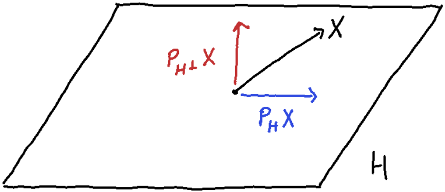
Distance Theorems and the Smallest Singular Value of Inhomogeneous Random Rectangular Matrices

Manuel Fernandez V
Online Asymptotic Geometric Analysis Seminar

October 30, 2024

Motivating Question

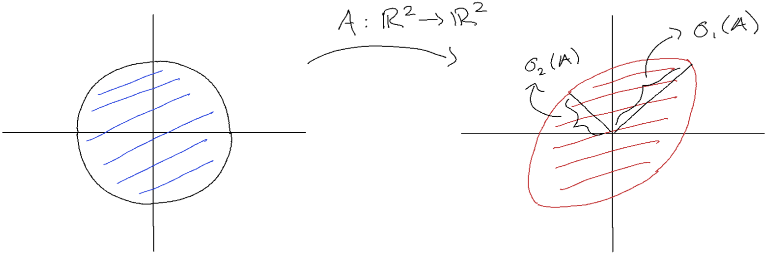
Let $X \in \mathbb{R}^n$ be a random vector, let $A \in \mathbb{R}^{n\times(n-d)}$ be a random matrix, and H be the subspace spanned by the columns of A.
Q: What can we say about $\mathrm{dist}(X, H)$?



In the context of non-asymptotic random matrix theory, this is an important question with a number of applications.

Preliminaries and Notation

- For a matrix $A \in \mathbb{R}^{N\times n}$ the singular values of A, $\sigma_1(A), \cdots, \sigma_n(A)$, are the square roots of the eigenvalues of $A^\top A$.
- The smallest singular value satisfies $\sigma_n(A) = \inf_{x\in\mathbb{S}^{n-1}} \|Ax\|_2$.
- The largest singular value satisfies $\sigma_1(A) = \sup_{x\in\mathbb{S}^{n-1}} \|Ax\|_2$.



- The Hilbert Schmidt norm of a matrix A is $\|A\|_{HS} = \sqrt{\sum_{i,j} a_{ij}^2}$
- A random variable X is uniformly anti-concentrated if $\sup_{z\in\mathbb{R}} \mathbb{P}(|X - z| < a) < b$.
- Given $\delta, \rho \in (0,1)$ we define $\mathbf{Comp}(\delta,\rho) := \{x \in \mathbb{S}^{n-1} \ : \ \mathrm{dist}(x, \mathbf{sparse}(\delta n)) \le \rho\}$.
- $\mathbf{Incomp}(\delta,\rho) := \mathbb{S}^{n-1} \setminus \mathbf{Comp}(\delta,\rho)$.

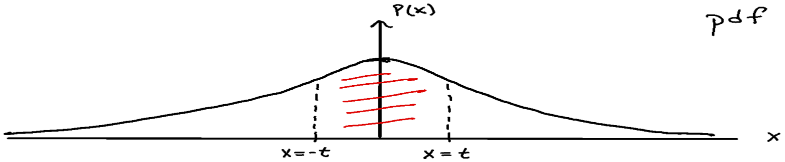
Distances in Non-asymptotic Random Matrix Theory

- Deriving lower tail estimates for the smallest singular value of rectangular matrices ([10]):
  Proving a lower bound on $\inf_{x \in \mathbb{S}^{n-1}} \|Ax\|_2$ reduces to proving a lowerbound on $\inf_{z \in \mathbf{spread}_J} \mathrm{dist}(A_J z, H_{J^c})$, for an appropriate choice of submatrix A_J and subspace H_{J^c}.
- Deriving upper tail estimate for smallest singular value of square matrices ([9]):
  Proving an upperbound on $\|A^{-1}\|$ reduces to proving a lowerbound on $\mathrm{dist}(A_k, H_{1,k})$ for all $1 \le k \le n$
- ℓ_∞ delocalization of eigenvectors [11]: Proving that eigenvectors have small ℓ_∞ norm requires proving lower bounds on the distance between anisotropic random vectors and random subspaces.
- Many more examples (rank deviation [8], no-gaps delocalization [12], etc.).

Levy concentration function

For a random variable $X \in \mathbb{R}$ the levy-concentration function is

$$\mathcal{L}(X,t) := \sup_{z \in \mathbb{R}} \mathbb{P}(|X - z| \leq t) \tag{1}$$



$$\mathcal{L}(X,t) := \sup_{z \in \mathbb{R}} Pr(|X - z| \leq t) \leq$$ 

For a random vector $X \in \mathbb{R}^n$ the levy-concentration function is

$$\mathcal{L}(X,t) := \sup_{v \in \mathbb{R}^n} \mathbb{P}\left(\|X - v\|_2 \leq t\right) \tag{2}$$

Importance for distance:

$$\mathbb{P}(\text{dist}(X,H) \leq t) = \mathbb{P}(\|P_{H^\perp} X\|_2 \leq t) \leq \mathcal{L}(P_{H^\perp} X, t)$$

Distance theorems

$X \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times (n-d)}$, $H = \text{colspan}(A)$

Theorem (Rudelson and Vershynin [10])

- $1 \leq d \leq cn$
- *X, A have i.i.d. subgaussian entries with unit variance.*

$$\mathcal{L}(P_{H^\perp} X, t\sqrt{d}) \leq (Ct)^d + e^{-cn}$$

Theorem (Livshyts [5])

- $1 \leq d \leq cn$
- *X has i.i.d. entries, A has i.i.d. rows*
- *UAC entries*
- $E\|A\|_{HS}^2 \leq Kn^2$

$$\mathcal{L}(P_{H^\perp} X, t\sqrt{d}) \leq (Ct)^d + e^{-cn}$$

Distance theorems cont.

$X \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times (n-d)}$, $H = \text{colspan}(A)$

Theorem (Livshyts, Tikhomirov, Vershynin [6])

- $d = 1$
- *UAC entries*
- $E\|X\|_2^2 \leq rn^2$
- $E\|A\|_{HS}^2 \leq Kn^2$

$$\mathcal{L}(P_{H^\perp}X, t\sqrt{d}) \leq (Ct)^d + e^{-cn}$$

Theorem (Rudelson, Vershynin [13])

- $\mathcal{L}(X_i, t) \leq p$ *for all* $1 \leq i \leq n$

$$\mathcal{L}(P_{H^\perp}X, t\sqrt{d}) \leq (Cp)^d$$

Distance theorem for inhomogeneous rectangular matrices

Theorem (F. [16])

$X \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times(n-d)}$, $H = \mathrm{colspan}(A)$

- $1 \le d \le \lambda n/(\log n)$
- *UAC entries*
- $E\|X\|_2^2 \le rn^2$
- $E\|A\|_{HS}^2 \le Kn^2$

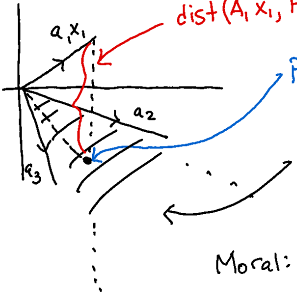
$$\mathcal{L}(P_{H^\perp}X, t\sqrt{d}) \le (Ct)^d + e^{-cn}.$$

Improves upon the square distance theorem in [6].
Remark about aspect ratio.

Relating Distance to Smallest Singular Value

Example in $\mathbb{R}^3$

$$A = \begin{bmatrix} | & | & | \\ A_1 & A_2 & A_3 \\ | & | & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$



$\mathrm{dist}(A_1 x_1, H_1) = |x_1| \, \|P_{H_1^\perp} A_1\|_2$

$p = P_{H_1} A_1$

$H_1 = \mathrm{span}(A_2, A_3)$

Moral: $\|Ax\|_2 \geq |x_1| \, \mathrm{dist}(A_1, H_1)$

Invertibility via distance (square case):

$$\Pr\left(\inf_{x \in \mathrm{Incomp}(\delta, \rho)} \|Ax\|_2 \leq \frac{\varepsilon}{\sqrt{n}} \right) \leq \frac{4}{\delta n} \sum_{i=1}^{n} \Pr\left(\mathrm{dist}(A_i, H_i) \leq \varepsilon \right)$$

Subgaussian rectangular matrices: some previous results

Let $A \in \mathbb{R}^{N\times n}$.
Standard gaussian matrices (Edelman. [2], Szarek [14]).

Theorem (Litvak, Pajor,Rudelson, Tomcazk-Jaegermann [4])

- *i.i.d. entries*
- *mean-0 subgaussian*
- $N - n \geq cn/(\log n)$

$$\mathbb{P}(\sigma_n(A) \leq c\sqrt{N}) \leq e^{-cN}$$

Theorem (Rudelson, Vershynin [10])

- *i.i.d. entries*
- *mean-0 subgaussian*

$$\mathbb{P}(\sigma_n(A) \leq \varepsilon(\sqrt{N+1} - \sqrt{n})) \leq (C\varepsilon)^{N+1-n} + e^{-cN}.$$

inhomogeneous matrices: some previous results

Let $A \in \mathbb{R}^{N \times n}$.

Theorem (Livshyts [5])

- *independent, mean-0, unit variance entries*
- *i.i.d rows*
- *uniformly anti-concentrated entries.*

$$\mathbb{P}(\sigma_n(A) \leq \varepsilon(\sqrt{N+1} - \sqrt{n})) \leq (C\varepsilon \log(1/\varepsilon))^{N+1-n} + e^{-cN}$$

Theorem (Livshyts, Tikhomirov, Vershynin [6])

- $\mathbb{E}\|A\|_{HS}^2 \leq Kn^2$
- *independent UAC entries*

$$\mathbb{P}(\sigma_n(A) \leq \varepsilon/\sqrt{n}) \leq C\varepsilon + e^{-cN}$$

Sparse inhomogeneous matrices: (Litvak, Rivasplata [3])

New Results

Let $A \in \mathbb{R}^{N \times n}$.

Theorem (Dabagia, F. [1])

- $\mathbb{E} A_i A_i^\top = I_N$
- *independent mean-0 entries*

1. *If A has UAC entries then*

$$\mathbb{P}\left(\sigma_n(A) \leq \varepsilon\left(\sqrt{N+1} - \sqrt{n}\right)\right) \leq (C\varepsilon \log(1/\varepsilon))^{N-n+1} + e^{-cN}$$

2. *If A has entries with bounded $2+\beta$ moments then*

$$\mathbb{P}\left(\sigma_n(A) \leq \varepsilon\left(\sqrt{N+1} - \sqrt{n}\right)\right) \leq (C\varepsilon)^{N-n+1} + e^{-cN}$$

Proving a distance theorem

Recall:

$$\mathcal{L}(Y,t) := \sup_{v\in\mathbb{R}^n} \mathbb{P}\left(\|Y-v\|_2 \le t\right)$$

Characteristic function: $\phi_Y(\theta) := \mathbb{E}\exp(2\pi i\langle Y,\theta\rangle)$.
Analytic tool: Esseen's inequality

$$\mathcal{L}(Y,t) \le C^d \int_{B(0,\sqrt{d})} |\phi_Y(\theta)| \, d\theta$$

Proving a distance theorem cont.

For $X \in \mathbb{R}^n$, $A \in \mathbb{R}^{n \times (n-d)}$ and $H = \mathrm{colspan}(A)$ we have $\mathrm{dist}(X, A) = \|P_{H^\perp} X\|_2$.
Write $U^\top U = P_{H^\perp}, Y = t^{-1} U X$

$$\begin{aligned}
\mathcal{L}(P_{H^\perp} X, t\sqrt{d}) &= \mathcal{L}(t^{-1} U X, \sqrt{d}) \\
&\leq C^d \int_{B(0,\sqrt{d})} |\phi_Y(\theta)| \, d\theta \\
&\leq \cdots \\
&\cdots \\
&\leq C^d \int_{B(0,\sqrt{d})} \exp\left(-4 \cdot \mathbb{E}(\mathrm{dist}^2(U^\top \theta / t) \star \bar{X}, \mathbb{Z}^n)^2\right) \, d\theta
\end{aligned}$$

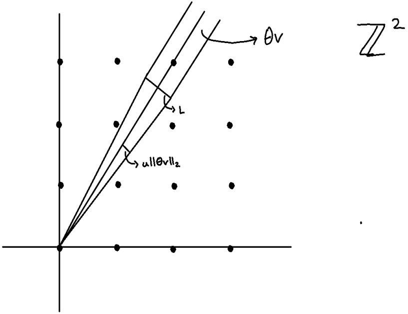
Q: How to bound this integral?

LCD of vectors 1

$\mathbb{E}(\text{dist}^2(U^\top\theta/t) \star \bar{X}, \mathbb{Z}^n)$ is large when $U^\top\theta$ lack 'arithmetic structure'.

- Essential Least Common Denominator (Rudelson and Vershynin [9],[10])

$$\mathbf{LCD}_{L,u}(V) = \inf\{\|\theta\|_2 > 0 \ : \ \text{dist}(V^\top\theta, \mathbb{Z}^n) < \min(u\|V^\top\theta\|_2, L)\} \tag{3}$$

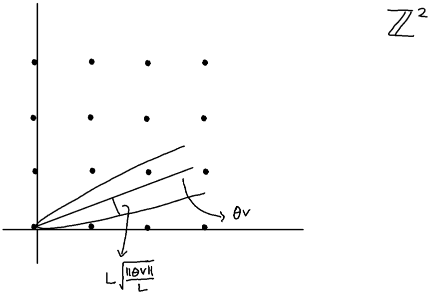


Pros: Successfully captures arithmetic structure.
Drawback: Difficult to use for $d > 1$, problematic for the inhomogeneous setting

LCD of vectors 2

- Logarithmic Least Common Denominator (Rudelson and Vershynin [12])

$$\mathbf{LCD}_{L,u}(V) = \inf\left\{\|\theta\|_2 > 0 \ : \ \text{dist}(V^\top \theta, \mathbb{Z}^n) < L\sqrt{\log_+ (\|V^\top \theta\|/L)}\right\} \qquad (4)$$



Pros: Easy to use for $d > 1$.
Drawback: Problematic for the inhomogeneous setting

LCD of vectors 3

- Randomized Least Common Denominator (Livshyts, Tikhomirov, Vershynin [6])

$$\mathbf{RLCD}^{X}_{L,u}(v) = \inf\{\theta > 0 \ : \ \mathbb{E}\mathrm{dist}^2(\theta v \star \bar{X}, \mathbb{Z}^n) < \min(u\|\theta v\|_2^2, L^2)\} \tag{5}$$

Pros: Applicable in the inhomogeneous setting.
Drawback: Specific for $d = 1$. Generalization maybe be difficult to use for $d > 1$.
Idea: Combine Randomized LCD and Logarithmic LCD

Randomized Logarithmic LCD

- Vector version: let $v \in \mathbb{R}^n$.

$$\mathbf{RLogLCD}^{X}_{L,u}(v) := \inf\left(\theta > 0 \ : \ \mathbb{E}\mathrm{dist}^2(\theta v \star \bar{X}, \mathbb{Z}^n) < L^2 \cdot \log_+\left(\frac{u\|\theta v\|_2}{L}\right)\right),$$
$$\mathbf{RLogLCD}^{A}_{L,u}(v) := \min_i \mathbf{RLogLCD}^{A_i}_{L,u}(v). \tag{6}$$

Randomized Logarithmic LCD cont.

- Matrix version: Let $V \in \mathbb{R}^{n \times N}$.

$$\mathbf{RLogLCD}_{L,u}^{X}(V) := \inf\left(\|\theta\| > 0 : \mathbb{E}\mathrm{dist}^2(V^{\intercal}\theta \star \bar{X}, \mathbb{Z}^n) < L^2 \cdot \log_+\left(\frac{u\|V^{\intercal}\theta\|_2}{L}\right)\right) \tag{7}$$

- Subspace version: Let $E \subseteq \mathbb{R}^N$ be a subspace.

$$\mathbf{RLogLCD}_{L,u}^{X}(E) := \inf_{v \in E \cap \mathbb{S}^{N-1}} \mathbf{RLogLCD}_{L,u}^{X}(v). \tag{8}$$

Large Randomized Logarithmic LCD $\Longrightarrow$ Distance Theorem

Takeaway: If $\mathbf{RLogLCD}^X_{L,u}(U)$ is large then the integral can be bounded nicely.
Recall:

$$\mathcal{L}(Y,t) := \sup_{v\in\mathbb{R}^n} \mathbb{P}(\|Y-v\|_2 \leq t)$$

Proposition (Levy concentration for projections)

- *$X \in \mathbb{R}^n$ is a random vector.*
- *$E \subseteq R^n$ is a subspace*
- $4L^2 \geq d+2$
- $D(E) := \boldsymbol{RLogLCD}^X_{L,u}(E)$

$$\mathcal{L}(P_E X, t) \leq \left(\frac{CL}{\sqrt{du}}\right)^d \cdot \max\left(t, \sqrt{d}/D(E)\right)^d \tag{9}$$

To recover our theorem, it suffices to prove that $D(E) \geq \sqrt{d}e^{cn/d}$ when $E = H^\perp$.

First steps: Proving Properties of Randomized Logarithmic LCD

Analogous to Randomized LCD [6], Randomized Logarithmic LCD enjoys some nice properties.

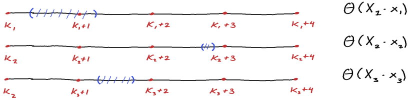
1. $\mathbf{RLogLCD}_{L,u}^{X}(v)$ is 'stable' under pertubations of v.
2. $\mathbf{RLogLCD}_{L,u}^{X}(v)$ is monotone in L (assuming $\mathbf{RLogLCD}_{L,u}^{X}(v)$ is large enough).
3. $\mathbf{RLogLCD}_{L,u}^{X}(v)$ isn't 'too small' when v is incompressible.
   Recall: $\mathbf{Incomp}(\delta, \rho) := \{x \in \mathbb{S}^{n-1} \; : \; \mathrm{dist}(x, \mathbf{sparse}(\delta n)) > \rho\}$.

Difficulty with lowerbound:

- Naive lower bound: L/u.
- Desired lower bound: $L/u + cn/\sqrt{\mathrm{Var}(X)}$.
- Problematic when $L/u \gg cn/\sqrt{\mathrm{Var}(X)}$

Idea: A better lowerbound when L is random

Issue: If you give me L, I can construct X so that $(L/u)v \star \bar{X} \in \mathbb{Z}^n$, where $v = (1/\sqrt{n}, \cdots, 1/\sqrt{n})$.
Workaround: Take $\tilde{L} \sim [L, 2L]$ uniformly.



Proposition

Let $x \in \boldsymbol{Incomp}(\delta, \rho)$. If $L^2 \lesssim n$ then there exists $\tilde{L} \in [L, 2L]$ such that

$$\boldsymbol{RLogLCD}^{X}_{\tilde{L},u}(x) \geq \frac{\tilde{L}}{u} + \frac{cn}{\sqrt{\mathrm{Var}(X)}}. \tag{10}$$

Caveat: $\tilde{L}$ depends on x.
Consolation: May pick $\tilde{L} \in \{L, 2L\} \cup \{L + i/10 \ : \ 1 \leq i \leq \lfloor 10L \rfloor\}$.

Main task: Lowerbounding $\mathbf{RLogLCD}^X_{L,u}(H^\perp)$ via recursion

Recall: $\mathbf{RLogLCD}^X_{L,u}(H^\perp) := \inf_{v \in H^\perp \cap \mathbb{S}^{N-1}} \mathbf{RLogLCD}^X_{L,u}(v)$.
Exploit: $H^\perp \in \text{null}(A^\top)$
Notation:

- $Q_i :=$ submatrix of A consisting of columns of A with variance at most $n^2/d - f(i)$.
- $\overline{D} = e^{cn/d}$
- $D_i = \overline{D}2^{-i}$
- $\underline{D} = D_{\max}$.

Recursive Formulation

$$\mathcal{H}_i := \{\exists x \in \textbf{Incomp}(\delta,\rho) \ s.t.\ A^\top x = 0, \textbf{RLogLCD}^{Q_i}_{\ell,u}(x) \in [\underline{D}, D_i)\},$$

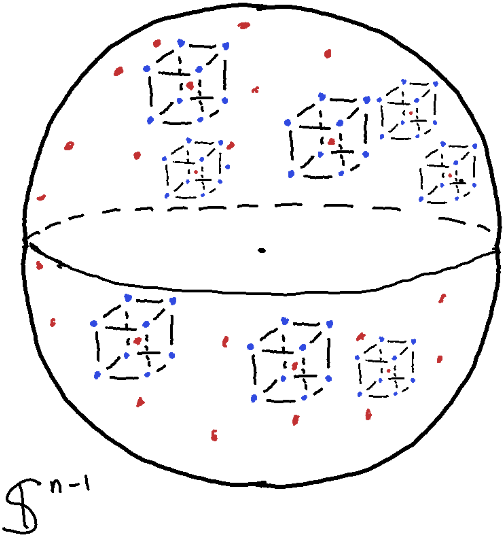
$$\mathcal{E}_i := \{\exists x \in \textbf{Incomp}(\delta,\rho) \ s.t.\ A^\top x = 0, \textbf{RLogLCD}^{Q_{i+1}}_{L,u}(x) \in [\underline{D}, D_i) \text{ and } \textbf{RLogLCD}^{Q_{i+1}}_{L,u}(x) \geq D_{i+1}\},$$

$$\mathcal{F}_i := \bigcup_{Y \in Q_i \setminus Q_{i+1}} \{\exists x \in \textbf{Incomp}(\delta,\rho) \ s.t.\ A^\top x = 0, \textbf{RLogLCD}^{Q_{i+1}|Y}_{L,u}(x) \in [\underline{D}, D_i) \text{ and } \textbf{RLogLCD}^{Q_{i+1}}_{L,u}(x) \geq D_{i+1}\}.$$

$$\mathcal{H}_0 \subseteq \bigcup_{1 \leq i \leq \max} (\mathcal{E}_i \cup \mathcal{F}_i)$$

$\mathbb{P}(\mathcal{H}_0) \leq$ Union Bound.
We bound $\textbf{RLogLCD}^{Q|X}_{L,u}(H^\perp)$ in a similar way.

Upper bounding $\mathbb{P}(\mathcal{E}_i), \mathbb{P}(\mathcal{F}_i)$



$\bullet$: $H^\perp \cap \mathbb{S}^{n-1}$

$\bullet$: $\mathcal{N}$

Observation:

$A^T x = 0$ but

$$\|A^T x\|_2 \geq \inf_{v \in \mathcal{N}} \|A^T v\|_2 - \varepsilon$$

(finer net for lager LCD)

Estimating the smallest singular value via distance

In view of previous work [10, 6, 5] showing that $\inf_{x\in\mathbf{Comp}(\delta,\rho)}\|Ax\|_2$ is large is not too difficult.
Crux of the matter: Lower bounding $\inf_{x\in\mathbf{Incomp}(\delta,\rho)}\|Ax\|_2$.

Definition (spread vector [10])

A vector $v \in \mathbb{S}^{d-1}$ is said to be spread if all entries satisfy $|v_i| \in [cd^{-1/2}, Cd^{-1/2}]$, for some positive constants c, C. We denote the set of spread vectors as spread_d.

Lemma (Invertibility via distance [10])

[10] Let $d \in \{1, 2\cdots, n\}$ and $A \in \mathbb{R}^{N\times n}$ be a random matrix. Then there exists $J \subset \{1, 2, \cdots, n\}$ of size d such that

$$\mathbb{P}\left(\inf_{x\in Incomp(\delta,\rho)}\|Ax\|_2 \le \frac{\varepsilon d}{\sqrt{n}}\right) \le C^d\mathbb{P}\left(\inf_{x\in spread_d} dist(A_J x, H_{J^c}) \le \varepsilon\sqrt{d}\right),$$

where C and the spread parameters depend only on δ and ρ.

Key ideas for obtaining estimates for inhomogeneous rectangular matrices

In view of previous tools and results ([10],[5],[6]), what do we need to prove our smallest singular value estimates?

1. A way to bound $\text{dist}(A_J x, H_{J^c})$ when A has no identical distribution assumptions. (Use our distance theorem)
2. An improved deviation inequality for regularized Hilbert-Schmidt norm.

Remark: The tall case.

Future Directions

Using the distance theorem from [16] a number of results known for i.i.d. or subgaussian matrices can be considered in the inhomogeneous setting.

- Upper tail estimates for smallest singular value [9, 15]
- Upper tail estimates for intermediate singular values [17]
- Delocalization type results for eigenvectors [12, 11, 7]
- Many more (rank [8]), (no-gaps [12]), etc.

For rectangular matrices, the need for isotropic columns seems fundamental in the case of small aspect ratios, but not for the square case. Can this condition be relaxed?

End

Thank you!

Max Dabagia and Manuel Fernandez.
The smallest singular value of inhomogenous random rectangular matrices, 2024.

Alan Edelman.
Eigenvalues and condition numbers of random matrices.
SIAM journal on matrix analysis and applications, 9(4):543–560, 1988.

Alexander Litvak and Omar Rivasplata.
Smallest singular value of sparse random matrices.
arXiv preprint arXiv:1106.0938, 2011.

Alexander E. Litvak, Alalin Pajor, Mark Rudelson, and Nicole Tomczak-Jaegermann.
Smallest singular value of random matrices and geometry of random polytopes.
Advances in Mathematics, 195:491–523, 2005.

Galyna V Livshyts.
The smallest singular value of heavy-tailed not necessarily iid random matrices via random rounding.
Journal d'Analyse Mathématique, 145(1):257–306, 2021.

Galyna V. Livshyts, Konstantin Tikhomirov, and Roman Vershynin.
The smallest singular value of inhomogeneous square random matrices, 2019.

Anna Lytova and Konstantin Tikhomirov.

On delocalization of eigenvectors of random non-hermitian matrices.
Probability Theory and Related Fields, 177(1):465–524, 2020.

Mark Rudelson.
A large deviation inequality for the rank of a random matrix.
The Annals of Probability, 52(5):1992 – 2018, 2024.

Mark Rudelson and Roman Vershynin.
The littlewood–offord problem and invertibility of random matrices.
Advances in Mathematics, 218(2):600–633, 2008.

Mark Rudelson and Roman Vershynin.
The smallest singular value of a random rectangular matrix.
2008.

Mark Rudelson and Roman Vershynin.
Delocalization of eigenvectors of random matrices with independent entries.
2015.

Mark Rudelson and Roman Vershynin.
No-gaps delocalization for general random matrices.
2015.

Mark Rudelson and Roman Vershynin.
Small ball probabilities for linear images of high-dimensional distributions.
International Mathematics Research Notices, 2015(19):9594–9617, 2015.

Stanislaw J Szarek.
Condition numbers of random matrices.
Journal of Complexity, 7(2):131–149, 1991.

Kateryna Tatarko.
An upper bound on the smallest singular value of a square random matrix.
J. Complexity, 48:119–128, 2018.

Manuel Fernandez V.
A distance theorem for inhomogenous random rectangular matrices, 2024.

Feng Wei.
Upper bound for intermediate singular values of random matrices.
Journal of Mathematical Analysis and Applications, 445(2):1530–1547, 2017.