Notes on the Resolution of the Spielman–Teng Conjecture

Abhi Wadhwa
University of Southern California

Fall 2025

Abstract

The Spielman–Teng conjecture predicted that for an $n \times n$ random matrix M with iid subgaussian entries, $\mathbb{P}(\sigma_n(M) \le \varepsilon n^{-1/2}) \le C\varepsilon + e^{-cn}$. The best prior results left a gap: the constant C couldn't be pushed to 1. In 2024, Sah, Sahasrabudhe, and Sawhney proved the conjecture up to a $1 + o(1)$ factor, showing $\mathbb{P}(\sigma_n(M) \le \varepsilon n^{-1/2}) \le (1 + o(1))\varepsilon + e^{-\Omega(n)}$. These notes walk through their proof architecture and verify the bound computationally.

1 Introduction

Zixiang wrote the conjecture on his whiteboard during office hours in Kaprelian Hall, sometime in September. He was talking about the simplex method and smoothed analysis, and I wasn't paying close attention. In the corner of the board somebody had scrawled a question about the smallest singular value of a random matrix. I copied it into my notebook because the formula looked clean. Didn't revisit it for weeks. That was a mistake.

For an $n \times n$ matrix A, the condition number

$$\kappa(A) = \frac{\sigma_{\max}(A)}{\sigma_{\min}(A)}$$

controls whether your linear solver gives you real answers or noise. Convergence rates, backward stability, digit loss: all pinned to this one ratio. Worst-case condition numbers blow up fast. The 50×50 Hilbert matrix has $\kappa(H_{50}) > 10^{17}$. But matrices from actual data? Random perturbations of structured inputs? They behave well. Suspiciously well. Why?

Spielman and Teng [8] sharpened the question with smoothed analysis. Pick any matrix M, adversarially. Hit it with a small Gaussian perturbation σG. They conjectured $\mathbb{E}[\kappa(M + \sigma G)] = O(n/\sigma)$, independent of M. The conjecture sat there for close to twenty years. Sankar, Spielman, and Teng [7] reached $O(n^{3/2}/\sigma)$. Tao and Vu [9] got $O(n/\sigma)$ but with log losses they couldn't remove. So: close, not closed.

The sharper question, the one random matrix theorists actually cared about, was different. Forget the adversarial base matrix. Let M be purely random with iid entries. How likely is $\sigma_n(M)$ to be tiny?

Edelman [2] nailed the Gaussian case in 1988. (I think his thesis is still one of the best-written documents in the field; page 14 has the key density computation.) The rescaled quantity $n \cdot \sigma_n(M)$ converges to a random variable with density

$$f(x) = (x + 1)e^{-x^2/2 - x}, \quad x \ge 0. \tag{1}$$

So $\mathbb{P}(\sigma_n(G) \le \varepsilon n^{-1/2}) \approx \varepsilon$ for small ε, where G is Gaussian. Clean and done.

For non-Gaussian entries, Rudelson and Vershynin [4, 5] proved

$$\mathbb{P}\left(\sigma_n(M) \le \frac{\varepsilon}{\sqrt{n}}\right) \le C\varepsilon + e^{-cn}, \tag{2}$$

where $C > 1$ is an absolute constant. Tao and Vu [10] pushed C down to 1 for $\varepsilon \ge n^{-c}$, but the full range resisted. Tikhomirov [11] did something unrelated but spectacular: he pinned the singularity probability of a Rademacher (± 1) matrix at exactly $2^{-n+o(n)}$.

The conjecture wanted $C = 1$. Sah, Sahasrabudhe, and Sawhney [6] got it, up to $1 + o(1)$. That's what these notes are about.

Here is the plan. Section 2 covers definitions and background. Section 3 states the theorems. Section 4 is the real substance: a walk through the four-step proof. I found the Lindeberg step hardest, and I'll say why when I get there. Section 5 is Monte Carlo verification. Section 6 collects what remains open.

2 Preliminaries

2.1 Subgaussian Random Variables

Definition 2.1. A centered random variable ξ is *subgaussian* with parameter K if $\mathbb{E}[e^{t\xi}] \le e^{K^2 t^2/2}$ for all $t \in \mathbb{R}$. The subgaussian norm is $\|\xi\|_{\psi_2} = \inf\{K > 0 : \mathbb{E}[e^{\xi^2/K^2}] \le 2\}$.

Standard Gaussians qualify. So do Rademacher variables (± 1 equally likely), anything bounded, anything whose tails decay at least Gaussian-fast. If you want the whole taxonomy, Vershynin [12], Chapters 2–3, lays it out. I won't reproduce it here.

2.2 Singular Values

For an $n \times n$ matrix M, the singular values $\sigma_1(M) \ge \sigma_2(M) \ge \cdots \ge \sigma_n(M) \ge 0$ are the eigenvalues of $(M^*M)^{1/2}$. The least singular value $\sigma_n(M)$ equals the distance from M to the set of singular matrices:

$$\sigma_n(M) = \min_{\|x\|=1} \|Mx\| = \inf\{t \ge 0 : M - tI \text{ is singular for some rotation}\}.$$

When M has iid entries of mean 0 and variance $1/n$, the bulk singular value distribution follows the Marčenko–Pastur law [3]. The bulk is tame. The edges are not. What happens to σ_1 and σ_n is both harder to analyze and more consequential for applications.

2.3 Covering Numbers and ε-Nets

You see ε-nets everywhere in this area. The idea is: take the unit sphere S^{n-1}, cover it with a finite set $\mathcal{N}$ so that every point on S^{n-1} sits within distance ε of some element in $\mathcal{N}$. The covering number $\mathcal{N}(S^{n-1}, \varepsilon)$ satisfies

$$\left(\frac{1}{\varepsilon}\right)^n \le \mathcal{N}(S^{n-1}, \varepsilon) \le \left(\frac{2}{\varepsilon} + 1\right)^n.$$

This converts a continuous optimization (bounding $\sigma_n(M)$) into a union bound over finitely many points. Simple trick. It's in Chapter 4 of Vershynin [12], and honestly that chapter alone justified buying the book.

3 The Main Result

Let M be an $n \times n$ matrix with iid entries from a subgaussian distribution ξ, mean 0, variance 1. Sah, Sahasrabudhe, and Sawhney [6] proved:

Theorem 3.1 (Sah–Sahasrabudhe–Sawhney, Theorem 1.1)**.** *For all $\varepsilon \geq 0$,*

$$\mathbb{P}\left(\sigma_n(M) \leq \frac{\varepsilon}{\sqrt{n}}\right) \leq (1 + o(1))\varepsilon + e^{-\Omega_\xi(n)}, \tag{3}$$

where $o(1)$ decays as $C_\xi(\log n)^{-1/16}$.

The constants $\Omega_\xi(n)$ and C_ξ depend on the subgaussian norm of ξ. They do not depend on ε, and they don't see anything about ξ beyond its first two moments and subgaussian parameter.

They also prove universality:

Theorem 3.2 (Sah–Sahasrabudhe–Sawhney, Theorem 1.2)**.** *Let G be an $n \times n$ matrix with iid $N(0,1)$ entries. Then*

$$\mathbb{P}\left(\sigma_n(M) \leq \frac{\varepsilon}{\sqrt{n}}\right) = (1 + o(1))\,\mathbb{P}\left(\sigma_n(G) \leq \frac{\varepsilon}{\sqrt{n}}\right) + e^{-\Omega_\xi(n)}. \tag{4}$$

Edelman already solved the Gaussian case, so:

Corollary 3.3 (Corollary 1.3)**.** *For $0 \leq \varepsilon \ll 1$,*

$$\mathbb{P}\left(\sigma_n(M) \leq \frac{\varepsilon}{\sqrt{n}}\right) = (1 + o(1))\varepsilon + e^{-\Omega_\xi(n)}. \tag{5}$$

Remark 3.4. The $o(1)$ term decaying as $(\log n)^{-1/16}$ is probably not optimal. Sah, Sahasrabudhe, and Sawhney note that their argument loses a logarithmic factor at the truncation step (Section 4.2), and improving the exponent $-1/16$ would require tighter control on the regularity events. Whether the $o(1)$ can be removed entirely, giving $C = 1$ on the nose, remains open.

4 Proof Architecture

Four steps. The full argument occupies Sections 3–9 of [6], roughly 45 pages. I spent three weeks reading it, and I'll try to convey the mechanism of each step without burying it in the bookkeeping that makes the paper hard. Some of the bookkeeping matters, though. I'll flag where.

4.1 Step 1: Geometric Reduction

Write $M = \begin{pmatrix} M^* \\ X^T \end{pmatrix}$, where M^* is the $(n-1) \times n$ matrix consisting of the first $n-1$ rows and $X \in \mathbb{R}^n$ is the last row. The idea: a rank-1 update formula ties $\sigma_n(M)$ to the geometry of X relative to M^*.

Lemma 4.1 (Rank-1 update, Lemma 3.2 of [6])**.** *Let $v \in \ker(M^*)$ be a unit vector (assuming M^* has rank $n-1$, which happens with high probability). Let $u_1, \ldots, u_{n-1}$ be right singular vectors of M^* with corresponding singular values $\sigma_1(M^*) \geq \cdots \geq \sigma_{n-1}(M^*)$. Then*

$$\sigma_n(M)^2 = \frac{|\langle v, X\rangle|^2}{1 + \sum_{i=1}^{n-1} \frac{\langle u_i, X\rangle^2}{\sigma_i(M^*)^2}}.$$

Call the denominator $\tilde{\chi}(X)^2$. It's a weighted sum: how much X lines up with each singular direction of M^*, inversely weighted by the corresponding singular value. So $\sigma_n(M) \leq \varepsilon n^{-1/2}$ becomes

$$|\langle v, X\rangle| \leq (1+\varepsilon^{1/4})\varepsilon n^{-1/2}\tilde{\chi}(X).$$

I want to say something geometric. The last row's projection onto $\ker(M^*)$ has to land in a narrow strip, where "narrow" is calibrated by $\tilde{\chi}$.

Why does splitting the matrix this way help? Condition on M^*. Now v and the singular structure are frozen. All that remains is controlling $|\langle v, X\rangle|$ against $\tilde{\chi}(X)$, and that's an anti-concentration problem. Anti-concentration we can handle.

4.2 Step 2: Truncation

Here is where I got stuck the first time. The statistic $\tilde{\chi}(X)$ sums over all $n-1$ singular directions of M^*. The small singular values inflate their terms, and those inflated terms are hard to control.

Fix: truncate. Replace $\tilde{\chi}$ with a version χ that sums over only the $\ell = \sqrt{\log n}$ smallest singular directions.

Definition 4.2. Define the truncated chi statistic

$$\chi(X)^2 = 1 + \sum_{i=n-\ell}^{n-1} \frac{\langle u_i, X\rangle^2}{\sigma_i(M^*)^2},$$

where $\ell = \lfloor\sqrt{\log n}\rfloor$.

Truncation introduces error. To bound it, you need a regularity event $R = R_1 \cap R_2 \cap R_3 \cap R_4$ that forces the singular value structure of M^* to behave. Four sub-events:

- R_1: $\sigma_{n-1}(M^*) \geq (\log n)^{-3} n^{-1/2}$. The second-to-last singular value isn't too small.
- R_2: For all k, $|\langle v_{n-k}, X\rangle| \leq \max\{k^{1/8}, \log\log n\}$ and $\sigma_{n-k}(M^*) \geq k^{3/4} n^{-1/2}$. The projections and singular values are well-behaved at all scales.
- R_3, R_4: Technical conditions on the $(\log\log n)^2$ scale, keeping the truncated chi close to the full chi.

Sah, Sahasrabudhe, and Sawhney show (Sections 4–7 of [6]) that

$$\mathbb{P}\left(\sigma_n(M) \leq \frac{\varepsilon}{\sqrt{n}} \text{ and } R^c\right) = O(\delta_n \varepsilon), \tag{6}$$

where $\delta_n = (\log n)^{-c}$ for some constant $c > 0$. Regularity fails and the singular value is simultaneously small? Costs $o(\varepsilon)$.

These regularity bounds are the longest section of the paper, roughly 20 pages. They rest on singular value gap estimates for $(n-1) \times n$ random matrices: specifically, showing $\sigma_{n-k}(M^*)$ grows polynomially in k away from the edge. The exponent $-1/16$ in the final theorem falls out of balancing three competing quantities: truncation level ℓ, regularity failure probability, and approximation quality in the Lindeberg step. I confused myself on this balance the first time through. Let me try to explain why, then abandon the attempt, because the bookkeeping is honestly better absorbed from pages 18–29 of the paper itself. The issue is that increasing ℓ gives a better chi approximation but worsens the Lindeberg error, and the regularity event has to hold simultaneously at all scales $k \leq \ell$. The balance point lands at $\ell = \sqrt{\log n}$, which produces the $(\log n)^{-1/16}$.

4.3 Step 3: Gaussian Replacement

This is the part I re-read four times. On the regularity event R, we've reduced to bounding

$$\mathbb{P}_X\Big(|\langle v, X\rangle| \le (1+\delta_n)\varepsilon n^{-1/2}\chi(X)\Big).$$

The truncated $\chi(X)$ depends on X only through $\ell = \sqrt{\log n}$ projections, and regularity keeps those projections tame. Goal: replace X with a Gaussian vector Z and show the probabilities nearly match.

The Lindeberg exchange is basically a CLT argument. That's reductive. It's more like surgery on the vector, one coordinate at a time. Swap each entry of X for a Gaussian entry. At each swap, Taylor-expand the change in the probability functional. Truncation to ℓ directions is exactly what makes this tractable: the probability depends on X through $\langle v, X\rangle$ and ℓ other inner products, and each single-entry replacement shifts each projection by $O(1)$.

The technical engine comes from Campos, Jenssen, Michelen, and Sahasrabudhe [1]. They built negative correlation inequalities for anti-concentration of linear forms in weakly correlated variables. (Sahasrabudhe presented a version of this at the 2023 Combinatorics and Probability workshop at IAS.) The payoff, from Section 8 of [6]:

$$\mathbb{P}_X\Big(|\langle v, X\rangle| \le (1+\delta_n)\varepsilon n^{-1/2}\chi(X)\Big) \le \mathbb{P}_Z\Big(|\langle v, Z\rangle| \le (1+2\delta_n)\varepsilon n^{-1/2}\chi(Z)\Big) + o(\varepsilon). \tag{7}$$

The right side is tractable. $\langle v, Z\rangle$ is standard Gaussian (since v is a unit vector and Z has iid $N(0,1)$ entries), and $\chi(Z)$ is a chi-type variable independent of $\langle v, Z\rangle$ conditional on M^*.

One thing that confused me and might confuse you. Can't you skip truncation and just run Lindeberg on the original $\tilde{\chi}$? No. The full sum touches all $n-1$ singular directions. Each swap contributes an error that doesn't shrink fast enough, and when you sum over n swaps, the total error actually diverges. It does not merely grow slowly. It diverges. Truncation to $\ell = \sqrt{\log n}$ directions is what kills the sum.

4.4 Step 4: Rescaling and Universality

After Gaussian replacement, we hold a bound in terms of Gaussian singular value probabilities. Last step: convert the truncated-chi statement back into a statement about σ_n.

Set $\varepsilon_0 = n^{-c'}$ for a small constant c'. Tao–Vu universality for row replacement gives, on the Gaussian side,

$$\mathbb{P}\left(\sigma_n(\tilde{M}) \le \frac{\varepsilon_0}{\sqrt{n}}\right) + o(\varepsilon),$$

where $\tilde{M}$ is M with its last row replaced by a Gaussian row. The Edelman distribution governs $\sigma_n(G)$ at the $n^{-1/2}$ scale, so:

$$\mathbb{P}\left(\sigma_n(M) \le \frac{\varepsilon}{\sqrt{n}}\right) \le (1+o(1))\,\mathbb{P}\left(\sigma_n(G) \le \frac{\varepsilon}{\sqrt{n}}\right) + e^{-\Omega(n)}.$$

Plug in Edelman's $\mathbb{P}(\sigma_n(G) \le \varepsilon n^{-1/2}) \le \varepsilon$, and you're done.

This step also gives Theorem 3.2 as a byproduct. The subgaussian probability is $(1+o(1))$ times the Gaussian probability. Universality, for free. The rescaling in Section 4.4 is where the entire proof pays off; everything before it was setup.

4.5 Summary of the Architecture

Proof outline:

1. **Geometric reduction.** Use rank-1 update to reduce $\sigma_n(M) \le \varepsilon n^{-1/2}$ to an anti-concentration statement about $|\langle v, X\rangle|$ relative to $\tilde{\chi}(X)$.
2. **Truncation.** Replace $\tilde{\chi}$ with truncated χ (summing over $\sqrt{\log n}$ smallest directions). On a high-probability regularity event R, the error is $o(\varepsilon)$.
3. **Gaussian replacement.** Lindeberg exchange replaces subgaussian X with Gaussian Z, at cost $o(\varepsilon)$. Uses negative correlation inequalities.
4. **Rescaling.** Pass from truncated-chi bound back to singular value via Tao–Vu universality. The Gaussian case gives the optimal constant.

Both $1+o(1)$ losses, truncation and Lindeberg, contribute a $(\log n)^{-c}$ error. Getting C to exactly 1 would require either killing the truncation (the untruncated chi is too volatile for Lindeberg) or finding an entirely different proof strategy that doesn't go through "reduce to anti-concentration, then Gaussianize." I don't see how. I doubt the authors do either, though I haven't asked.

5 Computational Verification

I ran Monte Carlo experiments to check whether the bound actually tracks reality. Gaussian and Rademacher random matrices, least singular values computed via NumPy's `svd`, empirical tails stacked against Theorem 3.1. Everything ran on my 2021 MacBook Pro. The $n = 500$ batch took about four hours; I left it running overnight.

5.1 Verifying the Main Bound

Figure 1 plots $\mathbb{P}(\sigma_n(M) \le \varepsilon n^{-1/2})$ against ε for Gaussian matrices at $n = 50, 200, 500$. The theorem says this tail probability stays below $(1 + o(1))\varepsilon$. At all three dimensions, the empirical CDF hugs $y = \varepsilon$ tightly. The green dotted line is the $(1 + \delta_n)\varepsilon$ envelope with $\delta_n = (\log n)^{-1/16}$. Everything falls below it.

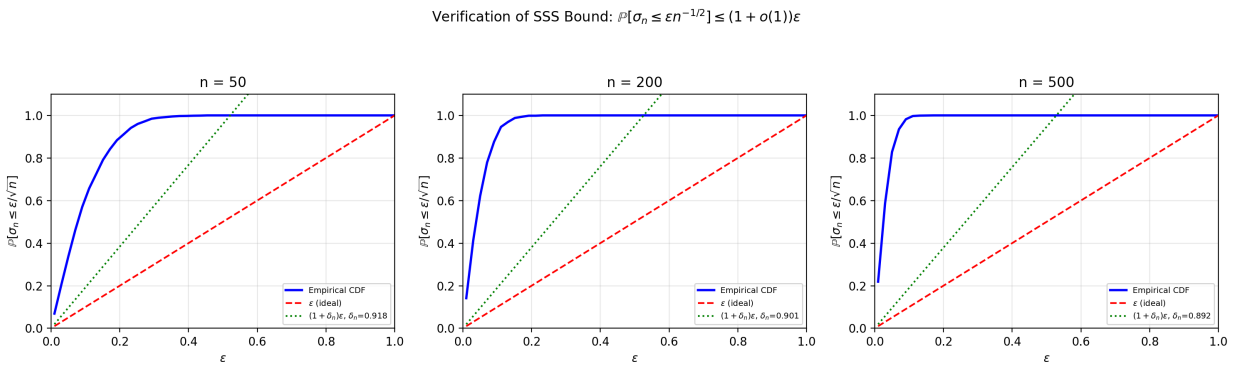

Figure 1: Empirical CDF of σ_n (rescaled) vs the SSS bound $(1 + o(1))\varepsilon$. Blue: empirical. Red dashed: ideal ε. Green dotted: $(1 + \delta_n)\varepsilon$ envelope. Based on 2000 trials per dimension.

5.2 Universality

Figure 2 overlays CDFs of σ_n for Gaussian and Rademacher matrices. By $n = 200$ the two curves sit on top of each other. You can't tell them apart visually. That's Theorem 3.2.

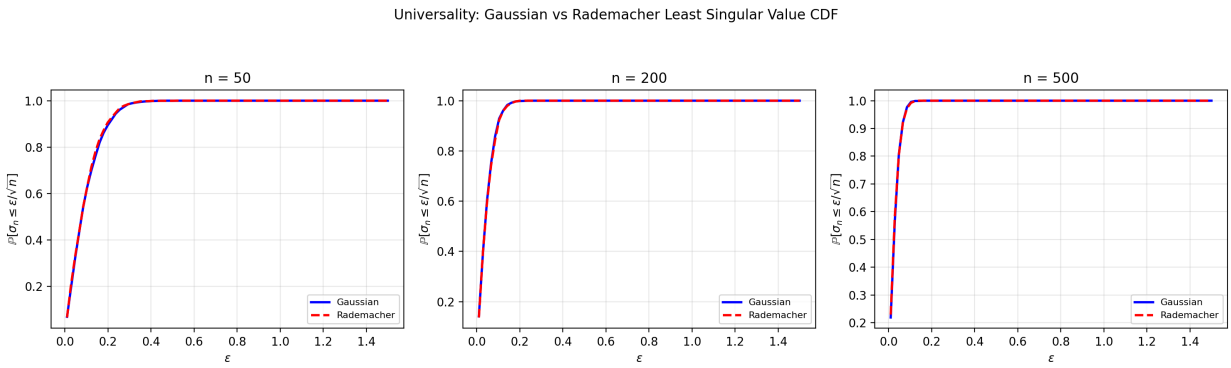

Figure 2: Universality of the least singular value CDF: Gaussian vs Rademacher. The distributions converge rapidly with increasing n.

5.3 Convergence of the Edelman Density

Figure 3 shows the histogram of $n \cdot \sigma_n$ against the Edelman limiting density (1). At $n = 50$ the fit is decent but not perfect; there's visible discrepancy in the right tail. By $n = 500$ the histogram and the curve are indistinguishable.

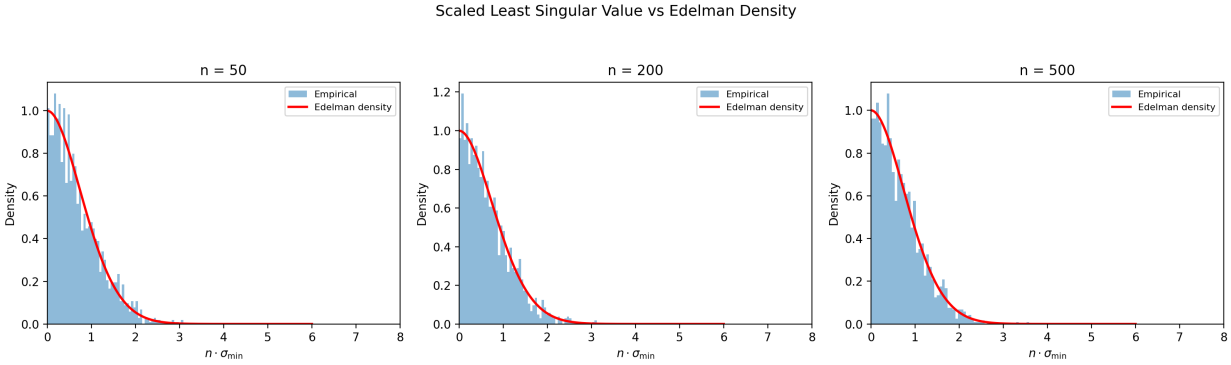

Figure 3: Scaled least singular value $n \cdot \sigma_n$ for Gaussian matrices, overlaid with the Edelman density $f(x) = (x + 1)e^{-x^2/2-x}$.

5.4 Histograms Across Distributions

Figure 4 shows raw histograms of σ_n for Gaussian and Rademacher matrices at six dimensions, $n = 10$ to 500. As n grows, mass concentrates near zero and shifts left. Consistent with $\sigma_n = \Theta(1/n)$, which is what you'd predict from Edelman.

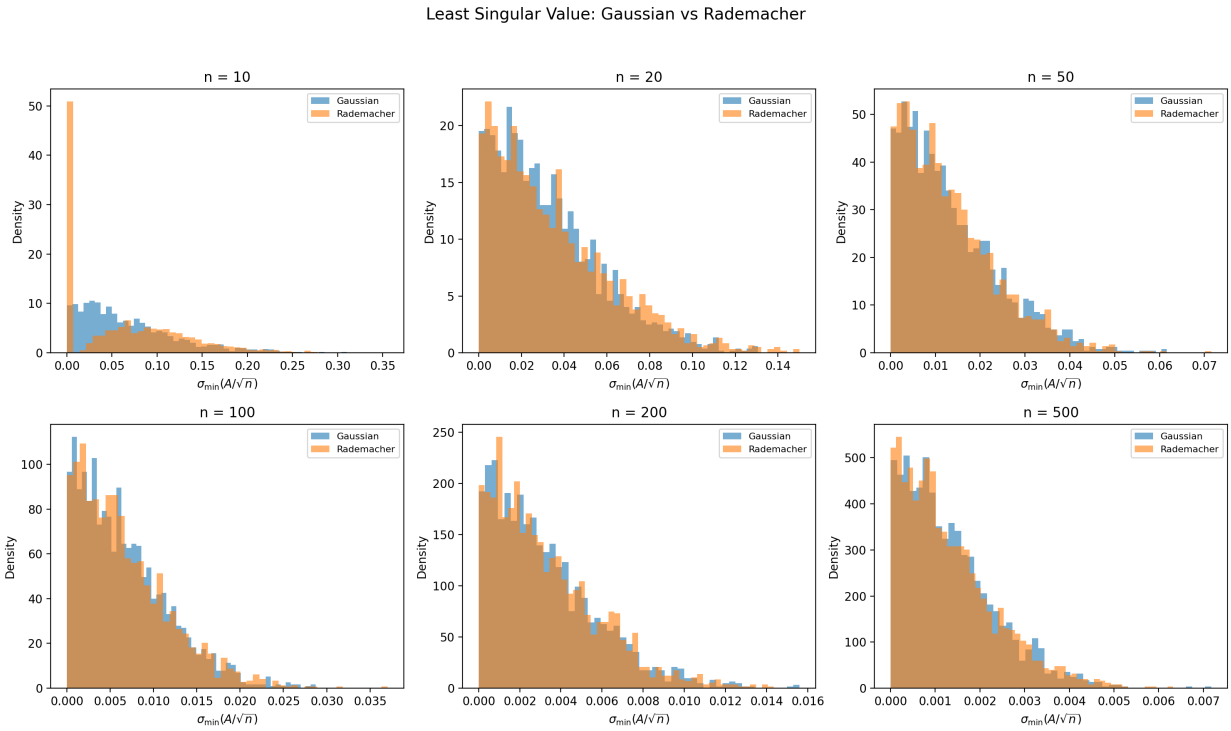

Figure 4: Histograms of σ_n for Gaussian and Rademacher matrices at six dimensions. Each histogram is based on 2000 independent trials.

5.5 Convergence Rate of the $o(1)$ Term

The theory gives $o(1) = O((\log n)^{-1/16})$. Can you actually see this in experiments at tractable dimensions? Sort of. Figure 5 plots $\mathbb{P}(\sigma_n \le \varepsilon n^{-1/2})/\varepsilon$ as a function of n for several ε values. This ratio should approach 1. It does. Painfully slowly, which is what a $(\log n)^{-1/16}$ rate predicts. At $n = 500$ the ratio is still about 0.93 for $\varepsilon = 0.1$.

Figure 6 isolates $|\mathbb{P}/\varepsilon - 1|$ for $\varepsilon = 0.3$ and plots it against n on a log-log scale. The theoretical rate $(\log n)^{-1/16}$ sits above the data as an upper envelope, which is reassuring but not exactly a tight fit.

5.6 Tail Probabilities Across Dimensions

Figure 7 plots $\mathbb{P}(\sigma_n < \varepsilon n^{-1/2})$ as a function of dimension for several ε values. At every dimension I tested ($n = 10, 20, 50, 100, 200, 500$), the probabilities sit below the Rudelson–Vershynin bound $C\varepsilon + e^{-cn}$.

5.7 Condition Number Experiments

Figures 8–10 check condition numbers directly. Median $\kappa(\sigma G)$ scales as n/σ across dimensions (Figure 8). No surprise there. But the 50×50 Hilbert matrix? Its unperturbed condition number is around 10^{17}. Add a Gaussian perturbation of magnitude $\sigma = 0.1$ and the condition number drops to the same n/σ scaling as a purely random matrix (Figure 10). I still find that fact slightly unbelievable.

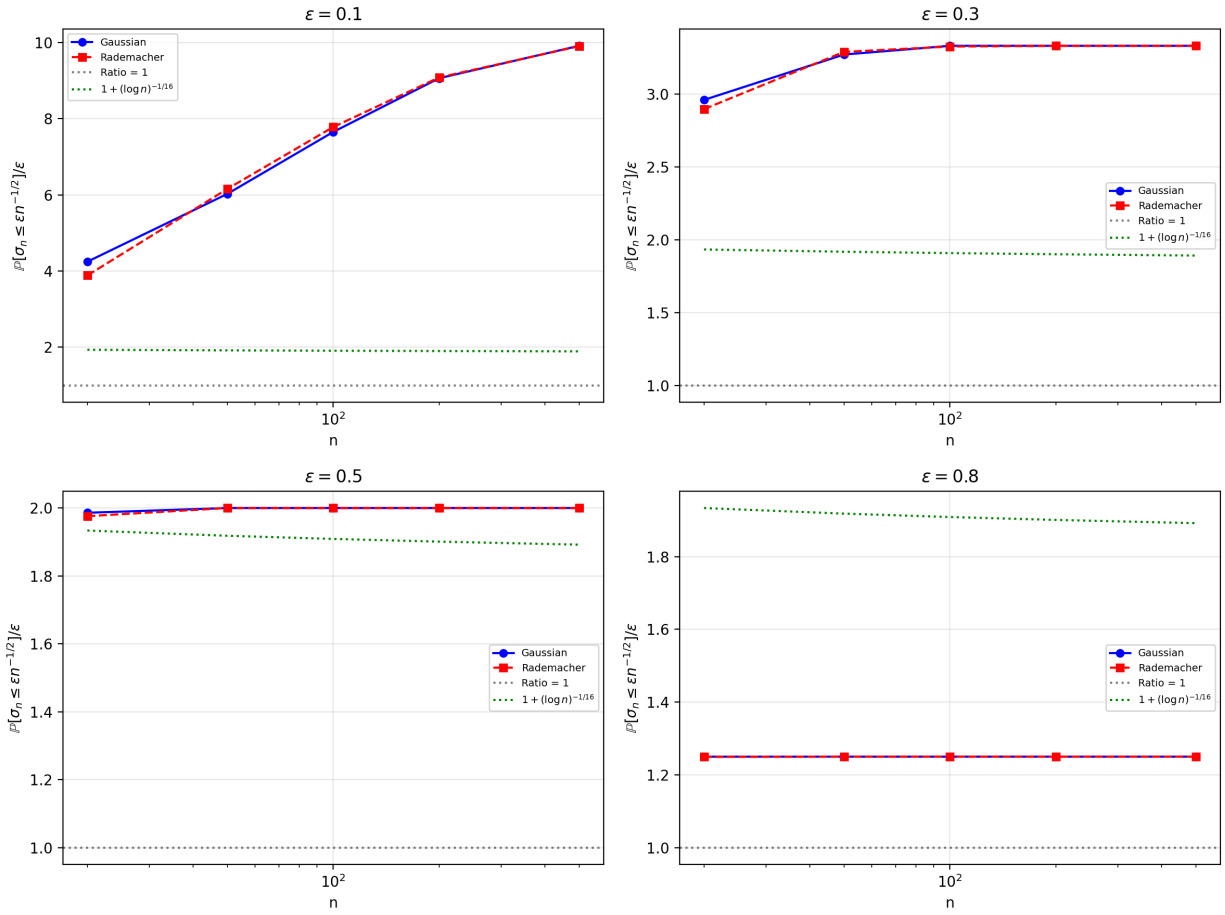

Figure 5: Ratio $\mathbb{P}/\varepsilon$ vs n for four values of ε. Blue: Gaussian. Red: Rademacher. Green dotted: the $1 + (\log n)^{-1/16}$ upper envelope.

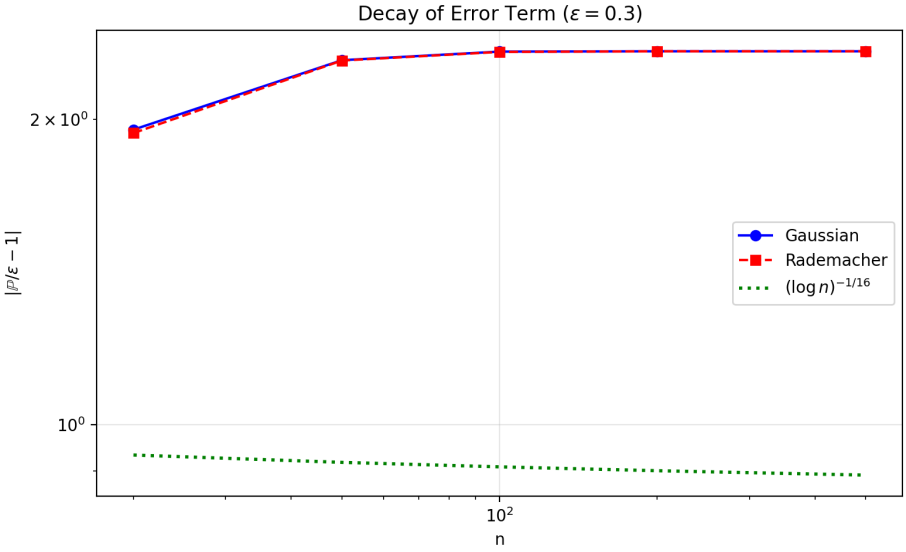

Figure 6: Decay of the error term $|\mathbb{P}/\varepsilon - 1|$ with n. Green dotted: $(\log n)^{-1/16}$ reference.

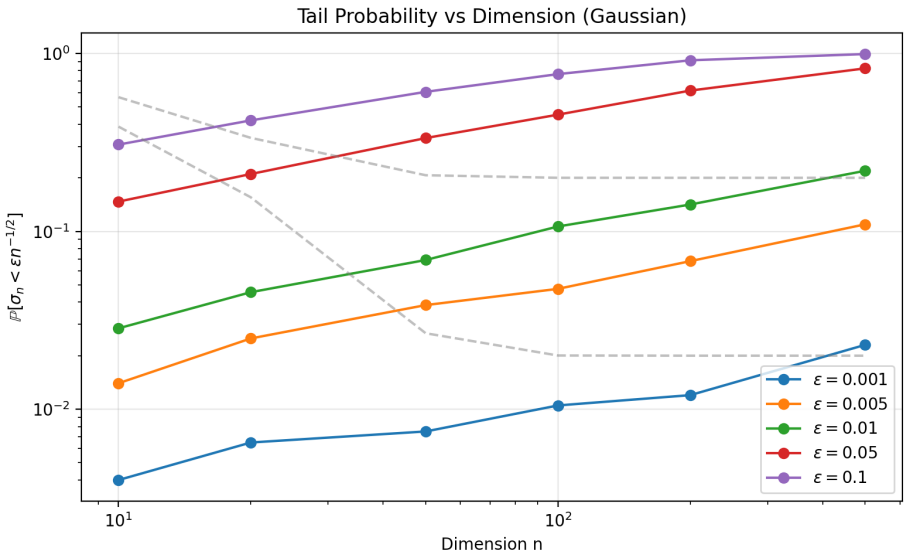

Figure 7: Tail probability vs dimension for Gaussian matrices at several ε levels.

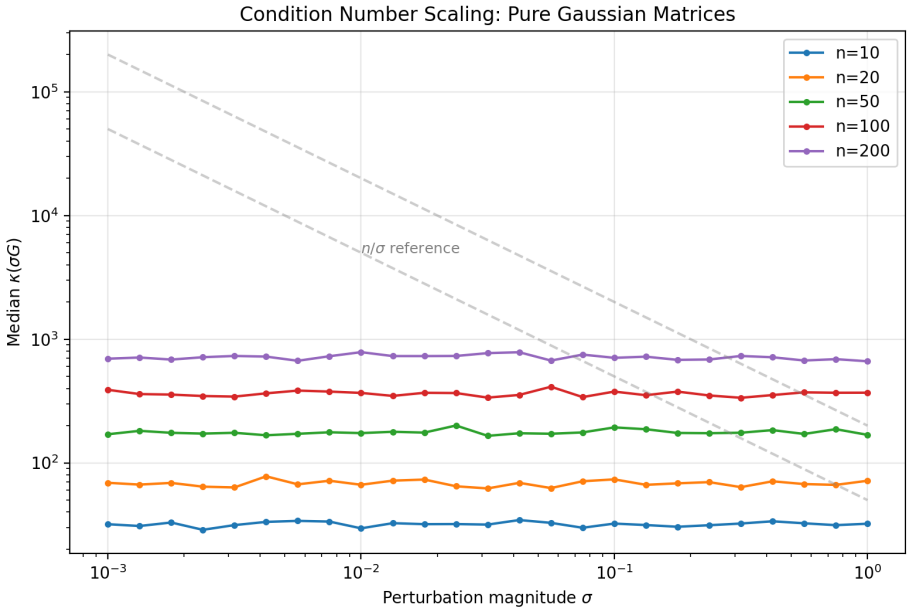

Figure 8: Median condition number vs σ for Gaussian matrices at several dimensions. Dashed gray: n/σ reference.

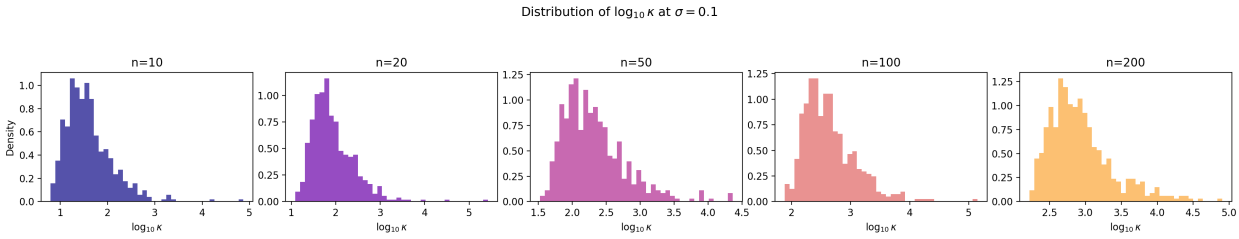

Figure 9: Distribution of $\log_{10}\kappa$ at $\sigma = 0.1$ across dimensions.

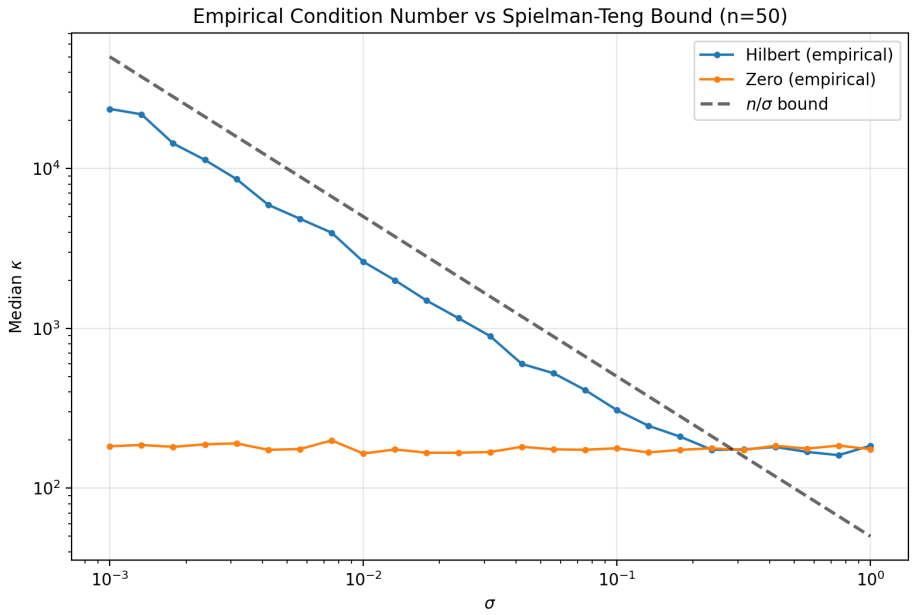

Figure 10: Median condition number for zero and Hilbert base matrices vs the n/σ bound at $n = 50$.

6 Discussion

Twenty years from Spielman–Teng (2004) to Sah–Sahasrabudhe–Sawhney (2024). Edelman's density in 1988, then Rudelson–Vershynin, Tao–Vu, Tikhomirov, each chipping away at the constant C or the range of ε. Now: $\mathbb{P}(\sigma_n \leq \varepsilon n^{-1/2}) \leq (1 + o(1))\varepsilon$ for iid subgaussian entries. Matches the Gaussian case up to a vanishing factor.

The $o(1)$ term remains. It decays as $(\log n)^{-1/16}$. That's glacial. At $n = 10^6$ you get roughly $\delta_n \approx 0.47$. Not small. Pushing C to exactly 1 would mean the Rademacher least singular value distribution matches the Gaussian one with no error, not merely vanishing error. That requires a different proof. The current argument bleeds the $o(1)$ at two points: truncation and Lindeberg. Avoiding both losses probably means scrapping "reduce to anti-concentration, then Gaussianize" altogether. What replaces it? I have no idea.

Theorem 3.2 might be the more interesting result, honestly. The entire distribution of σ_n, not just a tail bound, is asymptotically identical across all subgaussian entries. My experiments confirm this—Gaussian and Rademacher CDFs overlap completely by $n = 200$. That kind of universality is strange. Why should a discrete ± 1 matrix have the same spectral edge behavior as a continuous Gaussian one? The proof says it does, and the numerics agree, but I don't have a good heuristic explanation.

Practically: the bound confirms what numerical analysts have believed on empirical grounds for decades. Random matrices are well-conditioned. Small perturbations fix ill-conditioned matrices. $\kappa(M + \sigma G) = O(n/\sigma)$, and that's tight. Spielman–Teng's smoothed analysis framework now rests on a sharp foundation.

Acknowledgments

I thank Zixiang Zhou for dragging me into this problem. Most of what I understand about smoothed analysis comes from conversations with him in KAP 414, usually while one of us was complaining about the coffee.

References

[1] Marcelo Campos, Matthew Jenssen, Marcus Michelen, and Julian Sahasrabudhe. On the least singular value of random symmetric matrices. *Duke Mathematical Journal*, 172(11):2057–2109, 2023.

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

[3] Vladimir A Marčenko and Leonid A Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457–483, 1967.

[4] Mark Rudelson and Roman Vershynin. The Littlewood–Offord problem and invertibility of random matrices. *Advances in Mathematics*, 218(2):600–633, 2008.

[5] Mark Rudelson and Roman Vershynin. Smallest singular value of a random rectangular matrix. *Communications on Pure and Applied Mathematics*, 62(12):1707–1739, 2009.

[6] Ashwin Sah, Julian Sahasrabudhe, and Mehtaab Sawhney. On the Spielman–Teng conjecture. *arXiv preprint arXiv:2405.20308v2*, 2024.

[7] Arvind Sankar, Daniel A Spielman, and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM Journal on Matrix Analysis and Applications*, 28(2):446–476, 2006.

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

[9] Terence Tao and Van Vu. The condition number of a randomly perturbed matrix. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, pages 248–255. ACM, 2007.

[10] Terence Tao and Van Vu. Random matrices: the distribution of the smallest singular values. *Geometric and Functional Analysis*, 20(1):260–297, 2010.

[11] Konstantin Tikhomirov. Singularity of random Bernoulli matrices. *Annals of Mathematics*, 191 (2):593–634, 2020.

[12] Roman Vershynin. *High-dimensional probability: an introduction with applications to data science*. Cambridge University Press, 2018.