# Phase Transition of Eigenvalues of Covariances from the Spiked Mixture Model in High-dimensional Regimes

Paul-Louis Delacour, Raf Van de Plas, *Member, IEEE,*

**Abstract**

The spiked mixture model (SMM) has been introduced as a probabilistic model that generalizes the single-spike (Wishart) model to a mixture model form. With applications ranging from imaging mass spectrometry in the life sciences to hyperspectral imaging in computer vision, it is crucial to understand under which circumstances its signals can be recovered from noisy measurements. The highly multiplexed nature of these measurement types furthermore necessitates such analysis to hold in high-dimensional settings. In this paper, we prove that the extreme eigenvalues of the covariance matrix from the SMM exhibit a phase transition in high-dimensional regimes. We show that this phase transition, and thus signal recovery by extreme eigenvalues, depends on several interacting factors: the correlation between spikes (*i.e.*, how similar in content underlying signals are), the energy parameters (*i.e.*, the absolute strength of each underlying signal), and the mixture probabilities (*i.e.*, how likely it is to encounter each underlying signal). This work provides sharp information-theoretic bounds on the parameters needed to detect one or more spikes from extreme eigenvalues of the SMM covariance matrix, and these guarantees could potentially impact any application of the SMM. Understanding this interplay could serve as a tool for driving experimental design in analytical chemistry and life sciences.

**Index Terms**

Random matrix theory, spiked covariance model, spiked mixture model, phase transition, Marchenko–Pastur law, BBP transition, high-dimensional statistics, spectral estimation, signal detection.

## I. Introduction

Many high-dimensional systems and measurement types in, *e.g.*, neural networks [1], wireless communications systems [2], functional genomics [3], biomedical imaging [4], and finance [5] can be modeled as large random matrices whose spectral properties (eigenvalues and eigenvectors) encode essential information. As a result, tools from random matrix theory (RMT) are increasingly used to analyze such systems and data. For example, recent work used these tools to analyze weight matrices or Jacobians in deep neural networks in order to understand generalization and structure [6].

As technological advances deliver increasingly high-dimensional measurement types, *e.g.*, in fields such as spatial transcriptomics and imaging mass spectrometry (IMS) [7], the need to exploit RMT in a high-dimensional regime becomes essential. Therefore, in this work we focus specifically on random matrices whose dimensions grow without bound. A foundational result in this area is that for many classical ensembles (such as Wigner matrices), the empirical distribution of eigenvalues converges (in the weak sense) to a deterministic limit called the semicircle law [8]. Moreover, for sample covariance matrices, one similarly obtains in the limit the Marčhenko–Pastur (MP) law [9].

One particularly interesting phenomenon arises when a signal (*i.e.*, a finite low-rank deterministic component) is embedded in noise (*i.e.*, a noisy high-dimensional matrix). This scenario is relevant to questions of signal detectability, for example in the context of machine learning [10], and is known as the single-spike model. Above certain thresholds, the signal or 'spike' can be detected despite its noisy environment, through eigenvalues that escape from the bulk spectrum. This phenomenon is known as the Baik-Ben Arous-Péché (BBP) phase transition [11]. Subsequent work has further extended this idea to finitely many spikes [12], community detection models [13], and beyond [14, 15].

The spiked mixture model (SMM) [16] is a statistical model that generalizes the spiked Wishart model to a mixture model form. Equivalently, the covariance matrix from the SMM model can be viewed as a sum of covariance matrices from single-spiked Wishart models of random size. In this paper, we investigate the limiting spectral distribution of SMM covariance matrices. Our analysis highlights how the correlation between spikes (*i.e.*, how similar the underlying signals are), the energy parameters (*i.e.*, the absolute strength of each underlying signal), and the mixture probabilities (*i.e.*, how likely it is to encounter each underlying signal) govern the emergence of eigenvalues detaching from the bulk, a phenomenon often referred to as "eigenvalue push-out". Such a result provides sharp information-theoretic bounds on the parameters needed to detect one or

Paul-Louis Delacour is with the Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: p.l.delacour@tudelft.nl).

Raf Van de Plas is with the Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands, and with the Department of Biochemistry, Vanderbilt University, Nashville, TN 37232, U.S.A., and with the Mass Spectrometry Research Center, Vanderbilt University School of Medicine, Nashville, TN 37232, U.S.A. (e-mail: raf.vandeplas@tudelft.nl).

more spikes from the extreme eigenvalues of the covariance matrix. These guarantees potentially impact any application where the SMM can be used, including ones in IMS and hyperspectral imaging (HSI) [16].

Section II introduces the SMM and puts forward the SMM phase transition theorem. Furthermore, it includes practical demonstrations of the theorem, illustrating under which circumstances signals embedded in the noise are detectable from extreme eigenvalues. In section III, we provide a proof for the SMM phase transition theorem, with supporting argumentation provided in supplementary sections A and B.

### A. Related Works

Asymptotic spectra of sample covariance matrices have been studied extensively. In the linear setting $\mathbf{y} = \mathbf{\Sigma}^{1/2}\mathbf{z}$ with $\mathbf{z}$ having independent and identically distributed (i.i.d.) features, and $\mathbf{\Sigma}$ the associated covariance matrix, the limiting spectrum follows a (generalized) Marčhenko–Pastur law, and a Baik-Ben Arous-Péché phase transition governs the outliers: an isolated eigenvalue detaches from the bulk precisely when a population spike exceeds a critical threshold [11, 17, 18]. In [19], this picture was recently extended beyond settings where the dependence between features is encoded linearly through a covariance matrix $\mathbf{\Sigma}$. An arbitrary, possibly nonlinear dependence between features is allowed, requiring only that the quadratic forms $\mathbf{y}^T\mathbf{A}\mathbf{y}$ concentrate around their expectation uniformly over square matrices $\mathbf{A}$.

In this work, we derive the phase transition specific to the SMM using a low-rank perturbation approach akin to Benaych-Georges and Nadakuditi [20]. Conditional on the selected subpopulation spike within the mixture model, the SMM's covariance matrix is a finite-rank perturbation of a white Wishart matrix. Together with the bi-orthogonal invariance of Gaussian noise, this facilitates that $K$ spikes are captured jointly by a single small matrix whose singularity locates all outliers at once, and which we reduce to the $K \times K$ matrix $\mathbf{L}$. This yields a couple of benefits over the deterministic-equivalent analysis of [19]. First, the argument is lighter: rotational invariance reduces the problem to the $T$-transform of the limiting MP bulk, which is explicit in our isotropic-noise setting and which we obtain by Gaussian concentration. This avoids the resolvent local laws and fluctuation-averaging machinery required by the coordinate-general analysis in [19]. Second, treating the full rank-$K$ perturbation jointly lets us go beyond the distinct-spike regime of [19], whose results assume simple population spikes. Our approach covers repeated values of $\lambda_i(\mathbf{L})$ by a perturbation argument and recovers each outlier at its correct multiplicity. The ability to handle repeating eigenvalues is important since such degeneracies are intrinsic to mixture models, where symmetric configurations of the subpopulations (*e.g.*, balanced, equally spaced spikes) commonly produce repeated eigenvalues. Finally, our approach generalizes to regimes that do not satisfy the quadratic concentration assumption of [19]. As a demonstration, in appendix C, we show how our approach extends beyond scenarios with fixed signal strength spikes and can, in fact, handle a rare and intense regime in which the spike's signal strength diverges while the corresponding subpopulation probability vanishes, *i.e.*, spikes of unbounded energy in a vanishing fraction of observations.

### B. Notation

- For $k \in \{1, \ldots, n\}$, $\mathbf{e}_k \in \mathbb{R}^n$ denotes the $k$th standard basis vector.
- For $\mathbf{S} \in \mathbb{M}_d(\mathbb{C})$ a self adjoint matrix, we denote its $d$ eigenvalues by
$$\lambda_1(\mathbf{S}) \geq \lambda_2(\mathbf{S}) \geq \cdots \geq \lambda_d(\mathbf{S}).$$
- When $d \to \infty$ and $i \geq 1$ is fixed, the extreme eigenvalues of a sequence $\mathbf{S}_d \in \mathbb{M}_d(\mathbb{C})$ refer to $\lambda_i(\mathbf{S}_d)$ (largest eigenvalues) or $\lambda_{d-i+1}(\mathbf{S}_d)$ (smallest eigenvalues).
- For a vector $\mathbf{v}$, $\|\mathbf{v}\|_2$ and $\|\mathbf{v}\|_\infty$ denote the Euclidean norm and the $\ell_\infty$ norm, respectively.
- For a matrix $\mathbf{A}$, $\|\mathbf{A}\|_{\mathrm{op}}$ and $\|\mathbf{A}\|_{\mathrm{F}}$ denote the operator (spectral) norm and the Frobenius norm, respectively.
- For $u \in \mathbb{C}$, $E$ is a closed subset of $\mathbb{R}$, we denote the distance $d(u, E) = \min_{x \in E} |u - x|$.

## II. The Spiked Mixture Model (SMM) and Its Phase Transition

The SMM, introduced in [16], is a generalization of the single-spike model to a mixture model form. We consider $n$ independent observations $\mathbf{y}_1, \ldots, \mathbf{y}_n \in \mathbb{R}^d$, each sampled from the SMM:

$$\mathbf{y} = \begin{cases} \alpha\sqrt{\beta_1}\mathbf{v}_1 + \boldsymbol{\varepsilon} & \text{with probability } \pi_1 \\ \vdots & \\ \alpha\sqrt{\beta_K}\mathbf{v}_K + \boldsymbol{\varepsilon} & \text{with probability } \pi_K \end{cases}, \tag{1}$$

$$\alpha \sim \mathcal{N}(0, 1),\ \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),\ \sum_{k=1}^{K} \pi_k = 1,$$

$$\beta_1, \ldots, \beta_K \in \mathbb{R}^+,\ \mathbf{v}_1, \ldots, \mathbf{v}_K \in \mathbb{R}^d, \|\mathbf{v}_k\|_2 = 1,$$

where $\alpha$ is the random scaling factor of observation $\mathbf{y}$, $\sqrt{\beta_k}\mathbf{v}_k$ is the $k$-th subpopulation or spike with $\mathbf{v}_k$ as the normalized spike signal and $\sqrt{\beta_k}$ reporting the strength of the signal, $\boldsymbol{\varepsilon}$ is the random noise of observation $\mathbf{y}$, and $\pi_k$ is the probability of the $k$-th subpopulation. Let $z \in \{1, \ldots, K\}$ be a latent categorical variable indicating which of the spikes $\sqrt{\beta_1}\mathbf{v}_1, \ldots, \sqrt{\beta_K}\mathbf{v}_K$ was used to generate $\mathbf{y}$.

We will analyze the regime in which $d \to \infty, n \to \infty$, $d/n \to \gamma$, and $K$ remains fixed, hereafter referred to as the high-dimensional regime. In the following, we consider $\{\beta_k\}_{k=1}^K$ to be fixed, and further extend to diverging scenarios in appendix C. This paper adopts the Bayesian viewpoint in which the spikes are drawn from an arbitrary prior, but they are required to almost surely tend to a fixed correlation in the limit $d \to \infty$:

$$\mathbf{v}_l \cdot \mathbf{v}_k \xrightarrow[d\to\infty]{a.s.} \theta_{l,k} \in [0,1], \qquad\qquad \forall l, k \in [K]. \tag{2}$$

A simple example of a distribution satisfying (2) in the case of $K = 2$ is to sample $\mathbf{v}_1, \mathbf{v}_2$ as follows:

$$\begin{cases} \mathbf{v}_1 = \mathbf{u}_1/\|\mathbf{u}_1\|_2 \\ \mathbf{v}_2 = \dfrac{\theta\mathbf{u}_1 + \sqrt{1-\theta^2}\,\mathbf{u}_2}{\|\theta\mathbf{u}_1 + \sqrt{1-\theta^2}\,\mathbf{u}_2\|_2} \end{cases}, \tag{3}$$

for $\mathbf{u}_1, \mathbf{u}_2 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ and $\theta \in [0,1]$. The correlation then corresponds to

$$\mathbf{v}_1 \cdot \mathbf{v}_2 = \frac{\theta\|\mathbf{u}_1\|_2^2 + \sqrt{1-\theta^2}\,\mathbf{u}_1 \cdot \mathbf{u}_2}{\|\mathbf{u}_1\|_2\,\|\theta\mathbf{u}_1 + \sqrt{1-\theta^2}\mathbf{u}_2\|_2},$$

and, by using Gaussian concentration of measure, one gets $\mathbf{v}_1 \cdot \mathbf{v}_2 \xrightarrow[d\to\infty]{a.s.} \theta$. Furthermore, we consider the covariance matrices coming from SMM model (1):

$$\mathbf{S}_{d,n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{y}_i\mathbf{y}_i^\mathsf{T} \in \mathbb{R}^{d\times d}, \tag{4}$$

along with their ordered eigenvalues

$$\lambda_1(\mathbf{S}_{d,n}) \geq \ldots \geq \lambda_d(\mathbf{S}_{d,n}) \geq 0.$$

After introducing the SMM model and establishing the high-dimensional regime setting, we posit a corresponding phase transition theorem.

**Theorem II.1** (SMM phase transition)**.** *The extreme eigenvalues of $\mathbf{S}_{d,n}$ (4) exhibit the following behavior as $n \to \infty$ and $d \to \infty$ with $d/n \to \gamma$. Consider $\mathbf{L} \in \mathbb{R}^{K\times K}$ the Gram limiting matrix with entries $\mathbf{L}_{l,m} = \theta_{l,m}\sqrt{\pi_l\pi_m\beta_l\beta_m}$.*

- *For $1 \leq i \leq K$,*

$$\lambda_i(\mathbf{S}_{d,n}) \xrightarrow{a.s.} \begin{cases} T^{-1}(1/\lambda_i(\mathbf{L})) & \textit{if } \lambda_i(\mathbf{L}) > \sqrt{\gamma} \\ (1+\sqrt{\gamma})^2 & \textit{otherwise}, \end{cases}; \textit{ and}$$

- *for $i > K$,*

$$\lambda_i(\mathbf{S}_{d,n}) \xrightarrow{a.s.} (1+\sqrt{\gamma})^2.$$

*Here,*

$$T(z) = \int \frac{t}{z-t} d\mu_{\mathrm{MP}(\gamma)}(t) \qquad\qquad \textit{for } z \in \mathbb{C}\setminus[a,b],$$

*is the $T$-transform of $\mu$, and $\mu_{\mathrm{MP}(\gamma)}$ is the Marčhenko–Pastur distribution with parameter $\gamma > 0$. Its density is*

$$d\mu_{\mathrm{MP}(\gamma)}(x) = \frac{1}{2\pi\gamma x}\sqrt{(b-x)(x-a)}\mathbb{1}_{[a,b]}(x)dx + \max\left(0, 1-\frac{1}{\gamma}\right)\delta_0,$$

*where $\mathbb{1}_{[a,b]}$ is the indicator function, $a := (1-\sqrt{\gamma})^2$ and $b := (1+\sqrt{\gamma})^2$ denote the edges of the bulk of the distribution, and $\delta_0$ is the Dirac delta at location $0$. The inverse $T$-transform associated with $\mu_{\mathrm{MP}(\gamma)}$ is*

$$T^{-1}(l) = (1+1/l)(1+\gamma l).$$

*Note that for $l \in \mathbb{R}^+$ we indeed have $T^{-1}(1/l) \geq b$ with equality only for $l = \sqrt{\gamma}$.*

Importantly, the single-spike case of Theorem II.1 (equivalent to $K = 1$) is known as the BBP phase transition [11] and is well-studied. Here, we extend the characterization of the phase transition to a mixture model with $K \geq 1$ spikes, in which the underlying spikes almost surely tend to some correlation.

Also note that the Gram limiting matrix $\mathbf{L}$ in Theorem II.1 is positive semi-definite given that the matrix of correlations $\{\theta_{l,m}\}_{1\leq l,m\leq m}$ is the limit of Gram matrices.

Before addressing the proof of Theorem II.1, we illustrate this result empirically. We conducted a set of synthetic experiments in which datasets were constructed from the SMM model in (1) with $n = 2000$, $d = 1000$, $K = 2$, and $\gamma = 0.5$, and for which we sampled the $\mathbf{v}_1, \mathbf{v}_2$ spikes according to (3) and computed the covariance matrix (4). In Figure 1a, we plot the eigenvalue distribution of the sample covariance matrix of a particular dataset. In this example dataset, the signal strengths of the underlying spikes were set to $\beta_1 = 4$ and $\beta_2 = 5$, the probabilities of encountering these spikes were not equal, namely $\pi_1 = 0.4$ and $\pi_2 = 0.6$, and the correlation between the spikes' signal content was $\theta = 0.4$. For these particular values of the hyperparameters, we observe two eigenvalues successfuly popping out of the bulk of the Marčhenko–Pastur distribution. In Figure 1b, we demonstrate that not all hyperparameter value combinations lead to two escaping eigenvalues, but only a subset of them. This highlights the importance of understanding the interplay between the different SMM hyperparameters to determine under which circumstances spectral methods can detect the presence of underlying signals, and under which circumstances signals become effectively unrecoverable by spectral methods. This information will be crucial for effective experimental design in noisy environments and applications.

## III. Decomposition of SMM covariance matrices & SMM phase transition proof

In section III-A, we start by decomposing the covariance matrices that the SMM can generate, and then lay out a proof strategy for its phase transition. In sections III-B to III-E, we prove the necessary facts.

### A. Prelude (decomposition of the covariance matrices)

Let $z_i \in \{1, \ldots, K\}$ be a latent categorical variable indicating which component of the mixture (*i.e.*, spike) was used to generate the observation $\mathbf{y}_i$. Thus, $\mathbf{y}_i | (z_i = k) \sim \mathcal{N}(0, \beta_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T} + \mathbf{I}_d)$. With $\mathbf{x}_i \sim \mathcal{N}(0, \mathbf{I}_d)$, the covariance matrix from the SMM (1) can be expanded as:

$$\begin{aligned}\mathbf{S}_{d,n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{y}_i \mathbf{y}_i^\mathsf{T} &= \frac{1}{n}\sum_{k=1}^{K} (\beta_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T} + \mathbf{I}_d)^{1/2} \left( \sum_{i, z_i = k} \mathbf{x}_i \mathbf{x}_i^\mathsf{T} \right) (\beta_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T} + \mathbf{I}_d)^{1/2} \\ &= \sum_{k=1}^{K} \frac{n_k}{n} (\beta_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T} + \mathbf{I}_d)^{1/2} \mathbf{Z}_k (\beta_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T} + \mathbf{I}_d)^{1/2},\end{aligned} \tag{5}$$

where $n_k = |\{i \in [n] \text{ s.t. } z_i = k\}|$. Conditioned on $n_k$, $n_k \mathbf{Z}_k = \sum_{i, z_i = k} \mathbf{x}_i \mathbf{x}_i^\mathsf{T}$ follows a Wishart distribution $\mathcal{W}_d(n_k, \mathbf{I}_d)$. By expanding the square root, we find:

$$\begin{aligned}\mathbf{S}_{d,n} &= \sum_{k=1}^{K} \frac{n_k}{n} \left( \mathbf{I}_d + (\sqrt{\beta_k + 1} - 1) \mathbf{v}_k \mathbf{v}_k^\mathsf{T} \right) \mathbf{Z}_k \left( \mathbf{I}_d + (\sqrt{\beta_k + 1} - 1) \mathbf{v}_k \mathbf{v}_k^\mathsf{T} \right) \\ &= \mathbf{Z} + \sum_{k=1}^{K} \frac{n_k}{n} \left[ c_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T} \mathbf{Z}_k + c_k \mathbf{Z}_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T} + c_k^2 (\mathbf{v}_k^\mathsf{T} \mathbf{Z}_k \mathbf{v}_k) \mathbf{v}_k \mathbf{v}_k^\mathsf{T} \right] \\ &= \mathbf{Z} + \sum_{k=1}^{K} \frac{n_k}{n} \left[ \mathbf{v}_k \mathbf{v}_k^\mathsf{T} \mathbf{H}_k + \mathbf{H}_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T} \right],\end{aligned}$$

with $n\mathbf{Z} = \sum_{k=1}^{K} n_k \mathbf{Z}_k = \sum_{i=1}^{n} \mathbf{z}_i \mathbf{z}_i^\mathsf{T} \sim \mathcal{W}_d(n, \mathbf{I}_d)$, $c_k = \sqrt{\beta_k + 1} - 1$, and $\mathbf{H}_k = (c_k \mathbf{Z}_k + \frac{c_k^2}{2} (\mathbf{v}_k^\mathsf{T} \mathbf{Z}_k \mathbf{v}_k) \mathbf{I}_d) \in \mathbb{R}^{d \times d}$. If we define

$$\mathbf{P}_1 = \left[ \sqrt{\tfrac{n_1}{n}} \mathbf{H}_1 \mathbf{v}_1 \quad \sqrt{\tfrac{n_1}{n}} \mathbf{v}_1 \quad \ldots \quad \sqrt{\tfrac{n_K}{n}} \mathbf{H}_K \mathbf{v}_K \quad \sqrt{\tfrac{n_K}{n}} \mathbf{v}_K \right] \in \mathbb{R}^{d \times 2K}, \tag{6}$$

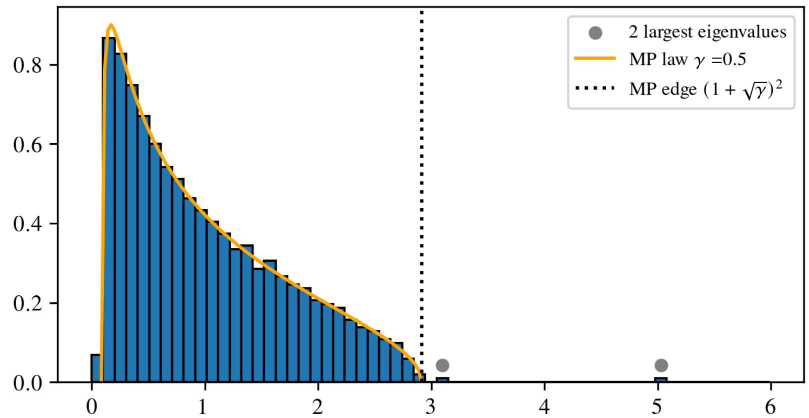
$$\mathbf{P}_2 = \left[ \sqrt{\tfrac{n_1}{n}} \mathbf{v}_1 \quad \sqrt{\tfrac{n_1}{n}} \mathbf{H}_1 \mathbf{v}_1 \quad \ldots \quad \sqrt{\tfrac{n_K}{n}} \mathbf{v}_K \quad \sqrt{\tfrac{n_K}{n}} \mathbf{H}_K \mathbf{v}_K \right] \in \mathbb{R}^{d \times 2K}, \tag{7}$$

we can decompose the covariance matrix $\mathbf{S}_{d,n}$ as follows:

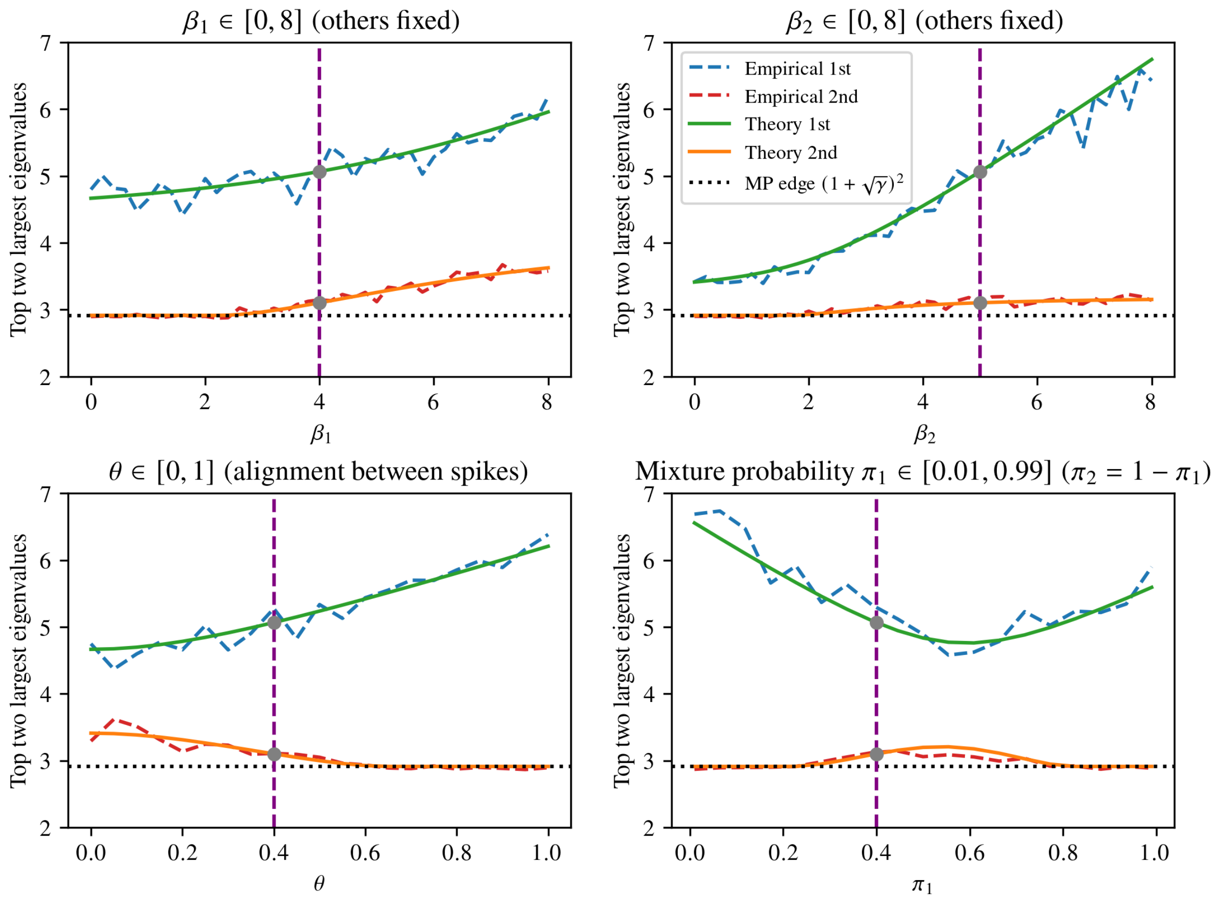
$$\mathbf{S}_{d,n} = \mathbf{Z} + \mathbf{P}_1 \mathbf{P}_2^\mathsf{T}. \tag{8}$$

For a proof of SMM phase transition, we first derive the extreme eigenvalues of (8) in a manner similar to [20]. Subsequently, we show that one can replace the $2K \times 2K$ limiting matrix by a simpler $K \times K$ matrix as described in Theorem II.1. The proof relies on the following facts:

1) Extreme eigenvalues after $K$ tend to the edge of the bulk of the MP distribution;
2) Extreme eigenvalues that do not escape the bulk of the MP distribution must tend to the edge of the bulk;
3) Similar to [20], we express extreme eigenvalues of $\mathbf{S}_{d,n}$ outside the bulk as $z$'s such that a $2K \times 2K$ matrix $\mathbf{M}_{d,n}(z)$ is singular;
4) The matrix $\mathbf{M}_{d,n}(z)$ converges almost surely to a matrix $\mathbf{M}(z)$;




(a) Example of two eigenvalues escaping the bulk of the Marčhenko–Pastur (MP) distribution. In this example dataset with $n = 2000$, $d = 1000$, $K = 2$, and $\gamma = 0.5$, the spike signal strengths were set to $\beta_1 = 4$ and $\beta_2 = 5$, the spike probabilities to $\pi_1 = 0.4$ and $\pi_2 = 0.6$, and the spike correlation to $\theta = 0.4$. This plot shows both the theoretical (orange line) and the empirical (blue bars) Marčhenko–Pastur distributions. For these particular hyperparameter values, we observe two eigenvalues (gray circles) successfully popping out of the bulk of the Marčhenko–Pastur distribution, reporting detectability.




(b) Effect of changing one of the hyperparameters. We show how varying a single hyperparameter, while keeping all others the same as in panel 1a impact the two largest eigenvalues. We explore varying the signal strength of the first spike, $\beta_1$ (top-left), the signal strength of the second spike, $\beta_2$ (top-right), the correlation between the spikes, $\theta$ (bottom-left), and the spike probabilities, $\pi_1$ and $\pi_2$ (bottom-right). In all plots, the situation of panel 1a is shown as a vertical purple line. We see that only a subset of all hyperparameter value combinations leads to two eigenvalues escaping the bulk of the Marčhenko–Pastur distribution.

Fig. 1: Empirical demonstrations of the SMM phase transition. Panel 1a shows one particular scenario, where two eigenvalues successfully escape the bulk of the Marčhenko–Pastur distribution, effectively reporting the spikes as recoverable from the noise. Panel 1b shows variations on the scenario of panel 1a, demonstrating how the interplay between the different SMM hyperparameters can result in either detectable or unrecoverable spikes.

5) The $z$'s in Fact 4 such that $\mathbf{M}(z)$ is singular are also the $z$'s such that the matrix $T(z)\mathbf{L}$ has an eigenvalue of 1, for a $K \times K$ real matrix $\mathbf{L}$; and
6) The $z$'s such that $\mathbf{M}_{d,n}(z)$ is singular converge to the $z$'s such that $\mathbf{M}(z)$ is singular (continuity lemma, adapted from [20]).

Fact 1 is proven in section III-B. In section III-C, we prove Facts 2 and 4, and recall Fact 3 from [20]. Facts 5 and 6 are proven in section III-D when $\mathbf{L}$ has a simple spectrum. In section III-E, we extend this result to the case where $\mathbf{L}$ may have equal eigenvalues.

### *B. Limits of the extreme eigenvalues $\lambda_i(\mathbf{S}_{d,n})$ for $i > K$*

In this section, we show that the extreme eigenvalues of $\mathbf{S}_{d,n}$ after $K$ tend almost surely to the edge of the bulk of the MP distribution. The proof relies on the following two lemmas.

**Lemma III.1.** *For matrices $\mathbf{P}_1$ and $\mathbf{P}_2$, defined in (6) and (7), the matrix $\mathbf{P} = \mathbf{P}_1\mathbf{P}_2^\mathsf{T}$ has at most $K$ non-negative eigenvalues.*

*Proof.* Reordering the columns of $\mathbf{P}_1$ and $\mathbf{P}_2$, we can decompose $\mathbf{P}$ as

$$\mathbf{P} = \mathbf{X}\mathbf{J}\mathbf{X}^\mathsf{T}, \quad \text{with}$$

$$\begin{aligned}\mathbf{X} &= \begin{bmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_K & \mathbf{A}\mathbf{v}_1 & \dots & \mathbf{A}\mathbf{v}_K\end{bmatrix} \in \mathbb{R}^{d\times 2K}, \quad \text{and}\\ \mathbf{J} &= \begin{bmatrix}0 & \mathbf{I}_K\\ \mathbf{I}_K & 0\end{bmatrix} \in \mathbb{R}^{2K\times 2K}.\end{aligned}$$

Let $W \subseteq \mathbb{R}^d$ be a subspace for which $\mathbf{y}^\mathsf{T}\mathbf{P}\mathbf{y} > 0$ for every non-zero $\mathbf{y}$. We start by proving that the mapping $T(\mathbf{y}) = \mathbf{X}^\mathsf{T}\mathbf{y}$ is injective on $W$. For $\mathbf{y}_0 \in W$ such that $T(\mathbf{y}_0) = 0$, we have

$$0 = \mathbf{y}_0^\mathsf{T}\mathbf{X}\mathbf{J}\mathbf{X}^\mathsf{T}\mathbf{y}_0 = \mathbf{y}_0^\mathsf{T}\mathbf{P}\mathbf{y}_0.$$

The definition of $W$ implies that $\mathbf{y}_0 = 0$, proving that the linear mapping $T$ is injective on $W$. Denoting $T|_W$ as the image of $T$ restricted to $W$, this means that

$$\dim(W) = \dim(T|_W). \tag{9}$$

Furthermore, any $\mathbf{z} \in T|_W$ satisfies $\mathbf{z}^\mathsf{T}\mathbf{J}\mathbf{z} > 0$. The definition of $\mathbf{J}$ implies that $\dim(T|_W) \leq K$ and thus by (9) that $\dim(W) \leq K$. ◻

**Lemma III.2** (Bounded extreme eigenvalues after $K$)**.** *For $\mathbf{S}_{d,n} = \mathbf{Z} + \mathbf{P}_1\mathbf{P}_2^\mathsf{T}$, defined in (8) and assuming that $d \geq 2K$, we have:*

$$\lambda_i(\mathbf{Z}) \leq \lambda_i(\mathbf{S}_{d,n}) \leq \lambda_{i-K}(\mathbf{Z}), \qquad \forall i > K.$$

*Proof.* The proof is based on Weyl's inequality:

$$\begin{aligned}\lambda_i(\mathbf{S}_{d,n}) &\leq \lambda_{i-K}(\mathbf{Z}) + \lambda_{K+1}(\mathbf{P}_1\mathbf{P}_2^\mathsf{T})\\ &\leq \lambda_{i-K}(\mathbf{Z}) && \text{using Lemma } III.1, \text{and}\\ \lambda_i(\mathbf{Z}) &\leq \lambda_i(\mathbf{S}_{d,n}) + \lambda_1(-\mathbf{P}_1\mathbf{P}_2^\mathsf{T})\\ &\leq \lambda_i(\mathbf{S}_{d,n}) - \lambda_d(\mathbf{P}_1\mathbf{P}_2^\mathsf{T})\\ &= \lambda_i(\mathbf{S}_{d,n}) && \text{since } \operatorname{rank}(\mathbf{P}_1\mathbf{P}_2^\mathsf{T}) \leq 2K.\end{aligned}$$

◻

Since the infimum and supremum of the bulk of the Marčhenko–Pastur (MP) distribution $\mu_{\mathrm{MP}(\gamma)}$ correspond to $a = (1-\sqrt{\gamma})^2$ and $b = (1+\sqrt{\gamma})^2$, it follows from [20] (section 6.2.1) that for every $i \geq 1$:

$$\begin{aligned}\lambda_i(\mathbf{Z}) &\xrightarrow{\text{a.s.}} b,\\ \lambda_{d-i+1}(\mathbf{Z}) &\xrightarrow{\text{a.s.}} a.\end{aligned}$$

Combining this with Lemma III.2, we obtain for every $i > K$:

$$\begin{aligned}\lambda_i(\mathbf{S}_{d,n}) &\xrightarrow{\text{a.s.}} b,\\ \lambda_{d-i+1}(\mathbf{S}_{d,n}) &\xrightarrow{\text{a.s.}} a.\end{aligned}$$

In other words, all extreme eigenvalues of $\mathbf{S}_{d,n}$ beyond the first $K$ almost surely tend to the edge of the bulk of the MP distribution.

### C. Limits of the extreme eigenvalues $\lambda_i(\mathbf{S}_{d,n})$ for $i \leq K$

Using the lower bound from Lemma III.2, we know for $i \geq 1$ that

$$\lambda_i(\mathbf{S}_{d,n}) \geq \lambda_i(\mathbf{Z}). \tag{10}$$

As in [20], the convergence of extreme eigenvalues of $\mathbf{Z}$ together with (10) implies that

$$\liminf_{n\to\infty} \lambda_i(\mathbf{S}_{d,n}) \geq b.$$

This shows that extreme eigenvalues of $\mathbf{S}_{d,n}$ that do not escape the bulk of the MP distribution must necessarily converge to its edge $b$.

In the following, we are interested in extreme eigenvalues that lie outside the bulk of the MP distribution. Similar to [20], we will show that the extreme eigenvalues of $\mathbf{S}_{d,n}$ are the $z$'s such that a certain matrix $\mathbf{M}_{d,n}(z)$ is singular. It is known that the eigenvalues of the matrix $\mathbf{Z}$ weakly converge to the MP distribution. Any $z \notin \{\lambda_1(\mathbf{Z}), \dots, \lambda_n(\mathbf{Z})\}$ is an eigenvalue of $\mathbf{S}_{d,n}$ iff:

$$\begin{aligned}
0 &= \det(z\mathbf{I}_d - \mathbf{S}_{d,n}) \\
&= \det(z\mathbf{I}_d - \mathbf{Z})\det(\mathbf{I}_d - (z\mathbf{I}_d - \mathbf{Z})^{-1}\mathbf{P}_1\mathbf{P}_2^\mathsf{T}) && \text{since } z \notin \{\lambda_1(\mathbf{Z}), \dots, \lambda_n(\mathbf{Z})\}, \\
&= \det(\mathbf{I}_{2K} - \mathbf{P}_2^\mathsf{T}(z\mathbf{I}_d - \mathbf{Z})^{-1}\mathbf{P}_1) && \text{using Sylvester's determinant identity.}
\end{aligned}$$

This shows that any $z \notin \{\lambda_1(\mathbf{Z}), \dots, \lambda_n(\mathbf{Z})\}$ is an eigenvalue of $\mathbf{S}_{d,n}$ if and only if the $2K \times 2K$ matrix $\mathbf{M}_{d,n}(z) = \mathbf{I}_{2K} - \mathbf{P}_2^\mathsf{T}(z\mathbf{I}_d - \mathbf{Z})^{-1}\mathbf{P}_1$ is singular. Since we know that $\lambda_d(\mathbf{Z})$ and $\lambda_1(\mathbf{Z})$ converge to the edge of the bulk of the MP distribution, the eigenvalues outside the bulk are the $z$'s from the set:

$$\mathcal{K}(\eta) := \{z \in \mathbb{C}; d(z, [a,b]) \geq \eta\}, \qquad \forall \eta > 0. \tag{11}$$

Using Theorem A.2 together with Lemmas B.2 and B.3, we get that the matrix $\mathbf{M}_{d,n}(z)$ converges almost surely to a matrix $\mathbf{M}(z)$ as $n, d \to \infty$ such that $\frac{d}{n} \to \gamma$.

$$\begin{aligned}
\mathbf{M}_{d,n}(z)_{2l-1,2m-1} &= \mathbb{1}_{l=m} - \frac{\sqrt{n_l n_m}}{n}\mathbf{v}_l^\mathsf{T}(z\mathbf{I}_d - \mathbf{Z})^{-1}\mathbf{H}_m\mathbf{v}_m \\
&= \mathbb{1}_{l=m} - \frac{\sqrt{n_l n_m}}{n}\left[c_m\mathbf{v}_l^\mathsf{T}(z\mathbf{I}_d - \mathbf{Z})^{-1}\mathbf{Z}_m\mathbf{v}_m + \frac{c_m^2}{2}\mathbf{v}_l^\mathsf{T}\mathbf{Z}_m\mathbf{v}_l\mathbf{v}_l^\mathsf{T}(z\mathbf{I}_d - \mathbf{Z})^{-1}\mathbf{v}_m\right] \\
&\to \mathbb{1}_{l=m} - \sqrt{\pi_l\pi_m}\theta_{l,m}\left[c_m T(z) + \frac{c_m^2}{2}T(z)(1-\gamma G(z))\right] \\
&= \mathbb{1}_{l=m} - \sqrt{\pi_l\pi_m}\theta_{l,m}T(z)c_m\left[1 + \frac{c_m}{2}(1-\gamma G(z))\right], \\
\mathbf{M}_{d,n}(z)_{2l-1,2m} &= -\frac{\sqrt{n_l n_m}}{n}\mathbf{v}_l^\mathsf{T}(z\mathbf{I}_d - \mathbf{Z})^{-1}\mathbf{v}_m \\
&\to -\sqrt{\pi_l\pi_m}\theta_{l,m}G(z), \\
\mathbf{M}_{d,n}(z)_{2l,2m-1} &= -\frac{\sqrt{n_l n_m}}{n}\mathbf{v}_l^\mathsf{T}\mathbf{H}_l(z\mathbf{I}_d - \mathbf{Z})^{-1}\mathbf{H}_m\mathbf{v}_m \\
&= -\frac{\sqrt{n_l n_m}}{n}c_l c_m\mathbf{v}_l^\mathsf{T}\left(\mathbf{Z}_l + \frac{c_l}{2}\mathbf{v}_l^\mathsf{T}\mathbf{Z}_l\mathbf{v}_l\mathbf{I}_d\right)(z\mathbf{I}_d - \mathbf{Z})^{-1}\left(\mathbf{Z}_m + \frac{c_m}{2}\mathbf{v}_m^\mathsf{T}\mathbf{Z}_m\mathbf{v}_m\mathbf{I}_d\right)\mathbf{v}_m \\
&\to -\sqrt{\pi_l\pi_m}\theta_{l,m}c_l c_m\left[T(z)\left(\frac{\gamma}{\pi_l}\delta_{l=m} + \frac{1}{1-\gamma G(z)}\right) + \frac{c_l + c_m}{2}T(z) + \frac{c_l c_m}{4}G(z)\right] \\
&= -\sqrt{\pi_l\pi_m}\left[\theta_{l,m}c_l c_m\frac{T(z)\gamma}{\pi_l}\delta_{l=m} + \theta_{l,m}c_l c_m\frac{T^2(z)}{G(z)}\left(1 + \frac{c_l}{2}(1-\gamma G(z))\right)\left(1 + \frac{c_m}{2}(1-\gamma G(z))\right)\right]
\end{aligned}$$

using the relation $G(z) = T(z)(1-\gamma G(z))$ in the last step, and

$$\begin{aligned}
\mathbf{M}_{d,n}(z)_{2l,2m} &= \mathbb{1}_{l=m} - \frac{\sqrt{n_l n_m}}{n}\mathbf{v}_l^\mathsf{T}\mathbf{H}_l(z\mathbf{I}_d - \mathbf{Z})^{-1}\mathbf{v}_m \\
&= \mathbb{1}_{l=m} - \frac{\sqrt{n_l n_m}}{n}\left[c_l\mathbf{v}_l^\mathsf{T}\mathbf{Z}_l(z\mathbf{I}_d - \mathbf{Z})^{-1}\mathbf{v}_m + \frac{c_l^2}{2}\mathbf{v}_l^\mathsf{T}\mathbf{Z}_l\mathbf{v}_l\mathbf{v}_l^\mathsf{T}(z\mathbf{I}_d - \mathbf{Z})^{-1}\mathbf{v}_m\right] \\
&\to \mathbb{1}_{l=m} - \sqrt{\pi_l\pi_m}\theta_{l,m}\left[c_l T(z) + \frac{c_l^2}{2}T(z)(1-\gamma G(z))\right] \\
&= \mathbb{1}_{l=m} - \sqrt{\pi_l\pi_m}\theta_{l,m}c_l T(z)\left[1 + \frac{c_l}{2}(1-\gamma G(z))\right],
\end{aligned}$$

with $G(z)$ the Cauchy Transform of the Marčhenko–Pastur distribution:

$$G(z) = \int \frac{1}{z-t} \mathrm{d}\mu_{\mathrm{MP}(\gamma)}(t).$$

Using $a_i = c_i(1 + \frac{c_i}{2}(1 - \gamma G(z))$, we define the matrices $\mathbf{A}_i, \mathbf{B}_{i,j} \in \mathbb{R}^{2\times 2}$ and the vectors $\mathbf{c}_i, \mathbf{d}_i \in \mathbb{R}^2$ as follows:

$$\begin{aligned}
\mathbf{A}_i &:= \begin{bmatrix} a_i T(z) & G(z) \\ c_i^2 T(z)\gamma/\pi_i + a_i^2 \frac{T^2(z)}{G(z)} & a_i T(z) \end{bmatrix}, \\
\mathbf{B}_{i,j} &:= \begin{bmatrix} a_j T(z) & G(z) \\ a_i a_j T^2(z)/G(z) & a_i T(z) \end{bmatrix} = \mathbf{c}_i \mathbf{d}_j^\mathsf{T}, \\
\mathbf{c}_i &:= \begin{bmatrix} 1 \\ a_i T(z)/G(z) \end{bmatrix}, \mathbf{d}_i := \begin{bmatrix} a_i T(z) \\ G(z) \end{bmatrix}.
\end{aligned}$$

With this notation, we obtain that $\mathbf{M}_{d,n}(z) \xrightarrow[d,n\to\infty,\, \frac{d}{n}\to\gamma]{\text{a.s.}} \mathbf{M}(z)$, with

$$\mathbf{M}(z) := \mathbf{I}_{2K} - \begin{bmatrix} \pi_1 \mathbf{A}_1 & \sqrt{\pi_1\pi_2}\theta_{1,2}\mathbf{B}_{12} & \dots & \sqrt{\pi_1\pi_K}\theta_{1,K}\mathbf{B}_{1K} \\ \sqrt{\pi_1\pi_2}\theta_{2,1}\mathbf{B}_{21} & \pi_2\mathbf{A}_2 & \dots & \sqrt{\pi_2\pi_K}\theta_{2,K}\mathbf{B}_{2K} \\ \dots & \dots & & \\ \sqrt{\pi_1\pi_K}\theta_{K,1}\mathbf{B}_{K,1} & \dots & \dots & \mathbf{I} - \pi_K \mathbf{A}_K \end{bmatrix} := \mathbf{I}_{2K} - \mathbf{L}. \tag{12}$$

Moreover, the convergence of $\mathbf{M}_{d,n}$ is uniform on $\mathcal{K}(\eta)$ for all $\eta > 0$, (11).

### *D. Equivalence using determinant relation*

In this section, we want to find $z$ such that the matrix $\mathbf{M}(z)$ in (12) is singular. We start by defining the matrix $\tilde{\mathbf{M}}(z) = \mathbf{I}_{2K} - \tilde{\mathbf{L}}(z)$ by its entries:

$$[\tilde{\mathbf{M}}(z)]_{i,j} := (-1)^{i+j}[\mathbf{M}(z)]_{i,j}.$$

By design, the matrix $\tilde{\mathbf{M}}(z)$ has the same determinant as the matrix $\mathbf{M}(z)$ since it is obtained by multiplying all even rows and even columns by $-1$. We can rewrite the absolute value of the determinant as:

$$\begin{aligned}
|\det(\mathbf{M}(z))| &= \left|\det(\mathbf{M}(z)\tilde{\mathbf{M}}(z))\right|^{1/2} \\
&= \left|\det\left((\mathbf{I}_{2K} - \mathbf{L}(z))(\mathbf{I}_{2K} - \tilde{\mathbf{L}}(z))\right)\right|^{1/2} \\
&= \left|\det\left(\mathbf{I}_{2K} - \left(\mathbf{L}(z) + \tilde{\mathbf{L}}(z) - \mathbf{L}(z)\tilde{\mathbf{L}}(z)\right)\right)\right|^{1/2}.
\end{aligned} \tag{13}$$

Defining $\mathbf{W} := \mathbf{L}(z) + \tilde{\mathbf{L}}(z) - \mathbf{L}(z)\tilde{\mathbf{L}}(z)$ results through computation (and cancelation) in:

$$\begin{aligned}
\mathbf{W}_{[2i:2i+2,2i:2i+2]} &= \pi_i\beta_i T(z) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, & i \in [0, K-1], \\
\mathbf{W}_{[2i:2i+2,2j:2j+2]} &= \sqrt{\pi_i\pi_j}\theta_{i,j} \begin{bmatrix} \beta_j T(z) & 0 \\ T^2(z)(c_i^2 a_j - c_j^2 a_i) & \beta_i T(z) \end{bmatrix}, & i, j \in [0, K-1], i \neq j.
\end{aligned}$$

We finish by linking the matrix $\mathbf{W} \in \mathbb{C}^{2K\times 2K}$ to the matrix $\mathbf{L} \in \mathbb{R}^{K\times K}$ defined in Theorem (II.1). Let

$$\mathbf{R} := \mathrm{Diag}\left(\sqrt{\beta_1}, \sqrt{\beta_1}, \sqrt{\beta_2}, \sqrt{\beta_2}, \dots, \sqrt{\beta_K}, \sqrt{\beta_K}\right) \in \mathbb{R}^{2K\times 2K}$$

be a rescaling matrix. Furthermore, let $\mathbf{Q} \in \mathbb{R}^{2K\times 2K}$ be a permutation matrix such that multiplication with $\mathbf{Q}$ on the left moves all even rows to the first $K$ rows and multiplication with $\mathbf{Q}^\mathsf{T}$ on the right moves all even columns to the first $K$ columns. More explicitly, with the permutation $\tilde{\pi}$ defined as

$$\tilde{\pi}(i) = \begin{cases} 2i & \text{if } i \in [0, K-1] \\ 2(i-K)+1 & \text{if } i \in [K, 2K-1] \end{cases},$$

the permutation matrix $\mathbf{Q}$ can be defined as

$$\mathbf{Q}_{i,j} = \begin{cases} 1 & \text{if } j = \tilde{\pi}(i) \\ 0 & \text{otherwise} \end{cases},$$

or equivalently:

$$\mathbf{Q} = \begin{bmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 & 0 \\ & & & \dots & & & \\ 0 & 0 & 0 & 0 & \dots & 1 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 & 0 \\ & & & \dots & & & \\ 0 & 0 & 0 & 0 & \dots & 0 & 1 \end{bmatrix}.$$

The matrix $\mathbf{QRWR^{-1}Q^\mathsf{T}}$ is then block lower triangular with all diagonal blocks equal and corresponding to $T(z)\mathbf{L}$. Knowing this, equation (13) can then be simplified to:

$$\begin{aligned} |\det(\mathbf{M}(z))| &= |\det(\mathbf{I}_{2K} - \mathbf{W})|^{1/2} \\ &= \left|\det\left(\mathbf{I}_{2K} - \mathbf{QRWR^{-1}Q^\mathsf{T}}\right)\right|^{1/2} \\ &= \left|\det(\mathbf{I}_K - T(z)\mathbf{L})^2\right|^{1/2} \\ &= |\det(\mathbf{I}_K - T(z)\mathbf{L})| \\ &= \prod_{i=1}^{K} |1 - T(z)\lambda_i(\mathbf{L})|\,. \end{aligned} \tag{14}$$

The phase transition for the extreme eigenvalues emerges since the MP distribution is compactly supported on $[a, b]$, and thus $T(z)$ is defined outside $[a, b]$. Let

$$\begin{aligned} T(a^-) &:= \lim_{z \to a} T(z) = -1/\sqrt{\gamma}, \text{and} \\ T(b^+) &:= \lim_{z \to b} T(z) = 1/\sqrt{\gamma} \end{aligned}$$

with $T(z)$ being a homeomorphism from $(-\infty, a)$ to $(T(a^-), 0)$ and from $(b, \infty)$ to $(0, T(b^+))$. Since $\lambda_i(\mathbf{L}) \geq 0$, there exists a solution to $T(z) = \frac{1}{\lambda_i(\mathbf{L})}$ only when $\frac{1}{\lambda_i(\mathbf{L})} < T(b^+)$, which is equivalent to $\lambda_i(\mathbf{L}) > \sqrt{\gamma}$.

When the eigenvalues of the limiting matrix $\mathbf{L}$ are distinct, the conclusion of Theorem II.1 follows from an application of Lemma III.3. When the eigenvalues are not distinct, we show in section III-E that, using a small perturbation, the conclusions of Theorem II.1 remain valid.

**Lemma III.3** (Continuity lemma, adapted from Lemma 6.1 in [20])**.** *For $\lambda_1(\mathbf{L}) > \ldots > \lambda_r(\mathbf{L}) > \sqrt{\gamma}$ with $r \leq K$, there exists $r$ complex sequences, $z_1(d, n), \ldots, z_r(d, n)$, such that* $\mathrm{Re}(z_1(d, n)) \geq \ldots \geq \mathrm{Re}(z_r(d, n))$ *converge respectively to the eigenvalues outside the bulk of the MP distribution:*

$$T^{-1}(1/\lambda_1(\mathbf{L})), \ldots, T^{-1}(1/\lambda_r(\mathbf{L})).$$

*Proof.* We start by proving that the $z$'s we are looking for do not escape the bulk of the MP distribution at infinity and that they lie in a compact set. Since $T(z) \xrightarrow{|z| \to \infty} 0$, there exists a real scalar $R \geq (1 + \sqrt{\gamma})^2$ such that for every $z \in \mathbb{C}, |z| \geq R$:

$$|T(z)| \leq \frac{1}{2\lambda_1(\mathbf{L})}.$$

As such, for all $i \in [K], |z| \geq R$ we have:

$$|1 - T(z)\lambda_i(\mathbf{L})| \geq \frac{1}{2}.$$

Combining this with (14), we get for $|z| \geq R$ that

$$|\det\left(\mathbf{M}(z)\right)| \geq 2^{-K}. \tag{15}$$

Moreover, we have shown in section III-C that $\mathbf{M}_{d,n}(z)$ converges uniformly on $\mathcal{K}(\eta) = \{z \in \mathbb{C}; \quad d(z, [a, b]) \geq \eta\}$ for all $\eta > 0$. Together with (15), this uniform convergence implies that for a large enough $n$ and $d$, the $z$'s such that $\mathbf{M}_{d,n}(z)$ is singular are supported on a compact subset $\mathcal{K}_1 \subset \mathcal{K}(\eta)$. This ensures that $z$ does not run away to infinity.

For $l \neq r \in \mathcal{K}_1$, we denote $\mathrm{Card}_{l,r}$ as the number of $z$'s in $B(l,r)$, the complex ball of diameter $[l,r]$, such that $\det(\mathbf{M}(z)) = 0$. Similarly, $\mathrm{Card}_{l,r}(d,n)$ denotes the numbers of $z$'s such that $\det(\mathbf{M}_{d,n}(z)) = 0$. With $l, r \notin \{\lambda_1(\mathbf{L}), \ldots, \lambda_r(\mathbf{L})\}$ ,$\gamma$ is a circle of diameter $[l,r]$. We finish by showing that for every $l, r$, $\mathrm{Card}_{l,r}(d,n)$ converges almost surely to $\mathrm{Card}_{l,r}$:

$$\mathrm{Card}_{l,r} = \frac{1}{2\pi i}\int_\gamma \frac{\partial_z \det \mathbf{M}(z)}{\det \mathbf{M}(z)} dz = \lim_{\substack{n\to\infty,\\ d\to\infty,\\ \frac{d}{n}\to\gamma}} \frac{1}{2\pi i}\int_\gamma \frac{\partial_z \det \mathbf{M}_{d,n}(z)}{\det \mathbf{M}_{d,n}(z)} dz.$$

This shows that for $d$ and $n$ large enough, there exist $r$ distinct solutions to $\det\left(\mathbf{M}_{d,n}(z)\right) = 0$, denoted as $z_1(d,n), \ldots, z_r(d,n)$ and with $\mathrm{Re}(z_1(d,n)) \geq \ldots \geq \mathrm{Re}(z_r(d,n))$. Furthermore, since we can choose $l$ and $r$ arbitrarily close to each other, we must have

$$z_i(d,n) \xrightarrow[\substack{n\to\infty,\\ d\to\infty,\\ \frac{d}{n}\to\gamma}]{\text{a.s.}} T^{-1}\left(\frac{1}{\lambda_i(\mathbf{L})}\right), \qquad \forall i \in [1,r].$$

$\square$

## E. Perturbation analysis

The purpose of this section is to extend Theorem II.1 to the case where $\mathbf{L}$ has repeated eigenvalues. We do so by introducing an arbitrarily small perturbation of the spikes $\mathbf{v}_1, \ldots, \mathbf{v}_K$ that yields a limiting matrix with a simple spectrum, by subsequently applying the results from section III-D, and finally by removing the perturbation by continuity.

Let $\mathbf{w}_1, \ldots, \mathbf{w}_K \sim \mathcal{N}(0, \frac{1}{d}\mathbf{I}_d)$ be $K$ independent vectors sampled from a multivariate Gaussian distribution. For arbitrary real scalars $t_1, \ldots, t_K \in \mathbb{R}$, we construct perturbed spike vectors $\tilde{\mathbf{v}}_i$ as follows:

$$\tilde{\mathbf{v}}_i := \frac{\mathbf{v}_i + t_i\mathbf{w}_i}{\|\mathbf{v}_i + t_i\mathbf{w}_i\|}, \qquad \forall i \in [K], \tag{16}$$

$$\tilde{\beta}_i := (1+t_i^2)\beta_i, \qquad \forall i \in [K]. \tag{17}$$

For $i \neq j$, we get

$$\tilde{\mathbf{v}}_i^\top \tilde{\mathbf{v}}_j \xrightarrow[d\to\infty]{\text{a.s.}} \frac{\theta_{i,j}}{\sqrt{(1+t_i^2)(1+t_j^2)}} := \tilde{\theta}_{i,j},$$

and

$$\mathbf{v}_i^\top \tilde{\mathbf{v}}_i \xrightarrow[d\to\infty]{\text{a.s.}} \frac{1}{\sqrt{1+t_i^2}}. \tag{18}$$

With $\tilde{\mathbf{D}} = \mathrm{Diag}(\pi_1\tilde{\beta}_1, \ldots, \pi_K\tilde{\beta}_K)$ and $\mathbf{t} = (t_1, \ldots, t_K)$, we define

$$\tilde{\mathbf{L}}(\mathbf{t}) := \tilde{\mathbf{D}}^{1/2}\tilde{\theta}\tilde{\mathbf{D}}^{1/2}.$$

The discriminant $\Delta(\mathbf{t})$ of the characteristic polynomial of $\tilde{\mathbf{L}}(\mathbf{t})$ is a real analytic function of the entries of $\tilde{\mathbf{L}}(\mathbf{t})$. As $\tilde{\mathbf{L}}(\mathbf{t})$ depends on $\mathbf{t}$ only in the diagonal, using the strengthened form of the Gershgorin circle theorem, we can choose $\mathbf{t}$ such that the Gershgorin discs do not intersect. Hence, $\tilde{L}(\mathbf{t})$ has a simple spectrum and $\Delta(\mathbf{t}) \neq 0$. Therefore, $\Delta$ is not identically zero. Since the zero set of a not identically zero, real analytic function has an empty interior, the set of perturbations yielding a simple spectrum is dense. Moreover, since $\tilde{\mathbf{L}}(\mathbf{t}) \to \mathbf{L}$ as $\|\mathbf{t}\|_\infty \to 0$, for every real scalar $\delta_1 > 0$, there exists an arbitrarily small $\mathbf{t}_0$ ($\|\mathbf{t}_0\|_\infty \leq \delta_2$, with $\delta_2$ a real scalar) such that $\tilde{\mathbf{L}}(\mathbf{t}_0)$ has a simple spectrum and

$$\|\mathbf{L} - \tilde{\mathbf{L}}(\mathbf{t}_0)\|_{\mathrm{op}} \leq \delta_1. \tag{19}$$

Let $\tilde{\mathbf{S}}_{d,n}$ be the covariance matrix associated with the perturbed vectors from (16) and (17) with $\mathbf{t}_0$. Furthermore, let

$$\rho(l) := \begin{cases} T^{-1}(1/l) & \text{if } l > \sqrt{\gamma} \\ (1+\sqrt{\gamma})^2 & \text{otherwise} \end{cases}.$$

Since $\tilde{\mathbf{L}}(\mathbf{t}_0)$ has distinct eigenvalues, the conclusion from section III-D shows that

$$\lambda_i(\tilde{\mathbf{S}}_{d,n}) \xrightarrow[\substack{n\to\infty,\\ d\to\infty,\\ \frac{d}{n}\to\gamma}]{\text{a.s.}} \tilde{\rho}_i := \rho\left(\lambda_i(\tilde{\mathbf{L}}(t_0))\right).$$

Using the triangle inequality, we get

$$|\lambda_i(\mathbf{S}_{d,n}) - \rho(\lambda_i(\mathbf{L}))| \leq \underbrace{\left|\lambda_i(\mathbf{S}_{d,n}) - \lambda_i(\tilde{\mathbf{S}}_{d,n})\right|}_{\text{(A)}} + \underbrace{\left|\lambda_i(\tilde{\mathbf{S}}_{d,n}) - \tilde{\rho}_i\right|}_{\text{(B)}} + \underbrace{\left|\tilde{\rho}_i - \rho(\lambda_i(\mathbf{L}))\right|}_{\text{(C)}}. \tag{20}$$

We show the following facts:

(C) is small by continuity of the function $\rho$;

(B) tends to zero by a previously established result in the case where the limiting matrix has distinct eigenvalues; and

(A) is small because the perturbation of the spikes is small. Hence, $\|\mathbf{S}_{d,n} - \tilde{\mathbf{S}}_{d,n}\|_{\text{op}}$ can be made arbitrarily small and Weyl's inequality applies.

In the following, we let the real scalar $\varepsilon > 0$ be arbitrary. Starting with the (C) term, using Weyl's inequality together with (19), we get that

$$|\lambda_i(\tilde{\mathbf{L}}(t_0)) - \lambda_i(\mathbf{L})| \leq \delta_1.$$

By selecting $\delta_1$ sufficiently small, we get by continuity of the function $\rho$ that

$$|\tilde{\rho}_i - \rho(\lambda_i(\mathbf{L}))| \leq \frac{\varepsilon}{3}. \tag{21}$$

For the (B) term, the results from section III-D imply that for a large enough $n$ and $d$, almost surely

$$|\lambda_i(\tilde{\mathbf{S}}_{d,n}) - \tilde{\rho}_i| \leq \frac{\varepsilon}{3}. \tag{22}$$

We finish by bounding the (A) term, using Weyl's inequality once again:

$$|\lambda_i(\mathbf{S}_{d,n}) - \lambda_i(\tilde{\mathbf{S}}_{d,n})| \leq \|\mathbf{S}_{d,n} - \tilde{\mathbf{S}}_{d,n}\|_{\text{op}}.$$

Combining this with the definitions of $\mathbf{S}_{d,n}$ and $\tilde{\mathbf{S}}_{d,n}$ in (5) and the triangle inequality, we obtain:

$$\begin{aligned}
&\left|\lambda_i(\mathbf{S}_{d,n}) - \lambda_i(\tilde{\mathbf{S}}_{d,n})\right| \\
&\leq \sum_{k=1}^{K} \frac{n_k}{n} \left\| \left(\mathbf{I}_d + \beta_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T}\right)^{1/2} \mathbf{Z}_k \left(\mathbf{I}_d + \beta_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T}\right)^{1/2} - \left(\mathbf{I}_d + \tilde{\beta}_k \tilde{\mathbf{v}}_k \tilde{\mathbf{v}}_k^\mathsf{T}\right)^{1/2} \mathbf{Z}_k \left(\mathbf{I}_d + \tilde{\beta}_k \tilde{\mathbf{v}}_k \tilde{\mathbf{v}}_k^\mathsf{T}\right)^{1/2} \right\|_{\text{op}} \\
&\leq \sum_{k=1}^{K} \frac{n_k}{n} \left\| \left(\mathbf{I}_d + \beta_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T}\right)^{1/2} - \left(\mathbf{I}_d + \tilde{\beta}_k \tilde{\mathbf{v}}_k \tilde{\mathbf{v}}_k^\mathsf{T}\right)^{1/2} \right\|_{\text{op}} \|\mathbf{Z}_k\|_{\text{op}} \left((1+\beta_k)^{1/2} + (1+\tilde{\beta}_k)^{1/2}\right).
\end{aligned}$$

We proceed by showing that $\left\| \left(\mathbf{I}_d + \beta_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T}\right)^{1/2} - \left(\mathbf{I}_d + \tilde{\beta}_k \tilde{\mathbf{v}}_k \tilde{\mathbf{v}}_k^\mathsf{T}\right)^{1/2} \right\|_{\text{op}}$ can be made arbitrarily small by an appropriate choice of $\delta_1$:

$$\begin{aligned}
&\left\| \left(\mathbf{I}_d + \beta_k \mathbf{v}_k \mathbf{v}_k^\mathsf{T}\right)^{1/2} - \left(\mathbf{I}_d + \tilde{\beta}_k \tilde{\mathbf{v}}_k \tilde{\mathbf{v}}_k^\mathsf{T}\right)^{1/2} \right\|_{\text{op}} \\
&= \left\| \left(\sqrt{\beta_k + 1} - 1\right) \mathbf{v}_k \mathbf{v}_k^\mathsf{T} - \left(\sqrt{\tilde{\beta}_k + 1} - 1\right) \tilde{\mathbf{v}}_k \tilde{\mathbf{v}}_k^\mathsf{T} \right\|_{\text{op}} \\
&\leq \left\| \left(\sqrt{\beta_k + 1} - 1\right) \mathbf{v}_k \mathbf{v}_k^\mathsf{T} - \left(\sqrt{\tilde{\beta}_k + 1} - 1\right) \mathbf{v}_k \mathbf{v}_k^\mathsf{T} \right\|_{\text{op}} + \left|\sqrt{\tilde{\beta}_k + 1} - 1\right| \|\mathbf{v}_k \mathbf{v}_k^\mathsf{T} - \tilde{\mathbf{v}}_k \tilde{\mathbf{v}}_k^\mathsf{T}\|_{\text{op}} \\
&\leq \left|\sqrt{\beta_k + 1} - \sqrt{\tilde{\beta}_k + 1}\right| + \left|\sqrt{\tilde{\beta}_k + 1} - 1\right| \sqrt{2} \sqrt{1 - (\mathbf{v}_k^\mathsf{T} \tilde{\mathbf{v}}_k)^2}.
\end{aligned}$$

Plugging in the definition of $\tilde{\beta}_k$ (17) and the asymptotic correlation of $\mathbf{v}_k^\mathsf{T} \tilde{\mathbf{v}}_k$ (18), for an appropriate choice of $\delta_2$, we get that for a large enough $n$ and $d$ almost surely

$$\left|\lambda_i(\mathbf{S}_{d,n}) - \lambda_i(\tilde{\mathbf{S}}_{d,n})\right| \leq \frac{\varepsilon}{3}. \tag{23}$$

Finally, using (21), (22), and (23) in (20), we get that for a large enough $n$ and $d$, almost surely

$$|\lambda_i(\mathbf{S}_{d,n}) - \rho(\lambda_i(\mathbf{L}))| \leq \varepsilon.$$

## IV. Conclusion

This work examines the recovery of signals from noisy measurements in high-dimensional regimes. In doing so, we extend beyond previous work on finitely many, low-rank perturbations of large random matrices. Our results generalize the phase transition behavior known for such perturbations, moving from a single-spike framework to a multi-spike mixture model setting. The latter facilitates the study of signal recovery in scenarios where multiple signals underlie a measurement set, which opens up a broad field of new applications. To our knowledge, this is the first study to characterize a multi-spike mixture scenario and to demonstrate the interaction of correlation, signal energy, and mixture probability in the phase transition. Consequently, our study builds on and extends prior work on the single-spike model's phase transition.

Specifically, our work examines the feasibility of signal recovery by extreme eigenvalues for measurements that can be modeled by a spiked mixture model (SMM). The SMM is not the first model to facilitate multiple spikes, with previous work, *e.g.*, on linear multi-spike models [12]. However, linear formulations do not always suit real-world measurements, particularly in technologies where the mixing of signals is not linear or deviates substantially from linear in certain scenarios. In such situations, the SMM can be essential: it accommodates multiple underlying signals yet does not require them to mix linearly, but instead selects the dominant signal per observation. Examples of such application areas can be found in our previous work on the SMM [16], where this model was successfully used in data types ranging from IMS in biomedicine to HSI in computer vision.

Our findings show that in this multi-spike mixture model setting, the phase transition, and thus signal recovery by extreme eigenvalues, depends on several interacting factors: the correlation between spikes (*i.e.*, how similar in content two underlying signals are), the energy parameter $\beta_k$ (*i.e.*, the absolute strength of each underlying signal), and the mixture probabilities (*i.e.*, how likely it is to encounter each underlying signal). Each of these parameters can individually impact the phase transition, and thus our ability to recover signals. However, the multi-spike setting becomes particularly interesting when we consider the hereto less-studied role of inter-spike-correlation and the interplay between the different parameters. Examples of this include strongly correlating spikes boosting mutual detectability despite making individual spike recognition harder, and the correlation between spikes substantially modulating the traditional role that signal strength plays in detectability. More broadly, our results highlight how the parameters of the mixture model interact and jointly shape the phase transition behaviour. Besides the methodological insights and the generalization of single-spike scenarios, we believe that understanding this interplay can become an essential tool for driving experimental design, *e.g.*, when developing physical IMS experiments in analytical chemistry and the life sciences.

## Acknowledgments

Research reported in this publication was supported by the National Institutes of Health (NIH)'s Common Fund, National Institute Of Diabetes And Digestive And Kidney Diseases (NIDDK), and the Office Of The Director (OD) under Award Numbers U54DK120058, U54DK134302, and U01DK133766 (R.V.), by NIH's Common Fund, National Eye Institute, and the Office Of The Director (OD) under Award Number U54EY032442 (R.V.), by NIH's National Institute Of Allergy And Infectious Diseases (NIAID) under Award Numbers R01AI138581 and R01AI145992 (R.V.), by NIH's National Institute On Aging (NIA) under Award Number R01AG078803 (R.V.), and by NIH's National Cancer Institute (NCI) under Award Number U01CA294527 (R.V.). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
The authors would like to thank Antoine Maillard (INRIA Paris & Département d'Informatique, École Normale Supérieure, Université PSL) for fruitful discussions and valuable suggestions.

# Appendix A
## Concentration bounds

**Theorem A.1** (Concentration of Lipschitz functions on the sphere [21] (section 5.1.2))**.** *Consider a random vector* $\mathbf{y} \sim \mathrm{Unif}(\mathbb{S}^{d-1})$ *and a Lipschitz function* $f\colon \mathbb{S}^{d-1} \to \mathbb{R}$*, with Lipschitz constant* $\|f\|_{\mathrm{Lip}}$*. For every* $t \geq 0$ *and* $d$ *large enough,*

$$\mathbb{P}\{|f(\mathbf{y}) - \mathbb{E}\, f(\mathbf{y})| \geq t\} \leq 2\exp\left(-\frac{cdt^2}{\|f\|_{\mathrm{Lip}}^2}\right),$$

*for an absolute positive constant* $c$*.*

**Theorem A.2** (Convergence orthogonal invariant bounded matrix)**.** *Consider an orthogonal invariant random matrix ensemble given by the probability measure* $\mathbb{P}_d$ *on sets* $\Omega_d$ *of* $d \times d$ *complex matrices. Let* $\mathbf{A} \in \mathbb{C}^{d\times d}$ *be a matrix sampled from such a model. Assume:*

(i) (operator norm bound in probability) *There exists a deterministic function* $C(d)$ *with*

$$C(d) \xrightarrow[d\to\infty]{} C < \infty,$$

*such that*

$$\sum_{d\geq 1} \mathbb{P}\left\{\|\mathbf{A}\|_{\mathrm{op}} \geq C(d)\right\} < \infty.$$

(ii) (convergence of the normalized trace) *The first moment converges almost surely:*

$$\frac{\mathrm{Tr}(\mathbf{A})}{d} \xrightarrow[d\to\infty]{\text{a.s.}} \mu.$$

*Then, for any pair of random unit vectors* $\mathbf{u}, \mathbf{v} \in \mathbb{S}^{d-1}$ *(independent of* $\mathbf{A}$*) such that*

$$\mathbf{u}^\mathsf{T}\mathbf{v} \xrightarrow[d\to\infty]{\text{a.s.}} \ell,$$

*we have*

$$\mathbf{u}^\mathsf{T}\mathbf{A}\mathbf{v} \xrightarrow[d\to\infty]{\text{a.s.}} \ell\,\mu.$$

*Proof.* Projecting $\mathbf{v}$ on $\mathbf{u}$, we get the decomposition

$$\mathbf{v} = \frac{(\mathbf{u}\cdot\mathbf{v})}{\|\mathbf{u}\|}\frac{\mathbf{u}}{\|\mathbf{u}\|} + \left(\mathbf{v} - \frac{(\mathbf{u}\cdot\mathbf{v})}{\|\mathbf{u}\|}\frac{\mathbf{u}}{\|\mathbf{u}\|}\right),$$

and thus $\mathbf{u}^\mathsf{T}\mathbf{A}\mathbf{v} = \mathbf{A}_1 + \mathbf{A}_2$ with

$$\begin{aligned}\mathbf{A}_1 &= (\mathbf{u}\cdot\mathbf{v})\mathbf{u}^\mathsf{T}\mathbf{A}\mathbf{u},\\ \mathbf{A}_2 &= \mathbf{u}^\mathsf{T}\mathbf{A}(\mathbf{v} - (\mathbf{u}\cdot\mathbf{v})\mathbf{u}).\end{aligned}$$

Since $\mathbf{A}$ is orthogonal invariant, $\mathbf{u}^\mathsf{T}\mathbf{A}\mathbf{u}$ has the same distribution as $\mathbf{y}^\mathsf{T}\mathbf{A}\mathbf{y}$ for any $\mathbf{y} \in \mathbb{S}^{d-1}$, which we denote as

$$\mathbf{A}_1 \sim (\mathbf{u}\cdot\mathbf{v})\mathbf{y}^\mathsf{T}\mathbf{A}\mathbf{y}.$$

In particular, we take $\mathbf{y} \sim \mathrm{Unif}(\mathbb{S}^{d-1})$. Similarly, $\mathbf{A}_2 \sim \sqrt{1-(\mathbf{u}\cdot\mathbf{v})^2}\mathbf{y}_1^\mathsf{T}\mathbf{A}\mathbf{y}_2$ for $\mathbf{y}_1 = (a_1, \ldots, a_d)$, $\mathbf{y}_2 = (b_1, \ldots b_d)$, the first two rows of a Haar distributed random orthogonal matrix. We then analyze the limits of $\mathbf{A}_1$ and $\mathbf{A}_2$ separately.

- **Step 1**: $\mathbf{u}^\mathsf{T}\mathbf{A}\mathbf{u}$ behaves like $g_1(\mathbf{y}) := \mathbf{y}^\mathsf{T}\mathbf{A}\mathbf{y}$ for $\mathbf{y} \sim \mathrm{Unif}(\mathbb{S}^{d-1})$. If we condition on $\mathbf{A}$, $g_1$ is Lipschitz on $\mathbb{S}^{d-1}$ since:

$$\begin{aligned}\nabla g_1(\mathbf{y}) &= (\mathbf{A} + \mathbf{A}^\mathsf{T})\mathbf{y}\\ \|\nabla g_1(\mathbf{y})\|_2 &\leq 2\|\mathbf{A}\|_{\mathrm{op}}.\end{aligned}$$

Furthermore, the conditional expectation gives

$$\mathbb{E}\left[g(\mathbf{y})|\mathbf{A}\right] = \frac{1}{d}\,\mathrm{Tr}(\mathbf{A}).$$

Applying Theorem A.1, we find the conditional probability bound:

$$\mathbb{P}\left\{\left|\mathbf{y}^\mathsf{T}\mathbf{A}\mathbf{y} - \frac{1}{d}\,\mathrm{Tr}(\mathbf{A})\right| \geq t\,|\mathbf{A}\right\} \leq 2\exp\left(-\frac{cdt^2}{4\|\mathbf{A}\|_{\mathrm{op}}^2}\right). \tag{24}$$

Let $D := \left\{\left|\mathbf{y}^\top \mathbf{A}\mathbf{y} - \frac{1}{d}\operatorname{Tr}(\mathbf{A})\right| \geq t\right\}$ and $B := \{\|\mathbf{A}\|_{\text{op}} \leq C(d)\}$, noting that $B$ is determined by $\mathbf{A}$, we get:

$$\mathbb{P}\{D|B\} = \frac{\mathbb{E}\{\mathbb{1}_D\mathbb{1}_B\}}{\mathbb{P}\{B\}} = \frac{\mathbb{E}_A\,\mathbb{E}_D\{\mathbb{1}_D\mathbb{1}_B|A\}}{\mathbb{P}\{B\}} = \frac{\mathbb{E}_A\,\mathbb{1}_B\mathbb{P}\{D|A\}}{\mathbb{P}\{B\}}.$$

Applying the bound (24) pointwise gives:

$$\mathbb{P}\left\{\left|\mathbf{y}^\top \mathbf{A}\mathbf{y} - \frac{1}{d}\operatorname{Tr}(\mathbf{A})\right| \geq t \mid \|\mathbf{A}\|_{\text{op}} \leq C(d)\right\} \leq 2\exp\left(-\frac{cdt^2}{4C(d)^2}\right).$$

The assumptions on $\mathbf{A}$ give:

$$\mathbb{P}\left\{\left|\mathbf{y}^\top \mathbf{A}\mathbf{y} - \frac{1}{d}\operatorname{Tr}(\mathbf{A})\right| \geq t\right\} \leq \exp\left(-\frac{c_2dt^2}{C(d)^2}\right) + \mathbb{P}\{\|\mathbf{A}\|_{\text{op}} > C(d)\}. \tag{25}$$

We continue by taking $t = t(d)$, a decreasing function, such that the exponential term in (25) is summable. For example with $t = d^{-1/4}$, for $d$ large enough we get

$$\mathbb{P}\left\{\left|\mathbf{y}^\top \mathbf{A}\mathbf{y} - \frac{1}{d}\operatorname{Tr}(\mathbf{A})\right| \geq d^{-1/4}\right\} = \exp\left(-c_3d^{1/2}\right) + \mathbb{P}\{\|\mathbf{A}\|_{\text{op}} > C(d)\}. \tag{26}$$

By assumption (i), the right hand side of (26) is summable in $d$. Therefore, the Borel-Cantelli lemma implies:

$$\left|\mathbf{y}^\top \mathbf{A}\mathbf{y} - \frac{1}{d}\operatorname{Tr}(\mathbf{A})\right| \xrightarrow[d\to\infty]{\text{a.s.}} 0.$$

Using the assumption of the almost sure convergence of the first moment, we get that:

$$\mathbf{y}^\top \mathbf{A}\mathbf{y} \xrightarrow[d\to\infty]{\text{a.s.}} \mu.$$

- **Step 2:** Analogous to step 1, we use the function $g_2(\mathbf{y}_1, \mathbf{y}_2) = \mathbf{y}_1^\top \mathbf{A}\mathbf{y}_2$, which is also Lipschitz on $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$ since:

$$\begin{aligned}&\nabla_{\mathbf{x}_1} g_2(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{A}\mathbf{x}_2,\ \nabla_{\mathbf{x}_2} g_2(\mathbf{x}_1, \mathbf{x}_2) = \mathbf{A}^\top\mathbf{x}_1\\ &\|\nabla g_2(\mathbf{x}_1, \mathbf{x}_2)\|_2^2 \leq \|\mathbf{A}^\top\mathbf{x}_1\|_2^2 + \|\mathbf{A}\mathbf{x}_2\|_2^2 \leq 2\|\mathbf{A}\|_{\text{op}}^2.\end{aligned}$$

Conditioned on $\mathbf{A}$, by symmetry $\mathbf{y}_1^\top \mathbf{A}\mathbf{y}_2$ and $-\mathbf{y}_1^\top \mathbf{A}\mathbf{y}_2$ give

$$\mathbb{E}\left[g_2(\mathbf{y}_1, \mathbf{y}_2)|\mathbf{A}\right] = 0,$$

and thus by Theorem A.1, we find

$$\mathbb{P}\left\{|\mathbf{y}_1^\top \mathbf{A}\mathbf{y}_2| \geq t\,|\mathbf{A}\right\} \leq 2\exp\left(-\frac{cdt^2}{\|\mathbf{A}\|_{\text{op}}^2}\right).$$

Similar to step 1, this allows us to conclude that

$$\mathbf{y}_1^\top \mathbf{A}\mathbf{y}_2 \xrightarrow[d\to\infty]{\text{a.s.}} 0.$$

□

## Appendix B
## Computation of limits

**Lemma B.1** (Herbst [22])**.** *Let $G$ be a Lipschitz function on $\mathbb{R}^m$, with Lipschitz constant $\|G\|_{Lip}$. Then, under the Gaussian measure $\gamma_m$, for all real scalars $\delta > 0$, we get the following concentration:*

$$\gamma_m\left(|G - \mathbb{E}[G]| \geq \delta\right) \leq 2\exp\left(-\frac{\delta^2}{2\|G\|_{Lip}^2}\right).$$

**Lemma B.2.** *Let $\mathbf{X} \in \mathbb{R}^{d\times n}$ have i.i.d. Gaussian entries $\mathcal{N}(0,1)$. Partition the columns as $X = \begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix}$, with $\mathbf{X}_1 \in \mathbb{R}^{d\times n_1}$, $\mathbf{X}_2 \in \mathbb{R}^{d\times(n-n_1)}$, and $n_1 \sim B(n, \pi_1)$ an independent random variable sampled from the Binomial distribution. Set*

$$\mathbf{S}_n := \frac{\mathbf{X}\mathbf{X}^\top}{n}, \qquad \mathbf{R}_n(z) := (z\mathbf{I} - \mathbf{S}_n)^{-1}, \text{ and} \qquad \mathbf{A}_n(z) := \mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\top}{n_1},$$

*where $z \in \mathbb{C} \setminus \{\lambda_1(\mathbf{S}_n), \ldots, \lambda_d(\mathbf{S}_n)\}$. Assume $d, n \to \infty$ with $d/n \to \gamma \in (0, \infty)$. Then,*

$$\frac{1}{d}Tr(\mathbf{A}_n(z)) \xrightarrow{a.s.} T(z),$$

*where $T(z) = \int_t \frac{t}{z-t} d\mu_{MP(\gamma)}(t)$ is the $T$-transform of the Marčhenko–Pastur law of parameter $\gamma$. Moreover, this convergence is uniform on $\mathcal{K}(\eta) = \{z \in \mathbb{C},\ d(z,[a,b]) > \eta\}$ for all real scalars $\eta > 0$.*

*Proof.* We define for the $j$th column $\mathbf{x}_j$ of $\mathbf{X}$:

$$a_j := \frac{1}{d}\mathbf{x}_j^\intercal \mathbf{R}_n(z)\mathbf{x}_j.$$

Then,

$$\frac{1}{d}\operatorname{Tr}\left(\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\intercal}{n_1}\right) = \frac{1}{n_1}\sum_{j=1}^{n_1} a_j, \qquad \text{and} \qquad \frac{1}{d}\operatorname{Tr}\left(\mathbf{R}_n(z)\frac{\mathbf{X}\mathbf{X}^\intercal}{n}\right) = \frac{1}{n}\sum_{j=1}^{n} a_j.$$

Since the columns of $\mathbf{X}$ are i.i.d., $\mathbb{E}[a_j]$ is the same for every $j$. Hence,

$$\mathbb{E}\left[\frac{1}{d}\operatorname{Tr}\left(\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\intercal}{n_1}\right)\right] = \mathbb{E}\left[\frac{1}{d}\operatorname{Tr}\left(\mathbf{R}_n(z)\frac{\mathbf{X}\mathbf{X}^\intercal}{n}\right)\right]. \tag{27}$$

The RHS of (27) corresponds to $\mathbb{E}\int_t \frac{t}{z-t} d\mu_{\mathbf{S}_{d,n}}(t)$, where $\mu_{\mathbf{S}_{d,n}}(t)$ is the empirical spectral distribution (ESD) of $\mathbf{S}_{d,n}$. By the Marčhenko–Pastur theorem, the ESD of $\mathbf{S}_{d,n}$ converges weakly (and in expectation) to the Marčhenko–Pastur law $\mu_{\mathrm{MP}(\gamma)}$. Therefore,

$$\mathbb{E}\left[\frac{1}{d}\operatorname{Tr}\left(\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\intercal}{n_1}\right)\right] \xrightarrow[d,n\to\infty,\, \frac{d}{n}\to\gamma]{} \int_t \frac{t}{z-t} d\mu_{\mathrm{MP}(\gamma)}(t) = T(z). \tag{28}$$

We then show concentration using Lemma B.1 for Lipschitz functions. Let $Y_n := \frac{1}{d}\operatorname{Tr}(\mathbf{A}_n(z)) = \frac{1}{dn_1}\sum_{j=1}^{n_1}\mathbf{x}_j^\intercal\mathbf{R}_n(z)\mathbf{x}_j$. Then,

$$\begin{aligned}
\frac{\partial \mathbf{R}_n(z)}{\partial \mathbf{X}_{kl}} &= \frac{1}{n}\mathbf{R}_n(z)\left(\mathbf{X}\mathbf{e}_l\mathbf{e}_k^\intercal + \mathbf{e}_k\mathbf{e}_l^\intercal\mathbf{X}^\intercal\right)\mathbf{R}_n(z) = \frac{1}{n}\mathbf{R}_n(z)\left(\mathbf{x}_l\mathbf{e}_k^\intercal + \mathbf{e}_k\mathbf{x}_l^\intercal\right)\mathbf{R}_n(z)\\
\frac{\partial Y_n}{\partial \mathbf{X}_{kl}} &= \frac{1}{dn_1}\sum_{j=1}^{n_1} 2\mathbf{e}_k^\intercal\mathbf{R}_n(z)\mathbf{x}_j\mathbb{1}_{\{j=l\}} + \mathbf{x}_j^\intercal\frac{\partial \mathbf{R}_n(z)}{\partial \mathbf{X}_{kl}}\mathbf{x}_j\\
&= \frac{2}{dn_1}\mathbf{e}_k^\intercal\mathbf{R}_n(z)\mathbf{x}_l\mathbb{1}_{\{l\le n_1\}} + \frac{1}{dn_1n}\sum_{j=1}^{n_1}\mathbf{x}_j^\intercal\mathbf{R}_n(z)(\mathbf{x}_l\mathbf{e}_k^\intercal + \mathbf{e}_k\mathbf{x}_l^\intercal)\mathbf{R}_n(z)\mathbf{x}_j\\
&= \frac{2}{dn_1}\mathbf{e}_k^\intercal\mathbf{R}_n(z)\mathbf{x}_l\mathbb{1}_{\{l\le n_1\}} + \frac{2}{dn_1n}\mathbf{e}_k^\intercal\mathbf{R}_n(z)\left(\sum_{j=1}^{n_1}\mathbf{x}_j\mathbf{x}_j^\intercal\right)\mathbf{R}_n(z)\mathbf{x}_l.
\end{aligned}$$

We use the operator norm bound for the resolvent:

$$\|\mathbf{R}(z)\|_{\mathrm{op}} \le \frac{1}{|\mathrm{Im}(z)|} := c(z).$$

This way, we can bound the absolute value of the partial derivatives of $Y$:

$$\left|\frac{\partial Y_n}{\partial \mathbf{X}_{kl}}\right|^2 \le \frac{c_1}{d^2n_1^2}\left((\mathbf{e}_k^\intercal\mathbf{R}_n(z)\mathbf{x}_l)^2\mathbb{1}_{\{l\le n_1\}} + \frac{1}{n^2}(\mathbf{e}_k^\intercal\mathbf{R}_n(z)\mathbf{X}_1\mathbf{X}_1^\intercal\mathbf{R}_n(z)\mathbf{x}_l)^2\right),$$

for some absolute constant $c_1 \ge 0$. Summing the last equation over $k, l$ (and switching indices), we get:

$$\begin{aligned}
\|\nabla Y_n\|_2^2 &\le \frac{c_1}{d^2n_1^2}\left(\sum_{i=1}^{n_1}\|\mathbf{R}_n(z)\mathbf{x}_i\|_2^2 + \frac{1}{n^2}\sum_{j=1}^{n}\|\mathbf{R}_n(z)\mathbf{X}_1\mathbf{X}_1^\intercal\mathbf{R}_n(z)\mathbf{x}_j\|_2^2\right)\\
&\le c_1\left(\frac{1}{d^2n_1^2}c(z)^2\sum_{i=1}^{n_1}\|\mathbf{x}_i\|_2^2 + \frac{1}{d^2n_1^2n^2}c^4(z)\|\mathbf{X}_1\|_{\mathrm{op}}^4\sum_{j=1}^{n}\|\mathbf{x}_j\|_2^2\right).
\end{aligned} \tag{29}$$

We now define the sets $\mathcal{A}, \mathcal{B}, \mathcal{C}$, and $\mathcal{D}$, where the gradient is bounded:

$$\begin{aligned}
&\mathcal{A} := \left\{\|\mathbf{x}_i\|_2 \le 2\sqrt{d},\ 1 \le i \le n\right\},\\
&\mathcal{B} := \left\{\|\mathbf{X}_1\|_{\mathrm{op}} \le c_3\left(\sqrt{d} + \sqrt{n_1}\right)\right\},\\
&\mathcal{C} := \left\{\left|\frac{n_1}{n} - \pi_1\right| \le \varepsilon\right\}, \qquad\qquad \varepsilon \in (0,1),\\
&\mathcal{D} := \mathcal{A} \cap \mathcal{B} \cap \mathcal{C},
\end{aligned}$$

which are such that

$$\mathbb{P}\left\{\mathcal{A}^c\right\} \leq n\exp(-c_2 d), \qquad \mathbb{P}\left\{\mathcal{B}^c\right\} \leq 2\exp(-d), \qquad \mathbb{P}\left\{\mathcal{C}^c\right\} \leq 2\exp\left(-c_6 n\varepsilon^2\right). \tag{30}$$

The probability bounds for $\mathcal{A}$ can be found in chapter 3 of [21], for $\mathcal{B}$ in chapter 4 of [21], and for $\mathcal{C}$ using the Chernoff bound. We thus get that

$$\|\nabla Y_n\|_2^2 \leq \mathcal{O}\left(\frac{1}{dn}\right) =: L_{d,n}^2,$$

with a probability of at least $1 - \mathcal{O}\left(\exp\left(-c_4 d\right)\right)$ for $n$ and $d$ large enough such that $\frac{d}{n} \to \gamma$ ($c_4$ is a constant dependent on $z$ and $\varepsilon$).

Let $f : \mathbb{R}^{d\times n} \to \mathbb{C}$ be such that $Y_n = f(\mathbf{X})$. We define $\tilde{f}$ to be the McShane's extension of $f$:

$$\tilde{f}(\mathbf{X}) = \inf_{\mathbf{Z}\in\mathcal{D}} f(\mathbf{Z}) + L_{d,n}\|\mathbf{X}-\mathbf{Z}\|_{\mathrm{F}}.$$

This extension guarantees that $\tilde{f}$ is $L_{d,n}$-Lipschitz and that

$$\tilde{f}(\mathbf{X}) = f(\mathbf{X}) \qquad \forall \mathbf{X} \in \mathcal{D}. \tag{31}$$

We will now show that this extension gives $\mathbb{E}\left[|f(\mathbf{X}) - \tilde{f}(\mathbf{X})|\right] = o(1)$. Using (31) and Cauchy-Schwarz, we get:

$$\begin{aligned}
\left|\mathbb{E}\left[f(\mathbf{X}) - \tilde{f}(\mathbf{X})\right]\right| &= \left|\mathbb{E}\left[(f(\mathbf{X}) - \tilde{f}(\mathbf{X}))\mathbb{1}_{\mathbf{X}\in\mathcal{D}^c}\right]\right| \\
&\leq \sqrt{\mathbb{E}\left[\left(f(\mathbf{X}) - \tilde{f}(\mathbf{X})\right)^2\right]\mathbb{P}\left[\mathcal{D}^c\right]} \\
&\leq \sqrt{2}\sqrt{\mathbb{E}\, f^2(\mathbf{X}) + \mathbb{E}\,\tilde{f}^2(\mathbf{X})}\sqrt{\mathbb{P}[\mathcal{D}^c]},
\end{aligned} \tag{32}$$

and

$$f^2(\mathbf{X}) = \left(\frac{1}{d}\operatorname{Tr}\left(\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^T}{n_1}\right)\right)^2 \leq \left\|\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^T}{n_1}\right\|_{\mathrm{op}}^2 \leq c(z)\frac{\|\mathbf{X}_1\|_{\mathrm{op}}^4}{n_1^2}.$$

Using the Dominated convergence theorem and the fact that $n_1/n \to \pi_1$ almost surely, we further obtain that

$$\lim_{\substack{n\to\infty,\\ d\to\infty,\\ \frac{d}{n}\to\gamma}} \mathbb{E}[f^2(\mathbf{X})] = \mathbb{E}\left[\lim_{d,n}\mathbb{E}\left[f^2(\mathbf{X})\,|n_1\right]\right] \leq c_2(z)\,\mathbb{E}\left[\lim_{d,n}\left(\sqrt{\frac{d}{n_1}}+1\right)^2\right] =: c_7, \tag{33}$$

$\tilde{f}(\mathbf{0}) = 0$ since $\mathbf{X} = 0$ gives $\mathbf{S}_n = \mathbf{0}$, and $\mathbf{A}_n = \mathbf{0}$. Since $\tilde{f}$ is Lipschitz we get:

$$\begin{aligned}
|\tilde{f}(\mathbf{X})| &\leq |f(\mathbf{0})| + L_{d,n}\|\mathbf{X}\|_{\mathrm{F}} = L_{d,n}\|\mathbf{X}\|_{\mathrm{F}} \\
\mathbb{E}\,\tilde{f}^2(\mathbf{X}) &\leq L_{d,n}^2\,\mathbb{E}\,\|\mathbf{X}\|_{\mathrm{F}}^2 = \mathcal{O}\left(1\right).
\end{aligned} \tag{34}$$

Plugging (30), (33), and (34) into (32), we get:

$$\mathbb{E}\left[|f(\mathbf{X}) - \tilde{f}(\mathbf{X})|\right] = o(1). \tag{35}$$

We are now ready to provide a concentration bound on $f(\mathbf{X})$ around its expectation:

$$\begin{aligned}
\mathbb{P}\left[|f(\mathbf{X}) - \mathbb{E}\,f(\mathbf{X})| \geq t\right] &= \mathbb{P}\left[|f(\mathbf{X}) - \mathbb{E}\,f(\mathbf{X})| \geq t, \mathbf{X}\in\mathcal{D}\right] + \mathbb{P}\left[|f(\mathbf{X}) - \mathbb{E}\,f(\mathbf{X})| \geq t, \mathbf{X}\in\mathcal{D}^c\right] \\
&\leq \mathbb{P}\left[|\tilde{f}(\mathbf{X}) - \mathbb{E}\,f(\mathbf{X})| \geq t, \mathbf{X}\in\mathcal{D}\right] + \mathbb{P}[\mathcal{D}^c] \\
&\leq \mathbb{P}\left[|\tilde{f}(\mathbf{X}) - \mathbb{E}\,f(\mathbf{X})| \geq t\right] + \mathbb{P}[\mathcal{D}^c] \\
&\leq \mathbb{P}\left[|\tilde{f}(\mathbf{X}) - \mathbb{E}\,\tilde{f}(\mathbf{X})| \geq t - |\,\mathbb{E}[\tilde{f}(\mathbf{X}) - f(\mathbf{X})]|\right] + \mathbb{P}[\mathcal{D}^c].
\end{aligned}$$

For $n$ and $d$ large enough, we know from (35) that $\mathbb{E}\left[|f(\mathbf{X}) - \tilde{f}(\mathbf{X})|\right] \leq \frac{t}{2}$. An application of Lemma B.1 gives (for $n$ and $d$ large enough):

$$\mathbb{P}\left\{|Y_n - \mathbb{E}\,Y_n| \geq t\right\} \leq 2\exp\left(\frac{-t^2 dn}{c_5}\right) + \mathcal{O}\left(\exp\left(-c_4 d\right)\right), \tag{36}$$

for some large enough constant $c_5$. By the Borel-Cantelli lemma, we get that $|Y_n - \mathbb{E}[Y_n]| \xrightarrow[d,n\to\infty,\frac{d}{n}\to\gamma]{\text{a.s.}} 0$, and therefore combining with (28) yields

$$Y_n \xrightarrow[d,n\to\infty,\frac{d}{n}\to\gamma]{\text{a.s.}} T(z).$$

Note that for $n, d$ large enough, the quantities are uniformly bounded and equicontinuous on the set $\mathcal{K}(\eta) = \{z \in \mathbb{C}, d(z, [a, b]) > \eta\}$ for all $\eta > 0$. An application of Arzela-Ascoli's theorem then gives the uniform convergence $\mathcal{K}(\eta)$. $\square$

**Lemma B.3.** *Let $\mathbf{X} \in \mathbb{R}^{d\times n}$ have i.i.d. Gaussian entries $\sim \mathcal{N}(0, 1)$. Partition the columns as $\mathbf{X} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2 & \dots & \mathbf{X}_K\end{bmatrix}$, with $\mathbf{X}_i \in \mathbb{R}^{d\times n_i}$, $(n_1, \dots, n_K) \sim$ Mult$(n, \pi_1, \dots, \pi_K)$, an independent random vector sampled from the multinomial distribution. For $i, j \in [K]$, set:*

$$\mathbf{S}_n := \frac{\mathbf{X}\mathbf{X}^\mathsf{T}}{n}, \qquad \mathbf{R}_n(z) := (z\mathbf{I} - \mathbf{S}_n)^{-1}, \text{and} \qquad \mathbf{B}_n(z) := \frac{\mathbf{X}_i\mathbf{X}_i^\mathsf{T}}{n_i}\mathbf{R}_n(z)\frac{\mathbf{X}_j\mathbf{X}_j^\mathsf{T}}{n_j},$$

*where $z \in \mathbb{C} \setminus \{\lambda_1(\mathbf{S}_n), \dots, \lambda_d(\mathbf{S}_n)\}$. Assume $d, n \to \infty$ with $d/n \to \gamma \in (0, \infty)$. Then,*

$$\frac{1}{d}\operatorname{Tr}(\mathbf{B}_n(z)) \xrightarrow{a.s.} T(z)\left(\frac{\gamma}{\pi_i}\mathbb{1}_{i=j} + \frac{1}{1-\gamma G(z)}\right),$$

*where $T(z) = \int_t \frac{t}{z-t} d\mu_{MP(\gamma)}(t)$, $G(z) = \int_t \frac{1}{z-t}\mu_{MP(\gamma)}(t)$ are respectively the $T$- and the Cauchy transform of the Marčhenko–Pastur law of parameter $\gamma$. Moreover, this convergence is uniform on $\mathcal{K}(\eta) = \{z \in \mathbb{C},\ d(z, [a, b]) > \eta\}$ for all $\eta > 0$.*

*Proof.* We start with the case $i = j$. Without loss of generality, we take $i = 1$.

$$\begin{aligned}
\frac{1}{d}\mathbb{E}\operatorname{Tr}[\mathbf{B}_n(z)] &= \frac{1}{dn_1^2}\mathbb{E}\operatorname{Tr}\left[\mathbf{X}_1\mathbf{X}_1^\mathsf{T}\left(zI - \frac{1}{n}\mathbf{X}\mathbf{X}^\mathsf{T}\right)^{-1}\mathbf{X}_1\mathbf{X}_1^\mathsf{T}\right] \\
&= \frac{1}{dn_1^2}\mathbb{E}\sum_{i=1}^{d}\sum_{j,k=1}^{n_1}[\mathbf{X}_1]_{i,j}\left[\mathbf{X}_1^\mathsf{T}\mathbf{R}_n(z)\mathbf{X}_1\right]_{j,k}[\mathbf{X}_1]_{i,k} \\
&= \frac{1}{dn_1^2}\mathbb{E}\sum_{i=1}^{d}\sum_{j,k=1}^{n_1}\delta_{k=j}\left[\mathbf{X}_1^\mathsf{T}\mathbf{R}_n(z)\mathbf{X}_1\right]_{j,k} + [\mathbf{X}_1]_{i,k}\frac{\partial}{\partial[\mathbf{X}_1]_{i,j}}\left[\mathbf{X}_1^\mathsf{T}\mathbf{R}_n(z)\mathbf{X}_1\right]_{j,k}.
\end{aligned} \tag{37}$$

We used Stein's identity in the last step. To continue, we need the differential of $g(\mathbf{X}_1) = \mathbf{X}_1^\mathsf{T}\mathbf{R}(z)\mathbf{X}_1$, which is given by

$$\mathcal{D}_g(\mathbf{X}_1)(\mathbf{H}) = \mathbf{H}^\mathsf{T}\mathbf{R}_n(z)\mathbf{X}_1 + \mathbf{X}_1^\mathsf{T}\mathbf{R}_n(z)\mathbf{H} + \frac{1}{n}\mathbf{X}_1^\mathsf{T}\mathbf{R}_n(z)\left(\mathbf{X}_1\mathbf{H}^\mathsf{T} + \mathbf{H}\mathbf{X}_1^\mathsf{T}\right)\mathbf{R}_n(z)\mathbf{X}_1. \tag{38}$$

Plugging this into (37), we get:

$$\begin{aligned}
\frac{1}{d}\mathbb{E}\operatorname{Tr}[\mathbf{B}_n(z)] &= \frac{1}{n_1}\mathbb{E}\operatorname{Tr}\left[\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\right] + \frac{1}{dn_1^2}\mathbb{E}\sum_{i=1}^{d}\sum_{j,k=1}^{n_1}[\mathbf{X}_1]_{i,k}[\mathbf{R}_n(z)\mathbf{X}_1]_{i,k} + [\mathbf{X}_1]_{i,k}[\mathbf{X}_1^\mathsf{T}\mathbf{R}_n(z)]_{j,i}\delta_{j=k} \\
&\quad + \frac{1}{n}\frac{1}{dn_1^2}\mathbb{E}\sum_{i=1}^{d}\sum_{j,k=1}^{n_1}\left([\mathbf{X}_1]_{i,k}[\mathbf{X}_1^\mathsf{T}\mathbf{R}_n(z)\mathbf{X}_1]_{j,j}\left[\mathbf{R}_n(z)\mathbf{X}_1\right]_{i,k} + [\mathbf{X}_1]_{i,k}[\mathbf{X}_1^\mathsf{T}\mathbf{R}_n(z)]_{j,i}\left[\mathbf{X}_1^\mathsf{T}\mathbf{R}_n(z)\mathbf{X}_1\right]_{j,k}\right) \\
&= \frac{1}{n_1}\mathbb{E}\operatorname{Tr}\left[\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\right] + \frac{1}{d}\mathbb{E}\operatorname{Tr}\left[\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\right] + \frac{1}{dn_1}\mathbb{E}\operatorname{Tr}\left[\mathbf{R}(z)\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\right] \\
&\quad + \frac{1}{dn}\mathbb{E}\operatorname{Tr}\left[\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\right]\operatorname{Tr}\left[\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\right] + \frac{1}{dn}\mathbb{E}\operatorname{Tr}\left[\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\mathbf{R}_n(z)\right].
\end{aligned}$$

The third and last term vanishes in the limit $d \to \infty, n \to \infty$ since by Lemma B.2 the expectation of $\frac{1}{d}\mathbb{E}\operatorname{Tr}\left[\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\right]$ is finite in the limit. Using Hoffman-Wielandt inequality in the last term, we find:

$$\operatorname{Tr}\left[\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\mathbf{R}_n(z)\right] \le \sum_{i=1}^{d}\lambda_i\left(\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\right)\lambda_i\left(\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\mathbf{R}_n(z)\right). \tag{39}$$

Furthermore, we use the fact that for two positive semi-definite matrices $\mathbf{A}, \mathbf{B}$ we have:

$$\lambda_i(\mathbf{A}\mathbf{B}) \le \lambda_1(\mathbf{A})\lambda_i(\mathbf{B}) \qquad \forall i.$$

Using this in (39), we obtain:

$$\operatorname{Tr}\left[\frac{\mathbf{X}_1\mathbf{X}_1^\intercal}{n_1}\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\intercal}{n_1}\mathbf{R}_n(z)\right] \leq \lambda_1^2\left(\mathbf{R}(z)\right)\sum_{i=1}^{d}\lambda_i^2\left(\frac{\mathbf{X}_1\mathbf{X}_1^\intercal}{n_1}\right). \tag{40}$$

The convergence in expectations on the moments of the Marčhenko–Pastur distribution together with the bound $\lambda_1(\mathbf{R}(z)) \leq \frac{1}{|\mathrm{Im}(z)|}$, we finally get from (40):

$$\frac{1}{dn}\mathbb{E}\operatorname{Tr}\left[\frac{\mathbf{X}_1\mathbf{X}_1^\intercal}{n_1}\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\intercal}{n_1}\mathbf{R}_n(z)\right] \xrightarrow[d,n\to\infty,\frac{d}{n}\to\gamma]{} 0.$$

Lastly, we show that the variance of the normalized trace of $Y_n := \frac{1}{d}\operatorname{Tr}\left(\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\intercal}{n_1}\right)$ goes to 0. By the Gaussian Poincaré inequality:

$$\operatorname{Var}\left(Y_n\right) \leq \mathbb{E}\left\|\nabla Y_n\right\|^2. \tag{41}$$

Using the bound (29), and the facts that $\mathbb{E}\left\|\mathbf{x}_i\right\|^2 = d$ and $\mathbb{E}\left\|\mathbf{X}_1\right\|_{\mathrm{op}}^4 = \mathcal{O}(n^2)$ for $n$ and $d$ large enough, we obtain:

$$\mathbb{E}\left\|\nabla Y_n\right\|^2 = \mathcal{O}\left(\frac{1}{dn}\right).$$

Together with (41), this implies:

$$\operatorname{Var}\left(Y_n\right) \xrightarrow[d,n\to\infty,\frac{d}{n}\gamma]{} 0.$$

This means that:

$$\mathbb{E}\left(\frac{1}{d}\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\intercal}{n_1}\right)^2 \to \left(\mathbb{E}\,\frac{1}{d}\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\intercal}{n_1}\right)^2.$$

All of this combined, we get:

$$\frac{1}{d}\mathbb{E}\operatorname{Tr}\left[\mathbf{B}_n(z)\right] \xrightarrow[d,n\to\infty,\frac{d}{n}\to\gamma]{} T(z)\left(\frac{\gamma}{\pi_1} + 1 + \gamma T(z)\right).$$

Using the Marčhenko–Pastur relation with $G(z)$, the Cauchy transform yields

$$T(z) = zG(z) - 1 = \frac{G(z)}{1-\gamma G(z)}, \tag{42}$$

and we find that

$$T(z)\left(\frac{\gamma}{\pi_1} + 1 + \gamma T\right) = T(z)\left(\frac{\gamma}{\pi_1} + \frac{1}{1-\gamma G(z)}\right).$$

This gives, with (42), the convergence in expectation.

We finish by using the same gradient-norm bound argument as in Lemma B.2 to show that $\|\nabla\frac{1}{d}\operatorname{Tr}(\mathbf{B}_n(z))\|^2 = \mathcal{O}((dn)^{-1})$ with exponentially high probability. Herbst' lemma B.1, together with the McShane extension and Borel–Cantelli lemma give almost surely convergence. The case for $i \neq j$ can be deduced by using what was proven for $i = j$, using the following relation:

$$\begin{aligned}
&\frac{1}{d}\operatorname{Tr}\left[\frac{\mathbf{X}_i\mathbf{X}_i^\intercal}{n_i}\left(z\mathbf{I}-\mathbf{S}_n\right)^{-1}\frac{\mathbf{X}_j\mathbf{X}_j^\intercal}{n_j}\right] \\
&= \frac{(n_i+n_j)^2}{2n_in_j}\frac{1}{d}\operatorname{Tr}\left[\frac{\mathbf{X}_i\mathbf{X}_i^\intercal + \mathbf{X}_j\mathbf{X}_j^\intercal}{n_i+n_j}(z\mathbf{I}-\mathbf{S}_n)^{-1}\frac{(\mathbf{X}_i\mathbf{X}_i^\intercal + \mathbf{X}_j\mathbf{X}_j^\intercal)}{n_i+n_j}\right] \\
&\quad - \frac{n_i^2}{2n_in_j}\frac{1}{d}\operatorname{Tr}\left[\frac{\mathbf{X}_i\mathbf{X}_i^\intercal}{n_i}(z\mathbf{I}-\mathbf{S}_n)^{-1}\frac{\mathbf{X}_i\mathbf{X}_i^\intercal}{n_i}\right] - \frac{n_j^2}{2n_in_j}\frac{1}{d}\operatorname{Tr}\left[\frac{\mathbf{X}_j\mathbf{X}_j^\intercal}{n_j}(z\mathbf{I}-\mathbf{S}_n)^{-1}\frac{\mathbf{X}_j\mathbf{X}_j^\intercal}{n_j}\right].
\end{aligned}$$

The convergence of each term can be deduced from what we proved in the case $i=j$ and gives:

$$\begin{aligned}
&\frac{1}{d}\operatorname{Tr}\left[\frac{\mathbf{X}_i\mathbf{X}_i^{\mathsf{T}}}{n_i}\left(z\mathbf{I}-\mathbf{S}_n\right)^{-1}\frac{\mathbf{X}_j\mathbf{X}_j^{\mathsf{T}}}{n_j}\right] \\
&\xrightarrow[d,n\to\infty,\,\frac{d}{n}\to\gamma]{\text{a.s.}} \frac{(\pi_i+\pi_j)^2}{2\pi_i\pi_j}T(z)\left(\frac{\gamma}{\pi_i+\pi_j}+\frac{1}{1-\gamma G(z)}\right) \\
&\qquad -\frac{1}{2}\frac{\pi_i}{\pi_j}T(z)\left(\frac{\gamma}{\pi_i}+\frac{1}{1-\gamma G(z)}\right)-\frac{1}{2}\frac{\pi_j}{\pi_i}T(z)\left(\frac{\gamma}{\pi_j}+\frac{1}{1-\gamma G(z)}\right) \\
&\qquad = T(z)\left(\frac{1}{1-\gamma G(z)}\right).
\end{aligned}$$

Similar to Lemma B.2, for $n,d$ large enough, the quantities are uniformly bounded and equicontinuous on the set $\mathcal{K}(\eta)=\{z\in\mathbb{C},\ d(z,[a,b])>\eta\}$ for all $\eta>0$. An application of Arzela-Ascoli's theorem thus gives the uniform convergence $\mathcal{K}(\eta)$. $\square$

## Appendix C
## Rare and Intense Subpopulations

The main paper establishes Theorem II.1 for fixed signal strengths $\beta_k$ in the SMM. In this appendix, we show that the same phase transition mechanism also extends to scenarios with diverging energy parameters, $\beta_k\to\infty$, provided that the probability-weighted energy $\pi_k\beta_k$ stays bounded. This signal model captures rare but intense subpopulations: a spike is seen in a vanishing fraction of observations ($\pi_k\to 0$) yet it is strong when present ($\beta_k\to\infty$), with a fixed limiting weighted energy contribution $\pi_k\beta_k\to\eta_k$. In an application domain such as IMS, this could represent a spatially scarce biological tissue structure with a strong mass spectral signature. The regime is of particular interest because it can potentially be encountered in application domains, yet it lies outside the framework provided by [19]. Specifically, the concentration of quadratic forms that [19] assumes fails in this scenario, while the finite-rank-based analysis provided in this paper still applies.

### *A. The regime*

We take $\beta_k=\beta_k(d)$ and $\pi_k=\pi_k(d)$ such that, as $d,n\to\infty$ with $d/n\to\gamma$,

$$\beta_k\to\infty,\qquad \pi_k\beta_k\to\eta_k\in(0,\infty),\qquad \sqrt{d}\ \ll\ \beta_k\ \ll\ d. \tag{43}$$

Since $\mathbf{L}_{l,m}=\theta_{l,m}\sqrt{\pi_l\pi_m\beta_l\beta_m}=\theta_{l,m}\sqrt{\eta_l\eta_m}$, the limiting matrix $\mathbf{L}$ depends on the spikes only through the weighted energy $\eta_k$ and stays bounded.

### *B. Failure of the concentration condition*

**Proposition C.1.** *Under (43), Assumption 5(c) of [19] is violated. More concretely, using $\pi_1\beta_1=\eta_1$, and $\mathbf{A}=\frac{1}{d}\mathbf{I}_d$, we have $\mathbf{g}^{\mathsf{T}}\mathbf{A}\mathbf{g}-\mathbb{E}[\mathbf{g}^{\mathsf{T}}\mathbf{A}\mathbf{g}]\not\prec\|\mathbf{A}\|_F$.*

*Proof.* Since $\operatorname{Tr}\mathbf{A}=1$ and $\|\mathbf{A}\|_F=d^{-1/2}$, we have $\operatorname{Tr}\mathbf{A}/\|\mathbf{A}\|_F=\sqrt{d}$. Writing $\mathbf{g}^{\mathsf{T}}\mathbf{A}\mathbf{g}=\frac{1}{d}\|\mathbf{g}\|^2$, the unconditional mean is $\mu:=\mathbb{E}[\mathbf{g}^{\mathsf{T}}\mathbf{A}\mathbf{g}]=1+\frac{1}{d}\sum_{k=1}^{K}\pi_k\beta_k=1+O(d^{-1})$.
Conditioned on the rare spike $\{z=1\}$, $\mathbf{g}=\alpha\sqrt{\beta_1}\mathbf{v}_1+\boldsymbol{\varepsilon}$ with $\alpha\sim\mathcal{N}(0,1)$, $\boldsymbol{\varepsilon}\sim\mathcal{N}(0,\mathbf{I}_d)$. Then,

$$\mathbf{g}^{\mathsf{T}}\mathbf{A}\mathbf{g}\,\big|\,\{z=1\}=\frac{\alpha^2\beta_1}{d}+\frac{2\alpha\sqrt{\beta_1}\,\mathbf{v}_1^{\mathsf{T}}\boldsymbol{\varepsilon}}{d}+\frac{\|\boldsymbol{\varepsilon}\|^2}{d}=1+\frac{\alpha^2\beta_1}{d}+O_{\prec}\!\left(d^{-1/2}\right),$$

using $\frac{1}{d}\|\boldsymbol{\varepsilon}\|^2=1+O_{\prec}(d^{-1/2})$ and $\frac{1}{d}2\alpha\sqrt{\beta_1}\mathbf{v}_1^T\boldsymbol{\varepsilon}=O_{\prec}(\sqrt{\beta_1}/d)=o(d^{-1/2})$ since $\beta_1\ll d$. Since $\alpha^2\sim\chi_1^2$, we have $\mathbb{P}[\alpha^2\ge\frac{1}{2}]\ge c_0>0$. Moreover, on $\{z=1\}\cap\{\alpha^2\ge\frac{1}{2}\}$ and for all large $d$,

$$\left|\mathbf{g}^{\mathsf{T}}\mathbf{A}\mathbf{g}-\mu\right|\ \ge\ \frac{\beta_1}{2d}-o(d^{-1/2}),\qquad \frac{\left|\mathbf{g}^{\mathsf{T}}\mathbf{A}\mathbf{g}-\mu\right|}{\|\mathbf{A}\|_F}\ \gtrsim\ \frac{\beta_1/d}{d^{-1/2}}=\frac{\beta_1}{\sqrt{d}}\ \xrightarrow{\beta_1\gg\sqrt{d}}\ \infty.$$

Hence, there exists $\varepsilon>0$ and large $d$ such that this event lies in $\{|\mathbf{g}^{\mathsf{T}}\mathbf{A}\mathbf{g}-\mathbb{E}[\mathbf{g}^{\mathsf{T}}\mathbf{A}\mathbf{g}]|>d^{\varepsilon}\|\mathbf{A}\|_F\}$, and its probability is

$$\mathbb{P}\left[z=1,\ \alpha^2\ge\frac{1}{2}\right]=\pi_1\mathbb{P}\left[\alpha^2\ge\frac{1}{2}\right]\ \ge\ \frac{c_0\eta_1}{\beta_1}\ \asymp\ \frac{1}{\beta_1}.$$

Taking a real scalar $D=1$, since $\beta_1\ll d$ we have $1/\beta_1\gg d^{-1}=d^{-D}$, so the bound $\mathbb{P}[\cdot]<d^{-D}$ required by $\prec$ fails for this $D$. Thus, for this particular choice of $\mathbf{A}$, Assumption 5(c) of [19] does not hold. $\square$

### C. Recovery of the phase transition: a worked example

We explore a concrete $K = 2$ instance in regime (43), which according to Proposition C.1 falls outside of the scope of [19], and for which the finite-rank-based argument provided in this paper still recovers the phase transition. The example is minimal: one rare-intense spike alongside an ordinary sub-critical one, with asymptotically orthogonal signals.

*a) Setup:* For $K = 2$, we draw $\mathbf{v}_1, \mathbf{v}_2$ from (3) with $\theta_{1,2} = 0$, so that $\mathbf{v}_1, \mathbf{v}_2$ are independent and uniform on $\mathbb{S}^{d-1}$. Conditioning on $\mathbf{v}_1$, the map $\mathbf{v}_2 \mapsto \mathbf{v}_1^\mathsf{T}\mathbf{v}_2$ is 1-Lipschitz with mean zero, so Theorem A.1 gives $\mathbf{v}_1^\mathsf{T}\mathbf{v}_2 = O_\prec(d^{-1/2})$.

$$\beta_1 = d^{1/2+\varepsilon}\ (\varepsilon \in (0, \tfrac{1}{2})), \quad \pi_1 = \frac{\eta_1}{\beta_1}, \qquad \beta_2 > 0 \text{ fixed}, \quad \pi_2 = 1 - \pi_1, \quad \pi_2\beta_2 \to \beta_2 := \eta_2 < \sqrt{\gamma}. \tag{44}$$

Then, $\beta_1 \to \infty$, $\pi_1\beta_1 = \eta_1 \in (0, \infty)$ fixed, and $\sqrt{d} \ll \beta_1 \ll d$, so (43) holds for the rare component. The second signal strength, $\beta_2$, adheres to the fixed-strength setting of Theorem II.1. The limiting Gram matrix is diagonal, $\mathbf{L} = \mathrm{Diag}(\eta_1, \eta_2)$.

*b) Block sizes:* With $(n_1, n_2) \sim \mathrm{Mult}(n; \pi_1, \pi_2)$, the rare count $n_1 \sim \mathrm{Binomial}(n, \pi_1)$ has mean $m_1 = n\pi_1 \asymp (\eta_1/\gamma)d^{1/2-\varepsilon} \to \infty$. Since $\pi_1 \to 0$, the additive control $\{|n_1/n - \pi_1| \le \varepsilon\}$ of Lemmas B.2–B.3 is vacuous (its exponent $n\pi_1^2 \asymp d^{-2\varepsilon} \to 0$). We replace it by the multiplicative Chernoff bound, valid for every $(n, \pi_1)$ and $\delta \in (0, 1)$,

$$\mathbb{P}\left(|n_1 - m_1| \ge \delta\, m_1\right) \le 2\exp\left(-\tfrac{\delta^2 m_1}{3}\right). \tag{45}$$

As $m_1$ is a diverging power of $d$, the right-hand side is summable, and the Borel-Cantelli lemma (letting $\delta$ to 0 along a countable sequence) yields

$$n_1 \xrightarrow{\text{a.s.}} \infty, \qquad \frac{n_1}{n\pi_1} \xrightarrow{a.s.} 1, \qquad \frac{n_1}{n}\beta_1 \xrightarrow{a.s.} 1, \qquad \frac{n_2}{n} \xrightarrow{a.s.} 1. \tag{46}$$

*c) Weighted trace (adapting the convergence results):* The diagonal limit of Lemma B.3 carries a factor $\gamma/\pi_i$ that diverges as $\pi_1 \to 0$, so the lemma cannot be applied directly to the rare block. It is, however, needed only in the $\frac{n_i}{n}$-weighted form, in which the weight cancels the divergence. Setting

$$\mathbf{B}_n(z) := \frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}\mathbf{R}_n(z)\frac{\mathbf{X}_1\mathbf{X}_1^\mathsf{T}}{n_1}, \qquad \mathbf{C}_n(z) := \frac{\mathbf{X}_2\mathbf{X}_2^\mathsf{T}}{n_2}\mathbf{R}_n(z)\frac{\mathbf{X}_2\mathbf{X}_2^\mathsf{T}}{n_2},$$

the proof of Lemma B.3—which uses $\pi_i$ only through $d/n_i \to \gamma/\pi_i$, all variance bounds needing merely $n_i \to \infty$ and $\|\mathbf{R}_n\| \le 1/|\mathrm{Im}(z)|$—together with (46) gives:

$$\frac{n_1}{n}\frac{1}{d}\mathrm{Tr}(\mathbf{B}_n(z)) \xrightarrow{\text{a.s.}} T(z)\gamma, \tag{47}$$

$$\frac{n_2}{n}\frac{1}{d}\mathrm{Tr}(\mathbf{C}_n(z)) \xrightarrow{\text{a.s.}} T(z)\left(\gamma + \frac{1}{1 - \gamma G(z)}\right), \tag{48}$$

the rare weight $\frac{n_1}{n} \asymp \pi_1$ absorbing the $\gamma/\pi_1$.

*d) Transform first, then take the limit:* As in the approach developed above, we let:

$$\mathbf{M}_{d,n}(z) := \mathbf{P}_2^\mathsf{T}(z\mathbf{I} - \mathbf{Z})^{-1}\mathbf{P}_1,$$
$$\tilde{\mathbf{M}}_{d,n}(z) := \tilde{\mathbf{P}}_2^\mathsf{T}(z\mathbf{I} - \mathbf{Z})^{-1}\tilde{\mathbf{P}}_1.$$

where $\tilde{\mathbf{P}}_i$ is obtained by scaling the even columns by $-1$. The eigenvalues popping out the MP spectra are the roots of

$$|\det(\mathbf{I}_{2K} - \mathbf{M}_{d,n})| = \left|\det\left(\mathbf{I}_{2K} - (\mathbf{M}_{d,n} + \tilde{\mathbf{M}}_{d,n} - \mathbf{M}_{d,n}\tilde{\mathbf{M}}_{d,n})\right)\right|^{1/2}. \tag{49}$$

In the approach developed in section III for fixed $\beta_k$, we first found the limiting $\mathbf{M}_{d,n}$, and manipulated the matrix afterward. Here, the entries of $\mathbf{M}_{d,n}$ diverge ($c_1 \sim \sqrt{\beta_1} \to \infty$), so we instead perform the manipulation at finite $d$ and limit afterwards. Conjugating by

$$\mathbf{R} := \mathrm{Diag}\left(\sqrt{\frac{n_2}{n}}, \sqrt{\frac{n_1}{n}}, \sqrt{\frac{n_1}{n}}, \sqrt{\frac{n_2}{n}}\right) \in \mathbb{R}^{2K\times 2K}$$

is a similarity, hence determinant-preserving for every finite $d$, and it places every divergent quantity behind a weight $\frac{n_k}{n} \asymp \pi_k$. With $\mathbf{H}_k = c_k\mathbf{Z}_k + \frac{c_k^2}{2}(\mathbf{v}_k^\top \mathbf{Z}_k \mathbf{v}_k)\mathbf{I}$, the strength $\beta_1$ (equivalently $c_1 \sim \sqrt{\beta_1}$) therefore enters every entry of $\mathbf{R}\left(\mathbf{M}_{d,n} + \tilde{\mathbf{M}}_{d,n} - \mathbf{M}_{d,n}\tilde{\mathbf{M}}_{d,n}\right)\mathbf{R}^{-1}$ only through the bounded combinations

$$\tfrac{n_1}{n}c_1^2 \to \eta_1, \qquad \frac{n_1}{n}c_1 \to 0,$$

together with the weighted traces (47), (48). Every entry is thus converging in the limit. The diagonal blocks converge to multiples of the identity, while the cross blocks carry the additional factor $\mathbf{v}_1^\top \mathbf{v}_2 = O_\prec(d^{-1/2}) \to 0$ and hence vanish. We then have

$$\begin{aligned}
&\mathbf{R}(\mathbf{M}+\tilde{\mathbf{M}})\mathbf{R}^{-1}\\
&:= \begin{bmatrix}
2\frac{n_1}{n}\mathbf{v}_1^\top(z\mathbf{I}-\mathbf{Z})^{-1}\mathbf{H}_1\mathbf{v}_1 & 0 & 2\frac{n_2}{n}\mathbf{v}_1^\top(z\mathbf{I}-\mathbf{Z})^{-1}\mathbf{H}_2\mathbf{v}_2 & 0\\
0 & 2\frac{n_1}{n}\mathbf{v}_1^\top\mathbf{H}_1(z\mathbf{I}-\mathbf{Z})^{-1}\mathbf{v}_1 & 0 & 2\frac{n_1}{n}\mathbf{v}_1^\top\mathbf{H}_1(z\mathbf{I}-\mathbf{Z})^{-1}\mathbf{v}_2\\
2\frac{n_1}{n}\mathbf{v}_2^\top(z\mathbf{I}-\mathbf{Z})^{-1}\mathbf{H}_1\mathbf{v}_1 & 0 & 2\frac{n_2}{n}\mathbf{v}_2^\top(z\mathbf{I}-\mathbf{Z})^{-1}\mathbf{H}_2\mathbf{v}_2 & 0\\
0 & 2\frac{n_2}{n}\mathbf{v}_2^\top\mathbf{H}_2(z\mathbf{I}-\mathbf{Z})^{-1}\mathbf{v}_1 & 0 & 2\frac{n_2}{n}\mathbf{v}_2^\top\mathbf{H}_2(z\mathbf{I}-\mathbf{Z})^{-1}\mathbf{v}_2
\end{bmatrix}\\
&\to \begin{bmatrix}
\eta_1 G(z) & 0 & 0 & 0\\
0 & \eta_1 G(z) & 0 & 0\\
0 & 0 & 2c_2T(z)+c_2^2G(z) & 0\\
0 & 0 & 0 & 2c_2T(z)+c_2^2G(z)
\end{bmatrix}\\
&\mathbf{R}\mathbf{M}\tilde{\mathbf{M}}\mathbf{R}^{-1}\\
&\to \begin{bmatrix}
\eta_1\gamma T(z)G(z) & 0 & 0 & 0\\
0 & \eta_1\gamma T(z)G(z) & 0 & \\
0 & 0 & -c_2^2T(z)\gamma G(z) & 0\\
0 & 0 & 0 & -c_2^2T(z)\gamma G(z)
\end{bmatrix}
\end{aligned}$$

$$\mathbf{R}(\mathbf{M}+\tilde{\mathbf{M}})\mathbf{R}^{-1} - \mathbf{R}\mathbf{M}\tilde{\mathbf{M}}\mathbf{R}^{-1} \to \begin{bmatrix}
\eta_1 T(z) & 0 & 0 & 0\\
0 & \eta_1 T(z) & 0 & 0\\
0 & 0 & \beta_2 T(z) & 0\\
0 & 0 & 0 & \beta_2 T(z)
\end{bmatrix}, \tag{50}$$

where we used the relation between $G(z)$ and $T(z)$ : $T(z) = G(z)/(1-\gamma G(z))$.

*e) Phase transition:* By (49) and (50), the limiting outlier equation is

$$|1-\eta_1 T(z)|\,|1-\eta_2 T(z)| = 0.$$

By the threshold analysis of Section III-D, $T(z) = 1/\eta_k$ has a root outside $[a,b]$ iff $\eta_k > \sqrt{\gamma}$. Since $\eta_2 < \sqrt{\gamma}$ by construction, the second factor gives no outlier and $\lambda_2(\mathbf{S}_{d,n}) \to (1+\sqrt{\gamma})^2$, while

$$\lambda_1(\mathbf{S}_{d,n}) \xrightarrow{\text{a.s.}} \begin{cases} T^{-1}(1/\eta_1) = (1+\eta_1)(1+\gamma/\eta_1), & \eta_1 > \sqrt{\gamma},\\ (1+\sqrt{\gamma})^2, & \eta_1 \le \sqrt{\gamma}. \end{cases}$$

Detectability of the rare spike's subpopulation is thus governed by its weighted energy $\eta_1 = \pi_1\beta_1$ alone, independently of the peak strength $\beta_1 = d^{1/2+\varepsilon}$, for every $\varepsilon \in (0, \frac{1}{2})$. Since this is a configuration on which Assumption 5(c) of [19] fails by Proposition C.1, it demonstrates merit for a finite-rank-based approach as pursued in this paper.

# REFERENCES


[1] M. Thamm, M. Staats, and B. Rosenow, “Random matrix analysis of deep neural network weight matrices,” *Physical Review E*, vol. 106, no. 5, p. 054124, 2022.

[2] A. Taherpour, A. Taherpour, T. Khattab, and X. Wang, “Large array antenna spectrum sensing in cognitive radio networks,” *IEEE Transactions on Cognitive Communications and Networking*, vol. 11, no. 2, pp. 1119–1132, 2025.

[3] M. Nitzan and M. P. Brenner, “Revealing lineage-related signals in single-cell gene expression using random matrix theory,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 11, p. e1913931118, 2021.

[4] J. Veraart, D. S. Novikov, D. Christiaens, B. Ades-aron, J. Sijbers, and E. Fieremans, “Denoising of diffusion mri using random matrix theory,” *NeuroImage*, vol. 142, pp. 394–406, 2016.

[5] L. Laloux, P. Cizeau, J.-P. Bouchaud, and M. Potters, “Noise dressing of financial correlation matrices,” *Phys. Rev. Lett.*, vol. 83, pp. 1467–1470, Aug 1999.

[6] L. Benigni and S. Péché, “Eigenvalue distribution of some nonlinear models of random matrices,” *Electronic Journal of Probability*, vol. 26, pp. 1–37, 2021.

[7] J. M. Spraggins, A. B. Hummon, S. R. Ellis, L. Li, B. M. Prentice, M. Marchetti-Deschmann, R. Van de Plas, and R. M. Caprioli, “Matrix-assisted laser desorption/ionization imaging mass spectrometry,” *Nature Reviews Methods Primers*, vol. 6, no. 1, p. 44, 2026.

[8] L. Erdős, B. Schlein, and H.-T. Yau, “Semicircle law on short scales and delocalization of eigenvectors for wigner random matrices,” *The Annals of Probability*, vol. 37, no. 3, pp. 815–852, May 2009.

[9] V. A. Marčenko and L. A. Pastur, “Distribution of eigenvalues for some sets of random matrices,” *Mathematics of the USSR-Sbornik*, vol. 1, no. 4, p. 457, apr 1967.

[10] A. Perry, A. S. Wein, A. S. Bandeira, and A. Moitra, “Optimality and sub-optimality of pca for spiked random matrices and synchronization,” *arXiv preprint arXiv:1609.05573*, 2016.

[11] J. Baik, G. Ben Arous, and S. Péché, “Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices,” *The Annals of Probability*, vol. 33, no. 5, pp. 1643–1697, 2005.

[12] G. Ben Arous, C. Gerbelot, and V. Piccolo, “High-dimensional optimization for multi-spiked tensor PCA,” *arXiv preprint arXiv:2408.06401*, 2024.

[13] S. Lee and J. O. Lee, “Phase transition for the generalized two-community stochastic block model,” *Journal of Applied Probability*, vol. 61, no. 2, p. 385–400, 2024.

[14] I. D. Landau, G. C. Mel, and S. Ganguli, “Singular vectors of sums of rectangular random matrices and optimal estimation of high-rank signals: The extensive spike model,” *Phys. Rev. E*, vol. 108, p. 054129, Nov 2023.

[15] J. H. Jung, H. W. Chung, and J. O. Lee, “Detection of signal in the spiked rectangular models,” in *Proceedings of the 38th International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, M. Meila and T. Zhang, Eds., vol. 139. PMLR, 18–24 Jul 2021, pp. 5158–5167. [Online]. Available: https://proceedings.mlr.press/v139/jung21a.html

[16] P.-L. Delacour, S. Wahls, J. M. Spraggins, L. Migas, and R. Van de Plas, “Signal recovery using a spiked mixture model,” *IEEE Transactions on Signal Processing*, vol. 73, pp. 3748–3761, 2025.

[17] A. Bloemendal, L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, “Isotropic local laws for sample covariance and generalized wigner matrices,” *Electronic Journal of Probability*, vol. 19, pp. 1–53, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:18262992

[18] A. Knowles and J. Yin, “Anisotropic local laws for random matrices,” *Probability Theory and Related Fields*, vol. 169, pp. 257–352, 2017. [Online]. Available: https://doi.org/10.1007/s00440-016-0730-4

[19] Z. Wang, D. Wu, and Z. Fan, “Nonlinear spiked covariance matrices and signal propagation in deep neural networks,” in *Proceedings of Thirty Seventh Conference on Learning Theory*, ser. Proceedings of Machine Learning Research, S. Agrawal and A. Roth, Eds., vol. 247. PMLR, 30 Jun–03 Jul 2024, pp. 4891–4957. [Online]. Available: https://proceedings.mlr.press/v247/wang24b.html

[20] F. Benaych-Georges and R. R. Nadakuditi, “The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices,” *Advances in Mathematics*, vol. 227, no. 1, pp. 494–521, 2011.

[21] R. Vershynin, *High-dimensional probability: An introduction with applications in data science*. Cambridge university press, 2018, vol. 47.

[22] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*. Cambridge university press, 2010, no. 118.