hexagon:2611.00662v1

SPECTRAL OUTLIERS OF SELF-COUPLED COVARIANCE MATRICES WITH EXACT ZERO WEIGHTS

SAMUIL PETKOV

ABSTRACT. We study the right spectral outliers of $M_d = n^{-1}A_d D A_d^{\mathsf{T}}$, where the columns of A_d are drawn from a Gaussian mixture with a fixed-dimensional signal subspace and each diagonal weight $D_{\ell\ell} = h_{b_\ell}(g_\ell)$ is a function of the class label and of the low-dimensional Gaussian coordinates of the same sample. The effective outlier theory of Ben Arous, Gheissari, Huang and Jagannath for such self-coupled matrices assumes that the weights do not accumulate at zero, which excludes hard gates such as the derivative of a rectified linear unit. Assuming $h_b(g) \in \{0\} \cup [\tau, M]$, we prove that exact zero weights do not change the positive right-outlier equation. The proof deletes inactive columns before any linearization, keeps their random normalization, and applies the invertible-weight theory to the exact conditional mark law of the active samples; we show, by conditioning on the marks, that this theory extends to non-Gaussian fixed-dimensional marks. The limiting spectrum acquires the atom $(1-\phi\pi)_+$ at zero, while positive outliers are the roots of a $q \times q$ determinant. We obtain multiplicity-preserving eigenvalue counts, absence of spurious right outliers, a projected-eigenvector normalization, an invariant-subspace formula at repeated roots, and the exact finite-dimensional kernel.

Mathematics Subject Classification (2020). Primary 60B20; Secondary 15B52, 62H25.

Keywords. Random matrices; weighted sample covariance matrices; spectral outliers; Gaussian mixtures; exact zero weights; rectified linear unit.

1. INTRODUCTION

Random covariance spectra separate macroscopic bulk behaviour from finitely many signal-bearing outliers. In classical spiked models this separation is controlled by deterministic transforms of the bulk law and produces the Baik–Ben Arous–Péché transition [8, 2, 4]. In *self-coupled* models, the diagonal weight of a sample depends on the same low-dimensional projection that enters the sample itself. Such matrices describe spectral estimators, information matrices, and blocks of neural-network curvature [3, 7].

An effective spectral theory for Gaussian-mixture self-coupling was developed by Ben Arous, Gheissari, Huang and Jagannath [3]. Its outlier proof linearizes the weighted covariance matrix using the inverse of the diagonal weights, and its assumptions forbid weights accumulating at zero. Consequently a hard gate, for example the derivative of a rectified linear unit (ReLU) [10], which vanishes on a set of positive probability, is excluded. We show that this singularity is algebraic rather than spectral: inactive columns can be removed exactly, without changing the effective equation for positive outliers.

Exact zero weights are already covered in a related centred setting. For isotropic standard Gaussian covariates, Kovačević, Zhang and Mondelli [7] treat spectral estimators $n^{-1}\sum_i T(y_i)a_i a_i^{\mathsf{T}}$ for multi-index models with a bounded preprocessing function T subject only to $\mathbb{P}\{T(y)=0\} < 1$, so that the weights may vanish with positive probability; they locate the top eigenvalues and determine the eigenvector overlaps with the signal subspace. The present note treats the noncentred Gaussian-mixture model of [3], and gives multiplicity-preserving counts on arbitrary compact intervals to the right of the bulk. Recent local-law and block-Wishart results [5, 9] provide broader context.

The argument has two parts. First, an exact finite-dimensional identity (Section 3) writes M_d as a random scalar multiple of an invertible-weight matrix built from the active columns only; these columns are i.i.d. from a truncated, generally non-Gaussian, mark law. Second, we show (Proposition 4.1) that the invertible-weight theory of [3] holds for arbitrary fixed-dimensional marks that are independent of the orthogonal Gaussian block: conditionally on the marks, every estimate used in the source proof concerns either the Gaussian block with a fixed weight matrix, or laws of large numbers for mark functionals.

Date: 20 July 2026.

Scope. All statements are asymptotic in the proportional regime $n/d \to \phi$ with fixed summary dimension q and fixed number of classes k. They require nonnegative weights and a fixed gap $\tau > 0$ between zero and the nonzero weights. Signed weights, positive weights accumulating continuously at zero, growing summary dimension, and uniformity along a data-dependent training trajectory are not covered. The numerical computations in Section 12 are illustrations and are not used in any proof. Whether the conclusions persist when the gap $\tau = \tau_d$ tends to zero with d, and at what quantitative rate the empirical outliers approach the effective roots, are open questions.

2. Model and main theorem

Fix $k, C, q \in \mathbb{N}$ with $q = C + k$, $\lambda > 0$, $\phi > 0$, probabilities $p_b > 0$ summing to one, and a positive-semidefinite matrix $G \in \mathbb{R}^{q\times q}$. For each large d, let

$$B_d = [x_1^{(d)}, \dots, x_C^{(d)}, \mu_1^{(d)}, \dots, \mu_k^{(d)}] \tag{2.1}$$

satisfy $B_d^{\mathsf{T}} B_d = G$. Choose $L_d \in \mathbb{R}^{d\times q}$ with orthonormal columns and

$$L_d^{\mathsf{T}} B_d = \sqrt{G}. \tag{2.2}$$

No inverse of G or $\sqrt{G}$ is used. If G is rank deficient, unused columns of L_d may be chosen arbitrarily outside $\operatorname{span} B_d$. Let $L_d^{\perp} \in \mathbb{R}^{d\times(d-q)}$ complete L_d to an orthogonal matrix.

For class b, let $m_b = L_d^{\mathsf{T}} \mu_b^{(d)}$. Independently over ℓ sample

$$b_\ell \sim \mathrm{Cat}(p_1, \dots, p_k), \quad g_\ell \sim N(0, I_q), \quad w_\ell \sim N(0, I_{d-q}), \tag{2.3}$$

and define

$$Y_\ell = L_d\Big(m_{b_\ell} + \lambda^{-1/2} g_\ell\Big) + \lambda^{-1/2} L_d^{\perp} w_\ell. \tag{2.4}$$

Let $A_d = [Y_1 \cdots Y_n]$, with $n/d \to \phi$, and let

$$D_{\ell\ell} = h_{b_\ell}(g_\ell), \qquad M_d = \frac{1}{n} A_d D A_d^{\mathsf{T}}, \tag{2.5}$$

where $h_b : \mathbb{R}^q \to [0, M]$ is measurable and

$$h_b(g) \in \{0\} \cup [\tau, M] \quad \text{a.s.} \tag{2.6}$$

for some $\tau > 0$. Put

$$\pi_b = \mathbb{P}\{h_b(g) > 0\}, \qquad \pi = \sum_{b=1}^{k} p_b \pi_b. \tag{2.7}$$

Let S_G be the physical Stieltjes branch

$$1 + z S_G(z) = \phi \sum_{b=1}^{k} p_b \mathbb{E}\left[\frac{S_G(z) h_b(g)}{\lambda\phi + S_G(z) h_b(g)}\right], \tag{2.8}$$

with $\operatorname{Im} S_G(z) > 0$ for $\operatorname{Im} z > 0$ and $z S_G(z) \to -1$. Let ν_G be its probability measure and

$$\lambda_+(G) = \sup \operatorname{supp} \nu_G. \tag{2.9}$$

For

$$v_b(g) = m_b + \lambda^{-1/2} g, \tag{2.10}$$

define, outside the support,

$$F_G(z) = \lambda\phi \sum_{b=1}^{k} p_b \mathbb{E}\left[\frac{h_b(g)}{\lambda\phi + S_G(z) h_b(g)} v_b(g) v_b(g)^{\mathsf{T}}\right], \tag{2.11}$$

and $B_G(z) = z I_q - F_G(z)$.

Theorem 2.1 (Zero-atom outlier theorem)**.** *If $\pi = 0$, then $M_d = 0$ almost surely and all assertions reduce to $\nu_G = \delta_0$, $S_G(z) = -1/z$, and $F_G = 0$. Assume $\pi > 0$.*

(1) *The empirical spectral distribution of M_d converges weakly in probability to ν_G, and*

$$\nu_G(\{0\}) = (1-\phi\pi)_+. \tag{2.12}$$

(2) *For every compact $K \subset (\lambda_+(G), \infty)$,*

$$\operatorname*{ess\,inf}_{\substack{z\in K,\, b\in[k]\\ g:\, h_b(g)>0}} |\lambda\phi + S_G(z)h_b(g)| > 0. \tag{2.13}$$

Hence F_G and B_G are real analytic to the right of the bulk.

(3) *Let $I = [a,b] \subset (\lambda_+(G), \infty)$ and assume $\det B_G(a) \det B_G(b) \neq 0$. Then, with probability tending to one,*

$$\#\{\sigma(M_d) \cap I\} = \#\{z \in I : \det B_G(z) = 0\}, \tag{2.14}$$

with algebraic and analytic multiplicity.

(4) *For each root $z_* > \lambda_+(G)$,*

$$\operatorname{ord}_{z_*} \det B_G(z) = \dim \operatorname{Ker} B_G(z_*). \tag{2.15}$$

The sum of right-root multiplicities is at most q.

(5) *Every isolated root of multiplicity m attracts exactly m empirical eigenvalues. Conversely, for every $\epsilon > 0$, all empirical eigenvalues above $\lambda_+(G) + \epsilon$ lie within $o_{\mathbb{P}}(1)$ of a bounded effective root.*

(6) *If z_* is simple, $z_d \to z_*$ is its empirical eigenvalue, u_d is a unit eigenvector, and $r_d = L_d^{\mathsf{T}} u_d$, then*

$$\|B_G(z_d) r_d\| = o_{\mathbb{P}}(1), \tag{2.16}$$

$$\left| r_d^{\mathsf{T}} B_G'(z_d) r_d - 1 \right| = o_{\mathbb{P}}(1). \tag{2.17}$$

If e_* is a unit vector spanning $\operatorname{Ker} B_G(z_*)$, then after sign choice,

$$r_d \xrightarrow{\mathbb{P}} \frac{e_*}{\sqrt{e_*^{\mathsf{T}} B_G'(z_*) e_*}}. \tag{2.18}$$

(7) *If z_* has multiplicity $m > 1$, choose an isolating interval and let Π_d be the corresponding empirical spectral projection. Let*

$$K_* = \operatorname{Ker} B_G(z_*), \quad P_* = \operatorname{Proj}_{K_*}, \quad J_* = P_* B_G'(z_*) P_*|_{K_*}. \tag{2.19}$$

Then $J_* \succ 0$, $\operatorname{rank} \Pi_d = m$ with probability tending to one, and

$$L_d^{\mathsf{T}} \Pi_d L_d \xrightarrow{\mathbb{P}} P_* J_*^{-1} P_*. \tag{2.20}$$

(8) *For every finite d, almost surely,*

$$\operatorname{rank} M_d = \min(d, N_{\text{active}}), \qquad N_{\text{active}} = \sum_{\ell=1}^{n} \mathbf{1}_{\{D_{\ell\ell}>0\}}. \tag{2.21}$$

Consequently,

$$\frac{1}{d} \dim \operatorname{Ker} M_d \longrightarrow (1-\phi\pi)_+ \quad \text{almost surely.} \tag{2.22}$$

The remainder of the note proves Theorem 2.1. Sections 3–4 contain the reduction and the extension of the invertible-weight theory; Sections 5–10 derive the individual assertions.

3. Exact active-column reduction

Define

$$T_\ell = h_{b_\ell}(g_\ell), \qquad I_\ell = \mathbf{1}_{\{T_\ell>0\}}, \qquad N_d = \sum_{\ell=1}^{n} I_\ell. \tag{3.1}$$

After a column permutation,

$$A_d = [A_{+,d}\ A_{0,d}], \qquad D = \begin{bmatrix} D_{+,d} & 0 \\ 0 & 0 \end{bmatrix}, \tag{3.2}$$

where $D_{+,d}$ is $N_d \times N_d$ and

$$\tau I \preceq D_{+,d} \preceq MI. \tag{3.3}$$

Therefore

$$M_d = \frac{1}{n} A_{+,d} D_{+,d} A_{+,d}^{\mathsf{T}} = \frac{N_d}{n} \widehat{M}_d, \qquad \widehat{M}_d = \frac{1}{N_d} A_{+,d} D_{+,d} A_{+,d}^{\mathsf{T}}. \tag{3.4}$$

This is an exact finite-dimensional identity. The inactive columns are absent before any determinant identity, Schur complement, or resolvent is introduced.

Lemma 3.1 (Conditional active law)**.** *Conditioned on activity, the low-dimensional mark has law*

$$\mathbb{P}_+(b, dg) = \frac{p_b \mathbf{1}_{\{h_b(g)>0\}} \gamma_q(dg)}{\pi}, \tag{3.5}$$

and w remains an independent standard Gaussian. Conditioned on $N_d = m$, the unordered active columns are distributed as m independent draws from this active law.

Proof. The activity event is measurable with respect to (b, g) and independent of w. Thus

$$\mathbb{P}\{b \in db, g \in dg, w \in dw \mid I = 1\} = \frac{p_b \mathbf{1}_{\{h_b(g)>0\}} \gamma_q(dg)}{\pi} \gamma_{d-q}(dw). \tag{3.6}$$

Independence across sample indices is preserved under coordinatewise conditioning. Given the number of successes, exchangeability makes their positions uniform, and the unordered successful marks are independent with law (3.5). □

Remark 3.2. The active g is generally truncated or reweighted. It must not be replaced by an unconditioned Gaussian vector: replacing (3.5) by the unconditioned law would alter the low-rank term. Thus the active columns are not, in general, an ordinary Gaussian mixture.

Because I_ℓ are independent Bernoulli variables of mean π,

$$\frac{N_d}{n} \to \pi, \qquad \frac{N_d}{d} \to \phi\pi \tag{3.7}$$

almost surely, for example by Hoeffding's inequality and Borel–Cantelli.

4. Invertible covariance with fixed-dimensional marks

The invertible-weight theory of [3] treats a weighted covariance matrix whose weights do not accumulate at zero and whose low-dimensional coordinates are Gaussian. After the reduction of Section 3 the weights are bounded below, but the low-dimensional mark of an active column has the truncated law (3.5). This section shows that Gaussianity of the mark is not used.

4.1. **Statement.** Fix q, $\lambda > 0$ and $0 < \tau \le M$. Let μ be a probability law on $[\tau, M] \times \mathbb{R}^q$ with

$$\int \|v\|^r \, \mu(dt, dv) < \infty \qquad \text{for every } r < \infty, \tag{4.1}$$

and let μ_T denote its first marginal. For each d let $m_d \in \mathbb{N}$ with $m_d/d \to c \in (0, \infty)$, let $L = L_d \in \mathbb{R}^{d \times q}$ have orthonormal columns, and let $L^\perp$ complete it to an orthogonal matrix. Let $(T_i, V_i)_{i \le m_d}$ be i.i.d. with law μ, and let $W_i \sim N(0, I_{d-q})$ be i.i.d. and independent of the marks. Define

$$Y_i = LV_i + \lambda^{-1/2} L^\perp W_i, \qquad \widehat{M}_d = \frac{1}{m_d} \sum_{i=1}^{m_d} T_i Y_i Y_i^{\mathsf{T}}. \tag{4.2}$$

Let s_c be the Stieltjes transform of the probability measure ν_c determined by

$$1 + x s_c(x) = c \, \mathbb{E}\left[\frac{s_c(x) T}{\lambda c + s_c(x) T} \right], \tag{4.3}$$

with $\operatorname{Im} s_c > 0$ on $\mathbb{C}^+$, let $\lambda_{+,c} = \sup \operatorname{supp} \nu_c$, and set

$$F_c(x) = \lambda c \, \mathbb{E}\left[\frac{T}{\lambda c + s_c(x) T} V V^{\mathsf{T}} \right], \qquad B_c(x) = x I_q - F_c(x). \tag{4.4}$$

Expectations are over $(T, V) \sim \mu$.

Almost-sure statements refer to one probability space carrying the variables for all d, on which the family of all marks (over all d) is independent of the family of all Gaussian vectors (over all d). The marks may be coupled across d arbitrarily; the Gaussian vectors are coupled across d in any way under which the almost-sure statements of [3] for a Gaussian block independent of a deterministic weight matrix hold (the source does not fix a coupling). Every later use of the proposition (Sections 8 and 10) is through probabilities at fixed d, which depend only on the law at that d and not on these couplings.

Proposition 4.1 (Invertible covariance with fixed-dimensional marks)**.** *In this setting the following hold.*

(i) *The empirical spectral distribution of $\widehat{M}_d$ converges weakly to ν_c almost surely.*

(ii) *For every compact $K \subset (\lambda_{+,c}, \infty)$, $\inf\{|\lambda c + s_c(x)t| : x \in K,\ t \in \operatorname{supp} \mu_T\} > 0$. Hence F_c and B_c are real analytic on $(\lambda_{+,c}, \infty)$.*

(iii) *Let $[a, b] \subset (\lambda_{+,c}, \infty)$ with $\det B_c(a) \det B_c(b) \neq 0$. Almost surely, for all large d, the number of eigenvalues of $\widehat{M}_d$ in $[a, b]$, counted with multiplicity, equals the number of zeros of $\det B_c$ in $[a, b]$, counted with order.*

(iv) *For every $[a, b] \subset (\lambda_{+,c}, \infty)$ there is $\eta > 0$ such that, with $\Omega = \{x \in \mathbb{C} : \operatorname{Re} x \in [a, b],\ |\operatorname{Im} x| \leq \eta\}$, there are random $q \times q$ matrix functions B_d, holomorphic on a neighbourhood of Ω for all large d, with*

$$L^{\mathsf{T}}(\widehat{M}_d - x)^{-1} L = -B_d(x)^{-1} \quad (x \in \Omega \setminus \sigma(\widehat{M}_d)), \tag{4.5}$$

and, almost surely, $\sup_\Omega \|B_d - B_c\| \to 0$ and $\sup_\Omega \|B_d' - B_c'\| \to 0$.

(v) *If $x_d \in \sigma(\widehat{M}_d) \cap [a, b]$, $x_d \to x_*$, u_d is a unit eigenvector and $r_d = L^{\mathsf{T}} u_d$, then almost surely $\|B_c(x_d) r_d\| \to 0$ and $r_d^{\mathsf{T}} B_c'(x_d) r_d \to 1$.*

4.2. **Two laws of large numbers.** The source proof invokes the strong law of large numbers for triangular arrays. We record the version used here; it requires no coupling between different values of d.

Lemma 4.2. *Let ξ be a real random variable with $\mathbb{E}|\xi|^6 < \infty$. For each d let $\xi_{d,1}, \dots, \xi_{d,m_d}$ be i.i.d. copies of ξ, with $m_d/d \to c > 0$. Then $m_d^{-1} \sum_{i \leq m_d} \xi_{d,i} \to \mathbb{E}\xi$ almost surely.*

Proof. Put $\zeta = \xi - \mathbb{E}\xi$ and $\Sigma_m = \sum_{i \leq m} \zeta_{d,i}$. In the expansion of $\mathbb{E}\Sigma_m^6$ over index tuples $(i_1, \dots, i_6)$, a tuple containing an index exactly once contributes zero. The remaining tuples take at most three distinct values, so there are at most $C_0 m^3$ of them for an absolute constant C_0, and each contributes at most $\mathbb{E}|\zeta|^6$ by Hölder's inequality. Markov's inequality gives $\mathbb{P}\{|\Sigma_{m_d}| > \epsilon m_d\} \leq C_0 \mathbb{E}|\zeta|^6 \epsilon^{-6} m_d^{-3}$, which is summable in d because $m_d \geq cd/2$ eventually. The Borel–Cantelli lemma completes the proof. □

Lemma 4.3. *Let $T_{d,1}, \dots, T_{d,m_d}$ be i.i.d. with law μ_T supported in $[\tau, M]$, and let $\hat{\mu}_d = m_d^{-1} \sum_i \delta_{T_{d,i}}$. Almost surely, $\hat{\mu}_d \to \mu_T$ weakly and the Hausdorff distance between $\operatorname{supp} \hat{\mu}_d$ and $\operatorname{supp} \mu_T$ tends to zero.*

Proof. Lemma 4.2 applied to the countably many bounded variables T^j, $j \in \mathbb{N}$, gives almost sure convergence of all moments; on $[\tau, M]$ this implies weak convergence by the Weierstrass approximation theorem. Almost surely every $T_{d,i}$ lies in $\operatorname{supp} \mu_T$, so $\operatorname{supp} \hat{\mu}_d \subset \operatorname{supp} \mu_T$. Conversely, fix $\epsilon > 0$ and cover the compact set $\operatorname{supp} \mu_T$ by finitely many open intervals $J_1, \dots, J_r$ of length ϵ centred at support points. Each has $p_j = \mu_T(J_j) > 0$, and the probability that some J_j contains no sample is at most $r(1 - \min_j p_j)^{m_d}$, which is summable in d. By Borel–Cantelli, eventually every point of $\operatorname{supp} \mu_T$ is within ϵ of a sample. Take $\epsilon = 1/k$, $k \in \mathbb{N}$. □

4.3. **The finite-d transform for a convergent aspect ratio.** The source's finite-d transform $S^{(d)}$ solves its equation (2.11), which carries the ratio $\phi_d = n/(d - q)$; this ratio is not constant even when $n/d = \phi$. Here $\phi_d = m_d/(d - q) \to c$. The following lemma gives the convergence on $\mathbb{C}^+$ in this generality.

Lemma 4.4. *Let ρ_d, ρ be probability measures on $[\tau, M]$ with $\rho_d \to \rho$ weakly, and let $\phi_d, c > 0$ with $\phi_d \to c$. For $z \in \mathbb{C}^+$ and a probability measure ϱ on $[\tau, M]$, $\varphi > 0$, consider*

$$\frac{1}{S} = -z + \varphi \int \frac{t}{1 + St}\, \varrho(dt), \qquad \operatorname{Im} S > 0. \tag{4.6}$$

(a) (4.6) *has at most one solution, and every solution satisfies* $|S| \le 1/\operatorname{Im} z$ *and* $\varphi \int t^2|1+St|^{-2}\varrho(dt) < |S|^{-2}$.
(b) *If* $S_d(z)$ *solves* (4.6) *with* $(\varrho,\varphi) = (\rho_d,\phi_d)$ *and* $S(z)$ *solves it with* (ρ, c), *then* $S_d(z) \to S(z)$ *for every* $z \in \mathbb{C}^+$. *If moreover the* S_d *are holomorphic on* $\mathbb{C}^+$, *the convergence is uniform on compact subsets of* $\mathbb{C}^+$.

Proof. For $\operatorname{Im} S > 0$ and $t > 0$, $\operatorname{Im}(1+St) = t \operatorname{Im} S > 0$, so the integrand is defined. Since $\operatorname{Im}(1/S) = -\operatorname{Im} S/|S|^2$ and $\operatorname{Im}\big(t/(1+St)\big) = -t^2 \operatorname{Im} S/|1+St|^2$, the imaginary part of (4.6) reads

$$\operatorname{Im} S\Big(\frac{1}{|S|^2} - \varphi \int \frac{t^2}{|1+St|^2}\varrho(dt)\Big) = \operatorname{Im} z > 0. \tag{4.7}$$

This gives the integral bound in (a), and also $\operatorname{Im} S \ge |S|^2 \operatorname{Im} z$, hence $|S| \le 1/\operatorname{Im} z$. If S_1, S_2 are solutions, subtracting the two equations gives

$$(S_1 - S_2)\Big(\frac{1}{S_1 S_2} - \varphi \int \frac{t^2}{(1+S_1 t)(1+S_2 t)}\varrho(dt)\Big) = 0.$$

By the Cauchy–Schwarz inequality and the integral bound, the modulus of the integral term is strictly less than $1/(|S_1||S_2|)$, so the bracket is nonzero and $S_1 = S_2$.

For (b), fix z and put $\Phi = \sup_d \phi_d < \infty$. If $|S_d| \le 1/(2M)$, then $|1+S_d t| \ge 1/2$ on $[\tau, M]$ and (4.6) gives $1/|S_d| \le |z| + 2\Phi M$. Hence $|S_d| \ge s_0 = \min\{1/(2M), 1/(|z|+2\Phi M)\}$, and by (4.7), $\operatorname{Im} S_d \ge s_0^2 \operatorname{Im} z$. Consequently $|1+S_d t| \ge t \operatorname{Im} S_d \ge \kappa = \tau s_0^2 \operatorname{Im} z$ for $t \in [\tau, M]$. The sequence S_d is bounded by (a); let S_∞ be a limit point along a subsequence. Then $\operatorname{Im} S_\infty \ge s_0^2 \operatorname{Im} z > 0$, so $|1+S_\infty t| \ge \kappa$ on $[\tau, M]$, and

$$\sup_{t\in[\tau,M]} \Big| \frac{t}{1+S_d t} - \frac{t}{1+S_\infty t} \Big| \le \frac{M^2 |S_d - S_\infty|}{\kappa^2} \to 0$$

along the subsequence. Since $t \mapsto t/(1+S_\infty t)$ is bounded and continuous on $[\tau, M]$ and $\rho_d \to \rho$ weakly, $\int t(1+S_d t)^{-1}\rho_d(dt) \to \int t(1+S_\infty t)^{-1}\rho(dt)$. With $\phi_d \to c$, S_∞ solves (4.6) for (ρ, c) and has positive imaginary part, so $S_\infty = S(z)$ by (a). Every subsequence has a further subsequence converging to $S(z)$, so $S_d(z) \to S(z)$. Under holomorphy, the bound $|S_d| \le 1/\operatorname{Im} z$ makes (S_d) locally bounded on $\mathbb{C}^+$, and Vitali's theorem upgrades pointwise to locally uniform convergence. □

The next lemma controls the right edge of the finite-d measures; it is used only in the paragraph *Avoiding the modification for (ii)–(v)* below.

Lemma 4.5. *Let* ρ_d *be finitely supported probability measures on* $[\tau/2, s]$, *where* $s = \max\operatorname{supp}\mu_T$, *with* $\rho_d \to \mu_T$ *weakly, and let* $\phi_d \to c > 0$. *Let* ν_d *and* ν *be the probability measures whose Stieltjes transforms solve* (4.6) *with* (ρ_d, ϕ_d) *and with* (μ_T, c), *respectively. Then* $\limsup_d \sup\operatorname{supp}\nu_d \le \sup\operatorname{supp}\nu$.

Proof. For a measure ϱ and $\varphi > 0$ write $f_{\varrho,\varphi}(y) = -1/y + \varphi \int t(1+yt)^{-1}\varrho(dt)$. Let $x > \sup\operatorname{supp}\nu$ and $y = S(x) = \int \nu(du)/(u-x) < 0$. By the Silverstein–Choi characterization used in the paragraph *Denominators* below, $1 + S(x')t \ne 0$ for $x' \ge x$ and $t \in \operatorname{supp}\mu_T$; since S is continuous on $[x, \infty)$ and $S(x') \to 0$ as $x' \to \infty$, in fact $1+yt > 0$ on $\operatorname{supp}\mu_T$, so $y \in (-1/s, 0)$. Letting $z \to x$ in (4.6) gives $x = f_{\mu_T, c}(y)$. Now let $s_d = \max\operatorname{supp}\rho_d \le s$. By [6, Lemmas 2.4–2.6], applied with $\pi = \rho_d$ and ratio ϕ_d, $\sup\operatorname{supp}\nu_d \le \max\{0, f_{\rho_d,\phi_d}(x_1)\}$, where x_1 is the unique critical point of f_{ρ_d,ϕ_d} in $(-1/s_d, 0)$. This function tends to $+\infty$ at both ends of that interval, because ρ_d has an atom at s_d, so $f_{\rho_d,\phi_d}(x_1)$ is its minimum there. Since $y \in (-1/s, 0) \subset (-1/s_d, 0)$,

$$\sup\operatorname{supp}\nu_d \le \max\{0, f_{\rho_d,\phi_d}(y)\}.$$

On $[\tau/2, s]$ the function $t \mapsto t/(1+yt)$ is bounded and continuous, because $1 + yt \ge 1 + ys > 0$. Weak convergence and $\phi_d \to c$ give $f_{\rho_d,\phi_d}(y) \to f_{\mu_T,c}(y) = x > 0$. Hence $\limsup_d \sup\operatorname{supp}\nu_d \le x$, and $x \downarrow \sup\operatorname{supp}\nu$ completes the proof. □

4.4. **Proof of Proposition 4.1.** All references in this proof are to the numbered statements of [3], version 3, Sections 1.4 and 2.

Normalization. The source works with $X_d = (\lambda/d)\sum_i T_i Y_i Y_i^{\mathsf T} = \lambda c_d \widehat{M}_d$, where $c_d = m_d/d$. Write

$$\sqrt{\frac{\lambda}{d}}\, Y_i = L\, r_i + L^{\perp}\frac{W_i}{\sqrt{d}}, \qquad r_i = \sqrt{\frac{\lambda}{d}}\, V_i. \tag{4.8}$$

Let $R \in \mathbb{R}^{m_d \times q}$ have rows $r_i^{\mathsf T}$, let $\mathcal{W} \in \mathbb{R}^{(d-q)\times m_d}$ have columns $W_i/\sqrt{d}$, and let $D = \operatorname{diag}(T_1, \dots, T_{m_d})$. This is the decomposition (2.34)–(2.36) of the source, with the Gaussian vector $g + \sqrt{\lambda}\, L^{\mathsf T}\mu_a$ there replaced by $\sqrt{\lambda}\, V_i$.

The conditioning. Let $\mathcal{G}$ be the σ-field generated by the marks $(T_i^{(d)}, V_i^{(d)})$ for all d and $i \le m_d$. For each d, $D = D^{(d)}$ and $R = R^{(d)}$ are $\mathcal{G}$-measurable, and, by the independence of the two families, conditionally on $\mathcal{G}$ the sequence $(\mathcal{W}^{(d)})_d$ has its unconditional joint law; in particular each $\mathcal{W}^{(d)}$ has i.i.d. $N(0, 1/d)$ entries. In the source, Assumption 1 (weights are functions of the label and of the Gaussian summary coordinates) enters the proof of Theorem 2.1 only through this structure and through the computation of a mean: it is used in Section 2.6 to conclude that $\mathcal{W}$ is independent of D, while R is a function of the same summary coordinates as D, and in Lemma 2.18 to write $\mathbb{E}\widetilde{F}_\delta^{(d)}$ (Step 3(d) below). The remark at the end of Section 2.6 that R has full rank by Gaussianity is not used in the proof of Theorem 2.1. We now go through the steps of the source proof and record, for each, what it uses.

Step 1: the weights alone. Because $T_i \in [\tau, M]$ almost surely, D is invertible and the empirical measure of D is τ'-regular in the sense of Definition 2.2 for every $\tau' \in (0, \min\{\tau, M^{-1}\})$, deterministically: its support lies in $[\tau, M] \subset [-\tau'^{-1}, \tau'^{-1}]$ and it gives mass $0 \le 1 - \tau'$ to $\{|x| \le \tau'\}$. (The endpoint $\tau' = \tau$ is excluded: if μ_T has an atom at τ, the mass of $\{|x| \le \tau\}$ can exceed $1 - \tau$.) Thus Lemma 2.3 holds with probability one. Assumption 2 holds because $\mu_T([0, \tau)) = 0$. Assumption 3 holds almost surely by Lemma 4.3. Lemmas 2.4–2.6 are deterministic statements about τ'-regular measures. Since τ' may be taken arbitrarily small, the ball $B(\tau'^{-1})$ appearing in Theorem 2.12 may be taken to contain any prescribed compact set.

Step 2: the Gaussian block with fixed weights. The following results of the source involve only $\mathcal{W}$ and D: Proposition 2.7 (bulk law of $\mathcal{W}D\mathcal{W}^{\mathsf T}$), Proposition 2.8 (no eigenvalues of $\mathcal{W}D\mathcal{W}^{\mathsf T}$ in E_δ eventually, a modification of [1]), Theorem 2.9 and Lemmas 2.10–2.11 (deterministic equivalent for the Green function of the linearization $\left[\begin{smallmatrix} -z & \mathcal{W} \\ \mathcal{W}^{\mathsf T} & -D^{-1} \end{smallmatrix}\right]$), Theorem 2.12, and Lemma 2.13 (convergence of the finite-d transform $S^{(d)}$). Their hypotheses on D are those verified in Step 1 (Proposition 2.7 needs only the weak convergence in Assumption 3; Proposition 2.8 and Theorem 2.12 also use its Hausdorff part, Theorem 2.9 uses invertibility and τ'-regularity, and Lemmas 2.10–2.11 use τ'-regularity and the event of Proposition 2.8), and their only hypothesis linking D and $\mathcal{W}$ is independence. Fix a realization of the marks in the $\mathcal{G}$-measurable probability-one event of Step 1 and Lemma 4.2. Along it the sequence $(D^{(d)})_d$ is deterministic and satisfies these hypotheses, and $(\mathcal{W}^{(d)})_d$ is Gaussian and independent of it, so these results apply and their conclusions hold with conditional probability one; integrating over the marks (Fubini's theorem), they hold almost surely. Theorem 2.12 allows the test vectors to be random provided they are independent of $\mathcal{W}$. We apply it with the q columns of $\widetilde{R}^{\mathsf T}$, where $\widetilde{R} = [0_{q\times(d-q)}\ R^{\mathsf T}]$; these are $\mathcal{G}$-measurable, hence independent of $\mathcal{W}$ conditionally on $\mathcal{G}$. Their norms are bounded almost surely, because $\|R\|_F^2 = \lambda d^{-1}\sum_i \|V_i\|^2 \to \lambda c\, \mathbb{E}\|V\|^2$ almost surely by Lemma 4.2 and (4.1).

Hypotheses of Theorem 2.9. Theorem 2.9 of the source is stated for a fixed $\tau > 0$, an invertible diagonal D whose empirical spectral measure is τ-regular, and an independent Gaussian matrix with i.i.d. entries. Its conclusion is a bound $Cd^{-1/4}(\operatorname{Im} z)^{-5}$ on $\langle v, (G(z) - \Pi(z))u\rangle$ for each pair of deterministic unit vectors, with probability $1 - O(n^{-100})$, uniformly in $|z| \le \tau^{-1}$. Its proof follows [6, Section 5 and Lemma 6.2]. The standing assumptions there [6, Section 2.1] are: the population matrix (here D) is deterministic, and positive definite in the diagonal case; the noise matrix has independent centred entries of variance $1/N$, where N is the dimension of the block carrying z, and every moment of $\sqrt{N}$ times an entry is bounded; the ratio of the dimension of the population matrix to N lies in $[\tau, \tau^{-1}]$; the population matrix has norm at most τ^{-1}, and its empirical eigenvalue distribution gives mass at most $1 - \tau$ to $[0, \tau]$. The high-probability bounds there are uniform over all data satisfying these conditions with the same τ [6,

Definition 3.4]. Here the population dimension is m_d and $N = d - q$; the letter M keeps its meaning as the upper bound of the weights. Fix a realization of the marks as in Step 2 and put $\tau_1 = \min\{\tau', c/2, 1/(2c)\}$. For all large d, D is deterministic, diagonal and positive definite (the source's extension to weights of either sign is not needed), $\|D\| \le M \le \tau_1^{-1}$, D has no eigenvalue in $[0, \tau_1]$, $\tau_1 \le m_d/(d-q) \le \tau_1^{-1}$, the matrix $\mathcal{W}$ is Gaussian and independent of the marks, and the test vectors used below are $\mathcal{G}$-measurable.

Two points are not covered literally. First, the normalization. The source states Theorem 2.9 for a matrix with standard Gaussian entries, and Propositions 2.7–2.8 for $n^{-1}WDW^{\mathsf{T}}$ with standard Gaussian W, but applies all three, in its Sections 2.9–2.10, to the matrix $\mathcal{W}$ of its Section 2.6. We read them in that normalization, in which the limiting measure is the one solving (2.11). The variance $1/(d-q)$ is the one under which (2.11), with ratio $m_d/(d-q)$, is the correct equation; $\mathcal{W}$ has variance $1/d$, as in Section 2.6 of the source. Write $\kappa_d = (d-q)/d$ and $\mathcal{W} = \kappa_d^{1/2}\widetilde{\mathcal{W}}$. Then

$$\begin{bmatrix} -z & \mathcal{W} \\ \mathcal{W}^{\mathsf{T}} & -D^{-1} \end{bmatrix} = \begin{bmatrix} I & 0 \\ 0 & \kappa_d^{1/2} I \end{bmatrix} \begin{bmatrix} -z & \widetilde{\mathcal{W}} \\ \widetilde{\mathcal{W}}^{\mathsf{T}} & -(\kappa_d D)^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & \kappa_d^{1/2} I \end{bmatrix},$$

and Theorem 2.9 applies to $\widetilde{\mathcal{W}}$ and $\kappa_d D$, which is τ_1-regular for large d because $\tau' < \tau$. The resulting deterministic equivalent for the original Green function has diagonal blocks $\widetilde{S}^{(d)} I_{d-q}$ and $-D(I + \kappa_d \widetilde{S}^{(d)} D)^{-1}$, where $\widetilde{S}^{(d)}$ solves (2.11) for the empirical law of the $\kappa_d T_i$. These laws converge weakly to μ_T and are supported in $[\tau/2, M]$ for large d, so Lemma 4.4, with $\tau/2$ in place of τ, gives $\widetilde{S}^{(d)} \to S$ and $\kappa_d \widetilde{S}^{(d)} \to S$ locally uniformly on $\mathbb{C}^+$, the same limit as $S^{(d)}$ in item (4) below. The extension to compact subsets of E_δ is as in item (4). Since $\mathcal{W}D\mathcal{W}^{\mathsf{T}} = \widetilde{\mathcal{W}}(\kappa_d D)\widetilde{\mathcal{W}}^{\mathsf{T}}$, the same rescaling applies to Propositions 2.7–2.8. This discrepancy, of relative order q/d, is already present in the source, and it changes no limit below. Second, the source's Theorem 2.9 does without the lower bound on $|1 + m(z)\sigma_i|$ and the stability condition assumed in [6, Section 3.4 and Definition 5.4]; the source replaces them by bounds in terms of $\operatorname{Im} z$ which it takes from [1, (3.20)–(3.21)]. This modification is internal to the source's proof and is not re-checked here. It is used only at spectral parameters with $\operatorname{Im} z \ge d^{-1/24}$ (Lemma 2.11) and $|z|$ bounded above and below, which lie in the spectral domain of [6, Section 3.1] after decreasing τ_1.

Avoiding the modification for (ii)–(v). The outlier assertions do not need the unverified part of the proof of Theorem 2.9. Work in the scale of X_d, let ν be the limiting measure of X_d, put $[a', b'] = \lambda c[a, b]$, and fix $\delta_0 \in (0, 1)$ with $a' - 3\delta_0 > \sup \operatorname{supp} \nu$. For (ii)–(v), Steps 2–5 need Lemma 2.11 and Theorem 2.12 only for z with $\operatorname{Re} z \in [a' - \delta_0, b' + \delta_0]$ and $0 \le \operatorname{Im} z$ small (the lower half-plane follows from $G(\bar{z}) = G(z)^*$), so they use Theorem 2.9 only at points with $\operatorname{Re} z \in [a' - \delta_0, b' + \delta_0]$ and $d^{-1/24} \le \operatorname{Im} z \le 1$. Let ϱ_d be the measure whose Stieltjes transform is $\widetilde{S}^{(d)}$; it is the measure ϱ of [6, Definition 2.3] for $\Sigma = \kappa_d D$ and ratio $m_d/(d-q)$. By Lemma 4.5, applied to the empirical law of the $\kappa_d T_i$ (supported in $[\tau/2, \max \operatorname{supp} \mu_T]$ for large d, and converging weakly to μ_T), $\operatorname{dist}(\operatorname{Re} z, \operatorname{supp} \varrho_d) \ge \delta_0$ at these points for all large d. These points therefore lie in the domain $\mathbf{D}^o$ of [6, Section 3.3], for suitable constants. Theorem 3.16(i) of [6], the anisotropic local law outside the spectrum, assumes only [6, Assumption 2.1 and (2.9)], which were checked above with $X = \widetilde{\mathcal{W}}^{\mathsf{T}}$ and $\Sigma = \kappa_d D$. It assumes no regularity of the spectrum, no lower bound on $|1 + m\sigma_i|$, and no stability condition. For fixed $\epsilon, D' > 0$ and deterministic unit vectors u, v, it gives

$$|\langle v, (G(z) - \Pi(z))u \rangle| \le N^{\epsilon} \max\{1, M\}^2 \Psi(z), \qquad \Psi(z) \le \frac{2}{N^{1/2} \operatorname{Im} z},$$

with probability at least $1 - N^{-D'}$, for all large d, uniformly in these z; here G and Π are written for $(\widetilde{\mathcal{W}}, \kappa_d D)$, and the bound $\Psi(z) \le 2N^{-1/2}(\operatorname{Im} z)^{-1}$ follows from $\operatorname{Im} m \le |m| \le (\operatorname{Im} z)^{-1}$. For large d this is smaller than the bound $Cd^{-1/4}(\operatorname{Im} z)^{-5}$ of Theorem 2.9. A union bound over a grid of mesh d^{-3} makes it simultaneous in z, because G and Π are Lipschitz in z there with constants polynomial in d on the event $\|\widetilde{\mathcal{W}}\| \le C_0$, whose complement has probability at most e^{-d} for C_0 large. Lemmas 2.10–2.11 and Theorem 2.12, restricted to a complex neighbourhood of $[a', b']$, then hold as stated; their proofs still use Proposition 2.8. The modified Theorem 2.9 therefore enters only the alternative argument for (i) in item (2) of the next paragraph.

Aspect ratio. The source states Theorem 2.1 with $n/d = \phi$; here $m_d/d \to c$. The ratio enters the proof in the following places. (1) Proposition 2.8 is stated for $n/d \to \phi$; in its proof the function ψ_d is formed with the finite-d ratio, and the uniform convergence $\psi_d \to \psi$, $\psi_d' \to \psi'$ on $[a, b]$ persists because $\phi_d \to c$.

(2) Proposition 2.7 is quoted by the source from Silverstein and Bai. Independently of how that theorem treats the ratio, its conclusion here also follows from results already listed: applying Theorem 2.9 to the $d-q$ coordinate vectors of the upper block, a union bound and the Borel–Cantelli lemma show that, for each fixed $z \in \mathbb{C}^+$, the normalized trace of $(\mathcal{W}D\mathcal{W}^\mathsf{T} - z)^{-1}$ minus $\widetilde{S}^{(d)}(z)$ (in the normalization above) tends to zero almost surely; $\widetilde{S}^{(d)}(z)$ converges to the limiting transform, as shown above, and convergence of Stieltjes transforms on a countable dense subset of $\mathbb{C}^+$ to the Stieltjes transform of a probability measure gives almost sure weak convergence. (3) Theorem 2.9 and Lemmas 2.4–2.6 and 2.10–2.11 are finite-d statements written with the finite-d ratio of (2.11), which varies with d already in the source. (4) Lemma 2.13: by Lemma 4.4(b), applied to the empirical weight laws (which converge weakly by Lemma 4.3) and to $\phi_d = m_d/(d-q) \to c$, the finite-d transforms converge to the limiting one locally uniformly on $\mathbb{C}^+$; Lemma 4.4(a) also identifies the source's $S^{(d)}$ as the unique solution of (2.11) in $\mathbb{C}^+$. The extension to compact subsets of E_δ is then as in the source, through Proposition 2.8 and Lemma 2.10. (5) Lemmas 2.16–2.18 contain the factor n/d; here it is $c_d \to c$, see Step 3.

Step 3: the reduced matrix. Theorem 2.14 states that $F^{(d)}(z) = R^\mathsf{T} D(I + S^{(d)}(z)D)^{-1}R$ converges to its deterministic limit uniformly on $E_{\delta,K,\eta}$. Its proof uses four ingredients.

(a) Lemma 2.15 compares $D(I + S^{(d)}D)^{-1}$ with a cut-off version; it uses only Step 1, Lemma 2.13 and Lemmas 2.5–2.6.

(b) Lemma 2.16 combines Lemma 2.15 with almost sure boundedness of

$$\frac{1}{d}\sum_\ell (\sqrt{d}R_{\ell i})^2(\sqrt{d}R_{\ell j})^2 = \frac{\lambda^2}{d}\sum_\ell V_{\ell i}^2 V_{\ell j}^2,$$

which holds by Lemma 4.2 and (4.1).

(c) Lemma 2.17 applies the law of large numbers, at fixed z, to i.i.d. summands $\lambda\,\theta_\delta(1+S(z)T_\ell)\,T_\ell(1+S(z)T_\ell)^{-1}V_\ell V_\ell^\mathsf{T}$, where S is the deterministic limit and θ_δ is the smooth cut-off; the cut-off bounds the scalar factor, so all moments are finite by (4.1) and Lemma 4.2 applies. Equicontinuity in z uses the same bound times $d^{-1}\sum_\ell \lambda\|V_\ell\|^2$, which converges almost surely. A countable dense set of z and the Arzelà–Ascoli theorem give local uniformity, as in the source.

(d) Lemmas 2.18–2.19 identify the mean. Because the summands are i.i.d. with law determined by μ,

$$\mathbb{E}\widetilde{F}_\delta^{(d)}(z) = c_d\,\lambda\,\mathbb{E}\left[\theta_\delta(1+S(z)T)\frac{T}{1+S(z)T}VV^\mathsf{T}\right], \tag{4.9}$$

which converges to $c\lambda\mathbb{E}[\cdots]$; since μ does not depend on d, Lemma 2.19 is not needed. The source uses Gaussianity of g only to rewrite this expectation in terms of g and G. For the removal of the cut-off as $\delta \to 0$, note that for positive weights $|1+S(z)t|$ is bounded below by some $\kappa > 0$ uniformly over $t \in \operatorname{supp}\mu_T$ and z in a compact subset of E_δ: for nonreal z, $\operatorname{Im}(1+S(z)t) = t\operatorname{Im}S(z) \neq 0$; for real z, $1+S(z)t \neq 0$ by the Silverstein–Choi characterization used in the paragraph *Denominators* below (which involves only the limiting equation); continuity and compactness give κ. For $\delta < \kappa$ the cut-off is identically one, so the truncated and untruncated limits coincide.

Therefore $F^{(d)} \to \widetilde{F}$ uniformly on $E_{\delta,K,\eta}$ almost surely, where

$$\widetilde{F}(z) = c\lambda\,\mathbb{E}\left[\frac{T}{1+S(z)T}VV^\mathsf{T}\right]. \tag{4.10}$$

Step 4: linearization and Schur complement. Since D is invertible, Section 2.10 of the source applies verbatim: rotating by $L \oplus L^\perp$, the linearization of X_d becomes $\left[\begin{smallmatrix} -zI_q & \widetilde{R} \\ \widetilde{R}^\mathsf{T} & \mathcal{N}(z) \end{smallmatrix}\right]$ with $\mathcal{N}(z) = \left[\begin{smallmatrix} -zI_{d-q} & \mathcal{W} \\ \mathcal{W}^\mathsf{T} & -D^{-1} \end{smallmatrix}\right]$. By Step 2, $\mathcal{N}(z)$ is invertible on E_δ for all large d. Put $\Delta_d(z) = zI_q + \widetilde{R}\mathcal{N}(z)^{-1}\widetilde{R}^\mathsf{T}$. Since $(X_d - z)^{-1}$ is the upper-left block of the inverse linearization, the Schur complement formula gives

$$L^\mathsf{T}(X_d - z)^{-1}L = -\Delta_d(z)^{-1}, \tag{4.11}$$

and the eigenvalues of X_d in $E_\delta \cap \mathbb{R}$ are the zeros of $\det\Delta_d$, with multiplicity equal to the order of the zero. By Step 2 and the bounded norms of the columns of $\widetilde{R}^\mathsf{T}$, $\Delta_d(z) - (zI_q - F^{(d)}(z)) \to 0$ uniformly, and by Step 3, $\Delta_d \to zI_q - \widetilde{F}$ uniformly on $E_{\delta,K,\eta}$, almost surely. Since these functions are holomorphic,

their derivatives converge uniformly on slightly smaller sets by the Cauchy estimates. Hurwitz's theorem gives the root counts, exactly as in the source.

Step 5: eigenvectors. Sections 2.11–2.12 of the source use only the linearization identity (2.51)–(2.52), the invertibility of $\mathcal{N}$, and the convergence of Δ_d and Δ_d' from Step 4. They give, for an eigenvalue $z_d \to z_* \neq 0$ of X_d with unit eigenvector u_d, that every limit point u of $L^{\mathsf{T}} u_d$ satisfies $(z_* - \widetilde{F}(z_*))u = 0$ and $\|u\|^2 - u^{\mathsf{T}} \widetilde{F}'(z_*) u = 1$. In fact the computation there gives, for $v_d = L^{\mathsf{T}} u_d$, the exact identities $\Delta_d(z_d) v_d = 0$ and $v_d^{\mathsf{T}} \Delta_d'(z_d) v_d = 1$: with $\mathcal{N}$ real symmetric for real z, $\Delta_d' = I_q + \widetilde{R} \mathcal{N}^{-1} (I_{d-q} \oplus 0) \mathcal{N}^{-1} \widetilde{R}^{\mathsf{T}}$, and the second term evaluated at v_d is the squared norm of the component of u_d orthogonal to $\operatorname{Ran} L$. Together with the uniform convergence of Δ_d and Δ_d' from Step 4 these identities give the limits in (v) directly.

Rescaling. Let $S(z)$ be the limiting transform of X_d, so $S(\lambda c x) = s_c(x)/(\lambda c)$. The equation $1 + zS = c\,\mathbb{E}[ST/(1+ST)]$ for X_d becomes (4.3), and (4.10) gives $\widetilde{F}(\lambda c x) = \lambda c F_c(x)$, hence $\widetilde{F}'(\lambda c x) = F_c'(x)$. Define $B_d(x) = (\lambda c_d)^{-1} \Delta_d(\lambda c_d x)$. Then (4.11) becomes (4.5), and, since $c_d \to c$, $B_d \to B_c$ together with the derivative, uniformly on a thin complex neighbourhood Ω of $[a, b]$ contained in the rescaled set $E_{\delta,K,\eta}$. This proves (iv); (iii) and (v) follow from Steps 4–5, and (i) from Section 2.9 of the source, which uses only Proposition 2.7 and Cauchy interlacing for a perturbation of rank at most $2q$.

Denominators. For (ii), the Silverstein–Choi characterization of the complement of the support [11, Theorem 4.1], applied to the bulk measure of X_d, which solves (2.11) with weight law μ_T, shows that $-1/S(z) \notin \operatorname{supp} \mu_T$ for real z outside the support; this characterization is also the one used in the proof of Proposition 2.8. Equivalently $\lambda c + s_c(x) t \neq 0$ for $x > \lambda_{+,c}$ and $t \in \operatorname{supp} \mu_T$. Continuity of s_c on K and compactness of $K \times \operatorname{supp} \mu_T$ give a positive infimum, and real analyticity follows by differentiating under the expectation. □

Remark 4.6 (Where the inverse weight matrix enters). In the source proof the diagonal inverse appears only in the lower-right block of the linearization, in the Schur complement producing the characteristic determinant, in the relation (2.52) between the eigenvector and the linearization vector, and in the differentiation of the Schur complement giving the norm identity. After (3.4), each occurrence is an inverse of $D_{+,d}$ and is bounded by τ^{-1}. The bulk theorem does not require the inverse and already permits a zero atom. No Moore–Penrose substitution is made. The rotation-invariance argument of Section 2.13 of the source, which uses the Gaussian law of g, is not used; its root bound is replaced by the monotonicity argument of Section 7.

5. Bulk law, exact scaling, and the zero atom

Under the active law (3.5), set

$$T = h_b(g), \qquad V = m_b + \lambda^{-1/2} g, \qquad c = \phi\pi. \tag{5.1}$$

This mark law satisfies the hypotheses of Proposition 4.1: $T \in [\tau, M]$, and $\mathbb{E}_+ \|V\|^r \leq \pi^{-1} \mathbb{E} \|m_b + \lambda^{-1/2} g\|^r < \infty$. Let s_+ solve

$$1 + x s_+(x) = c\,\mathbb{E}_+ \left[\frac{s_+(x) T}{\lambda c + s_+(x) T} \right]. \tag{5.2}$$

Define

$$S(z) = \frac{1}{\pi} s_+\Big(\frac{z}{\pi}\Big). \tag{5.3}$$

Since

$$\mathbb{E}_+[f(b,g)] = \frac{1}{\pi} \sum_b p_b \mathbb{E}\big[\mathbf{1}_{\{h_b(g)>0\}} f(b,g)\big], \tag{5.4}$$

substitution in (5.2) yields (2.8). Inactive terms vanish because the numerator is zero; no inverse of zero is used.

If ν_+ is the active bulk measure, then

$$\nu_G = (x \mapsto \pi x)_{\#} \nu_+, \qquad \operatorname{supp} \nu_G = \pi \operatorname{supp} \nu_+, \qquad \lambda_+(G) = \pi \lambda_{+,+}. \tag{5.5}$$

This exact pushforward supplies the right-edge statement without invoking continuity of support under weak convergence.

For the effective matrix,

$$F_+(x) = \lambda c\, \mathbb{E}_+\left[\frac{T}{\lambda c + s_+(x)T} V V^\intercal\right]. \tag{5.6}$$

Using $c = \phi\pi$ and (5.3),

$$F_G(z) = \pi F_+\Big(\frac{z}{\pi}\Big), \tag{5.7}$$

which expands exactly to (2.11). Hence

$$\det B_G(z) = \pi^q \det\Big[\frac{z}{\pi} I_q - F_+\Big(\frac{z}{\pi}\Big)\Big]. \tag{5.8}$$

Roots and analytic multiplicities correspond bijectively under $z = \pi x$.

Lemma 5.1 (Zero mass)**.** *The limiting measure satisfies* (2.12)*.*

Proof. The rank theorem proved in Section 6 gives the lower bound

$$\nu_G(\{0\}) \geq (1 - \phi\pi)_+. \tag{5.9}$$

Alternatively, the same lower bound follows from (5.5) and the standard active covariance rank deficit.

Let $m_0 = \nu_G(\{0\})$. For $t \downarrow 0$,

$$tS_G(-t) \longrightarrow m_0. \tag{5.10}$$

If $m_0 > 0$, then $S_G(-t) \to +\infty$. Dominated convergence in (2.8) gives

$$1 - m_0 = \phi \sum_b p_b \mathbb{P}\{h_b(g) > 0\} = \phi\pi. \tag{5.11}$$

Thus $m_0 = 1 - \phi\pi$ when $\phi\pi < 1$. If $\phi\pi \geq 1$, a positive m_0 would imply $1 - m_0 = \phi\pi \geq 1$, which is impossible. Therefore $m_0 = (1 - \phi\pi)_+$. □

6. Exact rank and kernel

Lemma 6.1 (General position of active columns)**.** *Conditioned on activity, each Y_ℓ has an absolutely continuous law on $\mathbb{R}^d$. Independent active columns are in general linear position almost surely.*

Proof. Under (3.5), g has a density on its active set, and w remains a full standard Gaussian. For each class b, the affine orthogonal map

$$(g, w) \mapsto L_d(m_b + \lambda^{-1/2} g) + \lambda^{-1/2} L_d^\perp w \tag{6.1}$$

has nonsingular linear part. Its image law is absolutely continuous on $\mathbb{R}^d$. Inductively, conditioned on the first $r < d$ active columns, their span is a proper linear subspace and hence has Lebesgue measure zero. The next active column avoids it almost surely. □

Therefore

$$\operatorname{rank} A_{+,d} = \min(d, N_d) \quad \text{a.s.} \tag{6.2}$$

Since $D_{+,d}^{1/2}$ is invertible,

$$\operatorname{rank} M_d = \operatorname{rank}(A_{+,d} D_{+,d}^{1/2}) = \min(d, N_d) \tag{6.3}$$

almost surely. Equation (2.22) follows from (3.7). The exact zeros are the finite-dimensional realization of the atom (2.12); they are not positive outliers.

7. Analyticity and root multiplicity

On a compact set to the right of the active bulk, Proposition 4.1(ii) supplies

$$\operatorname*{ess\,inf}_{T} |\lambda c + s_+(x)T| > 0. \tag{7.1}$$

Equations (5.3) and (5.7) imply (2.13). Thus F_G is real analytic on (λ_+, ∞).

For real $z > \lambda_+$,

$$S_G'(z) = \int \frac{\nu_G(dx)}{(x-z)^2} > 0. \tag{7.2}$$

Differentiating (2.11),

$$F_G'(z) = -\lambda\phi S_G'(z) \sum_b p_b \mathbb{E}\left[\frac{h_b(g)^2 v_b(g) v_b(g)^{\mathsf{T}}}{[\lambda\phi + S_G(z)h_b(g)]^2}\right] \preceq 0. \tag{7.3}$$

Hence

$$B_G'(z) = I_q - F_G'(z) \succeq I_q. \tag{7.4}$$

Every ordered eigenvalue of the Hermitian matrix $B_G(z)$ is strictly increasing, so each can cross zero at most once and the sum of root multiplicities is at most q.

Let z_* be a root and $K_* = \operatorname{Ker} B_G(z_*)$, $\dim K_* = m$. In the orthogonal decomposition $K_* \oplus K_*^{\perp}$,

$$B_G(z_* + t) = \begin{bmatrix} tJ_* + O(t^2) & tC + O(t^2) \\ tC^{\mathsf{T}} + O(t^2) & B_{22} + O(t) \end{bmatrix}, \tag{7.5}$$

where B_{22} is invertible and

$$J_* = P_* B_G'(z_*) P_*|_{K_*} \succ 0. \tag{7.6}$$

The Schur complement on K_* equals $tJ_* + O(t^2)$, and therefore

$$\det B_G(z_* + t) = \det(B_{22})\, t^m \det(J_*) + O(t^{m+1}). \tag{7.7}$$

This proves (2.15).

8. Random active size and empirical outlier counts

Lemma 8.1 (Random-index transfer)**.** *Let $r_{d,m} \in [0,1]$ be a failure probability for a model with m active columns. Suppose*

$$r_{d,m_d} \to 0 \tag{8.1}$$

for every deterministic sequence $m_d/d \to c$. If $N_d/d \to c$ in probability, then

$$\mathbb{E} r_{d,N_d} \to 0. \tag{8.2}$$

Proof. Choose $\eta_d \downarrow 0$ such that

$$\mathbb{P}\{|N_d/d - c| > \eta_d\} \to 0. \tag{8.3}$$

If

$$\sup_{|m/d - c| \le \eta_d} r_{d,m} \tag{8.4}$$

did not tend to zero, a subsequence and a choice $m_d/d \to c$ would contradict the deterministic-sequence hypothesis. Splitting the expectation on the typical and atypical events gives (8.2). □

For each deterministic $m_d/d \to c = \phi\pi$, Proposition 4.1 gives the exact active eigenvalue count on compact intervals separated from the active bulk; almost sure eventual statements imply that the corresponding failure probabilities tend to zero. Because

$$M_d = \alpha_d \widehat{M}_d, \qquad \alpha_d = \frac{N_d}{n} \to \pi, \tag{8.5}$$

one has

$$\#\{\sigma(M_d)\cap[a,b]\}=\#\left\{\sigma(\widehat{M}_d)\cap\left[\frac{a}{\alpha_d},\frac{b}{\alpha_d}\right]\right\}. \tag{8.6}$$

The active endpoints converge to a/π and b/π. Uniform determinant convergence in a small endpoint neighbourhood permits these converging endpoints. Equation (5.8) identifies the active root count with the target root count. The random-index lemma removes conditioning and proves (2.14).

9. No spurious right outliers and existence

The operator norm is tight because

$$0\preceq M_d\preceq\frac{M}{n}A_dA_d^{\mathsf{T}}. \tag{9.1}$$

Writing $A_d=Z_d+\mathcal{M}_d$, the Gaussian norm satisfies $\|Z_d\|/\sqrt{n}=O_{\mathbb{P}}(1)$, while

$$\|\mathcal{M}_d\|/\sqrt{n}\le\max_b\|\mu_b^{(d)}\|=O(1). \tag{9.2}$$

Thus $\|M_d\|=O_{\mathbb{P}}(1)$.

Also $F_G(z)$ remains bounded as $z\to+\infty$, while zI_q diverges, so effective roots lie in a bounded interval. Fix $\epsilon>0$ and a large deterministic K. The right roots in $[\lambda_++\epsilon,K]$ are finite. Remove small disjoint neighbourhoods of these roots. The complement is a finite union of compact root-free intervals, and the count theorem gives no empirical eigenvalues there with probability tending to one. This proves the no-spurious direction. Applying the count theorem to an isolating interval, and then to nested isolating intervals, proves existence and convergence at each effective root.

10. Simple roots and spectral projections

For the active model, Proposition 4.1(v) gives, near an isolated root $x_*=z_*/\pi$,

$$\|(x_dI-F_+(x_d))r_d\|=o_{\mathbb{P}}(1), \tag{10.1}$$

$$\left|r_d^{\mathsf{T}}[I-F_+'(x_d)]r_d-1\right|=o_{\mathbb{P}}(1). \tag{10.2}$$

Multiplication by α_d changes eigenvalues but not eigenvectors. From (5.7),

$$F_G'(z)=F_+'\left(\frac{z}{\pi}\right). \tag{10.3}$$

Local uniformity and $\alpha_d\to\pi$ yield (2.16)–(2.17). At a simple root, (2.15) makes the kernel one-dimensional, and (2.17) fixes the norm, proving (2.18).

For a repeated root, choose a positively oriented contour Γ enclosing z_* and no other effective root or bulk point, of radius smaller than the strip width in Proposition 4.1(iv). The active compressed resolvent identity (4.5), the uniform convergence $B_d\to B_c$, and the scaling (5.7) give

$$\sup_{z\in\Gamma}\left\|L_d^{\mathsf{T}}(M_d-zI)^{-1}L_d+B_G(z)^{-1}\right\|\xrightarrow{\mathbb{P}}0. \tag{10.4}$$

The full-space spectral projection is

$$\Pi_d=-\frac{1}{2\pi i}\oint_\Gamma(M_d-zI)^{-1}\,dz. \tag{10.5}$$

Thus

$$L_d^{\mathsf{T}}\Pi_dL_d\longrightarrow\frac{1}{2\pi i}\oint_\Gamma B_G(z)^{-1}\,dz. \tag{10.6}$$

The block expansion near z_* gives

$$B_G(z_*+t)^{-1}=\frac{1}{t}P_*J_*^{-1}P_*+O(1). \tag{10.7}$$

The residue is $P_*J_*^{-1}P_*$, proving (2.20). If U_d is any orthonormal basis of $\operatorname{Ran}\Pi_d$ and $R_d=L_d^{\mathsf{T}}U_d$, then

$$R_dR_d^{\mathsf{T}}\xrightarrow{\mathbb{P}}P_*J_*^{-1}P_*. \tag{10.8}$$

Only the invariant subspace is canonical; individual eigenvectors may rotate.

11. ReLU gates

Corollary 11.1. *Assume*

$$h_b(g) = a_b(g)\mathbf{1}_{\{\ell_b(g)>0\}}, \tag{11.1}$$

where ℓ_b is a nonconstant affine functional and

$$\tau \le a_b(g) \le M \quad \text{on } \{\ell_b(g) > 0\}. \tag{11.2}$$

Then Theorem 2.1 applies.

Proof. A nonconstant affine functional of a standard Gaussian is a nondegenerate one-dimensional Gaussian, so its positive half-space has probability strictly between zero and one. The weights belong to $\{0\} \cup [\tau, M]$ almost surely. □

For the pure gate $h_b(g) = \mathbf{1}_{\{\ell_b(g)>0\}}$,

$$M_d = \frac{1}{n}\sum_{\ell=1}^{n} \mathbf{1}_{\{\ell_{b_\ell}(g_\ell)>0\}} Y_\ell Y_\ell^{\mathsf{T}} \tag{11.3}$$

is a half-space-gated empirical second moment. For a one-unit ReLU model

$$f_w(Y) = a\,\max(w^{\mathsf{T}}Y, 0), \tag{11.4}$$

at a fixed parameter for which no sample lies at the kink,

$$\nabla_w f_w(Y) = a\,\mathbf{1}_{\{w^{\mathsf{T}}Y>0\}} Y. \tag{11.5}$$

The w–w Jacobian-Gram or Gauss–Newton block is therefore

$$\frac{a^2}{n}\sum_{\ell=1}^{n} \mathbf{1}_{\{w^{\mathsf{T}}Y_\ell>0\}} Y_\ell Y_\ell^{\mathsf{T}}. \tag{11.6}$$

For squared loss and fixed deterministic w, Gaussian data avoid the kink almost surely, the activation pattern is locally constant, and the classical Hessian block equals this matrix. This does not establish a trajectory-uniform theorem, cross-neuron blocks, or a generalized Hessian at a kink. Hard ReLU gates thus create a zero atom, not additional positive outliers.

12. Examples and numerical illustration

12.1. **Independent Bernoulli thinning.** If $h \sim \text{Bernoulli}(\pi)$ independently of the columns, then

$$M_d = \frac{N_d}{n}\left(\frac{1}{N_d}A_+A_+^{\mathsf{T}}\right), \qquad \frac{N_d}{d} \to c = \phi\pi. \tag{12.1}$$

The continuous bulk edges are

$$\lambda_\pm = \frac{(1 \pm \sqrt{c})^2}{\lambda\phi} = \frac{\pi}{\lambda}\left(1 \pm \frac{1}{\sqrt{c}}\right)^2, \tag{12.2}$$

and the zero mass is $(1 - c)_+$. Equation (2.8) becomes

$$zS(z)^2 + (1 + \lambda\phi z - \phi\pi)S(z) + \lambda\phi = 0. \tag{12.3}$$

12.2. **Centred pure gate.** Let $Y = \lambda^{-1/2}Z$ with $Z \sim N(0, I_d)$ and $h = \mathbf{1}_{\{\langle \ell, Z\rangle>0\}}$. Conditional on ZZ^{T}, the only Gaussian preimages are Z and $-Z$, with equal probability, and the gate takes complementary values. Therefore

$$\mathbb{P}\{h = 1 \mid YY^{\mathsf{T}}\} = \frac{1}{2}, \tag{12.4}$$

so h is independent of YY^{T}. The model is exactly independent Bernoulli thinning with $\pi = 1/2$.

12.3. **Noncentred scalar gate.** Let $g \sim N(0,1)$,

$$h(g) = \mathbf{1}_{\{a+bg>0\}}, \qquad b \neq 0, \tag{12.5}$$

and define

$$\delta = \frac{a}{|b|}, \quad s_b = \operatorname{sign}(b), \quad \pi = \Phi(\delta), \quad \kappa = \frac{\varphi(\delta)}{\Phi(\delta)}. \tag{12.6}$$

Then

$$\mathbb{E}[g \mid h = 1] = s_b\kappa, \qquad \mathbb{E}[g^2 \mid h = 1] = 1 - \delta\kappa. \tag{12.7}$$

For $v = m + \lambda^{-1/2}g$, set

$$c = \phi\pi, \qquad \beta = \lambda\mathbb{E}[v^2 \mid h = 1] = \lambda m^2 + 2\sqrt{\lambda}\, m s_b\kappa + 1 - \delta\kappa. \tag{12.8}$$

The scalar effective equation gives the standard spiked-covariance threshold

$$\beta > 1 + \frac{1}{\sqrt{c}} \tag{12.9}$$

and, above threshold,

$$z_{\text{out}} = \frac{\pi}{\lambda}\,\beta\left(1 + \frac{1}{c(\beta-1)}\right). \tag{12.10}$$

12.4. **Fixed-ϵ comparison.** For comparison only, define

$$D_\epsilon = D + \epsilon P_0, \qquad P_0 = \operatorname{diag}(\mathbf{1}_{\{D_{\ell\ell}=0\}}). \tag{12.11}$$

Then

$$M_{d,\epsilon} - M_d = \frac{\epsilon}{n} A_d P_0 A_d^{\mathsf{T}} \tag{12.12}$$

and

$$\|M_{d,\epsilon} - M_d\| \le \epsilon \frac{\|A_d\|^2}{n} = O_{\mathbb{P}}(\epsilon). \tag{12.13}$$

This perturbation estimate does not, by itself, prove support-edge or root-multiplicity stability. The exact active reduction avoids that missing limit interchange.

12.5. **Numerical illustration.** Figure 1 shows the noncentred scalar gate. The active conditional moment predicts an outlier at 4.6653, separated from the bulk edge 2. Direct simulations approach the prediction while the associated eigenvectors retain large overlap with the summary direction. Representative values from the same computations are listed below.

Quantity	Value
Bernoulli model rank	254 observed, 254 predicted
Bernoulli zero mass	0.20625 observed, 0.2 limiting
Bernoulli top eigenvalue	1.1641 observed, 1.1963 edge
Stieltjes residual	2.39×10^{-16}
Centred gate top eigenvalue	1.9728, edge 2
Noncentred predicted outlier	4.6653
Noncentred simulations	4.48–4.87
Determinant residual	1.62×10^{-10}
Regularization identity error	$< 1.1 \times 10^{-17}$
Two-construction operator difference	1.93×10^{-15}

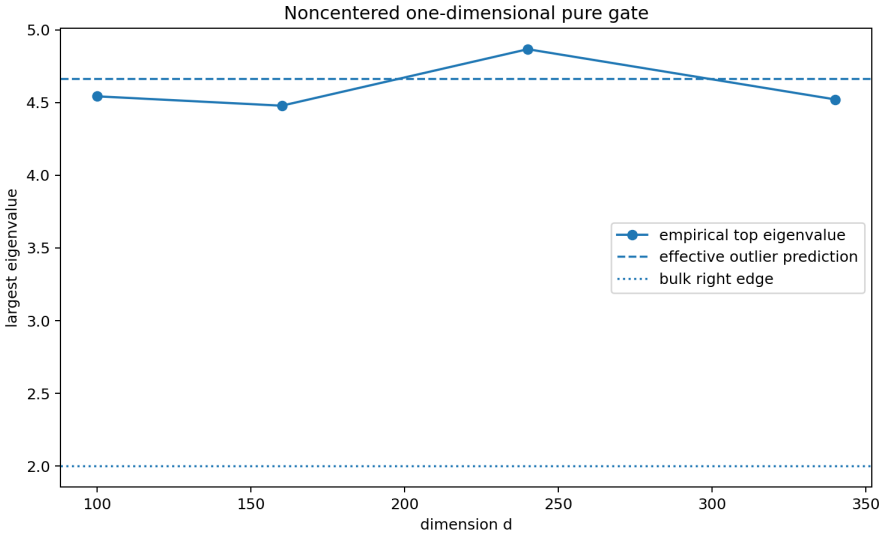


FIGURE 1. Largest eigenvalue for the gate $h(g) = \mathbf{1}_{\{g>0\}}$ with $\phi = 2$, $\lambda = 1$, and summary mean $m = 2$. The effective determinant predicts $z_{\text{out}} = 4.6653$ and the bulk edge is 2.

12.6. **A two-class check with non-constant weights.** The following check, added in the present revision, is numerical evidence only. Take $k = 2$, $C = 0$, $q = 2$, $p_1 = p_2 = 1/2$, $\lambda = 1$, $\phi = 1.2$ and $\sqrt{G} = 4I_2$, so that $m_1 = (4, 0)$ and $m_2 = (0, 4)$. Let

$$h_1(g) = a_1(g_1)\mathbf{1}_{\{g_2>-0.3\}}, \qquad h_2(g) = a_2(g_1)\mathbf{1}_{\{g_2<0.5\}},$$

where a_1 equals 0.5, 1, 1.5 on $(-\infty, -0.5)$, $[-0.5, 0.7)$, $[0.7, \infty)$, and a_2 equals 0.6, 1.2 on $(-\infty, 0]$, $(0, \infty)$. This is a gate of the form of Section 11 with $\tau = 0.5$ and $M = 1.5$. Here $\pi = 0.6547$, so $\nu_G(\{0\}) = 1 - \phi\pi = 0.2144$. Because the weights are piecewise constant, every expectation in (2.8) and (2.11) is a finite sum of truncated Gaussian moments. Inverting the map $S_G \mapsto z$ defined by (2.8) gives $\lambda_+(G) = 3.0976$. Bracketing the increasing eigenvalues of B_G gives two simple roots, $z_1 = 7.8129$ and $z_2 = 4.7294$; by (2.18) the predicted values of $\|r_d\|^2 = 1/(e_*^{\mathsf{T}} B_G'(z_*)e_*)$ are 0.8411 and 0.7348. Table 1 compares these with direct simulations of M_d, $n = 1.2d$. In all 80 eigenvector comparisons the

TABLE 1. Two-class model: mean ± standard deviation over independent runs of the three largest eigenvalues of M_d and of $\|r_d\|^2$ for the top two eigenvectors.

d	runs	λ_1	λ_2	λ_3	$\|r_{d,1}\|^2$	$\|r_{d,2}\|^2$
500	20	7.77 ± 0.48	4.68 ± 0.18	3.05 ± 0.08	0.842 ± 0.014	0.734 ± 0.015
1000	10	7.60 ± 0.29	4.72 ± 0.12	3.01 ± 0.05	0.837 ± 0.008	0.737 ± 0.014
2000	6	7.80 ± 0.40	4.80 ± 0.14	3.09 ± 0.05	0.842 ± 0.009	0.738 ± 0.008
4000	4	7.80 ± 0.10	4.70 ± 0.04	3.09 ± 0.02	0.841 ± 0.005	0.733 ± 0.003
prediction		7.8129	4.7294	edge 3.0976	0.8411	0.7348

cosine of the angle between r_d and the predicted kernel direction e_* was at least 0.98. For $d \le 1000$, $\dim \operatorname{Ker} M_d = d - N_{\text{active}}$ in all 30 runs, as (2.21) requires. The same code applied to the scalar gate of Figure 1 gives mean top eigenvalues 4.61 ± 0.13, 4.68 ± 0.10, 4.67 ± 0.06 and 4.62 ± 0.05 for $d = 500$, 1000, 2000, 4000 (20, 10, 6, 4 runs), against 4.6653. The mean of $\|r_d\|^2$ lies between 0.859 and 0.862 for each d, against 0.8609 from (2.18), which here coincides with the classical spiked-covariance overlap.

AI ASSISTANCE DISCLOSURE

OpenAI GPT-5.6 Sol, accessed through ChatGPT and Codex, was used to develop the reduction, the numerical checks and the first draft. Anthropic Claude Opus 5.5, used through Claude Code, was used to merge and edit the note and to write the conditioning argument of Section 4, including Lemma 4.4. Anthropic Claude Opus 5.5 (Claude Code) also produced Lemma 4.5, the paragraph of Section 4 that replaces the unverified part of Theorem 2.9 of [3] by [6, Theorem 3.16] for the outlier assertions, and the two-class numerical check of Section 12. The author reviewed and takes responsibility for every statement and proof. AI systems are not authors of this work, and the work has not been peer reviewed.

REFERENCES

[1] Z. D. Bai and Jack W. Silverstein, *No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices*, Annals of Probability **26** (1998), no. 1, 316–345.

[2] Jinho Baik, Gérard Ben Arous, and Sandrine Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Annals of Probability **33** (2005), no. 5, 1643–1697.

[3] Gérard Ben Arous, Reza Gheissari, Jiaoyang Huang, and Aukosh Jagannath, *Local geometry of high-dimensional mixture models: Effective spectral theory and dynamical transitions*, 2025, arXiv:2502.15655, version 3 (22 January 2026).

[4] Florent Benaych-Georges and Raj Rao Nadakuditi, *The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices*, Advances in Mathematics **227** (2011), no. 1, 494–521.

[5] Zhou Fan, Renyuan Ma, Elliot Paquette, and Zhichao Wang, *Anisotropic local law for non-separable sample covariance matrices*, 2026, arXiv:2602.17960.

[6] Antti Knowles and Jun Yin, *Anisotropic local laws for random matrices*, Probability Theory and Related Fields **169** (2017), no. 1-2, 257–352, arXiv:1410.3516; statement numbers as in the arXiv version.

[7] Filip Kovačević, Yihan Zhang, and Marco Mondelli, *Spectral estimators for multi-index models: Precise asymptotics and optimal weak recovery*, Proceedings of Thirty Eighth Conference on Learning Theory, Proceedings of Machine Learning Research, vol. 291, PMLR, 2025, https://proceedings.mlr.press/v291/kovacevic25a.html, pp. 3354–3404.

[8] V. A. Marčenko and L. A. Pastur, *Distribution of eigenvalues for some sets of random matrices*, Mathematics of the USSR-Sbornik **1** (1967), no. 4, 457–483.

[9] Andrea Montanari and Basil Saeed, *Variational formulas for the spectrum of block Wishart matrices*, 2026, arXiv:2606.27774.

[10] Vinod Nair and Geoffrey E. Hinton, *Rectified linear units improve restricted Boltzmann machines*, Proceedings of the 27th International Conference on Machine Learning (Haifa, Israel), 2010.

[11] Jack W. Silverstein and Sang-Il Choi, *Analysis of the limiting spectral distribution of large dimensional random matrices*, Journal of Multivariate Analysis **54** (1995), no. 2, 295–309.

DEPARTMENT OF PHYSICS, ÉCOLE NORMALE SUPÉRIEURE, UNIVERSITÉ PSL, PARIS, FRANCE
Email address: samuil.petkov@phys.ens.psl.eu