HOW WEAK ARE WEAK FACTORS? UNIFORM INFERENCE FOR SIGNAL STRENGTH IN SIGNAL PLUS NOISE MODELS

ANNA BYKHOVSKAYA, VADIM GORIN, AND SASHA SODIN

Abstract. The paper analyzes four classical signal-plus-noise models: the factor model, spiked sample covariance matrices, the sum of a Wigner matrix and a low-rank perturbation, and canonical correlation analysis with low-rank dependencies. The objective is to construct confidence intervals for the signal strength that are uniformly valid across all regimes – strong, weak, and critical signals. We demonstrate that traditional Gaussian approximations fail in the critical regime. Instead, we introduce a universal transitional distribution that enables valid inference across the entire spectrum of signal strengths. The approach is illustrated through applications in macroeconomics and finance.

1. Introduction

1.1. **Motivation.** In the modern era researchers increasingly have access to high-dimensional data sets across a wide range of fields. These data sets are inevitably contaminated by various forms of error and noise, making the separation of meaningful structure from background noise a central challenge. To address this analysts commonly employ dimension-reduction techniques. The two dominant approaches are low-rank methods, which assume that the underlying signal lies in a lower-dimensional subspace, and sparsity-based methods, which assume that only a small subset of variables or parameters are truly relevant, i.e., non zero. This paper adopts the low-rank perspective. For a discussion of settings where this assumption is appropriate, we refer to Udell and Townsend [2019], Giannone et al. [2021], and Thibeault et al. [2024]. In particular, Giannone et al. [2021] argue that numerous data sets in macroeconomics, microeconomics, and finance exhibit dense, rather than sparse, structures.

A prototypical example of a low rank setting is the factor model, where one observes an $N \times S$ data matrix X and assumes that it can be decomposed as

$$X = LF^{\mathsf{T}} + \mathcal{E}, \tag{1.1}$$

where F is an $S \times r$ matrix of factors, L is an $N \times r$ matrix of factor loadings, and LF^{T} represents the low-rank signal of interest. The signal rank r is small relative to the large dimensions N and S. The remainder $\mathcal{E}$ is a noise matrix, often assumed to have i.i.d. mean-zero entries in the simplest setting.

The feasibility of consistently estimating the signal component LF^{T} from the observed data X hinges on the strength of the signal, which can be quantified by the singular values of LF^{T}. When these singular values are large, the signal is strong and estimation is reliable, as can be directly predicted from the form of (1.1). As the signal weakens, the data X becomes less

Date: August 11, 2025.

Gorin's work was partially supported by NSF grant DMS-2246449. Sodin's work was partially supported by a Philip Leverhulme Prize (PLP-2020-064) and by the Morris Belkin Visiting Professor Program at the Weizmann Institute of Science.

informative, and below a certain critical threshold, accurate recovery becomes impossible. The relationships between the strength of the signal and feasibility of reconstruction of LF^{T} have been rigorously analyzed in a number of studies, see, e.g., Stock and Watson [2002], Bai and Ng [2002], Bai [2003], Paul [2007], Onatski [2012], Johnstone and Paul [2018], Bai and Ng [2023], Fan et al. [2024], Barigozzi and Hallin [2024] and references therein.

Given this behavior, applied work using factor models should begin by assessing the strength of the factors, since the validity of any inference on L or F critically depends on it. However, in practice, this step is often overlooked[1], and most studies tacitly assume that the factors are strong, without conducting any formal diagnostics.

This paper seeks to emphasize the importance of assessing signal strength in a broad class of "signal plus noise" models. To that end, we develop novel procedures for constructing confidence intervals for signal strength. Crucially, we do not assume that the signals are strong – an assumption often unjustified in empirical applications. Instead, our analysis remains valid across the full range of regimes: strong, weak, and critical signals.

1.2. **Models and results.** We analyze four classical high-dimensional statistical models: the factor model (1.1), spiked sample covariance, the spiked Wigner model, and spiked canonical correlations. These models correspond to three fundamental ensembles from random matrix theory: the Laguerre/Wishart ensemble for the first two, the Hermite/Gaussian/Wigner ensemble for the third, and the Jacobi ensemble for the fourth. Each of these models can be viewed as an instance of the signal-plus-noise framework – also known as spiked random matrices, a term originating with Johnstone [2001] – in which a low-rank signal matrix is embedded in a high-dimensional noisy environment. The goal is to detect and quantify the embedded signal.

The signal in each model can be decomposed into a sum of rank-one components. Each component is characterized by a positive scalar (its strength) and one or two unit-norm vectors (its direction), depending on the setup. In this work we focus solely on the signal strength and do not consider inference on directions.

Our analysis is based on spectral methods, whereby signal strength is inferred from the eigenvalues of certain model-specific matrices. In all four setups a well-documented phase transition phenomenon arises: the signal strength can be consistently estimated (in the high-dimensional asymptotic regime with proportional growth of data dimensions) only when it exceeds a critical threshold, see Jones et al. [1978], Baik and Silverstein [2006], Onatski [2012], Bao et al. [2019] and more references in Section 2. When the signal strength falls below the threshold, only partial probabilistic information, such as asymptotics of the likelihood ratio test can be recovered about the signal, but reliable point estimation becomes impossible, see, e.g. Onatski et al. [2013, 2014], Dobriban [2017], Johnstone and Onatski [2020], El Alaoui et al. [2020]. The intermediate regime, where the signal strength is close to the threshold, is typically referred to as the "critical" regime. This regime is particularly challenging for inference.

In the super-critical case, where the strength is significantly above the threshold, the estimation procedure is quite straightforward: one takes the largest eigenvalue, applies to it a certain explicit function (see Section 2 for the formulas) and gets the strength of the strongest signal. Repeating the same with the second, third, etc., eigenvalues one gets

[1]This pattern is evident in the vast majority of approximately 120 papers that employ factor models or PCA-related techniques, published in the five leading economics journals between 2015 and 2025.

strengths of the further components of the signal and the only question is when to stop, i.e., after which step one should declare that the following signals are too weak and can not be recovered. There are many results in the literature proposing various algorithms to choose the stopping point. We further remark that for very strong signals the function one should apply to the eigenvalues is close to identity ($f(x) = x$), whereas for weaker signals the function exhibits stronger dependence on the model of interest.

Once point estimates of the signal strengths are obtained, the next natural question is how to quantify uncertainty – specifically, how to construct confidence intervals for these estimates. The existing literature offers little guidance on this front – particularly guidance that is consistent across models and signal strengths – with most results focusing on strong signals. The technical challenge is rooted in the nonstandard asymptotic behavior of the eigenvalues near the phase transition threshold. While the fluctuations of the top eigenvalues are asymptotically Gaussian for well-separated (super-critical) signals, the limiting distribution becomes highly non-Gaussian and analytically intricate as the signal strength approaches the critical boundary (see Baik et al. [2005], Mo [2012], and Bloemendal and Virág [2013] for rigorous results in the sample covariance setting).

Our paper fills this gap by proposing a general procedure for constructing confidence intervals for signal strength. Remarkably, across all four models we study, the confidence intervals are characterized by a common limiting (stochastic) object, which we call the *Airy–Green function* and denote $\mathcal{G}(w)$. Our main contributions are: a rigorous construction of this function, a unified set of theorems linking it to the four canonical models, and tabulated confidence intervals based on $\mathcal{G}(w)$. The only model-specific components are a set of scaling constants, which we provide explicitly for each setting.

1.3. **Econometrics and statistics contributions.** In economics and finance it has long been observed that many data sets contain factors that are either non-informative or far from strong – see e.g. Giglio et al. [2023] and Kim et al. [2024] for overviews and extensive references. This concern is especially apparent in the vast "factor zoo" of potential variables proposed to explain stock returns. This empirical reality has motivated a line of theoretical research focused on inference for weaker factors. Broadly speaking, factors can be classified by their strength into three categories: strong (as in, e.g. Bai and Ng [2002], Stock and Watson [2002]), semi-strong (as in, e.g. Bai and Ng [2023], Fan et al. [2024]), and weak[2] (as in, e.g., Onatski [2012]). The literature also includes statistical procedures for testing and distinguishing between these types of factors (see, in particular, Kim et al. [2024]). Over the past decades a growing body of research has focused specifically on factor strength, including contributions by Chudik et al. [2011], Bailey et al. [2016], Wang and Fan [2017], Lettau and Pelger [2020], Cai et al. [2020], Bailey et al. [2021], Freyaldenhoven [2022], Uematsu and Yamagata [2022], and Pesaran and Smith [2025].

In comparison to this literature, our main methodological contribution is a unified procedure for constructing confidence intervals for signal strength across all four models and all signal ranges, as presented in Section 3. This approach does not rely on standard Gaussian quantiles, but instead uses a novel random *transition process* $\mathcal{T}(\Theta)$, whose quantiles are tabulated in Table 1. As shown in Figure 1a, the Gaussian approximation performs poorly near the critical threshold, making $\mathcal{T}(\Theta)$ essential for accurate inference in that regime. This is

[2]What we call semi-strong factors are sometimes referred to as weak, while weak factors may be termed weakly influential or extremely weak.

reminiscent of the construction of uniform confidence intervals for autoregressive models in Stock [1991], Mikusheva [2007], where the non-standard asymptotics near the unit root are smoothly connected to the standard normal behavior in the stationary region.

Beyond quantifying uncertainty in signal strength, our framework also enables signal detection and the assessment of factor informativeness. Specifically, one can check whether the uniform confidence intervals include zero and the identification cut-off, respectively.

A surprising finding is that the same transition process $\mathcal{T}(\Theta)$ governs all four models. In fact, the proofs in Section 9 follow different paths depending on the model, and only in the final step does a structural identity emerge, revealing that all four asymptotic distributions coincide. Random matrix theory has many universality theorems, and based on our results, we predict that the same transition process $\mathcal{T}(\Theta)$ governs a much wider class of signal-plus-noise models, beyond the ones analyzed here.

1.4. **Mathematical contributions.** From a mathematical perspective, we develop a new approach to analyzing critical spikes, grounded in perturbation theory equations that relate the eigenvalues of spiked and unspiked random matrices. This contrasts with earlier treatments of critical spikes in real symmetric matrices, which relied on Pfaffian point processes (as in Mo [2012]) or on tridiagonal matrix models (as in Bloemendal and Virág [2013, 2016], Lamarre and Shkolnikov [2019]). Our central technical contribution is to show that these perturbation equations admit a well-defined edge-scaling limit, which captures the asymptotic behavior of the largest eigenvalues. While our approach is novel in all four settings, we particularly emphasize the fourth – canonical correlations – where no prior results on critical spikes were available.

In Section 4 and 8 we establish this edge limit result under two key assumptions on the unspiked model: (i) the asymptotics of the largest eigenvalues converge to the Airy_1 point process, and (ii) a form of the local law holds for the Stieltjes transform near the spectral edge. These assumptions are known to hold for a wide range of random matrix ensembles, including the four models considered in this paper. A notable strength of our approach is its minimal reliance on model-specific structure: we require only the two inputs above.

We build on some of the ideas in Aizenman and Warzel [2015]. In contrast, however, we focus on the limit at the spectral edge—rather than in the bulk—which requires subtracting diverging counterterms. Moreover, we establish convergence in a stronger topology, which allows us to work directly on the real axis; see Appendix 8 for further details.

1.5. **Outline of the paper.** Section 2 introduces the four main signal-plus-noise models. Section 3 presents a unified procedure for constructing confidence intervals for signal strengths. Section 4 lays out the theoretical foundations underlying this procedure. Section 5 offers three empirical illustrations. Some extensions are discussed in Section 6. Section 7 concludes. All proofs are in Appendices 8 and 9.

2. Four signal plus noise models

In this section we present the four models, beginning with the simplest case – the spiked Wigner model – then proceeding to sample covariance and factor models based on PCA, and concluding with canonical correlation analysis (CCA). Although PCA-based models are the most widely used in practice, we adopt this order because the formulas are simpler in the Wigner case, making the key ideas more transparent.

2.1. **Spiked Wigner matrix.** Suppose we observe an $N \times N$ matrix $\mathbf{A}$ of the form

$$\mathbf{A} = \sum_{i=1}^{r} \theta_i \cdot \mathbf{u}_i^* (\mathbf{u}_i^*)^\mathsf{T} + \mathcal{E}, \tag{2.1}$$

where r is fixed (not growing with N) and $\theta_1 > \cdots > \theta_r > 0 \in \mathbb{R}$ are the strengths of r signals, with corresponding directions $\mathbf{u}_1^*, \ldots, \mathbf{u}_r^*$, which are assumed to be orthonormal N–dimensional vectors. The noise matrix $\mathcal{E}$ is a (Wigner) matrix sampled from the Gaussian Orthogonal Ensemble, meaning that $\mathcal{E} = \frac{1}{\sqrt{2N}}(\mathcal{Z} + \mathcal{Z}^\mathsf{T})$, where $\mathcal{Z}$ is an $N \times N$ matrix of i.i.d. $\mathcal{N}(0, \sigma^2)$ entries (see Section 6.1 for the discussion of non-Gaussian setting). We assume that θ_i and $\mathbf{u}_i^*$ are unknown deterministic parameters; one could alternatively allow $\mathbf{u}_i^*$ to be random, provided they are independent of $\mathcal{E}$. Our goal is to estimate the signal strengths $\theta_1, \ldots, \theta_r$.

We first assume that the variance of the underlying noise $\mathcal{Z}$, σ^2, is known and, without loss of generality, set it to 1 by rescaling the model.[3] In Section 3.3 we discuss adjustments for the case when σ^2 is unknown.

One common application of the spiked Wigner framework is modeling symmetric interaction networks, for instance, economic or social activity among N agents. In such cases each rank-one component $\theta_i \mathbf{u}_i^* (\mathbf{u}_i^*)^\top$ captures a latent structure in agent attributes $\mathbf{u}_i^*$, while the observed interactions are contaminated by noise $\mathcal{E}$. Low-rank approximations of this form underpin seminal network models including the stochastic block model of Holland et al. [1983], where communities are inferred from block-structured adjacency matrices, and latent space models.

The following result establishes the threshold for the estimation of θ_i via spectral methods.

Proposition (Jones et al. [1978], Füredi and Komlós [1981], Capitaine et al. [2009, 2012])**.** *Suppose that all θ_i are distinct and ordered $\theta_1 > \theta_2 > \cdots > \theta_r$, $\sigma^2 = 1$. Let $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N$ denote the eigenvalues of $\mathbf{A}$ sampled from* (2.1) *with $\sigma^2 = 1$. Denote*

$$\theta^c = 1, \qquad \lambda_+ = 2, \qquad \lambda(\theta) = \theta + \frac{1}{\theta}, \qquad V(\theta) = 2\,\frac{\theta^2 - 1}{\theta^2}. \tag{2.2}$$

For each $1 \leq i \leq r$, if $\theta_i > \theta^c$, then as $N \to \infty$, in the sense of convergence in distribution

$$\lambda_i = \lambda(\theta_i) + \frac{1}{\sqrt{N}} \mathcal{N}\big(0, V(\theta_i)\big) + o\left(\frac{1}{\sqrt{N}}\right), \tag{2.3}$$

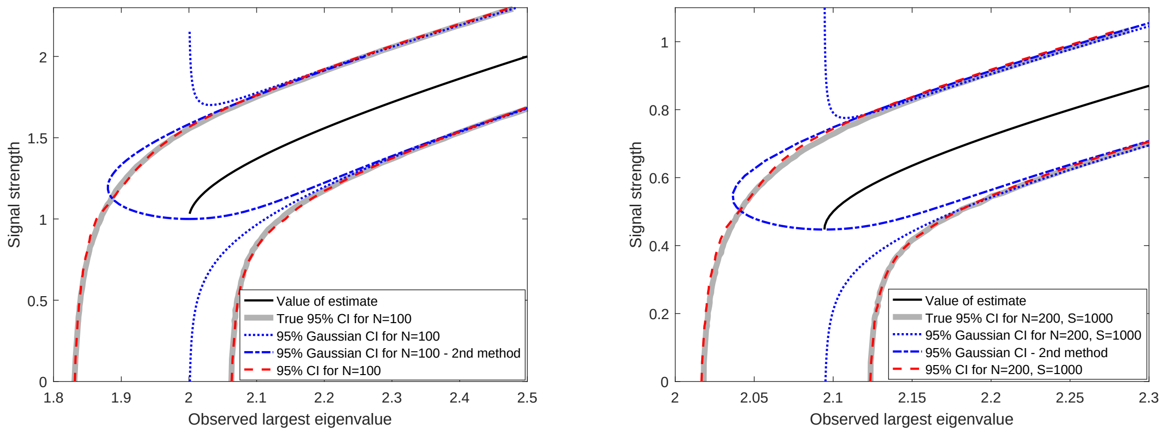
and the Gaussian limits $\mathcal{N}\big(0, V(\theta_i)\big)$ are independent over i. If $\theta_i \leq \theta^c$, then $\lim_{N\to\infty} \lambda_i = \lambda_+$, in probability.

Informally, the proposition says that "good" recovery of θ_i from the largest eigenvalues of $\mathbf{A}$ is possible if and only if θ_i is larger than the critical value $\theta^c = 1$. In this case, to estimate θ_i, one should take λ_i and apply the inverse of the mapping $\theta \mapsto \lambda(\theta)$, which is $\lambda \mapsto \frac{1}{2}\left(\lambda + \sqrt{\lambda^2 - 4}\right)$.

We assess the quality of estimating θ_i by constructing a confidence interval for it. Specifically, for each fixed i and significance level α we aim to find endpoints $\theta_i^-(\lambda_i, N, \alpha)$, $\theta_i^+(\lambda_i, N, \alpha)$ such that

$$\mathrm{Prob}\left(\theta_i \in [\theta_i^-(\lambda_i, N, \alpha), \theta_i^+(\lambda_i, N, \alpha)]\right) \approx 1 - \alpha, \tag{2.4}$$

[3]The prefactor $\frac{1}{\sqrt{2N}}$ in the definition of $\mathcal{E}$ ensures that its eigenvalues remain bounded and fill the interval $[-2, 2]$ as $N \to \infty$.



(A) Spiked Wigner matrix of Section 2.1 (B) Factor model of Section 2.3

FIGURE 1. Confidence intervals for θ as functions of the observed largest eigenvalue via Gaussian approximations and via our procedure of Section 3.

where $\approx$ denotes an $N \to \infty$ approximation, which should be uniform over the model parameters $\theta_1, \dots, \theta_r$ and $\mathbf{u}_1^*, \dots, \mathbf{u}_r^*$ in (2.1).

In principle, since we deal with multiple θ_i simultaneously, one could consider joint multi-dimensional confidence sets. However, due to the asymptotic independence of λ_i in (2.3), it is sufficient to construct separate intervals for each θ_i, which is the approach we take.[4]

The asymptotics (2.3) provides a way to construct confidence intervals by approximating θ_i in the argument of $V(\theta_i)$ with $\theta(\lambda_i) = \frac{1}{2}\left(\lambda_i + \sqrt{\lambda_i^2 - 4}\right)$ and then using Gaussian quantiles. This leads to the following formula for the confidence interval:

$$\theta_i \in \left[\frac{\lambda_i}{2} + \sqrt{\frac{\lambda_i^2}{4} - 1} - \frac{z_{\alpha/2}}{\sqrt{N}}\sqrt{1 + \frac{\lambda_i}{\sqrt{\lambda_i^2 - 4}}},\ \frac{\lambda_i}{2} + \sqrt{\frac{\lambda_i^2}{4} - 1} + \frac{z_{\alpha/2}}{\sqrt{N}}\sqrt{1 + \frac{\lambda_i}{\sqrt{\lambda_i^2 - 4}}}\right], \tag{2.5}$$

where $z_{\alpha/2}$ denotes the $\alpha/2$ quantile of $\mathcal{N}(0,1)$. E.g., to obtain a 95% confidence interval for a single fixed i, we set $z_{\alpha/2} = 1.96$.

The formula (2.5) reveals a problem as $\theta \to 1$ (i.e., $\lambda \to 2$): the confidence intervals diverge due to the $\sqrt{\lambda_i^2 - 4}$ singularity in the denominator. However, Monte Carlo simulations in Figure 1a indicate that no such explosion actually occurs. This suggests that the approximation error in the confidence interval (2.5) becomes non-negligible when λ_i is close to 2, making the formula unreliable in this regime. In contrast, our novel procedure, introduced in Section 3, closely matches the simulations across all values of λ_i.

Remark 2.1. *An alternative way to construct confidence intervals using Gaussian asymptotics is to rewrite* (2.3) *in the equivalent form*

$$\lambda_i \in \left[\theta_i + \frac{1}{\theta_i} - \frac{z_{\alpha/2}}{\sqrt{N}}\sqrt{2\frac{\theta_i^2 - 1}{\theta_i^2}} + o\left(\frac{1}{\sqrt{N}}\right),\quad \theta_i + \frac{1}{\theta_i} + \frac{z_{\alpha/2}}{\sqrt{N}}\sqrt{2\frac{\theta_i^2 - 1}{\theta_i^2}} + o\left(\frac{1}{\sqrt{N}}\right)\right].$$

[4]In contrast, if θ_i coincide, then the limits in (2.3) are neither Gaussian nor independent, cf. Capitaine et al. [2012, Theorem 3.3].

We drop $o\left(\frac{1}{\sqrt{N}}\right)$ terms, plot the intervals from the preceding formula on the (θ,λ)–plane, and then transpose the axes to obtain the desired confidence intervals on the (λ,θ)–plane; see the 2nd method in Figure 1a. For θ bounded away from 1 (equivalently, λ bounded away from 2), this procedure is equivalent to the intervals (2.5) as $N\to\infty$, though their finite-sample behavior differs near the cutoff. Compared to Monte Carlo intervals, both methods exhibit substantial bias, but in different directions.

2.2. **Spiked covariance model.** For the second setup we consider a deterministic $N\times N$ matrix

$$\Omega = \sigma^2 I_N + \sum_{i=1}^{r}(\theta_i-\sigma^2)\cdot \mathbf{u}_i^*(\mathbf{u}_i^*)^\mathsf{T} \tag{2.6}$$

where r is a fixed, small number, $\theta_1>\cdots>\theta_r>\sigma^2$ are the signal strengths, and $\mathbf{u}_1^*,\dots,\mathbf{u}_r^*$ are orthonormal N-dimensional vectors representing r signal directions. The eigenvalues of Ω are $\theta_1,\theta_2,\dots,\theta_r$, and σ^2 with multiplicity $(N-r)$. As before, we assume that σ^2 is known and set it to 1 without loss of generality; adjustments for unknown σ^2 are discussed in Section 3.3.

We observe an $N\times S$ data matrix X, whose columns are i.i.d. $\mathcal{N}(0,\Omega)$, and aim to estimate $\theta_1,\dots,\theta_r$ from the sample covariance matrix $\frac{1}{S}XX^\mathsf{T}$. This model has been central in statistics and random matrix theory since Johnstone [2001]; see Johnstone and Paul [2018] for a comprehensive overview, historical context, and many practical examples. Typically, the dimension S reflects multiple independent observations: across individuals, measurement points, time periods, etc. An exact analogue of (2.3) holds in this setting as well.

Proposition (Baik et al. [2005], Baik and Silverstein [2006], Paul [2007], Bai and Yao [2008])**.** *Suppose that $\sigma^2=1$ and $\theta_1>\theta_2>\cdots>\theta_r$ in (2.6). Let $\lambda_1\ge\lambda_2\cdots\ge\lambda_N$ denote the eigenvalues of $\frac{1}{S}XX^\mathsf{T}$ in (2.6). Assume*[5]*:*

$$\frac{N}{S}=\gamma^2+O\left(\frac{1}{N}\right),\quad N\to\infty,\qquad \gamma\in(0,1]. \tag{2.7}$$

Denote

$$\theta^c=1+\gamma,\qquad \lambda_+=(1+\gamma)^2,\qquad \lambda(\theta)=\theta+\frac{\gamma^2\theta}{\theta-1},\qquad V(\theta)=2\theta^2\gamma^2\left(1-\frac{\gamma^2}{(\theta-1)^2}\right). \tag{2.8}$$

For each $1\le i\le r$, if $\theta_i>\theta^c$, then as $N\to\infty$, in the sense of convergence in distribution

$$\lambda_i=\lambda(\theta_i)+\frac{1}{\sqrt{N}}\mathcal{N}\big(0,V(\theta_i)\big)+o\left(\frac{1}{\sqrt{N}}\right), \tag{2.9}$$

and the limits are independent over i. If $\theta_i\le\theta^c$, then $\lim_{N\to\infty}\lambda_i=\lambda_+$, in probability.

As in the previous section, we can use this Gaussian approximation to construct confidence intervals for each θ_i, yielding a modification of (2.5). However, this approach faces the same issue: the intervals become unreliable as λ_i approaches λ_+ and must be corrected using the procedures in Section 3.

[5] The case $\gamma>1$ can be also covered by similar methods.

2.3. **Factor model.** For the third setup we consider a random $N \times S$ matrix X defined by

$$X = \sum_{i=1}^{r} \sqrt{\theta_i S} \cdot \mathbf{u}_i^* (\mathbf{v}_i^*)^\mathsf{T} + \mathcal{E}, \tag{2.10}$$

where r is a fixed small number, $\theta_1, \dots, \theta_r > 0$ are the signal strengths, $\mathbf{u}_1^*, \dots, \mathbf{u}_r^*$ are N-dimensional orthonormal vectors of signal directions, called "loadings"[6], and $\mathbf{v}_1^*, \dots, \mathbf{v}_r^*$ are S-dimensional orthonormal vectors called "factors". The noise matrix $\mathcal{E}$ has independent $\mathcal{N}(0, \sigma^2)$ entries. For now, we assume σ^2 to be known and set it to 1; adjustments for unknown σ^2 are discussed in Section 3.3.

Our goal is to estimate $\theta_1, \dots, \theta_r$ from the eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_N$ of the sample covariance matrix $\frac{1}{S} X X^\mathsf{T}$. While the factor model has similarities to the spiked covariance model of the previous section, they are not equivalent, because we treat $\sqrt{\theta_i} \cdot \mathbf{u}_i^* (\mathbf{v}_i^*)^\mathsf{T}$ in (2.10) as deterministic parameters (the models would have been equivalent up to shift $\theta_i \to \theta_i + \sigma^2$, if each $\sqrt{S}\mathbf{v}_i^*$ were a mean 0 Gaussian vector with i.i.d. components). This distinction allows the factor model to capture complex structures along the S-dimension, which is essential in applications across finance, macroeconomics, natural sciences, and other fields. Once again, an analogue of (2.3) holds.

Proposition (Onatski [2012], Benaych-Georges and Nadakuditi [2012], Onatski [2018, Theorem 5])**.** *Suppose that $\sigma^2 = 1$ and $\theta_1 > \theta_2 > \dots > \theta_r$ in (2.10). Let $\lambda_1 \geq \lambda_2 \dots \geq \lambda_N$ denote the eigenvalues of $\frac{1}{S} X X^\mathsf{T}$. Assume[7]:*

$$\frac{N}{S} = \gamma^2 + O\left(\frac{1}{N}\right), \quad N \to \infty, \qquad \gamma \in (0, 1]. \tag{2.11}$$

Denote

$$\theta^c = \gamma, \quad \lambda_+ = (1+\gamma)^2, \quad \lambda(\theta) = (\theta+1)(1+\frac{\gamma^2}{\theta}), \quad V(\theta) = 2\gamma^2 \frac{(2\theta + 1 + \gamma^2)(\theta^2 - \gamma^2)}{\theta^2}. \tag{2.12}$$

For each $1 \leq i \leq r$, if $\theta_i > \theta^c$, then as $N \to \infty$, in the sense of convergence in distribution

$$\lambda_i = \lambda(\theta_i) + \frac{1}{\sqrt{N}} \mathcal{N}\big(0, V(\theta_i)\big) + o\left(\frac{1}{\sqrt{N}}\right), \tag{2.13}$$

and the limits are independent over i. If $\theta_i \leq \theta^c$, then $\lim_{N\to\infty} \lambda_i = \lambda_+$, in probability.

As in Section 2.1, the Gaussian approximation of λ_i leads to two methods for constructing confidence intervals. An analogue of (2.5) is

$$\theta_i \in \left[\theta(\lambda_i) - \frac{\sigma(\lambda_i)}{\sqrt{N}} z_{\alpha/2}, \theta(\lambda_i) + \frac{\sigma(\lambda_i)}{\sqrt{N}} z_{\alpha/2}\right],$$

where

$$\theta(\lambda) = \frac{\lambda - 1 - \gamma^2 + \sqrt{(1+\gamma^2-\lambda)^2 - 4\gamma^2}}{2}, \quad \sigma(\lambda) = \frac{\sqrt{2\gamma^2(2\theta(\lambda) + 1 + \gamma^2)(\theta(\lambda)^2 - \gamma^2)}}{\sqrt{(1+\gamma^2-\lambda)^2 - 4\gamma^2}}.$$

There is also a direct analogue of the second Gaussian method described in Remark 2.1. Figure 1b compares these two Gaussian-based intervals with our new approach, which we present in Section 3. The comparison reveals the same key features as in the spiked Wigner model.

[6]Sometimes $\{\sqrt{\theta_i}\mathbf{u}_i^*\}$ rather than $\{\mathbf{u}_i^*\}$ are referred to as loadings.

[7]Swapping the roles of N and S we also cover the case $\gamma > 1$.

2.4. **Canonical correlation analysis.** For the final setup we fix a small integer r and parameters $1 \geq \theta_1, \ldots, \theta_r \geq 0$. We consider a deterministic symmetric positive-definite $(N+M) \times (N+M)$ matrix Ω that satisfies

$$\begin{pmatrix} A & 0_{N\times M} \\ 0_{M\times N} & B \end{pmatrix} \Omega \begin{pmatrix} A^\mathsf{T} & 0_{N\times M} \\ 0_{M\times N} & B^\mathsf{T} \end{pmatrix} = \begin{pmatrix} I_N & \operatorname{diag}(\sqrt{\theta_1}, \ldots, \sqrt{\theta_r}) \\ \operatorname{diag}(\sqrt{\theta_1}, \ldots, \sqrt{\theta_r}) & I_M \end{pmatrix}, \tag{2.14}$$

where A and B are non-degenerate $N \times N$ and $M \times M$ matrices, respectively, I_N and I_M are identity matrices of $N \times N$ and $M \times M$ dimensions, respectively, and $\operatorname{diag}(\sqrt{\theta_1}, \ldots, \sqrt{\theta_r})$ is a rectangular matrix with $\sqrt{\theta_1}, \ldots, \sqrt{\theta_r}$ on the first r elements of the main diagonal and 0 everywhere else.

Let $\mathbf{x}$ be an $(N+M)$–dimensional Gaussian mean 0 random vector with covariance Ω, and let $\mathbf{u}$ and $\mathbf{v}$ denote its first N and last M coordinates, respectively. The parameters $\theta_1, \ldots, \theta_r$ are the squared canonical correlations between $\mathbf{u}$ and $\mathbf{v}$; see Bykhovskaya and Gorin [2024], as well as classical statistics references such as Thompson [1984], Gittins [1985], Anderson [2003], Muirhead [2009] for detailed introductions to canonical correlation analysis (CCA). Algorithmically, θ_i are the largest eigenvalues of the matrix $(\mathbb{E}\mathbf{u}\mathbf{u}^\mathsf{T})^{-1}\mathbb{E}\mathbf{u}\mathbf{v}^\mathsf{T}(\mathbb{E}\mathbf{v}\mathbf{v}^\mathsf{T})^{-1}\mathbb{E}\mathbf{v}\mathbf{u}^\mathsf{T}$.

Given S independent samples of $\mathbf{x}$, we construct two matrices: the $N \times S$ matrix $\mathbf{U}$ has S samples of $\mathbf{u}$ as its columns and the $M \times S$ matrix $\mathbf{V}$ has S samples of $\mathbf{v}$ as its columns. The sample squared canonical correlations $\lambda_1 \geq \lambda_2 \geq \ldots$ are the eigenvalues of the $N \times N$ matrix $(\mathbf{U}\mathbf{U}^\mathsf{T})^{-1}\mathbf{U}\mathbf{V}^\mathsf{T}(\mathbf{V}\mathbf{V}^\mathsf{T})^{-1}\mathbf{V}\mathbf{U}^\mathsf{T}$. Our goal is to estimate $\theta_1, \ldots, \theta_r$ from these observed eigenvalues.

In typical applications CCA is used to explore dependencies between two data sets, for example, two sets of individual characteristics, brain measurements versus behavioral scores, or two groups of stocks. The parameter θ_i quantify the strength of these dependencies. Once again, an analogue of (2.3) holds.

Proposition (Bao et al. [2019], Yang [2022b], Bai et al. [2022], Hou et al. [2023], Bykhovskaya and Gorin [2025])**.** *Suppose $\theta_1 > \theta_2 > \cdots > \theta_r$ in (2.14). Let $\lambda_1 \geq \lambda_2 \cdots \geq \lambda_N$ denote the sample squared canonical correlations. Assume*

$$\frac{S}{N} = \tau_N + O\left(\frac{1}{N}\right), \quad \frac{S}{M} = \tau_M + O\left(\frac{1}{N}\right), \quad N \to \infty, \qquad \tau_N, \tau_M > 1, \quad \tau_N^{-1} + \tau_M^{-1} < 1. \tag{2.15}$$

Denote

$$\theta^c = \frac{1}{\sqrt{(\tau_M - 1)(\tau_N - 1)}}, \qquad \lambda_+ = \left(\sqrt{\tau_M^{-1}(1-\tau_N^{-1})} + \sqrt{\tau_N^{-1}(1-\tau_M^{-1})}\right)^2,$$

$$\lambda(\theta) = \frac{\big((\tau_N - 1)\theta + 1\big)\big((\tau_M - 1)\theta + 1\big)}{\theta\tau_N\tau_M}, \tag{2.16}$$

$$V(\theta) = 2\frac{(1-\theta)^2}{\theta^2\tau_M^2\tau_N^3}\big(2(\tau_M - 1)(\tau_N - 1)\theta + \tau_M + \tau_N - 2\big)\big((\tau_M - 1)(\tau_N - 1)\theta^2 - 1\big).$$

For each $1 \leq i \leq r$, if $\theta_i > \theta^c$, then as $N \to \infty$, in the sense of convergence in distribution

$$\lambda_i = \lambda(\theta_i) + \frac{1}{\sqrt{N}}\mathcal{N}\big(0, V(\theta_i)\big) + o\left(\frac{1}{\sqrt{N}}\right), \tag{2.17}$$

and the limits are independent over i. If $\theta_i \leq \theta^c$, then $\lim_{N\to\infty} \lambda_i = \lambda_+$, in probability.

Remark 2.2. *The choice of $\frac{1}{\sqrt{N}}$ normalization introduces an asymmetry between M and N in the expression for the variance $V(\theta)$ in* (2.16)*.*

The same conclusion applies here: using (2.16) and Gaussian quantiles we can construct confidence intervals for θ_i that perform well when λ_i is bounded away from λ_+, but become inaccurate as λ_i approaches λ_+ and, therefore, require correction.

3. Construction of confidence intervals

In this section we present our algorithm for constructing confidence intervals and explain how they can be interpreted and used to distinguish between noise, non-informative signals, and meaningful signals. We begin by introducing the transition process $\mathcal{T}(\Theta)$ and its properties, and then show how to use it to construct confidence intervals. The underlying theorems will be presented in Section 4.

3.1. **Transition process.** As highlighted in Figure 1, the Gaussian limits in (2.3), (2.9), (2.13), and (2.17) ought to be replaced by a different limiting object, which we call the *transition process* $\mathcal{T}(\Theta)$. This is a random function of $\Theta \in \mathbb{R}$. Its formal definition is provided in Section 4.1, while for the purposes of constructing confidence intervals, the key quantities of interest are the quantiles of its distribution, which may be computed as follows:

- For $-3 \le \Theta \le 6$, quantiles are tabulated in Table 1 using the algorithm described in Section 4.1.
- For large positive values of Θ, the Gaussian approximation $\mathcal{T}(\Theta) \approx \mathcal{N}(\Theta^2, 4\Theta)$ should be used, i.e.,
$$\mathbb{P}\{\mathcal{T}(\Theta) \le t\} \approx \Phi((t-\Theta^2)/(2\sqrt{\Theta}))\ .$$
- For large negative values of Θ, the Tracy–Widom$_1$ approximation should be used:
$$\mathbb{P}\{\mathcal{T}(\Theta) \le t\} \approx F_1(t + 1/\Theta)\ ,$$
where the relevant Tracy–Widom quantiles are provided in Table 2.

The transition process $\mathcal{T}(\Theta)$, with appropriate centering and scaling, can be used to approximate the fluctuations of the largest eigenvalues, leading to the following algorithm.

Procedure 3.1. *For each of the four models in Section 2 with $\sigma^2 = 1$, the asymptotic distribution of the largest eigenvalues λ_i can be approximated as:*

$$\begin{cases} \lambda(\theta_i) - \frac{\kappa_2^{3/2}}{2}\sqrt{V(\theta_i)(\theta_i-\theta^c)^3} + \frac{\kappa_2^{-1/2}}{2N^{2/3}}\sqrt{\frac{V(\theta_i)}{\theta_i-\theta^c}}\,\mathcal{T}\Big(\kappa_2 N^{1/3}(\theta_i-\theta^c)\Big) + \frac{\kappa_3}{N}, & \text{if } \theta_i > \theta^c, \\ \lambda_+ + N^{-2/3}\kappa_1 \mathcal{T}\Big(\kappa_2 N^{1/3}(\theta_i-\theta^c)\Big) + \frac{\kappa_3}{N}, & \text{if } \theta_i \le \theta^c, \end{cases} \tag{3.1}$$

where the constants are taken from (2.2), (2.8), (2.12), (2.16)*; $\kappa_1 = \frac{1}{2}\frac{[V'(\theta^c)]^{2/3}}{[\lambda''(\theta^c)]^{1/3}}$, $\kappa_2 = \frac{[\lambda''(\theta^c)]^{2/3}}{[V'(\theta^c)]^{1/3}}$, $\kappa_3 = -\frac{3}{2}\frac{\kappa_1}{\kappa_2\theta^c}$, and we assume $\theta_{i-1} > \theta^c$.*

Theorem 4.6 and Corollary 4.8 establish that the approximation (3.1) is valid both when θ is bounded away from the critical value θ^c and when θ is close to θ^c. These results also show that, in many cases, the approximations for different λ_i are asymptotically independent. Hence, we can use (3.1) as a foundation for constructing confidence intervals.

$\Theta \backslash \alpha$	.005	.025	.05	.5	.95	.975	.995
-3.0	-3.85	-3.22	-2.89	-0.96	1.32	1.80	2.78
-2.9	-3.85	-3.22	-2.88	-0.95	1.32	1.80	2.79
-2.8	-3.84	-3.20	-2.86	-0.94	1.33	1.81	2.80
-2.7	-3.83	-3.20	-2.86	-0.93	1.35	1.83	2.82
-2.6	-3.82	-3.19	-2.85	-0.92	1.36	1.85	2.83
-2.5	-3.82	-3.18	-2.84	-0.91	1.38	1.86	2.83
-2.4	-3.80	-3.17	-2.83	-0.89	1.38	1.87	2.85
-2.3	-3.80	-3.16	-2.82	-0.89	1.42	1.91	2.90
-2.2	-3.79	-3.15	-2.82	-0.87	1.42	1.90	2.86
-2.1	-3.77	-3.13	-2.79	-0.85	1.44	1.93	2.94
-2.0	-3.75	-3.12	-2.78	-0.84	1.46	1.95	2.97
-1.9	-3.75	-3.11	-2.77	-0.82	1.49	1.98	2.98
-1.8	-3.74	-3.10	-2.75	-0.80	1.52	2.01	3.03
-1.7	-3.73	-3.09	-2.74	-0.78	1.54	2.03	3.05
-1.6	-3.71	-3.06	-2.72	-0.76	1.57	2.07	3.11
-1.5	-3.69	-3.05	-2.70	-0.74	1.61	2.11	3.14
-1.4	-3.67	-3.03	-2.69	-0.71	1.64	2.14	3.19
-1.3	-3.65	-3.01	-2.66	-0.69	1.68	2.19	3.26
-1.2	-3.64	-2.99	-2.65	-0.66	1.72	2.23	3.27
-1.1	-3.61	-2.97	-2.63	-0.63	1.77	2.29	3.37
-1.0	-3.58	-2.94	-2.59	-0.59	1.83	2.35	3.44
-0.9	-3.57	-2.91	-2.56	-0.56	1.88	2.41	3.52
-0.8	-3.55	-2.89	-2.54	-0.53	1.94	2.48	3.62
-0.7	-3.53	-2.87	-2.51	-0.48	2.01	2.57	3.70
-0.6	-3.51	-2.84	-2.49	-0.44	2.08	2.64	3.79
-0.5	-3.49	-2.82	-2.46	-0.39	2.17	2.74	3.93
-0.4	-3.45	-2.77	-2.42	-0.35	2.27	2.85	4.10
-0.3	-3.41	-2.74	-2.38	-0.28	2.38	2.97	4.25
-0.2	-3.37	-2.70	-2.33	-0.22	2.48	3.09	4.39
-0.1	-3.35	-2.66	-2.30	-0.15	2.61	3.24	4.59
0.0	-3.31	-2.62	-2.25	-0.08	2.75	3.40	4.76
0.1	-3.26	-2.57	-2.20	0.00	2.91	3.57	4.96
0.2	-3.22	-2.52	-2.14	0.09	3.08	3.77	5.24
0.3	-3.17	-2.47	-2.09	0.19	3.25	3.95	5.45
0.4	-3.13	-2.40	-2.02	0.29	3.46	4.19	5.72
0.5	-3.08	-2.34	-1.96	0.41	3.67	4.42	5.99
0.6	-3.01	-2.28	-1.89	0.54	3.90	4.68	6.29
0.7	-2.96	-2.21	-1.80	0.69	4.17	4.96	6.60
0.8	-2.90	-2.14	-1.73	0.83	4.41	5.23	6.91
0.9	-2.81	-2.04	-1.62	1.01	4.73	5.57	7.30
1.0	-2.71	-1.95	-1.53	1.18	5.02	5.88	7.62
1.1	-2.63	-1.86	-1.42	1.39	5.37	6.23	8.01
1.2	-2.56	-1.75	-1.31	1.60	5.69	6.59	8.38
1.3	-2.47	-1.64	-1.18	1.84	6.06	6.96	8.80
1.4	-2.37	-1.52	-1.05	2.10	6.44	7.37	9.22
1.5	-2.25	-1.38	-0.89	2.37	6.82	7.79	9.72

$\Theta \backslash \alpha$	.005	.025	.05	.5	.95	.975	.995
1.5	-2.25	-1.38	-0.89	2.37	6.82	7.79	9.72
1.6	-2.14	-1.23	-0.73	2.67	7.27	8.22	10.19
1.7	-2.03	-1.08	-0.55	3.00	7.67	8.65	10.66
1.8	-1.88	-0.91	-0.37	3.34	8.12	9.13	11.19
1.9	-1.73	-0.72	-0.14	3.70	8.59	9.62	11.71
2.0	-1.56	-0.53	0.07	4.09	9.09	10.14	12.20
2.1	-1.39	-0.30	0.34	4.50	9.61	10.66	12.84
2.2	-1.21	-0.05	0.62	4.91	10.12	11.20	13.31
2.3	-0.99	0.22	0.92	5.37	10.67	11.75	13.96
2.4	-0.77	0.51	1.24	5.84	11.24	12.36	14.52
2.5	-0.53	0.82	1.59	6.32	11.82	12.94	15.14
2.6	-0.24	1.16	1.96	6.82	12.45	13.57	15.88
2.7	0.04	1.52	2.37	7.35	13.05	14.22	16.52
2.8	0.36	1.91	2.78	7.90	13.68	14.85	17.18
2.9	0.73	2.34	3.23	8.47	14.35	15.55	17.90
3.0	1.08	2.77	3.70	9.04	15.02	16.21	18.60
3.1	1.52	3.27	4.22	9.67	15.71	16.92	19.38
3.2	1.93	3.72	4.72	10.28	16.42	17.63	20.07
3.3	2.41	4.25	5.26	10.93	17.15	18.40	20.85
3.4	2.93	4.79	5.82	11.61	17.91	19.18	21.74
3.5	3.40	5.36	6.41	12.29	18.65	19.93	22.48
3.6	3.96	5.95	7.03	13.01	19.48	20.78	23.37
3.7	4.52	6.59	7.68	13.72	20.26	21.56	24.14
3.8	5.12	7.18	8.30	14.48	21.13	22.47	25.05
3.9	5.76	7.89	9.01	15.25	21.95	23.31	25.98
4.0	6.38	8.57	9.72	16.05	22.83	24.17	26.87
4.1	7.04	9.25	10.41	16.85	23.70	25.06	27.75
4.2	7.70	9.93	11.15	17.68	24.61	25.96	28.68
4.3	8.41	10.69	11.91	18.52	25.55	26.94	29.71
4.4	9.14	11.49	12.73	19.41	26.49	27.88	30.62
4.5	9.93	12.27	13.52	20.28	27.43	28.84	31.60
4.6	10.68	13.09	14.35	21.20	28.43	29.81	32.63
4.7	11.46	13.91	15.19	22.12	29.41	30.84	33.67
4.8	12.23	14.76	16.07	23.06	30.44	31.88	34.68
4.9	13.12	15.63	16.94	24.03	31.52	32.99	35.90
5.0	13.97	16.54	17.87	25.04	32.55	34.02	36.98
5.1	14.88	17.44	18.80	26.04	33.64	35.13	38.05
5.2	15.76	18.39	19.73	27.06	34.74	36.22	39.18
5.3	16.70	19.36	20.71	28.11	35.83	37.34	40.37
5.4	17.67	20.35	21.75	29.19	37.01	38.52	41.57
5.5	18.57	21.35	22.75	30.28	38.14	39.70	42.71
5.6	19.63	22.36	23.78	31.40	39.34	40.91	43.98
5.7	20.67	23.43	24.85	32.54	40.53	42.08	45.15
5.8	21.65	24.46	25.91	33.67	41.72	43.31	46.43
5.9	22.78	25.52	27.00	34.83	42.97	44.54	47.70
6.0	23.82	26.63	28.11	36.03	44.25	45.86	48.95

TABLE 1. Quantiles of $\mathcal{T}(\Theta)$ for $-3 \leq \Theta \leq 6$ based on $MC = 10^6$ Monte Carlo simulations.

α	.005	.025	.05	.5	.95	.975	.995
quantile $F_1^{-1}(\alpha)$	-4.15	-3.52	-3.18	-1.27	0.98	1.45	2.42

TABLE 2. Quantiles of the Tracy–Widom$_1$ distribution from Bejan [2005].

Spiked Wigner matrix	**Spiked covariance model**
$\frac{1}{2\pi}\sqrt{4-x^2}\,\mathbf{1}_{[-2,2]}\,\mathrm{d}x$ *Semicircle law*	$\frac{1}{2\pi}\frac{\sqrt{(\lambda_+-x)(x-\lambda_-)}}{\gamma^2 x}\mathbf{1}_{[\lambda_-,\lambda_+]}\,\mathrm{d}x$ *Marchenko-Pastur law*
Factor model	**CCA**
$\frac{1}{2\pi}\frac{\sqrt{(\lambda_+-x)(x-\lambda_-)}}{\gamma^2 x}\mathbf{1}_{[\lambda_-,\lambda_+]}\,\mathrm{d}x$ *Marchenko-Pastur law*	$\frac{\tau_N}{2\pi}\frac{\sqrt{(\lambda_+-x)(x-\lambda_-)}}{x(1-x)}\mathbf{1}_{[\lambda_-,\lambda_+]}\,\mathrm{d}x$ *Wachter law*

TABLE 3. Limiting behavior of the empirical measures of eigenvalues, $\lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}$, in signal plus noise models.

3.2. Confidence intervals with known σ^2.

We begin with the case where the noise variance σ^2 is known, as specified in Section 2. A simple rescaling allows us to assume $\sigma^2 = 1$ without loss of generality. The algorithm then proceeds as follows:

The first step is to draw a histogram of all eigenvalues $\lambda_1, \lambda_2, \dots$. In the settings of (2.1), (2.6), (2.10), or (2.14), the histogram should closely resemble a known limiting shape; namely, the semicircle law, Marchenko-Pastur law, or Wachter law, depending on the model, as detailed in Table 3, with parameters specified in Table 4, see Appendix 8.4 for more details. If the histogram is reminiscent of one of these shapes, we regard the modelling assumptions as valid and apply Procedure 3.1 to construct confidence intervals. Section 6.2 discusses possible extensions when the empirical histogram deviates from the expected limit shape.

For the second step, we choose a significance level α (or confidence level $1-\alpha$) and, using Section 3.1, construct two deterministic functions $t_{\alpha/2,+}(\Theta)$ and $t_{\alpha/2,-}(\Theta)$ such that

$$\mathrm{Prob}\big(\mathcal{T}(\Theta) > t_{\alpha/2,+}(\Theta)\big) = \mathrm{Prob}\big(\mathcal{T}(\Theta) < t_{\alpha/2,-}(\Theta)\big) = \frac{\alpha}{2}. \tag{3.2}$$

Following (3.1) and using the parameter choices from Table 4, we rescale the functions $t_{\alpha/2,\pm}(\Theta)$ to obtain $\widehat{t}_\pm(\theta)$, defined as

$$\widehat{t}_\pm(\theta) = \begin{cases} \lambda(\theta) - \frac{\kappa_2^{3/2}}{2}\sqrt{V(\theta)(\theta-\theta^c)^3} + \frac{\kappa_2^{-1/2}}{2N^{2/3}}\sqrt{\frac{V(\theta)}{\theta-\theta^c}}\,t_{\alpha/2,\pm}\Big(\kappa_2 N^{1/3}(\theta-\theta^c)\Big) + \frac{\kappa_3}{N}, & \theta > \theta^c, \\ \lambda_+ + \kappa_1 N^{-2/3} t_{\alpha/2,\pm}\Big(\kappa_2 N^{1/3}(\theta-\theta^c)\Big) + \frac{\kappa_3}{N}, & \theta \le \theta^c. \end{cases} \tag{3.3}$$

For the third step, we fix an index i and consider the ith largest eigenvalue λ_i, such that $\lambda_i > \lambda_+$. We then determine two numbers $\theta_- < \theta_+$ such that

$$\widehat{t}_+(\theta_-) = \widehat{t}_-(\theta_+) = \lambda_i. \tag{3.4}$$

Visually, the procedure amounts to plotting the functions $\theta \mapsto \widehat{t}_\pm(\theta)$ and finding their intersection with the horizontal line $y = \lambda_i$. The resulting interval $[\theta_-, \theta_+]$ serves as the confidence

	Spiked Wigner matrix	**Spiked covariance model**
Parameters	–	$\gamma^2 = \frac{N}{S} \in (0,1]$
θ^c	1	$1+\gamma$
$\lambda_\pm$	± 2	$(1\pm\gamma)^2$
$\lambda(\theta)$	$\theta + \frac{1}{\theta}$	$\theta + \gamma^2 \frac{\theta}{\theta-1}$
$V(\theta)$	$2\frac{\theta^2-1}{\theta^2}$	$2\theta^2\gamma^2\left(1-\frac{\gamma^2}{(\theta-1)^2}\right)$
$\kappa_1,\ \kappa_2,\ \kappa_3$	$1,\ 1,\ -\frac{3}{2}$	$\gamma(1+\gamma)^{4/3},\ \frac{1}{\gamma(1+\gamma)^{2/3}},\ -\frac{3}{2}\gamma(1+\gamma)^2$
	Factor model	**CCA**
Parameters	$\gamma^2 = \frac{N}{S} \in (0,1]$	$\tau_N = \frac{S}{N} > 1$, $\tau_M = \frac{S}{M} > 1$, with $\tau_N^{-1} + \tau_M^{-1} < 1, \quad \tau_N \geq \tau_M$
θ^c	γ	$\frac{1}{\sqrt{(\tau_M-1)(\tau_N-1)}}$
$\lambda_\pm$	$(1\pm\gamma)^2$	$\left(\sqrt{\tau_M^{-1}(1-\tau_N^{-1})} \pm \sqrt{\tau_N^{-1}(1-\tau_M^{-1})}\right)^2$
$\lambda(\theta)$	$\theta + 1 + \gamma^2\frac{\theta+1}{\theta}$	$\frac{\big((\tau_N-1)\theta+1\big)\big((\tau_M-1)\theta+1\big)}{\theta\tau_N\tau_M}$
$V(\theta)$	$2\gamma^2\frac{(2\theta+1+\gamma^2)(\theta^2-\gamma^2)}{\theta^2}$	$2\frac{(1-\theta)^2}{\theta^2\tau_M^2\tau_N^3}\big(2(\tau_M-1)(\tau_N-1)\theta+\tau_M+\tau_N-2\big) \times \big((\tau_M-1)(\tau_N-1)\theta^2-1\big)$
$\kappa_1,\ \kappa_2,\ \kappa_3$	$\gamma(1+\gamma)^{4/3},\quad \frac{1}{\gamma(1+\gamma)^{2/3}},\quad -\frac{3}{2}\gamma(1+\gamma)^2$	$\frac{(\sqrt{\tau_N-1}\sqrt{\tau_M-1}-1)^{4/3}(\sqrt{\tau_N-1}+\sqrt{\tau_M-1})^{4/3}}{\tau_N^{5/3}\tau_M(\tau_N-1)^{1/6}(\tau_M-1)^{1/6}}$, $\frac{\tau_N^{1/3}(\tau_N-1)^{5/6}(\tau_M-1)^{5/6}}{(\sqrt{\tau_N-1}\sqrt{\tau_M-1}-1)^{2/3}(\sqrt{\tau_N-1}+\sqrt{\tau_M-1})^{2/3}}$, $-\frac{3}{2}\frac{(\sqrt{\tau_N-1}\sqrt{\tau_M-1}-1)^2(\sqrt{\tau_N-1}+\sqrt{\tau_M-1})^2}{\tau_N^2\tau_M\sqrt{\tau_N-1}\sqrt{\tau_M-1}}$

TABLE 4. Parameters. In the factor model the roles of S and N can be swapped when $\gamma^2 > 1$. In CCA N and M can be swapped when $\tau_N < \tau_M$.

interval for the ith largest signal strength θ_i. Corollary 4.8 ensures that as $N \to \infty$ we have $\mathrm{Prob}(\theta_i \in [\theta_-, \theta_+]) \to 1-\alpha$.

There are two special cases to consider at this step. First, it may happen that no value θ_- satisfies $\widehat{t}_+(\theta-) = \lambda_i$. This occurs when the shifted and rescaled λ_i falls below the $(1-\alpha/2)$ quantile of the Tracy-Widom distribution F_1. In this case the confidence interval becomes one-sided, and one should set $\theta_- = -\infty$ or, equivalently, to the lower bound of admissible values of θ_i, that is $\theta_i \geq \sigma^2$ for the spiked covariance and $\theta_i \geq 0$ for the others. Second, it may happen that θ_- exists, but lies below the lower bound for admissible values of θ_i. In this case θ_- should again be replaced by the appropriate lower bound. In terms of statistical consequences the two cases are equivalent.

3.3. **Unknown variance.** For the CCA setting in Section 2.4 the asymptotics in Theorem 4.6 do not depend on the noise covariance, i.e., the matrices A and B in (2.14). In contrast, for the three other settings, Sections 2.1, 2.2, and 2.3, the scaling depends on the noise variance, denoted by σ^2. Procedure 3.1 assumes $\sigma^2 = 1$. If $\sigma^2 \neq 1$ but is known, then the entries of the data matrix $\mathbf{A}$ or X should be divided by σ to reduce to the baseline case $\sigma^2 = 1$. If σ^2 is unknown, it must first be estimated.

We propose estimating the variance by discarding 25% of the eigenvalues from both the lower and upper ends, and matching sample moments to their theoretical values in order to solve for σ^2.

For the spiked Wigner model, let $\ell \approx 0.81$ denote the positive number such that

$$\int_{-2}^{-\ell} \frac{1}{2\pi}\sqrt{4-x^2}\mathrm{d}x = \int_{\ell}^{2} \frac{1}{2\pi}\sqrt{4-x^2}\mathrm{d}x = \frac{1}{4}, \tag{3.5}$$

and set

$$\sigma_0^2 = \int_{-\ell}^{\ell} \frac{x^2}{2\pi}\sqrt{4-x^2}\mathrm{d}x. \tag{3.6}$$

Given eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N$ of $\mathbf{A}$, we can form an estimate

$$\widehat{\sigma}^2 = \frac{1}{\sigma_0^2}\frac{1}{N}\sum_{i=\lfloor N/4\rfloor+1}^{\lfloor 3N/4\rfloor} \lambda_i^2. \tag{3.7}$$

The Wigner semicircle law for the GOE with explicit estimates for the remainders (see e.g., O'Rourke [2010]), combined with the interlacing inequalities between the eigenvalues of $\mathbf{A}$ and $\mathbf{B}$ in (2.1), as in Corollary 9.3, can be used to show that

$$\widehat{\sigma}^2 = \sigma^2 + O\left(\frac{\log(N)}{N}\right), \qquad N \to \infty. \tag{3.8}$$

Note that the scale of the random component in Theorem 4.6 is much larger than the error term in (3.8). As a result, our confidence intervals are much wider than this error term and normalizing the data by $\widehat{\sigma}$ does not change the validity of the confidence intervals constructed in the previous section.

For the spiked covariance and factor models, the procedure is analogous, but relies on the Marchenko-Pastur law (see Table 3) rather than the semicircle law. Fixing the parameter $\gamma^2 = \frac{N}{S} \in [0,1)$, we define ℓ_- and ℓ_+ as two positive numbers such that

$$\int_{\lambda_-}^{\ell_-} \frac{1}{2\pi}\frac{\sqrt{(\lambda_+ - x)(x-\lambda_-)}}{\gamma^2 x}\mathrm{d}x = \frac{1}{4}, \qquad \int_{\ell_+}^{\lambda_+} \frac{1}{2\pi}\frac{\sqrt{(\lambda_+ - x)(x-\lambda_-)}}{\gamma^2 x}\mathrm{d}x = \frac{1}{4}. \tag{3.9}$$

and set

$$\sigma_0^2 = \int_{\ell_1}^{\ell_2} \frac{x}{2\pi}\frac{\sqrt{(\lambda_+ - x)(x-\lambda_-)}}{\gamma^2 x}\mathrm{d}x. \tag{3.10}$$

Given eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N$ of $\frac{1}{S}XX^\mathsf{T}$, we can form an estimate

$$\widehat{\sigma}^2 = \frac{1}{\sigma_0^2}\frac{1}{N}\sum_{i=\lfloor N/4\rfloor+1}^{\lfloor 3N/4\rfloor} \lambda_i. \tag{3.11}$$

The Marchenko–Pastur law with explicit estimates for the reminders (see e.g., Bourgade et al. [2022]), combined with the interlacing inequalities between the eigenvalues of spiked and unspiked models, as in Corollaries 9.8 and 9.12, can be used to show that

$$\widehat{\sigma}^2 = \sigma^2 + O\left(\frac{\log(N)}{N}\right), \qquad N \to \infty. \tag{3.12}$$

Once again, normalizing the data by dividing by $\widehat{\sigma}$ does not affect the validity of the confidence intervals constructed in the previous section.

For a discussion of alternative procedures for estimating σ^2 see, for example, Kritchman and Nadler [2009, Section III.C], Shabalin and Nobel [2013, Section 4.1], Gavish and Donoho [2014, Section 3.E], or Ke et al. [2023, End of Section 2].

3.4. **Implications and interpretations.** Confidence intervals serve two important roles in the analysis of signal strength. First, they provide a measure of uncertainty: the narrower the confidence interval, the more precisely the signal strength is estimated.

Second, confidence intervals help assess the informativeness of estimated signals. If the lower bound of a confidence interval starts at $-\infty$ or at the minimal admissible value of θ, (which is σ^2 for the spiked covariance and zero for other models), then the signal may be spurious and could reflect pure noise – that is, a situation in which no true signal is present. Alternatively, if the confidence interval is bounded away from the minimal admissible value, but contains the identification threshold θ^c, then we know that a signal exists, but we cannot reject the null hypothesis that its strength falls below the identification cutoff. When $\theta \leq \theta^c$ the sample estimates of the signal directions $\mathbf{u}$ and $\mathbf{v}$ are asymptotically orthogonal to their true population counterparts (see, e.g., Paul [2007], Onatski [2012], Benaych-Georges and Nadakuditi [2012], Johnstone and Paul [2018], Bykhovskaya and Gorin [2025]), rendering the signal effectively non-informative.

4. Asymptotics through the Airy–Green function

The new asymptotics, which improves upon the Gaussian approximations (2.3), (2.9), (2.13), and (2.17), is based on a novel stochastic object we call the *Airy–Green function*. Its definition, along with the transition process $\mathcal{T}(\Theta)$ constructed from it, is presented in Section 4.1. Further discussion of its nature is provided in Section 4.2. Theorem on the convergence of the eigenvalue distributions in four signal plus noise models towards this object is stated in Section 4.3.

4.1. **Definition of $\mathcal{G}(w)$ and $\mathcal{T}(\Theta)$.** We begin by recalling the Airy$_1$ point process, a random sequence of points $\mathfrak{a}_1 \geq \mathfrak{a}_2 \geq \mathfrak{a}_3 \geq \dots$, which can be defined as the scaling limit of the largest eigenvalues of Wigner matrices.

Proposition (Forrester [1993], Tracy and Widom [1996]). *Let Y_N be an $N \times N$ matrix of i.i.d. $\mathcal{N}(0, \frac{2}{N})$ Gaussian random variables and let $\lambda_{1;N} \geq \lambda_{2;N} \geq \cdots \geq \lambda_{N;N}$ be the eigenvalues of $\mathbf{B} = \frac{1}{2}\left(Y_N + Y_N^{\mathsf{T}}\right)$. Then in the sense of convergence of finite-dimensional distributions*

$$\lim_{N\to\infty} \left\{N^{2/3}\left(\lambda_{i;N} - 2\right)\right\}_{i=1}^{N} = \{\mathfrak{a}_i\}_{i=1}^{\infty}. \tag{4.1}$$

Similar asymptotic results hold for the other models we consider, see further below. All existing formulae for the finite-dimensional distributions of $\{\mathfrak{a}_i\}_{i=1}^{\infty}$ are quite complicated and do not provide explicit distribution function, see, e.g., Forrester [2010]. Nevertheless, the distribution can be sampled and tabulated, see Bornemann [2009], Bykhovskaya et al. [2024].

In particular, Table 2 lists quantiles of the Tracy–Widom distribution, which describes the law of $\mathfrak{a}_1$.

The following theorem defines the *Airy–Green function* $\mathcal{G}(w)$; see Section 8.1 for the proof.

Theorem 4.1. *Let $\mathfrak{a}_1 \geq \mathfrak{a}_2 \geq \mathfrak{a}_3 \geq \dots$ be a realization of the Airy$_1$ point process and let $\{\xi_j\}_{j=1}^{\infty}$ be i.i.d. $\mathcal{N}(0,1)$ independent of $\{\mathfrak{a}_j\}_{j=1}^{\infty}$. Almost surely, for each $w \in \mathbb{C} \setminus \{\mathfrak{a}_j\}$ there exists a (random) limit*

$$\mathcal{G}(w) = \lim_{x \to -\infty} \left[\left(\sum_{j:\, \mathfrak{a}_j > x} \frac{\xi_j^2}{w - \mathfrak{a}_j} \right) - \frac{2}{\pi}\sqrt{-x} \right], \tag{4.2}$$

and, moreover, the convergence is uniform on any compact set $W \subset \mathbb{C}$ disjoint from $\{\mathfrak{a}_j\}$.

Note that any fixed $w \in \mathbb{C}$ is almost surely not in $\{\mathfrak{a}_j\}$, hence, for such w the convergence holds almost surely. Eq. (4.2) and Proposition 8.3 imply that $\mathcal{G}(w)$ changes monotonically from $+\infty$ to $-\infty$ over the interval $[\mathfrak{a}_1, +\infty)$, allowing us to state the following key definition.

Definition 4.2. *The* transition process $\mathcal{T}(\Theta)$, $\Theta \in \mathbb{R}$, *is a random function, defined as the unique solution to the equation $\mathcal{G}(w) = -\Theta$ satisfying $w \in [\mathfrak{a}_1, +\infty)$.*

Proposition 4.3. *Almost surely, $\Theta \mapsto \mathcal{T}(\Theta)$ is an increasing bijection of $\mathbb{R}$ onto $(\mathfrak{a}_1, \infty)$. As $\Theta \to +\infty$, $\mathcal{T}(\Theta)$ is asymptotically Gaussian:*

$$\lim_{\Theta \to +\infty} \frac{\mathcal{T}(\Theta) - \Theta^2}{2\sqrt{\Theta}} \stackrel{d}{=} \mathcal{N}(0,1). \tag{4.3}$$

Remark 4.4. *For large negative Θ we have distributional approximations:*

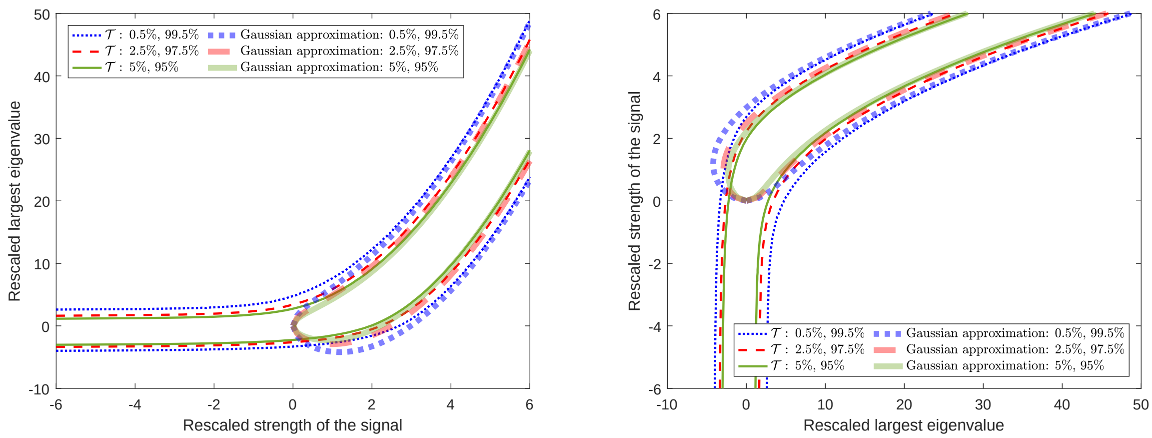
$$\mathcal{T}(\Theta) \stackrel{d}{=} \mathfrak{a}_1 - \frac{\xi_1^2}{\Theta} + O\left(\frac{1}{\Theta^2}\right) \stackrel{d}{=} \mathfrak{a}_1 - \frac{1}{\Theta} + O\left(\frac{1}{\Theta^2}\right), \qquad \Theta \to -\infty. \tag{4.4}$$

The first approximation follows directly from (4.2)*; the second from writing the distribution function as the expectation of the distribution function of $\mathfrak{a}_1$ shifted by random $\frac{\xi_1^2}{\Theta}$.*

Figure 2 shows the simulated quantiles for the random variables $\mathcal{T}(\Theta)$ as functions of Θ, or equivalently, confidence intervals for Θ as a function of $\mathcal{T}$. The underlying data is given in Table 1. These results are based on $MC = 10^6$ Monte Carlo simulations of the $\sqrt{N} \times \sqrt{N}$ top-left corners of $N \times N$ tridiagonal matrices of Dumitriu and Edelman [2002], with a perturbed $(1,1)$ matrix element and $N = 10^8$; see Edelman and Persson [2005, Section 1.1] and Johnstone et al. [2021, Lemma 5.2] for justifications of this approach. The figure shows that the Gaussian approximation from Proposition 4.3 performs very well for large Θ, but deteriorates near $\Theta = 0$.

4.2. **Discussion of the definition.** Let us clarify the terminology. The term "Airy" in the name $\mathcal{G}(w)$ refers to the Airy point process, whose points $\mathfrak{a}_i$ appear in its definition. The term "Green" stands from the tradition in random matrix theory to refer to matrix elements of the resolvent $(zI - D)^{-1}$ of a symmetric matrix D as the Green's function. Via eigenvalue decomposition, the $(1,1)$ matrix element of $(zI - D)^{-1}$ is $\sum_i \frac{u_{1i}^2}{z - d_i}$, where u_{1i} is the first coordinate of the ith normalized eigenvector of D corresponding to the eigenvalue d_i, making it reminiscent of the sum in (4.2).

The term "transition" in the name of $\mathcal{T}(\Theta)$ refers to its role in capturing the transition between subcritical $\theta < \theta^c$ and supercritical $\theta > \theta^c$ behavior in (2.3), (2.9), (2.13), (2.17).



(A) Quantiles of $\mathcal{T}(\Theta)$ as functions of Θ. (B) Confidence intervals for Θ (transposed axes).

FIGURE 2. Quantiles of $\mathcal{T}(\Theta)$ from Corollary 4.8 and from the Gaussian approximations based on Proposition 4.3.

This phenomenon is commonly known as the BBP phase transition, following Baik et al. [2005].

There are two other approaches to the transition process $\mathcal{T}(\Theta)$ in the literature. One is based on the limit of tridiagonal matrix model: Bloemendal and Virág [2013], Lamarre and Shkolnikov [2019] construct $\mathcal{T}(\Theta)$ as the largest eigenvalue of the Stochastic Airy Operator with Θ–dependent boundary condition. Another approach, developed by Mo [2012] using the framework of Pfaffian point processes, provides an integral representation for the one-dimensional marginal distribution of $\mathcal{T}(\Theta)$.

An advantage of our definition via the Airy–Green function is its robustness. Proving convergence to either of the two alternative definitions, requires finding delicate algebraic structures (tridiagonalization or Pfaffians) in the prelimit objects, which are not known in some cases (e.g., CCA). In contrast, our approach relies only on identifying the eigenvalues of a spiked model as solutions to an equation, that can be obtained in all spiked models via finite-rank perturbation theory.

Remark 4.5. *One can go beyond real matrices, and deal with complex, quaternionic, or even general β random matrix ensembles. In the latter setting the definition of the Airy–Green function should be extended to*

$$\mathcal{G}_\beta(w) = \lim_{x\to-\infty}\left[\left(\sum_{j:\,\mathfrak{a}_{j,\beta}>x}\frac{\beta^{-1}\xi_{j,\beta}^2}{w-\mathfrak{a}_{j,\beta}}\right)-\frac{2}{\pi}\sqrt{-x}\right], \tag{4.5}$$

where for $\beta>0$, $(\mathfrak{a}_{j,\beta})_{j=1}^\infty$ are the points of the Airy$_\beta$ point process (see e.g., Ramirez et al. [2011]) and $\xi_{j,\beta}^2$ are i.i.d. chi–squared random variables with β degrees of freedom, defined as Gamma-distributions for general β. For $\beta=1$ we are back to (4.2)*. For $\beta=2,4$, $\mathcal{G}_\beta(w)$ and the corresponding transition function, defined as in Definition 4.2, play the same role as $\mathcal{G}_1(w)$ in the signal plus noise models for complex and quaternionic matrices respectively.*

4.3. **Universal asymptotics for spiked models.** The next theorem presents the asymptotics of the largest eigenvalues for all signal plus noise models of Section 2.

Theorem 4.6. *Consider any of the four models of Section 2 with signal strengths $\theta_1 > \cdots > \theta_r$, with $\sigma^2 = 1$, and in the regime (2.7), (2.11), or (2.15). Fix an index $1 \le q \le r$ and suppose that as $N \to \infty$:*

(1) $\theta_1, \dots, \theta_{q-1}$ are fixed, distinct, and all larger than θ^c.
(2) $\theta_q = \theta^c + N^{-1/3}\tilde{\theta}$ for a fixed $\tilde{\theta} \in \mathbb{R}$.
(3) $\theta_{q+1}, \dots, \theta_r$ are fixed and all smaller than θ^c.

Then, in the sense of joint convergence in distribution,

$$\sqrt{N}(\lambda_i - \lambda(\theta_i)) \xrightarrow{d} \mathcal{N}(0, V(\theta_i)), \qquad 1 \le i \le q-1, \tag{4.6}$$

$$N^{2/3}(\lambda_q - \lambda_+) \xrightarrow{d} \kappa_1 \mathcal{T}(\kappa_2 \tilde{\theta}), \tag{4.7}$$

where the q limiting random variables in (4.6), (4.7) are jointly independent and the constants are as in (2.2),(2.8),(2.12),(2.16) with

$$\kappa_1 = \frac{1}{2}\left[\frac{[V'(\theta^c)]^2}{\lambda''(\theta^c)}\right]^{\frac{1}{3}}, \qquad \kappa_2 = \left[\frac{[\lambda''(\theta^c)]^2}{V'(\theta^c)}\right]^{\frac{1}{3}}. \tag{4.8}$$

If no signal strengths are close to θ^c, then the same limits hold without the (4.7) part.

Remark 4.7. *While the distributional limit of $\lambda_{q+1}, \dots, \lambda_r$ can be also computed, it is of no use for the confidence intervals: the limiting random variables would depend on $\tilde{\theta}$, but not on $\theta_{q+1}, \dots, \theta_r$.*

Note that the two limit regimes (4.6) and (4.7) heuristically agree with each other: if one sets $\tilde{\theta} = \varepsilon N^{1/3}$ with a small $\varepsilon > 0$, then using (4.7) and Proposition 4.3, we expect

$$\lambda_q \approx \lambda_+ + N^{-2/3}\kappa_1 \mathcal{T}(\kappa_2\tilde{\theta}) \approx \lambda_+ + \kappa_1\kappa_2^2\varepsilon^2 + 2N^{-1/2}\kappa_1\sqrt{\varepsilon\kappa_2}\mathcal{N}(0,1).$$

On the other hand, if one sets $\theta_i = \theta^c + \varepsilon$, then using (4.6) and Taylor expanding (noting $V(\theta^c) = \lambda'(\theta^c) = 0$), we expect

$$\lambda_i \approx \lambda_+ + \frac{\varepsilon^2}{2}\lambda''(\theta^c) + N^{-1/2}\sqrt{\varepsilon V'(\theta^c)}\mathcal{N}(0,1).$$

Using (4.8), we see that the last two asymptotic expansions are the same. In parallel, using Proposition 4.3 we can combine two asymptotic regimes of Theorem 4.6 into one (among several asymptotically equivalent formulas, we chose the one with the best finite sample performance):

Corollary 4.8. *The asymptotics (4.6) and (4.7) can be written in unified form as:*

$$\lambda_i \approx \lambda(\theta_i) - \frac{\kappa_2^{3/2}}{2}\sqrt{V(\theta_i)(\theta_i - \theta^c)^3} + \frac{\kappa_2^{-1/2}}{2N^{2/3}}\sqrt{\frac{V(\theta_i)}{\theta_i - \theta^c}}\mathcal{T}\Big(\kappa_2 N^{1/3}(\theta_i - \theta^c)\Big) + \frac{\kappa_3}{N}, \tag{4.9}$$

where the error is $o\big(N^{-2/3} + N^{-1/2}(V(\theta_i))^{1/2}\big)$ for $\theta_i > \theta^c$. For $\theta_i \le \theta^c$, one instead uses

$$\lambda_i \approx \lambda_+ + \frac{\kappa_1}{N^{2/3}}\mathcal{T}\left(\kappa_2\tilde{\theta}\right) + \frac{\kappa_3}{N}. \tag{4.10}$$

In (4.9) and (4.10) we use $\kappa_1 = \frac{1}{2}\frac{[V'(\theta^c)]^{2/3}}{[\lambda''(\theta^c)]^{1/3}}$, $\kappa_2 = \frac{[\lambda''(\theta^c)]^{2/3}}{[V'(\theta^c)]^{1/3}}$, and $\kappa_3 = -\frac{3}{2}\frac{\kappa_1}{\kappa_2\theta^c}$.

The formula (4.10) is a direct corollary of (4.7), while (4.9) combines (4.6) and (4.7) together. Indeed, when θ_i is bounded away from θ^c, (4.3) converts (4.9) into

$$\lambda(\theta_i)-\frac{\kappa_2^{3/2}}{2}\sqrt{V(\theta_i)(\theta_i-\theta^c)^3}+\frac{\kappa_2^{-1/2}}{2}\sqrt{\frac{V(\theta_i)}{\theta_i-\theta^c}}\kappa_2^2(\theta_i-\theta^c)^2+\frac{\kappa_2^{-1/2}}{N^{1/2}}\sqrt{\frac{V(\theta_i)}{\theta_i-\theta^c}}\sqrt{\kappa_2(\theta_i-\theta^c)}\mathcal{N}(0,1),$$

which is readily seen to be equivalent to (4.6). When θ_i is close to θ^c, $\theta_i=\theta^c+N^{-1/3}\tilde{\theta}$, Taylor expanding $\lambda(\cdot)$ and $V(\cdot)$ near θ^c, (4.9) turns into

$$\lambda_+ + \frac{\lambda''(\theta^c)}{2}(\theta_i-\theta^c)^2-\sqrt{V'(\theta_c)}\frac{\kappa_2^{3/2}}{2}(\theta_i-\theta^c)^2+\frac{\kappa_2^{-1/2}}{2N^{2/3}}\sqrt{V'(\theta^c)}\mathcal{T}\Big(\kappa_2N^{1/3}(\theta_i-\theta^c)\Big),$$

which is the same as (4.7).

Note that κ_3/N is $o(N^{-2/3})$, and therefore the choice of κ_3 does not affect the validity of the asymptotic formulas (4.9) and (4.10). These terms are introduced to further improve the performance of the formulas for intermediate values of N, cf. Johnstone [2008], Ma [2012], Johnstone and Ma [2012], which emphasize the importance of $1/N$ corrections for the practical applicability of limit theorems. The reasoning behind our choice of κ_3 is as follows. First, we require continuity at θ^c; hence, (4.9) and (4.10) use the same κ_3/N. Second, we leverage additional information available at $q=r=1$ and $\tilde{\theta}=-N^{1/3}\theta^c$ for the spiked Wigner, factor, and CCA models. (For the spiked covariance model, one instead takes $\tilde{\theta}=-N^{1/3}\gamma$ and adjusts the formula accordingly.) On one hand, combining (4.10) with the asymptotic approximation (4.4), we obtain

$$\lambda_1\approx\lambda_+ + \frac{\kappa_1}{N^{2/3}}\left(\mathfrak{a}_1+\frac{1}{\kappa_2N^{1/3}\theta^c}\right)+\frac{\kappa_3}{N}. \tag{4.11}$$

On the other hand, $\tilde{\theta}=-N^{1/3}\theta^c$ corresponds to $\theta=0$ (or $\theta=1$ for the spiked covariance model with $\tilde{\theta}=-N^{1/3}\gamma$), meaning that in all four models of interest, we are in the unspiked regime without a signal component. In this setting, the convergence of λ_1 to the Tracy–Widom distribution $\mathfrak{a}_1$ is well established, and $1/N$-order asymptotic corrections have been studied in Johnstone [2008], Ma [2012], Johnstone and Ma [2012]. From these works, one can extract

$$\lambda_1\approx\lambda_+ + \frac{\kappa_1}{N^{2/3}}\mathfrak{a}_1-\frac{1}{2N}\times\begin{cases}1 & \text{for spiked Wigner,}\\ \gamma(1+\gamma)^2 & \text{for spiked covariance and factors,}\\ \frac{(\sqrt{\tau_N-1}\sqrt{\tau_M-1}-1)^2(\sqrt{\tau_N-1}+\sqrt{\tau_M-1})^2}{\tau_N^2\tau_M\sqrt{\tau_N-1}\sqrt{\tau_M-1}} & \text{for CCA.}\end{cases} \tag{4.12}$$

Equating (4.11) with (4.12) yields the formula for κ_3, as recorded in Table 4. An interesting observation is that in each case the term $\frac{\kappa_1}{\kappa_2\theta^c}$ in (4.11) is twice the $\frac{1}{N}$ correction term in (4.12), which leads to the $\frac{3}{2}$ coefficient appearing in κ_3 across all four models.

Remark 4.9. *We expect that* (4.9) *also remains asymptotically valid on all mesoscropic scales, i.e., when* $\theta_i=\theta^c+N^{-\alpha}\tilde{\theta}$, $0<\alpha<1/3$. *We omit a detailed proof.*

The proof of Theorem 4.6 is given in Section 9. It begins with a rank-one perturbation equation, which expresses the eigenvalues in a signal plus noise model with r spikes (the "target model") as solutions to an algebraic equation involving a simpler model with $r-1$

spikes (the "base model"). Analyzing the asymptotic behavior of this equation leads to the following conclusion, stated informally below:

- If the strength of the added spike is subcritical, $\theta < \theta^c$, then the largest eigenvalues in the target model are very close to the largest eigenvalues in the base model.
- If the strength of the added spike is supercritical, $\theta > \theta^c$, then the largest eigenvalues in the target model are very close to the largest eigenvalues for the base model, except for one additional eigenvalue for the target model, which is close to $\lambda(\theta)$.
- If the strength of the added spike is critical, $\theta = \theta^c + N^{-1/3}\tilde{\theta}$, then in the target model eigenvalues which are (macrosopically) larger than λ_+ are very close to the eigenvalues in the base model. Near λ_+ the equations rescale to $\mathcal{G}(w) = -\kappa_2\tilde{\theta}$, where $\{\mathfrak{a}_j\}$ in the definition of $\mathcal{G}(w)$ arise as limits of the eigenvalues in the base model, and the eigenvalues in the target model converge to the roots of this equation.

On the technical level, the key novelty is in our ability to handle the most delicate case, when the spike is critical. If all spikes are subcritical or supercritical, the arguments are much simpler and follow ideas similar to those found in the references cited in Section 2. Some special cases of Theorem 4.6 can be handled by other methods, for example, the $r = 1$ case for the spiked Wigner and spiked covariance models is addressed in Mo [2012], Bloemendal and Virág [2013]; see also Bloemendal and Virág [2016], Lamarre and Shkolnikov [2019]. However, we believe that the level of generality achieved here – particularly our treatment of the factor model and CCA – was not previously available in the literature and is beyond the reach of those alternative methods.

Remark 4.10. *We expect that our methods can be extended to handle the case of multiple ($k > 1$) critical spikes, as well as the remaining largest eigenvalues $\lambda_{q+1}, \lambda_{q+2}, \dots$. The limiting behavior should be described by a higher-rank Airy point process, which we define recursively. The rank* 0 *process is the classical Airy point process $\{\mathfrak{a}_j\}$. The rank* 1 *process $\{\mathfrak{a}_j^{(\Theta)}\}$ consists of all real solutions to the equation $\mathcal{G}(w) = -\Theta$; in particular, the largest point $\mathfrak{a}_1^{(\Theta)}$ coincides with $\mathcal{T}(\Theta)$ from Definition 4.2. We then iterate this construction: given the rank k point process $\{\mathfrak{a}_j^{(\Theta_1,\dots,\Theta_k)}\}$ depending on k real parameters $\Theta_1,\dots,\Theta_k$, we define the rank $(k+1)$ process $\{\mathfrak{a}_j^{(\Theta_1,\dots,\Theta_k,\Theta_{k+1})}\}$ as the set of all real solutions to the equation*

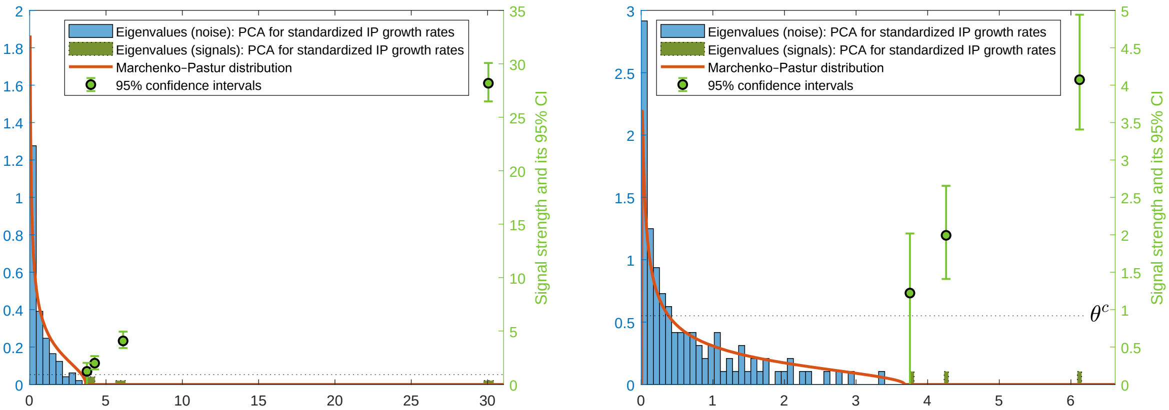
$$\lim_{x\to-\infty}\left[\left(\sum_{j:\,\mathfrak{a}_j^{(\Theta_1,\Theta_2,\dots,\Theta_k)}>x}\frac{[\xi_j^{(k)}]^2}{w-\mathfrak{a}_j^{(\Theta_1,\Theta_2,\dots,\Theta_k)}}\right)-\frac{2}{\pi}\sqrt{-x}\right]=-\Theta_{k+1}, \tag{4.13}$$

where $\xi_j^{(k)}$, $j = 1, 2, \dots$ are $\mathcal{N}(0,1)$, independent over j and k.

We anticipate that in a signal plus noise model with k critical spikes the eigenvalues near λ_+ converge, after the same recentering and rescaling as in Theorem 4.6, to the points $\{\mathfrak{a}_j^{(\tilde{\Theta}_1,\tilde{\Theta}_2,\dots,\tilde{\Theta}_k)}\}$. A different construction is provided in Bloemendal and Virág [2016], but it is ultimately expected to yield the same point process $\left\{\mathfrak{a}_j^{(\tilde{\Theta}_1,\tilde{\Theta}_2,\dots,\tilde{\Theta}_k)}\right\}_{j=1}^{\infty}$.

5. Empirical illustrations

In this section, we present three examples that illustrate the application of the procedure described in Section 3 to empirical data sets. In each case, the first step reveals a strong agreement between the histogram of eigenvalues and the corresponding theoretical



(A) All eigenvalues. (B) All except the largest ("market") signal.

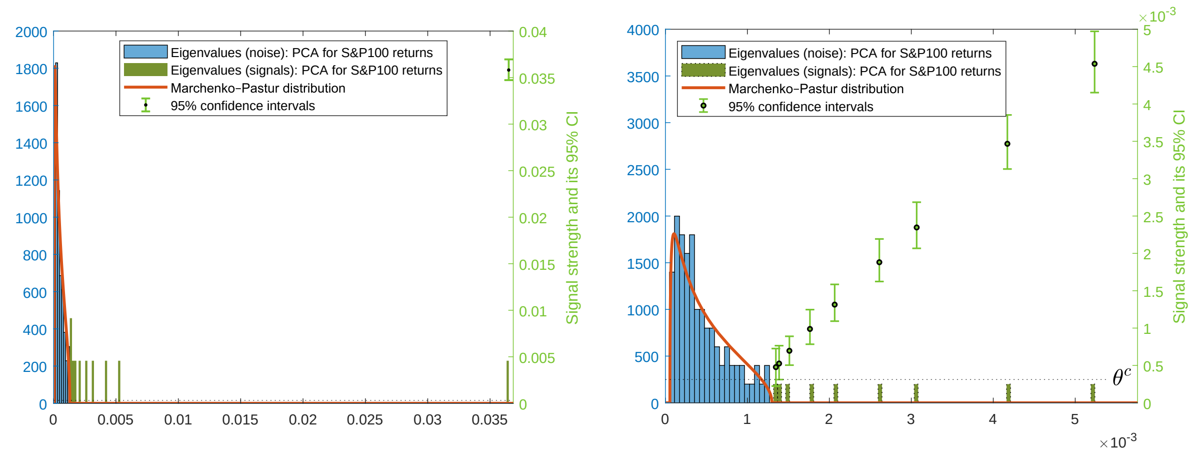
FIGURE 3. IP sample correlation eigenvalues: signals, their 95% confidence intervals, noise, and Marchenko-Pastur distribution with $N = 117$, $S = 139$.

curve—specifically, the Marchenko-Pastur law in the first two examples (factor models) and the Wachter law in the third (CCA). This suggests that the data aligns well with our modeling assumptions.

5.1. **Industrial Production.** Industrial production (IP) accounts for more than 10% of the United States' Gross Domestic Product (GDP), making it a significant component of total U.S. output. In this subsection we use data from Andreou et al. [2019], which investigates whether IP constitutes a dominant factor in U.S. economic activity. The data set contains quarterly IP growth rates across 117 sectors, spanning the period from 1977 : $Q1$ to 2011 : $Q4$. We de-mean the data and standardize each sector to have unit sample variance, thereby working with the sample correlation matrix of IP.

Figure 3 shows all eigenvalues of the standardized IP and highlights four of them (3.75, 4.26, 6.12, 30.05) that lie to the right of the theoretical Marchenko–Pastur upper edge, $\lambda_+ = 3.68$. The largest eigenvalue, 30.05, stands out markedly and represents a strong "market" factor. The two smallest among the four, being close to the cutoff, may reflect spurious signals arising from noise. To assess their significance, we construct 95% confidence intervals. Notably, the interval for the eigenvalue at 3.86 intersects the critical identification threshold $\theta^c = 0.92$, indicating that we cannot reject the null hypothesis that it represents noise (or non-informative signal). For the remaining eigenvalues, the null is rejected. We thus conclude that the IP growth rate is driven by three factors: strong, semi-strong, and weak.

5.2. **S&P100.** Analyzing stock returns is essential for understanding market dynamics, evaluating investment performance, and guiding both individual and institutional investment strategies. A key statistical object in this context is the covariance matrix of stock returns, which plays a central role in portfolio optimization, such as in the Markowitz mean–variance framework. The vast "factor zoo" – the large number of potential variables proposed to explain stock returns – highlights the practical challenge of distinguishing meaningful factors from noise in high-dimensional settings (see Cochrane [2011] for an influential discussion). Here we demonstrate how our methodology can be applied to the sample covariance matrix



(A) All eigenvalues.

(B) All except the largest ("market") signal.

FIGURE 4. S&P100 sample covariance: signals, their 95% confidence intervals, noise, and Marchenko-Pastur distribution with $N = 92$, $\gamma^2 = 0.4$, $\sigma = 0.02$.

of weekly S&P100 stock returns. We use data from Bykhovskaya and Gorin [2022], which covers 92 stocks over the period from January 1, 2010, to January 1, 2020.

Figure 4 presents the full spectrum of the sample covariance matrix and identifies ten eigenvalues – 0.0365, 0.005, 0.004, 0.003, 0.0026, 0.0021, 0.0018, 0.0015, 0.00139, 0.00136 – that lie to the right of the theoretical Marchenko–Pastur upper edge, $\lambda_+ = 0.0013$. To better fit the empirical data, we adopt an effective parameter value $\gamma^2 = 0.4$, in contrast to the true value $N/S = 0.18$. This adjustment may reflect temporal dependence in the data, which effectively reduces the sample size and increases $\gamma^2 = 0.4$. We also set $\sigma = 0.02$ to align the overall variance of the eigenvalues in the data.

Figure 4 reports the 95% confidence intervals for the ten candidate signals. The two smallest eigenvalues among them yield intervals that intersect the identification threshold $\theta^c = 0.0003$, indicating that they cannot be statistically distinguished from being non-informative. We therefore conclude that only eight of the ten observed spikes represent useful signals. As in previous examples, the largest eigenvalue corresponds to the "market" factor.

5.3. **Cyclical vs. non-cyclical stocks.** Financial stocks are typically classified into cyclical and non-cyclical (defensive) categories, depending on whether their performance tends to track economic business cycles. These groups are generally assumed to be uncorrelated, aside from exposure to a common "market" factor. Bykhovskaya and Gorin [2025] identify three non-zero canonical correlations between these two groups, suggesting the presence of three common factors. Here we revisit their analysis using the same data set to assess whether these observed correlations reflect genuine signals or could instead be attributed to noise.

The data set comprises weekly returns for 80 cyclical and 80 defensive stocks, spanning the period from January 1, 2010, to January 1, 2020. Figure 5 reproduces the canonical correlations reported by Bykhovskaya and Gorin [2025] and augments it with 95% confidence intervals. Bykhovskaya and Gorin [2025] used the signal strengths to estimate the

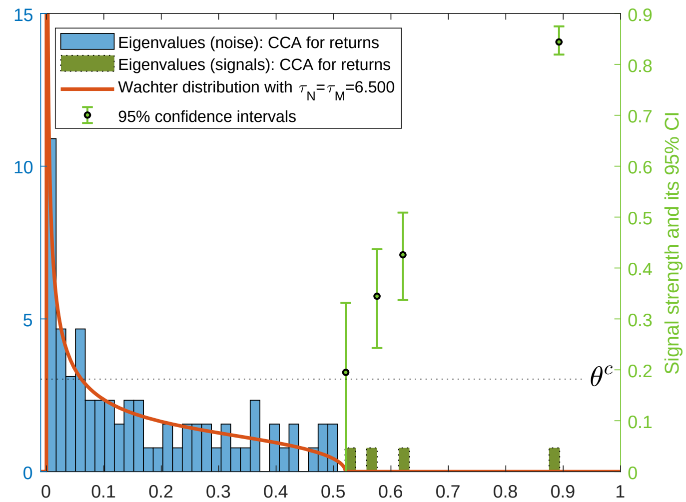


FIGURE 5. Squared sample canonical correlations between cyclical and non-cyclical stocks: signals, their 95% confidence intervals, noise, and Wachter distribution with $N = M = 80$, $S = 520$.

angles between true and estimated canonical variables. By incorporating our results, one can generate confidence intervals for these angles.

As shown in Figure 5, the 95% confidence intervals for three largest canonical correlations, 0.58, 0.62, and 0.89, lie above the cutoff $\theta^c = 0.18$, confirming them as true signals. In contrast, the confidence interval for the fourth largest value intersects the cutoff, indicating that it cannot be reliably distinguished from noise and is therefore classified as a non-informative component.

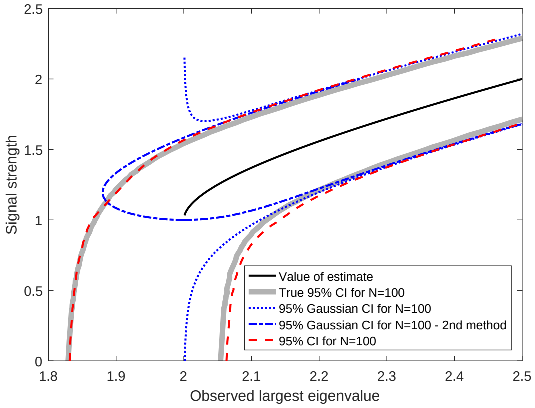
A comparison of Figures 3, 4, and 5 reveals an interesting pattern: in the factor models of the first two figures, the confidence intervals widen as the signal strengths increase, whereas in the CCA setting, the intervals become narrower as the signals approach 1. Theoretically, this behavior in CCA can be attributed to the factor $(1-\theta)^2$ in $V(\theta)$, as shown in Table 4.

6. Extensions

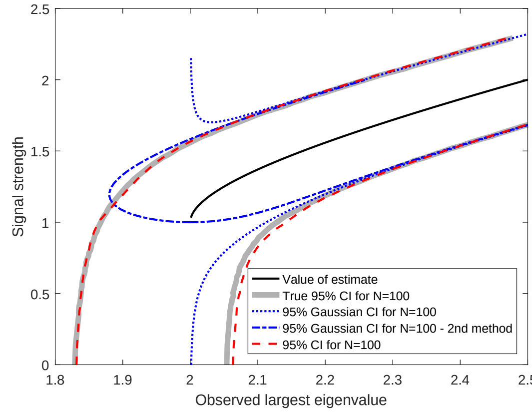
In this section we discuss possible extensions of our results, focusing on non-Gaussian data and broader classes of models than those considered in Section 2.

6.1. **Non-Gaussian noise.** The four models in Section 2 are all based on Gaussian noise matrices. A natural generalization is to replace the Gaussian vectors with more general random vectors having the same mean and covariance. It is well known (see Lee and Yin [2014], Ding and Yang [2018], Yang [2022a] and the broader reviews Deift and Gioev [2009], Tao and Vu [2012], Erdős and Yau [2017]) that for pure noise models without any signal, the distribution of the largest eigenvalues remains unchanged in many non-Gaussian settings. This raises the natural question of whether the same robustness extends to the confidence intervals discussed in Section 3.

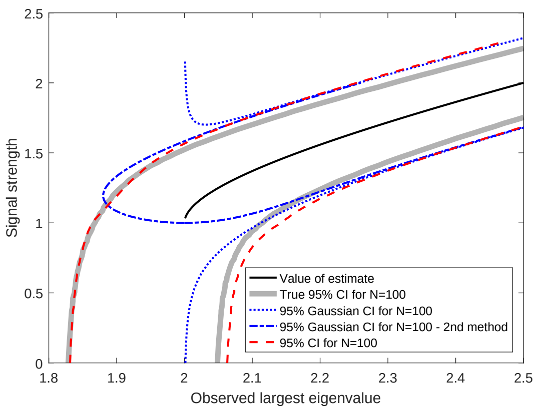
Figure 6 shows the results of Monte Carlo simulations of confidence intervals for Wigner matrices with a single spike and non-Gaussian noise. These can be compared to the Gaussian case in Figure 1a. The bold gray line depicts sample confidence intervals based on 10^5



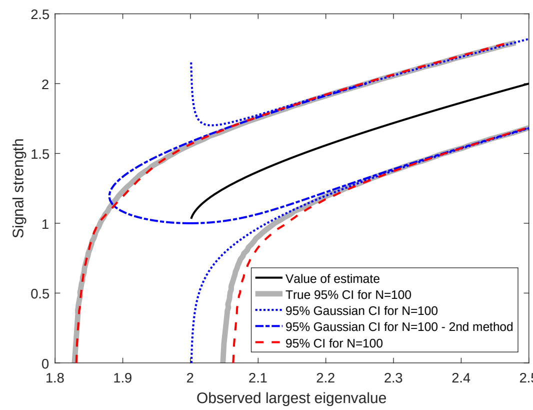
(A) Uniform noise, localized signal



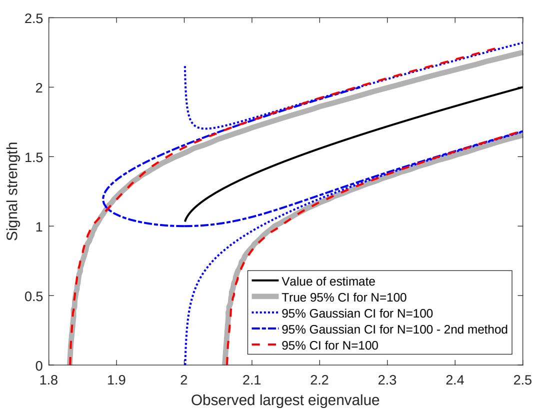
(B) Uniform noise, delocalized signal



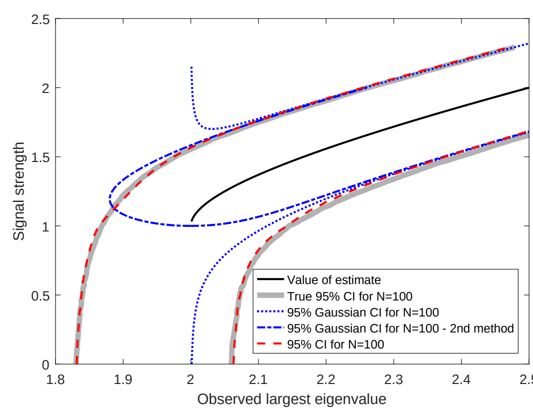
(C) Bernoulli $p = \frac{1}{2}$ noise, localized signal



(D) Bernoulli $p = \frac{1}{2}$ noise, delocalized signal



(E) Binomial$(2, 0.3)$ noise, localized signal



(F) Binomial$(2, 0.3)$ noise, delocalized signal

FIGURE 6. Confidence intervals for Wigner matrices with non-Gaussian noise.

simulations of 100×100 matrices. The noise matrix $\mathcal{E}$ in (2.1) is still defined as $\mathcal{E} = \frac{1}{\sqrt{2N}}(\mathcal{Z} + \mathcal{Z}^{\mathsf{T}})$, with $\mathcal{Z}$ having i.i.d. entries. We vary the distribution of these entries, each rescaled to have mean 0 and variance 1, to be: (i) uniform on $[0, 1]$, (ii) Bernoulli with success probability $p = 1/2$, or (iii) Binomial with parameters $n = 2$, $p = 0.3$. We also consider two signal vectors: a localized signal $\mathbf{u}^* = (1, 0, \ldots, 0)^\mathsf{T}$ and a delocalized signal $\mathbf{u}^* = \frac{1}{\sqrt{N}}(1, 1, \ldots, 1)^\mathsf{T}$. While in the Gaussian setting the choice of $\mathbf{u}^*$ is irrelevant due to rotational invariance, this is no longer true for general noise distributions.

Figure 6 shows that, while the sample confidence intervals remain reasonably close to those constructed via the procedure in Section 3, the agreement is notably better for delocalized signals. For the localized signal significant deviations appear at larger values of the observed eigenvalue λ_1. A heuristic explanation follows directly from the model $\mathbf{A} = \theta \cdot \mathbf{u}^*(\mathbf{u}^*)^\mathsf{T} + \mathcal{E}$ in (2.1). When $\mathbf{u}^* = (1, 0, \ldots, 0)^\mathsf{T}$, the parameter θ enters $\mathbf{A}$ only through its sum with the $(1, 1)$ entry of $\mathcal{E}$, so the distribution of that single element directly affects any estimate of θ. By contrast, when $\mathbf{u}^* = \frac{1}{\sqrt{N}}(1, 1, \ldots, 1)^\mathsf{T}$, the projection of the noise onto the signal direction aggregates many independent entries. Thus, by the central limit theorem, the influence of the individual noise distribution diminishes as N grows, leading to behavior indistinguishable from the Gaussian case. Similar robustness is expected for other delocalized signals, though a precise definition of "delocalized" can be delicate.

For spiked Wigner matrices with supercritical signals ($\theta > \theta^c$), non-Gaussian cases have been rigorously analyzed in several papers. Non-trivial dependence of the asymptotics of λ_1 on the distribution of the noise has been established in Capitaine et al. [2009, 2012], Pizzo et al. [2013], Knowles and Yin [2013, 2014] for localized signals. In contrast, universality of the limit for delocalized signals has been shown under various conditions (including different definitions of "delocalized") in Féral and Péché [2007], Benaych-Georges et al. [2011], Capitaine et al. [2012], Pizzo et al. [2013], Renfrew and Soshnikov [2013], Knowles and Yin [2013, 2014]. In addition, Knowles and Yin [2013, Remark 2.18] explain that the dependence of the limit on the noise distribution is washed out as θ approaches θ^c. Similar phenomenology is expected (and in some cases proven) for other signal plus noise models.

This discussion, together with the simulation results, leads us to conjecture that for all signal plus noise models of interest, the confidence intervals constructed via the procedure in Section 3 remain good approximations even under non-Gaussian noise, particularly when the signal is delocalized. From a practical standpoint, this reinforces the robustness of our method, especially in light of Giannone et al. [2021], who argue that many economic data sets are non-sparse and should therefore be modeled using delocalized signals.

6.2. **General models and empirical distributions of eigenvalues.** In all four models from Section 2 the empirical distributions of eigenvalues take specific parametric forms, as summarized in Table 3. Although our estimation procedure relies only on the largest eigenvalues, the formulas in Table 4 were derived using information from the empirical distributions in Table 3.

In other signal plus noise models and some empirical data sets the eigenvalue histograms may differ substantially from the four cases we consider. For $\theta > \theta^c$, the fluctuations of the largest eigenvalues have been analyzed in various alternative settings; see, e.g., Benaych-Georges et al. [2011], Benaych-Georges and Nadakuditi [2012], Onatski [2012]. These works develop analogues of (2.3), (2.9), and (2.13), but the formulas for θ^c, λ_+, $\lambda(\theta)$, and $V(\theta)$ become more complex. We conjecture that the confidence interval procedure from Section 3 remains valid in such broader contexts, with only the model-dependent parameters from Table 4 needing to be updated.

One applied setting of particular interest is the approximate factor model, which resembles the setup in Section 2.3 but allows the noise matrix $\mathcal{E}$ to have a more complex correlation structure rather than being i.i.d. To apply our confidence interval procedure in this context, one can proceed as follows: first, select a value for λ_+; then use all eigenvalues below λ_+ to estimate the empirical distribution of noise eigenvalues (replacing the parametric forms in Table 3); next, substitute this estimate into the formulas from Onatski [2012] for $\lambda(\theta)$ and $V(\theta)$; and finally, apply our method to construct confidence intervals for the eigenvalues exceeding λ_+. Choosing λ_+ optimally is delicate; one approach, suggested in Onatski [2010, Section 4], is to select it based on the characteristic $\sqrt{x}$ behavior of the eigenvalue density near the edge – a feature clearly visible in the four models of Table 3 and present in many other cases.

7. Conclusion

The paper presents a unified framework for conducting inference on signal strength in high-dimensional signal plus noise models, with a particular focus on the critical regime where standard Gaussian approximations fail. We demonstrate that the limiting distribution of top eigenvalues is governed by a universal stochastic process, the transition process $\mathcal{T}(\Theta)$, whose quantiles can be tabulated and used to construct valid confidence intervals. This approach applies uniformly across four canonical models: spiked Wigner matrices, spiked sample covariance matrices, factor models, and canonical correlation analysis.

Our procedure is robust to both weak and critical signals, enabling practitioners to distinguish between informative and non-informative components without imposing assumptions on signal strength. Moreover, our methodology reveals a surprising universality: despite differences in the statistical structure of the models, the same transition process governs the fluctuations of their top eigenvalues. We believe that this suggests deeper underlying principles in high-dimensional inference and opens a broad avenue for future research in more general signal plus noise settings.

8. Appendix A: Random Stieltjes transform at the edge and general asymptotic theorem

Our goal in this section is to rigorously introduce the Airy–Green function $\mathcal{G}(w)$ and limit theorems related to it. Universally, in all settings we study, the equations connecting the parameters of the spiked model with observed largest eigenvalues will be asymptotically written in terms of $\mathcal{G}(w)$.

Let us recall the definitions of meromorphic functions and their convergence, which will be used in the proofs. A function $f(z)$ of a complex variable is called meromorphic if it is defined and analytic on $\mathbb{C} \setminus \{x_j\}$, where $\{x_j\}$ is at most a countable set of isolated points, each of which is a pole of finite order. Equivalently, f is an analytic function from $\mathbb{C}$ to the Riemann sphere $\overline{\mathbb{C}}$. If f_n and f are meromorphic, we write $f_n \overset{\text{mer}}{\to} f$ if for any compact $W \subset \mathbb{C}$ we have

$$\sup_{z \in W} d(f_n(z), f(z)) \to 0 \ ,$$

where

$$d(\zeta, \zeta') = \frac{|\zeta - \zeta'|}{\sqrt{1 + |\zeta|^2}\sqrt{1 + |\zeta'|^2}}$$

is the spherical distance. Note that if f_n and f are meromorphic functions, then $f_n \overset{\text{mer}}{\to} f$ if and only if $f_n \to f$ uniformly on any compact $W \subset \mathbb{C}$ not containing the poles of f.

8.1. **The function $\mathcal{G}(w)$.** In this subsection we restate, elaborate on, and prove the statements from Section 4.1.

Theorem 8.1. *Let $\mathfrak{a}_1 \geq \mathfrak{a}_2 \geq \mathfrak{a}_3 \geq \dots$ be a realization of the Airy$_1$ point process and let $\{\xi_j\}_{j=1}^{\infty}$ be i.i.d. Gaussian $\mathcal{N}(0,1)$ independent of $\{\mathfrak{a}_j\}_{j=1}^{\infty}$. Almost surely, for every $w \in \mathbb{C} \setminus \{\mathfrak{a}_j\}$ there exists a limit*

$$\mathcal{G}(w) = \lim_{x \to -\infty} \left[\left(\sum_{j\colon \mathfrak{a}_j > x} \frac{\xi_j^2}{w - \mathfrak{a}_j} \right) - \frac{2}{\pi}\sqrt{-x} \right], \tag{8.1}$$

and, moreover, the convergence is uniform on any compact $W \subset \mathbb{C} \setminus \{\mathfrak{a}_j\}$.

Remark 8.2. *The conditional expectation $\mathbb{E}(\mathcal{G}_\beta(w)|\{\mathfrak{a}_j\})$ (equivalent to replacing $\frac{1}{\beta}\xi_j^2$ by 1) was also used recently in another context by Huang and Zhang [2024].*

In the notation above, the theorem asserts that almost surely the functions converge in the topology $\overset{\text{mer}}{\to}$.

Proposition 8.3. *$\mathcal{G}(w)$ is a meromorphic function with poles at $\{\mathfrak{a}_j\}_{j=1}^{\infty}$ and satisfying:*

$$\lim_{\substack{w \to \infty \\ \mathrm{Re}(w) \geq 0}} \left| \mathcal{G}(w) + \sqrt{w} \right| = 0, \qquad \textit{almost surely.} \tag{8.2}$$

For real w, in the sense of convergence in distribution, we also have

$$\lim_{w \to +\infty} w^{1/4} \big(\mathcal{G}(w) + \sqrt{w} \big) \overset{d}{=} \mathcal{N}(0,1). \tag{8.3}$$

Remark 8.4. *For $\sqrt{w}$ in (8.2), one should use the branch of the square root which is positive on positive reals. In particular, $\sqrt{\mathbf{i}R} = \frac{1+\mathbf{i}}{\sqrt{2}}\sqrt{R}$, $R > 0$, where $\mathbf{i} = \sqrt{-1}$. While we do not need this, the asymptotics (8.2) can be extended from $\mathrm{Re}(w) \geq 0$ to all w such that $|\arg(w)| < \pi - \varepsilon$ for a fixed $\varepsilon > 0$. Similarly, (8.3) can be extended to complex w.*

Proposition 8.5. *Recall $\mathcal{T}(\Theta)$ of Definition 4.2 that solves $\mathcal{G}(w) = -\Theta$. Almost surely, $\Theta \mapsto \mathcal{T}(\Theta)$ is an increasing bijection of $\mathbb{R}$ onto $(\mathfrak{a}_1, \infty)$. As $\Theta \to +\infty$, $\mathcal{T}(\Theta)$ is asymptotically Gaussian: in distribution*

$$\lim_{\Theta \to +\infty} \frac{\mathcal{T}(\Theta) - \Theta^2}{2\sqrt{\Theta}} \overset{d}{=} \mathcal{N}(0,1). \tag{8.4}$$

The proofs are based on four lemmas describing the asymptotics of $\mathfrak{a}_j$.

Lemma 8.6. *Let $\rho(x)\mathrm{d}x$ be the first correlation measure of $\{\mathfrak{a}_j\}$, which means that for any compactly supported bounded $f(x)$ with finitely many discontinuity points, we have*

$$\mathbb{E}\sum_{j=1}^{\infty} f(\mathfrak{a}_j) = \int_{-\infty}^{\infty} f(x)\rho(x)\mathrm{d}x.$$

Then:

(1) $\rho(x)$ is a bounded continuous function of x;
(2) $\rho(x)$ decays faster than $\exp(-Cx)$ for any $C > 0$ as $x \to +\infty$;
(3) At $-\infty$, $\rho(x)$ has the asymptotics

$$\rho(x) = \frac{(-x)^{1/2}}{\pi} + O(1/|x|), \quad x \to -\infty\ . \tag{8.5}$$

Proof. Recall that the Airy function $\mathrm{Ai}(x)$ is a solution of the differential equation $\mathrm{Ai}''(x) - x\,\mathrm{Ai}(x) = 0$, and is given by the improper integral $\mathrm{Ai}(x) = \frac{1}{\pi}\int_0^\infty \cos(t^3 + xt)\mathrm{d}t$. We need two asymptotic expansions for $\mathrm{Ai}(x)$, which can be found in Abramowitz and Stegun [1972, Section 10.4]:

$$\mathrm{Ai}(x) = \frac{1}{2\sqrt{\pi}\cdot x^{1/4}} \exp\left(-\frac{2}{3}x^{3/2}\right)\cdot \left[1 + O\left(x^{-3/2}\right)\right], \qquad x \to +\infty, \tag{8.6}$$

$$\mathrm{Ai}(x) = \frac{1}{\sqrt{\pi}(-x)^{1/4}} \sin\left(\frac{2}{3}(-x)^{3/2} + \frac{\pi}{4}\right)\cdot \left[1 + O\left((-x)^{-3/2}\right)\right], \qquad x \to -\infty. \tag{8.7}$$

The expansions are valid in a complex neighborhood of the real axis and, hence, can be differentiated to get

$$\mathrm{Ai}'(x) = -\frac{x^{1/4}}{2\sqrt{\pi}} \exp\left(-\frac{2}{3}x^{3/2}\right)\cdot \left[1 + O\left(x^{-3/2}\right)\right], \qquad x \to +\infty, \tag{8.8}$$

$$\mathrm{Ai}'(x) = -\frac{(-x)^{1/4}}{\sqrt{\pi}} \cos\left(\frac{2}{3}(-x)^{3/2} + \frac{\pi}{4}\right)\cdot \left[1 + O\left((-x)^{-3/2}\right)\right], \qquad x \to -\infty. \tag{8.9}$$

Pastur and Shcherbina [2011, (6.3.2), (6.1.18), (5.3.6)] give an explicit formula for the first correlation measure of the Airy$_1$ point process in terms of the Airy function:

$$\rho(x) = [\mathrm{Ai}'(x)]^2 - x[\mathrm{Ai}(x)]^2 + \frac{1}{2}\mathrm{Ai}(x)\left(1 - \int_x^{+\infty} \mathrm{Ai}(z)\mathrm{d}z\right) \tag{8.10}$$

$$= \int_0^\infty \mathrm{Ai}(z+x)^2 \mathrm{d}z + \frac{1}{2}\mathrm{Ai}(x)\left(1 - \int_x^{+\infty} \mathrm{Ai}(z)\mathrm{d}z\right). \tag{8.11}$$

Plugging the asymptotics of $\mathrm{Ai}(x)$ into the definition of $\rho(x)$, we see that it decays super-exponentially as $x \to +\infty$, while for $x \to -\infty$

$$\rho(x) = \frac{(-x)^{1/2}}{\pi}\cdot \left[1 + O\left((-x)^{-3/2}\right)\right] + O\big((-x)^{-1}\big). \tag{8.12}$$

□

Lemma 8.7. *For $\mathfrak{a}_1 > \mathfrak{a}_2 > \mathfrak{a}_3 > \dots$ being (a realisation of) the Airy$_1$ point process,*

$$\mathbb{E}\big(\#\{j \ge 1 \mid \mathfrak{a}_j > -T\}\big) = \frac{2}{3\pi}T^{3/2} + O(\ln(T)), \quad T \to +\infty\ . \tag{8.13}$$

Proof. We can write the expectation in terms of the first correlation measure of $\{\mathfrak{a}_j\}$:

$$\mathbb{E}\big(\#\{j \ge 1 \mid \mathfrak{a}_j > -T\}\big) = \int_{-T}^{+\infty} \rho(x)\mathrm{d}x. \tag{8.14}$$

Plugging the result of Lemma 8.6 into (8.14) and integrating, we get (8.13). □

Lemma 8.8. *For $\mathfrak{a}_1 > \mathfrak{a}_2 > \mathfrak{a}_3 > \dots$ being the Airy$_1$ point process,*

$$\lim_{T\to+\infty} \frac{\mathrm{Var}\big(\#\{j \ge 1 \mid \mathfrak{a}_j > -T\}\big)}{\ln(T)} = \frac{11}{6\pi^2}. \tag{8.15}$$

Proof. We use a trick from O'Rourke [2010]. The edge scaling limit of Forrester and Rains [2001, Theorem 4.3] is the following identity in law:

$$\mathrm{even}\big(\{\mathfrak{a}_i\}_{i=1}^\infty \cup \{\mathfrak{a}'_i\}_{i=1}^\infty\big) \overset{d}{=} \{\mathfrak{a}_i^{\beta=2}\}_{i=1}^\infty, \tag{8.16}$$

where $\{\mathfrak{a}_i\}$ and $\{\mathfrak{a}'_i\}$ are two independent copies of the Airy$_1$ point process, $\mathfrak{a}_i^{\beta=2}$ is the Airy$_2$ point process, and "even" is the operation of removing all particles with odd indices, i.e., keeping the 2nd, 4th, 6th, etc, largest ones. From Soshnikov [2000, Theorem 1], it is known that as $T \to +\infty$, $\mathrm{Var}\big(\#\{j \ge 1 \mid \mathfrak{a}_j^{\beta=2} > -T\}\big) \sim \frac{11}{12\pi^2}\ln(T)$. Through (8.16) this implies

$$\lim_{T\to+\infty} \frac{\mathrm{Var}\big(\#\{j \ge 1 \mid \mathfrak{a}_j > -T\} + \#\{j \ge 1 \mid \mathfrak{a}'_j > -T\}\big)}{\ln(T)} = 4 \cdot \frac{11}{12\pi^2} = \frac{11}{3\pi^2},$$

because the "even" operator divides the number of particles by two (up to error of at most 1), and therefore divides the variance by four (up to an error negligible as $T \to \infty$). Since variances for independent random variables are added, we get (8.15). $\square$

Lemma 8.9. *For each $\varepsilon > 0$ there exists a random variable $\mathfrak{J} = \mathfrak{J}(\varepsilon)$, such that almost surely we have*

$$\left| \mathfrak{a}_j + \left(\frac{3\pi j}{2}\right)^{2/3} \right| \le j^{\varepsilon}, \qquad \textit{for all } j > \mathfrak{J}. \tag{8.17}$$

Proof. Choose $0 < \varepsilon < \frac{1}{3}$. For $n = 1, 2, \dots$, let A_n be the event

$$\big|\#\{j \ge 1 \mid \mathfrak{a}_j > -n\} - \mathbb{E}\#\{j \ge 1 \mid \mathfrak{a}_j > -n\}\big| > n^{1/2+\varepsilon}.$$

Using Chebyshev's inequality, we have

$$\mathrm{Prob}(A_n) \le \frac{\mathrm{Var}\big(\#\{j \ge 1 \mid \mathfrak{a}_j > -n\}\big)}{n^{1+2\varepsilon}}.$$

Combining with (8.15), we conclude that $\sum_{n=1}^{\infty} \mathrm{Prob}(A_n) < \infty$. Therefore, by the Borel–Cantelli lemma, there exists a random variable $\mathfrak{n}$, such that

$$\big|\#\{j \ge 1 \mid \mathfrak{a}_j > -n\} - \mathbb{E}\#\{j \ge 1 \mid \mathfrak{a}_j > -n\}\big| \le n^{1/2+\varepsilon}, \qquad \text{for all } n > \mathfrak{n}.$$

Combining with (8.13) and increasing $\mathfrak{n}$, if necessary, we conclude that almost surely

$$\left|\#\{j \ge 1 \mid \mathfrak{a}_j > -n\} - \frac{2}{3\pi} n^{3/2}\right| \le n^{1/2+\varepsilon}, \qquad \text{for all } n > \mathfrak{n}.$$

Therefore,

$$\mathfrak{a}_{\lfloor \frac{2}{3\pi} n^{3/2} + n^{1/2+\varepsilon} \rfloor} \le -n \qquad \text{and} \qquad \mathfrak{a}_{\lfloor \frac{2}{3\pi} n^{3/2} - n^{1/2+\varepsilon} \rfloor} > -n-1, \qquad \text{for all } n > \mathfrak{n}.$$

Denoting $k = \lfloor \frac{2}{3\pi} n^{3/2} + n^{1/2+\varepsilon} \rfloor$, $\ell = \lfloor \frac{2}{3\pi} n^{3/2} - n^{1/2+\varepsilon} \rfloor$ we conclude that for large k and ℓ,

$$\mathfrak{a}_k < -\left(\frac{3\pi k}{2}\right)^{2/3} + \frac{1}{2} k^{\varepsilon} \qquad \text{and} \qquad \mathfrak{a}_\ell > -\left(\frac{3\pi \ell}{2}\right)^{2/3} - \frac{1}{2}\ell^{\varepsilon}. \tag{8.18}$$

In order to extend the inequalities from k and ℓ of special form we used to all large k and ℓ, note that the distance between adjacent allowed values of k is (assuming k is large):

$$\frac{2}{3\pi}(n+1)^{3/2} + (n+1)^{1/2+\varepsilon} - \left(\frac{2}{3\pi} n^{3/2} + n^{1/2+\varepsilon}\right) < \frac{1}{3} n^{1/2} < k^{1/3},$$

and similarly for ℓ. Hence, the monotonicity of $\mathfrak{a}_k$ in k and the first inequality in (8.18) imply that for a (random) $\mathfrak{K}$, we have almost surely

$$\mathfrak{a}_k < -\left(\frac{3(k - k^{1/3})}{2\pi}\right)^{2/3} + \frac{1}{2}k^{\varepsilon} < -\left(\frac{3k}{2\pi}\right)^{2/3} + k^{\varepsilon}, \qquad \text{for all } k > \mathfrak{K}.$$

Similarly producing a corollary of the second inequality in (8.18), we get (8.17). □

Proof of Theorem 8.1. We split $\mathcal{G}(w)$ into two parts:

$$\mathcal{G}(w) = \lim_{x\to-\infty}\left[\sum_{j:\,\mathfrak{a}_j>x}\frac{1}{w-\mathfrak{a}_j} - \frac{2}{\pi}\sqrt{-x}\right] + \lim_{x\to-\infty}\left[\sum_{j:\,\mathfrak{a}_j>x}\frac{\xi_j^2-1}{w-\mathfrak{a}_j}\right] = \mathcal{G}_1(w)+\mathcal{G}_2(w). \tag{8.19}$$

For $\mathcal{G}_1(w)$, we further write it as:

$$\mathcal{G}_1(w) = \lim_{x\to-\infty}\sum_{j:\,\mathfrak{a}_j>x}\left[\frac{1}{w-\mathfrak{a}_j}-\frac{1}{\mathbf{i}-\mathfrak{a}_j}\right] + \lim_{x\to-\infty}\left[\left(\sum_{j:\,\mathfrak{a}_j>x}\frac{1}{\mathbf{i}-\mathfrak{a}_j}\right)-\frac{2}{\pi}\sqrt{-x}\right]. \tag{8.20}$$

Note that

$$\frac{1}{w-\mathfrak{a}_j}-\frac{1}{\mathbf{i}-\mathfrak{a}_j} = \frac{\mathbf{i}-w}{(w-\mathfrak{a}_j)(\mathbf{i}-\mathfrak{a}_j)}.$$

Hence, using (8.17), the j–th term in the first sum of (8.20) decays as $j^{-4/3}$ and the sum is absolutely convergent, uniformly in w bounded away from $\mathfrak{a}_j$. For the second sum, its imaginary part is

$$\lim_{x\to-\infty}\left[\sum_{j:\,\mathfrak{a}_j>x}\frac{-\mathbf{i}}{1+\mathfrak{a}_j^2}\right],$$

which is again absolutely convergent. The real part is

$$\lim_{x\to-\infty}\left[\left(\sum_{j:\,\mathfrak{a}_j>x}\frac{-\mathfrak{a}_j}{1+\mathfrak{a}_j^2}\right)-\frac{2}{\pi}\sqrt{-x}\right]. \tag{8.21}$$

Using (8.17), for large j we have $-\mathfrak{a}_j = \left(\frac{3\pi j}{2}\right)^{2/3}+O(j^{\varepsilon})$, and, therefore, for large m and n:

$$\begin{aligned}\sum_{j=m}^{n}\frac{-\mathfrak{a}_j}{1+\mathfrak{a}_j^2} &= \sum_{j=m}^{n}\frac{1}{\left(\frac{3\pi j}{2}\right)^{2/3}+O(j^{\varepsilon})} = \sum_{j=m}^{n}\left[\frac{1}{\left(\frac{3\pi j}{2}\right)^{2/3}}+O(j^{-4/3+\varepsilon})\right]\\ &= \left(\frac{2}{3\pi}\right)^{2/3}\int_m^n x^{-2/3}\mathrm{d}x+o(1) = 3\left(\frac{2}{3\pi}\right)^{2/3}\left(n^{1/3}-m^{1/3}\right)+o(1).\end{aligned} \tag{8.22}$$

We check the Cauchy criterion for (8.21) and compute the difference of its values at $x=-y$ and $x=-z$ for large $y>z>0$. Using (8.22), we get

$$3\left(\frac{2}{3\pi}\right)^{2/3}\left(n(y)^{1/3}-m(z)^{1/3}\right)-\frac{2}{\pi}\sqrt{y}+\frac{2}{\pi}\sqrt{z}+o(1), \tag{8.23}$$

where $n(y)$ is the index j for the closest to $-y$ point $\mathfrak{a}_j$ and $m(z)$ is the index for the closest to $-z$ point $\mathfrak{a}_j$. Assuming y and z large enough, so that $n(y)>\mathfrak{J}$ and $m(z)>\mathfrak{J}$, we can use (8.17) and get

$$n(y) = \frac{2}{3\pi}\big(y+O(y^{\varepsilon})\big)^{3/2} = \frac{2}{3\pi}y^{3/2}+O\big(y^{1/2+\varepsilon}\big), \qquad m(z) = \frac{2}{3\pi}z^{3/2}+O\big(z^{1/2+\varepsilon}\big).$$

Plugging into (8.23) and choosing ε to be small enough, we get

$$3\left(\frac{2}{3\pi}\right)^{2/3}\left(\left(\frac{2}{3\pi}y^{3/2}+O\big(y^{1/2+\varepsilon}\big)\right)^{1/3}-\left(\frac{2}{3\pi}z^{3/2}+O\big(z^{1/2+\varepsilon}\big)\right)^{1/3}\right)-\frac{2}{\pi}\sqrt{y}+\frac{2}{\pi}\sqrt{z}+o(1)$$
$$=O(y^{-1/2+\varepsilon})+O(z^{-1/2+\varepsilon})+o(1)\to 0, \qquad \text{as} \quad y>z\to\infty.$$

Therefore, (8.21) has an almost sure limit and $\mathcal{G}_1(w)$ is well-defined. We proceed to $\mathcal{G}_2(w)$ and again split it into two parts:

$$\mathcal{G}_2(w)=\lim_{x\to-\infty}\left[\sum_{j:\,\mathfrak{a}_j>x}\frac{\xi_j^2-1}{\mathbf{i}-\mathfrak{a}_j}\right]+\lim_{x\to-\infty}\left[\sum_{j:\,\mathfrak{a}_j>x}\frac{(\xi_j^2-1)(\mathbf{i}-w)}{(w-\mathfrak{a}_j)(\mathbf{i}-\mathfrak{a}_j)}\right]. \tag{8.24}$$

We would like to condition on the (typical) values of $\{\mathfrak{a}_j\}$, and then prove that both limits exist almost surely with respect to the randomness coming from ξ_j. For the imaginary part of the sum in the first limit, we notice that

$$\left|\operatorname{Im}\left(\frac{\xi_j^2-1}{\mathbf{i}-\mathfrak{a}_j}\right)\right|=\frac{|\xi_j^2-1|}{1+(\mathfrak{a}_j)^2}.$$

Using (8.17) and the monotone convergence theorem (conditionally on $\{\mathfrak{a}_j\}$ the sum of expectations with respect to ξ_j is finite), we see that almost surely

$$\sum_{j=1}^{\infty}\frac{|\xi_j^2-1|}{1+(\mathfrak{a}_j)^2}<\infty.$$

Hence, the imaginary part of the first sum in (8.24) is absolutely convergent and $x\to-\infty$ limit is well-defined. The same monotone convergence argument shows that the second sum in (8.24) is dominated by a convergent series, and, therefore, it is absolutely convergent uniformly over w in compact sets bounded away from $\mathfrak{a}_j$. It remains to deal with the real part of the first sum in (8.24):

$$\lim_{x\to-\infty}\operatorname{Re}\left[\sum_{j:\,\mathfrak{a}_j>x}\frac{\xi_j^2-1}{\mathbf{i}-\mathfrak{a}_j}\right]=\lim_{x\to-\infty}\left[\sum_{j:\,\mathfrak{a}_j>x}(\xi_j^2-1)\frac{-\mathfrak{a}_j}{1+(\mathfrak{a}_j)^2}\right].$$

We condition on $\{\mathfrak{a}_j\}$ and note that we deal with a sum of independent mean 0 random variables. Hence, by the Kolmogorov two-series theorem (see, e.g., Durrett [2019, Theorem 2.5.6]), the almost sure convergence would follow from the convergence of the sum of (conditional) variances, i.e., convergence of the series

$$\sum_{j=1}^{\infty}\mathbb{E}\big[(\xi_j^2-1)^2\big]\frac{(\mathfrak{a}_j)^2}{(1+(\mathfrak{a}_j)^2)^2},$$

which readily follows from (8.17). We conclude that $\mathcal{G}_2(w)$ is also well-defined. □

Proof of Proposition 8.3. We rewrite the definition (8.1) of $\mathcal{G}(w)$ as

$$\mathcal{G}(w) = \lim_{x\to-\infty}\left[\sum_{j:\,\mathfrak{a}_j>x}\left(\frac{1}{w+\left(\frac{3\pi j}{2}\right)^{2/3}}+\frac{\xi_j^2-1}{w+\left(\frac{3\pi j}{2}\right)^{2/3}}+\xi_j^2\left[\frac{\left(\frac{3\pi j}{2}\right)^{2/3}+\mathfrak{a}_j}{(w-\mathfrak{a}_j)(w+\left(\frac{3\pi j}{2}\right)^{2/3})}\right]\right)-\frac{2}{\pi}\sqrt{-x}\right].$$

Splitting the sum into three and using Lemma 8.9, the last expression is transformed into

$$\mathcal{G}(w) = \lim_{x\to-\infty}\left[\left(\sum_{j:\,\left(\frac{3\pi j}{2}\right)^{2/3}<-x}\frac{1}{w+\left(\frac{3\pi j}{2}\right)^{2/3}}\right)-\frac{2}{\pi}\sqrt{-x}\right]+\sum_{j=1}^{\infty}\frac{\xi_j^2-1}{w+\left(\frac{3\pi j}{2}\right)^{2/3}} + \sum_{j=1}^{\infty}\xi_j^2\left[\frac{\left(\frac{3\pi j}{2}\right)^{2/3}+\mathfrak{a}_j}{(w-\mathfrak{a}_j)(w+\left(\frac{3\pi j}{2}\right)^{2/3})}\right]. \tag{8.25}$$

Let us show that as w becomes large (constrained by $\mathrm{Re}(w)\ge 0$), the second and third sums in (8.25) tend to 0.

For the third sum, let us show that it is $o(|w|^{-1/4})$, in the sense that there exists a random variable $\mathfrak{c}$, such that for all w with $\mathrm{Re}(w)\ge 0$ and $|w|\ge 1$, we have:

$$\left|\sum_{j=1}^{\infty}\xi_j^2\left[\frac{\left(\frac{3\pi j}{2}\right)^{2/3}+\mathfrak{a}_j}{(w-\mathfrak{a}_j)(w+\left(\frac{3\pi j}{2}\right)^{2/3})}\right]\right|\le \mathfrak{c}|w|^{-1/4}. \tag{8.26}$$

We use Lemma 8.9 and note that for $j>\mathfrak{J}$, the numerator satisfies $|\left(\frac{3\pi j}{2}\right)^{2/3}+\mathfrak{a}_j|\le j^{\varepsilon}$. In addition, since ξ_j^2 has exponential tails, the Borel–Cantelli lemma implies that there exists a random $\mathfrak{C}=\mathfrak{C}(\varepsilon)>0$, such that almost surely $\xi_j^2<\mathfrak{C}j^{\varepsilon}$ for all $j=1,2,\dots$. We choose ε to be small enough and upper-bound the series (8.26) by three sums:

$$\sum_{j=1}^{\mathfrak{J}}\frac{\xi_j^2\left|\left(\frac{3\pi j}{2}\right)^{2/3}+\mathfrak{a}_j\right|}{|w-\mathfrak{a}_j||w+\left(\frac{3\pi j}{2}\right)^{2/3}|}+\sum_{j=\mathfrak{J}+1}^{\lfloor |w|^{3/2}\rfloor}\frac{2\mathfrak{C}j^{2\varepsilon}}{|w+\left(\frac{3\pi j}{2}\right)^{2/3}|^2}+\sum_{j=\lfloor |w|^{3/2}\rfloor+1}^{\infty}\frac{2\mathfrak{C}j^{2\varepsilon}}{|w+\left(\frac{3\pi j}{2}\right)^{2/3}|^2}. \tag{8.27}$$

The first sum is finite, and, therefore, almost surely converges to 0 at speed $|w|^{-2}$ as $|w|\to\infty$. The second sum has at most $|w|^{3/2}$ terms and each term is upper-bounded as $O(|w|^{-2})$; hence, the sum is $O(w^{-1/2})$. The terms of the last sum can be upper-bounded by a constant times $j^{2\varepsilon-4/3}$, and therefore the sum is upper bounded by a constant times $|w|^{\frac{3}{2}(2\varepsilon-1/3)}$. Combining all three bounds, we arrive at (8.26).

For the second sum in (8.25), we do summation by parts:

$$\sum_{j=1}^{M}\frac{\xi_j^2-1}{w+\left(\frac{3\pi j}{2}\right)^{2/3}} = \frac{1}{w+\left(\frac{3\pi(M+1)}{2}\right)^{2/3}}\sum_{k=1}^{M}(\xi_k^2-1) - \sum_{j=1}^{M}\left[\sum_{k=1}^{j}[\xi_k^2-1]\right]\left(\frac{1}{w+\left(\frac{3\pi(j+1)}{2}\right)^{2/3}}-\frac{1}{w+\left(\frac{3\pi j}{2}\right)^{2/3}}\right).$$

Taking absolute values, we get an upper-bound for a deterministic constant $C > 0$

$$\left|\sum_{j=1}^{M} \frac{\xi_j^2 - 1}{w + \left(\frac{3\pi j}{2}\right)^{2/3}}\right| \leq \frac{C}{|w| + M^{2/3}} \left|\sum_{k=1}^{M} (\xi_k^2 - 1)\right| + \sum_{j=1}^{M} \left|\sum_{k=1}^{j} [\xi_k^2 - 1]\right| \frac{C j^{-1/3}}{(|w| + j^{2/3})^2}.$$

Applying the Law of Iterated Logarithm to the sums $\sum_{k=1}^{j}[\xi_k^2 - 1]$, we find another random variable $\mathfrak{C}' > 0$, not dependent on w, such that

$$\left|\sum_{j=1}^{M} \frac{\xi_j^2 - 1}{w + \left(\frac{3\pi j}{2}\right)^{2/3}}\right| \leq \frac{\mathfrak{C}'}{|w| + M^{2/3}} \sqrt{M \ln\ln M} + \sum_{j=1}^{M} \sqrt{j \ln\ln(j+2)} \frac{\mathfrak{C}' j^{-1/3}}{(|w| + j^{2/3})^2}.$$

The last expression clearly tends to 0 as $|w| \to \infty$, uniformly in M. We conclude that for large w, up to $o(1)$ error only the first term in (8.25) contributes to the $w \to \infty$ asymptotics. This term is deterministic and we can analyze it by converting sums into integrals. For large x and large w with $\mathrm{Re}(w) \geq 0$, we have

$$\sum_{j:\, \left(\frac{3\pi j}{2}\right)^{2/3} < -x} \frac{1}{w + \left(\frac{3\pi j}{2}\right)^{2/3}} = \sum_{j:\, \left(\frac{3\pi j}{2}\right)^{2/3} < -x} \left(\int_{j-1}^{j} \frac{\mathrm{d}y}{w + \left(\frac{3\pi y}{2}\right)^{2/3}} + O\left(\frac{j^{-1/3}}{\left|w + \left(\frac{3\pi j}{2}\right)^{2/3}\right|^2} \right) \right).$$

Let us upper bound the sum of the $O(\cdot)$ terms:

$$\begin{aligned} \sum_{j=1}^{\infty} \frac{j^{-1/3}}{\left|w + \left(\frac{3\pi j}{2}\right)^{2/3}\right|^2} &= \sum_{j=1}^{\lfloor |w|^{3/2} \rfloor} \frac{j^{-1/3}}{\left|w + \left(\frac{3\pi j}{2}\right)^{2/3}\right|^2} + \sum_{j=\lfloor |w|^{3/2} \rfloor + 1}^{\infty} \frac{j^{-1/3}}{\left|w + \left(\frac{3\pi j}{2}\right)^{2/3}\right|^2} \\ &\leq |w|^{3/2}|w|^{-2} + \text{const} \cdot \sum_{j=\lfloor |w|^{3/2} \rfloor + 1}^{\infty} j^{-5/3} \leq |w|^{-1/2} + \text{const} \cdot |w|^{-\frac{3}{2}\cdot\frac{2}{3}} = O(|w|^{-1/2}). \end{aligned} \tag{8.28}$$

It remains to analyze the integral, for which we change the variables $v = \left(\frac{3\pi y}{2}\right)^{2/3}$:

$$\begin{aligned} \int_0^{\frac{2}{3\pi}(-x)^{3/2}} \frac{\mathrm{d}y}{w + \left(\frac{3\pi y}{2}\right)^{2/3}} &= \frac{1}{\pi} \int_0^{-x} \frac{\sqrt{v}}{w + v} \mathrm{d}v = \frac{1}{\pi} \left[2\sqrt{v} - 2\sqrt{w} \arctan\left(\frac{\sqrt{v}}{\sqrt{w}} \right) \right]_{v=0}^{v=-x} \\ &= \frac{2}{\pi}\sqrt{-x} - \frac{2}{\pi}\sqrt{w} \arctan\left(\frac{\sqrt{-x}}{\sqrt{w}} \right). \end{aligned}$$

We conclude that

$$\begin{aligned} &\lim_{x \to -\infty} \left[\left(\sum_{j:\, \left(\frac{3\pi j}{2}\right)^{2/3} < -x} \frac{1}{w + \left(\frac{3\pi j}{2}\right)^{2/3}} \right) - \frac{2}{\pi}\sqrt{-x} \right] \\ &= -\frac{2}{\pi}\sqrt{w} \lim_{x \to -\infty} \arctan\left(\frac{\sqrt{-x}}{\sqrt{w}} \right) + O(|w|^{-1/2}) \\ &= -\frac{2}{\pi}\sqrt{w} \lim_{x \to -\infty} \left(\frac{\pi}{2} - \arctan\left(\frac{\sqrt{w}}{\sqrt{-x}} \right) \right) + O(|w|^{-1/2}) = -\sqrt{w} + O(|w|^{-1/2}). \end{aligned} \tag{8.29}$$

Plugging back into (8.25), we arrive at (8.2):

$$\mathcal{G}(w) = -\sqrt{w} + o(1), \qquad \text{as } w \to \infty \text{ with } \mathrm{Re}(w) \geq 0.$$

In order to prove (8.3), we again use (8.25). (8.26) and (8.29) imply that the sum of the first and the third terms is $-\sqrt{w} + o\left(w^{-1/4}\right)$ and it remains to analyze the second term. Note that it is a sum of mean 0 independent random variables, and therefore the Central Limit Theorem applies. Hence, it remains to compute the asymptotic variance of the sum as $w \to \infty$, which is

$$\sum_{j=1}^{\infty} \frac{2}{\left(w + \left(\frac{3\pi j}{2}\right)^{2/3}\right)^2} = (2 + o(1)) \int_0^{\infty} \frac{\mathrm{d}x}{\left(w + \left(\frac{3\pi x}{2}\right)^{2/3}\right)^2} = (2 + o(1)) \frac{1}{2\sqrt{w}}, \tag{8.30}$$

which matches the claim of (8.2). □

Proof of Proposition 8.5. An equivalent statement with a different proof can be found in Bloemendal [2011, Theorem 4.1.1]. Our proof is based on (8.3), which we restate as

$$\mathcal{G}(w) = -\sqrt{w} + w^{-1/4}\mathcal{N}(0,1) + o\left(w^{-1/4}\right), \qquad w \to +\infty. \tag{8.31}$$

Using Definition 4.2, the main computation of the proof is to replace $\mathcal{G}(w)$ with $-\Theta$ and then solve (8.31), viewed as an equation on unknown w, and treating Θ as a parameter. In this way we get the desired equivalent form of (8.4):

$$\mathcal{T}(\Theta) = w = \Theta^2 + 2\Theta^{1/2}\mathcal{N}(0,1) + o\left(\Theta^{1/2}\right), \qquad \Theta \to +\infty.$$

In order to justify the validity of this computation, we use the monotonicity of $\mathcal{G}(w)$ on $[\mathfrak{a}_1, +\infty)$. The distributional limit of $\frac{\mathcal{T}(\Theta) - \Theta^2}{2\sqrt{\Theta}}$ is obtained from the computation of the following probabilities for $t \in \mathbb{R}$, in which we used Definition 4.2:

$$\mathrm{Prob}\left(\mathcal{T}(\Theta) \le 2t\sqrt{\Theta} + \Theta^2\right) = \mathrm{Prob}\left(-\Theta \ge \mathcal{G}(2t\sqrt{\Theta} + \Theta^2), \quad \mathfrak{a}_1 \le 2t\sqrt{\Theta} + \Theta^2\right).$$

The second condition $\mathfrak{a}_1 \le 2t\sqrt{\Theta} + \Theta^2$ has probability approaching 1 as $\Theta \to \infty$, and, therefore, can be dropped. For the first condition, we use (8.31) to transform it as $\Theta \to +\infty$:

$$\begin{aligned} &\mathrm{Prob}\left(-\Theta \ge -\sqrt{2t\sqrt{\Theta} + \Theta^2} + (2t\sqrt{\Theta} + \Theta^2)^{-1/4}\mathcal{N}(0,1) + o\left((2t\sqrt{\Theta} + \Theta^2)^{-1/4}\right)\right) \\ &\quad = \mathrm{Prob}\left(-\Theta \ge -\Theta - t\Theta^{-1/2} + \Theta^{-1/2}\mathcal{N}(0,1) + o\left(\Theta^{-1/2}\right)\right) = \mathrm{Prob}\big(t \ge \mathcal{N}(0,1) + o(1)\big), \end{aligned}$$

and the last probability clearly tends to the Gaussian distribution function as $\Theta \to \infty$. □

8.2. **A class of meromorphic functions.** In this subsection we first consider deterministic meromorphic complex functions. We then introduce randomness and establish general theorems that aid in proving convergence to $\mathcal{G}(w)$.

Definition 8.10. *Given a real number $\gamma \in \mathbb{R}$, a sequence of real numbers $x_1 \ge x_2 \ge x_3 \ge \dots$ with $\lim_{n\to\infty} x_n = -\infty$, and a sequence of non-negative weights $\{w_j\}_{j=1}^{\infty}$, satisfying*

$$\sum_{j=1}^{\infty} \frac{w_j}{1 + x_j^2} < \infty, \tag{8.32}$$

we define a complex function

$$f(z) = \gamma + \sum_{j=1}^{\infty} w_j \left(\frac{1}{z - x_j} + \frac{x_j}{1 + x_j^2}\right), \qquad z \in \mathbb{C} \setminus \{x_j\}_{j=1}^{\infty}. \tag{8.33}$$

We let Ω_- denote the convex cone of all complex functions of this form. The minus in the notation Ω_- indicates that $x_n \to -\infty$.

Note that we allow some w_j to vanish, so (8.33) may reduce to a finite sum.

Lemma 8.11. *The sum* (8.33) *converges uniformly on any compact subset of $\mathbb{C} \setminus \{x_j\}_{j=1}^{\infty}$; that is, it converges in the topology $\stackrel{mer}{\to}$.*

Proof. For z in a compact set $\mathfrak{Z} \subset \mathbb{C} \setminus \{x_j\}_{j=1}^{\infty}$, we have

$$\left| w_j \left(\frac{1}{z - x_j} + \frac{x_j}{1 + x_j^2} \right) \right| = \left| \frac{1 + z x_j}{z - x_j} \cdot \frac{w_j}{1 + x_j^2} \right| \leq C(\mathfrak{Z}) \cdot \frac{w_j}{1 + x_j^2}.$$

Hence, the absolute and uniform convergence of (8.33) follows from (8.32). □

Our arguments crucially use the sublinearity of functions in Ω_-:

Lemma 8.12. *For any $f \in \Omega_-$, we have*

$$\lim_{R \to \pm\infty} \frac{f(\mathbf{i}R)}{R} = 0.$$

Proof. We have

$$f(\mathbf{i}R) - f(\mathbf{i}) = \mathbf{i}(1 - R) \sum_{j=1}^{\infty} \frac{w_j}{(\mathbf{i}R - x_j)(\mathbf{i} - x_j)}.$$

For $R > 1$, the magnitude of the sum is bounded by (8.32) and each term goes to 0 as $R \to \infty$. Hence, by the dominated convergence theorem the sum goes to 0 and $f(\mathbf{i}R) - f(\mathbf{i}) = o(R)$. Since obviously also $f(\mathbf{i}) = o(R)$ as $R \to \infty$, the conclusion follows. □

Now we state a distributional convergence theorem for random functions from Ω_-. Note that we make no assumptions on γ_n or γ.

Theorem 8.13. *Take random functions f_n, $n = 1, 2, \dots,$ and f from Ω_-, corresponding to random $(\gamma_n, \{x_{j;n}\}_{j=1}^{\infty}, \{w_{j;n}\}_{j=1}^{\infty})$ and $(\gamma, \{x_j\}_{j=1}^{\infty}, \{w_j\}_{j=1}^{\infty})$. Suppose that, in the sense of convergence in finite-dimensional distributions:*

$$\lim_{n \to \infty} x_{j;n} = x_j, \quad \textit{and} \quad \lim_{n \to \infty} w_{j;n} = w_j, \ \textit{for each } j = 1, 2, \dots, \tag{8.34}$$

and there exists a deterministic function $\phi : \mathbb{R}_+ \to \mathbb{C}$ such that for any $\varepsilon > 0$ and $R > 0$

$$\lim_{R \to +\infty} \operatorname{Prob}(|f(\mathbf{i}R) - \phi(R)| > \varepsilon) = \lim_{R \to +\infty} \limsup_{n \to \infty} \operatorname{Prob}(|f_n(\mathbf{i}R) - \phi(R)| > \varepsilon) = 0. \tag{8.35}$$

Then there exists a coupling that places all f_n and f on the same probability space such that almost surely $f_n \stackrel{mer}{\to} f$.

As noted earlier, this implies convergence in distribution $f_n \stackrel{\text{mer, d}}{\to} f$, and thus convergence in distribution of $(f_n(z_j))_{j=1}^k$ to $(f(z_j))_{j=1}^k$ at any finite set of points.

Theorem 8.13 is inspired by Aizenman and Warzel [2015], see Theorem 3.1 and Section 6 there, as well as Sodin [2018, Section 1.4]. There are, however, important differences: in Aizenman and Warzel [2015] the function $\phi(R)$ was constant – this does not hold in our main application. Additionally, our topology of convergence is stronger than that of Aizenman and Warzel [2015], which will be crucial for us when we want to solve equations of the form $f_n(z) = \theta$. Unlike Sodin [2018], we do not make use of any results from complex analysis, beyond the basic theory.

Remark 8.14. *One can also deal with several sequences of functions $f_n^{[k]}(z)$ converging towards $f^{[k]}(z)$, $k = 1, 2, \dots, K$. The tail condition* (8.35)*, the conclusion of the theorem, and the proof remain exactly the same for such extension.*

In the rest of this subsection we prove Theorem 8.13. We start with a deterministic statement.

Lemma 8.15. *Let f_n, $n = 1, 2, \dots$, and f be deterministic functions from Ω_-, corresponding to $(\gamma^n, \{x_{j;n}\}_{j=1}^{\infty}, \{w_{j;n}\}_{j=1}^{\infty})$, and $(\gamma, \{x_j\}_{j=1}^{\infty}, \{w_j\}_{j=1}^{\infty})$. Suppose that:*

$$\lim_{n\to\infty} w_{j;n} = w_j, \quad \text{and} \quad \lim_{n\to\infty} x_{j;n} = x_j, \text{ for each } j = 1, 2, \dots, \qquad \text{and} \tag{8.36}$$

$$\lim_{n\to\infty} f_n(\mathbf{i}) = f(\mathbf{i}). \tag{8.37}$$

Then $f_n \overset{mer}{\to} f$, i.e., $f_n(z) \to f(z)$, uniformly over z in compact subsets of $\mathbb{C} \setminus \{x_j\}_{j=1}^{\infty}$.

Proof. We have

$$f_n(z) - f_n(\mathbf{i}) = \sum_{j=1}^{\infty} w_{j;n} \left(\frac{1}{z - x_{j;n}} - \frac{1}{\mathbf{i} - x_{j;n}} \right) = (\mathbf{i} - z) \sum_{j=1}^{\infty} \frac{w_{j;n}}{(z - x_{j;n})(\mathbf{i} - x_{j;n})}.$$

Each term in the last series converges towards its counterpart for the series of $f(z) - f(\mathbf{i})$, and we need to produce a uniform tail bound. For that we note

$$\left| \sum_{j=M}^{\infty} \frac{w_{j;n}}{(z - x_{j;n})(\mathbf{i} - x_{j;n})} \right| \le \sum_{j=M}^{\infty} \left| \frac{\mathbf{i} + x_{j;n}}{z - x_{j;n}} \right| \cdot \frac{w_{j;n}}{1 + x_{j;n}^2} \le \left[1 + \max_{j \ge M} \left\{ \frac{|z| + 1}{|z - x_{j;n}|} \right\} \right] \sum_{j=M}^{\infty} \frac{w_{j;n}}{1 + x_{j;n}^2}.$$

The first factor stays uniformly bounded as $n \to \infty$, and it remains to show that the second factor tends to 0 as $M \to \infty$ (uniformly in n). For that we observe

$$\sum_{j=M}^{\infty} \frac{w_{j;n}}{1 + x_{j;n}^2} = -\mathrm{Im} f_n(\mathbf{i}) - \sum_{j=1}^{M-1} \frac{w_{j;n}}{1 + x_{j;n}^2}. \tag{8.38}$$

For an arbitrary $\varepsilon > 0$, we choose M large enough, so that

$$\sum_{j=M}^{\infty} \frac{w_j}{1 + x_j^2} = -\mathrm{Im} f(\mathbf{i}) - \sum_{j=1}^{M-1} \frac{w_j}{1 + x_j^2} < \varepsilon.$$

Then sending $n \to \infty$, in the finite sum in the right-hand side of (8.38), we deduce existence of n_0 such that for all $n > n_0$,

$$\sum_{j=M}^{\infty} \frac{w_{j;n}}{1 + x_{j;n}^2} < 2\varepsilon. \tag{8.39}$$

Because $\sum_{j=1}^{\infty} \frac{w_{j;n}}{1 + x_{j;n}^2} < \infty$ for each n, at the expense of increasing M, we can guarantee that (8.39) holds for all $n = 1, 2, \dots$. □

The next step is to study the value $f_n(\mathbf{i})$ which appeared in the previous lemma.

Lemma 8.16. *Under the conditions of Theorem 8.13, the random variables $f_n(\mathbf{i})$ converge in distribution as $n \to \infty$ towards $f(\mathbf{i})$, jointly with the convergence* (8.34).

Proof. We take two constants $Q > q > 1$, which are both large. Consider

$$f_n(\mathbf{i}q) - f_n(\mathbf{i}) = \sum_{j=1}^{\infty} w_{j;n} \left[\frac{1}{\mathbf{i}q - x_{j;n}} - \frac{1}{\mathbf{i} - x_{j;n}} \right] = \mathbf{i}(1-q) \sum_{j=1}^{\infty} \frac{w_{j;n}}{(\mathbf{i}q - x_{j;n}) \cdot (\mathbf{i} - x_{j;n})}. \tag{8.40}$$

Using (8.34), term-by-term the last series converges as $n \to \infty$ towards

$$f(\mathbf{i}q) - f(\mathbf{i}) = \mathbf{i}(1-q) \sum_{j=1}^{\infty} \frac{w_j}{(\mathbf{i}q - x_j) \cdot (\mathbf{i} - x_j)},$$

which is a well-defined finite random variable due to $f \in \Omega_-$ and (8.32). In order to justify the interchange of the order of summation and taking the $n \to \infty$ limit, we need to produce an additional tail bound. We observe that for large $Q - q$ we have

$$\left| \sum_{j:\, |x_{j;n}| > Q} \frac{w_{j;n}}{(\mathbf{i}q - x_{j;n}) \cdot (\mathbf{i} - x_{j;n})} \right| \le 2 \sum_{j:\, |x_{j;n}| > Q} \frac{w_{j;n}}{Q^2 + x_{j;n}^2} \le -\frac{2}{Q} \mathrm{Im}\big[f_n(\mathbf{i}Q) \big].$$

We claim that by choosing first large Q, and then large n we can make $\frac{2}{Q}\mathrm{Im}\big[f_n(\mathbf{i}Q)\big]$ arbitrarily small with probability arbitrary close to 1. Indeed, combining Lemma 8.12 with the first limit of (8.35), we conclude that $\phi(Q) = o(Q)$ for large Q. Then the second limit in (8.35) implies that for large n also $f_n(\mathbf{i}Q) = o(Q)$ with high probability.

On the other hand, using (8.34) and $\lim_{n\to\infty} |x_n| = \infty$ from the definition of Ω_-, we conclude that for large n, the part of the sum (8.40) with $|x_{j;n}| \le Q$ has only finitely many (with the number dependent only on Q) terms with high probability. It follows that we are allowed to pass to the limit $n \to \infty$ in (8.34) and we have proven that for each $q > 1$

$$\lim_{n\to\infty} \Big(f_n(\mathbf{i}q) - f_n(\mathbf{i});\, (x_{j;n}, w_{j;n})_{j=1}^{\infty} \Big) \stackrel{d}{=} \Big(f(\mathbf{i}q) - f(\mathbf{i});\, \big(x_j, w_j\big)_{j=1}^{\infty} \Big). \tag{8.41}$$

In order to finish the proof of the lemma, it remains to get rid of $f_n(\mathbf{i}q)$. For that we notice that by (8.35), $f_n(\mathbf{i}q) - f_n(\mathbf{i}) \approx \phi(q) - f_n(\mathbf{i})$ and also $f(\mathbf{i}q) - f(\mathbf{i}) \approx \phi(q) - f(\mathbf{i})$, i.e., the differences of the left-hand sides and the right-hand sides are small with probability close to 1 when q and n are large. Subtracting $\phi(q)$ from both sides, we are done. □

Proof of Theorem 8.13. By Lemma 8.16, in distribution,

$$\lim_{n\to\infty} \Big(f_n(\mathbf{i});\, (x_{j;n}, w_{j;n})_{j=1}^{\infty} \Big) \stackrel{d}{=} \Big(f(\mathbf{i});\, \big(x_j, w_j\big)_{j=1}^{\infty} \Big). \tag{8.42}$$

By the Skorokhod's representation theorem (see, e.g., Billingsley [1999, Chapter 1, Theorem 6.7]) one can construct a coupling, so that (8.42) becomes almost sure convergence. Then Lemma 8.15 implies $f_n \stackrel{\mathrm{mer}}{\to} f$. □

The setting of Theorem 8.13 is tailored to our main application – convergence to $\mathcal{G}(w)$ – which is why we assume $x_1 \ge x_2 \ge \dots$. However, we could instead consider functions of the form (8.33) constructed from a sequence $x_1 \le x_2 \le \dots$ with $x_n \to +\infty$; we denote the class of such functions by Ω_+. Similarly, we could allow both types of sequences and work with sums of the form $f(z) = f^-(z) + f^+(z)$, where $f^{\pm} \in \Omega_{\pm}$. The following theorem can be proved by repeating the argument of Theorem 8.13 word for word.

Theorem 8.17. *In Theorem 8.13, instead of $f_n, f \in \Omega_-$, we can take $f_n = f_n^+ + f_n^-$, $f = f^+ + f^-$, with $f_n^\pm, f^\pm \in \Omega_\pm$ corresponding to $(\gamma_n^\pm, \{x_{j;n^\pm}\}_{j=1}^\infty, \{w_{j;n}^\pm\}_{j=1}^\infty)$ and $(\gamma^\pm, \{x_j^\pm\}_{j=1}^\infty, \{w_j^\pm\}_{j=1}^\infty)$. Replacing* (8.34) *with*

$$\lim_{n\to\infty} x_{j;n}^\pm = x_j^\pm, \quad \text{and} \quad \lim_{n\to\infty} w_{j;n}^\pm = w_j^\pm, \text{ for each } j = 1, 2, \dots, \tag{8.43}$$

and keeping (8.35)*, the conclusion continues to hold in the same sense as in Theorem 8.13: almost surely $f_n \overset{mer}{\to} f$.*

Remark 8.18. *For both $f \in \Omega_-$ and $f \in \Omega_+$, the function $-f(z)$ belongs to the Herglotz-Nevanlinna class, meaning it maps the upper half-plane to itself. Our proofs can be adapted to handle most other Herglotz–Nevanlinna functions, as long as they exhibit sublinear growth. However, the linear Herglotz-Nevanlinna function az, $a > 0$, is excluded from our definitions: Lemma 8.12 is essential for the arguments.*

Theorems 8.13 and 8.17 can be applied in various situations in random matrix theory, where $x_{j;n}$ are eigenvalues converging to various universal scaling limits (such as Airy, sine, or Bessel point processes). When all $w_{j;n}$ equal 1, we deal with the Stieltjes transform or the log-derivative of the charateristic polynomial, connecting us to the vast literature on the latter, see, e.g., Lambert and Paquette [2020], Johnstone et al. [2025], Collins-Woodfin and Le [2025] for the most recent results related to the edge limits. In our main application, instead, $w_{j;n}$ are i.i.d. random variables.

8.3. **Convergence to $\mathcal{G}(w)$.** In this section we use Theorem 8.13 to derive sufficient conditions for convergence towards $\mathcal{G}(w)$ of Theorem 8.1. For each $N = 1, 2, \dots$, we are given an N–tuple of random variables $\lambda_{1;N} \geq \lambda_{2;N} \geq \dots \geq \lambda_{N;N}$; we assume that for each N these variables are defined on its own probability space, which therefore depends on N. In addition, for each N we are given a sequence of i.i.d. Gaussian random variables $\xi_1, \xi_2, \dots, \xi_N$ on the same N–dependent probability space (we omit an additional N from the index of ξ_i, because the distribution of ξ_i does not depend on it), which are independent of $\{\lambda_{i;N}\}_{i=1}^N$. In addition, we are given a continuous non-negative function $h(x) \geq 0$. We encode the data via the empirical Stieltjes transform:

$$m_N(z) = \frac{1}{N} \sum_{i=1}^N \frac{h(\lambda_{i;N})}{z - \lambda_{i;N}}.$$

The data is assumed to satisfy the following:

Assumption 8.19. *There exists a function $m(z)$ and constants $\lambda_+ \in \mathbb{R}$, $\mathfrak{s} > 0$, $\mathfrak{m} \in \mathbb{R}$, $\varepsilon > 0$, and $C > 0$ such that*

(i) In the sense of convergence of finite-dimensional distributions:

$$\lim_{N\to\infty} \left\{ N^{2/3} \mathfrak{s}^{2/3} \left(\lambda_{j;N} - \lambda_+\right)\right\}_{j=1}^N \overset{d}{=} \{\mathfrak{a}_j\}_{j=1}^\infty, \tag{8.44}$$

where $\mathfrak{a}_1 \geq \mathfrak{a}_2 \geq \dots$ is the Airy$_1$ point process.

(ii) As $R \to 0$ we have

$$m(\lambda_+ + \mathbf{i}R) = \mathfrak{m} - \mathfrak{s}\frac{1+\mathbf{i}}{\sqrt{2}}\sqrt{R} + o\left(1\right). \tag{8.45}$$

(iii) For all $N = 1, 2, \ldots$ and all $N^{-2/3} < R < N^{-2/3+\varepsilon}$ we should have

$$\mathbb{E}\left[|m_N(\lambda_+ + \mathbf{i}R) - m(\lambda_+ + \mathbf{i}R)|^2\right] \le \frac{C}{N^2 R^2}. \tag{8.46}$$

While not required by the assumptions, it is typical to have

$$m(z) = \int_{\mathbb{R}} \frac{h(x)}{z - x} \mu(x) \mathrm{d}x,$$

where $\mu(x)\mathrm{d}x$ is an independent of N probability measure $\mu(x)\mathrm{d}x$, supported on a finite interval $[\lambda_-, \lambda_+] \subset \mathbb{R}$, and which is a limit of the empirical measures of $\lambda_{i;N}$.

The first condition in Assumption 8.19 specifies the edge limit of $\lambda_{i;N}$; the second condition is related to the square root behavior $h(x)\mu(x) \approx \frac{\mathfrak{s}}{\pi}\sqrt{\lambda_+ - x}$ for x close to λ_+ from the left; the third condition is the optimal local law meaning that the empirical measure of $\lambda_{i;N}$ is close to the limit given by $\mu(x)\mathrm{d}x$.

We denote

$$\mathcal{G}_N(w) = N^{1/3}\mathfrak{s}^{-2/3} \left(\frac{1}{N} \sum_{j=1}^{N} \frac{h(\lambda_{j;N})\, \xi_j^2}{\lambda_+ + N^{-2/3}\mathfrak{s}^{-2/3} w - \lambda_{j;N}} - \mathfrak{m} \right). \tag{8.47}$$

The next theorem asserts that one can couple the processes defined for the different values of N in such a way that $\mathcal{G}_N$ converge to the Airy-Green function.

Theorem 8.20. *Under Assumption 8.19, there exists a coupling that places random variables $\left(\{\lambda_{j;N}\}_{j=1}^N, \{\xi_j\}_{j=1}^N\right)_{N=1}^{\infty}$ and the Airy-Green function $\mathcal{G}(w)$ on the same probability space, such that almost surely*

$$\mathcal{G}_N(w) \overset{mer}{\to} h(\lambda_+) G(w), \qquad N \to \infty. \tag{8.48}$$

Corollary 8.21. *For any finitely many deterministic $w_1, \ldots, w_k \in \mathbb{C}$, the convergence in (8.48) holds in joint distribution over $w = w_1, \ldots, w = w_k$.*

Corollary 8.22. *Fix $\theta \in \mathbb{R}$ and let $\tilde{w}_N$ denote the largest real solution of the equation $\mathcal{G}_N(w) = \theta$ and let $\tilde{w}$ denote the largest real solution of the equation $\mathcal{G}(w) = \theta$. Under Assumption 8.19,*

$$\lim_{N \to \infty} \tilde{w}_N \overset{d}{=} \tilde{w}. \tag{8.49}$$

Remark 8.23. *For some settings we might have more than one sequence ξ_j and more than one choice for $h(x)$. One can take finite K and consider simultaneous edge limits for*

$$\frac{1}{N} \sum_{j=1}^{N} \frac{h^{(k)}(\lambda_{j;N}) \cdot (\xi_j^{(k)})^2}{z - \lambda_{j;N}}, \qquad k = 1, 2, \ldots, K,$$

where $\{\xi_j^{(k)}\}$ are all i.i.d. Gaussian $\mathcal{N}(0,1)$ over j, but might be correlated over k. Theorem 8.20 and Corollaries 8.21, 8.22 have immediate extensions to such setting, where each of the k functions converges to its own $\mathcal{G}^{(k)}(w)$, constructed by the same $\{\mathfrak{a}_j\}$, but distinct sequences $\{\xi_j\} = \{\xi_j^{(k)}\}$, which might be correlated over k. The convergence would be joint over $k = 1, 2, \ldots, K$, while the proof remains exactly the same.

Proof of Theorem 8.20. We want to apply Theorem 8.13 with $n=N$, $f_n=\mathcal{G}_N$, $f=h(\lambda_+)\cdot\mathcal{G}$.

Condition 1. We need to check that the functions $\mathcal{G}_N$ and $\mathcal{G}$ are almost surely in the class Ω_-. For $\mathcal{G}_N$ this is immediate. $\mathcal{G}$ also belongs to Ω_-, as can be seen by rewriting the definition:

$$\begin{aligned}\mathcal{G}(z) &= \lim_{x\to-\infty}\left[\left(\sum_{j:\,\mathfrak{a}_j>x}\frac{\xi_j^2}{z-\mathfrak{a}_j}\right)-\frac{2}{\pi}\sqrt{-x}\right]\\ &= \lim_{x\to-\infty}\sum_{j:\,\mathfrak{a}_j>x}\xi_j^2\left[\frac{1}{z-\mathfrak{a}_j}+\frac{\mathfrak{a}_j}{1+\mathfrak{a}_j^2}\right]+\lim_{x\to-\infty}\left(\sum_{j:\,\mathfrak{a}_j>x}\frac{-\mathfrak{a}_j}{1+\mathfrak{a}_j^2}-\frac{2}{\pi}\sqrt{-x}\right)\end{aligned}$$

and noting that almost sure convergence of $\sum_{j=1}^{\infty}\frac{\xi_j^2}{1+\mathfrak{a}_j^2}$ follows from Lemma 8.9 and monotone convergence theorem (applied conditionally on the values of $\{\mathfrak{a}_j\}_{j=1}^{\infty}$).

Condition 2. Limits (8.34) are included in assumption (8.44) for the particle positions, while the distribution of $w_{j;n}=h(\lambda_{j;N})\xi_j^2$ converges to that of $h(\lambda_+)\xi_j^2$ by continuity of $h(x)$.

Condition 3. It remains to check (8.35). We set $\phi(R)=-\frac{1+\mathbf{i}}{\sqrt{2}}\sqrt{R}$. Proposition 8.3 yields that $\mathcal{G}(\mathbf{i}R)+\frac{1+\mathbf{i}}{\sqrt{2}}\sqrt{R}$ goes to 0 as $R\to\infty$ almost surely, and, therefore, also in probability, verifying the first limit in (8.35). For the second limit we prove that

$$\lim_{R\to+\infty}\limsup_{N\to\infty}\mathbb{E}\left|\mathcal{G}_N(\mathbf{i}R)+\frac{1+\mathbf{i}}{\sqrt{2}}\sqrt{R}\right|^2=0. \tag{8.50}$$

By the Markov inequality, (8.50) is sufficient for establishing (8.35). The limit (8.50) is a corollary of the assumptions (8.45) and (8.46). Indeed, taking the expectation with respect to ξ_j first and using $\mathbb{E}(\xi_j^2-1)^2=2$, we have

$$\begin{aligned}\mathbb{E}\left|\mathcal{G}_N(\mathbf{i}R)+\frac{1+\mathbf{i}}{\sqrt{2}}\sqrt{R}\right|^2 &= \mathbb{E}\left(\mathcal{G}_N(\mathbf{i}R)+\frac{1+\mathbf{i}}{\sqrt{2}}\sqrt{R}\right)\left(\overline{\mathcal{G}_N(\mathbf{i}R)}+\frac{1-\mathbf{i}}{\sqrt{2}}\sqrt{R}\right)\\ &= \mathbb{E}\left|N^{1/3}\mathfrak{s}^{-2/3}m_N\left(\lambda_++\frac{\mathbf{i}R}{N^{2/3}\mathfrak{s}^{2/3}}\right)-\mathfrak{m}N^{1/3}\mathfrak{s}^{-2/3}+\frac{1+\mathbf{i}}{\sqrt{2}}\sqrt{R}\right|^2\\ &\quad+\mathbb{E}\sum_{j=1}^{N}\frac{2h^2(\lambda_{j;N})}{\left|\mathbf{i}R-N^{2/3}\mathfrak{s}^{2/3}(\lambda_{j;N}-\lambda_+)\right|^2}\\ &= \mathbb{E}\Bigg|N^{1/3}\mathfrak{s}^{-2/3}m\left(\lambda_++\frac{\mathbf{i}R}{N^{2/3}\mathfrak{s}^{2/3}}\right)-\mathfrak{m}N^{1/3}\mathfrak{s}^{-2/3}+\frac{1+\mathbf{i}}{\sqrt{2}}\sqrt{R}\\ &\quad+N^{1/3}\mathfrak{s}^{-2/3}\left(m_N\left(\lambda_++\frac{\mathbf{i}R}{N^{2/3}\mathfrak{s}^{2/3}}\right)-m\left(\lambda_++\frac{\mathbf{i}R}{N^{2/3}\mathfrak{s}^{2/3}}\right)\right)\Bigg|^2\\ &\quad+\frac{N^{1/3}\mathfrak{s}^{-2/3}}{\mathbf{i}R}\mathbb{E}\left[m_N\left(\lambda_+-\frac{\mathbf{i}R}{N^{2/3}\mathfrak{s}^{2/3}}\right)-m_N\left(\lambda_++\frac{\mathbf{i}R}{N^{2/3}\mathfrak{s}^{2/3}}\right)\right].\end{aligned} \tag{8.51}$$

The fourth line of (8.51) becomes small as $R \to \infty$ by (8.45), in which one needs to rescale R. The fifth line becomes small by (8.46), again with rescaled R. The sixth line becomes small as $R \to \infty$ by a combination of (8.45) and (8.46). $\square$

Proof of Corollary 8.21. Take $W = \{w_1, w_2, \dots, w_k\}$. The distribution of each $\mathfrak{a}_j$ is absolutely continuous (e.g., because there is a well-defined density of the first correlation measure (8.10)), and therefore almost surely no $\mathfrak{a}_j$ belong to W. Hence, $(\mathcal{G}_N(w_1), \dots, \mathcal{G}_N(w_k))$ converges towards $(\mathcal{G}(w_1), \dots, \mathcal{G}(w_k))$ almost surely and, therefore, also in distribution. $\square$

Proof of Corollary 8.22. Directly subtracting the values at two points inside the definition of $\mathcal{G}(w)$, we see that it is a strictly monotone decreasing function of real $w > \mathfrak{a}_1$. Proposition 8.3 implies that it varies from $+\infty$ to $-\infty$ on $(\mathfrak{a}_1, +\infty)$. We conclude that for each $\theta \in \mathbb{R}$, there exists a unique random $\tilde{w} > \mathfrak{a}_1$, such that $\mathcal{G}(\tilde{w}) = \theta$.

Next, we consider the coupling of Theorem 8.20 and note that by Rouché's theorem (see e.g., Ahlfors [1979]) locally-uniform convergence of meromorphic functions implies the convergence of their zeros. Hence, $\tilde{w}_N \to \tilde{w}$ almost surely, and, therefore, also in distribution. $\square$

8.4. **Universal bound for β–ensembles.** In order to verify the conditions of Theorem 8.20, we are going to use the following universal bound valid for all β–ensembles. Take $\beta > 0$, a function $V(x) : [a, b] \to \mathbb{R}$, often called a potential, and consider for each $N = 1, 2, \dots,$ a probability measure on $\{\lambda_i\}_{i=1}^N$, such that $b \geq \lambda_1 > \lambda_2 > \dots > \lambda_N \geq a$, with probability density proportional to

$$\prod_{1 \leq i < j \leq N} (\lambda_i - \lambda_j)^\beta \prod_{i=1}^N \exp\left(-\frac{\beta N}{2} V_N(\lambda_i)\right). \tag{8.52}$$

We will be interested in the case $\beta = 1$ and the following three options for $V_N(\lambda)$, although the theorems we use hold in much greater generality.

- GOE ensemble. The distribution of the eigenvalues of $\mathcal{E} = \frac{1}{\sqrt{2N}}(\mathcal{Z} + \mathcal{Z}^\mathsf{T})$ with $\sigma^2 = 1$ as in Section 2.1 takes the form (8.52) with (see, e.g., Forrester [2010, Chapter 1], Pastur and Shcherbina [2011, Chapter 4]):
$$a = -\infty, \quad b = +\infty, \quad V_N(\lambda) = \frac{\lambda^2}{2}.$$
- LOE ensemble. The distribution of the eigenvalues of $\frac{1}{S} X X^\mathsf{T}$, where $S \geq N$ and X is $N \times S$ matrix of i.i.d. $\mathcal{N}(0,1)$, as in Sections 2.2 and 2.3 in the case of no signals θ, takes the form (8.52) with (see, e.g., Forrester [2010, Chapter 1], Pastur and Shcherbina [2011, Chapter 7]):
$$a = 0, \quad b = +\infty, \quad V_N(\lambda) = -\frac{S - N - 1}{N} \ln(\lambda) + \frac{S}{N} \lambda.$$
- JOE ensemble. The distribution of the squared sample canonical correlations between independent matrices $\mathbf{U}$ and $\mathbf{V}$ of dimensions $N \times S$ and $M \times S$, respectively, and filled with i.i.d. $\mathcal{N}(0,1)$, as in Section 2.4 in the case of no signals θ, with $N \leq M$, $M + N \leq S$ takes the form (8.52) with (see e.g. Forrester [2010, Section 3.6.1])
$$a = 0, \quad b = 1, \quad V_N(\lambda) = -\frac{M - N - 1}{N} \ln(\lambda) - \frac{S - N - M - 1}{N} \ln(1 - \lambda).$$

We let μ_V denote the equilibrium measure or limit shape for each of the ensembles, which is a deterministic measure, approximating the (random) empirical measure $\frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i}$ for large N. Explicitly, its density is given by:

- For GOE, μ_V is the semicirle law of density $\frac{1}{2\pi}\sqrt{4-x^2}\,\mathbf{1}_{[-2,2]}\,\mathrm{d}x$.
- For LOE, μ_V is the Marchenko-Pastur law of density $\frac{1}{2\pi}\frac{\sqrt{(\lambda_+-x)(x-\lambda_-)}}{\gamma^2 x}$, where $(1\pm\gamma)^2$ and the parameter γ satisfies $\frac{N}{S}=\gamma^2+O(1)$ as $N\to\infty$.
- For JOE, μ_V is the Wachter law of density $\frac{\tau_N}{2\pi}\frac{\sqrt{(x-\lambda_-)(\lambda_+-x)}}{x(1-x)}\mathbf{1}_{[\lambda_-,\lambda_+]}\,\mathrm{d}x$, where $\lambda_\pm = \left(\sqrt{\tau_M^{-1}(1-\tau_N^{-1})}\pm\sqrt{\tau_N^{-1}(1-\tau_M^{-1})}\right)^2$ and the parameters τ_N, τ_N satisfy $\frac{S}{N}=\tau_N+O(1)$, $\frac{S}{M}=\tau_M+O(1)$ as $N\to\infty$.

In all three cases $[\lambda_-,\lambda_+]$ denotes the support of μ_V. Let us also introduce the empirical and limiting Stieltjes transforms:

$$m_N(z)=\frac{1}{N}\sum_{i=1}^N\frac{1}{z-\lambda_i},\qquad m_V(z)=\int_{\lambda_-}^{\lambda_+}\frac{\mu_V(\mathrm{d}x)}{z-x}. \tag{8.53}$$

In all situation of interest the function $m_V(z)$ is explicit and satisfies (8.45). We will present the formulas for $m_V(z)$ when they are needed.

Theorem 8.24. *For each of GOE, LOE, JOE, there exist constants $\tilde\eta>0$ and $C>0$ (depending on γ, τ_N, τ_M in a continuous way), such that for all $N\ge 1$, $q=1,2,\dots,$ and $z=E+\mathbf{i}\eta$ with $0<\eta<\tilde\eta$, $\lambda_- -\tilde\eta<E<\lambda_+ +\tilde\eta$, we have*

$$\mathbb{E}\,|m_N(z)-m_V(z)|^q\le\left(\frac{Cq^2}{N\eta}\right)^q. \tag{8.54}$$

This theorem under the name "Optimal Local Law" can be found in Bourgade et al. [2022, Proposition 3.5], [Huang and Zhang, 2024, Remark 7.8], and Alt et al. [2025], where general β-ensembles (8.52) are analyzed. Related statements can also be found in many other sources, e.g., Bao et al. [2019, Lemma B.2 on arXiv or Lemma S5.2 in the supplement to the published version] and Yang [2022a, Theorem 2.14] contain slightly weaker bounds for the CCA setting.

Theorem 8.24 readily implies the condition (8.46) of Assumption 8.19 for the case $h(x)=1$. Several more forms of $h(x)$ will also be needed.

Corollary 8.25. *For LOE or JOE ensembles, the bound* (8.54) *extends to the case $h(x)=\sqrt{x}$, in the following form: there exist $\tilde\eta>0$ and $C>0$, such that for all $N\ge 1$, and $z=E+\mathbf{i}\eta$ with $0<\eta<\tilde\eta$, $\max(\lambda_- -\tilde\eta,0)<E<\lambda_+ +\tilde\eta$, we have*

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^N\frac{\sqrt{\lambda_i}}{z-\lambda_i}-\int_{\lambda_-}^{\lambda_+}\frac{\sqrt{x}\mu_V(\mathrm{d}x)}{z-x}\right]^2\le\frac{C}{N^2\eta^2}+\frac{C\ln^2(N)}{N^2}. \tag{8.55}$$

Remark 8.26. *It is plausible that the second term in the right-hand side of* (8.55) *can be dropped. Yet, the current form is sufficient for us, as it clearly implies* (8.46).

Proof of Corollary 8.25. Note a deterministic bound $\left|\frac{1}{z-\lambda_i}\right| \le \frac{1}{\eta}$. This bound allows us to discard (for the purpose of proving (8.55)) various events, whose probabilities are decaying fast with N. We will do this several times in the proof.

Let us first assume that $\lambda_- > 0$ (which happens whenever p is bounded away from 0). By the large deviations principle for the support of the empirical measure of eigenvalues (see, e.g., Borot and Guionnet [2013, Proposition 2.1]) for each $\delta > 0$, with probability exponentially close to 1 for large N all the eigenvalues $\lambda_1, \dots, \lambda_N$ lie in the positive segment $[\lambda_- - \delta, \lambda_+ + \delta] \subset (0, +\infty)$. We choose a contour Γ enclosing both this segment and a point z, intersecting real axis in positive points, and lying inside the domain of validity of (8.54). We write using (8.53) and residue expansion of the contour integral:

$$\frac{1}{N}\sum_{i=1}^{N} \frac{\sqrt{\lambda_i}}{z-\lambda_i} = \frac{1}{2\pi \mathbf{i}} \oint_\Gamma \frac{m_N(u)\sqrt{u}}{u-z} \mathrm{d}u - m_N(z)\sqrt{z}, \tag{8.56}$$

$$\int_{\lambda_-}^{\lambda_+} \frac{\sqrt{x}\mu_V(\mathrm{d}x)}{z-x} = \frac{1}{2\pi \mathbf{i}} \oint_\Gamma \frac{m_V(u)\sqrt{u}}{u-z} \mathrm{d}u - m_V(z)\sqrt{z}. \tag{8.57}$$

Using the inequality $|a+b|^2 \le 2|a|^2 + 2|b|^2$, it is sufficient to separately analyze the first and second terms. The second moment of the difference of the second terms in (8.56) and (8.57) is upper-bounded by (8.54). For the first terms, we split the integration contour into two parts: Γ_1 is the part at distance at most $1/N$ from the real axis and Γ_2 is the remaining part away from the real axis. Hence, it remains to upper-bound

$$\mathbb{E}\left|\int_{\Gamma_1} \frac{(m_N(u)-m_V(u))\sqrt{u}}{u-z}\mathrm{d}u\right|^2 + \mathbb{E}\left|\int_{\Gamma_2} \frac{(m_N(u)-m_V(u))\sqrt{u}}{u-z}\mathrm{d}u\right|^2. \tag{8.58}$$

For the first term in (8.58), the integrand is upper-bounded by a deterministic constant, because Γ_1 is away from $[\lambda_- - \delta, \lambda_+ + \delta]$ by construction and we ignore the event when eigenvalues are not in this segment, as explained at the beginning of the proof. The length of Γ_1 is of order $1/N$, and therefore the first term is upper-bounded by $C_1 N^{-2}$.

For the second term in (8.58), using $\mathbb{E}|ab| \le (\mathbb{E}|a|^2)^{1/2}(\mathbb{E}|b|^2)^{1/2}$ and the bound (8.54), we produce an estimate:

$$\begin{aligned} \mathbb{E}\int_{\Gamma_2} &\frac{(m_N(u)-m_V(u))\sqrt{u}}{u-z}\mathrm{d}u \int_{\Gamma_2} \frac{\overline{(m_N(v)-m_V(v))\sqrt{v}}}{\overline{v}-\overline{z}}\mathrm{d}\overline{v} \\ &= \iint_{\Gamma_2\times\Gamma_2} \mathbb{E}\left[(m_N(u)-m_V(u))\overline{(m_N(v)-m_V(v))}\right] \frac{\sqrt{u}\overline{\sqrt{v}}}{(u-z)(\overline{v}-\overline{z})}\mathrm{d}u\mathrm{d}\overline{v} \\ &\qquad\qquad \le \frac{C_2}{N^2}\iint_{\Gamma_2\times\Gamma_2} \frac{1}{|\mathrm{Re}\, u||\mathrm{Re}\, v|}|\mathrm{d}u||\mathrm{d}v| \le C_3 \frac{\ln^2(N)}{N^2}. \end{aligned} \tag{8.59}$$

Summing three upper bounds, we arrive at (8.55).

In the case $\lambda_- = 0$ we need to be more careful, because if we argue in the same way as for $\lambda_- > 0$, then Γ has to loop around 0 and $\sqrt{u}$ is no longer holomorphic on the contour. Instead, we choose the contour Γ passing directly through 0 and enclosing $(0, \lambda_+ + \delta)$. Compared to the previous argument, it is no longer true that on the part of the contour at distance $\frac{1}{N}$ from 0 the integrand is upper-bounded by a constant. However, in this case the splitting of the contour Γ is not required at all: the $1/\eta$ singularity of the bound (8.54) is compensated by $\sqrt{u}$ factor in the integrand, and the resulting expression $\eta^{-1/2}$ is integrable at $\eta = 0$. $\square$

Corollary 8.27. *For JOE ensembles, the bound* (8.54) *further extends to the cases* $h(x) = \sqrt{1-x}$ *and* $h(x) = \sqrt{x(1-x)}$, *in the same form: there exist* $\tilde\eta > 0$ *and* $C > 0$, *such that for all* $N \geq 1$, *and* $z = E + \mathbf{i}\eta$ *with* $0 < \eta < \tilde\eta$, $\max(\lambda_- - \tilde\eta, 0) < E < \min(\lambda_+ + \tilde\eta, 1)$, *we have*

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N}\frac{h(\lambda_i)}{z-\lambda_i} - \int_{\lambda_-}^{\lambda_+}\frac{h(x)\mu_V(\mathrm{d}x)}{z-x}\right]^2 \leq \frac{C}{N^2\eta^2} + \frac{C\ln^2(N)}{N^2}. \tag{8.60}$$

The proof for Corollary 8.27 is the same as for Corollary 8.25 and is omitted.

We also need to check the same condition for finite-rank perturbations of β–ensembles (8.52), which, in fact, follows automatically as the following lemma asserts.

Lemma 8.28. *There exist two positive constants* e_1 *and* e_2 *such that the following holds. Suppose that* λ_i^N, $1 \leq i \leq N$, *satisfies* (8.46) *of Assumption 8.19 with* $C = C_1$ *and* μ_i^N *interlaces with it, which means either*

$$\mu_1^N \geq \lambda_1^N \geq \mu_2^N \cdots \geq \mu_N^N \geq \lambda_N^N \qquad \text{or} \qquad \lambda_1^N \geq \mu_1^N \geq \lambda_2^N \cdots \geq \lambda_N^N \geq \mu_N^N. \tag{8.61}$$

Then μ_i^N, $1 \leq i \leq N$, *satisfies* (8.46) *with* $C = e_1C_1 + e_2$. *This holds with* $h(x) = 1$; *additionally with* $h(x) = \sqrt{x}$ *if the eigenvalues are positive; and with* $h(x) = \sqrt{1-x}$ *and* $h(x) = \sqrt{x(1-x)}$ *if the eigenvalues lie in the interval* $[0, 1]$.

Proof. We only consider the case $h(x) = 1$, as other cases are similar. We would like to upper bound the absolute value of the difference,

$$\frac{1}{N}\sum_{i=1}^{N}\frac{1}{z-\lambda_{i;N}} - \frac{1}{N}\sum_{i=1}^{N}\frac{1}{z-\mu_{i;N}}.$$

The real part of the difference is

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{Re}z-\lambda_{i;N}}{(\mathrm{Re}z-\lambda_{i;N})^2+(\mathrm{Im}z)^2} - \frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{Re}z-\mu_{i;N}}{(\mathrm{Re}z-\mu_{i;N})^2+(\mathrm{Im}z)^2}. \tag{8.62}$$

Note that the function $x \mapsto \frac{\mathrm{Re}z-x}{(\mathrm{Re}z-x)^2+(\mathrm{Im}z)^2}$ is monotone on each of the three segments $x \in (-\infty, \mathrm{Re}z - |\mathrm{Im}z|]$, $x \in (\mathrm{Re}z - |\mathrm{Im}z|, \mathrm{Re}z + |\mathrm{Im}z|)$, $x \in [\mathrm{Re}z + |\mathrm{Im}z|, +\infty)$. We split the first sum in (8.62) into three parts, corresponding to $\lambda_{i;N}$ in each of these three segments. Let N_1, N_2, N_3 be the numbers of terms in the corresponding parts. Using the interlacement (8.61) we match the N_1 terms in the first segment to $N_1 - 1$ terms in the second sum in (8.62) corresponding to $\mu_{i;N}$ which are between those N_1 terms. By monotonicity, the difference between the two subsums is at most $\frac{1}{N}\max_{x\in\mathbb{R}}\frac{\mathrm{Re}z-x}{(\mathrm{Re}z-x)^2+(\mathrm{Im}z)^2} = \frac{1}{2N|\mathrm{Im}z|}$. Similarly, the N_2 terms in the second segment are matched to $N_2 - 1$ terms in the second sum in (8.62) again leading to the difference between the two subsums at most $\frac{1}{2N|\mathrm{Im}z|}$. The same is true for the remaining N_3 terms in the third segment, matched to $N_3 - 1$ terms in the second sum in (8.62). Note that as a result of this procedure, three terms in the second sum of (8.62) remained unmatched. Their contribution is upper bounded by $3 \cdot \frac{1}{2N|\mathrm{Im}z|}$. We conclude that

$$\left|\mathrm{Re}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{z-\lambda_{i;N}} - \frac{1}{N}\sum_{i=1}^{N}\frac{1}{z-\mu_{i;N}}\right)\right| \leq \frac{3}{N|\mathrm{Im}z|}.$$

For the imaginary part, the argument is similar: we need to bound

$$\frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{Im}z}{(\mathrm{Re}z-\lambda_{i;N})^2+(\mathrm{Im}z)^2}-\frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{Im}z}{(\mathrm{Re}z-\mu_{i;N})^2+(\mathrm{Im}z)^2},$$

which is done by noticing that the function $x\mapsto\frac{\mathrm{Re}z-x}{(\mathrm{Re}z-x)^2+(\mathrm{Im}z)^2}$ is monotone on each of the two segments $x\in(-\infty,\mathrm{Re}z]$, $x\in(\mathrm{Re}z,+\infty)$. Repeating the real part arguments, we get

$$\left|\mathrm{Im}\left(\frac{1}{N}\sum_{i=1}^{N}\frac{1}{z-\lambda_{i;N}}-\frac{1}{N}\sum_{i=1}^{N}\frac{1}{z-\mu_{i;N}}\right)\right|\le\frac{3}{N|\mathrm{Im}z|}.$$

Hence, using the inequality $|a+b|^2\le 2|a|^2+2|b|^2$, we get

$$\left|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_++\mathbf{i}R-\mu_{i;N}}-m(\lambda_++\mathbf{i}R)\right|^2$$
$$\le 2\left|\frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_++\mathbf{i}R-\lambda_{i;N}}-m(\lambda_++\mathbf{i}R)\right|^2+\frac{36}{N^2R^2}.$$

Taking expectation and using (8.46) for $\lambda_{i;N}$, we are done. □

9. Appendix B: Applications of general theorem to four models

In this section we use Theorem 8.13 to prove Theorem 4.6. The proofs for all four settings follow similar outlines, and we present the most detailed exposition for spiked Wigner matrices. The difficulty of the argument increases steadily from the Wigner to the CCA case, and we recommend sequential reading.

9.1. **Spiked Wigner matrices.**

For the spiked Wigner case of Theorem 4.6, our first task is to introduce an equation governing the change of the eigenvalues of a symmetric matrix under a rank one perturbation.

Take $N\times N$ real symmetric matrix $\mathbf{B}$ with eigenvalues $\lambda_1\ge\lambda_2\ge\dots\ge\lambda_N$ and normalized eigenvectors $\mathbf{u}_i$, $1\le i\le N$, satisfying $\langle\mathbf{u}_i,\mathbf{u}_j\rangle=\delta_{i=j}$. In addition, take a column-vector $\mathbf{u}^*$. For a real constant $\theta\ne 0$ we define

$$\mathbf{A}=\theta\cdot\mathbf{u}^*(\mathbf{u}^*)^{\mathsf{T}}+\mathbf{B}.$$

Proposition 9.1. *For each eigenvalue a of $\mathbf{A}$, either*

$$\frac{1}{\theta}=\sum_{i=1}^{N}\frac{\langle\mathbf{u}_i,\mathbf{u}^*\rangle^2}{a-\lambda_i};\tag{9.1}$$

or $a=\lambda_j$ for $1\le j\le N$, where λ_j is an eigenvalue of $\mathbf{B}$ of multiplicity one, $\langle\mathbf{u}^*,\mathbf{u}_j\rangle=0$, and the equation (9.1) holds with the j-th term excluded; or $a=\lambda_j$, where λ_j is an eigenvalue of $\mathbf{B}$ of multiplicity larger than one.

We omit the proof, see, e.g., Jones et al. [1978] or Arbenz et al. [1988].

Remark 9.2. *We mostly use Proposition 9.1 in the generic situations of all distinct λ_i and all non-zero $\langle\mathbf{u}_i,\mathbf{u}^*\rangle$. Hence (9.1) will hold.*

Corollary 9.3. *The eigenvalues of $\mathbf{A}$ and $\mathbf{B}$ interlace: if $\mu_1\ge\dots\ge\mu_N$ are the eigenvalues of $\mathbf{A}$, then $\mu_1\ge\lambda_1\ge\mu_2\ge\dots$ for $\theta>0$ and $\lambda_1\ge\mu_1\ge\lambda_2\ge\dots$ for $\theta<0$.*

Proof. Assuming without loss of generality that all λ_i are distinct and for all i we have $\langle \mathbf{u}_i, \mathbf{u}^* \rangle \neq 0$ (general case is obtained by a limit transition), note that (9.1) is a polynomial equation on a of degree N, and therefore has N roots which are $\mu_1, \ldots, \mu_N$. Tracking the sign changes of the difference between the RHS and the LHS on intervals $(-\infty, \lambda_N)$, $(\lambda_N, \lambda_{N-1})$,..., $(\lambda_1, +\infty)$, we localize the roots, so that they satisfy the desired interlacements. □

The plan for the rest of the section is to recursively use (9.1) to produce an inductive proof of Theorem 4.6 for the spiked Wigner model by growing r. On each step, we need to know the law of the scalar products $\langle \mathbf{u}_i, \mathbf{u}^* \rangle$ appearing in the equation. This is simplified by the following observation:

Lemma 9.4. *One can replace the vectors $\mathbf{u}_i^*$ in* (2.1) *with r vectors uniformly distributed on the N–dimensional unit sphere, independent from each other and from $\mathcal{E}$: Theorem 4.6 for deterministic orthonormal vectors as in Section 2.1 is equivalent to the same theorem for independent unit vectors.*

Proof. We first note that the asymptotics in Theorem 4.6 does not depend on the choice of the unit vectors $\mathbf{u}_i^*$ in (2.1) as long as they are orthogonal to each other. Indeed, any r orthogonal vectors can be obtained from any other r orthogonal vectors by an orthogonal transformation of the space. Such transformation does not change the eigenvalues of $\mathbf{A}$, and also does not change the probability distribution of the matrix $\mathcal{E}$ (which uses Gaussianity of its matrix elements).

Now let $\mathbf{v}_1^*, \ldots, \mathbf{v}_r^*$ be r independent vectors uniformly distributed on the unit sphere. We consider the matrix $M_r = \sum_{i=1}^r \theta_i \mathbf{v}_i^* (\mathbf{v}_i^*)^\mathsf{T}$ and would like to decompose it as $M_r = \sum_{i=1}^r \theta_i' \mathbf{w}_i^* (\mathbf{w}_i^*)^\mathsf{T}$ with orthonormal vectors $\mathbf{w}_i^*$; $\mathbf{v}_i^*$ were not orthonormal. Clearly, θ_i' are non-zero eigenvalues of M_r and $\mathbf{w}_i^*$ are corresponding eigenvectors. We claim that

$$\theta_i = \theta_i' + O\left(\frac{1}{N}\right), \qquad N \to \infty. \tag{9.2}$$

Note that (9.2) implies the statement of Lemma 9.4, because addition of $O(1/N)$ does not change any of the asymptotics statements of Theorem 4.6. Hence, it remains to prove (9.2). This can be done by induction on r. Using (9.1) with $\mathbf{B} = M_{r-1}$, the non-zero eigenvalues of M_r solve an equation

$$\frac{1}{\theta_r} = \sum_{i=1}^{r-1} \frac{\langle \mathbf{w}_i, \mathbf{v}_r^* \rangle^2}{a - \theta_i'} + \sum_{i=r}^{N} \frac{\langle \mathbf{w}_i, \mathbf{v}_r^* \rangle^2}{a}, \tag{9.3}$$

where $\mathbf{w}_i$ are eigenvectors of M_{r-1} and θ_i' are non-zero eigenvalues. Representing the unit vector $\mathbf{v}_r^*$ as a vector with i.i.d. $\mathcal{N}(0,1)$ components divided by its length, using the Law of Large Numbers and the induction hypothesis, the equation is rewritten as

$$\frac{1}{\theta_r} = \frac{1}{N} \sum_{i=1}^{r-1} \frac{\chi_i^2 (1 + O(\frac{1}{\sqrt{N}}))}{a - \theta_i + O(\frac{1}{N})} + \frac{1 + O(\frac{1}{N})}{a}, \tag{9.4}$$

where χ_i are i.i.d. $\mathcal{N}(0,1)$ random variables. (9.4) is a polynomial equation on a of degree r, which clearly has r roots of the form $\theta_i + O(\frac{1}{N})$, thus proving (9.2). □

The induction can proceed in various orders of spike addition, and we choose to first add all subcritical and supercritical spikes, and then add the critical spike with index q at the very end. Hence, as an intermediate statement we have the following:

Proposition 9.5. *Consider the spiked Wigner model* $\mathbf{A} = \sum_{i=1}^r \theta_i \cdot \mathbf{u}_i^* (\mathbf{u}_i^*)^\mathsf{T} + \mathcal{E}$ *of Section 2.1, where* $\mathcal{E} = \frac{1}{\sqrt{2N}}(\mathcal{Z} + \mathcal{Z}^\mathsf{T})$*, with* $\mathcal{Z}$ *being* $N \times N$ *matrix of i.i.d.* $\mathcal{N}(0,1)$*, and* $\theta_1 > \theta_2 > \cdots > \theta_r$ *split into two groups:* $\theta_1, \dots, \theta_{q-1} > \theta^c = 1$ *and* $\theta_q, \dots, \theta_r < \theta^c = 1$*. Then, in the sense of convergence in joint distribution and using* (2.2)*:*

$$\lim_{N\to\infty} \sqrt{N}(\lambda_i - \lambda(\theta_i)) \stackrel{d}{=} \mathcal{N}(0, V(\theta_i)), \qquad 1 \le i \le q-1, \tag{9.5}$$

$$\lim_{N\to\infty} N^{2/3}(\lambda_i - 2) \stackrel{d}{=} \mathfrak{a}_{i-q+1}, \qquad i \ge q, \tag{9.6}$$

where $\mathcal{N}(0, V(\theta_i))$*,* $0 \le i \le q-1$*, are independent and* $\{\mathfrak{a}_j\}_{j\ge1}$ *are points of the Airy*$_1$ *point process independent from* (9.5)*. In addition,* (8.46) *with* $h(x) = 1$ *holds for the eigenvalues of* $\mathbf{A}$*.*

Proof. The final statement, (8.46) is proven by induction on r, starting from Theorem 8.24 for $r = 0$, and using Lemma 8.28 with Corollary 9.3 for the induction step. Here

$$\mu(x)\mathrm{d}x = \frac{1}{2\pi}\sqrt{4 - x^2}\, \mathbf{1}_{[-2,2]}\, \mathrm{d}x, \qquad m(z) = \frac{z - \sqrt{z^2 - 4}}{2}, \tag{9.7}$$

and the constants of (8.45) are computed as $\mathfrak{s} = 1$, $\mathfrak{m} = 1$, $\lambda_+ = 2$.

Statements close to (9.5), (9.6) are known from Capitaine et al. [2012], Benaych-Georges et al. [2011], Knowles and Yin [2013]; we sketch the proof in order to be self-contained.

We proceed by induction on r; the base case $r = 0$ is Proposition 4.1. We analyze the equation (9.1) with $\theta = \theta_1$ and λ_i being eigenvalues of $\mathbf{A} = \sum_{i=2}^r \theta_i \cdot \mathbf{u}_i^*(\mathbf{u}_i^*)^\mathsf{T} + \mathcal{E}$. We start by considering the case $\theta_1 > \theta_c$ (i.e., $q > 1$) and comment on the changes for the case $q = 1$ at the end. We first look at the interval $a \in [\lambda_1, +\infty)$. The right-hand side of the equation (9.1) is a monotone-decreasing function of a in this interval, changing from $+\infty$ to 0. Hence, there is a unique $\hat{a} \in (\lambda_1, +\infty)$ solving (9.1). In order to locate this $\hat{a}$, we make an asymptotic expansion of the equation. Using Lemma 9.4, we can assume $\mathbf{u}^*$ in (9.1) to be a uniformly random vector on the unit sphere, independent from everything else. Then $\langle \mathbf{u}_i, \mathbf{u}^* \rangle$ are coordinates of a similar vector, because $\mathbf{u}_i$ are orthonormal. Hence, the equation (9.1) is recast as

$$\frac{1}{\theta_1} = \sum_{j=1}^N \frac{\zeta_j^2}{a - \lambda_j}, \qquad (\zeta_1, \dots, \zeta_N) \sim \text{Uniform on unit sphere } \mathbb{S}^{N-1}. \tag{9.8}$$

We further approximate the RHS as $N \to \infty$ for $a > \lambda_j$, by writing it as

$$\frac{1}{N}\sum_{j=1}^N \frac{1}{a - \lambda_j} + \sum_{j=1}^N \frac{\zeta_j^2 - \frac{1}{N}}{a - \lambda_j}. \tag{9.9}$$

Let us first imagine that $(\lambda_1, \dots, \lambda_N)$ are eigenvalues of GOE, i.e., of the matrix $\mathcal{E} = \frac{1}{\sqrt{2N}}(\mathcal{Z} + \mathcal{Z}^\mathsf{T})$ from the statement of the proposition. Then, using the semicircle law (see Theorem 8.24 or Bai and Silverstein [2010, Chapter 2 and Theorem 9.2]), the first term in (9.9) becomes

$$\int_{-2}^{2} \frac{1}{2\pi}\sqrt{4-x^2}\frac{\mathrm{d}x}{a-x} + O\left(\frac{1}{N}\right) = \frac{1}{2}\left(a-\sqrt{a^2-4}\right) + O\left(\frac{1}{N}\right). \tag{9.10}$$

For the second term, note that $\zeta_j^2 - \frac{1}{N}$, $j = 1, 2, \dots, N$, are weakly dependent mean 0 random variables. Hence, conditionally on $\lambda_1, \dots, \lambda_N$, CLT applies and the sum is asymptotically Gaussian. The limit of the variance can be computed using

$$\mathbb{E}\left(\zeta_j^2 - \tfrac{1}{N}\right)^2 = \frac{2}{N^2} + o\left(\frac{1}{N^2}\right), \qquad \mathbb{E}\left(\zeta_i^2 - \tfrac{1}{N}\right)\left(\zeta_j^2 - \tfrac{1}{N}\right) = -\frac{2}{N^3} + o\left(\frac{1}{N^3}\right),$$

where the first identity comes from writing $\zeta_j^2 \stackrel{d}{=} \frac{\xi_j^2}{\sum_{\ell=1}^N \xi_\ell^2}$ with i.i.d. $\mathcal{N}(0,1)$ random variables ξ_j and the second identity comes from combining the first one with $\mathbb{E}\left[\sum_{j=1}^N (\zeta_j^2 - \frac{1}{N})\right]^2 = \mathbb{E}[0]^2 = 0$. Hence,

$$\mathbb{E}\left[\left(\sum_{j=1}^N \frac{\zeta_j^2 - \frac{1}{N}}{a-\lambda_j}\right)^2 \middle| \lambda_1, \dots, \lambda_N\right] = \frac{2}{N^2}\sum_{j=1}^N \frac{1}{(a-\lambda_j)^2} - \frac{2}{N^3}\left[\sum_{j=1}^N \frac{1}{a-\lambda_j}\right]^2 + o\left(\frac{1}{N}\right),$$

and plugging in the semicircle law, we further approximate the variance as

$$\begin{aligned} &\frac{2}{N}\left(\int_{-2}^{2}\frac{1}{2\pi}\sqrt{4-x^2}\frac{\mathrm{d}x}{(z-x)^2} - \left[\int_{-2}^{2}\frac{1}{2\pi}\sqrt{4-x^2}\frac{\mathrm{d}x}{z-x}\right]^2\right) + o\left(\frac{1}{N}\right) \\ &= \frac{1}{N}\left(-1 + \frac{a}{\sqrt{a^2-4}}\right) - \frac{1}{2N}\left(a - \sqrt{a^2-4}\right)^2 + o\left(\frac{1}{N}\right) = \frac{(a-\sqrt{a^2-4})^3}{4N\sqrt{a^2-4}} + o\left(\frac{1}{N}\right). \end{aligned}$$

We conclude that on the interval $(\lambda_1, +\infty)$, the equation (9.8) is approximated by

$$\frac{1}{\theta_1} = \frac{1}{2}\left(a - \sqrt{a^2-4}\right) + \frac{1}{\sqrt{N}}\sqrt{\frac{(a-\sqrt{a^2-4})^3}{4\sqrt{a^2-4}}}\mathcal{N}(0,1) + o\left(\frac{1}{\sqrt{N}}\right). \tag{9.11}$$

Recall that this approximation was obtained assuming λ_i to be eigenvalues of GOE. What we actually need for them is instead to be coming from a deformation of GOE, i.e., to be the eigenvalues of $\sum_{i=2}^r \theta_i \cdot \mathbf{u}_i^*(\mathbf{u}_i^*)^\mathsf{T} + \mathcal{E}$. The approximation (9.11) remains true for such a finite rank deformation, as follows from the interlacements of Corollary 9.3, by repeating the arguments in the proof of Lemma 8.28.

Solving the equation (9.11) as $N \to \infty$, we get

$$a = \theta_1 + \frac{1}{\theta_1} + \mathcal{N}(0,1)\frac{\sqrt{2}}{\sqrt{N}}\sqrt{\frac{\theta_1^2-1}{\theta_1^2}} + o\left(\frac{1}{\sqrt{N}}\right), \tag{9.12}$$

which matches (9.5) and proves the desired asymptotics for the largest eigenvalue.

For the remaining eigenvalues, the idea is to show that the $(i+1)$st largest root of (9.8) is very close to λ_i and then use the induction hypothesis. By Corollary 9.3, the $(i+1)$st largest root is the unique root in the interval $(\lambda_{i+1}, \lambda_i)$ and our task is to show that it is much closer to the right end-point of this segment rather than to the left end-point.

Case 1: $\theta_{i+1} > \theta^c$, so that λ_i is bounded away from λ_+. Let us approximate (9.8) as $N \to \infty$ near λ_i. For the sum over $j \neq i$, the same arguments leading to (9.11) continue to hold and the equation turns into:

$$\frac{1}{\theta_1} - \frac{1}{2}\left(a - \sqrt{a^2-4}\right) - \frac{1}{\sqrt{N}}\sqrt{\frac{(a-\sqrt{a^2-4})^3}{4\sqrt{a^2-4}}}\mathcal{N}(0,1) = \frac{\zeta_i^2}{a-\lambda_i} + o\left(\frac{1}{\sqrt{N}}\right). \tag{9.13}$$

Note that for a close to λ_i, the value of $\frac{1}{2}\left(a - \sqrt{a^2-4}\right)$ is close to $1/\theta_{i+1} > 1/\theta_1$ by induction assumption (9.5). Hence, for large N the left-hand side of (9.13) is negative and bounded away from zero. On the other hand $\zeta_i^2 = O(\frac{1}{N})$. We conclude that a should be smaller than λ_i, at distance $O(\frac{1}{N})$, in order for (9.13) to hold. Hence, this root a satisfies the asymptotics (9.5).

Case 2: $\theta_{i+1} < \theta^c$, so that λ_i is close to λ_+. We approximate the right-hand side of (9.8) as $N \to \infty$ near λ_+ by writing $\zeta_j^2 \stackrel{d}{=} \frac{\xi_j^2}{\sum_{\ell=1}^N \xi_\ell^2}$ with i.i.d. $\mathcal{N}(0,1)$ random variables ξ_j and then using Theorem 8.20 with $\mathfrak{s} = \mathfrak{m} = 1$, $h(x) = 1$, whose assumptions hold by the induction hypothesis. The right-hand side of (9.8) has asymptotics $1 + N^{-1/3} \cdot \mathcal{G}(b) + o(N^{-1/3})$ where the rescaled variable is $b = N^{2/3}(a-2)$. The interval $a \in (\lambda_{i+1}, \lambda_i)$ turns asymptotically into $b \in (\mathfrak{a}_{i-q+2}, \mathfrak{a}_{i-q+1})$, and the equation becomes $\mathcal{G}(b) = N^{1/3}(1/\theta_1 - 1) + o(N^{-1/3})$. Since $1/\theta_1 < 1/\theta^c = 1$, we are looking for a value of $b \in (\mathfrak{a}_{i-q+2}, \mathfrak{a}_{i-q+1})$, where $\mathcal{G}(b)$ would be large and negative. Clearly, then b needs to be close to the right end-point of the segment, i.e., to $\mathfrak{a}_{i-q+1}$ and we achieve (9.6).

To finish the proof it remains to analize the case $\theta_1 < \theta^c$, i.e., $q = 1$. The argument then repeats the just presented Case 2, with the only difference being that we now look for a value of $b \in (\mathfrak{a}_{i-q+2}, \mathfrak{a}_{i-q+1})$, where $\mathcal{G}(b)$ would be large and *positive*. Then b needs to be close to the left end-point of the segment[8], which is $\mathfrak{a}_{i-q+2} = \mathfrak{a}_{i+1}$ and we arrive at (9.6). □

Proof of Theorem 4.6 for the spiked Wigner model of Section 2.1. We note that the constants (4.8) simplify to $\kappa_1 = \kappa_2 = 1$. We analyze the equation (9.1) for $\theta = \theta_q$, $(\lambda_i, \mathbf{u}_i)_{i=1}^N$ being eigenvalues and eigenvectors of

$$\mathbf{A} = \sum_{\substack{1 \le i \le r \\ i \neq q}} \theta_i \cdot \mathbf{u}_i^* (\mathbf{u}_i^*)^\mathsf{T} + \mathcal{E},$$

and $\mathbf{u}^*$ being a uniformly random unit vector independent from the rest. By Proposition 9.1 and Lemma 9.4 the N solutions of this equation denoted $a_1 \ge a_2 \ge \cdots \ge a_N$, are precisely the eigenvalues of Theorem 4.6 (which were λ_i there) and we need to establish (4.6) and (4.7). We rely on Proposition 9.5 for the asymptotics of $\lambda_1, \lambda_2, \dots$ and therefore know that the largest ones satisfy (4.6) (equivalently, (9.5)), while the next ones converge to the points of the Airy$_1$ point process by (9.6).

We first claim that

$$a_i - \lambda_i = O\left(\frac{1}{N}\right), \qquad 1 \le i \le q-1. \tag{9.14}$$

The proof of (9.14) is exactly the same as Case 1 in the proof of Proposition 9.5, i.e., using (9.13) (with θ_1 replaced by θ_q this time). (9.14) combined with (9.5) implies the desired asymptotics (4.6) for $a_1, \dots, a_{q-1}$.

[8]For $i = q = 1$ the segment of interest is $(\mathfrak{a}_1, +\infty)$.

It remains to compute the asymptotics of a_q. Because $\mathbf{u}^*$ is uniformly random and $[\mathbf{u}_i]_{i=1}^N$ are orthonormal, the vector $\langle \mathbf{u}_j, \mathbf{u}^*\rangle^2$, $1 \le j \le N$, has the same distribution as $\frac{\xi_j^2}{\sum_{\ell=1}^N \xi_\ell^2}$ with i.i.d. $\mathcal{N}(0,1)$ random variables ξ_j. Hence, recalling $\theta_q = \theta^c + N^{-1/3}\tilde{\theta} = 1 + N^{-1/3}\tilde{\theta}$, (9.1) becomes:

$$\frac{1}{1+N^{-1/3}\tilde{\theta}} = \frac{1}{\sum_{\ell=1}^N \xi_\ell^2} \sum_{i=q}^{N} \frac{\xi_i^2}{a-\lambda_i} + \frac{1}{\sum_{\ell=1}^N \xi_\ell^2} \sum_{i=1}^{q-1} \frac{\xi_i^2}{a-\lambda_i}. \tag{9.15}$$

By Corollary 9.3, a_q is the unique root of (9.15) in the interval[9] $(\lambda_q, \lambda_{q-1})$. In order to locate this root, we change the variables

$$b = N^{2/3}(a-2), \qquad a = 2 + N^{-2/3} b, \tag{9.16}$$

and investigate the asymptotics of (9.15) for finite b, i.e., for a close to $\lambda_+ = 2$. Because $\lambda_1, \dots, \lambda_{q-1}$ are bounded away from 2 by Proposition 9.5, the second sum in the right-hand side of (9.15) is $O(\frac{1}{N})$ and can be omitted. For the first sum we apply Theorem 8.20 with $\mathfrak{s} = \mathfrak{m} = 1$, $h(x) = 1$. Hence, (9.15) turns into

$$\frac{1}{1+N^{-1/3}\tilde{\theta}} - 1 = N^{-1/3}\mathcal{G}(b) + o\left(N^{-1/3}\right). \tag{9.17}$$

Taylor expanding the left-hand side in small $N^{-1/3}\tilde{\theta}$ and using Corollary 8.22, we conclude that the desired root b converges towards the largest root of the equation $-\tilde{\theta} = \mathcal{G}(b)$. Comparing with Definition 4.2, we are done. □

9.2. **Spiked covariance model.** The proof of Theorem 4.6 for the spiked covariance model of Section 2.2 follows the same outline as the argument in the previous section, but with a different computational part. We explain the new computations, but otherwise try being brief on the technical details which repeat the previous section. We start with an analogue of Proposition 9.1.

Suppose that we are given $N \times S$ matrix $\mathbf{U}$, in which the rows are indexed by $i = 0, 1, \dots, N-1$ and $S \ge N$. We let λ_* denote the squared length of the zeroth row of $\mathbf{U}$ (treated as an S–dimensional vector) and let $\mathbf{v}^*$ be the unit vector in the direction of this row. Let $\widetilde{\mathbf{U}}$ denote the $(N-1) \times S$ matrix formed by rows $i = 1, 2, \dots, N-1$ of $\mathbf{U}$.

We would like to connect the singular values and singular vectors of $\mathbf{U}$ to: singular values and vectors of $\widetilde{\mathbf{U}}$, λ_*, and $\mathbf{v}^*$. We let $(\mathbf{u}_i, \mathbf{v}_i, \sqrt{\lambda_i})$, $1 \le i \le N-1$, be the left singular vector (of $(N-1)\times 1$ dimensions), right singular vector (of $S \times 1$ dimensions), singular value triplets for $\widetilde{\mathbf{U}}$, which means that

$$\widetilde{\mathbf{U}} = \left(\mathbf{u}_1; \mathbf{u}_2; \dots; \mathbf{u}_{N-1}\right) \begin{pmatrix} \sqrt{\lambda_1} & 0 & \dots & \\ 0 & \sqrt{\lambda_2} & 0 & \\ & 0 & \ddots & \\ & & 0 & \sqrt{\lambda_{N-1}} \end{pmatrix} \begin{pmatrix} \mathbf{v}_1^{\mathsf{T}} \\ \mathbf{v}_2^{\mathsf{T}} \\ \vdots \\ \mathbf{v}_{N-1}^{\mathsf{T}} \end{pmatrix} = \sum_{i=1}^{N-1} \sqrt{\lambda_i}\mathbf{u}_i \mathbf{v}_i^{\mathsf{T}}$$

and $\langle \mathbf{u}_i, \mathbf{u}_j\rangle = \delta_{i=j}$, $\langle \mathbf{v}_i, \mathbf{v}_j\rangle = \delta_{i=j}$. We order the singular values so that $\lambda_1 \ge \dots \ge \lambda_{N-1} \ge 0$.

[9] If $q = 1$, then we should set $\lambda_{q-1} = +\infty$.

Proposition 9.6. *Suppose that $a \geq 0$ is an eigenvalue of $\mathbf{U}\mathbf{U}^\mathsf{T}$. Then either*

$$\lambda_* \left(1 + \sum_{i=1}^{N-1} \frac{\lambda_i \langle \mathbf{v}^*, \mathbf{v}_i\rangle^2}{a - \lambda_i}\right) = a, \tag{9.18}$$

or $a = \lambda_j$ for $1 \leq j \leq N$, where $\sqrt{\lambda_j}$ is a singular value of multiplicity one, $\langle \mathbf{v}^*, \mathbf{v}_j \rangle = 0$, and (9.18) *holds with the jth term excluded, or $a = \lambda_j$, where $\sqrt{\lambda_j}$ is a singular value of multiplicity greater than* 1.

We omit the proof, see, e.g., Bykhovskaya and Gorin [2025, Appendix A] (which has λ_i squared and $\mathbf{v}$'s replaced with $\mathbf{u}$'s).

In order to see how Proposition 9.6 is relevant for the setting of Section 2.2, we choose one index $1 \leq k \leq r$ and a deterministic orthogonal matrix O, which maps $\mathbf{u}_k^*$ into the zeroth basis vector and orthogonal complement of $\mathbf{u}_k^*$ to the span of the basis vectors with labels $1, 2, \dots, N-1$, so that

$$O\Omega O^\mathsf{T} = \begin{pmatrix} \theta_k & 0 & \dots & 0 \\ 0 & & & \\ \vdots & & \Omega' & \\ 0 & & & \end{pmatrix},$$

where Ω' is $(N-1)\times(N-1)$ symmetric matrix with eigenvalues $\{\theta_i\}_{i \neq k}$ and 1 of multiplicity $N-r$. Conjugating with O does not change the eigenvalues of the sample covariance matrix $\frac{1}{S}XX^\mathsf{T}$; it also does not change the fact that the columns of X are i.i.d. On the other hand, after transformation by O, we can use Proposition 9.6 for $\mathbf{U} = \Omega X$. We reach the following conclusion:

Corollary 9.7. *For each $1 \leq k \leq r$, the eigenvalues of $\frac{1}{S}XX^\mathsf{T}$ solve an equation in a*

$$\frac{1}{S}\frac{\sum_{i=N}^S \xi_i^2}{a} + \frac{1}{S}\sum_{i=1}^{N-1} \frac{\xi_i^2}{a - \lambda_i} = \frac{1}{\theta_k}, \tag{9.19}$$

where $\lambda_1^2 \geq \dots \geq \lambda_{N-1}^2$ are eigenvalues of $\frac{1}{S}YY^\mathsf{T}$, with Y being $(N-1)\times S$ matrix with i.i.d. Gaussian columns of covariance Ω', and $\xi_1, \dots, \xi_S$ are i.i.d. $\mathcal{N}(0,1)$ independent from Y.

Proof. After we rotate by O defined above, the zeroth row of OX is a vector with i.i.d. $\sqrt{\theta_k}\mathcal{N}(0,1)$ random variables as its components, independent from the remaining rows of X. Hence, the scalar products of the zeroth row with orthonormal vectors $\mathbf{v}_i$ are again i.i.d. $\sqrt{\theta_k}\mathcal{N}(0,1)$ and we can denote them $\sqrt{\theta_k}\xi_i$. The values of $\langle \mathbf{v}^*, \mathbf{v}_i\rangle$ in (9.18) differ from these scalar product by the normalization of $\mathbf{v}^*$, i.e., they are

$$\frac{\theta_k \xi_i^2}{\sum_{l=1}^S \theta_k \xi_l^2} = \frac{\xi_i^2}{\sum_{l=1}^S \xi_l^2}.$$

On the other hand, the value of λ^* in (9.18) is the squared length of the zeroth row, $\|\mathbf{v}^*\|^2 = \sum_{l=1}^S \theta_k \xi_l^2$. Finally, Proposition 9.6 deals with $\mathbf{U}\mathbf{U}^\mathsf{T}$, while in the corollary the matrix XX^T is divided by S and, therefore, we should rescale by S both λ_i and a. Hence, dividing (9.18) by a and θ_k, we get

$$\frac{1}{a}\left(\frac{\sum_{i=1}^S \xi_i^2}{S} + \frac{1}{S}\sum_{i=1}^{N-1}\frac{\lambda_i \xi_i^2}{a - \lambda_i}\right) = \frac{1}{\theta_k}.$$

Rearranging the terms, we arrive at (9.19). ∎

Corollary 9.8. *In Corollary 9.7, the eigenvalues* $a_1 \geq \cdots \geq a_N$ *of* $\frac{1}{S}XX^\mathsf{T}$ *interlace with those of* $\frac{1}{S}YY^\mathsf{T}$*:* $a_1 \geq \lambda_1 \geq a_2 \geq \cdots \geq \lambda_{N-1} \geq a_N$*.*

Proof. After multiplying by denominators, (9.19) is a degree N polynomial equation and we locate its roots by keeping track of the signs of the difference between left-hand and right-hand sides on intervals $(\lambda_1, +\infty)$, (λ_2, λ_1), ..., $(\lambda_{N-1}, \lambda_N)$. □

Proposition 9.9. *Consider the spiked covariance model* $\Omega = \sigma^2 I_N + (\theta_i - \sigma^2) \cdot \mathbf{u}_i^*(\mathbf{u}_i^*)^\mathsf{T}$ *of Section 2.2 with* $\sigma^2 = 1$*, and* $\theta_1 > \theta_2 > \cdots > \theta_r \geq 0$ *split into two groups:* $\theta_1, \ldots, \theta_{q-1} > \theta^c = 1 + \gamma$ *and* $\theta_q, \ldots, \theta_r < \theta^c = 1 + \gamma$*. Then, with* $\frac{N}{S} = \gamma^2 + O\left(\frac{1}{N}\right)$*, in the sense of convergence in joint distribution and using* (2.8)*:*

$$\lim_{N\to\infty} \sqrt{N}(\lambda_i - \lambda(\theta_i)) \stackrel{d}{=} \mathcal{N}(0, V(\theta_i)), \qquad 1 \leq i \leq q-1, \tag{9.20}$$

$$\lim_{N\to\infty} N^{2/3} \frac{\lambda_i - (1+\gamma)^2}{\gamma(1+\gamma)^{4/3}} \stackrel{d}{=} \mathfrak{a}_{i-q+1}, \qquad i \geq q, \tag{9.21}$$

where $\mathcal{N}(0, V(\theta_i))$*,* $0 \leq i \leq q-1$*, are independent and* $\{\mathfrak{a}_j\}_{j\geq 1}$ *are points of the Airy*$_1$ *point process independent from* (9.20)*. In addition,* (8.46) *with* $h(x) = 1$ *holds for the eigenvalues of the sample covariance matrix* $\frac{1}{S}XX^\mathsf{T}$*.*

Proof. The final statement, (8.46) with $h(x) = 1$ follows by induction on r from Theorem 8.24 for $r = 0$, and using Lemma 8.28 with Corollary 9.8 for the induction step. Here $\lambda_\pm = (1 \pm \gamma)^2$,

$$\mu(x)\mathrm{d}x = \frac{1}{2\pi} \frac{\sqrt{(\lambda_+ - x)(x - \lambda_-)}}{\gamma^2 x} \mathbf{1}_{[\lambda_-, \lambda_+]}\, \mathrm{d}x, \qquad m(z) = \frac{z + \gamma^2 - 1 - \sqrt{(z-\lambda_+)(z-\lambda_-)}}{2\gamma^2 z}, \tag{9.22}$$

which are the Marchenko-Pastur law and its Stieljes transform, respectively, and the constants of (8.45) are computed to be $\mathfrak{s} = \frac{\sqrt{\lambda_+ - \lambda_-}}{2\gamma^2\lambda_+} = \frac{1}{\gamma^{3/2}(1+\gamma)^2}$, $\mathfrak{m} = \frac{(1+\gamma)^2 + \gamma^2 - 1}{2\gamma^2(1+\gamma)^2} = \frac{1}{\gamma(1+\gamma)}$,

Statements close to (9.20), (9.21) are known from Paul [2007], Bai and Yao [2008], Bloemendal et al. [2016]. Alternatively, the proof can be obtained by the same argument as in Proposition 9.5, by induction on r with the base case $r = 0$ given in Johnstone [2001], Soshnikov [2002] and the step based on Corollary 9.7. We only highlight the key computation, which is an analogue of (9.11) and (9.12).

We rewrite (9.19) as

$$\frac{1}{S} \cdot \frac{S - N + 1}{a} + \frac{1}{S} \sum_{i=1}^{N-1} \frac{1}{a - \lambda_i} + \frac{1}{S} \frac{\sum_{i=N}^{S}(\xi_i^2 - 1)}{a} + \frac{1}{S} \sum_{i=1}^{N-1} \frac{\xi_i^2 - 1}{a - \lambda_i} = \frac{1}{\theta_k}, \tag{9.23}$$

and analyze its asymptotics, assuming that $\lambda_1, \ldots, \lambda_N$ are eigenvalues of $\frac{1}{S}XX^\mathsf{T}$, where X is $(N-1) \times S$ matrix of i.i.d. $\mathcal{N}(0,1)$. Then, using the Marchenko-Pastur law (see Theorem 8.24 or Bai and Silverstein [2010, Chapter 3 and Theorem 9.10]), formulas for the limit (9.22), and CLT, the first two terms in (9.23) become deterministic as $S \to \infty$, while the third and fourth terms become independent Gaussians. Hence, (9.23) becomes

$$\frac{1 - \gamma^2}{a} + \gamma^2 m(a) + \frac{\sqrt{2}}{\sqrt{N}} \mathcal{N}(0,1) \sqrt{\frac{\gamma^2(1-\gamma^2)}{a^2} - \gamma^4 m'(a)} + o\left(\frac{1}{\sqrt{N}}\right) = \frac{1}{\theta_k}. \tag{9.24}$$

Plugging the formula for $m(z)$ from (9.22), we get

$$\frac{a+1-\gamma^2-\sqrt{(a-\lambda_+)(a-\lambda_-)}}{2a} + \frac{\gamma}{a\sqrt{N}}\mathcal{N}(0,1)\sqrt{1-\gamma^2-\frac{(1-\gamma^2)^2-a(1+\gamma^2)}{\sqrt{(a-\lambda_+)(a-\lambda_-)}}}+o\left(\frac{1}{\sqrt{N}}\right)=\frac{1}{\theta_k}. \tag{9.25}$$

Treating the last identity as an equation on a, we solve it asymptotically as $N\to\infty$, getting:

$$a=\theta_k\left(1+\frac{\gamma^2}{\theta_k-1}\right)+\mathcal{N}(0,1)\frac{\gamma\sqrt{2}}{\sqrt{N}}\theta_k\sqrt{1-\frac{\gamma^2}{(\theta_k-1)^2}}+o\left(\frac{1}{\sqrt{N}}\right), \tag{9.26}$$

which matches (2.8). □

Proof of Theorem 4.6 for the spiked covariance model of Section 2.2. The constants (4.8) simplify to $\kappa_1=\gamma(1+\gamma)^{4/3}$, $\kappa_2=\frac{1}{\gamma(1+\gamma)^{2/3}}$. Note that κ_1 matches the denominator in (9.21), as it should; it also equals $\mathfrak{s}^{-2/3}$, where $\mathfrak{s}$ is the constant in (8.45), as computed after (9.22). Hence, Assumption 8.19 will be satisfied.

We analyze the equation (9.19) for $k=q$. In this situation the asymptotics of $(\lambda_i)_{i=1}^N$ is given to us by Proposition 9.9. Arguing as in the previous section, the $q-1$ largest roots of the equation are close to $\lambda_1,\dots,\lambda_{q-1}$, resulting in (4.6). In order to establish (4.7), we need to approximate (9.19) for a close to $\lambda_+=(1+\gamma)^2$ and locate the root of the equation in the $(\lambda_q,\lambda_{q-1})$ interval. We change the variables

$$b=N^{2/3}\frac{a-\lambda_+}{\kappa_1}=N^{2/3}\frac{a-(1+\gamma)^2}{\gamma(1+\gamma)^{4/3}},\qquad a=(1+\gamma)^2+N^{-2/3}\gamma(1+\gamma)^{4/3}b, \tag{9.27}$$

and apply Theorem 8.20 with $h(x)=1$, converting (9.19) into (recall that $\theta^c=1+\gamma$ and $\mathfrak{m}=\frac{1}{\gamma(1+\gamma)}$, as computed after (9.22)):

$$\frac{S+1-N}{S}\cdot\frac{1}{(1+\gamma)^2}+\frac{N}{S}\left[\frac{1}{\gamma(1+\gamma)}+N^{-1/3}\frac{1}{\gamma(1+\gamma)^{4/3}}\mathcal{G}(b)\right]=\frac{1}{1+\gamma+N^{-1/3}\tilde\theta}+o\left(N^{-1/3}\right).$$

Recalling that $\frac{N}{S}=\gamma^2+O\left(\frac{1}{N}\right)$, we convert the last equation into

$$\frac{1-\gamma}{1+\gamma}+\frac{\gamma}{(1+\gamma)}+N^{-1/3}\frac{\gamma}{(1+\gamma)^{4/3}}\mathcal{G}(b)=\frac{1}{1+\gamma}\left(1-N^{-1/3}\frac{\tilde\theta}{1+\gamma}\right)+o\left(N^{-1/3}\right).$$

The finite order term cancel out and we finally get after multiplying by $N^{1/3}$ the equation

$$\frac{\gamma}{(1+\gamma)^{4/3}}\mathcal{G}(b)=-\tilde\theta\frac{1}{(1+\gamma)^2}+o(1)\quad\Longleftrightarrow\quad\mathcal{G}(b)=-\tilde\theta\frac{1}{\gamma(1+\gamma)^{2/3}}+o(1).$$

Recognizing the constant κ_2 in the right-hand side of the last formula and comparing with Definition 4.2, we arrive at (4.7). □

9.3. **Factor models.** The proof of Theorem 4.6 for the factor model of Section 2.3 follows the same outline as in the previous two sections. However, an analogue of Propositions 9.1 and 9.6 becomes more complicated in this case.

Proposition 9.10. *For $N \le S$, let X be an $N \times S$ matrix of the form*

$$X = \sqrt{\theta S} \cdot \mathbf{u}^* (\mathbf{v}^*)^{\mathsf{T}} + Y, \qquad \textit{with} \qquad Y = \sum_{i=1}^{N} \sqrt{\lambda_i S} \cdot \mathbf{u}_i (\mathbf{v}_i)^{\mathsf{T}}, \tag{9.28}$$

where $\mathbf{u}^*$ and $\mathbf{u}_i$ are N–dimensional unit vectors; $\mathbf{u}_i$, $1 \le i \le N$, are pairwise orthogonal; $\mathbf{v}^*$ and $\mathbf{v}_i$ are S–dimensional unit vectors; $\mathbf{v}_i$, $1 \le i \le N$, are pairwise orthogonal. If a is a non-zero squared singular value of $X/\sqrt{S}$, i.e., an eigenvalue of $\frac{1}{S} X X^{\mathsf{T}}$, then either it solves an equation:

$$\left(1 - \sqrt{\theta} \sum_{i=1}^{N} \frac{\sqrt{\lambda_i} \langle \mathbf{u}^*, \mathbf{u}_i \rangle \langle \mathbf{v}^*, \mathbf{v}_i \rangle}{a - \lambda_i}\right)^2 = a\theta \left(\sum_{i=1}^{N} \frac{\langle \mathbf{u}^*, \mathbf{u}_i \rangle^2}{a - \lambda_i}\right) \left(\sum_{j=1}^{S} \frac{\langle \mathbf{v}^*, \mathbf{v}_j \rangle^2}{a - \lambda_j}\right), \tag{9.29}$$

where $\mathbf{v}_{N+1}, \dots \mathbf{v}_S$ are arbitrary vectors complimenting $\mathbf{v}_1, \dots, \mathbf{v}_N$ to an orthonormal basis, $\langle \cdot, \cdot \rangle$ is the scalar product, and we set $\lambda_j = 0$ for $N < j \le S$. Or $a = \lambda_i$ for $1 \le i \le N$, where λ_i has multiplicity one, $\langle \mathbf{u}^*, \mathbf{u}_i \rangle = \langle \mathbf{v}^*, \mathbf{v}_i \rangle = 0$, and the equation (9.29) *holds with the i-th terms excluded; or $a = \lambda_i$, where λ_i has multiplicity larger than one.*

Remark 9.11. *A somewhat similar statement can be found in Benaych-Georges and Nadakuditi [2012, Lemma 4.1]. Also if for some i we have $\lambda_i = 0$, $\mathbf{u}^* = \mathbf{u}_i$, and $\langle \mathbf{u}^*, \mathbf{u}_{i'} \rangle = 0$ for all $i \ne i'$, then* (9.29) *turns into* (9.18) *of the previous subsection.*

Proof of Proposition 9.10. We only describe the situation when all λ_i are distinct and for each index i either $\langle \mathbf{u}^*, \mathbf{u}_i \rangle \ne 0$ or $\langle \mathbf{v}^*, \mathbf{v}_i \rangle \ne 0$. Other cases can be obtained by continuously deforming the parameters.

Let a be a squared singular value of $X/\sqrt{S}$ corresponding to vectors $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$. Then $a = pq/S$, where p and q solve:

$$\begin{cases} X\hat{\mathbf{v}} = p\hat{\mathbf{u}}, \\ X^{\mathsf{T}} \hat{\mathbf{u}} = q\hat{\mathbf{v}}. \end{cases} \qquad \Longleftrightarrow \qquad \begin{cases} \langle X\hat{\mathbf{v}}, \mathbf{u}_i \rangle = p \langle \hat{\mathbf{u}}, \mathbf{u}_i \rangle, & 1 \le i \le N, \\ \langle X^{\mathsf{T}} \hat{\mathbf{u}}, \mathbf{v}_j \rangle = q \langle \hat{\mathbf{v}}, \hat{\mathbf{v}}_j \rangle, & 1 \le j \le S. \end{cases}$$

Denoting $\alpha_i = \langle \hat{\mathbf{u}}, \mathbf{u}_i \rangle$ and $\beta_j = \langle \hat{\mathbf{v}}, \hat{\mathbf{v}}_j \rangle$, we rewrite these $N + S$ equations as:

$$\begin{cases} \langle X \sum_{j'=1}^{S} \beta_{j'} \mathbf{v}_{j'}, \mathbf{u}_i \rangle = \sqrt{\theta S} \sum_{j'=1}^{S} \beta_{j'} \langle \mathbf{v}^*, \mathbf{v}_{j'} \rangle \langle \mathbf{u}^*, \mathbf{u}_i \rangle + \sqrt{\lambda_i S} \beta_i = p\alpha_i, & 1 \le i \le N, \\ \langle X^{\mathsf{T}} \sum_{i'=1}^{N} \alpha_{i'} \mathbf{u}_{i'}, \mathbf{v}_j \rangle = \sqrt{\theta S} \sum_{i'=1}^{N} \alpha_{i'} \langle \mathbf{u}^*, \mathbf{u}_{i'} \rangle \langle \mathbf{v}^*, \mathbf{v}_j \rangle + \sqrt{\lambda_j S} \alpha_j = q\beta_j, & 1 \le j \le N, \\ \langle X^{\mathsf{T}} \sum_{i'=1}^{N} \alpha_{i'} \mathbf{u}_{i'}, \mathbf{v}_j \rangle = \sqrt{\theta S} \sum_{i'=1}^{N} \alpha_{i'} \langle \mathbf{u}^*, \mathbf{u}_{i'} \rangle \langle \mathbf{v}^*, \mathbf{v}_j \rangle = q\beta_j, & N+1 \le j \le S. \end{cases}$$

Combining the equations corresponding to the same $i = j$ and solving as two linear equations in two variables α_i, β_i, we get

$$\begin{cases} \alpha_i = \sqrt{\theta S} \frac{q \langle \mathbf{u}^*, \mathbf{u}_i \rangle \tilde{\beta} + \sqrt{\lambda_i S} \langle \mathbf{v}^*, \mathbf{v}_i \rangle \tilde{\alpha}}{pq - \lambda_i S}, & 1 \le i \le N, \\ \beta_i = \sqrt{\theta S} \frac{\sqrt{\lambda_i S} \langle \mathbf{u}^*, \mathbf{u}_i \rangle \tilde{\beta} + p \langle \mathbf{v}^*, \mathbf{v}_i \rangle \tilde{\alpha}}{pq - \lambda_i S}, & 1 \le i \le S, \end{cases} \tag{9.30}$$

where in the last formula for $N < i \le S$ we should use $\lambda_i = 0$ and

$$\tilde{\alpha} = \sum_{i'=1}^{N} \alpha_{i'} \langle \mathbf{u}^*, \mathbf{u}_{i'} \rangle, \qquad \tilde{\beta} = \sum_{j'=1}^{S} \beta_{j'} \langle \mathbf{v}^*, \mathbf{v}_{j'} \rangle.$$

We should further identity the values of $\tilde{\alpha}$ and $\tilde{\beta}$, for which we plug (9.30) back into their definitions, getting a system of two equations:

$$\begin{cases} \tilde{\alpha} = \tilde{\alpha}\sqrt{\theta S}\sum_{i=1}^{N} \frac{\sqrt{\lambda_i S}\langle \mathbf{u}^*, \mathbf{u}_i\rangle\langle \mathbf{v}^*, \mathbf{v}_i\rangle}{pq-\lambda_i S} + \tilde{\beta}\sqrt{\theta S}\sum_{i=1}^{N} \frac{q\langle \mathbf{u}^*, \mathbf{u}_i\rangle^2}{pq-\lambda_i S}, \\ \tilde{\beta} = \tilde{\alpha}\sqrt{\theta S}\sum_{j=1}^{S} \frac{p\langle \mathbf{v}^*, \mathbf{v}_j\rangle^2}{pq-\lambda_j S} + \tilde{\beta}\sqrt{\theta S}\sum_{j=1}^{S} \frac{\sqrt{\lambda_j S}\langle \mathbf{u}^*, \mathbf{u}_j\rangle\langle \mathbf{v}^*, \mathbf{v}_j\rangle}{pq-\lambda_j S}. \end{cases}$$

The system of two homogeneous linear equations has a non-zero solution if and only if the determinant of the matrix coefficients is zero, which is precisely the condition (9.29). □

An analogue of Corollaries 9.3 and 9.8 becomes more delicate for the factor model, which, however, does not lead to any significant changes in the ways we use it in our arguments.

Corollary 9.12. *In Proposition 9.10, let $a_1 \geq \cdots \geq a_N$ be the eigenvalues of $\frac{1}{S}XX^\mathsf{T}$ and let $\lambda_1 \geq \cdots \geq \lambda_N$ be the eigenvalues of $\frac{1}{S}YY^\mathsf{T}$. Then there exists another set of N eigenvalues, $\mu_1 \geq \cdots \geq \mu_N$, such that*

$$a_1 \geq \mu_1 \geq a_2 \geq \cdots \geq a_N \geq \mu_N, \qquad \text{and} \qquad \lambda_1 \geq \mu_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq \mu_N. \tag{9.31}$$

Proof. Using (9.28), we write

$$XX^\mathsf{T} = \theta S \cdot \mathbf{u}^*(\mathbf{u}^*)^\mathsf{T} + \sqrt{\theta S}\cdot\left[\mathbf{u}^*(\mathbf{v}^*)^\mathsf{T} Y^* + Y\mathbf{v}^*(\mathbf{u}^*)^\mathsf{T}\right] + YY^*,$$

which implies that $\frac{1}{S}XX^\mathsf{T}$ is a sum of $\frac{1}{S}YY^\mathsf{T}$ and a rank two symmetric matrix. In the (non-orthogonal) basis $(\mathbf{u}^*, Y\mathbf{v}^*)$, this matrix has the form

$$\begin{pmatrix} \theta S + c\sqrt{\theta S} & c\theta S + d\sqrt{\theta S} \\ \sqrt{\theta S} & c\sqrt{\theta S} \end{pmatrix}, \qquad c = \langle \mathbf{u}^*, Y\mathbf{v}^*\rangle, \quad d = \langle Y\mathbf{v}^*, Y\mathbf{v}^*\rangle.$$

The determinant of this matrix is $(c^2 - d)\theta S < 0$, because $\mathbf{u}^*$ is a unit vector. Hence, the matrix has one positive and one negative eigenvalues. It remains to use Corollary 9.3 twice. □

Next we state an analogue of Lemma 9.4.

Lemma 9.13. *One can replace the vectors $\mathbf{u}_i^*$ and $\mathbf{v}_i^*$ in the factor model $X = \sum_{i=1}^r \sqrt{\theta_i}\sqrt{S}\cdot \mathbf{u}_i^*(\mathbf{v}_i^*)^\mathsf{T} + \mathcal{E}$ of (2.10) with $2r$ vectors uniformly distributed on the N–dimensional and S–dimensional unit spheres, respectively, independent from each other and from $\mathcal{E}$: Theorem 4.6 for deterministic orthonormal vectors as in Section 2.3 is equivalent to the same theorem for independent unit vectors.*

Proof. We first note that the asymptotics in Theorem 4.6 does not depend on the choice of the unit vectors $\{\mathbf{u}_i^*\}_{i=1}^r$ and $\{\mathbf{v}_i^*\}_{i=1}^r$ in (2.10) as long as they form two orthonormal systems. Indeed, any r orthonormal vectors can be obtained from any other r orthonormal vectors by an orthogonal transformation of the space. Such transformation does not change the eigenvalues of $\frac{1}{S}XX^\mathsf{T}$, and also does not change the probability distribution of the matrix $\mathcal{E}$ (which uses Gaussianity of its matrix elements).

Now let $\tilde{\mathbf{u}}_1^*, \dots, \tilde{\mathbf{u}}_r^*$ be r independent vectors uniformly distributed on the N–dimensional unit sphere and let $\tilde{\mathbf{v}}_1^*, \dots, \tilde{\mathbf{v}}_r^*$ be r independent vectors uniformly distributed on the S–dimensional unit sphere. We consider the matrix $M_r = \sum_{i=1}^r \sqrt{\theta_i}\sqrt{S}\tilde{\mathbf{u}}_i^*(\tilde{\mathbf{v}}_i^*)^\mathsf{T}$ and would like to decompose it as $M_r = \sum_{i=1}^r \sqrt{\theta_i'}\sqrt{S}\mathbf{u}_i^*(\mathbf{v}_i^*)^\mathsf{T}$ with orthonormal vectors $\mathbf{u}_i^*$ and $\mathbf{v}_i^*$; $\tilde{\mathbf{u}}_i^*$ and

$\tilde{\mathbf{v}}_i^*$ were not orthonormal. Clearly, $\sqrt{\theta_i'}$ are non-zero singular values of M_r and $\mathbf{u}_i^*$, $\mathbf{v}_i^*$ are corresponding left and right singular vectors. We claim that

$$\theta_i = \theta_i' + O\left(\frac{1}{N} + \frac{1}{S}\right), \qquad N, S \to \infty. \tag{9.32}$$

Note that (9.32) implies the statement of Lemma 9.13, because addition of $O(1/N)$ does not change any of the asymptotics statements of Theorem 4.6. Hence, it remains to prove (9.32). This can be done by induction on r. Using (9.29) with $Y = M_{r-1}$, the non-zero eigenvalues of M_r solve an equation

$$\begin{aligned} &\left(1 - \sqrt{\theta_r} \sum_{i=1}^{r-1} \frac{\sqrt{\theta_i'} \langle \tilde{\mathbf{u}}_r^*, \mathbf{u}_i \rangle \langle \tilde{\mathbf{v}}_r^*, \mathbf{v}_i \rangle}{a - \theta_i'}\right)^2 \\ &\qquad = a\theta_r \left(\sum_{i=1}^{r-1} \frac{\langle \tilde{\mathbf{u}}_r^*, \mathbf{u}_i \rangle^2}{a - \theta_i'} + \sum_{i=r}^{N} \frac{\langle \tilde{\mathbf{u}}_r^*, \mathbf{u}_i \rangle^2}{a}\right) \left(\sum_{j=1}^{r-1} \frac{\langle \tilde{\mathbf{v}}_r^*, \mathbf{v}_j \rangle^2}{a - \theta_j'} + \sum_{j=r}^{S} \frac{\langle \tilde{\mathbf{v}}_r^*, \mathbf{v}_j \rangle^2}{a}\right), \end{aligned} \tag{9.33}$$

where $(\mathbf{u}_i, \mathbf{v}_i, \theta_i')$ are singular vectors and values of M_{r-1} for $1 \le i \le r-1$, and $\mathbf{u}_i$, $\mathbf{v}_i$ with $i \ge r$ complement those to othonormal bases of $N-$ and $S-$dimensional spaces, respectively. We note that after multiplying by the denominators a, $a - \theta_1'$, $\dots$, $a - \theta_{r-1}'$, (9.33) becomes a polynomial equation of degree r (there is a cancelation between the left and right hand sides which guarantees that denominators $(a - \theta_i')^2$ do not appear) and therefore it has r roots. Let us locate these roots.

Representing the unit vectors $\tilde{\mathbf{u}}_r^*$ and $\tilde{\mathbf{v}}_r^*$ as vectors with i.i.d. $\mathcal{N}(0,1)$ components divided by their lengths, using the Law of Large Numbers and the induction hypothesis, the equation is rewritten as $N, S \to \infty$ as

$$\begin{aligned} &\left(1 - \sqrt{\frac{\theta_r}{NS}} \sum_{i=1}^{r-1} \frac{\sqrt{\theta_i + O\left(\frac{1}{N} + \frac{1}{S}\right)} \chi_i \xi_i \left(1 + O\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{S}}\right)\right)}{a - \theta_i + O\left(\frac{1}{N} + \frac{1}{S}\right)}\right)^2 = a\theta_r \\ &\times \left(\frac{1}{N} \sum_{i=1}^{r-1} \frac{\chi_i^2 \left(1 + O\left(\frac{1}{\sqrt{N}}\right)\right)}{a - \theta_i + O\left(\frac{1}{N} + \frac{1}{S}\right)} + \frac{1 + O\left(\frac{1}{N}\right)}{a}\right) \left(\frac{1}{S} \sum_{j=1}^{r-1} \frac{\xi_i^2 \left(1 + O\left(\frac{1}{\sqrt{S}}\right)\right)}{a - \theta_j + O\left(\frac{1}{N} + \frac{1}{S}\right)} + \frac{1 + O\left(\frac{1}{S}\right)}{a}\right), \end{aligned} \tag{9.34}$$

where χ_i and ξ_i are i.i.d. $\mathcal{N}(0,1)$ random variables. (9.34) clearly has r roots of the form $\theta_i + O(\frac{1}{N} + \frac{1}{S})$, $1 \le i \le r$, thus proving (9.32). $\square$

Next, we prove an analogue of Propositions 9.1 and 9.9.

Proposition 9.14. *Consider the factor model $X = \sum_{i=1}^r \sqrt{\theta_i S} \cdot \mathbf{u}_i^* (\mathbf{v}_i^*)^\mathsf{T} + \mathcal{E}$ of Section 2.2 with $\sigma^2 = 1$, and $\theta_1 > \theta_2 > \cdots > \theta_r \ge 0$ split into two groups: $\theta_1, \dots, \theta_{q-1} > \theta^c = \gamma$ and $\theta_q, \dots, \theta_r < \theta^c = \gamma$. Then, with $\frac{N}{S} = \gamma^2 + O\left(\frac{1}{N}\right)$, in the sense of convergence in joint distribution and using* (2.12)*:*

$$\lim_{N \to \infty} \sqrt{N}(\lambda_i - \lambda(\theta_i)) \stackrel{d}{=} \mathcal{N}(0, V(\theta_i)), \qquad 1 \le i \le q-1, \tag{9.35}$$

$$\lim_{N \to \infty} N^{2/3} \frac{\lambda_i - (1+\gamma)^2}{\gamma(1+\gamma)^{4/3}} \stackrel{d}{=} \mathfrak{a}_{i-q+1}, \qquad i \ge q, \tag{9.36}$$

where $\mathcal{N}(0, V(\theta_i))$, $0 \le i \le q-1$, are independent and $\{\mathfrak{a}_j\}_{j\ge 1}$ are points of the Airy$_1$ point process independent from (9.35). In addition, (8.46) holds for the eigenvalues of $\frac{1}{S}XX^\mathsf{T}$ both with $h(x) = 1$ and with $h(x) = \sqrt{x}$.

Proof. The final statement, (8.46) with $h(x) = 1$ and $h(x) = \sqrt{x}$ follows by induction on r from Theorem 8.24 and Corollary 8.25 for $r = 0$, and using Lemma 8.28 with Corollary 9.12 for the induction step. Here $\lambda_\pm = (1 \pm \gamma)^2$, and for $h(x) = 1$

(9.37)
$$\mu(x)\mathrm{d}x = \frac{1}{2\pi}\frac{\sqrt{(\lambda_+ - x)(x-\lambda_-)}}{\gamma^2 x}\,\mathbf{1}_{[\lambda_-,\lambda_+]}\,\mathrm{d}x, \qquad m(z) = \frac{z+\gamma^2-1-\sqrt{(z-\lambda_+)(z-\lambda_-)}}{2\gamma^2 z},$$

which are the same as in (9.22), and the constants of (8.45) are computed to be $\mathfrak{s} = \frac{\sqrt{\lambda_+-\lambda_-}}{2\gamma^2\lambda_+} = \frac{1}{\gamma^{3/2}(1+\gamma)^2}$, $\mathfrak{m} = \frac{(1+\gamma)^2+\gamma^2-1}{2\gamma^2(1+\gamma)^2} = \frac{1}{\gamma(1+\gamma)}$. For the choice $h(x) = \sqrt{x}$, the corresponding function $m(z)$ can be also computed, but we do not need this expression.

Statements similar to (9.35) are known from Onatski [2012] and Benaych-Georges and Nadakuditi [2012]; (9.36) is harder to locate in the literature (although it is also probably known to the specialists). Alternatively, the proof can be obtained by the same argument as in Proposition 9.5, by induction on r with the base case $r = 0$ given in Johnstone [2001], Soshnikov [2002] and the step based on Proposition 9.10. We only highlight the key computation, which is an analogue of (9.11) and (9.12).

In view of Lemma 9.13, and representing uniformly random unit vectors as Gaussian vectors with i.i.d. components divided by their norms, we rewrite (9.29) as

(9.38)
$$\left(1 - \sqrt{\frac{\theta}{\sum_{i=1}^N \xi_i^2 \sum_{j=1}^S \eta_j^2}}\sum_{i=1}^N \frac{\sqrt{\lambda_i}\xi_i\eta_i}{a-\lambda_i}\right)^2 = \frac{a\theta}{\sum_{i=1}^N \xi_i^2 \sum_{j=1}^S \eta_j^2}\left(\sum_{i=1}^N \frac{\xi_i^2}{a-\lambda_i}\right)\left(\sum_{j=1}^S \frac{\eta_j^2}{a-\lambda_j}\right),$$

where ξ_i and η_j are i.i.d. $\mathcal{N}(0,1)$. We analyze the asymptotics of (9.38) assuming that $\lambda_1, \dots, \lambda_N$ are eigenvalues of $\frac{1}{S}\mathcal{E}\mathcal{E}^\mathsf{T}$, where $\mathcal{E}$ is $N \times S$ matrix of i.i.d. $\mathcal{N}(0,1)$. We multiply (9.38) by $\frac{\sum_{i=1}^N \xi_i^2 \sum_{j=1}^S \eta_j^2}{NS}$ and then rewrite separating the terms of different orders as $N \to \infty$:

$$\text{(9.39)}\quad \frac{\sum_{i=1}^N 1 \sum_{j=1}^S 1}{NS} + \sum_{i=1}^N \frac{\xi_i^2-1}{N} + \sum_{j=1}^S \frac{\eta_j^2-1}{S} - 2\sqrt{\frac{\theta}{NS}}\sum_{i=1}^N \frac{\sqrt{\lambda_i}\xi_i\eta_i}{a-\lambda_i} + o\left(\frac{1}{\sqrt{N}}\right)$$
$$= \frac{a\theta}{NS}\left[\sum_{i=1}^N \frac{1}{a-\lambda_i}\sum_{j=1}^S \frac{1}{a-\lambda_j} + \sum_{i=1}^N \frac{1}{a-\lambda_i}\sum_{j=1}^S \frac{\eta_j^2-1}{a-\lambda_j} + \sum_{j=1}^S \frac{1}{a-\lambda_j}\sum_{i=1}^N \frac{\xi_i^2-1}{a-\lambda_i}\right].$$

Using the Marchenko-Pastur law (see Theorem 8.24 or Bai and Silverstein [2010, Chapter 3 and Theorem 9.10]), formulas for the limit (9.22), and $\frac{N}{S} = \gamma^2 + O\left(\frac{1}{N}\right)$ the equation (9.39) becomes:

$$\text{(9.40)}\quad 1 - a\theta\, m(a)\tilde{m}(a) + \sum_{i=1}^N \frac{\xi_i^2-1}{N}\left[1 - \frac{a\theta}{a-\lambda_i}\frac{1}{S}\sum_{j=1}^S \frac{1}{a-\lambda_j}\right]$$
$$+ \sum_{j=1}^S \frac{\eta_j^2-1}{S}\left[1 - \frac{a\theta}{a-\lambda_j}\frac{1}{N}\sum_{i=1}^N \frac{1}{a-\lambda_i}\right] - 2\sqrt{\frac{\theta}{NS}}\sum_{i=1}^N \frac{\sqrt{\lambda_i}\xi_i\eta_i}{a-\lambda_i} = o\left(\frac{1}{\sqrt{N}}\right),$$

where we used the notation

$$\tilde{m}(a) = \left[\gamma^2 m(a) + \frac{1-\gamma^2}{a}\right].$$

Since ξ_i^2-1, η_j^2-1, $\xi_i\eta_i$ are three mean 0 i.i.d. in i sequences, we are in a position to apply the Central Limit Theorem; note that all pairwise covariances between these random variables vanish (using $\mathbb{E}(\xi_i^2-1)\xi_i\eta_i = 0$), and therefore different sums give rise to independent Gaussian limits. Using $\mathbb{E}(\xi_i^2-1)^2 = \mathbb{E}(\xi_j^2-1)^2 = 2$, $\mathbb{E}\xi_i^2\eta_i^2 = 1$ and applying CLT, we further transform (9.40) into

$$\begin{aligned}(9.41)\quad 1 - a\theta\, m(a)\tilde{m}(a) + \frac{\mathcal{N}(0,1)}{\sqrt{N}}\Bigg[\frac{2}{N}\sum_{i=1}^{N}\left(1 - \frac{a\theta}{a-\lambda_i}\frac{1}{S}\sum_{j=1}^{S}\frac{1}{a-\lambda_j}\right)^2 + \\ + \frac{2\gamma^2}{S}\sum_{j=1}^{S}\left(1 - \frac{a\theta}{a-\lambda_j}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{a-\lambda_i}\right)^2 + 4\frac{\theta}{S}\sum_{i=1}^{N}\frac{\lambda_i}{(a-\lambda_i)^2}\Bigg]^{1/2} = o\left(\frac{1}{\sqrt{N}}\right).\end{aligned}$$

Applying the Marchenko-Pastur law, we finally get

$$\begin{aligned}(9.42)\quad 1 - a\theta\, m(a)\tilde{m}(a) + \frac{\mathcal{N}(0,1)}{\sqrt{N}}\Big[[2 - 4a\theta\, m(a)\tilde{m}(a) - 2a^2\theta^2\, m'(a)\tilde{m}^2(a)] \\ + [2\gamma^2 - 4\gamma^2 a\theta\, m(a)\tilde{m}(a) - 2\gamma^2 a^2\theta^2\, \tilde{m}'(a)m^2(a)] - [4\gamma^2\theta m(a) + 4\gamma^2 a\,\theta m'(a)]\Big]^{1/2} = o\left(\frac{1}{\sqrt{N}}\right).\end{aligned}$$

In the leading order, the solution a is found from the equation $1 - a\theta\, m(a)\tilde{m}(a) = 0$. Plugging the formula for $m(z)$ from (9.22), this becomes

$$\begin{aligned}(9.43)\quad 1 = a\theta\, \frac{a - \sqrt{(a-\lambda_+)(a-\lambda_-)} + \gamma^2 - 1}{2\gamma^2 a} \cdot \frac{a - \sqrt{(a-\lambda_+)(a-\lambda_-)} + 1 - \gamma^2}{2a} \\ = \theta\, \frac{a - 1 - \gamma^2 - \sqrt{(a-\lambda_+)(a-\lambda_-)}}{2\gamma^2},\end{aligned}$$

which is equivalent to

$$a = \theta + 1 + \gamma^2 + \frac{\gamma^2}{\theta} = \frac{1}{\theta}(\gamma^2+\theta)(1+\theta), \tag{9.44}$$

which matches the formula for $\lambda(\theta)$ in (2.12).

Further, continuing to treat (9.42) as an equation on a, we develop the second order expansion of the solution as $N \to \infty$. For that we note the following simplifications whenever a is given by (9.44):

$$\sqrt{(a-\lambda_+)(a-\lambda_-)} = \theta - \frac{\gamma^2}{\theta}, \qquad m\,(a) = \frac{1}{\theta+\gamma^2}, \qquad \tilde{m}(a) = \frac{1}{1+\theta}. \tag{9.45}$$

$$m'(a) = \frac{-\theta^2}{(\gamma^2+\theta)^2\,(\theta^2-\gamma^2)}, \qquad \tilde{m}'(a) = \frac{-\theta^2}{(\theta^2-\gamma^2)\,(1+\theta)^2}. \tag{9.46}$$

Hence, if we assume that a is close to (9.44) and define small Δa through

$$a = \theta + 1 + \gamma^2 + \frac{\gamma^2}{\theta} + \Delta a, \tag{9.47}$$

then expanding in Δa using (9.45), (9.46), we get

$$1 - a\theta\, m(a)\tilde{m}(a) = \Delta a \frac{\theta}{\theta^2 - \gamma^2} + O(\Delta a^2).$$

In addition, the expression after $\mathcal{N}(0,1)$ under $[\cdot]^{1/2}$ in (9.42) simplifies upon plugging a from (9.44): using (9.45), (9.46), we get

$$2\gamma^2 \, \frac{1 + \gamma^2 + 2\theta}{\theta^2 - \gamma^2}.$$

Hence, in terms of Δa the equation (9.42) becomes

$$\Delta a \frac{\theta}{\theta^2 - \gamma^2} + \frac{\mathcal{N}(0,1)}{\sqrt{N}} \left[2\gamma^2 \, \frac{1 + \gamma^2 + 2\theta}{\theta^2 - \gamma^2} \right]^{1/2} + O(\Delta a^2) + o\left(\frac{1}{\sqrt{N}} \right) = 0.$$

Its solution gives the desired asymptotic statement (9.35), matching the formula for $V(\theta)$ in (2.12). $\square$

Proof of Theorem 4.6 for the factor model of Section 2.3. The constants (4.8) simplify to $\kappa_1 = \gamma(1+\gamma)^{4/3}$, $\kappa_2 = \frac{1}{\gamma(1+\gamma)^{2/3}}$ – exactly the same as in the proofs for the spiked covariance model in Section 2.2. Hence, Assumption 8.19 will be satisfied.

We analyze the equation (9.29) for $\theta = \theta_q$ in the form of (9.38), i.e., we study

$$\left(1 - \sqrt{\frac{\theta_q}{\sum_{i=1}^N \xi_i^2 \sum_{j=1}^S \eta_j^2}} \sum_{i=1}^N \frac{\sqrt{\lambda_i}\xi_i \eta_i}{a - \lambda_i} \right)^2 = \frac{a\theta_q}{\sum_{i=1}^N \xi_i^2 \sum_{j=1}^S \eta_j^2} \left(\sum_{i=1}^N \frac{\xi_i^2}{a - \lambda_i} \right) \left(\sum_{j=1}^S \frac{\eta_j^2}{a - \lambda_j} \right), \tag{9.48}$$

where the asymptotics of $(\lambda_i)_{i=1}^N$ is given to us by Proposition 9.14 and ξ_i and η_j are i.i.d. $\mathcal{N}(0,1)$. Arguing as in the previous sections, the $q-1$ largest roots of the equation are close to $\lambda_1, \ldots, \lambda_{q-1}$, resulting in (4.6). In order to establish (4.7), we need to approximate (9.48) for a close to $\lambda_+ = (1+\gamma)^2$ and locate the root of the equation in the $(\lambda_q, \lambda_{q-1})$ interval. We change the variables

$$b = N^{2/3} \frac{a - \lambda_+}{\kappa_1} = N^{2/3} \frac{a - (1+\gamma)^2}{\gamma(1+\gamma)^{4/3}}, \qquad a = (1+\gamma)^2 + N^{-2/3}\gamma(1+\gamma)^{4/3} b, \tag{9.49}$$

recall that $\theta_c = \gamma$, $\theta_q = \gamma + N^{-1/3}\tilde{\theta}$, and aim to apply Theorem 8.20 to (9.48). For this computation, we can approximate $\sum_{i=1}^N \xi_i^2 \approx N$ and $\sum_{j=1}^S \eta_j^2 \approx S$, because the relative errors in these approximations are of order $N^{-1/2}$, which is smaller than $N^{-1/3}$, the scale of our interest. Hence, (9.48) becomes

$$\begin{aligned} &\left(1 - \sqrt{\frac{\gamma + N^{-1/3}\tilde{\theta}}{NS}} \sum_{i=1}^N \frac{\sqrt{\lambda_i}\xi_i \eta_i}{a - \lambda_i} \right)^2 \\ &\quad = \frac{\left((1+\gamma)^2 + N^{-2/3}\gamma(1+\gamma)^{4/3} b \right) (\gamma + N^{-1/3}\tilde{\theta})}{NS} \left(\sum_{i=1}^N \frac{\xi_i^2}{a - \lambda_i} \right) \left(\sum_{j=1}^S \frac{\eta_j^2}{a - \lambda_j} \right). \end{aligned} \tag{9.50}$$

For the factors in the right-hand side of (9.50) we apply Theorem 8.20 with $h(x) = 1$. For the left-hand side we write

$$\xi_i\eta_i = \frac{1}{2}\left[\left(\frac{\xi_i+\eta_i}{\sqrt{2}}\right)^2 - \left(\frac{\xi_i-\eta_i}{\sqrt{2}}\right)^2\right]$$

and apply Theorem 8.20 twice with $h(x) = \sqrt{x}$. In view of Remark 8.23, the convergence is joint over all four applications of Theorem 8.20, and we get a limit expressed in terms of the four (correlated through the choices of ξ_j sequences) copies of $\mathcal{G}(w)$. As noted after (9.37), we use $\mathfrak{s} = \frac{1}{\gamma^{3/2}(1+\gamma)^2}$, $\mathfrak{m} = \frac{1}{\gamma(1+\gamma)}$; we also recall $\frac{N}{S} \to \gamma^2 + O\left(\frac{1}{N}\right)$. As a result, after dropping all $o(N^{-1/3})$ terms, (9.50) becomes:

$$\begin{aligned}(9.51)\quad 1 - N^{-1/3}\gamma^{1/2}(1+\gamma)^{-4/3}\sqrt{(1+\gamma)^2}\left(\mathcal{G}^{(1)}(b) - \mathcal{G}^{(2)}(b)\right) + o(N^{-1/3})\\
= o(N^{-1/3}) + N^{-1/3}\tilde{\theta}(1+\gamma)^2\gamma^2\frac{1}{\gamma(1+\gamma)}\left(\frac{1}{\gamma(1+\gamma)} + \frac{\frac{1}{\gamma^2}-1}{(1+\gamma)^2}\right)\\
+(1+\gamma)^2\gamma^3\left(\frac{1}{\gamma(1+\gamma)} + \frac{N^{-1/3}}{\gamma(1+\gamma)^{4/3}}\mathcal{G}^{(3)}(b)\right)\left(\frac{1}{\gamma(1+\gamma)} + \frac{N^{-1/3}}{\gamma(1+\gamma)^{4/3}}\mathcal{G}^{(4)}(b) + \frac{\frac{1}{\gamma^2}-1}{(1+\gamma)^2}\right).\end{aligned}$$

The order 1 terms cancel out and multiplying (9.51) by $N^{1/3}$, we finally get:

$$-\tilde{\theta}\frac{1}{\gamma} + o(1) = \frac{1}{(1+\gamma)^{1/3}}\mathcal{G}^{(3)}(b) + \frac{\gamma}{(1+\gamma)^{1/3}}\mathcal{G}^{(4)}(b) + \gamma^{1/2}\frac{\left(\mathcal{G}^{(1)}(b) - \mathcal{G}^{(2)}(b)\right)}{(1+\gamma)^{1/3}}. \tag{9.52}$$

At this step an algebraic miracle happens; the following claim is responsible for the appearance of exactly the same function $\mathcal{T}(\Theta)$ in the asymptotics of the factor model:

Claim. The right-hand side of (9.52) is the same random function as $(1+\gamma)^{2/3}\mathcal{G}(b)$.

The claim is established by recalling which noises enter into the functions $\mathcal{G}^{(k)}$, $k = 1, 2, 3, 4$: these are $\left(\frac{\xi_j+\eta_j}{\sqrt{2}}\right)^2$, $\left(\frac{\xi_j-\eta_j}{\sqrt{2}}\right)^2$, ξ_j^2, η_j^2, respectively. Hence, in the linear combination of $\mathcal{G}^{(k)}$ of (9.52), the noises combine into

$$\frac{1}{(1+\gamma)^{1/3}}\xi_j^2 + \frac{\gamma}{(1+\gamma)^{1/3}}\eta_j^2 + \gamma^{1/2}\frac{2\xi_j\eta_j}{(1+\gamma)^{1/3}} = \frac{\left(\xi_j + \gamma^{1/2}\eta_j\right)^2}{(1+\gamma)^{1/3}} \stackrel{d}{=} (1+\gamma)^{2/3}\left[\mathcal{N}(0,1)\right]^2.$$

Using the claim, we conclude that (in distribution) the solution of the equation (9.52) converges as $N \to \infty$ to the solution of

$$-\tilde{\theta}\frac{1}{\gamma(1+\gamma)^{2/3}} = \mathcal{G}(b). \tag{9.53}$$

Recognizing the constant κ_2 in the left-hand side of the last formula and comparing with Definition 4.2, we arrive at (4.7). $\square$

9.4. **Canonical Correlation Analysis.** The proof of Theorem 4.6 for CCA of Section 2.4 follows the same outline as in the previous three sections. However, an analogue of Propositions 9.1, 9.6, 9.10 becomes even more complicated, which introduces new challenges in the proofs. Note that the CCA setting is symmetric under $N \leftrightarrow M$ and there is no loss of generality to assume $N \leq M$, as we do throughout this section. The reader may consult Bykhovskaya and Gorin [2024] for general information on CCA.

Suppose that we are given $N \times S$ matrix $\mathbf{U}$ and $M \times S$ matrix $\mathbf{V}$, in which the first rows are denoted $\mathbf{u}^*$ and $\mathbf{v}^*$, respectively. The remaining rows form matrices $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$ of sizes $(N-1) \times S$ and $(M-1) \times S$, respectively. We would like to connect sample canonical correlations between $\mathbf{U}$ and $\mathbf{V}$ to $\mathbf{u}^*$, $\mathbf{v}^*$, and the sample canonical correlations and variables between $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$. The latter are denoted $c_1, \dots, c_{N-1}$, $\mathbf{u}_1, \dots, \mathbf{u}_{N-1}$, $\mathbf{v}_1, \dots, \mathbf{v}_{M-1}$, where $1 \ge c_1 \ge \dots \ge c_{N-1} \ge 0$, $\mathbf{u}_i$ are S–dimensional vectors forming an orthonormal basis in the space spanned by rows of $\widetilde{\mathbf{U}}$, $\mathbf{v}_j$ are S–dimensional vectors forming an orthonormal basis in the space spanned by the rows of $\widetilde{\mathbf{V}}$, and $\langle \mathbf{u}_i, \mathbf{v}_j \rangle = c_i \delta_{i=j}$, $1 \le i \le N-1$, $1 \le j \le M-1$. For the convenience of notation, we also introduce numbers $c_N = c_{N+1} = \dots = c_{M-1} = 0$.

Proposition 9.15. *For each squared sample canonical correlation a between $\mathbf{U}$ and $\mathbf{V}$, either*

$$\begin{aligned}
&\Bigg[\langle \mathbf{u}^*, \mathbf{v}^* \rangle + \sum_{j=1}^{M-1} \frac{\langle \mathbf{u}^*, \mathbf{v}_j \rangle (c_j \langle \mathbf{v}^*, \mathbf{u}_j \rangle - a \langle \mathbf{v}^*, \mathbf{v}_j \rangle)}{a - c_j^2} \\
&\qquad - a \sum_{i=1}^{N-1} \frac{\langle \mathbf{u}^*, \mathbf{u}_i \rangle (\langle \mathbf{v}^*, \mathbf{u}_i \rangle - c_i \langle \mathbf{v}^*, \mathbf{v}_i \rangle)}{a - c_i^2} \Bigg]^2 \\
&= a \left[-\langle \mathbf{u}^*, \mathbf{u}^* \rangle + \sum_{j=1}^{M-1} \frac{\langle \mathbf{u}^*, \mathbf{v}_j \rangle^2 - 2c_j \langle \mathbf{u}^*, \mathbf{v}_j \rangle \langle \mathbf{u}^*, \mathbf{u}_j \rangle}{a - c_j^2} + a \sum_{i=1}^{N-1} \frac{\langle \mathbf{u}^*, \mathbf{u}_i \rangle^2}{a - c_i^2} \right] \\
&\quad \times \left[-\langle \mathbf{v}^*, \mathbf{v}^* \rangle + \sum_{i=1}^{N-1} \frac{\langle \mathbf{v}^*, \mathbf{u}_i \rangle^2 - 2c_i \langle \mathbf{v}^*, \mathbf{u}_i \rangle \langle \mathbf{v}^*, \mathbf{v}_i \rangle}{a - c_i^2} + a \sum_{j=1}^{M-1} \frac{\langle \mathbf{v}^*, \mathbf{v}_j \rangle^2}{a - c_j^2} \right],
\end{aligned} \tag{9.54}$$

or $a = c_i^2$ for $1 \le i \le N-1$, where c_i has multiplicity one, $\langle \mathbf{u}^*, \mathbf{u}_i \rangle = \langle \mathbf{u}^*, \mathbf{v}_i \rangle = \langle \mathbf{v}^*, \mathbf{u}_i \rangle = \langle \mathbf{v}^*, \mathbf{v}_i \rangle = 0$, and the equation (9.54) *holds with the i-th terms excluded; or $a = c_i^2$, where c_i has multiplicity larger than one.*

We omit the proof, see Bykhovskaya and Gorin [2025, Appendix A], with $N = K$. The next equation replaces (9.8), (9.19), (9.38) in the context of CCA.

Corollary 9.16. *In the setting of Section 2.4, for each $1 \le k \le r$, the squared sample canonical correlations of $\mathbf{U}$ and $\mathbf{V}$ solve an equation in variable a*

$$\begin{aligned}
&\Bigg[\sum_{i=1}^{S} \xi_i \eta_i + \sum_{j=1}^{M-1} \frac{((1-c_j^2)^{\frac12} \xi_{j+N-1} + c_j \xi_j)((1-a) c_j \eta_j - a(1-c_j^2)^{\frac12} \eta_{j+N-1})}{a - c_j^2} \\
&\qquad - a \sum_{i=1}^{N-1} \frac{(1-c_i^2)^{1/2} \xi_i \big((1-c_i^2)^{1/2} \eta_i - c_i \eta_{i+N-1}\big)}{a - c_i^2} \Bigg]^2 \\
&= a \left[-\sum_{i=1}^{S} \xi_i^2 + \sum_{j=1}^{M-1} \frac{\big((1-c_j^2)^{1/2} \xi_{j+N-1} + c_j \xi_j\big)^2 - 2c_j \xi_j \big((1-c_j^2)^{1/2} \xi_{j+N-1} + c_j \xi_j\big)}{a - c_j^2} + a \sum_{i=1}^{N-1} \frac{\xi_i^2}{a - c_i^2} \right] \\
&\times \left[-\sum_{i=1}^{S} \eta_i^2 + \sum_{i=1}^{N-1} \frac{\eta_i^2 - 2c_i \eta_i ((1-c_i^2)^{1/2} \eta_{i+N-1} + c_i \eta_i)}{a - c_i^2} + a \sum_{j=1}^{M-1} \frac{\big((1-c_j^2)^{1/2} \eta_{j+N-1} + c_j \eta_j\big)^2}{a - c_j^2} \right],
\end{aligned} \tag{9.55}$$

where (ξ_i, η_i) are i.i.d. Gaussian mean 0 two-dimensional vectors with covariance matrix

$$\begin{pmatrix} \mathbb{E}\xi_i^2 & \mathbb{E}\xi_i\eta_i \\ \mathbb{E}\eta_i\xi_i & \mathbb{E}\eta_i^2 \end{pmatrix} = \begin{pmatrix} C_{uu} & C_{uv} \\ C_{uv} & C_{vv} \end{pmatrix}, \qquad \frac{C_{uv}^2}{C_{uu}C_{vv}} = \theta_k, \tag{9.56}$$

and c_i are squared sample canonical correlations between $\tilde{\mathbf{U}}$, $\tilde{\mathbf{V}}$, which are $(N-1)\times S$ and $(M-1)\times S$ matrices, independent from (ξ_i, η_i) and produced trough the same mechanism as $\mathbf{U}$, $\mathbf{V}$, but with r smaller by 1 and θ_k removed from $\{\theta_1, \theta_2, \dots, \theta_r\}$.

Remark 9.17. *Due to invariance of (9.55) under multiplications of ξ_i or η_j by constants, there is no loss of generality in assuming $C_{uu} = C_{vv} = 1$.*

Proof of Corollary 9.16. We fix the index $1 \le k \le r$ and connect Proposition 9.15 to the setting of Section 2.4. For that we choose two deterministic invertible matrices $\widetilde{A}$ of size $N \times N$ and $\widetilde{B}$ of size $M \times M$, so that:

- The first coordinate of $\widetilde{A}\mathbf{u}$ is independent from all the coordinates other than the first in $\widetilde{A}\mathbf{u}$ and in $\widetilde{B}\mathbf{v}$.
- The first coordinate of $\widetilde{B}\mathbf{v}$ is independent from all the coordinates other than the first in $\widetilde{A}\mathbf{u}$ and in $\widetilde{B}\mathbf{v}$.
- The variances of the first coordinates of $\widetilde{A}\mathbf{u}$ and $\widetilde{B}\mathbf{v}$ are 1.
- The squared correlation coefficient between the first coordinates of $\widetilde{A}\mathbf{u}$ and $\widetilde{B}\mathbf{v}$ is θ_k.

The existence of such $\widetilde{A}$ and $\widetilde{B}$ readily follows from the existence of the decomposition (2.14), which is the basic statement in CCA – the first coordinates of $\widetilde{A}\mathbf{u}$ and $\widetilde{B}\mathbf{v}$ are canonical variables, corresponding to the canonical correlation θ_k.

Multiplying the matrices $\mathbf{U}$ and $\mathbf{V}$ by $\widetilde{A}$ and $\widetilde{B}$, respectively, does not change the squared sample canonical correlations, and brings them to the form of Proposition 9.15. It remains to explain that (9.54) is the same as (9.55). The correlation structure between the last $N-1$ rows of $\widetilde{A}\mathbf{U}$ and the last $M-1$ rows of $\widetilde{B}\mathbf{V}$ is as in (2.14) but with θ_k removed from $\{\theta_1, \theta_2, \dots, \theta_r\}$. Hence c_i^2 in (9.54) match their description in Corollary 9.16. The first rows of $\widetilde{A}\mathbf{U}$ and $\widetilde{B}\mathbf{V}$ are S–dimensional vectors with i.i.d. components, and with correlation structure as in (9.56). We should further understand the joint distribution of the four arrays of random scalar products appearing in (9.54).

$$\langle \mathbf{u}^*, \mathbf{u}_i\rangle, \quad \langle \mathbf{v}^*, \mathbf{u}_i\rangle, \quad 1 \le i \le N-1, \qquad \langle \mathbf{u}^*, \mathbf{v}_j\rangle, \quad \langle \mathbf{v}^*, \mathbf{v}_j\rangle, \quad 1 \le j \le M-1. \tag{9.57}$$

If all $\mathbf{u}_i$,$\mathbf{v}_j$ were orthonormal, then computing scalar products would be easy: scalar products of i.i.d. mean 0 Gaussian vectors with orthonormal basis are again i.i.d. Gaussian vectors. However, $\langle \mathbf{u}_i, \mathbf{v}_i\rangle \ne 0$. In order to fix this, we introduce new vectors $\mathbf{v}'_j$, $1 \le j \le M-1$, which are:

$$\frac{\mathbf{v}_1 - c_1\mathbf{u}_1}{\sqrt{1-c_1^2}}, \frac{\mathbf{v}_2 - c_2\mathbf{u}_2}{\sqrt{1-c_2^2}}, \dots, \frac{\mathbf{v}_{N-1} - c_{N-1}\mathbf{u}_{N-1}}{\sqrt{1-c_{N-1}^2}}, \qquad \mathbf{v}_N, \dots, \mathbf{v}_{M-1}.$$

Since $\{\mathbf{u}_i\}$ are orthonormal, $\{\mathbf{v}_j\}$ are orthonormal, and $\langle \mathbf{u}_i, \mathbf{v}_j\rangle = c_i\delta_{i=j}$, we conclude that the $N+M-2$ vectors $\mathbf{u}_1, \dots, \mathbf{u}_{N-1}$, $\mathbf{v}'_1, \dots, \mathbf{v}'_{M-1}$ are also orthonormal.

Since $\mathbf{u}^*, \mathbf{v}^*$ are Gaussian and independent from all $\mathbf{u}_j$, $\mathbf{v}_j$, the scalar products $\langle \mathbf{u}^*, \mathbf{u}_i\rangle, \langle \mathbf{v}^*, \mathbf{u}_i\rangle, \langle \mathbf{u}^*, \mathbf{v}'_j\rangle, \langle \mathbf{v}^*, \mathbf{v}'_j\rangle$ are also Gaussian and independent from all $\mathbf{u}_j$ and $\mathbf{v}_j$.

Among these scalar products, most are pairwise independent, with the only non-zero covariances being:

$$\mathbb{E}[\langle \mathbf{u}^*, \mathbf{u}_i\rangle]^2 = \mathbb{E}[\langle \mathbf{v}^*, \mathbf{u}_i\rangle]^2 = \mathbb{E}[\langle \mathbf{u}^*, \mathbf{v}'_j\rangle]^2 = \mathbb{E}[\langle \mathbf{v}^*, \mathbf{v}'_j\rangle]^2 = 1.$$

$$\mathbb{E}\langle \mathbf{u}^*, \mathbf{u}_i\rangle\langle \mathbf{v}^*, \mathbf{u}_i\rangle = \mathbb{E}\langle \mathbf{u}^*, \mathbf{v}'_j\rangle\langle \mathbf{v}^*, \mathbf{v}'_j\rangle = \sqrt{\theta_k}.$$

Let us choose an orthonormal basis $\mathbf{w}_1, \dots, \mathbf{w}_S$ of the S–dimensional space, such that the first $N+M-2$ vectors are $\mathbf{u}_1, \dots, \mathbf{u}_{N-1}$, $\mathbf{v}'_1, \dots, \mathbf{v}'_{M-1}$ and the rest are arbitrary (the choice is independent from $\mathbf{u}^*$ and $\mathbf{v}^*$). Define

$$\xi_i = \langle \mathbf{u}^*, \mathbf{w}_i\rangle, \qquad \eta_i = \langle \mathbf{v}^*, \mathbf{w}_i\rangle.$$

Then (ξ_i, η_i), $1 \le i \le S$, are i.i.d. vectors with correlation structure (9.56). We express various scalar products in (9.54) trough ξ_i and η_i:

$$\langle \mathbf{u}^*, \mathbf{v}^*\rangle = \sum_{i=1}^{S} \xi_i \eta_i, \qquad \langle \mathbf{u}^*, \mathbf{u}^*\rangle = \sum_{i=1}^{S} \xi_i^2, \qquad \langle \mathbf{v}^*, \mathbf{v}^*\rangle = \sum_{i=1}^{S} \eta_i^2.$$

$$\langle \mathbf{u}^*, \mathbf{u}_i\rangle = \xi_i, \quad \langle \mathbf{v}^*, \mathbf{u}_i\rangle = \eta_i,$$

$$\langle \mathbf{u}^*, \mathbf{v}_j\rangle = \sqrt{1-c_j^2}\,\xi_{j+N-1} + c_j\xi_j, \quad \langle \mathbf{v}^*, \mathbf{v}_j\rangle = \sqrt{1-c_j^2}\,\eta_{j+N-1} + c_j\eta_j.$$

Plugging these expressions into (9.54), we get (9.55). □

We record a version of Corollaries 9.3, 9.8, 9.12, which is slightly more complicated, leading to an additional step in the proofs, on which we comment in Remark 9.20.

Corollary 9.18. *Let N roots of (9.54) be denoted $a_1, \dots, a_N$. Then all a_i are real numbers between 0 and 1. Moreover, if we arrange a_i in the decreasing order, then there exists another sequence of $N-1$ real numbers $y_1 \ge y_2 \ge \dots \ge y_{N-1}$, such that two interlacing conditions hold:*

$$a_1 \ge y_1 \ge a_2 \ge \dots \ge y_{N-1} \ge a_N, \qquad \text{and} \qquad y_1 \ge c_1^2 \ge y_2 \ge \dots \ge y_{N-1} \ge c_{N-1}^2. \tag{9.58}$$

In the notation of the proof of Corollary 9.16, the numbers $y_1, \dots, y_{N-1}$ are squared sample canonical correlations between $\widetilde{\mathbf{U}}$ =(last $N-1$ rows of $\widetilde{A}\mathbf{U}$) and $\mathbf{V}$.

Proof. This is a version of Lemma A.6 in Bykhovskaya and Gorin [2025]. $a_1, \dots, a_N$ are eigenvalues of the product of projections on $\mathbf{U}$ and on $\mathbf{V}$ in S–dimensional space, i.e., $P_{\mathbf{U}}P_{\mathbf{V}}$ (or $P_{\mathbf{V}}P_{\mathbf{U}}P_{\mathbf{V}}$ or $P_{\mathbf{U}}P_{\mathbf{V}}P_{\mathbf{U}}$), while $y_1, \dots, y_{N-1}$ are eigenvalues of the product of projections on smaller $\widetilde{\mathbf{U}}$ and $\mathbf{V}$. Then (9.58) are two instances interlacing inequalities between eigenvalues of a matrix and its principal submatrix, as in Bhatia [1997, Corollry III.1.5]. □

Next, we prove an analogue of Propositions 9.5, 9.9, 9.14.

Proposition 9.19. *Consider the CCA model, as in Section 2.4, with $\theta_1 > \theta_2 > \dots > \theta_r \ge 0$ split into two groups: $\theta_1, \dots, \theta_{q-1} > \theta^c = \frac{1}{\sqrt{(\tau_M-1)(\tau_N-1)}}$ and $\theta_q, \dots, \theta_r < \theta^c = \frac{1}{\sqrt{(\tau_M-1)(\tau_N-1)}}$. Then, with $\frac{S}{N} = \tau_N + O\left(\frac{1}{N}\right)$, $\frac{S}{M} = \tau_M + O\left(\frac{1}{N}\right)$, in the sense of convergence in joint distribution and using (2.16):*

$$\lim_{N\to\infty} \sqrt{N}(\lambda_i - \lambda(\theta_i)) \stackrel{d}{=} \mathcal{N}(0, V(\theta_i)), \qquad 1 \le i \le q-1, \tag{9.59}$$

$$\lim_{N\to\infty} N^{2/3} \left[\frac{(\sqrt{\tau_N-1}+\sqrt{\tau_M-1})^{4/3}(\sqrt{\tau_N-1}\sqrt{\tau_M-1}-1)^{4/3}}{\tau_N^{5/3}\tau_M(\tau_N-1)^{1/6}(\tau_M-1)^{1/6}}\right]^{-1} (\lambda_i - \lambda_+) \stackrel{d}{=} \mathfrak{a}_{i-q+1}, \qquad i \ge q, \tag{9.60}$$

where $\mathcal{N}(0, V(\theta_i))$, $0 \le i \le q-1$, are independent and $\{\mathfrak{a}_j\}_{j\ge 1}$ are points of the Airy$_1$ point process independent from (9.59). *In addition,* (8.46) *hold for the sample squared canonical correlations between* $\mathbf{U}$ *and* $\mathbf{V}$ *with* $h(x) = 1$, $h(x) = \sqrt{x}$, $h(x) = \sqrt{1-x}$, *and* $h(x) = \sqrt{x(1-x)}$.

Proof. The final statement, (8.46) with various choices of $h(x)$, follows by induction on r from Theorem 8.24, Corollary 8.25, Corollary 8.27 for $r = 0$, and using Lemma 8.28 with Corollary 9.18 for the induction step. Further, various parameters are:

$$\begin{aligned}
\lambda_\pm &= \left(\sqrt{\tau_M^{-1}(1-\tau_N^{-1})} \pm \sqrt{\tau_N^{-1}(1-\tau_M^{-1})}\right)^2 = \frac{\left(\sqrt{\tau_N - 1} \pm \sqrt{\tau_M - 1}\right)^2}{\tau_N \tau_M}, \\
\mu(x)\mathrm{d}x &= \frac{\tau_N}{2\pi} \frac{\sqrt{(x-\lambda_-)(\lambda_+ - x)}}{x(1-x)} \mathbf{1}_{[\lambda_-,\lambda_+]}\,\mathrm{d}x, \\
m(z) &= \frac{\tau_M^{-1} + \tau_N^{-1} - z + \sqrt{(z-\lambda_-)(z-\lambda_+)}}{2\tau_N^{-1} z(z-1)} + \frac{1}{z},
\end{aligned} \tag{9.61}$$

which are the Wachter law and its Stieljes transform, respectively (see, e.g., Bykhovskaya and Gorin [2024, Section 3]), and the constants of (8.45) are computed[10] to be

(9.62)
$$\mathfrak{s} = \frac{\sqrt{\lambda_+ - \lambda_-}}{2\tau_N^{-1}\lambda_+(1-\lambda_+)} = \tau_N^{5/2}\tau_M^{3/2} \frac{\sqrt[4]{(\tau_N - 1)(\tau_M - 1)}}{(\sqrt{\tau_N - 1} + \sqrt{\tau_M - 1})^2(\sqrt{\tau_N - 1}\sqrt{\tau_M - 1} - 1)^2},$$

(9.63)
$$\mathfrak{m} = \frac{\tau_M^{-1} + \tau_N^{-1} - \lambda_+}{2\tau_N^{-1}\lambda_+(\lambda_+ - 1)} + \frac{1}{\lambda_+} = \frac{\tau_N^{-1} - \tau_M^{-1} + (1 - 2\tau_N^{-1})\lambda_+}{2\tau_N^{-1}\lambda_+(1-\lambda_+)} = \frac{\tau_M - \tau_N + (\tau_N - 2)\tau_M\lambda_+}{2\tau_M\lambda_+(1-\lambda_+)}.$$

Statements close to (9.59), (9.60) are known from Bao et al. [2019], Yang [2022b]. Alternatively, the proof can be obtained by the same argument as in Proposition 9.5, by induction on r with the base case $r = 0$ given in Johnstone [2008], Han et al. [2018] and the step based on Corollary 9.16. We only highlight the key computation, which is an analogue of (9.11), (9.12).

We recall (9.55) and analyze its asymptotic behavior for $a > \lambda_+$ and assuming that $\{c_i^2\}$ are squared sample canonical correlations between independent $\mathbf{U}$ and $\mathbf{V}$ of sizes $(N-1)\times S$ and $(M-1)\times S$, respectively, and with i.i.d. $\mathcal{N}(0,1)$ random variables as their elements. In this situation the joint distribution of $\{c_i^2\}$ is the JOE ensemble (see Section 8.4), and their empirical measure converges to the Wachter law (9.61) as $N \to \infty$. According to Remark 9.17, we use (9.56) with $C_{uu} = C_{vv} = 1$, $C_{uv} = \sqrt{\theta_k}$; we also recall $c_j = 0$ for $j \ge N$. We divide (9.55) by N^2 and split its sums into their expectations (conditional on $\{c_i\}$) and the mean 0 parts. The left-hand side of (9.55) is transformed into:

[10] Note that $1 - \lambda_+ = \frac{(\sqrt{\tau_N - 1}\sqrt{\tau_M - 1} - 1)^2}{\tau_N \tau_M}$.

$$\text{(9.64)}\quad \Bigg[\sqrt{\theta_k}\left(\frac{S-M+1}{N}-\sum_{i=1}^{N-1}\frac{a(1-a)}{N(a-c_i^2)}-a\frac{N-1}{N}\right)$$
$$+\sum_{i=1}^{S}\frac{\xi_i\eta_i-\sqrt{\theta_k}}{N}+\sum_{j=1}^{N-1}\left[\frac{((1-c_j^2)^{\frac12}\xi_{j+N-1}+c_j\xi_j)((1-a)c_j\eta_j-a(1-c_j^2)^{\frac12}\eta_{j+N-1})}{N(a-c_j^2)}+\frac{\sqrt{\theta_k}}{N}\right]$$
$$-\sum_{j=N}^{M-1}\frac{\xi_{j+N-1}\eta_{j+N-1}-\sqrt{\theta_k}}{N}-a\sum_{i=1}^{N-1}\frac{(1-c_i^2)^{1/2}\xi_i\big((1-c_i^2)^{1/2}\eta_i-c_i\eta_{i+N-1}\big)-(1-c_i^2)\sqrt{\theta_k}}{N(a-c_i^2)}\Bigg]^2,$$

where the first line is of constant order and deterministic as $N\to\infty$; the second and third lines are of order $N^{-1/2}$ and become Gaussian as $N\to\infty$. The $[\cdot]$ factor in the third line of (9.55) is similarly transformed into:

$$\text{(9.65)}\quad \left[\frac{2N-S-2}{N}+\frac{M-N}{N}\cdot\frac{1}{a}+\sum_{i=1}^{N-1}\frac{1-a}{N(a-c_i^2)}\right]-\sum_{i=1}^{S}\frac{\xi_i^2-1}{N}$$
$$+\sum_{j=1}^{M-1}\frac{\big((1-c_j^2)^{1/2}\xi_{j+N-1}+c_j\xi_j\big)^2-2c_j\xi_j\big((1-c_j^2)^{1/2}\xi_{j+N-1}+c_j\xi_j\big)-1+2c_j^2}{N(a-c_j^2)}+a\sum_{i=1}^{N-1}\frac{\xi_i^2-1}{N(a-c_i^2)}.$$

The fourth line of (9.55) is transformed into:

$$\text{(9.66)}\quad \left[\frac{M+N-S-2}{N}+\sum_{i=1}^{N-1}\frac{1-a}{N(a-c_i^2)}\right]-\sum_{i=1}^{S}\frac{\eta_i^2-1}{N}$$
$$+\sum_{i=1}^{N-1}\frac{\eta_i^2-2c_i\eta_i((1-c_i^2)^{1/2}\eta_{i+N-1}+c_i\eta_i)-1+2c_i^2}{N(a-c_i^2)}+a\sum_{j=1}^{M-1}\frac{\big((1-c_j^2)^{1/2}\eta_{j+N-1}+c_j\eta_j\big)^2-1}{N(a-c_j^2)}.$$

Hence, using the convergence towards the Wachter law (see Theorem 8.24 or Bai and Silverstein [2010, Sections 4.4, 9.13, 9.14, 9.15] or Dumitriu and Paquette [2012]), in the leading order the equation (9.55) becomes as $N\to\infty$:

$$\text{(9.67)}\quad \theta_k\big[\big(\tau_N-\tau_N\tau_M^{-1}-a(1-a)m(a)-a\big)\big]^2$$
$$=a\big[2-\tau_N+\frac{\tau_N\tau_M^{-1}-1}{a}+(1-a)m(a)\big]\big[\tau_N\tau_M^{-1}+1-\tau_N+(1-a)m(a)\big]+O\big(N^{-1/2}\big).$$

Plugging $m(a)$ from (9.61), simplifying (see Bykhovskaya and Gorin [2025, Lemma B.6] for such a computation), and solving the resulting equation in a, we express the solution as $a=\frac{\big((\tau_N-1)\theta+1\big)\big((\tau_M-1)\theta+1\big)}{\theta\tau_N\tau_M}+O(N^{-1/2})$, thus matching the expression for $\lambda(\theta)$ in (2.16).

Further, similarly to the proofs of Propositions 9.5, 9.9, and 9.14, applying CLT to the second-order terms in (9.64), (9.65), (9.66), we replace $O(N^{-1/2})$ in (9.67) with $N^{-1/2}$ times $\mathcal{N}(0,1)$ times an explicit function of a. Then solving (9.67) again, we arrive at the statement on the asymptotic Gaussianity of the solution a, which matches (2.16). □

Remark 9.20. *In the proof of Proposition 9.5 we located all the largest eigenvalues by analyzing the equation* (9.1) *for the variable a near each $\lambda(\theta_i)$ and near λ_+. In addition, we used the interlacements of Corollary 9.3 to make sure that we did not miss any other* *large eigenvalues. Similarly in the above proof, by analyzing (9.55), we locate a solving the equation near each $\lambda(\theta_i)$ and near λ_+. The argument for not missing other eigenvalues needs to be modified this time: if we use (9.58) naïvely, then, say, in the $r = 1$ case it would allow two eigenvalues $\lambda_1 \geq \lambda_2$ larger than c_1^2, rather than only one in Corollary 9.3.*

The remedy is to still use (9.58), but compare with $y_1, \dots, y_{N-1}$ instead of $c_1^2, \dots, c_N^2$. Note that in the induction step of the argument of Proposition 9.19, the largest eigenvalues $y_1, y_2, \dots,$ satisfy exactly the same induction assumption and therefore exactly the same asymptotics as $c_1^2, c_2^2, \dots$ (the only difference is that N got decreased by 1, but this does not change the asymptotic behavior). With this update, the same arguments, as before, go through. The same remark applies to the following proof.

Proof of Theorem 4.6 for the CCA setting of Section 2.4. The constants (4.8) evaluate to:

$$V'(\theta^c) = \frac{4\sqrt{\tau_N - 1}\sqrt{\tau_M - 1}\big(\sqrt{\tau_N - 1}\sqrt{\tau_M - 1} - 1\big)^2\big(\sqrt{\tau_N - 1} + \sqrt{\tau_M - 1}\big)^2}{\tau_N^3 \tau_M^2}, \tag{9.68}$$

$$\lambda''(\theta^c) = 2\frac{(\tau_N - 1)^{3/2}(\tau_M - 1)^{3/2}}{\tau_N \tau_M}, \tag{9.69}$$

$$\kappa_1 = \frac{1}{2}\frac{[V'(\theta^c)]^{\frac{2}{3}}}{\lambda''(\theta^c)^{\frac{1}{3}}} = \frac{\big(\sqrt{\tau_N - 1}\sqrt{\tau_M - 1} - 1\big)^{4/3}\big(\sqrt{\tau_N - 1} + \sqrt{\tau_M - 1}\big)^{4/3}}{\tau_N^{5/3}\tau_M(\tau_N - 1)^{1/6}(\tau_M - 1)^{1/6}}, \tag{9.70}$$

$$\kappa_2 = \frac{[\lambda''(\theta^c)]^{\frac{2}{3}}}{V'(\theta^c)^{\frac{1}{3}}} = \frac{(\tau_N - 1)^{5/6}(\tau_M - 1)^{5/6}\tau_N^{1/3}}{\big(\sqrt{\tau_N - 1}\sqrt{\tau_M - 1} - 1\big)^{2/3}\big(\sqrt{\tau_N - 1} + \sqrt{\tau_M - 1}\big)^{2/3}}. \tag{9.71}$$

This matches the constants in Proposition 9.19 and, hence, Assumption 8.19 will be satisfied. We analyze the equation (9.55) for $k = q$, i.e., we study

$$\begin{aligned}
&\Bigg[\sum_{i=1}^{S}\xi_i\eta_i + \sum_{j=1}^{M-1}\frac{((1-c_j^2)^{\frac{1}{2}}\xi_{j+N-1} + c_j\xi_j)((1-a)c_j\eta_j - a(1-c_j^2)^{\frac{1}{2}}\eta_{j+N-1})}{a - c_j^2} \\
&\qquad - a\sum_{i=1}^{N-1}\frac{(1-c_i^2)^{1/2}\xi_i\big((1-c_i^2)^{1/2}\eta_i - c_i\eta_{i+N-1}\big)}{a - c_i^2}\Bigg]^2 \\
&= a\Bigg[-\sum_{i=1}^{S}\xi_i^2 + a\sum_{i=1}^{N-1}\frac{\xi_i^2}{a - c_i^2} \\
&\qquad + \sum_{j=1}^{M-1}\frac{\big((1-c_j^2)^{1/2}\xi_{j+N-1} + c_j\xi_j\big)^2 - 2c_j\xi_j\big((1-c_j^2)^{1/2}\xi_{j+N-1} + c_j\xi_j\big)}{a - c_j^2}\Bigg] \\
&\times\Bigg[-\sum_{i=1}^{S}\eta_i^2 + a\sum_{j=1}^{M-1}\frac{\big((1-c_j^2)^{1/2}\eta_{j+N-1} + c_j\eta_j\big)^2}{a - c_j^2} \\
&\qquad + \sum_{i=1}^{N-1}\frac{\eta_i^2 - 2c_i\eta_i((1-c_i^2)^{1/2}\eta_{i+N-1} + c_i\eta_i)}{a - c_i^2}\Bigg],
\end{aligned} \tag{9.72}$$

where (ξ_i, η_i) are mean 0, variance 1 Gaussian i.i.d. in i random variables with the squared correlation coefficient of ξ_i and η_i equal to θ_q, the asymptotics of $(c_i^2)_{i=1}^{N-1}$ is given to us by Proposition 9.19; $c_j = 0$ for $j \geq N$. Arguing as in the previous sections, the $q - 1$ largest roots of the equation are close to $\lambda_1, \dots, \lambda_{q-1}$, resulting in (4.6). In order to establish (4.7), we need to investigate (9.72) for a close to $\lambda_+ = \frac{(\sqrt{\tau_N - 1} + \sqrt{\tau_N - 1})^2}{\tau_N \tau_M}$ and locate the root of the equation in the $(\lambda_q, \lambda_{q-1})$ interval. For this computation, we can approximate $\sum_{i=1}^S \xi_i^2 \approx S$, $\sum_{j=1}^S \eta_j^2 \approx S$, and $\sum_{i=1}^S \xi_i \eta_i \approx S\sqrt{\theta_q}$ because the relative errors in these approximations are of order $N^{-1/2}$, which is smaller than $N^{-1/3}$ scale of our interest. Similarly, the sums $\sum_{j=N}^{M-1}$ can be replaced with their expectations. Hence, up to $O(N^{-1/2})$ error, (9.72) becomes

$$
\begin{aligned}
&\Bigg[\frac{S+N-M}{N}\sqrt{\theta_q} + \sum_{j=1}^{N-1} \frac{((1-c_j^2)^{\frac12}\xi_{j+N-1} + c_j\xi_j)((1-a)c_j\eta_j - a(1-c_j^2)^{\frac12}\eta_{j+N-1})}{N(a-c_j^2)} \\
&\qquad - a\sum_{i=1}^{N-1} \frac{(1-c_i^2)^{1/2}\xi_i\big((1-c_i^2)^{1/2}\eta_i - c_i\eta_{i+N-1}\big)}{N(a-c_i^2)}\Bigg]^2 \\
&= a\Bigg[-\frac{S}{N} + \frac{M-N}{Na} + \sum_{j=1}^{N-1} \frac{\big((1-c_j^2)^{1/2}\xi_{j+N-1} + c_j\xi_j\big)^2 - 2c_j\xi_j\big((1-c_j^2)^{1/2}\xi_{j+N-1} + c_j\xi_j\big)}{N(a-c_j^2)} \\
&\qquad + a\sum_{i=1}^{N-1}\frac{\xi_i^2}{N(a-c_i^2)}\Bigg] \\
&\times\Bigg[\frac{-S+M-N}{N} + \sum_{i=1}^{N-1}\frac{\eta_i^2 - 2c_i\eta_i((1-c_i^2)^{1/2}\eta_{i+N-1} + c_i\eta_i)}{N(a-c_i^2)} + a\sum_{j=1}^{N-1}\frac{\big((1-c_j^2)^{1/2}\eta_{j+N-1} + c_j\eta_j\big)^2}{N(a-c_j^2)}\Bigg].
\end{aligned} \tag{9.73}
$$

We change the variables

$$
b = N^{2/3}\frac{a-\lambda_+}{\kappa_1}, \qquad a = \lambda_+ + N^{-2/3}\kappa_1 b, \tag{9.74}
$$

recall that $\theta^c = \frac{1}{\sqrt{(\tau_N-1)(\tau_M-1)}}$, $\theta_q = \frac{1}{\sqrt{(\tau_N-1)(\tau_M-1)}} + N^{-1/3}\tilde\theta$, and apply Theorem 8.20 to (9.73). Arguing exactly as for the factor model in the previous section, we need several applications of the theorem, leading to several functions $\mathcal{G}^{(1)}(b)$, $\mathcal{G}^{(2)}(b)$, ..., which are then recombined together. The leading deterministic terms recombine into the same expression as (9.67) evaluated at $a = \lambda_+$. Recalling that $m(\lambda_+) = \mathfrak{m}$, as computed in the proof of Proposition 9.19, the first two lines of (9.73), up to $o(N^{-1/3})$ error, become:

$$
\begin{aligned}
&\Bigg[\sqrt{(\tau_N-1)^{-1/2}(\tau_M-1)^{-1/2} + N^{-1/3}\tilde\theta}\,\left(\tau_N - \tau_N\tau_M^{-1} - \lambda_+(1-\lambda_+)\mathfrak{m} - \lambda_+\right) \\
&\quad + \frac{N^{-1/3}}{\kappa_1}\sum_{j=1}^{\infty}\frac{\Big((1-\lambda_+)^{\frac12}\check\xi_j + \sqrt{\lambda_+}\xi_j\Big)\Big((1-\lambda_+)\sqrt{\lambda_+}\eta_j - \lambda_+(1-\lambda_+)^{\frac12}\check\eta_j\Big) - \lambda_+(1-\lambda_+)^{1/2}\xi_j\Big((1-\lambda_+)^{1/2}\eta_j - \sqrt{\lambda_+}\check\eta_j\Big)}{b-\mathfrak{a}_j}\Bigg]^2 \\
&= (\tau_N-1)^{-1/2}(\tau_M-1)^{-1/2}\Big(\tau_N - \tau_N\tau_M^{-1} - \lambda_+(1-\lambda_+)\mathfrak{m} - \lambda_+\Big)^2\Bigg[1 + 4\tilde\theta N^{-1/3}(\tau_N-1)^{1/2}(\tau_M-1)^{1/2} \\
&\quad + 2N^{-1/3}\frac{(\tau_N-1)^{1/2}(\tau_M-1)^{1/2}(1-\lambda_+)\sqrt{\lambda_+}}{\kappa_1\Big(\tau_N - \tau_N\tau_M^{-1} - \lambda_+(1-\lambda_+)\mathfrak{m} - \lambda_+\Big)}\sum_{j=1}^{\infty}\frac{\check\xi_j\Big((1-\lambda_+)^{1/2}\eta_j - \sqrt{\lambda_+}\check\eta_j\Big)}{b-\mathfrak{a}_j} + o(N^{-1/3})\Bigg],
\end{aligned} \tag{9.75}
$$

where the sum $\sum_{j=1}^{\infty}$ is a shortcut for linear combinations of several functions $\mathcal{G}(b)$ which should be formally understood as in (4.2); and $(\xi_j, \eta_j, \check{\xi}_j, \check{\eta}_j)$ are i.i.d. Gaussian random vectors which are coordinate-wise $\mathcal{N}(0,1)$, $\mathbb{E}\xi_j\eta_j = \mathbb{E}\check{\xi}_j\check{\eta}_j = \sqrt{\theta^c}$, and all the other covariances are zero. Similarly, the third and fourth lines of (9.73) turn, up to $o(N^{-1/3})$ error, into:

$$\lambda_+ \left(2-\tau_N+\frac{\tau_N\tau_M^{-1}-1}{\lambda_+}+(1-\lambda_+)\mathfrak{m}\right)\Bigg[1+\frac{N^{-1/3}}{\kappa_1(2-\tau_N-\frac{\tau_N\tau_M^{-1}-1}{\lambda_+}+(1-\lambda_+)\mathfrak{m})} \tag{9.76}$$
$$\times\sum_{j=1}^{\infty}\frac{\left((1-\lambda_+)^{1/2}\check{\xi}_j+\sqrt{\lambda_+}\xi_j\right)^2-2\sqrt{\lambda_+}\xi_j\left((1-\lambda_+)^{1/2}\check{\xi}_j+\sqrt{\lambda_+}\xi_j\right)+\lambda_+\xi_j^2}{b-\mathfrak{a}_j}\Bigg]$$
$$=\lambda_+ \left(2-\tau_N+\frac{\tau_N\tau_M^{-1}-1}{\lambda_+}+(1-\lambda_+)\mathfrak{m}\right)\left[1+\frac{N^{-1/3}(1-\lambda_+)}{\kappa_1(2-\tau_N-\frac{\tau_N\tau_M^{-1}-1}{\lambda_+}+(1-\lambda_+)\mathfrak{m})}\sum_{j=1}^{\infty}\frac{\check{\xi}_j^2}{b-\mathfrak{a}_j}\right].$$

The last line of (9.73) becomes, up to $o(N^{-1/3})$ error:

$$\left(\tau_N\tau_M^{-1}+1-\tau_N+(1-\lambda_+)\mathfrak{m}\right)\Bigg[1+\frac{N^{-1/3}}{\kappa_1(\tau_N\tau_M^{-1}+1-\tau_N+(1-\lambda_+)\mathfrak{m})} \tag{9.77}$$
$$\times\sum_{j=1}^{\infty}\frac{\eta_j^2-2\sqrt{\lambda_+}\eta_j((1-\lambda_+)^{1/2}\check{\eta}_j+\sqrt{\lambda_+}\eta_j)+\lambda_+((1-\lambda_+)^{1/2}\check{\eta}_j+\sqrt{\lambda_+}\eta_j)^2}{b-\mathfrak{a}_j}\Bigg]$$
$$=\left(\tau_N\tau_M^{-1}+1-\tau_N+(1-\lambda_+)\mathfrak{m}\right)\left[1+\frac{N^{-1/3}(1-\lambda_+)}{\kappa_1(\tau_N\tau_M^{-1}+1-\tau_N+(1-\lambda_+)\mathfrak{m})}\sum_{j=1}^{\infty}\frac{((1-\lambda_+)^{1/2}\eta_j-\sqrt{\lambda_+}\check{\eta}_j)^2}{b-\mathfrak{a}_j}\right].$$

Equating (9.75) = (9.76) · (9.77), noting that the leading term cancels (this is precisely the equation relating θ^c with λ_+, cf. (9.67) and the paragraph after it) and multiplying by $N^{1/3}$, we get:

$$-\tilde{\theta}\kappa_1(\tau_N-1)^{1/2}(\tau_M-1)^{1/2}=2\frac{(\tau_N-1)^{1/2}(\tau_M-1)^{1/2}(1-\lambda_+)\sqrt{\lambda_+}}{\tau_N-\tau_N\tau_M^{-1}-\lambda_+(1-\lambda_+)\mathfrak{m}-\lambda_+}\sum_{j=1}^{\infty}\frac{\check{\xi}_j\left((1-\lambda_+)^{1/2}\eta_j-\sqrt{\lambda_+}\check{\eta}_j\right)}{b-\mathfrak{a}_j} \tag{9.78}$$
$$-\frac{(1-\lambda_+)\lambda_+}{\tau_N\tau_M^{-1}-1+\lambda_+(2-\tau_N)+\lambda_+(1-\lambda_+)\mathfrak{m})}\sum_{j=1}^{\infty}\frac{\check{\xi}_j^2}{b-\mathfrak{a}_j}-\frac{(1-\lambda_+)}{\tau_N\tau_M^{-1}+1-\tau_N+(1-\lambda_+)\mathfrak{m}}\sum_{j=1}^{\infty}\frac{((1-\lambda_+)^{1/2}\eta_j-\sqrt{\lambda_+}\check{\eta}_j)^2}{b-\mathfrak{a}_j}+o(1).$$

Similarly to the factor model in the previous section, at this step an algebraic miracle happens, leading to the appearance of exactly the same function $\mathcal{T}(\Theta)$ in the asymptotics. In order to see that we simplify the right-hand side of (9.78). Let us plug the value of $\mathfrak{m}$ from (9.63) and analyze the coefficient of $\frac{1}{b-\mathfrak{a}_j}$ in (9.78), which is:

$$\frac{4\tau_M(\tau_N-1)^{1/4}(\tau_M-1)^{1/4}\sqrt{\lambda_+}(1-\lambda_+)}{2\tau_M\tau_N-\tau_M-\tau_N-\tau_N\tau_M\lambda_+}\check{\xi}_j\left((1-\lambda_+)^{1/2}\eta_j-\sqrt{\lambda_+}\check{\eta}_j\right) \tag{9.79}$$
$$+\frac{2\tau_M\lambda_+(1-\lambda_+)}{\tau_M-\tau_N+(\tau_N-2)\tau_M\lambda_+}\check{\xi}_j^2+\frac{2\tau_M\lambda_+(1-\lambda_+)}{\tau_N-\tau_M+\tau_N(\tau_M-2)\lambda_+}((1-\lambda_+)^{1/2}\eta_j-\sqrt{\lambda_+}\check{\eta}_j)^2.$$

We observe a complete square in the last equation, which would follow from the identity:

$$\left[\frac{4\tau_M(\tau_N-1)^{1/4}(\tau_M-1)^{1/4}\sqrt{\lambda_+}(1-\lambda_+)}{2\tau_M\tau_N-\tau_M-\tau_N-\tau_N\tau_M\lambda_+}\right]^2 \stackrel{?}{=} 4\left[\frac{2\tau_M\lambda_+(1-\lambda_+)}{\tau_M-\tau_N+(\tau_N-2)\tau_M\lambda_+}\right]\left[\frac{2\tau_M\lambda_+(1-\lambda_+)}{\tau_N-\tau_M+\tau_N(\tau_M-2)\lambda_+}\right]. \tag{9.80}$$

The last identity is equivalent to

$$\frac{(\tau_N-1)^{1/2}(\tau_M-1)^{1/2}}{[2\tau_M\tau_N-\tau_M-\tau_N-\tau_N\tau_M\lambda_+]^2} \stackrel{?}{=} \frac{\lambda_+}{[\tau_M-\tau_N+(\tau_N-2)\tau_M\lambda_+][\tau_N-\tau_M+\tau_N(\tau_M-2)\lambda_+]}.$$

Plugging the value of λ_+ from (9.61), we need to show:

$$\frac{\sqrt{\tau_N-1}\sqrt{\tau_M-1}}{[2\tau_M\tau_N-\tau_M-\tau_N-\left(\sqrt{\tau_N-1}+\sqrt{\tau_M-1}\right)^2]^2} \stackrel{?}{=} \frac{\left(\sqrt{\tau_N-1}+\sqrt{\tau_M-1}\right)^2}{[\tau_M\tau_N-\tau_N^2+(\tau_N-2)\left(\sqrt{\tau_N-1}+\sqrt{\tau_M-1}\right)^2][\tau_M\tau_N-\tau_M^2+(\tau_M-2)\left(\sqrt{\tau_N-1}+\sqrt{\tau_M-1}\right)^2]}, \tag{9.81}$$

which can be directly seen to be true by transforming the denominators:

$$\begin{aligned} 2\tau_M\tau_N-\tau_M-\tau_N-\left(\sqrt{\tau_N-1}+\sqrt{\tau_M-1}\right)^2 &= 2\sqrt{\tau_N-1}\sqrt{\tau_M-1}(\sqrt{\tau_N-1}\sqrt{\tau_M-1}-1), \\ \tau_M\tau_N-\tau_N^2+(\tau_N-2)\left(\sqrt{\tau_N-1}+\sqrt{\tau_M-1}\right)^2 &= 2\sqrt{\tau_N-1}(\sqrt{\tau_N-1}\sqrt{\tau_M-1}-1)(\sqrt{\tau_M-1}+\sqrt{\tau_N-1}), \\ \tau_N\tau_M-\tau_M^2+(\tau_M-2)\left(\sqrt{\tau_N-1}+\sqrt{\tau_M-1}\right)^2 &= 2\sqrt{\tau_M-1}(\sqrt{\tau_N-1}\sqrt{\tau_M-1}-1)(\sqrt{\tau_M-1}+\sqrt{\tau_N-1}). \end{aligned} \tag{9.82}$$

Therefore, crucially, (9.79) is the square of a mean 0 Gaussian random variable. The variance of this random variable equals the expectation of (9.79). Recalling that $\check{\xi}_j$, η_j, $\check{\eta}_j$ are $\mathcal{N}(0,1)$ with covariances

$$\mathbb{E}\check{\xi}_j\check{\eta}_j=\sqrt{\theta^c}=(\tau_N-1)^{-1/4}(\tau_M-1)^{-1/4}, \qquad \mathbb{E}\eta_j\check{\xi}_j=\mathbb{E}\eta_j\check{\eta}_j=0,$$

using (9.61) and (9.82) we compute this expectation to be:

$$\begin{aligned} &2\tau_M(1-\lambda_+)\lambda_+\left[\frac{-2}{2\tau_M\tau_N-\tau_M-\tau_N-\tau_N\tau_M\lambda_+}+\frac{1}{\tau_M-\tau_N+(\tau_N-2)\tau_M\lambda_+}+\frac{1}{\tau_N-\tau_M+\tau_N(\tau_M-2)\lambda_+}\right] \\ &\quad= \tau_M^{-1}\tau_N^{-2}(\sqrt{\tau_N-1}\sqrt{\tau_M-1}-1)\left(\sqrt{\tau_N-1}+\sqrt{\tau_M-1}\right)^2 \\ &\quad\times\left[\frac{-2}{\sqrt{\tau_N-1}\sqrt{\tau_M-1}}+\frac{\tau_N}{\sqrt{\tau_N-1}(\sqrt{\tau_M-1}+\sqrt{\tau_N-1})}+\frac{\tau_M}{\sqrt{\tau_M-1}(\sqrt{\tau_M-1}+\sqrt{\tau_N-1})}\right] \\ &\quad= \tau_M^{-1}\tau_N^{-2}\frac{(\sqrt{\tau_N-1}\sqrt{\tau_M-1}-1)^2(\sqrt{\tau_M-1}+\sqrt{\tau_N-1})^2}{\sqrt{\tau_N-1}\sqrt{\tau_M-1}}. \end{aligned} \tag{9.83}$$

We conclude that the equation (9.78) can be rewritten as

$$-\tilde{\theta}\kappa_1\tau_M\tau_N^2\frac{(\tau_N-1)(\tau_M-1)}{(\sqrt{\tau_N-1}\sqrt{\tau_M-1}-1)^2(\sqrt{\tau_M-1}+\sqrt{\tau_N-1})^2}=\mathcal{G}(b)+o(1). \tag{9.84}$$

Plugging the value of κ_1 from (9.70), we get

$$-\tilde{\theta}\frac{\tau_N^{1/3}(\tau_N-1)^{5/6}(\tau_M-1)^{5/6}}{(\sqrt{\tau_N-1}\sqrt{\tau_M-1}-1)^{2/3}(\sqrt{\tau_M-1}+\sqrt{\tau_N-1})^{2/3}}=\mathcal{G}(b)+o(1). \tag{9.85}$$

Using (9.71), recognizing the constant κ_2 in the left-hand side of (9.85), and comparing with Definition 4.2, we arrive at (4.7). □

References

M. Abramowitz and I. A. Stegun. *Handbook of mathematical functions with formulas, graphs, and mathematical tables, Tenth Printing.* US Government printing office, 1972.

L. V. Ahlfors. *Complex analysis.* McGraw-Hill, New York, 3rd edition, 1979.

M. Aizenman and S. Warzel. On the ubiquity of the Cauchy distribution in spectral problems. *Probability Theory and Related Fields*, 163:61–87, 2015.

J. Alt, P. Bourgade, and M. Pain. Local law for β–ensembles with a hard edge. *In preparation*, 2025.

T. W. Anderson. *Introduction to multivariate statistical analysis, 3rd edition.* John Wiley & Sons, 2003.

E. Andreou, P. Gagliardini, E. Ghysels, and M. Rubin. Inference in group factor models with an application to mixed-frequency data. *Econometrica*, 87(4):1267–1305, 2019.

P. Arbenz, W. Gander, and G. H. Golub. Restricted rank modification of the symmetric eigenvalue problem: Theoretical considerations. *Linear Algebra and its Applications*, 104: 75–95, 1988.

J. Bai. Inferential theory for factor models of large dimensions. *Econometrica*, 71(1):135–171, 2003.

J. Bai and S. Ng. Determining the number of factors in approximate factor models. *Econometrica*, 70(1):191–221, 2002.

J. Bai and S. Ng. Approximate factor models with weaker loadings. *Journal of Econometrics*, 235(2):1893–1916, 2023.

Z. Bai and J. W. Silverstein. *Spectral Analysis of Large Dimensional Random Matrices.* Springer, 2nd edition, 2010.

Z. Bai and J.-F. Yao. Central limit theorems for eigenvalues in a spiked population model. *Annales de l'IHP Probabilités et statistiques*, 44:447–474, 2008.

Z. Bai, Z. Hou, J. Hu, D. Jiang, and X. Zhang. Limiting canonical distribution of two large-dimensional random vectors. In *Methodology and Applications of Statistics: A Volume in Honor of CR Rao on the Occasion of his 100th Birthday*, pages 213–238. Springer, 2022.

J. Baik and J. W. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of multivariate analysis*, 97(6):1382–1408, 2006.

J. Baik, G. Ben Arous, and S. Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Annals of Probability*, pages 1643–1697, 2005.

N. Bailey, G. Kapetanios, and M. H. Pesaran. Exponent of cross-sectional dependence: Estimation and inference. *Journal of Applied Econometrics*, 31(6):929–960, 2016.

N. Bailey, G. Kapetanios, and M. H. Pesaran. Measurement of factor strength: Theory and practice. *Journal of Applied Econometrics*, 36(5):587–613, 2021.

Z. Bao, J. Hu, G. Pan, and W. Zhou. Canonical correlation coefficients of high-dimensional gaussian vectors: Finite rank case. *Annals of Statistics*, 47(1):612–640, 2019.

M. Barigozzi and M. Hallin. The dynamic, the static, and the weak factor models and the analysis of high-dimensional time series. *arXiv preprint arXiv:2407.10653*, 2024.

A. Bejan. Largest eigenvalues and sample covariance matrices. Tracy-Widom and Painlevé ii: computational aspects and realization in s-plus with applications. *Preprint:* *http://users.stat.umn.edu/~jiang040/downloadpapers/largesteigen/largesteigen.pdf*, 2005.

F. Benaych-Georges and R. R. Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, 2012.

F. Benaych-Georges, A. Guionnet, and M. Maida. Fluctuations of the extreme eigenvalues of finite rank deformations of random matrices. *Electronic Journal of Probability*, 16: 1621–1662, 2011.

R. Bhatia. *Matrix analysis*. Springer, 1997.

P. Billingsley. *Convergence of probability measures*. John Wiley & Sons Inc., New York, 1999.

A. Bloemendal. *Finite rank perturbations of random matrices and their continuum limits, PhD Thesis*. University of Toronto, 2011.

A. Bloemendal and B. Virág. Limits of spiked random matrices I. *Probability Theory and Related Fields*, 156:795–825, 2013.

A. Bloemendal and B. Virág. Limits of spiked random matrices II. *The Annals of Probability*, pages 2726–2769, 2016.

A. Bloemendal, A. Knowles, H.-T. Yau, and J. Yin. On the principal components of sample covariance matrices. *Probability theory and related fields*, 164(1-2):459–552, 2016.

F. Bornemann. On the numerical evaluation of distributions in random matrix theory: a review. *arXiv preprint arXiv:0904.1581*, 2009.

G. Borot and A. Guionnet. Asymptotic expansion of β matrix models in the one-cut regime. *Communications in Mathematical Physics*, 317:447–483, 2013.

P. Bourgade, K. Mody, and M. Pain. Optimal local law and central limit theorem for β-ensembles. *Communications in Mathematical Physics*, 390(3):1017–1079, 2022.

A. Bykhovskaya and V. Gorin. Cointegration in large VARs. *Annals of Statistics*, 50(3): 1593–1617, 2022.

A. Bykhovskaya and V. Gorin. Canonical correlation analysis: Review. *arXiv preprint arXiv:2411.15625*, 2024.

A. Bykhovskaya and V. Gorin. High-dimensional canonical correlation analysis. *arXiv preprint arXiv:2306.16393*, 2025.

A. Bykhovskaya, V. Gorin, and E. Kiss. Largevars: an R package for testing large VARs for the presence of cointegration. 2024. `https://github.com/eszter-kiss/Largevars`.

T. T. Cai, X. Han, and G. Pan. Limiting laws for divergent spiked eigenvalues and largest nonspiked eigenvalue of sample covariance matrices. *The Annals of Statistics*, 48(3):1255–1280, 2020.

M. Capitaine, C. Donati-Martin, and D. Féral. The largest eigenvalues of finite rank deformation of large Wigner matrices: Convergence and nonuniversality of the fluctuations. *The Annals of Probability*, pages 1–47, 2009.

M. Capitaine, C. Donati-Martin, and D. Féral. Central limit theorems for eigenvalues of deformations of Wigner matrices. In *Annales de l'IHP Probabilités et statistiques*, volume 48, pages 107–133, 2012.

A. Chudik, M. Pesaran, and E. Tosetti. Weak and strong cross-section dependence and estimation of large panels. *Econometrics Journal*, 14:C45–C90, 2011.

J. H. Cochrane. Presidential address: Discount rates. *The Journal of finance*, 66(4):1047–1108, 2011.

E. Collins-Woodfin and H. G. Le. An edge CLT for the log determinant of Laguerre beta ensembles. In *Annales de l'Institut Henri Poincare (B) Probabilites et statistiques*, volume 61, pages 83–128. Institut Henri Poincaré, 2025.

P. Deift and D. Gioev. *Random matrix theory: invariant ensembles and universality*. American Mathematical Soc., 2009.

X. Ding and F. Yang. A necessary and sufficient condition for edge universality at the largest singular values of covariance matrices. *The Annals of Applied Probability*, 28(3): 1679–1738, 2018.

E. Dobriban. Sharp detection in PCA under correlations: All eigenvalues matter. *The Annals of Statistics*, 45(4):1810–1833, 2017.

I. Dumitriu and A. Edelman. Matrix models for beta ensembles. *Journal of Mathematical Physics*, 43(11):5830–5847, 2002.

I. Dumitriu and E. Paquette. Global fluctuations for linear statistics of β-jacobi ensembles. *Random Matrices: Theory and Applications*, 1(04):1250013, 2012.

R. Durrett. *Probability: theory and examples, 5th edition*. Cambridge university press, 2019.

A. Edelman and P.-O. Persson. Numerical methods for eigenvalue distributions of random matrices. *arXiv preprint math-ph/0501068*, 2005.

A. El Alaoui, F. Krzakala, and M. Jordan. Fundamental limits of detection in the spiked Wigner model. *The Annals of Statistics*, 48(2):863–885, 2020.

L. Erdős and H.-T. Yau. *A dynamical approach to random matrix theory*. American Mathematical Soc., 2017.

J. Fan, Y. Yan, and Y. Zheng. When can weak latent factors be statistically inferred? *arXiv preprint arXiv:2407.03616*, 2024.

D. Féral and S. Péché. The largest eigenvalue of rank one deformation of large wigner matrices. *Communications in mathematical physics*, 272(1):185–228, 2007.

P. Forrester. *Log-gases and random matrices*. Princeton University Press, 2010.

P. J. Forrester. The spectrum edge of random matrix ensembles. *Nuclear Physics B*, 402(3): 709–728, 1993.

P. J. Forrester and E. M. Rains. Interrelationships between orthogonal, unitary and symplectic matrix ensembles. In *Random Matrix Models and Their Applications*, pages 171–207. Cambridge University Press, 2001.

S. Freyaldenhoven. Factor models with local factors—determining the number of relevant factors. *Journal of Econometrics*, 229(1):80–102, 2022.

Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1:233–241, 1981.

M. Gavish and D. L. Donoho. The optimal hard threshold for singular values is $4/\sqrt{3}$. *IEEE Transactions on Information Theory*, 60(8):5040–5053, 2014.

D. Giannone, M. Lenza, and G. E. Primiceri. Economic predictions with big data: The illusion of sparsity. *Econometrica*, 89(5):2409–2437, 2021.

S. Giglio, D. Xiu, and D. Zhang. Prediction when factors are weak. *Preprint*, 2023.

R. Gittins. *Canonical Analysis: A Review with Applications in Ecology*. Springer, 1985.

C. Han, G. Pan, and Q. Yang. A unified matrix model including both CCA and F matrices in multivariate analysis: The largest eigenvalue and its applications. *Bernoulli*, 24(4B): 3447–3468, 2018.

P. W. Holland, K. B. Laskey, and S. Leinhardt. Stochastic blockmodels: First steps. *Social networks*, 5(2):109–137, 1983.

Z. Hou, X. Zhang, Z. Bai, and J. Hu. Spiked eigenvalues of noncentral fisher matrix with applications. *Bernoulli*, 29(4):3171–3197, 2023.

J. Huang and L. Zhang. A convergence framework for Airy$_\beta$ line ensemble via pole evolution. *arXiv preprint arXiv:2411.10586*, 2024.

I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of Statistics*, 29(2):295–327, 2001.

I. M. Johnstone. Multivariate analysis and Jacobi ensembles: largest eigenvalue, Tracy-Widom limits and rates of convergence. *Annals of Statistics*, 36(6):2638–2716, 2008.

I. M. Johnstone and Z. Ma. Fast approach to the Tracy-Widom law at the edge of GOE and GUE. *The Annals of Applied Probability*, 22(5):1962, 2012.

I. M. Johnstone and A. Onatski. Testing in high-dimensional spiked models. *The Annals of Statistics*, 48(3):1231–1254, 2020.

I. M. Johnstone and D. Paul. PCA in high dimensions: An orientation. *Proceedings of the IEEE*, 106(8):1277–1292, 2018.

I. M. Johnstone, Y. Klochkov, A. Onatski, and D. Pavlyshyn. Spin glass to paramagnetic transition in spherical Sherrington-Kirkpatrick model with ferromagnetic interaction. *arXiv preprint arXiv:2104.07629*, 2021.

I. M. Johnstone, Y. Klochkov, A. Onatski, and D. Pavlyshyn. An edge CLT for the log determinant of Wigner ensembles. *Bernoulli*, 31(1):55–80, 2025.

R. Jones, J. Kosterlitz, and D. Thouless. The eigenvalue spectrum of a large symmetric random matrix with a finite mean. *Journal of Physics A: Mathematical and General*, 11 (3):L45, 1978.

Z. T. Ke, Y. Ma, and X. Lin. Estimation of the number of spiked eigenvalues in a covariance matrix by bulk eigenvalue matching analysis. *Journal of the American Statistical Association*, 118(541):374–392, 2023.

S. Kim, V. Raponi, and P. Zaffaroni. Testing for Weak Factors in Asset Pricing. *Preprint*, 2024.

A. Knowles and J. Yin. The isotropic semicircle law and deformation of Wigner matrices. *Communications on Pure and Applied Mathematics*, 66(11):1663–1749, 2013.

A. Knowles and J. Yin. The outliers of a deformed Wigner matrix. *The Annals of Probability*, 42(5):1980–2031, 2014.

S. Kritchman and B. Nadler. Non-parametric detection of the number of signals: Hypothesis testing and random matrix theory. *IEEE Transactions on Signal Processing*, 57(10):3930–3941, 2009.

P. Y. G. Lamarre and M. Shkolnikov. Edge of spiked beta ensembles, stochastic Airy semigroups and reflected Brownian motions. 55(3):1402–1438, 2019.

G. Lambert and E. Paquette. Strong approximation of Gaussian β–ensemble characteristic polynomials: the edge regime and the stochastic Airy function. *arXiv preprint arXiv:2009.05003*, 2020.

J. O. Lee and J. Yin. A necessary and sufficient condition for edge universality of Wigner matrices. *Duke Mathematical Journal*, 163(1), 2014.

M. Lettau and M. Pelger. Estimating latent asset-pricing factors. *Journal of Econometrics*, 218(1):1–31, 2020.

Z. Ma. Accuracy of the Tracy-Widom limits for the extreme eigenvalues in white Wishart matrices. *Bernoulli*, pages 322–359, 2012.

A. Mikusheva. Uniform inference in autoregressive models. *Econometrica*, 75(5):1411–1452, 2007.

M. Mo. Rank 1 real Wishart spiked model. *Communications on Pure and Applied Mathematics*, 65(11):1528–1638, 2012.

R. J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons, 2009.

A. Onatski. Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4):1004–1016, 2010.

A. Onatski. Asymptotics of the principal components estimator of large factor models with weakly influential factors. *Journal of Econometrics*, 168(2):244–258, 2012.

A. Onatski. Asymptotics of the principal components estimator of large factor models with weak factors and iid Gaussian noise. *Cambridge Working Papers in Economics, Cambridge-INET Working Paper Series*, 1808, 2018.

A. Onatski, M. J. Moreira, and M. Hallin. Asymptotic power of sphericity tests for high-dimensional data. *Annals of statistics*, 41(3):1204–1231, 2013.

A. Onatski, M. J. Moreira, and M. Hallin. Signal detection in high dimension: The multispiked case. *The Annals of Statistics*, 42(1):225–254, 2014.

S. O'Rourke. Gaussian fluctuations of eigenvalues in Wigner random matrices. *Journal of Statistical Physics*, 138(6):1045–1067, 2010.

L. Pastur and M. Shcherbina. *Eigenvalue distribution of large random matrices*. Number 171. American Mathematical Soc., 2011.

D. Paul. Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statistica Sinica*, pages 1617–1642, 2007.

M. H. Pesaran and R. P. Smith. Identifying and exploiting alpha in linear asset pricing models with strong, semi-strong, and latent factors. *Journal of Financial Econometrics*, 23(3):nbae029, 2025.

A. Pizzo, D. Renfrew, and A. Soshnikov. On finite rank deformations of wigner matrices. In *Annales de l'IHP Probabilités et statistiques*, volume 49, pages 64–94, 2013.

J. Ramirez, B. Rider, and B. Virág. Beta ensembles, stochastic Airy spectrum, and a diffusion. *Journal of the American Mathematical Society*, 24(4):919–944, 2011.

D. Renfrew and A. Soshnikov. On finite rank deformations of wigner matrices ii: Delocalized perturbations. *Random Matrices: Theory and Applications*, 2(01):1250015, 2013.

A. A. Shabalin and A. B. Nobel. Reconstruction of a low-rank matrix in the presence of gaussian noise. *Journal of Multivariate Analysis*, 118:67–76, 2013.

S. Sodin. On the critical points of random matrix characteristic polynomials and of the riemann ξ-function. *The Quarterly Journal of Mathematics*, 69(1):183–210, 2018.

A. Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *Journal of Statistical Physics*, 108:1033–1056, 2002.

A. B. Soshnikov. Gaussian fluctuation for the number of particles in Airy, Bessel, sine, and other determinantal random point fields. *Journal of Statistical Physics*, 100:491–522, 2000.

J. H. Stock. Confidence intervals for the largest autoregressive root in us macroeconomic time series. *Journal of monetary economics*, 28(3):435–459, 1991.

J. H. Stock and M. W. Watson. Forecasting using principal components from a large number of predictors. *Journal of the American statistical association*, 97(460):1167–1179, 2002.

T. Tao and V. Vu. Random matrices: the universality phenomenon for Wigner ensembles. *arXiv preprint arXiv:1202.0068*, 2012.

V. Thibeault, A. Allard, and P. Desrosiers. The low-rank hypothesis of complex systems. *Nature Physics*, 20(2):294–302, 2024.

B. Thompson. *Canonical correlation analysis: Uses and interpretation*. Number 47. Sage, 1984.

C. A. Tracy and H. Widom. On orthogonal and symplectic matrix ensembles. *Communications in Mathematical Physics*, 177:727–754, 1996.

M. Udell and A. Townsend. Why are big data matrices approximately low rank? *SIAM Journal on Mathematics of Data Science*, 1(1):144–160, 2019.

Y. Uematsu and T. Yamagata. Estimation of sparsity-induced weak factor models. *Journal of Business & Economic Statistics*, 41(1):213–227, 2022.

W. Wang and J. Fan. Asymptotics of empirical eigenstructure for high dimensional spiked covariance. *Annals of statistics*, 45(3):1342, 2017.

F. Yang. Sample canonical correlation coefficients of high-dimensional random vectors: Local law and Tracy-Widom limit. *Random Matrices: Theory and Applications*, 11(1):2250007, 2022a.

F. Yang. Limiting distribution of the sample canonical correlation coefficients of high-dimensional random vectors. *Electronic Journal of Probability*, 27:1–71, 2022b.

(Anna Bykhovskaya) Duke University
Email address: anna.bykhovskaya@duke.edu

(Vadim Gorin) University of California at Berkeley
Email address: vadicgor@gmail.com

(Sasha Sodin) The Hebrew University of Jerusalem and Queen Mary University of London
Email address: alexander.sodin@mail.huji.ac.il