with iid $N(0,1)$ entries; kernel or image of a random Bernoulli matrix with iid ± 1 entries [Kashin'77], [Litvak et al'05, Rudelson'06]

Limit laws in random matrix theory

Some classical results on the limiting spectrum of random matrices (as dimentions $\to \infty$):

- Wigner's semicircle law
- Marchenko-Pastur law
- Circular law

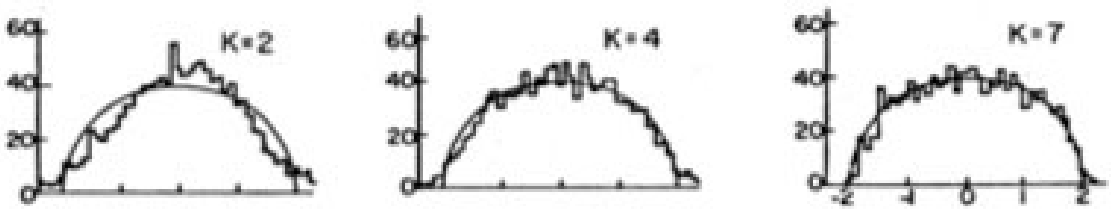
Limit laws in random matrix theory

A_n = random symmetric Gaussian matrix, whose above diagonal entries are independent $N(0,1)$.

Wigner's semicircle law '58

As dimension $n \to \infty$, the spectrum of $\frac{1}{\sqrt{n}} A_n$ is distributed according to the semicircle law with density

$$\frac{1}{2\pi}\sqrt{4-x^2} \quad \text{on } [-2,2].$$



[J. French, S. Wong, Phys. Lett. B. 35 (1971), 5]

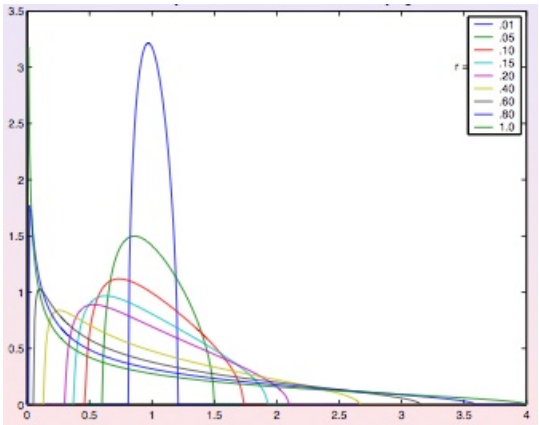
Limit laws in random matrix theory

$A = A_{N,n}$ is an $N \times n$ random Gaussian matrix with i.i.d. $N(0,1)$ entries; $W_{N,n} = A^*A$ is called Wishart matrix.

Marchenko-Pastur law '67

As the dimensions $N, n \to \infty$ while aspect ratio $n/N \to y \in (0,1]$, the spectrum of $\frac{1}{N} W_{N,n}$ has limiting density

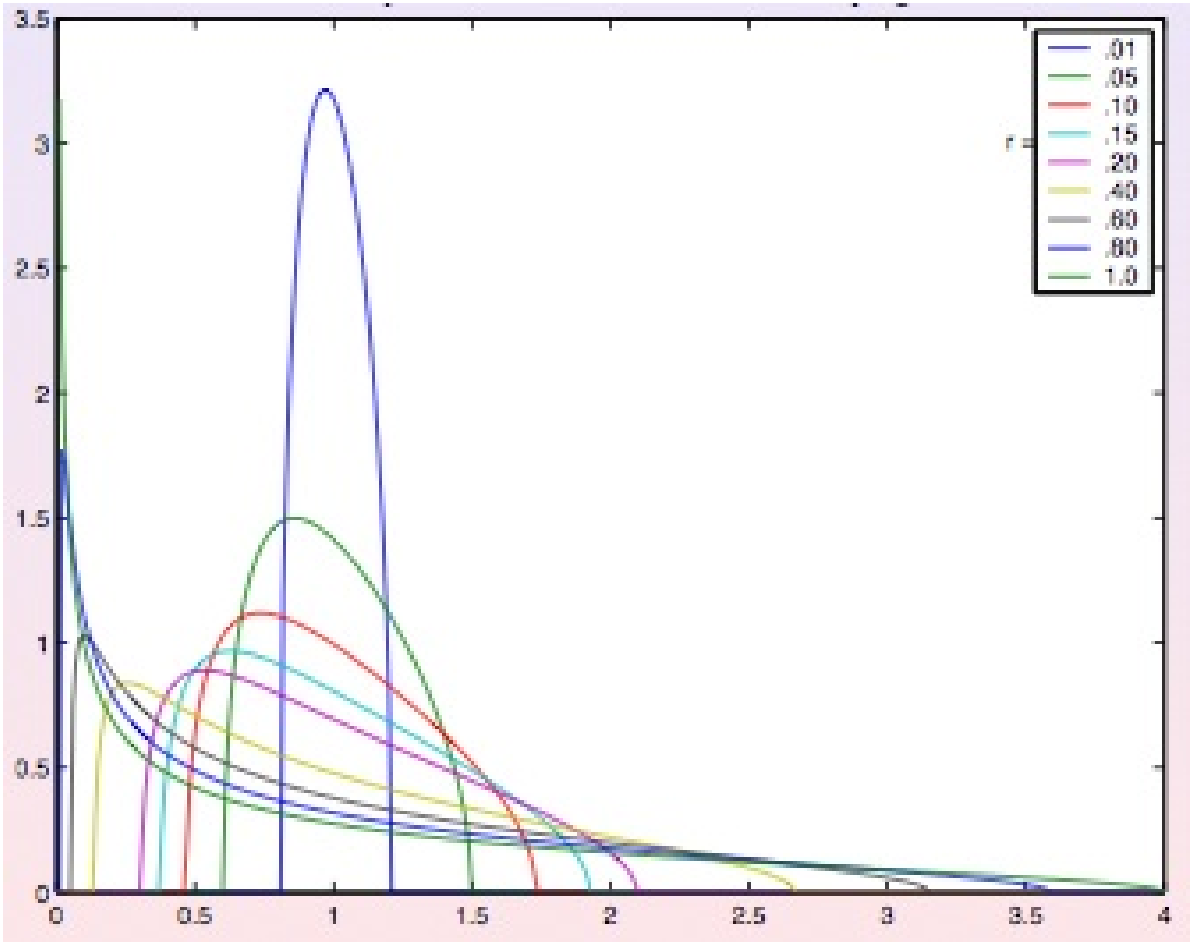
$$\frac{1}{2\pi xy}\sqrt{(b-x)(x-a)} \quad \text{where } a = (1-\sqrt{y})^2, b = (1+\sqrt{y})^2.$$



[El Karoui, Estimation of large dimensional sparse covariance matrices, 2009]

Limit laws in random matrix theory

Limiting density in Marchenko-Pastur law:



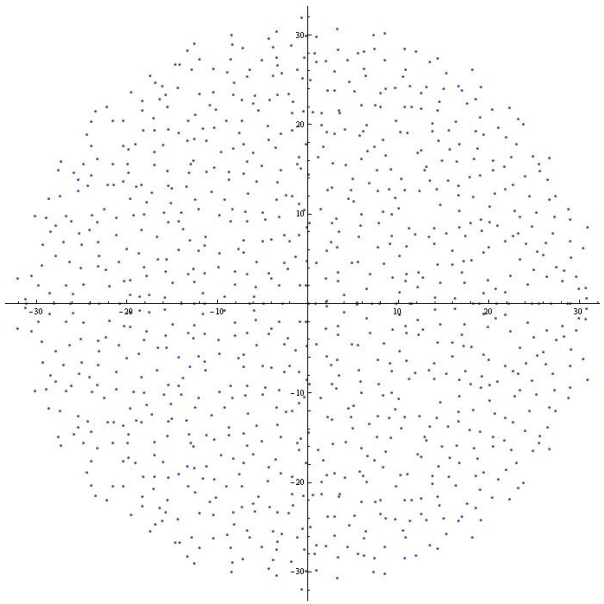
[El Karoui, Estimation of large dimensional sparse covariance matrices, 2009]

Limit laws in random matrix theory

$A_n =$ is an $n \times n$ random Gaussian matrix.

Circular law [Mehta'67]

As the dimension $n \to \infty$ the spectrum of $\frac{1}{\sqrt{n}} A_n$ is distributed according to the uniform measure on the unit disc $\{z \in \mathbb{C} : \ |z| = 1\}$.



[B.Valkó, A course on random matrices, `math.wisc.edu/~valko/courses/833/833.html`]

Universality

- It is widely believed that phenomena typically observed in statistical physics and in asymptotic random matrix theory are universal – independent of the distribution of the entries.
- Universality in classical probability: Central Limit Theorem. For i.i.d. random variables Z_i, the normalized sums

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i \to N(0,1)$$

  regardless of the distribution of Z_i.
- Universality in random matrix theory:
  - Wigner's semicircle law [Pastur'73, Bai-Silverstein'10]
  - Marchenko-Pastur law [Watcher'78, see Bai'99]
  - circular law [Girko'84, Bai'97, Götze-Tikhomirov'08, Tao-Vu'08-10]

Asymptotic and non-asymptotic regimes

- Asymptotic random matrix theory offers remarkable predictions as dimensions grow to infinity. This fits very well the purposes of statistical physics.
- However, there is often lack of understanding of finite (fixed but large) dimensions. Many applications operate there:
  - statistics (number of parameters is fixed),
  - numerical analysis of algorithms (number of variables and equations is fixed),
  - functional analysis (operators act on fixed spaces).
- Asymptotic regime = dimensions grow to infinity; precise limiting phenomena
- Non-asymptotic regime = any fixed dimensions; results are optimal up to constants

Extreme singular values

- Functional analytic view. Suppose we are given a linear operator A on a Banach space. Basic questions we ask are:
  - whether A is bounded, and find a bound $\|A\| \le M$
  - whether A is invertible, and find a bound $\|A^{-1}\| \le 1/m$
  - More generally, we are trying to obtain best bounds

$$m\|x\| \le \|Ax\| \le M\|x\| \quad \text{for all } x.$$

- General problem (largely open): estimate the norm, and the norm of inverse, of a random matrix $A : X \to Y$ with i.i.d. entries, acting as an operator between general Banach spaces.

Extreme singular values

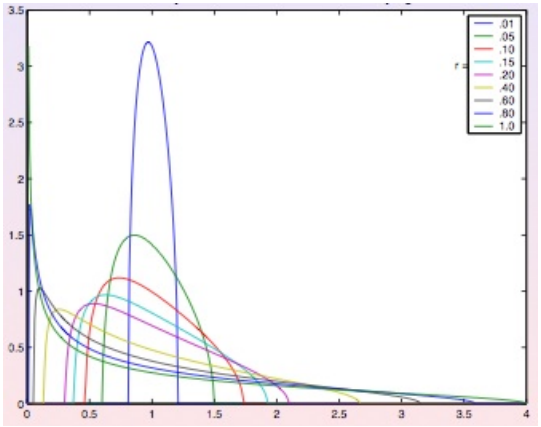
- Trying to obtain best bounds $m\|x\| \le \|Ax\| \le M\|x\|$ for all x.
- Simplest case: operators on Euclidean spaces. Then m, M are the extreme singular values of A (eigenvalues of $|A| = \sqrt{A^*A}$):

$$m = s_{\min}(A), \quad M = s_{\max}(A).$$

Problem. What are $s_{\min}(A)$, $s_{\max}(A)$ for random A, iid entries?

- And what is the condition number $\kappa(A) = s_{\max}(A)/s_{\min}(A)$? Important in the analysis of algorithms (linear solvers etc.)
- $s_{\max}(A) =$ "soft edge" of spectrum, easier to analyze.
  $s_{\min}(A) =$ "hard edge", more difficult.
- Technical assumptions throughout: entries have zero mean, unit variance, and subgaussian tails: $\mathbb{P}(|a_{ij}| > t) \le 2e^{-ct^2}$.

Soft edge: asymptotic regime



- Heuristics: from the endpoints of Marchenko-Pastur density,

$$s_{\min}(A) \sim \sqrt{N} - \sqrt{n}, \quad s_{\max}(A) \sim \sqrt{N} + \sqrt{n}.$$

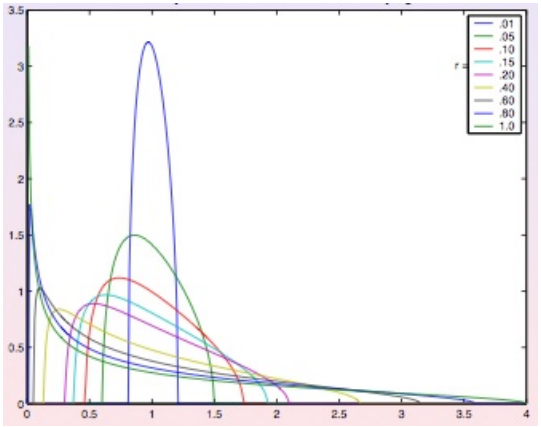
- This heuristics is valid in the asymptotic regime:

Bai-Yin law [Geman'80 for Gaussian,... see Bai-Yin'93]

As dimensions $N, n \to \infty$ while aspect ratio $n/N \to y \in (0, 1]$,

$$\frac{1}{\sqrt{N}} s_{\min}(A) \to 1 - \sqrt{y}, \quad \frac{1}{\sqrt{N}} s_{\max}(A) \to 1 + \sqrt{y} \quad \text{almost surely.}$$

Soft edge: non-asymptotic regime



- Heuristics: from the endpoints of Marchenko-Pastur density,

$$s_{\min}(A) \sim \sqrt{N} - \sqrt{n}, \quad s_{\max}(A) \sim \sqrt{N} + \sqrt{n}.$$

- This heuristics is valid in the non-asymptotic regime:

Soft edge: non-asymptotic bound

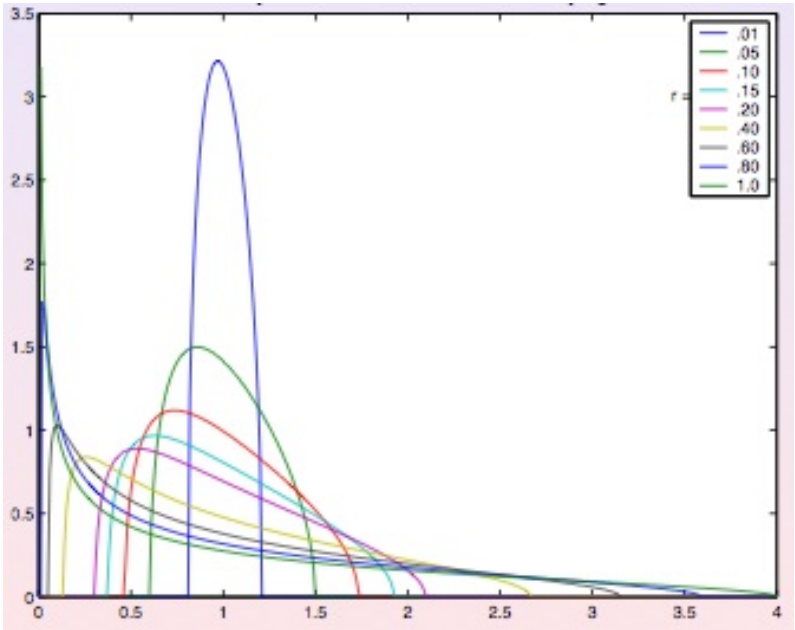
In all dimensions N, n, we have

$$\mathbb{E}\, s_{\max}(A) \le C(\sqrt{N} + \sqrt{n}).$$

Proof: a standard covering argument.

Hard edge: Gaussian matrices

- The hard edge is more difficult: $s_{\min}(A) \gtrsim \sqrt{N} - \sqrt{n}$?
- And even if true, it is useless for square matrices ($N = n$). They are ill-conditioned – Marchenko-Pastur density blows up:



- [von Neumann et al'48, Smale'85] predicted that $s_{\min}(A) \sim \sqrt{n} - \sqrt{n-1} \sim n^{-1/2}$.
- Since $s_{\max}(A) \sim n^{1/2}$, this would imply that the condition number of $n \times n$ random matrices is $\kappa(A) \sim n$.

Hard edge: Gaussian matrices

- Edelman-Szarek: von Neumann, Smale's prediction is valid for Gaussian matrices, $s_{\min}(A) \sim \sqrt{n} - \sqrt{n-1} \sim n^{-1/2}$:

Edelman's law [Edelman'88, Szarek'91, after Smale's question'85, von Neumann-Goldstine's prediction'47]

For square random Gaussian matrix $A = A_n$, for every fixed $\varepsilon \geq 0$:

$$\mathbb{P}\big(s_{\min}(A) \leq \varepsilon n^{-1/2}\big) \to 1 - \exp(-\varepsilon - \varepsilon^2/2) \quad \text{as } n \to \infty.$$

Non-asymptotically, in all dimensions n:

$$\mathbb{P}\big(s_{\min}(A) \leq \varepsilon n^{-1/2}\big) \leq \varepsilon \quad \text{for } \varepsilon \geq 0.$$

Hard edge: universality

- Edelman: $\mathbb{P}\big(s_{\min}(A) \le \varepsilon n^{-1/2}\big) \le \varepsilon$ for random Gaussian A. But the method is specific to Gaussian distribution (uses the explicit joint density of eigenvalues).

Conjecture [Spielman-Teng ICM'02]

Edelman's law holds for random Bernoulli matrix A, except for the singularity probability correction:

$$\mathbb{P}\big(s_{\min}(A) \le \varepsilon n^{-1/2}\big) \le \varepsilon + c^n$$

where $c \in (0,1)$ is a constant.

- [Kahn-Komlos-Szemeredi '95]: the singularity probability is indeed exponentially small: $\mathbb{P}\big(s_{\min}(A) = 0\big) \le c^n$.
- [R-V'08]: Spielman-Teng's Conjecture (and thus von Neumann, Smale prediction) holds up to a constant factor, and for general random matrices:

Hard edge: universality

Hard edge, non-asymptotic universality [R-V'08]

For general $n \times n$ random matrices A with iid entries (mean 0, variance 1, subgaussian),

$$\mathbb{P}\left(s_{\min}(A) \le \varepsilon n^{-1/2}\right) \le C\varepsilon + c^n \quad \text{for } \varepsilon \ge 0$$

where $C > 0$ and $c \in (0,1)$ are constants.

- Hence $s_{\min}(A) \sim n^{-1/2}$ whp ($\lesssim$ from above; $\gtrsim$ from [R-V'08a]). So the condition number or random square matrices is $\kappa(A) \sim n$. This validates von Neumann, Smale's predictions for general random matrices.
- Non-asymptotic universality: the result is independent of the distribution of entries.
- The best constants are unknown: $C = 1$ and $c = 1/2 + o(1)$?

Smallest singular value: general random matrices

Hard edge, asymptotic universality [Tao-Vu'09]

For every $\varepsilon \geq 0$,

$$\mathbb{P}\big(s_{\min}(A) \leq \varepsilon n^{-1/2}\big) \to \mathbb{P}\big(s_{\min}(G) \leq \varepsilon n^{-1/2}\big) \quad \text{as } n \to \infty$$

where G is the $n \times n$ Gaussian random matrix.

Smallest singular value: general random matrices

- For rectangular matrices, $s_{\min}(A) \sim \sqrt{N} - \sqrt{n-1}$.
- This contains the result for square matrices: $s_{\min}(A) \sim \sqrt{n} - \sqrt{n-1} \sim n^{-1/2}$.

Smallest singular value of rectangular matrices [Rudelson-V'09]

For general $N \times n$ random matrices A with iid entries (mean 0, variance 1, subgaussian),

$$\mathbb{P}\big(s_{\min}(A) \le \varepsilon(\sqrt{N} - \sqrt{n-1})\big) \le (C\varepsilon)^{N-n+1} + c^N \quad \text{for } \varepsilon \ge 0$$

where $C > 0$ and $c \in (0,1)$ are constants.

- Optimal up to constants. New even for Gaussian matrices.
- Partial cases / weaker results [Bennett et al'77, Litvak et al'05, Artstein et al'06, Rudelson'06].

Tail distribution

Definition

A random variable ξ is called subgaussian if for any $t > 0$

$$\mathbb{P}\left(|\xi| > t\right) \leq 2e^{-\alpha t^2}.$$

Theorem

Let A be an $n \times n$ random matrix with i.i.d. subgaussian entries. Then for any $\varepsilon > 0$

$$\mathbb{P}\left(s_n(A) \leq \varepsilon n^{-1/2}\right) \leq C\varepsilon + c^{-n}.$$

First approach: ε-net argument

1. Individual estimate (probability):

$$\mathbb{P}\left(\|Ay\| < t\right) \text{ is small}$$

This bound should hold uniformly over $y \in S^{n-1}$.

First approach: ε-net argument

1. Individual estimate (probability):

$$\mathbb{P}\left(\|Ay\| < t\right) \text{ is small}$$

This bound should hold uniformly over $y \in S^{n-1}$.

2. Discretization (geometry): find a small set $\mathcal{N} \subset S^{n-1}$ such that

$$\forall x \in S^{n-1} \; \exists y \in \mathcal{N} \; \|x - y\| < \varepsilon.$$

Union bound:

$$\mathbb{P}\left(\exists y \in \mathcal{N} \;\; \|Ay\| < t\right) \text{ is small } \cdot |\mathcal{N}|$$

First approach: ε-net argument

1. Individual estimate (probability):

$$\mathbb{P}\left(\|Ay\| < t\right) \text{ is small}$$

This bound should hold uniformly over $y \in S^{n-1}$.

2. Discretization (geometry): find a small set $\mathcal{N} \subset S^{n-1}$ such that

$$\forall x \in S^{n-1}\ \exists y \in \mathcal{N}\ \|x - y\| < \varepsilon.$$

Union bound:

$$\mathbb{P}\left(\exists y \in \mathcal{N}\ \ \|Ay\| < t\right) \text{ is small } \cdot\, |\mathcal{N}|$$

3. Approximation: assume that $\|Ay\| \geq t$ for all $y \in \mathcal{N}$, then $\|Ax\| \geq t/2$ for all $x \in S^{n-1}$.

First approach: ε-net argument

1. Individual estimate (probability):

$$\mathbb{P}\left(\|Ay\| < t\right) \text{ is small}$$

This bound should hold uniformly over $y \in S^{n-1}$.

2. Discretization (geometry): find a small set $\mathcal{N} \subset S^{n-1}$ such that

$$\forall x \in S^{n-1}\ \exists y \in \mathcal{N}\ \|x - y\| < \varepsilon.$$

Union bound:

$$\mathbb{P}\left(\exists y \in \mathcal{N}\ \ \|Ay\| < t\right) \text{ is small } \cdot\, |\mathcal{N}|$$

3. Approximation: assume that $\|Ay\| \geq t$ for all $y \in \mathcal{N}$,
then $\|Ax\| \geq t/2$ for all $x \in S^{n-1}$.

- Main principle: "anticoncentration" beats the metric entropy of the set.

First approach: ε-net argument

1. Individual estimate (probability):

$$\mathbb{P}\left(\|Ay\| < t\right) \text{ is small}$$

This bound should hold uniformly over $y \in S^{n-1}$.

2. Discretization (geometry): find a small set $\mathcal{N} \subset S^{n-1}$ such that

$$\forall x \in S^{n-1}\ \exists y \in \mathcal{N}\ \|x - y\| < \varepsilon.$$

Union bound:

$$\mathbb{P}\left(\exists y \in \mathcal{N}\ \ \|Ay\| < t\right) \text{ is small } \cdot\, |\mathcal{N}|$$

3. Approximation: assume that $\|Ay\| \geq t$ for all $y \in \mathcal{N}$,
then $\|Ax\| \geq t/2$ for all $x \in S^{n-1}$.

- Main principle: "anticoncentration" beats the metric entropy of the set.
- Advantage: provides exponential bounds for probability.

First approach: ε-net argument

1. Individual estimate (probability):

$$\mathbb{P}\left(\|Ay\| < t\right) \text{ is small}$$

This bound should hold uniformly over $y \in S^{n-1}$.

2. Discretization (geometry): find a small set $\mathcal{N} \subset S^{n-1}$ such that

$$\forall x \in S^{n-1}\ \exists y \in \mathcal{N}\ \|x - y\| < \varepsilon.$$

Union bound:

$$\mathbb{P}\left(\exists y \in \mathcal{N}\ \ \|Ay\| < t\right) \text{ is small } \cdot\, |\mathcal{N}|$$

3. Approximation: assume that $\|Ay\| \geq t$ for all $y \in \mathcal{N}$, then $\|Ax\| \geq t/2$ for all $x \in S^{n-1}$.

- Main principle: "anticoncentration" beats the metric entropy of the set.
- Advantage: provides exponential bounds for probability.
- Fails when the exponential bounds are not available.

First approach: ε-net argument

1. Individual estimate (probability):

$$\mathbb{P}\left(\|Ay\| < t\right) \text{ is small}$$

This bound should hold uniformly over $y \in S^{n-1}$.

2. Discretization (geometry): find a small set $\mathcal{N} \subset S^{n-1}$ such that

$$\forall x \in S^{n-1}\ \exists y \in \mathcal{N}\ \|x - y\| < \varepsilon.$$

Union bound:

$$\mathbb{P}\left(\exists y \in \mathcal{N}\ \ \|Ay\| < t\right) \text{ is small } \cdot\, |\mathcal{N}|$$

3. Approximation: assume that $\|Ay\| \geq t$ for all $y \in \mathcal{N}$, then $\|Ax\| \geq t/2$ for all $x \in S^{n-1}$.

- Main principle: "anticoncentration" beats the metric entropy of the set.
- Advantage: provides exponential bounds for probability.
- Fails when the exponential bounds are not available.
- Can be applied to parts of the sphere having *low complexity*.

Second approach: invertibility via distance

- Let $x \in S^{n-1}$ be a vector such that $|x_1| \geq n^{-1/2}$.

Second approach: invertibility via distance

- Let $x \in S^{n-1}$ be a vector such that $|x_1| \geq n^{-1/2}$.
- Set $W = \text{span}(Y_k \mid k \neq 1)$, where Y_k are columns of A. Then

$$\|Ax\| \geq \text{dist}(Ax, W) = \text{dist}(x_1 \cdot Y_1, W).$$

Hence,

$$\|Ax\| \geq n^{-1/2} \cdot \text{dist}(Y_1, W),$$

where the right hand side is independent of x.

Second approach: invertibility via distance

- Let $x \in S^{n-1}$ be a vector such that $|x_1| \geq n^{-1/2}$.
- Set $W = \text{span}(Y_k \mid k \neq 1)$, where Y_k are columns of A. Then

$$\|Ax\| \geq \text{dist}(Ax, W) = \text{dist}(x_1 \cdot Y_1, W).$$

  Hence,

$$\|Ax\| \geq n^{-1/2} \cdot \text{dist}(Y_1, W),$$

  where the right hand side is independent of x.
- What if only $n/2$ coordinates of the vector x are $\sim n^{-1/2}$?
  Solution: choose a random coordinate.

Second approach: invertibility via distance

- Let $x \in S^{n-1}$ be a vector such that $|x_1| \geq n^{-1/2}$.
- Set $W = \text{span}(Y_k \mid k \neq 1)$, where Y_k are columns of A. Then

$$\|Ax\| \geq \text{dist}(Ax, W) = \text{dist}(x_1 \cdot Y_1, W).$$

  Hence,

$$\|Ax\| \geq n^{-1/2} \cdot \text{dist}(Y_1, W),$$

  where the right hand side is independent of x.
- What if only $n/2$ coordinates of the vector x are $\sim n^{-1/2}$?
  Solution: choose a random coordinate.
- What if less than $n/2$ coordinates of the vector x are $\sim n^{-1/2}$?
  Solution: the set of such points has low complexity.
  Use the ε-net argument.

Second approach: invertibility via distance

- Let $x \in S^{n-1}$ be a vector such that $|x_1| \geq n^{-1/2}$.
- Set $W = \text{span}(Y_k \mid k \neq 1)$, where Y_k are columns of A. Then

$$\|Ax\| \geq \text{dist}(Ax, W) = \text{dist}(x_1 \cdot Y_1, W).$$

  Hence,

$$\|Ax\| \geq n^{-1/2} \cdot \text{dist}(Y_1, W),$$

  where the right hand side is independent of x.
- What if only $n/2$ coordinates of the vector x are $\sim n^{-1/2}$?
  Solution: choose a random coordinate.
- What if less than $n/2$ coordinates of the vector x are $\sim n^{-1/2}$?
  Solution: the set of such points has low complexity.
  Use the ε-net argument.
- We reduced the problem to estimating $\text{dist}(Y_1, W)$ from below.

Distance to a random hyperplane

- We have to estimate $\operatorname{dist}(Y_1, W)$, where $W = \operatorname{span}(Y_k \mid k \neq 1$, and Y_k are columns of A.

Distance to a random hyperplane

- We have to estimate $\operatorname{dist}(Y_1, W)$, where $W = \operatorname{span}(Y_k \mid k \neq 1$, and Y_k are columns of A.
- Y_1 is independent of $W \Rightarrow$ condition on W.

Distance to a random hyperplane

- We have to estimate $\operatorname{dist}(Y_1, W)$, where $W = \operatorname{span}(Y_k \mid k \neq 1$, and Y_k are columns of A.
- Y_1 is independent of $W \Rightarrow$ condition on W.
  1. Estimate the distance from a random vector to a fixed subspace.

Distance to a random hyperplane

- We have to estimate $\operatorname{dist}(Y_1, W)$, where $W = \operatorname{span}(Y_k \mid k \neq 1$, and Y_k are columns of A.
- Y_1 is independent of $W \Rightarrow$ condition on W.
  1. Estimate the distance from a random vector to a fixed subspace.
  2. Use the fact that W is random to exclude exceptional subspaces.

Distance to a random hyperplane

- We have to estimate $\operatorname{dist}(Y_1, W)$, where $W = \operatorname{span}(Y_k \mid k \neq 1$, and Y_k are columns of A.
- Y_1 is independent of $W \;\Rightarrow\;$ condition on W.
  1. Estimate the distance from a random vector to a fixed subspace.
  2. Use the fact that W is random to exclude exceptional subspaces.
- Let $w \in S^{n-1}$ be a unit normal to W. If Y_1 is a Gaussian vector, then

$$\mathbb{P}\left(|\langle Y_1, w\rangle| < \varepsilon\right) = \mathbb{P}\left(|g| < \varepsilon\right) \leq C\varepsilon.$$

Distance to a random hyperplane

- We have to estimate $\operatorname{dist}(Y_1, W)$, where $W = \operatorname{span}(Y_k \mid k \neq 1$, and Y_k are columns of A.
- Y_1 is independent of $W \;\Rightarrow$ condition on W.
  1. Estimate the distance from a random vector to a fixed subspace.
  2. Use the fact that W is random to exclude exceptional subspaces.
- Let $w \in S^{n-1}$ be a unit normal to W. If Y_1 is a Gaussian vector, then

$$\mathbb{P}\left(|\langle Y_1, w\rangle| < \varepsilon\right) = \mathbb{P}\left(|g| < \varepsilon\right) \leq C\varepsilon.$$

- General case: obstacles of the *arithmetic* nature.
  How to identify them?
  Solution: additive combinatorics / harmonic analysis.

Distance to a random hyperplane

- We have to estimate $\operatorname{dist}(Y_1, W)$, where $W = \operatorname{span}(Y_k \mid k \neq 1$, and Y_k are columns of A.
- Y_1 is independent of $W \;\Rightarrow$ condition on W.
  1. Estimate the distance from a random vector to a fixed subspace.
  2. Use the fact that W is random to exclude exceptional subspaces.
- Let $w \in S^{n-1}$ be a unit normal to W. If Y_1 is a Gaussian vector, then

$$\mathbb{P}\left(|\langle Y_1, w\rangle| < \varepsilon\right) = \mathbb{P}\left(|g| < \varepsilon\right) \leq C\varepsilon.$$

- General case: obstacles of the *arithmetic* nature.
  How to identify them?
  Solution: additive combinatorics / harmonic analysis.
- How to avoid arithmetic obstacles?
  Solution: w is a random vector. It cannot have any arithmetic structure with probability close to 1.

Distance to a random hyperplane

- We have to estimate $\operatorname{dist}(Y_1, W)$, where $W = \operatorname{span}(Y_k \mid k \neq 1$, and Y_k are columns of A.
- Y_1 is independent of $W \;\Rightarrow$ condition on W.
  1. Estimate the distance from a random vector to a fixed subspace.
  2. Use the fact that W is random to exclude exceptional subspaces.
- Let $w \in S^{n-1}$ be a unit normal to W. If Y_1 is a Gaussian vector, then

$$\mathbb{P}\left(|\langle Y_1, w\rangle| < \varepsilon\right) = \mathbb{P}\left(|g| < \varepsilon\right) \leq C\varepsilon.$$

- General case: obstacles of the *arithmetic* nature.
  How to identify them?
  Solution: additive combinatorics / harmonic analysis.
- How to avoid arithmetic obstacles?
  Solution: w is a random vector. It cannot have any arithmetic structure with probability close to 1.
- How to use randomness?
  Solution: Arithmetic structure $\Rightarrow$ low complexity $\Rightarrow$ use a chain of ε-net arguments.

Small ball probability

- $Y = (\xi_1, \dots, \xi_n)$ – a vector with independent random coordinates;
- $a = (a_1, \dots, a_n)$ a fixed vector.

How to bound

$$\mathbb{P}\left(|\langle Y, a\rangle| < \varepsilon\right) = \mathbb{P}\left(|\sum_{k=1}^{n} a_k \xi_k| < \varepsilon\right)?$$

Study of the small ball probability for sums of independent random variables goes back to Kolmogorov, Lévy and Esseen *(local limit theorems)*.

- *Large deviation inequalities* demonstrate that S concentrates nicely about its mean.
- *Small ball probability* goes in the opposite direction: S can not concentrate too much.

Small ball probability

$$S = \sum_{k=1}^{N} a_k \xi_k, \qquad a_k \in \mathbb{R}.$$

- Examples (for Bernoulli random variables $\xi_k = \pm 1$):

For $a = (1, 1, 0, \ldots, 0)$, we have $\mathbb{P}(S = 0) = \frac{1}{2}$.

Small ball probability

$$S = \sum_{k=1}^{N} a_k \xi_k, \qquad a_k \in \mathbb{R}.$$

- Examples (for Bernoulli random variables $\xi_k = \pm 1$):

For $a = (1, 1, 0, \ldots, 0)$, we have $\mathbb{P}(S = 0) = \frac{1}{2}$.

For $a = (1, 1, 1, \ldots, 1)$, we have $\mathbb{P}(S = 0) \sim N^{-1/2}$.

Small ball probability

$$S = \sum_{k=1}^{N} a_k \xi_k, \qquad a_k \in \mathbb{R}.$$

- Examples (for Bernoulli random variables $\xi_k = \pm 1$):

For $a = (1, 1, 0, \dots, 0)$, we have $\mathbb{P}(S = 0) = \frac{1}{2}$.

For $a = (1, 1, 1, \dots, 1)$, we have $\mathbb{P}(S = 0) \sim N^{-1/2}$.

For $a = (1, 2, 3, \dots, N)$, we have $\mathbb{P}(S = 0) \sim N^{-3/2}$.

Small ball probability

$$S = \sum_{k=1}^{N} a_k \xi_k, \qquad a_k \in \mathbb{R}.$$

- Examples (for Bernoulli random variables $\xi_k = \pm 1$):

For $a = (1, 1, 0, \dots, 0)$, we have $\mathbb{P}(S = 0) = \frac{1}{2}$.

For $a = (1, 1, 1, \dots, 1)$, we have $\mathbb{P}(S = 0) \sim N^{-1/2}$.

For $a = (1, 2, 3, \dots, N)$, we have $\mathbb{P}(S = 0) \sim N^{-3/2}$.

- If the small ball probability is big, then the coefficient vector a has strong *additive structure*.

Littlewood-Offord Theory

$$S = \sum_{k=1}^{N} a_k \xi_k, \qquad a = (a_1, \dots, a_N) \in \mathbb{R}^N.$$

We want to construct a measure of the arithmetic structure of a the vector a so that

1. it controls the small ball probability;
2. it is relatively easy to evaluate.

Littlewood-Offord Theory

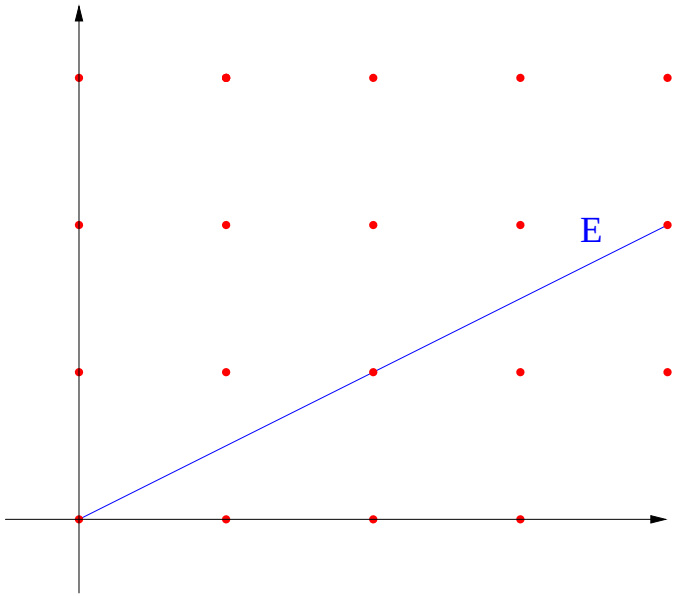
$$S = \sum_{k=1}^{N} a_k \xi_k, \qquad a = (a_1, \ldots, a_N) \in \mathbb{R}^N.$$

We want to construct a measure of the arithmetic structure of a the vector a so that

1. it controls the small ball probability;
2. it is relatively easy to evaluate.

Least common denominator

$$\mathrm{lcd}(a) = \inf\{\theta > 0 \mid \theta a \in \mathbb{Z}^N\}.$$



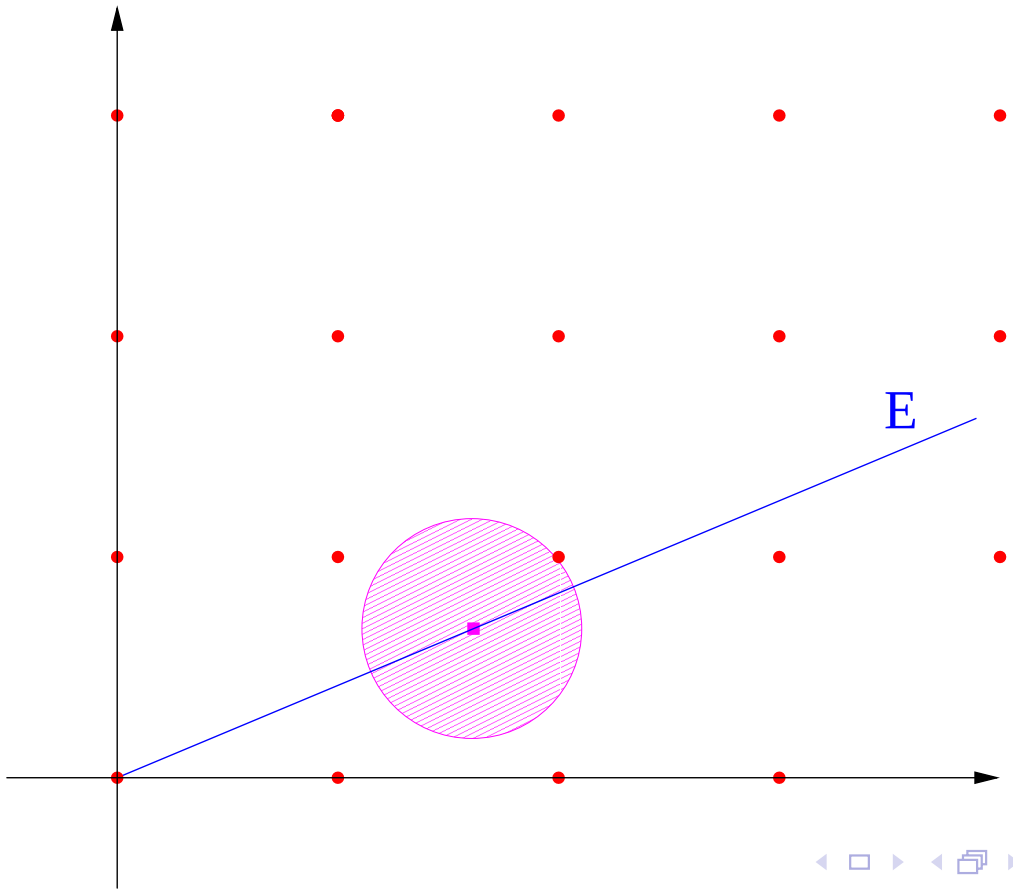
- LCD is "inversely proportional" to the amount of structure in a.
- A related measure of structure: *length of the shortest arithmetic progression into which a embeds* (exactly or approximately).

Essential least common denominator

Definition

Let $\alpha > 0$ (typically, $\alpha = c\sqrt{N}$). Define the least common denominator of the vector $a \in S^{n-1}$ by

$$\mathrm{LCD}(E) := \inf\{\theta > 0 \mid \mathrm{dist}(\theta a, \mathbb{Z}^N) < \alpha\}.$$

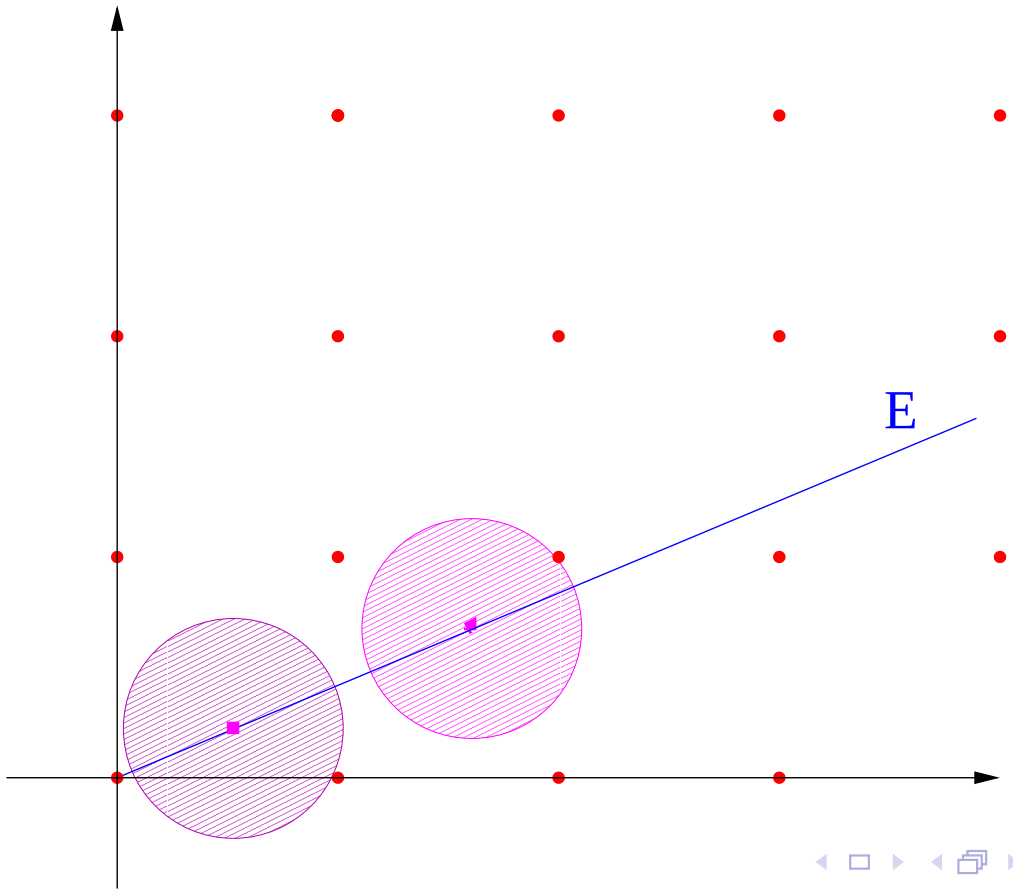


Essential least common denominator

Definition

Let $\alpha > 0$ (typically, $\alpha = c\sqrt{N}$). Define the least common denominator of the vector $a \in S^{n-1}$ by

$$\mathrm{LCD}(E) := \inf\{\theta > 0 \mid \mathrm{dist}(\theta a, \mathbb{Z}^N) < \alpha\}.$$

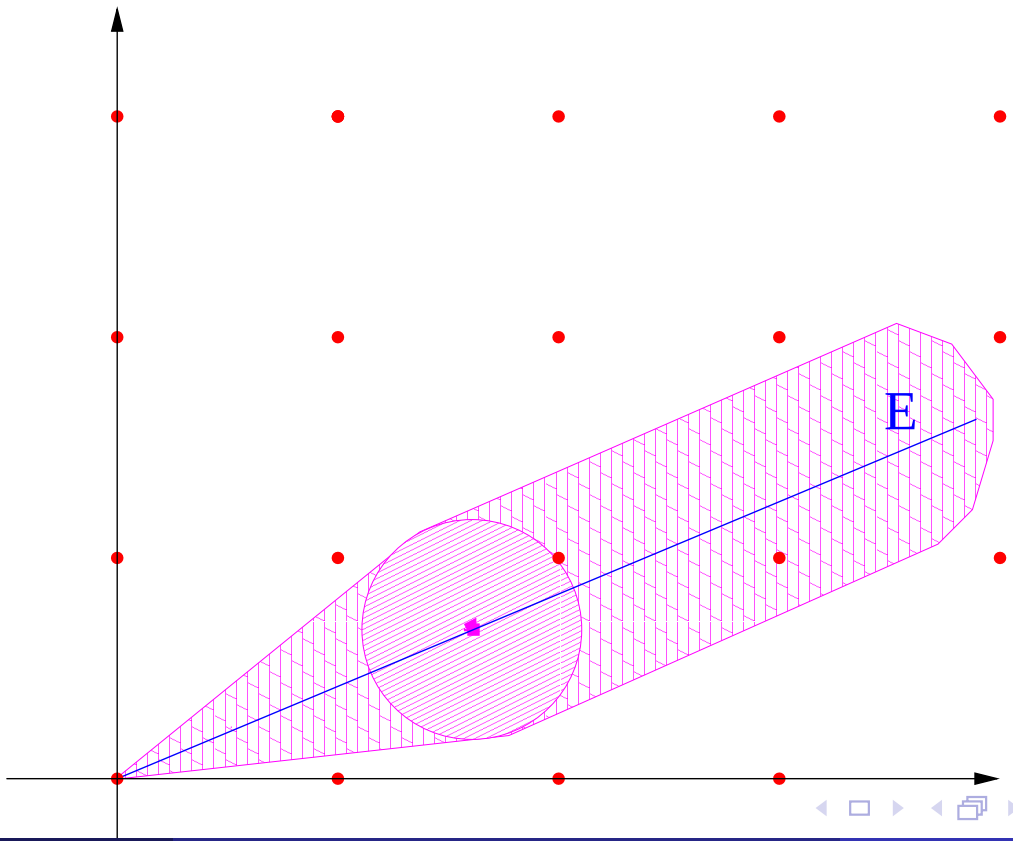


Essential least common denominator

Definition

Let $\alpha > 0$ (typically, $\alpha = c\sqrt{N}$). Define the least common denominator of the vector $a \in S^{n-1}$ by

$$\mathrm{LCD}(E) := \inf\{\theta > 0 \mid \mathrm{dist}(\theta a, \mathbb{Z}^N) < \min(\theta/2, \alpha)\}.$$



Littlewood-Offord Theory

$$S = \sum_{k=1}^{N} a_k \xi_k, \qquad a = (a_1, \dots, a_N) \in \mathbb{R}^N, \qquad \|a\|_2 = 1.$$

Small Ball Probability Theorem

$$\mathbb{P}\left(|S| \le \varepsilon\right) \lesssim \varepsilon + \frac{1}{\mathrm{LCD}_\alpha(a)} + e^{-\alpha^2}.$$

Littlewood-Offord Theory

$$S = \sum_{k=1}^{N} a_k \xi_k, \qquad a = (a_1, \dots, a_N) \in \mathbb{R}^N, \qquad \|a\|_2 = 1.$$

Small Ball Probability Theorem

$$\mathbb{P}\left(|S| \le \varepsilon\right) \lesssim \varepsilon + \frac{1}{\mathrm{LCD}_\alpha(a)} + e^{-\alpha^2}.$$

- This corresponds to the "ideal" estimate, valid for continuous distributions. For example, for Gaussian ξ_k, we have $\mathbb{P}\left(|S| \le \varepsilon\right) \sim \varepsilon$.

Littlewood-Offord Theory

$$S = \sum_{k=1}^{N} a_k \xi_k, \qquad a = (a_1, \dots, a_N) \in \mathbb{R}^N, \qquad \|a\|_2 = 1.$$

Small Ball Probability Theorem

$$\mathbb{P}\left(|S| \le \varepsilon\right) \lesssim \varepsilon + \frac{1}{\mathrm{LCD}_\alpha(a)} + e^{-\alpha^2}.$$

- This corresponds to the "ideal" estimate, valid for continuous distributions. For example, for Gaussian ξ_k, we have $\mathbb{P}\left(|S| \le \varepsilon\right) \sim \varepsilon$.
- The only obstacle to small ball probability is *structure*. If the small ball probability is large then $\mathrm{LCD}(a)$ is small, thus a has strong additive structure (and in particular embeds into a short arithmetic progression).

Littlewood-Offord Theory

$$S = \sum_{k=1}^{N} a_k \xi_k, \qquad a = (a_1, \dots, a_N) \in \mathbb{R}^N, \qquad \|a\|_2 = 1.$$

Small Ball Probability Theorem

$$\mathbb{P}\left(|S| \le \varepsilon\right) \lesssim \varepsilon + \frac{1}{\mathrm{LCD}_\alpha(a)} + e^{-\alpha^2}.$$

- This corresponds to the "ideal" estimate, valid for continuous distributions. For example, for Gaussian ξ_k, we have $\mathbb{P}\left(|S| \le \varepsilon\right) \sim \varepsilon$.
- The only obstacle to small ball probability is *structure*. If the small ball probability is large then $\mathrm{LCD}(a)$ is small, thus a has strong additive structure (and in particular embeds into a short arithmetic progression).
- This term is typically *exponentially small* in N. $(\alpha \sim 0.01\sqrt{N})$

Additive Structure of Random Vectors

No Structure Theorem

Let H be a hyperplane in $\mathbb{R}^N$ spanned by $N-1$ independent vectors with i.i.d. random coefficients. Then the normal a of H has no structure:

$$\mathrm{LCD}(a) \geq e^{cN}$$

with probability at least $1 - e^{-cN}$.

- Nontrivial: the coefficients of a are not independent.

Distance to a hyperplane

Distance Theorem

1. Arbitrary hyperplane. Let H be a hyperplane in $\mathbb{R}^n$. Then, for $\varepsilon \geq 1/\operatorname{LCD}(H^\perp)$,

$$\mathbb{P}\left(\operatorname{dist}(X, H) \leq \varepsilon\right) \lesssim \varepsilon + c^n.$$

2. Random hyperplanes. H is spanned by $n-1$ independent random vectors. Then this bound holds for *arbitrary* $\varepsilon > 0$.

1. Small Ball Probability Theorem.
2. No Structure Theorem $\Rightarrow$ LCD of random $H^\perp$ is exponentially large.