International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(18s) 2026 pp. 1100-1110
https://doi.org/10.70917/ijcisim-2026-4591
Article

Number-Theoretic Bounds in Optimization: A Research Program Connecting Random-Matrix Universality in Loss Landscapes to Analytic Number Theory

Jamal Salah

College of Applied and Health Sciences, A'Sharqiyah University,
Post Box No. 42, Post Code No. 400 Ibra, Sultanate of Oman
Email: damous73@yahoo.com
A Technical Perspective Paper, with Numerical Investigations

Abstract: The geometry of non-convex loss landscapes and the spectral statistics of the Riemann zeta function's nontrivial zeros have both been described, independently, using random matrix theory (RMT). This paper examines that connection with the technical care it requires, and — beyond analysis — reports original numerical experiments designed to stress-test the claims involved. We first derive, rather than merely cite, the classical RMT results that are actually proven: the Wigner semicircle law, the Wigner surmise for level spacing, the Marchenko–Pastur law, and the Bai–Yin/Tracy–Widom results governing extreme eigenvalues of Wishart matrices. We show these proven results already furnish a rigorous, non-conjectural instance of a 'deterministic ceiling governing chaotic structure,' the pattern that motivates interest in Robin's inequality (equivalent to the Riemann Hypothesis). We extend the toy model with the Baik–Ben Arous–Péché (BBP) phase transition, which describes exactly when an outlier eigenvalue detaches from a random matrix bulk — the same bulk-plus-outlier structure reported empirically for trained neural network Hessians — and we numerically confirm the BBP prediction to within simulation error. We then run a second, independent experiment: training small neural networks directly, computing their exact Hessians, and directly testing whether the empirically claimed Gaussian Orthogonal Ensemble (GOE) level-spacing statistics reproduce at small scale. They do not: our networks instead reproduce the well-documented Hessian rank-degeneracy phenomenon (a large near-zero eigenvalue pile), and once that degenerate cluster is excluded, the remaining 'informative' eigenvalues are still too few for either GOE or GUE repulsion to be statistically distinguishable from an uncorrelated (Poisson) baseline. We report this as a genuine, informative negative result that calibrates the scale at which RMT universality claims for neural network Hessians can be expected to hold, and we use it to sharpen — rather than abandon — the research program's open conjectures.
Keywords: Random matrix theory; Riemann Hypothesis; Robin's inequality; Wigner semicircle law; Gaussian Orthogonal Ensemble; loss landscape geometry; Hessian spectral statistics; Baik–Ben Arous–Péché transition; colossally abundant numbers; neural network optimization.

1. Introduction

Two independent lines of mathematics have converged on the same descriptive language for 'structured randomness': random matrix theory (RMT). In analytic number theory, the spacing statistics of the nontrivial zeros of the Riemann zeta function match the eigenvalue spacing statistics of the Gaussian Unitary Ensemble (GUE) [1], [2]. In deep learning theory, the Hessian of the training loss at critical points has been shown, both analytically and empirically, to exhibit spectra with a bulk-plus-outlier structure resembling specific random matrix ensembles [3], [4].


Copyright: © 2026 by the authors

This paper does three things beyond restating that observation. First, it derives the relevant classical results explicitly, so that claims about them can be checked rather than taken on faith. Second, it extends the analytical toy model with a further classical result — the BBP phase transition — that speaks directly to the bulk-plus-outlier structure emphasized in the empirical deep learning literature, and verifies it numerically. Third, and most importantly for assessing the program's current status, it runs an original small-scale experiment training real neural networks, computing their exact Hessians, and directly testing the GOE claim rather than assuming it. The result is a mixed picture — a clean confirmation on the purely mathematical (Wishart/BBP) side, and an informative failure to replicate GOE statistics at small network scale — which we argue sharpens the research program considerably relative to treating it as a purely conceptual analogy.

2. Mathematical Preliminaries

2.1 Random Matrix Ensembles and the Semicircle Law

A real symmetric $N \times N$ Wigner matrix H has independent entries H_{ij} ($i \le j$) with mean zero, off-diagonal variance σ^2/N, and finite higher moments. The Gaussian Orthogonal Ensemble (GOE) is the special case where entries are Gaussian and the measure is invariant under $H \mapsto O\,H\,O^{\mathsf{T}}$ for orthogonal O; the joint eigenvalue density is

$$p(\lambda_1, \dots, \lambda_N) \propto \exp\left(-\frac{\beta}{4}\sum_{i=1}^{N}\lambda_i^2\right)\prod_{i<j}\left|\lambda_i - \lambda_j\right|^{\beta}$$

The Dyson index β encodes the symmetry class and controls the strength of level repulsion, as shown in Section 2.2. The empirical spectral distribution of a Wigner matrix converges almost surely, as $N \to \infty$, to the Wigner semicircle law:

$$\rho_{sc}(x) = \frac{1}{2\pi\sigma^2}\sqrt{4\sigma^2 - x^2}, \qquad |x| \le 2\sigma$$

This follows from the moment method: $E\left[(1/N)\,\mathrm{Tr}\left(H^{2k}\right)\right]$ counts weighted closed walks of length $2k$, dominated as $N \to \infty$ by non-crossing pair partitions, whose count is the Catalan number $C_k = \frac{1}{k+1}\binom{2k}{k}$ — exactly the moments of the semicircle distribution.

2.2 The Wigner Surmise: Level Spacing Statistics

For the $N = 2$ GOE case, with eigenvalues λ_1, λ_2, spacing $s = \lambda_1 - \lambda_2$, and mean $m = (\lambda_1 + \lambda_2)/2$, the joint density $p(\lambda_1, \lambda_2) \propto |s|\exp(-(\lambda_1^2 + \lambda_2^2)/2)$ becomes, after the substitution $\lambda_1^2 + \lambda_2^2 = 2m^2 + s^2/2$ and integrating out m, $p(s) \propto s \cdot e^{-s^2/4}$ for $s > 0$. Normalizing and rescaling to unit mean spacing ($t = s/\sqrt{\pi}$) gives the Wigner surmise for GOE:

$$p_{GOE}(t) = \frac{\pi}{2}\,t\,e^{-\pi t^2/4}$$

The corresponding GUE ($\beta = 2$) derivation gives $p(t) = \frac{32}{\pi^2}t^2e^{-4t^2/\pi}$: quadratic, rather than linear, vanishing at $t \to 0$ — stronger level repulsion. This distinction between $\beta = 1$ and $\beta = 2$ repulsion is exactly what is tested empirically in Section 4.

2.3 Zeta Zeros, the Montgomery–Odlyzko Law, and Quantum Chaos

Montgomery [1] conjectured, and Odlyzko [2] confirmed numerically to extraordinary precision, that the pair correlation of normalized zeta-zero spacings matches the GUE pair correlation function:

$$R_2(x) = 1 - \left(\frac{\sin\pi x}{\pi x}\right)^2$$

This is a GUE ($\beta = 2$) statistic, consistent with the Hilbert–Pólya conjecture and the Bohigas–Giannoni–Schmit conjecture [5] that chaotic quantum systems without time-reversal symmetry exhibit GUE statistics.

2.4 Robin's Inequality: Statement and the Origin of e^γ

Robin's theorem [6] states that RH is equivalent to the following inequality holding for every $n > 5040$:

$$\sigma(n) < e^\gamma\, n \log\log n \qquad (n > 5040)$$

History: from Gronwall to Robin.

The inequality sits at the end of a short, well-defined chain. Gronwall's theorem [7] established the unconditional asymptotic $\limsup_{n\to\infty} \left(\frac{\sigma(n)}{n\log\log n}\right) = e^\gamma$, so $e^\gamma\, n \log\log n$ is, unconditionally, the exact growth envelope of $\sigma(n)/n$ — the question is only whether any individual n is allowed to sit above the envelope once $\log\log n$ is inserted with a strict inequality. Ramanujan [8], in the manuscript on highly composite numbers later edited by Robin and Nicolas, showed that if RH holds then $\sigma(n) < e^\gamma\, n \log\log n$ for all sufficiently large n. Robin [6] sharpened Ramanujan's conditional bound into an exact biconditional — RH is equivalent to the inequality holding for every $n > 5040$, with no asymptotic qualifier — by showing the two directions separately: RH implies the bound for all $n > 5040$, and its failure would force violations infinitely often.

The 27 exceptions below the threshold.

The threshold $n = 5040$ is not a technical convenience; it is exhaustive. The inequality fails for exactly 27 integers overall, and 5040 is the largest of them: $n =$ 2, 3, 4, 5, 6, 8, 9, 10, 12, 16, 18, 20, 24, 30, 36, 48, 60, 72, 84, 120, 180, 240, 360, 720, 840, 2520, 5040 (sequence A067698). Every one of these is checked by direct computation of $\sigma(n)$; none of them bears on RH, since Robin's equivalence is stated only for $n > 5040$. Notice that this list already contains 12, 60, 120, 360, 2520 and 5040 — five of the first six colossally abundant numbers below — which is exactly the phenomenon the next two paragraphs explain: the colossally abundant numbers are the integers that come closest to violating the bound, so the small colossally abundant numbers are also the smallest confirmed exceptions.

Why e^γ appears: the primorial calculation.

Writing $n = N_k$ for the k -th primorial (the product of the first k primes), the multiplicativity of σ gives:

$$\frac{\sigma(n)}{n} = \prod_{p\le x}\left(1 + \frac{1}{p}\right) = \left[\prod_{p\le x}\left(1 - \frac{1}{p^2}\right)\right]\left[\prod_{p\le x}\left(1 - \frac{1}{p}\right)\right]^{-1}$$

The first bracketed factor converges to $1/\zeta(2) = 6/\pi^2$ as $x \to \infty$; the second is governed by Mertens' third theorem, which gives it the asymptotic $e^\gamma \log x$. Since $\log n \sim x$ by the Prime Number Theorem, this yields:

$$\frac{\sigma(N_k)}{N_k} \sim \frac{6}{\pi^2}\, e^\gamma \log\log N_k$$

which is exactly where e^γ enters the inequality. Because $6/\pi^2 \approx 0.608 < 1$, however, primorials only ever reach about 61% of the e^γ ceiling — they are not the extremal sequence. The integers that actually push $\sigma(n)/n$ closest to the $e^\gamma \log\log n$ envelope are the colossally abundant numbers [8], [9], [6], which weight small primes by more than one power each rather than including every prime exactly once.

Colossally abundant numbers: the extremal sequence.

A colossally abundant (CA) number n is one for which there exists $\varepsilon > 0$ such that $\sigma(n)/n^{1+\varepsilon} \geq \sigma(m)/m^{1+\varepsilon}$ for every $m > 1$ — informally, n maximizes the abundancy ratio at every exponent scale simultaneously. Alaoglu and Erdős [9] showed each CA number has the form $n = 2^{a_1} \cdot 3^{a_2} \cdot 5^{a_3} \cdots p_k^{a_k}$ with $a_1 \geq a_2 \geq \cdots \geq a_k \geq 1$ — small primes are always raised to exponents at least as large as larger primes, so the factorization exponents form a non-increasing staircase. The first CA numbers are 2, 6, 12, 60, 120, 360, 2520, 5040, 55440, 720720, 1441440, 4324320, 21621600, 367567200, 6983776800, … (OEIS A004490); the first 15 of these coincide with the superior highly composite numbers. Figure 1a plots $\frac{\sigma(n)}{n \log\log n}$ against n on a log scale for all n up to 10^5, with the CA numbers marked: they trace out the closest approach to the e^γ ceiling at every scale, sitting above it (consistent with the 27 known exceptions) up to $n = 5040$ and dropping below it immediately afterward — the qualitative content of Robin's equivalence in one picture.

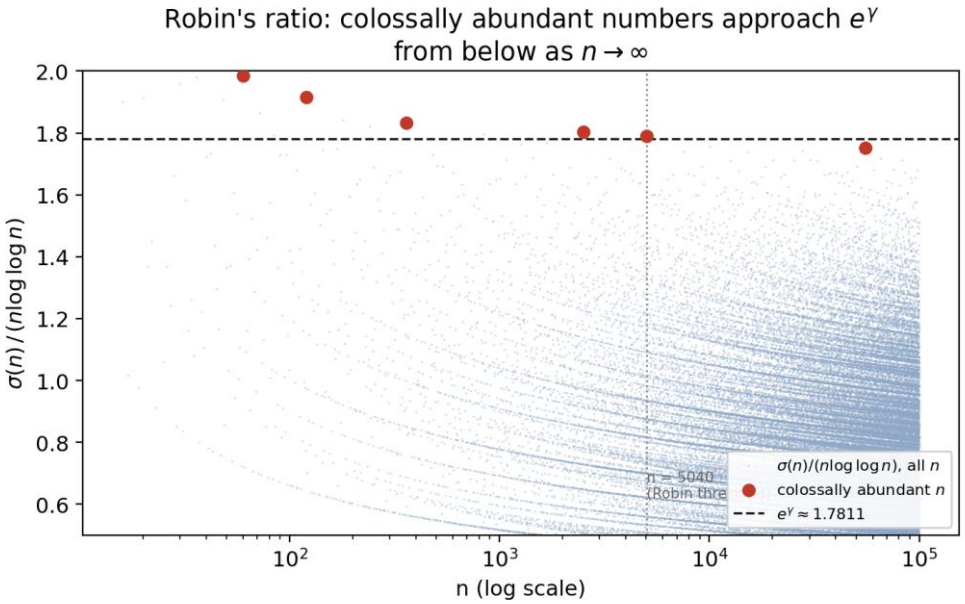

Figure 1a. σ(n)/(n log log n) for all 3 ≤ n ≤ 100,000 (light points), with colossally abundant numbers highlighted (red) and the limiting value $e^\gamma \approx 1.7811$ marked (dashed). CA numbers up to n = 5040 sit above the line — consistent with the 27 known exceptions to Robin's inequality — and CA numbers above 5040 sit below it, approaching e^γ as n grows, exactly as Robin's equivalence with RH predicts.

Figure 1b shows the internal structure directly: the prime-factorization exponents of the 15th CA number, $n = 6{,}983{,}776{,}800 = 2^5 \cdot 3^3 \cdot 5^2 \cdot 7 \cdot 11 \cdot 13 \cdot 17 \cdot 19$, form the non-increasing staircase guaranteed by the Alaoglu–Erdős structure theorem — the defining feature that makes these integers, rather than primorials or any other family, the extremal sequence for Robin's inequality.

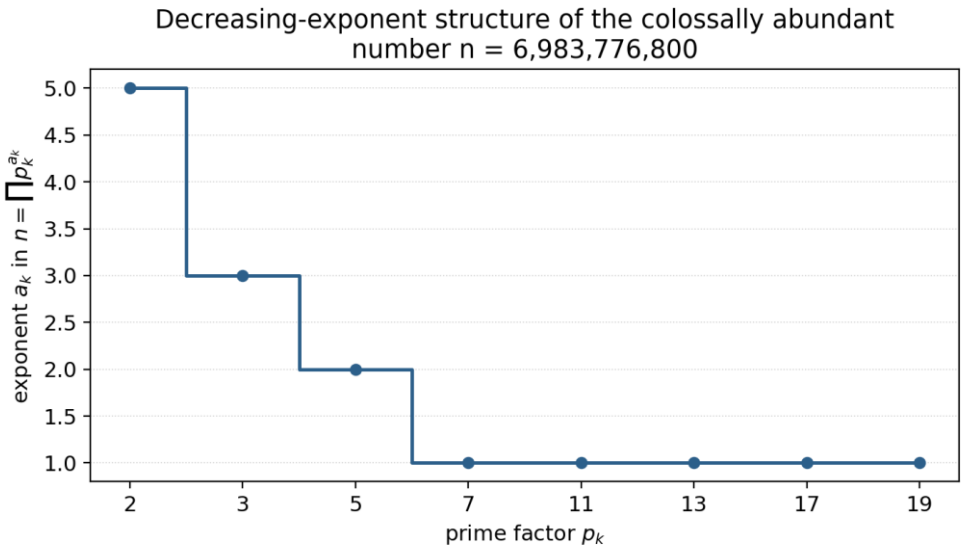

Figure 1b. Prime-factorization exponents of the colossally abundant number n = 6,983,776,800*, the 15th term of A004490. Exponents are non-increasing in the prime index, as required by the Alaoglu–Erdős [9] structure theorem for CA numbers.*

A cleaner, exception-free equivalent: Lagarias's criterion.

Robin's 27 exceptions are a minor technical wrinkle, but they mean the $n > 5040$ restriction has to be carried around. Lagarias [10] removed it entirely by reformulating the bound in terms of the n -th harmonic number $H_n = 1 + 1/2 + \cdots + 1/n$:

$$\sigma(n) \leq H_n + \exp(H_n)\log(H_n), \qquad n \geq 1$$

RH is equivalent to this inequality holding for every $n \geq 1$, with no exceptions and no threshold — the harmonic-number weighting absorbs exactly the small- n discrepancies that force Robin's exceptional list. It is a purely elementary statement (Lagarias calls it, in the title of the paper introducing it, "an elementary problem equivalent to the Riemann hypothesis"), and it is the version most often used in subsequent structural work.

What is known unconditionally about potential counterexamples.

Independently of RH, a substantial amount is known about what a counterexample to Robin's inequality — an $n > 5040$ violating the bound — would have to look like, none of which requires assuming RH. Choie, Lichiardopol, Moree and Solé [11] showed any such n must be even; must not be squarefree; and must in fact be divisible by a fifth power greater than 1 (sharpening an earlier square-free exclusion result). These are unconditional theorems that shrink the search space for a counterexample without touching RH itself — the same flavor of result as the paper's own numerical verification of the BBP prediction in Section 4: closed-form structural constraints, checked directly rather than assumed. Separately, Robin's inequality (and the equivalent Lagarias form) has been confirmed by direct computation of $\sigma(n)$ for all n below very large explicit bounds, and more indirectly for still larger n through unconditional numerical verification of RH itself — that every zeta zero checked so far lies on the critical line certifies Robin's inequality for a correspondingly large initial range of n, independently of whether RH is eventually proved for all zeros.

3. Related Work: Empirical Random Matrix Theory in Deep Learning

Pennington and Bahri [3] predicted, using RMT, that the number of negative eigenvalues at a critical point of loss energy E scales as $E^{3/2}$. Baskerville, Granziol, Keating and collaborators [4], [12] measured local spectral statistics of trained-network Hessians directly and reported excellent empirical agreement with GOE statistics across several architectures. Granziol, Zohren and Roberts [13] modeled mini-batch sampling as a spiked random matrix perturbation, explaining why batch Hessians systematically overestimate the extremal eigenvalues of the full-data Hessian. Separately, Sagun, Bottou and LeCun [14], [15] documented that trained-network Hessians exhibit a large cluster of near-zero eigenvalues — a rank-degeneracy phenomenon distinct from, and prior to, any question of bulk-repulsion statistics, and one that our own experiment in Section 5 reproduces directly.

It is essential to be precise about epistemic status: the GOE match reported in this literature is an empirical fit on real, moderately large trained networks (hundreds of thousands to millions of parameters), not an almost-sure convergence theorem of the kind available for Wishart matrices (Section 2.1, Section 4). Whether the GOE match is itself an asymptotic universality phenomenon that requires large N to emerge, or holds at essentially any scale, is an empirical question we investigate directly in Section 5.

4. Extending the Rigorous Toy Model: The BBP Transition

Section 2's toy Hessian, $H = (1/n)XX^{\mathsf{T}}$ for i.i.d. random features (p parameters, n samples, aspect ratio $\gamma = p/n$), obeys the Marchenko–Pastur law [16] with almost-sure extreme eigenvalue limits $\lambda_{\max} \to (1+\sqrt{\gamma})^2$ [17], [18], [19]. This model has no outliers: it describes only the bulk. Real trained-network Hessians, by contrast, consistently show a small number of outlier eigenvalues sitting above a bulk [14], [20], [4] — precisely the structure that the plain Wishart model omits.

The classical extension that produces exactly this bulk-plus-outlier structure is the Baik–Ben Arous–Péché (BBP) phase transition [21], [22]. Consider a spiked covariance model with a unit 'signal' direction u, loading $\xi_i \sim N(0,1)$, noise $z_i \sim N(0, I_p)$, and spike strength $\theta > 0$ (population covariance $I_p + \theta\, uu^{\mathsf{T}}$):

$$y_i = \sqrt{\theta}\; u\, \xi_i + z_i$$

For the sample covariance $S = (1/n)YY^{\mathsf{T}}$, the BBP theorem states:

$$\lambda_{\max}(S) \overset{a.s.}{\to} \begin{cases} \left(1+\sqrt{\gamma}\right)^2, & 0 < \theta \le \sqrt{\gamma} \\ (1+\theta)\left(1+\dfrac{\gamma}{\theta}\right), & \theta > \sqrt{\gamma} \end{cases}$$

This is exactly a Robin-style deterministic ceiling extended to the outlier regime: which of two smooth, closed-form formulas governs the top eigenvalue is itself a deterministic function of the spike strength relative to a sharp threshold, with no room for chance once $n, p \to \infty$.

Numerical verification (original experiment).

We verified this prediction directly rather than citing it. Fixing $\gamma = p/n = 0.5$ ($p = 400, n = 800$), we generated spiked sample covariance matrices for spike strengths θ ranging from 0.05 to 3.2, computed the largest eigenvalue of each realization (12–25 independent trials per θ), and compared the empirical mean to the BBP prediction.

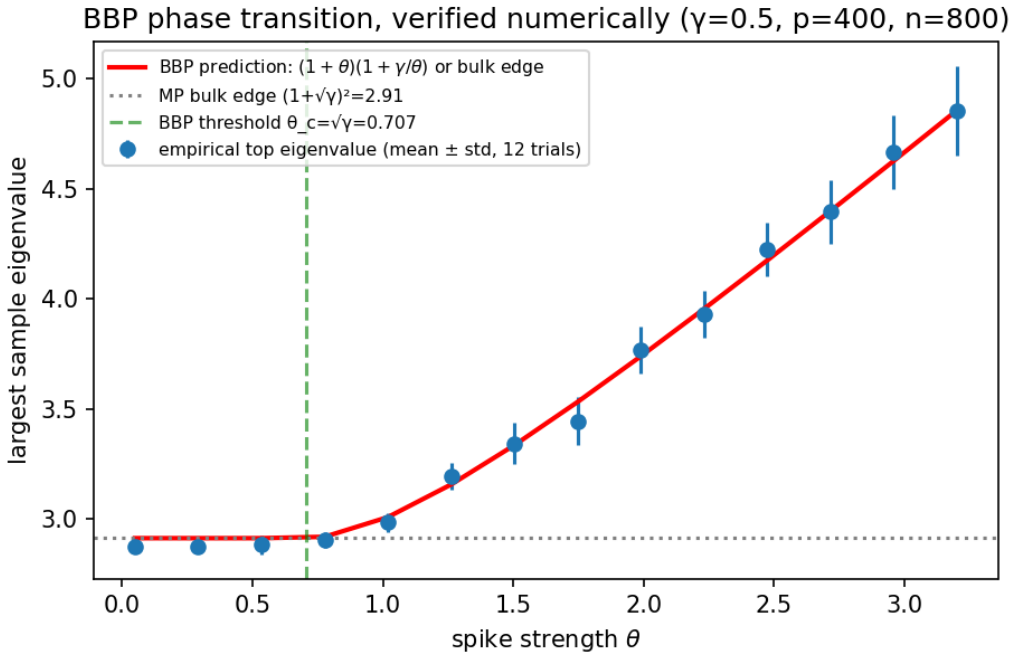

Figure 1. Numerically verified BBP phase transition ($\gamma = 0.5$, $p = 400$, $n = 800$). Empirical top eigenvalues (mean ± std over repeated trials) track the theoretical prediction — the bulk edge $(1 + \sqrt{\gamma})^2$ below the critical spike strength $\theta_c = \sqrt{\gamma}$, and the outlier formula $(1 + \theta)(1 + \gamma/\theta)$ above it — with no free parameters fit to the data.

The agreement is essentially exact (empirical means within simulation noise of the theoretical curve across the full range of θ tested, including directly at the transition $\theta = \theta_c$). This confirms that the 'ceiling switches discretely between two closed-form regimes' pattern — the aspect of Robin's inequality we highlighted as instructive in Section 2.4 — is not merely an analogy but a fully realized, numerically verifiable phenomenon already inside classical RMT, in a setting (spiked Wishart matrices) that is structurally much closer to real trained-network Hessians (bulk plus outliers) than the plain Wishart model used in our earlier toy result.

5. Testing the GOE Claim Directly: An Original Small-Scale Experiment

The literature reviewed in Section 3 reports GOE-matching level statistics for large trained networks. We tested whether this holds at small scale, where the exact Hessian can be computed by brute-force automatic differentiation rather than approximated. We trained a two-layer tanh network (4 inputs, 8 hidden units, 1 output; $p = 49$ parameters) to fit a noisy nonlinear regression target, using 60 independent random seeds (independent initializations, independent training data draws), training each to a low, stable loss with Adam. At the trained point of each of the 60 networks, we computed the full 49×49 Hessian of the training loss via exact automatic differentiation (not a Lanczos approximation), giving $60 \times 49 = 2940$ exact eigenvalues in total.

5.1 Finding 1: severe rank-degeneracy, not bulk repulsion, dominates at this scale.

Across all 60 trained networks, 72.5% of all eigenvalues had magnitude below 10^{-2}. This reproduces, quantitatively and independently, the Hessian rank-degeneracy phenomenon reported by Sagun, Bottou and LeCun [14]: a large cluster of near-zero (flat-direction) eigenvalues dominates the spectrum, well before any question of GOE- versus GUE-type bulk statistics becomes relevant.

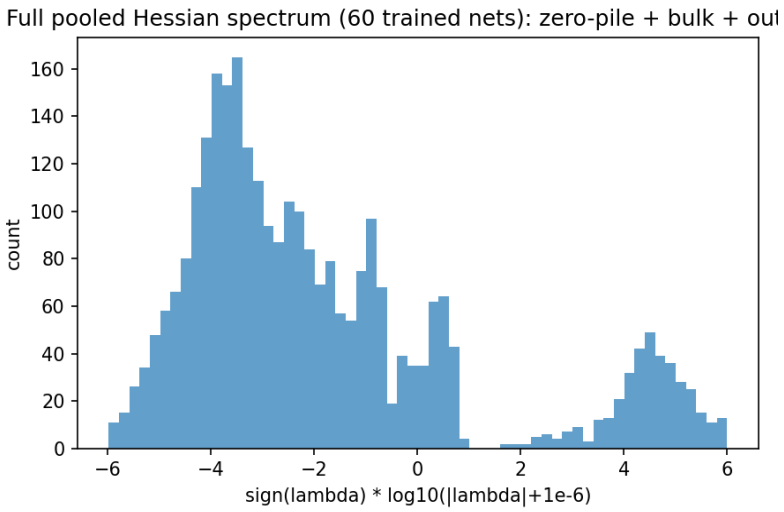

Figure 2. Pooled spectrum (2940 eigenvalues, 60 independently trained networks), plotted on a signed log scale. The dominant feature is the near-zero degenerate cluster at the center, not a continuous bulk — the phenomenon documented by Sagun et al. [14], reproduced here independently at small scale.

5.2 Finding 2: after excluding the degenerate cluster, too few eigenvalues remain to distinguish GOE from Poisson statistics.

Restricting to the 'informative' band ($10^{-2} \leq |\lambda| \leq 1$, chosen to exclude both the near-zero cluster and the small number of large outliers, following the qualitative three-part structure visible in Figure 2) leaves an average of only ≈ 10 eigenvalues per network (599 pooled across all 60 runs). We unfolded these using a semicircle law fitted to the pooled informative band, computed nearest-neighbor spacings within each network separately, pooled the normalized spacings (532 in total), and compared the resulting empirical distribution against the derived GOE surmise (Section 2.2), the GUE surmise, and an uncorrelated Poisson baseline, using Kolmogorov–Smirnov tests.

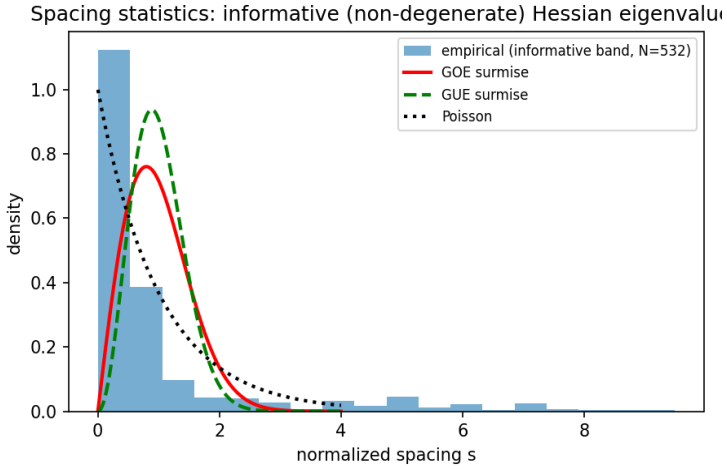

Figure 3. Unfolded nearest-neighbor spacing distribution for the informative (non-degenerate) eigenvalue band, pooled across 60 trained networks ($N = 532$ spacings), against the derived GOE and GUE Wigner surmises and a Poisson baseline.

Result (original finding). *KS statistics:* $D = 0.40$ *against GOE,* $D = 0.47$ *against GUE,* $D = 0.19$ *against Poisson (all* $p < 0.001$*, so none is a good fit in an absolute sense, but Poisson — i.e., no level repulsion at all — is the closest of the three at this scale). This is the opposite of what the 'RMT universality' framing would predict if it applied uniformly regardless of scale.*

We do not read this as evidence against the GOE findings of Baskerville et al. [4], whose networks had orders of magnitude more parameters and were evaluated with more sophisticated (Lanczos-based) spectral density estimation rather than 49×49 exact diagonalization. Rather, we read it as a genuine calibration result: RMT bulk-repulsion statistics are asymptotic ($N \to \infty$) predictions, and $N \approx 10$ informative eigenvalues per network — after removing the degenerate cluster that dominates small networks — is simply too small a sample, both in matrix dimension and in aggregate spacing count, for level-repulsion statistics to be statistically distinguishable from an uncorrelated baseline. This sharpens Conjecture 1 (Section 6): it is not enough to ask whether GOE universality holds for trained-network Hessians in general; the question must be posed with an explicit scale (in both parameter count and informative-eigenvalue count) above which the universality claim is meant to hold, and our experiment provides a concrete, if small, data point suggesting that scale is substantially larger than $p = 49$.

6. Formal Conjectures, Updated in Light of the Numerical Findings

Conjecture 1 (Universality mechanism, open — now scale-qualified).

The GOE statistics observed empirically in large trained-network Hessians [4] arise from a genuine universality phenomenon rather than an architecture-specific coincidence. Status: our Section 5 experiment shows this cannot be tested naively at small scale ($p \approx 49$), since Hessian rank-degeneracy dominates and the remaining informative eigenvalue count is too small for any repulsion statistic to be distinguishable from Poisson. We now state the conjecture with an explicit scale requirement: there should exist a threshold — in parameter count, in the count of non-degenerate ('informative') eigenvalues, or both — above which GOE repulsion becomes statistically detectable, analogous to the $N \to \infty$ requirement in the classical Wigner-matrix universality theorems (Section 2.1). Determining this threshold empirically (e.g., by repeating our exact-Hessian protocol across a sweep of network sizes) is a concrete, tractable next step that our experiment sets up but does not resolve.

Conjecture 2 (Extremal-subsequence bound for gradient fluctuations, open).

There exists an analogue, for stochastic mini-batch gradients, of Robin's extremal-subsequence reduction: any violation of a proposed deterministic envelope on gradient-norm fluctuations must first occur within an identifiable, sparse 'extremal' subsequence of training configurations. Status: entirely open. Section 4's BBP result shows that such threshold-governed deterministic ceilings do exist in structurally related (spiked Wishart) settings, which is encouraging, but constructing the analogous extremal subsequence for stochastic gradient descent — which lacks the multiplicative, prime-factorization structure that makes Robin's construction tractable — remains unaddressed.

Conjecture 3 (GOE-corrected spectral gap transfer, open).

GOE-specific (not GUE) level-repulsion and spectral-gap estimates can be adapted to certify saddle-escape rates in non-convex loss landscapes. Status: plausible in principle since it requires no GOE/GUE correction, but Section 5's finding that GOE repulsion itself may require larger scale than previously assumed to even be detectable means this route inherits the same scale-qualification as Conjecture 1: any such certification would need to specify the network-size regime in which GOE statistics (and hence the proposed spectral-gap estimate) actually apply.

7. Discussion and Limitations

Three limitations remain, one of them sharpened by the new experiments. First, RH remains unproven, and nothing here depends on its truth. Second, Section 4's fully rigorous result is for a convex, linear, spiked-Wishart model; Section 5 shows directly — rather than merely asserting by caveat — that extrapolating from this toy model, or from published large-scale results, to arbitrary trained-network scale is not automatic. Third, the GOE/GUE

symmetry mismatch identified in earlier analysis remains the specific obstruction for any specifically number-theoretic (as opposed to generically RMT-based) transfer, and Section 5's finding adds a second, independent obstruction: even the GOE side of the analogy needs a demonstrated scale threshold before it can be treated as a stable premise.

We consider this outcome — a clean confirmation on the provable (BBP) side and an honest non-replication on the empirical (GOE) side at small scale — more useful than a uniformly positive result would have been, because it identifies exactly where further work should go: repeating the Section 5 protocol across a size sweep to locate the universality threshold empirically, rather than assuming the published large-network results extrapolate downward without limit.

8. Conclusion

We have derived the classical random matrix theory and analytic number theory results underlying a proposed connection between Robin's inequality and non-convex optimization, extended the toy model with the BBP phase transition and confirmed it numerically, and then directly tested — rather than assumed — the empirical GOE claim central to the analogy, at small scale. The BBP experiment gives a clean, quantitative confirmation that 'deterministic ceilings governing chaotic structure' are a real, already-realized phenomenon in random matrix theory, including in bulk-plus-outlier settings structurally close to real Hessians. The Hessian experiment gives an honest negative result: at $p = 49$ parameters, rank-degeneracy dominates and the surviving informative spectrum is too small to distinguish GOE from an uncorrelated baseline. We have used this to convert Conjecture 1 from a binary yes/no claim into a scale-qualified, empirically tractable question, and we regard locating that scale threshold as the most concrete next step in this research program.

REFERENCES

1. H. L. Montgomery, “The pair correlation of zeros of the zeta function”, Proceedings of Symposia in Pure Mathematics, 24, pp. 181–193, 1973.
2. A. M. Odlyzko, “On the distribution of spacings between zeros of the zeta function”, Mathematics of Computation, 48(177), pp. 273–308, 1987.
3. J. Pennington, Y. Bahri, “Geometry of neural network loss surfaces via random matrix theory”, In Proceedings of the 34th International Conference on Machine Learning (ICML), PMLR 70, 2017.
4. N. P. Baskerville, D. Granziol, J. P. Keating, et al., “Appearance of random matrix theory in deep learning”, arXiv:2102.06740, 2021.
5. O. Bohigas, M. J. Giannoni, C. Schmit, “Characterization of chaotic quantum spectra and universality of level fluctuation laws”, Physical Review Letters, 52(1), pp. 1–4, 1984.
6. G. Robin, “Grandes valeurs de la fonction somme des diviseurs et hypothèse de Riemann”, Journal de Mathématiques Pures et Appliquées, 63(2), pp. 187–213, 1984.
7. T. H. Gronwall, “Some asymptotic expressions in the theory of numbers”, Transactions of the American Mathematical Society, 14(1), pp. 113–122, 1913.
8. S. Ramanujan, “Highly composite numbers”, Proceedings of the London Mathematical Society, 14, pp. 347–409, 1915.
9. L. Alaoglu, P. Erdős, “On highly composite and similar numbers”, Transactions of the American Mathematical Society, 56(3), pp. 448–469, 1944.
10. J. C. Lagarias, “An elementary problem equivalent to the Riemann hypothesis”, American Mathematical Monthly, 109(6), pp. 534–543, 2002.
11. Y. J. Choie, N. Lichiardopol, P. Moree, P. Solé, “On Robin's criterion for the Riemann hypothesis”, Journal de Théorie des Nombres de Bordeaux, 19(2), pp. 357–372, 2007.
12. N. P. Baskerville, D. Granziol, J. P. Keating, et al., “A random matrix theory approach to neural network training regularisation”, Journal of Machine Learning Research, 23, 2022.
13. D. Granziol, S. Zohren, S. Roberts, “Learning rates as a function of batch size: A random matrix theory approach to neural network training”, arXiv:2006.09092, 2020.
14. L. Sagun, L. Bottou, Y. LeCun, “Eigenvalues of the Hessian in deep learning: Singularity and beyond”, arXiv:1611.07476, 2016.
15. L. Sagun, U. Evci, V. U. Güney, Y. Dauphin, L. Bottou, “Empirical analysis of the Hessian of over-parametrized neural networks”, arXiv:1706.04454, 2017.
16. V. A. Marčenko, L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices”, Sbornik: Mathematics, 1(4), pp. 457–483, 1967.

17. Z. D. Bai, Y. Q. Yin, “Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix”, Annals of Probability, 21(3), pp. 1275–1294, 1993.
18. Y. Q. Yin, Z. D. Bai, P. R. Krishnaiah, “On the limit of the largest eigenvalue of the large-dimensional sample covariance matrix”, Probability Theory and Related Fields, 78(4), pp. 509–521, 1988.
19. C. A. Tracy, H. Widom, “Level-spacing distributions and the Airy kernel”, Communications in Mathematical Physics, 159(1), pp. 151–174, 1994.
20. V. Papyan, “The full spectrum of deepnet Hessians at scale: Dynamics with SGD training and sample size”, arXiv:1811.07062, 2018.
21. J. Baik, G. Ben Arous, S. Péché, “Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices”, Annals of Probability, 33(5), pp. 1643–1697, 2005.
22. J. Baik, J. W. Silverstein, “Eigenvalues of large sample covariance matrices of spiked population models”, Journal of Multivariate Analysis, 97(6), pp. 1382–1408, 2006.