LIKELIHOOD RATIO TESTING IN CRITICALLY-SPIKED WIGNER MODELS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF STATISTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Damian Pavlyshyn
May 2022

Preface

Spiked random matrix models are widely used to model data in which a low-rank signal exists alongside high-dimensional noise. When the eigenvalues of this signal, known as spikes, are either above or below a certain critical threshold, the corresponding models have been widely studied and are well understood.

However, the behavior of models with spikes at the critical threshold is more difficult to study, and has often remained elusive, despite the existence of data that is not well-explained by either sub- or supercritical models.

This thesis contains the results of series of projects that investigate the likelihood ratios for Gaussian models with critical spikes. It includes rigorous results illustrating the transition between the qualitatively different sub- and supercritical regimes, which has applications not only in statistics, but also in the statistical physics of the SSK model for magnetism. It also contains edge universality results, demonstrating how the quantities crucial for understanding the likelihood of critically-spiked Gaussian random matrices can be extended to their Wigner counterparts.

The thesis concludes with a presentation of results about phenomena occurring exactly at the critical threshold, determining the contiguous set of alternatives for testing a critical spike and providing the first description of the limiting likelihood ratio for such tests.

Acknowledgments

I first want to express my infinite gratitude to my advisor and role model, Iain Johnstone. His thoughtfulness and deep insights inspired me to always seek out and develop understanding of newer and harder problems, and his kindness and humility gave me the support and confidence needed to solve them. Whether to work or simply to chat, it was always a privilege and a pleasure to spend time with him.

I would also like to thank my long-time collaborators Alexei Onatski and Yegor Klochkov. Throughout our collaboration, they were an ongoing and invaluable source of ideas, suggestions and commentary, and discussions with them fueled much of my progress.

This thesis, and my Ph. D. in general would not have been possible without the tireless work of the Sequoia Hall student services manager, Susie Ementon. An email from her was the first contact that I had with the Stanford statistics department, and since then she has been an unwavering pillar of support. Her door was always open for long conversations about anything on my mind, and I could always be confident that she would resolve any issue or complication with incredible efficiency, and most likely before I even became aware of it.

The statisticians that shaped me most in the last five years have been my fellow students. I would especially like to thank my cohort — Chenyang, Elena, Ismael, Maxime, Michael, Shuangning, Sky, Sohom, Suyash, Swarnadip, Theo and Zijun — through our community and friendship, we somehow figured this crazy process out together. I will also miss my office, which I shared with Ben, Jaime, Kenneth and Kevin, for the fun and/or thought-provoking discussions that we had there.

My greatest source of strength throughout my studies has always been my family. Although I have been separated from them, whether by disease or war, for years, the love and faith in me that they sent from great distances has kept me going. I owe everything to my parents, Olya and Marko, who have encouraged, taught and supported me throughout my life and in every one of my endeavors. Я також невимовно вдячний своїй родині в Україні. Вони постійно про мене дбали, цікавилися мною і передавали свою підтримку навіть тоді, коли самі були у дуже скрутних ситуаціях.

I shared nearly day of these last five years with my partner, Julia, and it is hard to imagine how I could have kept going without her ongoing and tireless comfort and advice. I am incredibly thankful to have met her and to have spent this experience with her.

Contents

Chapter 1

Introduction

1.1 Gaussian matrix ensembles

Gaussian matrices, which are symmetric or Hermitian matrices with iid Gaussian entries, are the simplest examples of random matrix ensembles. Because of this, they are extensively studied, and most familiar random matrix results were established first for Gaussian matrices, and then extended to other ensembles.

The eigenvalues of Gaussian matrices behave similarly to the eigenvalues of random matrices with more direct applications in statistics, such as sample covariance matrices for Gaussian data, which follow the Wishart distribution. This similarity, stemming from the similarity between the joint densities of eigenvalues in the two models, makes the study of Gaussian matrices useful not only in its own right, but also as a preliminary step for establishing results about the eigenvalues of other kinds of random matrix ensembles.

Indeed, for questions of statistical testing, whose resolution invariably require a detailed understanding of the model likelihood, this analogy between Gaussian matrices and other random matrix ensembles is particularly useful. In treatments such as [JO20; DJO18], analyses of large classes of random matrix ensembles have followed the basic pattern of the corresponding Gaussian analysis.

In addition to their use as a simple proxy for investigating other random matrix ensembles, Gaussian matrices have explicit uses in statistical physics. It is common for statistical physics models to be defined in terms of matrices of independent $\{\pm 1\}$-valued random variables, whose spectral properties are well-approximated by Gaussians. An example of such an approximation, which will be examined in this thesis, is the Spherical Sherrington-Kirkpatrick (SSK) model, which is a Gaussian approximation of the Sherrington-Kirkpatrick (SK) model that retains many of the important features of the SK model while being substantially easier to analyze using Gaussian matrix results.

1.1.1 Definitions of GUE and GOE.

We recall here the definitions of the GUE and GOE. For $1 \leq i \leq j \leq N$, let ξ_{ij}, η_{ij} be independent $\mathcal{N}(0,1)$ random variables. Then define a Hermitian matrix Z_1 with entries

$$Z_{1,ij} = \begin{cases} \xi_{ij} & \text{if } i = j, \\ \frac{1}{\sqrt{2}}(\xi_{ij} + \mathrm{i}\eta_{ij}) & \text{if } i < j, \\ \overline{Z_{1,ji}} & \text{if } i > j. \end{cases}$$

Similarly, define a symmetric matrix Z_2 by

$$Z_{2,ij} = \begin{cases} \sqrt{2}\xi_{ii} & \text{if } i = j, \\ \xi_{ij} & \text{if } i < j \\ Z_{2,ji} & \text{if } i > j. \end{cases}$$

For $\alpha \in \{1, 2\}$, we call the distribution of Z_α the Gaussian Unitary Ensemble (GUE) if $\alpha = 1$, and the Gaussian Orthogonal Ensemble (GOE) if $\alpha = 2$.

The parameter $2/\alpha$, typically denoted by β, indexes the number field of which the matrix entries are elements, and is called the *Dyson parameter*. Although the Dyson parameterization is commonly used in the literature, for the results presented in this thesis, it is more convenient to instead use the parameter α, sometimes known as the *Jack parameter* [see, e.g. Ful04], and we will use this notation throughout.

We will also use the term *scaled Gaussian matrix* to refer to $Z_\alpha/\sqrt{N}$, where Z_α is an $N \times N$ Gaussian matrix.

1.1.2 Spiked models

An extension of the Gaussian random matrix model that has been extensively studied in the statistics literature is the rank-one deformation of a Gaussian matrix Z_N given by

$$W_N := \frac{Z_N}{\sqrt{N}} + hvv^*, \tag{1.1}$$

where $\|v\|_2 = 1$. Such distributions, parameterized by the value h, called the spike, are known as *spiked ensembles*. Since the distribution of the eigenvalues of Z_N is rotationally invariant, the direction v has no effect on the distribution of W_N, and so is commonly taken to be the standard basis vector e_1, in which case W_N differs from $Z_N/\sqrt{N}$ only in its $(1,1)$ entry.

Spiked ensembles were first studied in a statistical context in [Joh01], where the spike parameterized a rank-one deviation from an identity covariance in a Wishart ensemble, which arise, for example, in the study of high-dimensional covariance estimation and principal components analysis. Eigenvalues of matrix ensembles of the form of Eq. (1.1) display similar limiting behavior to that of spiked Wishart ensembles, so it is useful to extend the spike terminology to these "deformed Gaussian" ensembles.

Many spiked random matrix ensembles exhibit a phase transition in the limiting behavior of the largest eigenvalue. Specifically, spikes h that are strictly smaller than a certain critical threshold h_c do not influence the distribution of the largest eigenvalue, and are known as subcritical. In ensembles with supercritical spikes $h > h_c$, the largest eigenvalue is separated from the bulk and converges almost surely to a value strictly above the support of the eigenvalue bulk.

Such a phase transition was originally identified for spiked Wishart ensembles in [BBP05], but has since been described in other spiked models. Relevant to the Gaussian ensembles that are investigate in this thesis, the critical threshold was shown to be $h_c = 1$ for spiked G(U/O)E in [Péc06] and [Mai07] respectively.

In the supercritical case, that is, when $h > 1$, the largest eigenvalue of a scaled G(U/O)E matrix was found in [CDF09] to have Gaussian fluctuations given by

$$N^{1/2}(\lambda_1 - (h + h^{-1})) \xrightarrow{\mathrm{d}} \mathcal{N}(0, \alpha(1 - h^{-2})).$$

On the other hand, [Péc06; Mai07] showed that in the subcritical case, when $h < 1$, the largest eigenvalue has Tracy-Widom fluctuations around 2. In particular,

$$N^{2/3}(\lambda_1 - 2) \xrightarrow{\mathrm{d}} \mathrm{TW}_{2/\alpha}\,.$$

Moreover, [Péc06] described the limiting distribution of the largest eigenvalue for $h = 1$ in terms of a Fredholm determinant. In [BV13], Bloemendal and Viràg considered the largest eigenvalues of Gaussian matrices with critical spike $h = 1 + b_0 N^{-1/3}$, describing the corresponding one-parameter family of limiting distributions. Throughout this thesis, we will refer to this limiting distribution as the BV, or Bloemendal-Viràg distribution so that, according to [BV13, Theorem 1.5],

$$N^{-2/3}(2 - \lambda_1^{(b_0)}) \xrightarrow{\mathrm{d}} \mathrm{BV}(b_0)_{2/\alpha}, \tag{1.2}$$

where $\lambda_1^{(b_0)}$ is the largest eigenvalue of a scaled Gaussian matrix with Dyson parameter $2/\alpha$ and critical spike $1 + b_0 N^{-1/3}$.

1.2 Testing and local alternatives in spiked models

For a sequence of models $\{\mathbf{P}_{N,h}\}$ parameterized by h and with null hypotheses

$$H_{N,0} \colon h = h_{N,0},$$

some sequences of possible alternative hypotheses are asymptotically impossible to distinguish from the null while others can be distinguished with probability approaching 1.

The threshold between these two regimes is a set of sequences of alternatives

$$H_{N,1} \colon h = h_{N,1} := h_{N,0} + \theta a_N$$

indexed by θ, where (a_N) is a sequence chosen such that the corresponding sequence of likelihood ratios converges to a non-degenerate limit. In this way, the local alternatives $h_{N,1}$ describe the alternatives that can be tested for with power asymptotically falling strictly between 0 and 1.

This notion was first formalized with the name "contiguity" in [Le 60]. A sequence of simple experiments was defined as contiguous if the supports of the null and alternative distributions asymptotically coincided. The connection between contiguity and the limiting behavior of the likelihood ratio was codified in the celebrated result known as Le Cam's first lemma, and it is this result that will be used in this thesis to establish contiguity.

1.2.1 Likelihood of spiked Gaussians

In context of spiked Gaussian matrices, we let $\mathbf{P}_{N,h}$ be the distribution of a spiked Gaussian matrix as described in Eq. (1.1).

Due to the rotational symmetry of W_N, it is enough for the purposes of testing h to understand the likelihood of the eigenvalues of W_N. For a large family of random matrix distributions, the joint density of the eigenvalues Λ of W_N follow a common pattern, which was extensively cataloged in [Jam64]. In the spiked Gaussian case, the corresponding representation was shown in [JO20, Lemma 14 (supplementary material)] to be

$$p_N(\Lambda; h) = c(\Lambda) \cdot e^{\frac{2N}{\alpha} \cdot \frac{h^2}{4}} \cdot \int_{S_\alpha^{N-1}} \exp\Big\{ \frac{N}{\alpha} h \cdot u^* \Lambda u \Big\} (\mathrm{d}u), \tag{1.3}$$

where $c(\Lambda)$ is a function of the eigenvalues only, and where the integral is taken over the $(N-1)$-sphere with respect to the normalised Haar measure.

As discussed in [Ona14] and cataloged for a much larger family of spiked models in [JO20], the joint density of eigenvalues Λ of a spiked Gaussian ensemble with a spike of size h is

$$p_N(\Lambda; h) = c(\Lambda) \cdot e^{\frac{2N}{\alpha} \cdot \frac{h^2}{4}} \cdot \frac{C_N}{2\pi \mathrm{i}} \int_{\mathcal{K}} \exp\Big\{ \frac{N}{\alpha} \Big[hz - \frac{1}{N} \sum_{j=1}^{N} \log(z - \lambda_j) \Big] \Big\} \,\mathrm{d}z \tag{1.4}$$

for a constant C_N and where $\mathcal{K}$ is a contour running from $-\mathrm{i}\infty$ to $+\mathrm{i}\infty$ and passing to the right of all the eigenvalues Λ.

1.2.2 Contiguity in spiked random matrix models

In classical statistical models, it is common for contiguous alternatives to be separated by an $O(N^{-1/2})$ gap. For Gaussian matrix models of the form Eq. (1.1), this is not necessarily so, and in the papers [JO20; DJO18] the authors carried out a program of identifying contiguous alternatives or a large number of spiked random matrix models including the Gaussian case.

The paper [JO20] investigated tests of subcritical spikes. In the spiked GOE case, this amounted to testing, for $\theta \in (0, 1)$.

$$H_{N,0}\colon h = 0 \qquad \text{vs.} \qquad H_{N,1}\colon h = \theta. \tag{1.5}$$

The likelihood ratio for this experiment was shown to asymptotically depend on the eigenvalues only through the function

$$\sum_{j=1}^{N} \log(z_0(\theta) - \lambda_j), \tag{1.6}$$

where $z_0(\theta) = \theta + \theta^{-1}$.

Objects like that in Eq. (1.6), known as logarithmic linear statistics, appear often in the analysis of random matrix likelihoods due to the appearance of a similar quantity in the integrand of Eq. (1.4). When z_0 is strictly greater than 2, the spectral central limit theorem established for Gaussian matrices in [BY05], implies that the log-likelihood ratio for Eq. (1.5) is asymptotically Gaussian.

The likelihood ratio in the supercritical case was analyzed in an as-yet unpublished extension of [DJO18]. In this case, for a supercritical null spike $h_0 > 1$, the contiguous experiments were shown to have hypotheses

$$H_{N,0} \colon h = h_0 \qquad \text{vs.} \qquad H_{N,1} \colon h = h_0 + \theta N^{-1/2}.$$

In this case, the likelihood ratio depends asymptotically only on the largest eigenvalue, which in the supercritical case is asymptotically Gaussian.

The "vanishingly supercritical" case, has a null spike of $h_0 = 1 + b_0 N^{-\alpha}$ for $\alpha < 1/3$ and $b_0 > 0$. This sequence of experiments can be analysed with the same general techniques as in the supercritical case, while in some sense approaching the critical regime at $\alpha = 1/3$. Indeed, the appropriate vanishingly supercritical contiguous experiments are

$$H_{N,0} \colon h = h_0 := 1 + b_0 N^{-\alpha} \qquad \text{vs.} \qquad H_{N,1} \colon h = h_0 + \theta N^{(\alpha-1)/2}.$$

Naively, by plugging $\alpha = 1/3$ into the above, one might expect that contiguous local alternative in the critical case should be $h_1 = h_0 + \theta N^{-1/3}$. Demonstrating this fact will be a goal of this thesis, and only after establishing substantial tools for analysing likelihood ratios of Gaussian matrices in the first few chapters will such contiguity be proved on the critical scale in Chapter 4.

1.3 Log-determinant Central limit theorems

As mentioned in the previous section, logarithmic linear statistics — that is, functions of the form

$$\mathcal{L}_N = \sum_{j=1}^{N} \log(E - \lambda_j)$$

arise often in the study of random matrix likelihoods. Indeed, understanding these quantities is essential for establishing the likelihood ratio results which are the focus of this thesis.

When $E > 2$ is fixed, the limiting distribution of $\mathcal{L}_N$ is well-known, following from the spectral central limit theorem of [BY05]. However, for the results that will subsequently discussed in this thesis, we require Gaussian behavior near, at, or just inside the edge

$$E = E_N = 2 + \sigma_N N^{-2/3}, \qquad -\gamma \leq \sigma_N \ll \log^2 N \tag{1.7}$$

for some fixed $\gamma > 0$.

Here E is sufficiently close to the edge that the functions $f_N(z) = \log|z - E|$ do not appear to be covered even by recent mesoscopic CLTs (e.g. [LS19; LSX20]).

The basic identity

$$L_N = \sum_{j=1}^{N} \log|\lambda_j - E| = \log|\det(W_N - E)|$$

casts the linear statistic (now with the absolute value under the logarithm) as a log determinant, i.e. in terms of the characteristic polynomial of W_N. The latter is the subject of a substantial literature. In particular, as pioneered by Tao and Vu [TV12] for $E = 0$, for Gaussian ensembles W_N drawn from GUE or GOE, one can use the Trotter equivalence to cast the matrix in tridiagonal Jacobi form and analyze the recurrence satisfied by the principal minors.

Chapter 2 contains a summary of the result of [JKOP20], in which this program is carried out at the edge of the spectrum to arrive at the following result.

Theorem 1.1. *Let W_N be a Gaussian matrix whose off-diagonal moments match GUE ($\alpha = 1$) or GOE ($\alpha = 2$) to third order. For edge values $E = E_N$ satisfying Eq.* (2.1)*, we* *have*

$$(\log|\det(W_N - E)| - \mu_N)/\tau_N \xrightarrow{\mathrm{d}} \mathcal{N}(0,1), \tag{1.8}$$

with

$$\mu_N = \frac{1}{2}N + \sigma_N N^{1/3} - \tfrac{2}{3}\sigma_N^{3/2} - \frac{1}{6}(\alpha - 1)\log N, \qquad \tau_N = \sqrt{\frac{\alpha}{3}\log N}. \tag{1.9}$$

1.4 Eigenvalue statistic universality

An important generalization of Gaussian matrix ensembles is Wigner matrix ensembles. Informally, these are symmetric or Hermitian matrices whose entries are independent and subject to certain distributional restrictions. The first investigation of such matrices was performed in [Wig58], which proved the semicircle law for Wigner matrices whose entries have symmetric, variance-one distributions.

Subsequent authors have used the term "Wigner" to refer to similar matrices with varying technical conditions on the entry distributions. Throughout this thesis, we will use the following definition:

Definition 1.2 (Wigner matrix)**.** A Wigner matrix is an $N \times N$ matrix $W_N = [\xi_{ij}/\sqrt{N}]_{ij}$ which is either Hermitian ($\alpha = 1$) or symmetric ($\alpha = 2$) and for which the components $\{\operatorname{Re}\xi_{ij}, \operatorname{Im}\xi_{ij}\}_{i<j}$ and $\{\xi_{ii}\}$ are independent random variables with mean zero and satisfy some of the following conditions:

W1 $\mathbf{E}|\xi_{ij}|^2 = 1$ for $i \neq j$ and $\mathbf{E}\xi_{ii}^2 \leq B$ for some absolute constant B;

W2 In the Hermitian case, $\mathbf{E}\xi_{ij}^2 = 0$ for $i \neq j$;

W3 The moments of $\operatorname{Im}\xi_{ij}$, $\operatorname{Re}\xi_{ij}$ are bounded uniformly in N. That is, for all $p \in \mathbf{Z}_{>0}$, there is a constant C_p such that

$$\mathbf{E}|\operatorname{Im}\xi_{ij}|^p, \mathbf{E}|\operatorname{Re}\xi_{ij}|^p \leq C_p;$$

W4 $\mathbf{E}(\operatorname{Re}\xi_{ij})^3 = \mathbf{E}(\operatorname{Im}\xi_{ij})^3 = 0$ for $i \neq j$.

As in the Gaussian case, we refer to $W_N + hvv^*$, where $\|v\|_2 = 1$, as a spiked Wigner matrix with a spike of h.

In the series of papers [TV10; TV11a; TV11b; TV12], Tao and Vu developed a number of so-called "four-moment theorems." These concluded that the limiting distributions of

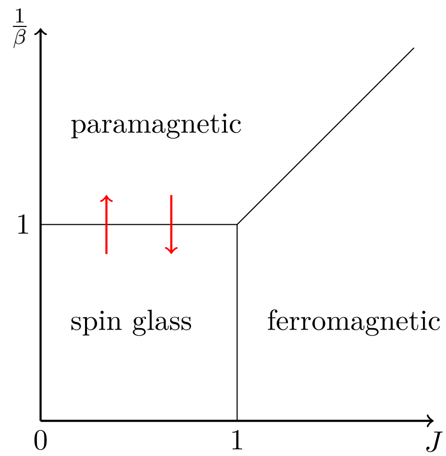


Figure 1.1: Phase diagram showing the Spin glass, Paramagnetic, and Ferromagnetic regimes. Red arrows indicate the transition between Spin glass and Paramagnetic regimes, which is the focus of this paper.

various eigenvalue statistics were the same when evaluated at the eigenvalues of a Gaussian matrix as with those of Wigner matrices whose entries match those of a Gaussian matrix up to order four.

It was under this four-moment assumption that, the universality of the log-determinant $\log|\det W_N|$ was established in [TV12]. Another important development was in [EKYY12; KY13a], in which the authors established two-moments theorem for certain eigenvalue statistics near the edge. These took advantage of the fact that scaled Gaussian eigenvalues are separated from each other by distance of $O(N^{-2/3})$ at the edge rather than $O(N^{-1})$ as in the bulk of the spectrum. This sparsity allowed a reduction in the required number of matching moments.

The main universality result presented in this thesis is Proposition 2.14 and takes the form of a three-moment theorem at the edge for a family of eigenvalue statistics that can be approximated by an integral of the Stieltjes transform. Among these statistics is the log-determinant, which allows us to extend the log-determinant CLT of Theorem 1.1 to the Wigner case. In addition to establishing marginal results, this treatment also includes a method for establishing the joint universality of these statistics.

1.5 The SSK model

In addition to its statistical significance, the integral term in Eq. (1.3) has an important physical interpretation. The Spherical Sherrington-Kirkpatrick model with Curie-Weiss ferromagnetic interaction is a model of magnetism with Hamiltonian

$$H_N(\sigma) = \frac{1}{2}\Big(\frac{1}{\sqrt{N}}\sum_{i,j=1}^{N} A_{ij}\sigma_i\sigma_j + \frac{J}{N}\sum_{i,j=1}^{N}\sigma_i\sigma_j\Big), \tag{1.10}$$

where $\sigma \in \sqrt{N}\mathcal{S}_2^{N-1}$, $J \geq 0$ is known as the coupling constant, and A is a real symmetric $N \times N$ matrix with zeroes on the diagonal and independent upper triangular entries A_{ij} with mean zero and variance 1. It was introduced in [KTJ76] as a tractable variant of the original Sherrington-Kirkpatrick model that has discrete spins $\sigma \in \{\pm 1\}^N$.

[KTJ76] considered Gaussian A_{ij}, but since that paper this assumption has been substantially weakened. For example, the papers [BL16; BL17; BLW18] considered general Wigner matrices A. Nevertheless, the Gaussian model remains the benchmark. In fact, [BDG01] called the model with A from GOE "the standard SSK model" (see also [Tal06] and [PT07]).

The free energy of the Wigner SSK model is then closely related to the likelihood Eq. (1.3) with $\alpha = 2$. Precisely, let W_N be a Wigner matrix with a spike of J and zeroes on the diagonal. Letting Λ be the diagonal matrix of eigenvalues of W_N, write

$$\begin{aligned} Z_{\alpha,N} &= \int_{S_\alpha^{N-1}} \exp\Big\{\frac{N}{\alpha} h \cdot u^*\Lambda u\Big\}(\mathrm{d}u), \\ F_{\alpha,N} &= \frac{\alpha}{2N}\log Z_{\alpha,N} \end{aligned} \tag{1.11}$$

The papers [BL16; BL17; BLW18] made a thorough study of the fluctuations of the free energy $F_{2,N}$ in the spin glass, para- and ferro-magnetic regimes.[1] The fluctuations of the free energy in the three regimes are shown to be

1. (Spin glass) If $\beta > 1$ and $J < 1$, then

$$\frac{2N^{2/3}}{\beta - 1}(F_{2,N} - F(\beta)) \xrightarrow{\mathrm{d}} \mathrm{TW}_1\,.$$

[1] In these papers, the parameterization is slightly different from the one presented here, so that in their case, the critical threshold is $\beta = 1/2$ instead of $\beta = 1$.

2. (Paramagnetic) If $\beta < 1$ and $\beta < 1/J$, then

$$N(F_{2,N} - F(\beta)) \xrightarrow{\mathrm{d}} \mathcal{N}(f_1, a_1),$$

where a_1 depends on β but not on J, while f_1 depends on both β and J.

3. (Ferromagnetic) If $J > 1$ and $\beta > 1/J$, then

$$N^{1/2}(F_{2,N} - F(\beta)) \xrightarrow{\mathrm{d}} \mathcal{N}(0, a_2),$$

where a_2 depends on β.

The leading order term $F(\beta)$ differs across the regimes:

$$F(\beta) = \begin{cases} \beta - \dfrac{1}{2}\log\beta - \dfrac{3}{4} & \text{for spin glass} \\ \dfrac{1}{4}\beta^2 & \text{for paramagnetic} \\ \dfrac{\beta}{2}(J + J^{-1}) - \dfrac{1}{2}\log(\beta J) - \dfrac{1}{4}J^{-2} - \dfrac{1}{2} & \text{for ferromagnetic.} \end{cases} \tag{1.12}$$

This result characterizes the fluctuations of $F_{2,N}$ in models lying strictly within the three regimes. The results for the transitions studied in [BL17] and [BLW18] are summarized in the following table.

Transition	Transition window	Fluctuations of $F_{2,N}$
Spin glass - Ferromagnetic	$J = 1 + b_0 N^{-1/3}$, $\beta > 1$	$N^{2/3}(F_{2,N} - F(\beta)) \xrightarrow{d} \frac{\beta-1}{2}\,\mathrm{BV}(-b_0)_1$
Paramagnetic - Ferromagnetic	$\beta = \frac{1}{J} + BN^{-1/2}$, $J > 1$	$N(F_{2,N} - F(\beta)) \xrightarrow{d} G_1 + Q_B(G_2)$.

Here (G_1, G_2) has a bivariate Gaussian distribution that depends on J but not on B, and Q_B is a non-linear function that depends both on J and B.

Concerning the remaining transition between the spin glass and the paramagnetic regimes, [BL16] and [BL17] conjectureed that the critical window of temperatures for this transition is $\beta = 1 + O(N^{-1/3}\sqrt{\log N})$ for any $J < 1$. They arrived at this conjecture by matching the orders of the variance of $F_{2,N}$ as $\beta \to 1$ from above and below.

The discussion in Chapter 3 contains results from [JKOP21] that confirm this conjecture.

Namely, if

$$\beta = 1 + bN^{-1/3}\sqrt{\log N}, \qquad 0 \le J < 1,$$

then $F_{\alpha,N}$ has fluctuations of order $N/\sqrt{\log N}$. Moreover, as b increases from $-\infty$ to ∞, we describe the transition of the limiting distribution of $F_{\alpha,N}$ from Gaussian to the Tracy-Widom.

Precisely, the main result is as follows.

Theorem 1.3. *Consider $F_{\alpha,N}$ with $\alpha = 1$ or $\alpha = 2$, as defined in Eq.* (1.11)*. Let $\beta = 1 + bN^{-1/3}\sqrt{\log N}$ for a constant $b \in \mathbf{R}$ and let $0 \le J < 1$. Further let $b_+ = \max\{0, b\}$ be the positive part of b. Then*

$$\frac{N}{\sqrt{\frac{\alpha}{12}\log N}}\left(F_{\alpha,N} - F(\beta) + \frac{\log N}{12N}\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1) + \sqrt{\frac{3}{\alpha}}b_+\mathrm{TW}_{2/\alpha}, \tag{1.13}$$

where TW_2 and TW_1 are the complex and real Tracy-Widom distributions, respectively, independent from the $\mathcal{N}(0,1)$, and where $F(\beta)$ is as in Eq. (1.12)*, that is*

$$F(\beta) = \begin{cases} \beta - \frac{1}{2}\log\beta - \frac{3}{4} & \text{if } b \ge 0, \\ \frac{1}{4}\beta^2 & \text{if } b < 0. \end{cases}$$

1.5.1 Relevance to statistical testing

Recalling Eq. (1.11) together with Eq. (1.3), we find that for $J = 0$, $F_{\alpha,N}$ is distributed as the scaled log-likelihood ratio for testing

$$H_0\colon h = 0 \qquad \text{vs.} \qquad H_1\colon h = \beta$$

in the spiked Gaussian model, under the null hypothesis. Specifically, for $\beta \le 1$

$$\log\frac{p_N(\Lambda;\beta)}{p_N(\Lambda;0)} = \frac{2N}{\alpha}[F_{\alpha,N} - F(\beta)].$$

Theorem 1.3 therefore gives the limiting behavior of the null distribution of the likelihood ratio. The mean shift and variance, both growing on order $\log N$, verify that the null and alternative distributions fail to be contiguous, and so we cannot directly obtain the limiting distribution of the likelihood ratio under alternative hypotheses β near 1.

1.6 The stochastic Airy function

The spectral analysis of certain stochastic differential operators can be a powerful tool for the understanding of eigenvalues of random matrices. Introduced in [ES07] and studied rigorously in [RRV11], the operator relevant to the edge eigenvalues of Gaussian matrices is the so-called stochastic Airy operator

$$\mathcal{H}_{2/\alpha} = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + x + \sqrt{2\alpha}B'(x),$$

where B' is the distributional derivative of a standard Brownian motion. Here, $\mathcal{H}_{2/\alpha}$ acts on $L^2([0,\infty))$ functions f satisfying the Dirichlet initial condition $f(0) = 0$.

The random operator $\mathcal{H}_{2/\alpha}$ was shown in [ES07; RRV11] to have a simple, lower-bounded spectrum $\Lambda_0 < \Lambda_1 < \cdots$, and that for any $k \in \mathbf{Z}_{>0}$, the eigenvalue distributions of a scaled $N \times N$ Gaussian matrix W_N satisfy

$$(N^{-2/3}(\lambda_j - 2))_{1\le j\le k} \xrightarrow{\mathrm{d}} (-\Lambda_{j-1})_{1\le j\le k}.$$

This analysis was extended to spiked matrices in [BV13]. In the case of Gaussian matrices with a critical spike of size $1 + b_0N^{-1/3}$, the largest eigenvalues converge in distribution to the eigenvalues of $\mathcal{H}_{2/\alpha}$ acting on functions satisfying the Robin initial condition $f'(0) = -b_0f(0)$.

In [LP21], Lambert and Paquette made a significant development in the analysis of the stochastic Airy operator. They studied solutions $\phi_\lambda \in L^2([0,\infty))$ of $-\mathcal{H}_{2/\alpha}\phi_\lambda = \lambda\phi_\lambda$ for each $\lambda \in \mathbf{C}$, keeping track of the dependence of these solutions on their behavior at 0. Lambert and Paquette phrased this problem in terms of the "Stochastic Airy Equation," which was the SDE

$$\mathrm{d}\phi'_\lambda(t) = (t+\lambda)\phi_\lambda(t)\,\mathrm{d}t + \phi_\lambda(t)\sqrt{2\alpha}\,\mathrm{d}B(t),$$

showing that this equation has a unique (up to a constant multiple) solution in $L^2([0,\infty))$ and called this solution the "Stochastic Airy Function," denoted by SAi_λ.

In this way, the results of [ES07; RRV11; LP21] imply that the spectrum of $-\mathcal{H}_{2/\alpha}$, and so the limiting distribution of the largest eigenvalues of a Gaussian matrix, is given by

$$\{\lambda \in \mathbf{C} : \mathrm{SAi}_\lambda(0) = 0\}.$$

In fact, the main result of [LP21] is substantially stronger, stating that, if $\varphi_N(z) = \prod_{j=1}^N (z - \lambda_j)$ is the characteristic polynomial of W_N, then, in the sense of uniform convergence in compact sets,

$$\Big(w_N(1 + \lambda N^{-2/3}/2)\varphi_N(2 + \lambda N^{-2/3})\frac{\mathbf{E}e^{G_N}}{e^{G_N}} : \lambda \in \mathbf{R}\Big) \xrightarrow{\mathrm{d}} (\mathrm{SAi}_\lambda(0) : \lambda \in \mathbf{R}), \tag{1.14}$$

where G_N is a centered Gaussian random variable with $\mathbf{E}G_N^2 = \frac{\alpha}{3}\log N + O(1)$, and where w_N is the weight function

$$w_N(z) = \Big((2\pi)^{1/4} e^{Nz^2} 2^{-N} (Nz^2)^{-1/2} \sqrt{\frac{N!}{N^N}}\Big)^{-1}.$$

This result is closely connected with the classical Plancherel-Rotach asymptotics established in [PR29], which state that

$$(w_N(1 + \lambda N^{-2/3})\pi_N(1 + \lambda N^{-2/3}) : \lambda \in \mathbf{C}) \to (\mathrm{Ai}(\lambda) : \lambda \in \mathbf{C}), \tag{1.15}$$

where π_N are the orthogonal Hermite polynomials with respect to the measure $e^{-2Nx^2}\,\mathrm{d}x$.

The convergence of Eq. (1.14) over the whole complex plane was conjectured in [LP21], and subsequently established in [Ass22] though a complex analytic argument.

In addition to the theoretical results of this thesis that make use of the stochastic Airy function, Section 5.2 contains the descriptions of algorithms for efficiently solving the stochastic Airy equations as a function of t or of λ.

1.6.1 The critically-spiked case

The results of [BV13] suggest that the limiting distribution of the largest eigenvalues of a scaled Gaussian matrix $W_N^{(b_0)}$ with critical spike $1 + b_0 N^{-1/3}$ should be given by

$$\{\lambda \in \mathbf{C} : \mathrm{SAi}'_\lambda(0) = -b_0\,\mathrm{SAi}_\lambda(0)\}.$$

In Chapter 4, we establish this rigorously, through an extension of Eq. (1.15). Namely, if $\varphi_N^{(b_0)}$ is the characteristic polynomial of $W_N^{(b_0)}$, then

$$\Big(N^{1/3} w_N (1+\lambda N^{-2/3}/2) \varphi_N^{(b_0)}(2+\lambda N^{-2/3}) \frac{\mathbf{E} e^{G_N}}{e^{G_N}} : \lambda \in \mathbf{C}\Big) \xrightarrow{\mathrm{d}} (-b_0 \,\mathrm{SAi}_\lambda(0) - \mathrm{SAi}'_\lambda(0) : \lambda \in \mathbf{C}). \quad (1.16)$$

This convergence allows us to extend many theorems proved for subcritical Gaussian matrices in Chapters 2 and 3 to their critical equivalent. Relatively straightforward is the following central limit theorem, which is the equivalent to Theorem 1.1 in the case that both the spike and E are exactly on the critical scale:

Proposition 1.4. *Let $\lambda_1^{(b_0)} \geq \cdots \geq \lambda_N^{(b_0)}$ be the eigenvalues of a scaled Gaussian matrix with Dyson parameter $2/\alpha$ and critical spike $1 + b_0 N^{-1/3}$. Let $\gamma = 2 + \lambda N^{-2/3}$ for some $\lambda \in \mathbf{R}$. Then*

$$\frac{\sum_{j=1}^N \log|\gamma - \lambda_j^{(b)}| - \frac{N}{2} - N^{1/3} C + \frac{\alpha+1}{6} \log N}{\sqrt{\frac{\alpha}{3}\log N}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

Another important result established in Chapter 4 is a partial description of the limiting behaviour the log-partition function of the SSK model near the triple point.

Theorem 1.5. *Consider $F_{\alpha,N}$ with $\alpha = 1$ or $\alpha = 2$. Let $\beta = 1 + bN^{-1/3}\sqrt{\log N}$ for a constant $b \geq 0$ and let $J = 1 + b_0 N^{-1/3}$ for a constant $b_0 \in \mathbf{R}$.*

Then

$$\frac{N}{\sqrt{\frac{\alpha}{12}\log N}} \Big(F_{\alpha,N} - F(\beta) - \frac{\log N}{12N}\Big) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1) + \sqrt{\frac{3}{\alpha}} b \cdot \mathrm{BV}(-b_0)_{2/\alpha},$$

where $\mathrm{BV}(-b_0)_2$ and $\mathrm{BV}(-b_0)_1$ are the complex and real Bloemendal-Viràg distributions respectively, independent from the $\mathcal{N}(0,1)$, and where

$$F(\beta) = \beta - \frac{1}{2}\log\beta - \frac{3}{4}.$$

This result effectively describes half of the critical point of the SSK model, as the half-plane of parameters $\{(b_0, b) : b_0 \in \mathbf{R}, b < 0\}$ is not yet covered.

Finally, and most importantly, the stochastic Airy machinery will allow us to establish a representation of the limiting likelihood ratio for testing for the difference between two different spiked Gaussian matrices.

In particular, we will demonstrate the following theorem, which is phrased in terms of the quantity $s_{b_0}^{(\alpha)}$. This informally represents the holomorphic αth root of $-b_0\,\mathrm{SAi}_\lambda(0) - \mathrm{SAi}'_\lambda(0)$, and is defined rigorously in Section 4.3.4.

Theorem 1.6. *Let $\alpha \in \{1,2\}$ and $N \in \mathbf{Z}_{>0}$. Let $b, b_0 \in \mathbf{R}$ and let $\beta = 1 + bN^{-1/3}, \beta_0 = 1 + b_0 N^{-1/3}$.*

Let $p_N(\,\cdot\,; h)$ be the density of the eigenvalues of an $N \times N$ GUE if $\alpha = 1$ or GOE if $\alpha = 2$ with a spike of h. If $\Lambda \sim p_N(\,\cdot\,; \beta_0)$, then

$$\frac{p_N(\Lambda; \beta)}{p_N(\Lambda; \beta_0)} \xrightarrow{\mathrm{d}} \frac{\int_{\mathcal{K}} e^{bw/\alpha} s_{b_0}^{(\alpha)}(w)^{-1}\,\mathrm{d}w}{\int_{\mathcal{K}} e^{b_0 w/\alpha} s_{b_0}^{(\alpha)}(w)^{-1}\,\mathrm{d}w}, \tag{1.17}$$

where $\mathcal{K}$ is a contour that runs from $-\mathrm{i}\infty$ to $+\mathrm{i}\infty$ and passes on the positive side of the largest zero of $s_{b_0}^{(\alpha)}$.

The limiting object on the left-hand side of Eq. (1.17) is extremely complicated. However, the convergence of the likelihood ratio without centering or scaling to a non-trivial limit is an essential result for the contiguity of the experiments

$$H_0\colon h = 1 + b_0 N^{-1/3} \qquad \text{vs.} \qquad H_1\colon h = 1 + bN^{-1/3},$$

which resolves the question raised in Section 1.2.2 and was the motivation for the direction of inquiry of this thesis. An important question left for future investigation is a more explicit description of the limiting quantity on the right-hand side of Eq. (1.17).

1.7 Efficient simulations of random matrix objects

The eigenvalue distribution of G(U/O)E matrices are special cases of the eigenvalue distributions of so-called “Generalized Gaussian β-ensembles.” Introduced in [DE02], these are a family of ensembles of tridiagonal matrices parameterized by the Dyson parameter β. When $\beta = 1$ or 2, these eigenvalue distributions match those of GOE and GUE matrices respectively.

These tridiagonal matrices can be easier to analyse than dense Gaussian matrices, and indeed this tridiagonal representation is a crucial element in the analysis performed in [JKOP20; JKOP21]. In addition to their theoretical uses, tridiagonal matrices require less memory to store in a computer and can be diagonalized more quickly than dense matrices. This makes the tridiagonal representation essential when simulating the eigenvalues of large Gaussian matrices.

In Chapter 5, we extend the ideas of [DE02] to cover spiked matrices. In particular, for $d \in \mathbf{Z}_{>0}$, Section 5.1 describes a family of random banded matrix ensembles with bandwidth $2d+1$ whose eigenvalue distributions match those of a Gaussian matrix with d spikes. It also contains a similar banded representation of Wishart matrices with d spikes, extending the tridiagonal representation of Wisharts also given in [DE02].

1.8 Notation and definitions

We establish here some notational and terminological conventions that will be used throughout this thesis.

If M_N is a matrix with real eigenvalues $\lambda_1 \geq \cdots \geq \lambda_N$ and $I \subseteq \mathbf{R}$, then we denote the *eigenvalue counting function* $\mathcal{N}_{M_N}$ by

$$\mathcal{N}_{M_N}(I) = \#\{j : \lambda_j \in I\}.$$

The *one-point correlation function* of M_N, denoted by ρ_N, is then the density of $\mathbf{E}\mathcal{N}_{M_N}$ with respect to the Lebesgue measure. That is, for any measurable $I \subseteq \mathbf{R}$, we have

$$\mathbf{E}\mathcal{N}_{M_N}(I) = \int_I \rho_N(x)\,\mathrm{d}x. \tag{1.18}$$

The notation $a_N \lesssim b_N$ means that $a_N \leq C b_N$ for some C and N large. Let (B_N) be a sequence of events. We then say that

1. B_N holds *asymptotically almost surely* (a.a.s.) if $\mathbf{P}(B_N) \to 1$ as $N \to \infty$.

2. B_N holds *with high probability* if there exists a $d > 0$ such that

$$\mathbf{P}(B_N^c) \lesssim N^{-d}.$$

3. B_N holds *with overwhelming probability* (w.o.p.) if, for all $A > 0$,

$$\mathbf{P}(B_N^c) \lesssim N^{-A}.$$

If $X_N \lesssim c_N$ w.o.p. and there are constants C_0, C_1 such that eventually $c_N \geq N^{-C_0}$ and $\mathbf{E}X_N \leq N^{C_1}$, then $\mathbf{E}X_N \lesssim c_N$ (for proof, see e.g. [BK18, Lemma 7.1].) Here and later "$X_N \lesssim c_N$ w.o.p." means that there exists C such that event $X_N \leq Cc_N$ holds w.o.p. Similarly for statements like $X_N = O(c_N)$ w.o.p.

We say that θ_N is a $\Theta_{\mathbf{P}}(1)$ variable if θ_N is a.s. positive and θ_N, θ_N^{-1} are $O_{\mathbf{P}}(1)$. Clearly $\exp\{O_{\mathbf{P}}(1)\} = \Theta_{\mathbf{P}}(1)$ and $\log\{\Theta_{\mathbf{P}}(1)\} = O_{\mathbf{P}}(1)$. If θ_{N1}, θ_{N2} are $\Theta_{\mathbf{P}}(1)$ then so are $\theta_{N1}\theta_{N2}$ and θ_{N1}/θ_{N2}.

We collect for later use some elementary criteria for convergence in probability of a sequence of random variables $\{X_N\}$.

C1 If for each c large, $X_N = Y_{N1}(c) + Y_{N2}(c)$ with $Y_{N1}(c) = o_{\mathbf{P}}(1)$ and $\mathbf{E}Y_{N2}(c)^2 \leq 1/c^2$ for $N > N(c)$, then $X_N = o_{\mathbf{P}}(1)$.

C2 If for each ε small there exist events $\mathcal{E}_{N,\varepsilon}$ of probability at least $1-\varepsilon$ for $N > N(\varepsilon)$ such that on $\mathcal{E}_{N,\varepsilon}$ we have $X_N = Y_{N1}(\varepsilon)+Y_{N2}(\varepsilon)$ with $Y_{Nk}(\varepsilon) = O_{\mathbf{P}}(1)$, then $X_N = O_{\mathbf{P}}(1)$.

Chapter 2

CLT for the log-determinant

This chapter is drawn from [JKOP20], which was co-authored with Iain Johnstone, Alexei Onastki and Yegor Klochkov. The majority of this chapter comprises results developed by the present author. Section 2.2 contains a minimal set of results proved by co-authors that is required to understand the results of this chapter. These results are stated without proofs, but with appropriate references to [JKOP20].

2.1 Introduction

Let $\lambda_1 \geq \cdots \geq \lambda_N$ be the eigenvalues of an $N \times N$ Wigner matrix. The logarithmic linear statistic

$$\mathcal{L}_N = \sum_{j=1}^{N} f(\lambda_j) = \sum_{j=1}^{N} \log(E - \lambda_j)$$

arises in several applications; we focus in particular on statistical testing in spiked models and on the fluctuation behavior of the free energy in the spherical Sherrington-Kirkpatrick (SSK) model of statistical physics. Suppose initially that $E > 2$ is fixed. In this case $\mathcal{L}_N - N \int f \, \mathrm{d}\rho_{\mathrm{SC}}$, where ρ_{SC} denotes the semicircle density is asymptotically Gaussian with finite variance that depends on the first four moments of the entries of W_N. Since $f(z) = \log(E - z)$ is analytic in a neighborhood of the semi-circle support, this follows from general CLTs for linear statistics, e.g. [BY05].

This chapter concerns Gaussian behavior of the log determinant

$$L_N = \sum_{j=1}^{N} \log|\lambda_j - E| = \log|\det(W_N - E)|$$

near, at, or just inside the edge:

$$E = E_N = 2 + \sigma_N N^{-2/3}, \qquad -\gamma \leq \sigma_N \ll \log^2 N \tag{2.1}$$

for some fixed $\gamma > 0$. Specifically, we will prove the following result:

Theorem 2.1. *Let W_N be a Wigner matrix whose off-diagonal moments match GUE ($\alpha = 1$) or GOE ($\alpha = 2$) to third order. For edge values $E = E_N$ satisfying Eq. (2.1), we have*

$$\left(\log|\det(W_N - E)| - \mu_N\right)/\tau_N \xrightarrow{\mathrm{d}} \mathcal{N}(0,1), \tag{2.2}$$

with

$$\mu_N = \frac{1}{2}N + \sigma_N N^{1/3} - \tfrac{2}{3}\sigma_N^{3/2} - \frac{1}{6}(\alpha - 1)\log N, \qquad \tau_N = \sqrt{\frac{\alpha}{3}\log N}. \tag{2.3}$$

2.2 Gaussian case

Theorem 2.1 is proved in [JKOP20] by first establishing the result for GUEs, extending it to GOEs using an eigenvalue interlacing result of [FR99], and then using a Lindeberg swapping argument to conclude the complete result for Wigner matrices whose entry distributions have moments that match those of Gaussian matrices up to appropriate order.

The following Theorem 2.2 was developed for GUE matrices by Alexei Onatski and Yegor Klochkov in [JKOP20]. In this section, we show how to extend this result to the GOE case before proceeding to the more general Wigner case in Section 2.3.

Theorem 2.2 ([JKOP20, Theorem 2])**.** *Consider a matrix $\hat{W}_N$ from a (scaled) general Gaussian β-ensemble with $\beta = 2/\alpha$. Let $D_N = \det(\hat{W}_N - 2\theta_N)$, where $2\theta_N \equiv E = 2 + N^{-2/3}\sigma_N$ with $(\log\log N)^2 \ll \sigma_N \ll (\log N)^2$. Then,*

$$(\log|D_N| - \mu_N)/\tilde{\tau}_N \xrightarrow{d} \mathcal{N}(0,1),$$

where

$$\tilde{\tau}_N = \sqrt{\alpha r(\theta_N^{-2})} \quad \textit{with} \quad r(x) = \log\frac{1}{2}[1 + (1-x)^{-1/2}].$$

Remark 2.3. Theorem 2.2 concerns "general Gaussian β-ensembles." These are ensembles of tridiagonal matrices introduced in [DE02] that generalize the eigenvalue behavior of Gaussian matrices. In particular, when $\beta = 1, 2, 4$, the eigenvalues of a "Gaussian β-ensemble" have the same distribution as the eigenvalues of a GOE, GUE or Gaussian Symplectic Ensemble (GSE) respectively. Here, the GSE is a symmetric matrix ensemble with quaternion entries, whose distribution is invariant under symplectic transformation.

Theorem 2.2 covers singularities $2\theta_N = E = 2 + N^{-2/3}\sigma_N$ in the range $(\log\log N)^2 \ll \sigma_N \ll \log^2 N$ for all positive α. Section 3 of [JKOP20] details how to extend the result to singularities at a distance of exact order $N^{-2/3}$ away from the edge, or even (just) inside the bulk. That is, it concerns sequences σ_N satisfying

$$-\gamma \le \sigma_N \le \bar{\sigma}_N := (\log\log N)^3 \qquad \text{for some } \gamma > 0. \tag{2.4}$$

The extension to this case relies on the properties of GUE, and so allows us to conclude Theorem 2.1 only for GUE matrices.

In order to further extend the result to GOE matrices, we make use of the following proposition, which connects the fluctuations of a linear statistic of a GUE with those of a GOE.

In the following few results, we will use the following notation: for a function $f\colon \mathbf{R} \to \mathbf{R}$ and an $N \times N$ matrix W with real eigenvalues, we write

$$f(W) := \sum_{j=1}^{N} f(\lambda_j),$$

where $\lambda_1 \ge \cdots \ge \lambda_N$ are the eigenvalues of W.

Proposition 2.4. *Let $W_N^{\mathbf{C}}$ and $W_N^{\mathbf{R}}$ be $N \times N$ scaled GUE and GOE matrices, respectively. Suppose that f_N is a series of linear statistics such that*

$$f_N(W_N^{\mathbf{C}}) = a_N + O_{\mathbf{P}}(b_N),$$

for some sequences a_N and b_N. Then,

$$f_N(W_N^{\mathbf{R}}) = a_N + O_{\mathbf{P}}(b_N + \mathrm{TV}(f_N)),$$

where $\mathrm{TV}(f_N)$ *is the total variation of the univariate function* $f_N \colon \mathbf{R} \to \mathbf{R}$.

If the f_N are uniformly bounded, this result is uninformative about the fluctuations of $f_N(W_N^{\mathbf{R}})$, since b_N and $\mathrm{TV}(f_N)$ are both $O(1)$. It becomes more useful when the f_N have singularities in or near the bulk distribution. In particular, a critical technical result of [JKOP20] is lemma 18. It considers the linear statistics $f_c^l(\lambda) = (E-\lambda)^{-l}\mathbf{1}_{\{|E-\lambda|>cN^{-2/3}\}}$, stating that $f(W_N^{\mathbf{C}})$ has fluctuations of order $N^{-1+\frac{2}{3}l}$. In this case, $\mathrm{TV}(f_c^l) = O(N^{\frac{2}{3}l})$, and so Proposition 2.4 allows the extension of [JKOP20, lemma 18] to GOEs.

The main engine for proving Proposition 2.4 is an identity stated in [FR99], which relates the eigenvalues of a GUE to the eigenvalues of two independent GOEs. In particular, we use it in the following lemma.

Lemma 2.5. *Let* $M_N^{\mathbf{C}}$ *be an* $N \times N$ *GUE, and let* f *be a function with total variation* $\mathrm{TV}(f) < \infty$*. If* $M_N^{\mathbf{R}}, \tilde{M}_N^{\mathbf{R}}$ *are two independent* $N \times N$ *GOEs, then*

$$f(M_N^{\mathbf{C}}) \overset{\mathrm{d}}{=} \frac{1}{2}(f(M_N^{\mathbf{R}}) + f(\tilde{M}_N^{\mathbf{R}})) + X_N, \tag{2.5}$$

where $|X_N| \leq \mathrm{TV}(f)$*, and* $\overset{\mathrm{d}}{=}$ *denotes equality in distribution.*

Proof. Let $M_N^{\mathbf{R}}, \tilde{M}_{N+1}^{\mathbf{R}}$ be independent $N\times N$ and $(N+1)\times(N+1)$ GOEs. Call the eigenvalues of $M_N^{\mathbf{R}}$ and $\tilde{M}_{N+1}^{\mathbf{R}}$ $\{\lambda_i\}_{i=1}^N$ and $\{\tilde{\lambda}_i\}_{i=1}^{N+1}$, respectively. Further, denote the combined set of eigenvalues $\{\lambda_i\}_{i=1}^N \cup \{\tilde{\lambda}_i\}_{i=1}^{N+1}$ by Λ^+, and enumerate its elements in decreasing order

$$\Lambda^+ = \{\lambda_1^+ \geq \cdots \geq \lambda_{2N+1}^+\}.$$

[FR99, Theorem 5.2] implies that the even elements of this set (that is, $\{\lambda_2, \lambda_4, \ldots, \Lambda_{2N}\}$) are equal in distribution to the eigenvalues of an $N \times N$ GUE.

Thus, if $M_N^{\mathbf{C}}$ is an $N \times N$ GUE, we have

$$\begin{aligned}
f(M_N^{\mathbf{C}}) &\overset{\mathrm{d}}{=} \sum_{i=1}^N f(\lambda_{2i}^+) \\
&= \frac{1}{2}\Big(\sum_{j=1}^{2N+1} f(\lambda_j^+) + \sum_{i=1}^N [f(\lambda_{2i}^+) - f(\lambda_{2i-1}^+)] - f(\lambda_{2N+1}^+)\Big) \\
&= \frac{1}{2}\Big(f(M_N^{\mathbf{R}}) + f(\tilde{M}_{N+1}^{\mathbf{R}}) - f(\lambda_{2N+1}^+) + \sum_{i=1}^N [f(\lambda_{2i}^+) - f(\lambda_{2i-1}^+)]\Big).
\end{aligned}$$

Notice that, since λ_j^+ are ordered, we have

$$\Big|\sum_{i=1}^{N}[f(\lambda_{2i}^+) - f(\lambda_{2i-1}^+)]\Big| \leq \mathrm{TV}(f).$$

Further, let $\tilde{M}_N^{\mathbf{R}}$ be the principal submatrix of $\tilde{M}_{N+1}^{\mathbf{R}}$, which is thus independent and equal in distribution to $M_N^{\mathbf{R}}$. If we let $\tilde{\mu}_1, \dots, \tilde{\mu}_N$ be the eigenvalues of $\tilde{M}_N^{\mathbf{R}}$, then Cauchy's interlacing theorem yields

$$\tilde{\lambda}_1 \geq \tilde{\mu}_1 \geq \tilde{\lambda}_2 \geq \cdots \geq \tilde{\lambda}_N \geq \tilde{\mu}_N \geq \tilde{\lambda}_{N+1},$$

and so we have

$$\begin{aligned}
|f(\tilde{M}_{N+1}^{\mathbf{R}}) - f(\lambda_{2N+1}^+) - f(\tilde{M}_N^{\mathbf{R}})| &= \Big|\sum_{i=1}^{N} f(\tilde{\lambda}_i) - \sum_{i=1}^{N} f(\tilde{\mu}_i) + (f(\tilde{\lambda}_{N+1}) - f(\lambda_{2N+1}^+))\Big| \\
&\leq \sum_{i=1}^{N}|f(\tilde{\lambda}_i) - f(\tilde{\mu}_i)| + |f(\tilde{\lambda}_{N+1}) - f(\lambda_{2N+1}^+)| \\
&\leq \mathrm{TV}(f).
\end{aligned}$$

We conclude that Eq. (2.5) holds. □

An immediate useful corollary is as follows.

Corollary 2.6. *Under the assumptions of Lemma 2.5,*

$$\begin{aligned}
\mathbf{E} f(M_N^{\mathbf{R}}) &= \mathbf{E} f(M_N^{\mathbf{C}}) + O(\mathrm{TV}(f)), \\
\operatorname{Var} f(M_N^{\mathbf{R}}) &\leq 2 \operatorname{Var} f(M_N^{\mathbf{C}}) + 2\mathrm{TV}^2(f).
\end{aligned}$$

Remark 2.7. Notice that Lemma 2.5 and Corollary 2.6 also hold for scaled Gaussian matrices $W_N^{\mathbf{R}/\mathbf{C}} = M_N^{\mathbf{R}/\mathbf{C}}/\sqrt{N}$, since $f(W_N^{\mathbf{R}/\mathbf{C}}) = g(M_N^{\mathbf{R}/\mathbf{C}})$ for $g(\lambda) = f(\lambda/\sqrt{N})$, which satisfy $\mathrm{TV}(f) = \mathrm{TV}(g)$.

However, to finish proving Proposition 2.4 in its generality, we require the following technical lemma about tightness.

Lemma 2.8. *Let X_N, Y_N be iid sequences of random variables such that $X_N + Y_N$ is tight. Then X_N (and thus also Y_N) is tight.*

Proof. For any constant K, we have

$$\mathbf{P}(X_N > K) = \mathbf{P}(X_N > K, Y_N > K)^{1/2} \leq \mathbf{P}(|X_N + Y_N| > K)^{1/2},$$

and similarly,

$$\mathbf{P}(X_N < -K) \leq \mathbf{P}(|X_N + Y_N| > K)^{1/2},$$

which yields

$$\sup_N \mathbf{P}(|X_N| > K) \leq 2 \sup_N \mathbf{P}(|X_N + Y_N| > K)^{1/2},$$

where the right hand side of the latter inequality can be made arbitrarily small by the tightness of $X_N + Y_N$. □

With all these results in hand, we are ready to complete the proof of Proposition 2.4. We have

$$\begin{aligned}\left|\frac{f_N(W_N^{\mathbf{R}}) - a_N}{b_N + \mathrm{TV}(f_N)} + \frac{f_N(\tilde{W}_N^{\mathbf{R}}) - a_N}{b_N + \mathrm{TV}(f_N)}\right| &= 2\left|\frac{(f_N(W_N^{\mathbf{R}}) + f_N(\tilde{W}_N^{\mathbf{R}}))/2 - a_N}{b_N + \mathrm{TV}(f_N)}\right| \\ &\leq 2\left|\frac{(f_N(W_N^{\mathbf{R}}) + f_N(\tilde{W}_N^{\mathbf{R}}))/2 + X_N - a_N}{b_N}\right| + 2\left|\frac{X_N}{\mathrm{TV}(f_N)}\right|.\end{aligned} \tag{2.6}$$

By Lemma 2.5, the first term in the latter sum is tight, since it is equal in distribution to $(f_N(W_N^{\mathbf{C}}) - a_N)/b_N$, whereas the second term is no larger than 2. But since the two terms on the left hand side of Eq. (2.6) are iid, Lemma 2.8 yields that they must be tight, and so

$$f_N(W_N^{\mathbf{R}}) = a_N + O_{\mathbf{P}}(b_N + \mathrm{TV}(f_N)).$$

2.3 Extension to Wigner matrices

2.3.1 Outline of approach

Proving that a Wigner matrix W_N' satisfies a certain property as long as a matrix W_N from G(U/O)E satisfies this property is often based on the Lindeberg swapping process, where elements of W_N are replaced by the elements of W_N' one by one without losing the property in question. Typically, one needs to show that any individual swap does not change the expectation $\mathbf{E}Q(M)$ of some smooth function $Q(\cdot)$ of the matrix M participating in the swapping process too much.

Although our initial interest is in the asymptotic normality of the log-determinant, we will eventually need to use Lindeberg swapping for several functionals which depend on the Stieltjes transform evaluated at $z = E + \mathrm{i}\eta$ for E near the edge and η distant at least $N^{-2/3-\delta}$ from the real axis — here the gross $N^{-2/3}$ scale is that appropriate for working at the edge of the spectrum. We outline the swapping approach for the log-determinant example but with the general class of "Stieltjes edge functionals" in mind.

We adopt the method of [TV12], with modifications to work at the edge, and under weakened assumptions, as described below. We call a quantity $S(W_N)$ *insensitive* at rate δ_N if $S(W_N) - S(W_N') = O(\delta_N)$. Let $L_N(W_N) = \log|\det(W_N - E)|$. To extend the asymptotic normality of $L_N(W_N)$ to $L_N(W_N')$ it is sufficient, via a standard smoothing argument, to show that $\mathbf{E}G \circ L_N(W_N)$ is insensitive at rate δ_N for scalar functions for which $\|G^{(j)}\|_\infty \le b_N^j$. For the log-determinant $\delta_N = b_N \asymp (\log N)^{-1/4}$ will work.

In an initial *approximation* step, we show that it suffices to replace $L_N(W)$ by a function of the Stieltjes transform $s_W = \frac{1}{N}\operatorname{tr}(W - z)^{-1}$

$$g(W) = N \int_{\gamma_N}^{N^{100}} \operatorname{Im} s_W(E + \mathrm{i}\eta)\,\mathrm{d}\eta. \tag{2.7}$$

Here $\gamma_N = N^{-2/3-\delta}$: to show that values $0 \le \eta \le \gamma_N$ can be neglected, we use an anti-concentration result that guarantees that with high probability, all eigenvalues are at least $N^{-2/3-\zeta}$–distant from E. This too is proved by Lindeberg swapping, now with a second Stieltjes functional, in Proposition 2.15.

The swapping argument is now applied to show that $\mathbf{E}Q(W_N)$ is insensitive for Q of the form $(G \circ g)(W_N)$. To review this in outline, let γ index an ordering of the independent components $\{\operatorname{Re}\xi_{ij}, \operatorname{Im}\xi_{ij}\}_{i<j}$ and $\{\xi_{ii}\}$ of W_N. Thus γ runs over N^2 and $N(N+1)/2$ elements in the Hermitian and symmetric cases respectively. By convention in each case, the first N values of γ index the diagonal matrix entries. Thus W^γ will refer to a matrix in which the elements prior to γ come from W_N' while those at γ or later come from W_N.

At stage γ in the swapping process, we can write $W^{(0)} = W^\gamma$, $W^{(1)} = W^{\gamma+1}$, and

$$W^{(0)} = W_0 + \frac{\xi^{(0)}}{\sqrt{N}} V, \qquad W^{(1)} = W_0 + \frac{\xi^{(1)}}{\sqrt{N}} V, \tag{2.8}$$

and $W_0 = W_0^\gamma$ is independent of both $\xi^{(0)}$ and $\xi^{(1)}$. In the symmetric case, V is one of the elementary matrices of the form $e_a e_a^*$ or $e_a e_b^* + e_b e_a^*$, for $1 \le a < b \le N$. In the Hermitian case, we add matrices $\mathrm{i}e_a e_b^* - \mathrm{i}e_b e_a^*$. The variables $\xi^{(0)}$ and $\xi^{(1)}$ correspond to the γth components of W_N and W'_N respectively. All matrices W^γ are Wigner matrices.

To focus on individual swaps, write

$$\mathbf{E}Q(W) - \mathbf{E}Q(W') = \sum_\gamma \mathbf{E}\Delta_\gamma,$$

with $\Delta_\gamma = Q(W^\gamma) - Q(W^{\gamma+1}) = Q(W^{(0)}) - Q(W^{(1)})$.

We consider $W^{(0)}$ and $W^{(1)}$ as perturbations of W_0. Thus, set $W_t^\gamma = W_0^\gamma + tN^{-1/2}V_\gamma$, and introduce $Q_\gamma(t) = Q(W_t^\gamma)$. Note that this function is independent of $\xi^{(i)}$, and that

$$\Delta_\gamma = Q_\gamma(\xi^{(0)}) - Q_\gamma(\xi^{(1)}).$$

In a Taylor expansion of Q_γ, formal for now, this independence implies

$$\mathbf{E}[Q_\gamma(\xi^{(i)})] = \sum_j \frac{1}{j!}\mathbf{E}[Q_\gamma^{(j)}(0)]\,\mathbf{E}([\xi^{(i)}]^j).$$

If moments match at order $j \le k-1$, that is, $\mathbf{E}([\xi^{(0)}]^j) = \mathbf{E}\left([\xi^{(1)}]^j\right)$, then the jth order term in $\mathbf{E}\Delta_\gamma$ vanishes. If, as one expects, $Q_\gamma^{(k)}(t)$ is of order $N^{-k/2}b_N$, and bounding the remainder term appropriately leads to the required bounds on $\mathbf{E}\Delta_\gamma$. This is formalized in Proposition 2.9.

To show that such derivative bounds hold specifically for $Q = G \circ g$ when g is as in Eq. (2.7), we need good bounds for $\partial_t^j g^\gamma(t)$ when $g^\gamma(t) = g(W_t^\gamma)$. Introduce notation for the resolvent and Stieltjes transforms

$$R_t^\gamma = R_t^\gamma(z) = (W_t^\gamma - z)^{-1}, \qquad s_t^\gamma(\eta) = N^{-1}\operatorname{tr} R_t^\gamma(E + \mathrm{i}\eta). \tag{2.9}$$

The standard resolvent perturbation argument (equations Eq. (2.18)-Eq. (2.21)) shows that $\partial_t^j s_t^\gamma = c_j N^{-j/2-1}\operatorname{tr}[(R_t^\gamma V)^j R_t^\gamma]$.

This is bounded for E near the edge and $\eta > N^{-2/3-\delta}$ using the entrywise local law (see Proposition 2.12(i)). Working at the edge allows, through use of the Ward identity, improvements in bounds because $\operatorname{Im} R$ is small. What results (see the proof of Proposition 2.14) are bounds $\|\partial_t^j g^\gamma(t)\|_\infty \lesssim N^{-j/2}a_N$ with $a_N = 1$ in the log-determinant case. These bounds are useful both for reducing the number of matching moments required to three (for off-diagonal entries) and requiring only bounded variances (for diagonal entries). Combining with the derivative bounds on G, the chain rule shows that we obtain the desired insensitivity with $\delta_N = a_N b_N = b_N$.

2.3.2 Lindeberg swapping formalism for asymptotically flat Q

Throughout this chapter, we will use the term "Wigner matrix" to refer to Wigner matrices satisfying **W1**–**W3** as defined in Definition 1.2.

Moreover, we will say that the moments of two Wigner matrices W_N, W_N' *match to order* m if, for an integer $0 < a \le m$,

$$\mathbf{E}(\operatorname{Re}\xi_{ij})^a = \mathbf{E}(\operatorname{Re}\xi_{ij}')^a, \qquad \mathbf{E}(\operatorname{Im}\xi_{ij})^a = \mathbf{E}(\operatorname{Im}\xi_{ij}')^a$$

for all $1 \le i < j \le N$. Note that this constrains only the *off-diagonal* entries. The diagonal entries already match to order one by assumption, which is all that we need.

This means that **W4** can be rephrased as the condition that the entries of a matrix W_N match those of a Gaussian matrix up to order 3.

In Proposition 2.9 and its consequence Proposition 2.10, we make the swapping argument explicit for abstract Q satisfying generic asymptotic 'flatness' derivative bounds. In the next section we assemble tools — resolvent perturbation and local law — with the goal of establishing, in Proposition 2.14, the necessary flatness bounds for some specific choices of Q needed for our later applications.

Fix $c_0 > 0$ and set $\|F\|_{c_0} = \sup\{|F(t)|, |t| \le N^{c_0}\}$. Let $\delta_N \to 0$ in such a way that $\delta_N \gtrsim N^{-c_1}$ for some $c_1 > 0$. Let Q be a function on $N \times N$ Hermitian/symmetric matrices taking values in $[0,1]$. Let Wigner matrices W_N, W_N' be given and define $Q_\gamma(t) = Q(W_t^\gamma)$ as in Section 2.3.1. We say that Q satisfies **condition F** or $F(\delta_N)$ if for all γ and $1 \le k \le 4$ we have w.o.p. that

$$\|Q_\gamma^{(k)}\|_{c_0} \lesssim N^{-\frac{k}{2}}\delta_N. \tag{F}$$

Proposition 2.9. *Let W_N, W_N' be Wigner matrices whose moments match to third order. Let $c_0, c_1 > 0$ be fixed and for each $j = 1, \dots, m$, let $Q_j \colon \mathbf{C}^{N\times N} \to [0,1]$ satisfy condition*

$F(\delta_{j,N})$. *If* $Q = \prod_{j=1}^m Q_j$, *then,*

$$\mathbf{E}Q(W_N) - \mathbf{E}Q(W_N') \lesssim \max_{j=1,\dots,m} \delta_{j,N}. \tag{2.10}$$

Proof. Consider first the case $m = 1$. We set $\Delta_{\gamma i} = Q(W^{(i)}) - Q(W_0)$, and decompose

$$\mathbf{E}Q(W_N) - \mathbf{E}Q(W_N') = \sum_\gamma \mathbf{E}(\Delta_{\gamma 0} - \Delta_{\gamma 1}).$$

Let $E_N = E_N(W_0^\gamma)$ denote the overwhelming probability event Eq. (F) and then introduce 'good' events $G_{Ni} = E_N \cap \{|\xi^{(i)}| \le N^{c_0}\}$. Let A be a fixed constant such that $N^{2-A} \lesssim \delta_N$. Using boundedness of Q and the moment bound **W3**, with p chosen so that $pc_0 > A$, we have

$$\mathbf{E}(\Delta_{\gamma 0} - \Delta_{\gamma 1}) = \mathbf{E}(\Delta_{\gamma 0}\mathbf{1}(G_{N0})) - \mathbf{E}(\Delta_{\gamma 1}\mathbf{1}(G_{N1})) + O(N^{-A}). \tag{2.11}$$

As before, set $Q_\gamma(t) = Q(W_t^\gamma)$, so that $\Delta_{\gamma i} = Q_\gamma(\xi^{(i)}) - Q_\gamma(0)$. By Taylor expansion, we have

$$\Delta_{\gamma i} = \sum_{j=1}^{k-1} \frac{1}{j!} Q_\gamma^{(j)}(0)(\xi^{(i)})^j + \frac{1}{k!} Q_\gamma^{(k)}(\xi^*)(\xi^{(i)})^k,$$

for some ξ^* with $|\xi^*| \le |\xi^{(i)}|$. Noting that $Q_\gamma(t)$ and event E_N are independent of $\xi^{(i)}$, we have

$$\begin{aligned} \mathbf{E}[Q_\gamma^{(j)}(0)(\xi^{(i)})^j\mathbf{1}(G_{Ni})] &= \mathbf{E}[Q_\gamma^{(j)}(0)\mathbf{1}(E_N)]\,\mathbf{E}[|\xi^{(i)}|^j\mathbf{1}(|\xi^{(i)}| \le N^{c_0})] \\ &= \mathbf{E}[Q_\gamma^{(j)}(0)\mathbf{1}(E_N)]\,\mathbf{E}[|\xi^{(i)}|^j] + O(N^{-j/2}\delta_N \cdot N^{-A}), \end{aligned}$$

where we used the fact that $\mathbf{E}[|\xi^{(i)}|^j\mathbf{1}(|\xi^{(i)}| > N^{c_0})] \le C_p N^{-c_0(p-j)} = O(N^{-A})$ for suitable p, as follows from **W3** and the Hölder inequality. For the remainder, on event G_{Ni} we also have $|Q_\gamma^{(k)}(\xi^*)| \le \|Q_\gamma^{(k)}\|_{c_0} \lesssim N^{-k/2}\delta_N$, and hence

$$|\mathbf{E}[Q_\gamma^{(k)}(\xi^*)(\xi^{(i)})^k\mathbf{1}(G_{Ni})]| \lesssim N^{-k/2}\delta_N.$$

Summarizing, we have

$$\mathbf{E}[\Delta_{\gamma i}\mathbf{1}(G_{Ni})] = \sum_{j=1}^{k-1} \frac{1}{j!}\mathbf{E}[Q_\gamma^{(j)}(0)\mathbf{1}(E_N)]\,\mathbf{E}[|\xi^{(i)}|^j] + O(N^{-k/2}\delta_N + N^{-A}).$$

Choose $k = k(\gamma)$ so that $\mathbf{E}(\xi^{(1)})^j = \mathbf{E}(\xi^{(0)})^j$ for $1 \le j \le k-1$. Then the sums cancel and Eq. (2.11) yields

$$\mathbf{E}(\Delta_{\gamma 0} - \Delta_{\gamma 1}) = O(N^{-k/2}\delta_N + N^{-A}).$$

For the $O(N^2)$ off-diagonal terms, moment matching to third order allows $k(\gamma) = 4$, while for the N diagonal terms, we take $k(\gamma) = 2$, since then only $\mathbf{E}\xi^{(i)} = 0$. Summing over all γ, we obtain

$$\mathbf{E}Q(W_N) - \mathbf{E}Q(W_N') = O(\delta_N + N^{2-A}) = O(\delta_N) \tag{2.12}$$

from the choice of A.

For $m > 1$, apply the product rule, use Eq. (F) and $\|Q_\gamma\|_{c_0} \le 1$:

$$\|Q_\gamma^{(k)}\|_{c_0} \lesssim \sum_{\ell_1+\cdots+\ell_m=k} \binom{k}{\ell_1, \ldots, \ell_m} \prod_{\substack{1\le j\le m \\ \ell_i \ge 1}} N^{-\frac{\ell_j}{2}} \delta_{j,N} \lesssim N^{-\frac{k}{2}} \max_{j=1,\ldots,m} \delta_{j,N},$$

Thus Q satisfies $F(\max_j \delta_{j,N})$ and the result follows from Eq. (2.12). □

We use Proposition 2.9 to formulate a criterion that allows joint convergence in distribution of vector functions of W_N to be transferred to the corresponding functions of W_N'.

Proposition 2.10. *Let W_N, W_N' be Wigner matrices whose moments match up to third order. Let $\boldsymbol{\xi}_N = \boldsymbol{\xi}_N(W_N)$ and $\boldsymbol{\xi}_N' = \boldsymbol{\xi}_N(W_N')$ both be $\mathbf{R}^m$ valued random vectors. Suppose that $\boldsymbol{\xi}_N \xrightarrow{\text{d}} \boldsymbol{\xi}$, and that each component ξ_j of the limit has a continuous distribution function.*

Let $\eta_N \to 0$ be given, and suppose that for each $1 \le j \le m$ and $s \in \mathbf{R}$ there exist functions $Q_j^{\pm}(\cdot, s)$ satisfying condition $F(\delta_{j,N})$ such that for $W = W_N, W_N'$, w.o.p.

$$\mathbf{1}\{\xi_{Nj}(W) \le s\} \le Q_j^+(W, s) \le \mathbf{1}\{\xi_{Nj}(W) \le s + \eta_N\} \tag{2.13}$$

$$\mathbf{1}\{\xi_{Nj}(W) \le s - \eta_N\} \le Q_j^-(W, s) \le \mathbf{1}\{\xi_{Nj}(W) \le s\} \tag{2.14}$$

Then we also have (joint) convergence $\boldsymbol{\xi}_N' \xrightarrow{\text{d}} \boldsymbol{\xi}$.

Proof. Let $\delta_N = \max_j \delta_{j,N}$. It suffices to show that for each $\mathbf{s} = (s_j)$ that

$$\mathbf{P}(\boldsymbol{\xi}_N \le \mathbf{s} - \eta_N) - O(\delta_N) \le \mathbf{P}(\boldsymbol{\xi}_N' \le \mathbf{s}) \le \mathbf{P}(\boldsymbol{\xi}_N \le \mathbf{s} + \eta_N) + O(\delta_N). \tag{2.15}$$

Indeed, we then have

$$|\mathbf{P}(\boldsymbol{\xi}'_N \le \mathbf{s}) - \mathbf{P}(\boldsymbol{\xi}_N \le \mathbf{s})| \le \sum_j \mathbf{P}(|\boldsymbol{\xi}_{Nj} - s_j| \le \eta_N) + O(\delta_N) \to 0,$$

since each limiting component ξ_j has a continuous distribution function.

We verify the upper bound in Eq. (2.15). For each $A > 0$ large, we have from Eq. (2.13) for W'_N, then Proposition 2.9 and then Eq. (2.13) again, now for W_N, that

$$\begin{aligned}\mathbf{P}(\boldsymbol{\xi}'_N \le \mathbf{s}) &\le \mathbf{E}\prod_j Q_j^+(W'_N, s_j) + O(N^{-A}) \\ &\le \mathbf{E}\prod_j Q_j^+(W_N, s_j) + O(\delta_N) \le \mathbf{P}(\boldsymbol{\xi}_N \le \mathbf{s} + \eta_N) + O(\delta_N).\end{aligned}$$

The lower bound in Eq. (2.15) follows similarly, using Eq. (2.14). □

2.3.3 Flatness for Stieltjes functionals

Resolvent perturbation: deterministic bounds

We recall and modify some bounds of [TV12] on stability of Hermitian matrices with respect to perturbation in one or two entries, using Ward's identity to improve the bounds at the edge.

Let M_0 be a Hermitian $N \times N$ matrix, $z = E + \mathrm{i}\eta \in \mathbf{C}_+$, where $\mathbf{C}_+$ is the open half-plane $\mathbf{C}_+ = \{z \in \mathbf{C} : \mathrm{Re}(z) > 0\}$, and V an elementary matrix as defined immediately after Eq. (2.8). Set $M_t = M_0 + tN^{-1/2}V$ and $R_t = R_t(z) = (M_t - z)^{-1}$, and $s_t(z) = N^{-1}\operatorname{tr} R_t(z)$. Recall from [TV12] the definitions of the matrix norms $\|A\|_{(q,p)}$, and in particular

$$\|A\|_{(\infty,1)} = \max_{1\le i,j\le N} |A_{ij}|, \qquad \|A\|_{(\infty,2)} = \max_i \Big(\sum_j |A_{ij}|^2\Big)^{1/2}.$$

Note also that if V is an elementary matrix, then

$$|\operatorname{tr}(AV)| = |\operatorname{tr}(VA)| \le 2|A|_{(\infty,1)} \tag{2.16}$$

$$\|AVB\|_{(\infty,1)} \le 2\|A\|_{(\infty,1)}\|V\|_{(\infty,1)}\|B\|_{(\infty,1)}. \tag{2.17}$$

Let $\kappa_N(z,t) = tN^{-1/2}\|R_t\|_{(\infty,1)}$. Lemma 12 of [TV12] says that if $\kappa_N(z,t) \to 0$ as

$N \to \infty$, then for large N

$$R_{t+u} = R_t + \sum_{j=1}^{\infty} \Big(\frac{-u}{\sqrt{N}}\Big)^j (R_t V)^j R_t, \tag{2.18}$$

with the right side being absolutely convergent. In addition, for $1 \le p \le \infty$,

$$\|R_t\|_{(\infty,p)} \le \|R_0\|_{(\infty,p)} \exp\{2|t|N^{-1/2}\|R_0\|_{(\infty,1)}\}. \tag{2.19}$$

Here the factor 2 arises from the use of Eq. (2.17) in the Tao-Vu argument. The same bound holds with the roles of R_0 and R_t reversed.

Expansion Eq. (2.18) allows evaluation of t-derivatives of $s_t(z)$. Indeed

$$\partial_t^j s_t(z) = j! N^{-j/2} c_j(z,t) \tag{2.20}$$

$$c_j(z,t) = (-1)^j N^{-1} \operatorname{tr}((R_t V)^j R_t). \tag{2.21}$$

The following variant of [TV12, Proposition 13] yields uniform bounds on c_j in terms of $\|\operatorname{Im} R\|_\infty = \max_{1\le i\le N} |\operatorname{Im} R_{ii}|$, which allows tighter bounds near the edge.

Proposition 2.11. *Let c_0 and A be small and positive, $\check{\sigma}_N = (\log N)^{O(\log\log N)}$, and define*

$$\mathbf{S}_e(A) = \{z = E + \mathrm{i}\eta \in \mathbf{C} :\ |E-2| \le N^{-2/3}\check{\sigma}_N,\ \eta > N^{-2/3-A}\} \tag{2.22}$$

$$\kappa_N = \sup_{|t|\le N^{c_0}, z\in \mathbf{S}_e(A)} |t|\|R_0\|_{(\infty,1)}/\sqrt{N}.$$

Then for $z \in \mathbf{S}_e(A)$ and $|t| \le N^{c_0}$,

$$|c_j(z,t)| \le (N\eta)^{-1} 2^j e^{2(j+1)\kappa_N} \|R_0\|_{(\infty,1)}^{j-1} \|\operatorname{Im} R_0\|_\infty. \tag{2.23}$$

Proof. From the cyclic property of traces, then Eq. (2.16) and Eq. (2.17), we have

$$|\operatorname{tr}((R_t V)^j R_t)| = |\operatorname{tr}(V (R_t V)^{j-1} R_t^2)| \le 2\|(R_t V)^{j-1} R_t^2\|_{(\infty,1)} \le 2^j \|R_t\|_{(\infty,1)}^{j-1} \|R_t^2\|_{(\infty,1)}.$$

The Ward identity, e.g. [BK18, eq. (3.6)] says that

$$\sum_j |R_{ij}|^2 = \eta^{-1} \operatorname{Im} R_{ii}$$

is valid for any resolvent matrix $R = (W - E - \mathrm{i}\eta)^{-1}$ with $\eta \neq 0$ and Hermitian (or symmetric) W. For $\eta > 0$, we have

$$\|R\|_{(\infty,2)}^2 = \sup_i \sum_j |R_{ij}|^2 = \eta^{-1} \|\operatorname{Im} R\|_\infty. \tag{2.24}$$

If B is a normal matrix, (i.e. $B^*B = BB^*$), then

$$\|AB\|_{(\infty,1)} \leq \|A\|_{(\infty,2)} \|B\|_{(\infty,2)}. \tag{2.25}$$

This uses the Cauchy-Schwarz bound $|(AB)_{ij}|^2 \leq \sum_k |A_{ik}|^2 \sum_k |B_{kj}|^2$, since B normal implies $\sum_k |B_{kj}|^2 = \sum_k |B_{jk}|^2 \leq \|B\|_{(\infty,2)}^2$.

The resolvent of a Hermitian matrix is normal, so from Eq. (2.25), Eq. (2.19), and then Eq. (2.24) we have

$$\|R_t^2\|_{(\infty,1)} \leq \|R_t\|_{(\infty,2)}^2 \leq e^{4\kappa_N} \|R_0\|_{(\infty,2)}^2 = \eta^{-1} e^{4\kappa_N} \|\operatorname{Im} R_0\|_\infty.$$

Combine the last two displays and use Eq. (2.19) to bound $\|R_t\|_{(\infty,1)} \leq e^{2\kappa_N} \|R_0\|_{(\infty,1)}$ to arrive at Eq. (2.23). □

Local law

We will need the local law for Wigner matrices and some of its important consequences, in particular at the spectral edge.

Proposition 2.12. *Let W_N be a Wigner matrix.*

(i) (local law) Let $R(z) = (W_N - zI)^{-1}$ denote the resolvent matrix and $s_{sc}(z)$ the Stieltjes transform of the semicircle law. Fix $\tau > 0$ small. For each $\varepsilon > 0$, we have w.o.p.

$$R_{ij} = s_{sc}(z)\delta_{ij} + O(N^\varepsilon \Psi(z)),$$

uniformly for $z \in \mathbf{S}(\tau) = \{E + \mathrm{i}\eta : |E| < \tau^{-1}, N^{-1+\tau} \leq \eta \leq \tau^{-1}\}$ and $i, j = 1, \ldots, N$, where

$$\Psi(z) = \sqrt{\frac{\operatorname{Im} s_{sc}(z)}{N\eta}} + \frac{1}{N\eta}.$$

(ii) (semi-circle law on small scales) For each $\varepsilon > 0$, we have w.o.p. that

$$\mathcal{N}_{W_N}(I) = N \int_I \rho_{\rm sc}(\mathrm{d}x) + O(N^\varepsilon),$$

uniformly for all intervals $I \subset \mathbf{R}$, where $\mathcal{N}_{W_N}(I)$ denotes the number of eigenvalues of W_N in I and ρ_{sc} denotes the semi-circle law.

(iii) (at the edge.) Let $A > 0$ be small and fixed, $\check{\sigma}_N = (\log N)^{O(\log\log N)}$, and let $\mathbf{S}_e(A)$ be the edge domain Eq. (2.22). For each $\varepsilon > 0$ and uniformly for $z = E + \mathrm{i}\eta \in \mathbf{S}_e(A)$, we have w.o.p.

$$\|R\|_{(\infty,1)} \lesssim 1 \wedge \eta^{-1}, \qquad \|\operatorname{Im} R\|_\infty \lesssim (\eta^{1/2} + N^{-1/3+\varepsilon+A}) \wedge \eta^{-1}. \tag{2.26}$$

*Let $W_0 = W - \xi N^{-1/2} V$ with V an elementary matrix and ξ satisfying moment bounds **W3**. Set $R_0 = (W_0 - zI)^{-1}$. Then the bounds Eq. (2.26) apply to R_0 also.*

Remark 2.13. For clarity, we emphasize that these are simultaneous high probability bounds for all z in the indicated ranges. For example, then w.o.p.

$$\sup_{z\in\mathbf{S}_e(A)} (\operatorname{Im} z \vee 1)|\operatorname{Im} R(z)|_\infty \lesssim 1.$$

Such statements follow from the N^2-Lipschitz continuity of $R_{ij}(z), s_{\rm sc}(z)$ and of the right side bounds over the indicated ranges, c.f. e.g. [BK18, Remark 2.7].

Proof. For (i) and (ii), see e.g. [BK18, Theorems 2.6, 2.8].

We turn to (iii). Basic bounds on $s_{sc}(z)$, e.g. [EY17, Lemma 6.2], establish for $\eta > 0$, $|E| \le 10$ and $\kappa = ||E| - 2|$ that

$$|s_{sc}(z)| \le 1, \qquad \operatorname{Im} s_{sc}(z) \lesssim \sqrt{\kappa + \eta}.$$

For $N^{-2/3-A} \le \eta \le 1$, we have $\Psi(z) \lesssim (N\eta)^{-1/2} \le N^{-1/6+A/2}$ and so from the local law $\|R\|_{(\infty,1)} \lesssim 1$. For $\eta \ge 1$, just use the elementary bound $|R_{jk}| \le \eta^{-1}$ arising from the spectral decomposition

$$R_{jk}(E + \mathrm{i}\eta) = \sum_{l=1}^N \frac{u_l(j)u_l^*(k)}{\lambda_l - E - \mathrm{i}\eta}, \tag{2.27}$$

where $u_l(j)$ denotes the j-th component of the eigenvector u_l corresponding to $\lambda_l(W_N)$.

For $(\operatorname{Im} R)_{jj}$ we exploit the improved bounds on $\operatorname{Im} s_{sc}$ at the edge. Since $\kappa \leq N^{-2/3}\check{\sigma}_N$, we have $\operatorname{Im} s_{sc} \lesssim N^{-1/3}\check{\sigma}^{1/2} + \eta^{1/2}$ and $N\eta \geq N^{1/3-A}$, and conclude

$$\Psi(z) \lesssim \frac{N^{-1/6}\check{\sigma}^{1/4}}{\sqrt{N\eta}} + \frac{1}{\sqrt{N\eta^{1/2}}} + \frac{1}{N\eta} \lesssim N^{-1/3+A}.$$

Hence, the second part of Eq. (2.26) follows from the local law.

Turning to R_0, we put $\Delta = W - W_0 = N^{-1/2}\xi V$ and use the resolvent identity $R_0 = R + R\Delta R + R_0(\Delta R)^2$. Write $\|\cdot\|_*$ for $\|\cdot\|_{(\infty,1)}$. Even for R_0, the bound $\|R_0\|_* \leq \eta^{-1}$ follows from Eq. (2.27) as before. So to conclude the rest of Eq. (2.26) for R_0, it suffices to show that w.o.p. $\|R_0 - R\|_* \lesssim N^{-1/3+\varepsilon+A}$ for $N^{-2/3-A} \leq \eta \leq 1$.

We have the trivial bound $\|R_0\|_* \leq \eta^{-1} \leq N^{2/3+A}$. Since **W3** implies that $|\xi| \leq N^{\varepsilon/2}$ w.o.p., we have that $\|\Delta\| \lesssim N^{-1/2+\varepsilon/2}$, and along with $\|R\|_* \lesssim 1$, and bound Eq. (2.17) for elementary matrices, we find that w.o.p. both

$$\|R\Delta R\|_* \lesssim N^{-1/2+\varepsilon/2}, \qquad \|R_0(\Delta R)^2\|_* \lesssim N^{2/3+A-1+\varepsilon} \lesssim N^{-1/3+\varepsilon+A}. \qquad \square$$

Stieltjes functionals

We return to establishing flatness condition F for certain functionals $Q = G \circ g$. Let W be an Hermitian matrix and $s_W(z)$ its empirical Stieltjes transform. In the following proposition, we consider examples of *Stieltjes functionals* $g(W) = \Lambda(s_W)$ for some continuous linear functional Λ acting on functions holomorphic on $\mathbf{C}_+$.

The first two of these examples will be used in the next subsection to extend the non-concentration property for the eigenvalues of G(U/O)E matrices to Wigner matrices (Proposition 2.15) and, using this, to extend the log determinant CLT to Wigner matrices. The last two examples are key to the analysis of the SSK model in Chapter 3, which details results from [JKOP21]. There, we need to extend results on the k-th largest eigenvalue and the trace of the inverse powers of $z - W_N$ from G(U/O)E to Wigner matrices.

Proposition 2.14. *Let W be a Wigner matrix. Let $\varepsilon > 0$, $0 < c_0 < 1/2$ and let $E \in \mathbf{R}$ be such that $|E - 2| \lesssim \check{\sigma}_N N^{-\frac{2}{3}+A}$.*

For each of the following statistics, define functions $g\colon \mathbf{C}^{N\times N} \to \mathbf{R}$, $G\colon \mathbf{R} \to \mathbf{R}$ and a sequence δ_N according to the following specification in each case for $1 \leq j \leq 4$:

1. *Log-determinant: with* $\gamma_N = N^{-2/3-\varepsilon}$,

$$g(W) = N\int_{\gamma_N}^{N^{100}} \operatorname{Im} s_W(E+\mathrm{i}\eta)\,\mathrm{d}\eta, \qquad \|G^{(j)}\|_\infty \le (\log N)^{-j/4}, \qquad \delta_N = (\log N)^{-1/4}.$$

2. *Eigenvalue counting: with* $\eta = N^{-2/3-9\varepsilon}$, $E_1 < E_2$ *such that* $\max_i |E_i - 2| \lesssim N^{-2/3+10\varepsilon}$, *and constants* $C_J > 0$,

$$g(W) = \frac{N}{\pi}\int_{E_1}^{E_2} \operatorname{Im} s_W(x+\mathrm{i}\eta)\,\mathrm{d}x, \qquad \|G^{(j)}\|_\infty \le (\log N)^{C_j}, \qquad \delta_N = N^{-1/3+O(\varepsilon)}.$$

3. *Inverse moments: with* $\eta = N^{-2/3-\varepsilon}$ *and* $l \in \mathbb{Z}_{>0}$,

$$g(W) = N^{-\frac{2}{3}l+1} \operatorname{Re} s_W^{(l-1)}(E+\mathrm{i}\eta), \qquad \|G^{(j)}\|_\infty \le (\log N)^{Cj}, \qquad \delta_N = N^{-\frac{1}{3}+O(\varepsilon)}.$$

In each of the cases listed above, the corresponding function $Q = G \circ g$ *satisfies the condition of Eq.* (F)*. That is, for* $1 \le k \le 4$*, it follows w.o.p. that*

$$\|Q_\gamma^{(k)}\|_{c_0} \lesssim N^{-\frac{k}{2}}\delta_N.$$

Proof. Define $g^\gamma(t) = g(W_t^\gamma)$ so that $Q_\gamma(t) = G(g_\gamma(t))$. In order to bound $Q_\gamma^{(k)}(t)$ we start with bounds for $\partial_t^j g^\gamma$. Recalling Eq. (2.9), we have $g^\gamma(t) = \Lambda(s_t^\gamma)$. Standard results on differentiation of integrals and then Eq. (2.20) imply that

$$\partial_t^j g^\gamma(t) = \Lambda(\partial_t^j s_t^\gamma) = j! N^{-\frac{j}{2}} \Lambda(c_j^\gamma(\,\cdot\,, t)),$$

where from Eq. (2.21) and Eq. (2.9))

$$c_j^\gamma(z,t) = (-1)^j N^{-1} \operatorname{tr}((R_t^\gamma V_\gamma)^j R_t^\gamma).$$

Hence, to bound $\|\partial_t^j g^\gamma\|$, it suffices to use bounds on the coefficients c_j^γ. We will omit the superscript γ to simplify notations. From Propositions 2.11 and 2.12, for fixed $A > 0$,

$$N|c_j(E+\mathrm{i}\eta, t)| \lesssim \begin{cases} \eta^{-1/2} + N^{-1/3+\varepsilon+A}\eta^{-1} & N^{-2/3-A} \le \eta \le 1 \\ \eta^{-j-1} & \eta \ge 1, \end{cases} \tag{2.28}$$

uniformly in $|t| \le N^{c_0}$. Note that there is no dependence on j for $\eta \le 1$.

In the log-determinant case, we have that

$$\Lambda(f) = \int_{\gamma_N}^{N^{100}} N \operatorname{Im} f(y + \mathrm{i}\eta) \, \mathrm{d}\eta.$$

Evaluated at $c_j(\,\cdot\,, t)$ with $A = 2\varepsilon$, we use Eq. (2.28) to obtain

$$\begin{aligned} \int_{\gamma_N}^{N^{100}} N|c_j(E + \mathrm{i}\eta, t)| \, \mathrm{d}\eta &\lesssim \int_{\gamma_N}^{1} (\eta^{-1/2} + N^{-1/3+\varepsilon+A}\eta^{-1}) \, \mathrm{d}\eta + \int_{1}^{N^{100}} \eta^{-j-1} \, \mathrm{d}\eta \\ &\lesssim 1. \end{aligned}$$

For the remaining integrals, we need only the following consequence of Eq. (2.28).

$$N|c_j(E + \mathrm{i}\eta, t)| \lesssim N^{\frac{1}{3}+\varepsilon+2A}. \tag{2.29}$$

Set $A = 10\varepsilon$ in the eigenvalue counting case. This yields

$$\int_{E_1}^{E_2} N|c_j(y + \mathrm{i}\eta, t)| \, \mathrm{d}y \lesssim N^{\frac{1}{3}+\varepsilon+2A} N^{-\frac{2}{3}+\varepsilon} = N^{-\frac{1}{3}+O(\varepsilon)}.$$

For inverse moments, we have $\Lambda(c_j(\cdot, t)) = N^{-2l/3+1} \operatorname{Re} c_j^{(l-1)}(E+\mathrm{i}\eta)$. Let Γ be a contour of radius $N^{-\frac{2}{3}-2\varepsilon}$ around $E + \mathrm{i}\eta$. In this way, each c_j is analytic on the interior of Γ, and so we use Cauchy's integral formula and Eq. (2.29) with $A = 2\varepsilon$ to see that

$$N^{-\frac{2}{3}l+1}|c_j^{(l-1)}(E + \mathrm{i}\eta)| \le \frac{(l-1)!}{2\pi} \oint_{\Gamma} \frac{N|c_j(w)|}{N^{\frac{2}{3}l}|w - E - \mathrm{i}\eta|^l} |\mathrm{d}w| \lesssim N^{-\frac{1}{3}+O(\varepsilon)}.$$

In sum suppose that, for sequences a_N and b_N such that $a_N b_N \to 0$, we have w.o.p.

$$\|\partial_t^j g^\gamma(t)\|_{c_0} \lesssim N^{-\frac{j}{2}} a_N, \qquad \|G^{(j)}\|_\infty \lesssim b_N^j$$

In the proof so far, we have seen that the above conditions hold with the following values of a_N and b_N for some constant C:

1. Log-determinant: $a_N = 1$, $b_N = (\log N)^{-1/4}$.
2. Eigenvalue counting: $a_N = N^{-\frac{1}{3}+O(\varepsilon)}$, $b_N = (\log N)^C$.
3. Inverse moments: $a_N = N^{-\frac{1}{3}+O(\varepsilon)}$, $b_N = (\log N)^C$.

We apply Faà di Bruno's formula to compute bounds for $\partial_t^k(G\circ g^\gamma)(t)$. Let $\mathcal{M}_k = \{m \in \mathbf{Z}_{\geq 0}^k : \sum_{j=1}^k jm_j = k\}$, so that $m_+ = m_1 + \cdots + m_k \geq 1$ for each $m \in \mathcal{M}_k$. Then for certain combinatorial constants C_{km} we have that, uniformly in $|t| \leq N^{c_0}$,

$$\begin{aligned}|\partial_t^k(G\circ g^\gamma)(t)| &\leq \sum_{m\in\mathcal{M}_k} C_{km}|G^{(m_+)}(g^\gamma(t))| \cdot \prod_{j=1}^k |g^{(j)}(t)^{m_j}| \\ &\lesssim \sum_{m\in\mathcal{M}_k} C_{km} b_N^{m_+} \prod_{j=1}^k N^{-\frac{jm_j}{2}} a_N^{m_j} = N^{-\frac{k}{2}} \sum_{m\in\mathcal{M}_k} C_{km}(a_N b_N)^{m_+} \lesssim N^{-\frac{k}{2}} a_N b_N,\end{aligned}$$

Hence, the conclusion in each case follows with $\delta_N = a_N b_N$. □

2.3.4 Conclusions for Wigner matrices

The last important result required to conclude Theorem 2.1 for Wigner matrices is eigenvalue non-concentration. That is, we require a result that indicates that the eigenvalues of a Wigner matrix a reasonable separated from E with high probability.

We proceed with this as before: first establishing the result for Gaussian matrices, and then using Proposition 2.14 to conclude that it still holds in the Wigner case.

Non-concentration

Let us introduce new notation

$$\check{\sigma}_N = (\log N)^{O(\log\log N)}. \tag{2.30}$$

The specific non-concentration result required is the next proposition. The version of it for G(U/O)E matrices appears in [JKOP20, Lemma 17].

Proposition 2.15. *Let W_N' be a Wigner matrix whose off-diagonal moments match GOE or GUE to third order. Call its eigenvalues $\lambda_1', \ldots, \lambda_N'$. Let $E \in \mathbf{R}$ be such that $|E-2| \lesssim N^{-\frac{2}{3}}\check{\sigma}_N$. Then there exists a c_1 such that, for each $c_0 \in (0, c_1)$, there exists $d > 0$ such that, for N large,*

$$\mathbf{P}(\min_{j=1,\ldots,N} |\lambda_j' - E| \leq N^{-\frac{2}{3}-c_0}) \leq N^{-d}. \tag{2.31}$$

Proof. Define the eigenvalue counting function $\mathcal{N}_W(E_1, E_2) = \#\{j : E_1 \leq \lambda_j(W) \leq E_2\}$.

The event in Eq. (2.31) has the form $\mathcal{N}_W(E_1, E_2) \geq 1$. The first step is to approximate this using the Stieltjes transform.

Let $\varepsilon = 2c_0$ and define $\ell = \frac{1}{2}N^{-\frac{2}{3}-\varepsilon}$, and $\eta = N^{-\frac{2}{3}-9\varepsilon}$. Let $E_1, E_2 \in \mathbf{R}$ be such that $|E_1 - 2|, |E_2 - 2| \lesssim N^{-\frac{2}{3}}\check{\sigma}_N$ and $E_2 - E_1 \geq 2\ell$.

A suitable approximation is given by Corollary 17.3 of [EY17] (based on the local law and eigenvalue rigidity), which we apply twice with $E = E_1$ and E_2 respectively. Subtracting the latter bounds from the former, this yields w.o.p. that

$$\frac{N}{\pi}\int_{E_1+\ell}^{E_2-\ell} \operatorname{Im} s_W(y+\mathrm{i}\eta)\,\mathrm{d}y - 2N^{-\varepsilon} \leq \mathcal{N}_W(E_1,E_2) \leq \frac{N}{\pi}\int_{E_1-\ell}^{E_2+\ell} \operatorname{Im} s_W(y+\mathrm{i}\eta)\,\mathrm{d}y + 2N^{-\varepsilon}.$$

Let $E^{\pm} = E \pm 2N^{-\frac{2}{3}-c_0}$, and define the function

$$g(W) = \frac{N}{\pi}\int_{E^-+\ell}^{E^+-\ell} \operatorname{Im} s_W(y+\mathrm{i}\eta)\,\mathrm{d}y,$$

Applying these bounds with $(E_1, E_2) = (E^-, E^+)$ and $(E^- + 2\ell, E^+ - 2\ell)$, we conclude that, w.o.p.,

$$\mathcal{N}_W(E^- + 2\ell, E^+ - 2\ell) - 2N^{-\varepsilon} \leq g(W) \leq \mathcal{N}_W(E^-, E^+) + 2N^{-\varepsilon}. \tag{2.32}$$

Let G be a smooth increasing function such that

$$G(x) = \begin{cases} 1 & \text{if } x \geq 2/3, \\ 0 & \text{if } x \leq 1/3. \end{cases}$$

Taking $Q = G \circ g$ and applying G to each side of Eq. (2.32), we then have that, w.o.p.,

$$\mathbf{1}\{\mathcal{N}_W(E^- + 2\ell, E^+ - 2\ell) \geq 1\} \leq Q(W) \leq \mathbf{1}\{\mathcal{N}_W(E^-, E^+) \geq 1\}.$$

Now we can use Propositions 2.9 and 2.14(2) to compare $Q(W_N')$ with $Q(W_N)$, for W_N drawn from G(U/O)E with eigenvalues λ_j. For any $A > 0$, we have

$$\begin{aligned}\mathbf{P}(\min_j|\lambda_j' - E| \leq 2N^{-\frac{2}{3}-c_0} - 2\ell) &= \mathbf{P}\{\mathcal{N}_{W_N'}(E^- + 2\ell, E^+ - 2\ell) \geq 1\} \\ &\leq \mathbf{E}Q(W_N') + O(N^{-A}) \\ &\leq \mathbf{E}Q(W_N) + O(N^{-\frac{1}{3}+O(\varepsilon)})\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbf{P}(\min_j |\lambda_j - E| \leq 2N^{-\frac{2}{3}-c_0}) + O(N^{-\frac{1}{3}+O(\varepsilon)}) \\
&\leq \frac{1}{2}N^{-d} + O(N^{-1/3+O(\varepsilon)}) \leq N^{-d}.
\end{aligned}
$$

At the last line we applied the non-concentration bound for G(U/O)E of [JKOP20, Lemma 17].

For N large, we have $2N^{-\frac{2}{3}-c_0} - 2\ell \geq N^{-2/3-c_0}$ and so the final bound Eq. (2.31) follows from these inequalities. □

The next lemma shows that non-concentration implies control of inverse power sums at around their typical magnitude. The proof is by standard dyadic decomposition.

Lemma 2.16. *Let $\{\lambda_j\}$ be the eigenvalues of a Wigner matrix W_N whose off-diagonal moments match GOE or GUE to third order. Suppose that $|E-2| \leq N^{-2/3}\check{\sigma}_N$. Then there exist constants $\{C_r\}$ such that for each $\varepsilon > 0$ small, with high probability we have*

$$
S_r(E) := \sum_{j=1}^{N} \frac{1}{|\lambda_j - E|^r} \leq \begin{cases} C_1 N & \text{if } \; r = 1 \\ C_r N^{2r/3+(r+1)\varepsilon} & \text{if } \; r \geq 2. \end{cases}
$$

The bounds also hold for $S_r(E')$ uniformly in $|E'-E| \leq \delta/2$ with $\delta = N^{-2/3-\varepsilon}$, by increasing C_r to $2^r C_r$.

Proof. Let $\delta = N^{-2/3-\varepsilon}$ and $A_N = \{\min_j |E-\lambda_j| > \delta\}$: by Proposition 2.15 this event has probability at least $1 - N^{-\varepsilon/2}$. We will work on event A_N, and show that there the claims hold w.o.p. On A_N the interval $I_0 = [E-\delta, E+\delta]$ contains no eigenvalues. Consider the 'coronae' defined by $I_k = \{x \in \mathbf{R} \ : \ 2^{k-1}\delta < |x-E| \leq 2^k\delta\}$ for $1 \leq k \leq k' = \min\{k : E - 2^k\delta \leq 1\}$, and add two half-infinite intervals I_{-1} and $I_{k'+1}$ to obtain a disjoint cover of $\mathbf{R}$. We may then bound (on event A_N)

$$
S_r(E) \leq \sum_{k=1}^{k'} \frac{\mathcal{N}_{W_N}(I_k)}{(2^{k-1}\delta)^r} + \frac{N}{(2^{k'}\delta)^r}. \tag{2.33}
$$

The semicircle density is bounded by $\sqrt{2-x}\mathbf{1}_{x\leq 2}$ and so $\rho_{\mathrm{sc}}([2-a, 2-b]) \leq a^{3/2}$. The lower endpoint of I_k is $E - 2^k\delta \geq 2 - 2^k\delta - \check{\sigma}_N N^{-2/3}$. Since $\check{\sigma}_N^{3/2} \leq N^{\varepsilon}$ for large N,

$$
\rho_{\mathrm{sc}}(I_k) \leq \sqrt{2}\left((2^k\delta)^{3/2} + N^{\varepsilon-1}\right).
$$

Proposition 2.12 (ii) says that, with overwhelming probability, simultaneously for all $k \le k' = O(\log N)$, we have

$$\mathcal{N}_{W_N}(I_k) \le N\rho_{\mathrm{sc}}(I_k) + O(N^\varepsilon) \le \sqrt{2}N(2^k\delta)^{3/2} + CN^\varepsilon.$$

Putting this into Eq. (2.33) and noting that $2^{k'}\delta \in [\frac{1}{2}, 3]$ we obtain w.o.p.

$$S_r(E) \le 2^{r+1/2}N\sum_1^{k'}(2^k\delta)^{3/2-r} + CN^\varepsilon\delta^{-r} + 2^rN.$$

The sum may be bounded using

$$N\delta^{3/2-r}\sum_1^{k'}2^{(3/2-r)k} \le \begin{cases} 3^{1/2}N & \text{if } r = 1 \\ 4N^{-3\varepsilon/2}\delta^{-r} & \text{if } r \ge 2. \end{cases}$$

Observe that $N^\varepsilon\delta^{-r} = N^{(2/3+\varepsilon)r+\varepsilon}$. For $r = 1$, this is $o(N^{-1})$ and so $S_1(E) \le C_1N$ on A_N w.o.p. For $r \ge 2$ this is the dominant term, so that $S_r(E) \le C_rN^{2r/3+(r+1)\varepsilon}$.

The bounds also hold for $S_r(E')$ uniformly in $|E' - E| \le \delta/2$, by increasing C_r to 2^rC_r. Indeed, for such E', on event A_N we have $|\lambda_j - E'| \ge \frac{1}{2}|\lambda_j - E|$ for all j. □

Log-determinant

We derive the central limit theorem for the log-determinant

$$L_N(W_N') = \log|\det(W_N' - E)|$$

for a Wigner matrix W_N' and $E = E_N = 2 + \sigma_N N^{-2/3}$. Recall the scaling constants μ_N, τ_N from Eq. (2.3). Let W_N be drawn from (scaled) GOE or GUE. From Theorem 2.1, which we have already established for the Gaussian ensembles, we have

$$\check{L}_N(W_N) = \tau_N^{-1}(L_N(W_N) - \mu_N) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1). \tag{2.34}$$

Proposition 2.17 (Log determinant CLT). *Let W_N' be a Wigner matrix whose off-diagonal moments match GOE or GUE to third order. Let $E = E_N = 2 + \sigma_N N^{-2/3}$ with with $-\gamma \le \sigma_N \ll \log^2 N$.*

Then

$$\tau_N^{-1}(\log|\det(W_N' - E)| - \mu_N) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1). \tag{2.35}$$

Proof. To rewrite the log-determinant in terms of an integral of the Stieltjes transform, note that $(\mathrm{d}/\,\mathrm{d}\eta)\log|\lambda - E - \mathrm{i}\eta| = \mathrm{Im}[(\lambda - E - \mathrm{i}\eta)^{-1}]$, which yields [TV12, eq. (46)]

$$L_N(W) = \log|\det(W - E - \mathrm{i}N^{100})| - N\int_0^{N^{100}} \mathrm{Im}\, s_W(E + \mathrm{i}\eta)\,\mathrm{d}\eta.$$

The uniform moment bounds **W3** imply that

$$\log|\det(W - E - \mathrm{i}N^{100})| = N\log(N^{100}) + O_{\mathbf{P}}(N^{-50}).$$

Moreover, for each $\varepsilon > 0$ small, if we take $\gamma_N = N^{-\frac{2}{3}-2\varepsilon}$, then non-concentration implies that the contribution to the integral from $\eta \le \gamma_N$ is negligible. Indeed, with $\lambda_j = \lambda_j(W_N)$,

$$N\,\mathrm{Im}\, s_W(E + \mathrm{i}\eta) = \eta\sum_{j=1}^N \frac{1}{(\lambda_j - E)^2 + \eta^2} \le \eta\sum_{j=1}^N \frac{1}{(\lambda_j - E)^2}.$$

By Lemma 2.16 we thus have

$$\left|N\int_0^{\gamma_N} \mathrm{Im}\, s_W(E + \mathrm{i}\eta)\,\mathrm{d}\eta\right| \le \tfrac{1}{2}\gamma_N^2 S_2(E) = O_{\mathbf{P}}(\gamma_N^2 N^{\frac{4}{3}+3\varepsilon}) = o_{\mathbf{P}}(1).$$

To summarize, if we define the Stieltjes functional

$$g(W) = N\int_{\gamma_N}^{N^{100}} \mathrm{Im}\, s_W(E + \mathrm{i}\eta)\,\mathrm{d}\eta,$$

set $\bar{\mu}_N = \mu_N + N\log(N^{100})$ and define

$$\xi_N(W) = \tau_N^{-1}(g(W_N) - \bar{\mu}_N),$$

then we have shown that $\check{L}_N(W_N) = \xi_N(W_N) + o_{\mathbf{P}}(1)$.

We carry out the Lindeberg swapping with $g(W)$. Let $H^+ : \mathbf{R} \to [0,1]$ be a smooth decreasing function such that

$$H^+(x) = \begin{cases} 1 & \text{if } \ x \le 0 \\ 0 & \text{if } \ x \ge \eta_N, \end{cases}$$

and let $H^-(x) = H^+(x - \eta_N)$. For $s \in \mathbf{R}$ define $G_s^\pm(x) = H^\pm(\tau_N^{-1}(x - \bar{\mu}_N) - s)$. One verifies that

$$\mathbf{1}\{\xi_N(W) \leq s\} \leq G_s^+(g(W)) \leq \mathbf{1}\{\xi_N(W) \leq s + \eta_N\}. \tag{2.36}$$

Setting $Q^+(W, s) = G_s^+(g(W))$, we obtain bound Eq. (2.13). Bound Eq. (2.14) follows similarly using instead H^- with $Q^-(W, s) = G_s^-(g(W))$.

Observe that $\|G_s^{\pm(j)}\|_\infty \lesssim (\tau_N \eta_N)^{-j} \lesssim (\log N)^{-j/4}$ if we choose $\eta_N = \tau_N^{-1/2}$. Then Proposition 2.14 (1) implies that $Q_s^\pm$ satisfy condition F with $\delta_N = (\log N)^{-1/4}$. From Proposition 2.10 we conclude that $\xi_N(W_N')$ and hence $\check{L}(W_N')$ have the same limiting distribution as $\xi_N(W_N)$ and $\check{L}_N(W_N)$. Thus the validity of Theorem 2.1 for Gaussian ensembles implies Eq. (2.35). $\square$

Chapter 3

Spin glass to paramagnetic transition in the SSK model

This chapter presents results from [JKOP21]. Similarly to Chapter 2, it contains results developed in coauthorship with Iain Johnstone, Alexei Onastki and Yegor Klochkov. Technical results about Gaussian and spiked matrices were mostly developed by coauthors, and so such results are stated without proof, but with the relevant citations.

3.1 Introduction

We study the large-N behavior of the spherical integrals

$$Z_{\alpha,N} = \int_{\mathcal{S}_\alpha^{N-1}} \exp\{\frac{N\beta}{\alpha} \cdot u^* M_\alpha u\}(\mathrm{d}u), \tag{3.1}$$

where

$$M_\alpha = J \cdot ww^* + W_N \tag{3.2}$$

is a random $N \times N$ matrix such that W_N is a real (when $\alpha = 1$) or complex (when $\alpha = 2$) Wigner matrix; J is a constant from $[0, 1)$; and w is an arbitrary unit-length vector from $\mathbb{C}^N$ if $\alpha = 1$ or from $\mathbb{R}^N$ if $\alpha = 2$. In (3.1), $\mathcal{S}_\alpha^{N-1}$ denotes the unit sphere in $\mathbb{C}^N$ if $\alpha = 1$ or in $\mathbb{R}^N$ if $\alpha = 2$, $(\mathrm{d}u)$ denotes the normalized uniform measure over $\mathcal{S}_\alpha^{N-1}$, and the symbol * denotes combined transposition and complex conjugation. We investigate the limiting distributions of the quantities

$$F_{\alpha,N} = \frac{\alpha}{2N} \log Z_{\alpha,N} \tag{3.3}$$

for β in the so-called "critical regime" of $\beta = 1 + O(N^{-1/3}\sqrt{\log N})$.

Our original motivation stems from the fact that integrals (3.1) appear in the likelihood ratio in statistical tests of spiked models in multivariate statistics, as seen in Eq. (1.3). In such models, J and β play the roles of the size of the spike under the null and under alternative hypotheses, respectively.

Recall from Section 1.5 that the Spherical Sherrington-Kirkpatrick model with Curie-Weiss ferromagnetic interaction is a model of magnetism with Hamiltonian

$$H_N(\sigma) = \frac{1}{2}\Big(\frac{1}{\sqrt{N}}\sum_{i,j=1}^{N} A_{ij}\sigma_i\sigma_j + \frac{J}{N}\sum_{i,j=1}^{N}\sigma_i\sigma_j\Big), \tag{3.4}$$

where $\sigma \in \sqrt{N}\mathcal{S}_2^{N-1}$, $J \geq 0$ is known as the coupling constant, and A is a real symmetric $N \times N$ matrix with zeroes on the diagonal and independent upper triangular entries A_{ij} with mean zero and variance 1.

Within this context, $F_{\alpha,N}$ can be interpreted as the free energy of the SSK model, where A is the corresponding Wigner matrix.

3.1.1 Main result

In this chapter, we investigate the regime

$$\beta = 1 + bN^{-1/3}\sqrt{\log N}, \qquad 0 \leq J < 1,$$

which, as discussed in Section 1.5, is the paramagnetic to spin-glass transition in the SSK model.

We show that, in this case, $F_{\alpha,N}$ has fluctuations of order $N/\sqrt{\log N}$. Moreover, as b increases from $-\infty$ to ∞, we describe the transition of the limiting distribution of $F_{\alpha,N}$ from Gaussian to the Tracy-Widom.

Precisely, our main result is as follows.

Theorem 3.1. *Consider* $F_{\alpha,N}$ *with* $\alpha = 1$ *or* $\alpha = 2$*, as defined in Eq.* (3.1) *– Eq.* (3.3)*. Let* $\beta = 1 + bN^{-1/3}\sqrt{\log N}$ *for a constant* $b \in \mathbf{R}$ *and let* $0 \leq J < 1$*. Further let* $b_+ = \max\{0, b\}$

be the positive part of b*. Then*

$$\frac{N}{\sqrt{\frac{\alpha}{12}\log N}}\left(F_{\alpha,N}-F(\beta)+\frac{\log N}{12N}\right)\xrightarrow{\mathrm{d}}\mathcal{N}(0,1)+\sqrt{\frac{3}{\alpha}}b_+\,\mathrm{TW}_{2/\alpha}, \tag{3.5}$$

where TW_2 *and* TW_1 *are the complex and real Tracy-Widom distributions, respectively, independent from the* $\mathcal{N}(0,1)$*, and where* $F(\beta)$ *is as in Eq.* (1.12)*, that is*

$$F(\beta)=\begin{cases}\beta-\frac{1}{2}\log\beta-\frac{3}{4} & \text{if } b\geq 0,\\ \frac{1}{4}\beta^2 & \text{if } b<0.\end{cases}$$

3.2 Preliminary results

Contour integral representations. Our analysis is based on the now well known contour integral representation of $Z_{\alpha,N}$, cf. [JKOP21, Section 9.6]

$$Z_{\alpha,N}=\frac{C_{\alpha,N}}{2\pi\mathrm{i}}\int_{\mathcal{K}}\exp\{(N/\alpha)G(z)\}\,\mathrm{d}z,\qquad G(z)=\beta z-\frac{1}{N}\sum_{j=1}^{N}\log(z-\lambda_j), \tag{3.6}$$

where for now the integration contour $\mathcal{K}$ is the vertical line from $\gamma-\mathrm{i}\infty$ to $\gamma+\mathrm{i}\infty$ for any constant $\gamma>\lambda_{1,\alpha}$, $\lambda_1\geq\cdots\geq\lambda_N$ are the eigenvalues of M_α, and

$$C_{\alpha,N}=\frac{\Gamma(N/\alpha)}{(\beta N/\alpha)^{N/\alpha-1}}.$$

Notice that the integrand is an analytic function in $\mathbf{C}\setminus(-\infty,\lambda_1]$ and that the integral along the circular arc

$$C_{R,K}=\{z\in\mathbf{C}:|z|=R,\mathrm{Re}(z)\leq K\}$$

satisfies, for large enough R,

$$\left|\int_{C_{R,K}}\exp\{(N/\alpha)G(z)\}\,\mathrm{d}z\right|\leq 2\pi R\cdot\frac{e^{N\beta K/\alpha}}{(R/2)^{N/\alpha}}\xrightarrow{R\to\infty}0.$$

In particular, Cauchy's theorem implies that $\mathcal{K}$ can be deformed without affecting the value of the integral as long as λ_j are never intersected and as long as the resulting contour has real part bounded above.

Log determinant CLT.

In this section, we will use the following version of the log-determinant CLT. It is a generalization of Theorem 2.1 to spiked Wigner matrices whose proof can be found in [JKOP20, Proposition 29] but is omitted from this thesis:

Proposition 3.2 (Spiked log determinant CLT)**.** *Let $\tilde{W}_N$ be a subcritically-spiked Wigner matrix. Let $E = E_N = 2 + \tilde{\sigma}_N N^{-2/3}$ with $\tilde{\sigma}_N$ a monotone sequence for which $\tilde{\sigma}_N \geq -C$ for some positive constant C and $\tilde{\sigma}_N = o(\log^2 N)$. Then*

$$\tau_N^{-1}(\log|\det(\tilde{W}_N - E)| - \mu_N) \xrightarrow{\mathrm{d}} N(0,1). \tag{3.7}$$

One-point correlation function. Let ρ_N be the level density or one-point function of GUE. Then the expectation of a linear spectral statistic is given by

$$\mathbf{E}\Big[N^{-1}\sum_{i=1}^{N} f(\lambda_i)\Big] = \int f(\lambda)\rho_N(\lambda)\,\mathrm{d}\lambda. \tag{3.8}$$

A key tool in approximating such expectations will be a uniform bound, due to Götze and Tikhomirov, for the deviation of the one-point function in GUE from the semicircle density $\rho_{\mathrm{SC}}(x) = (2\pi)^{-1}\sqrt{4-x^2}\mathbf{1}_{|x|\leq 2}$. Indeed, [GT05, Theorem 1.2] shows the existence of absolute constants $\gamma, C > 0$ such that for all $|x| \leq 2 - \gamma N^{-2/3}$,

$$|\rho_N(x) - \rho_{\mathrm{SC}}(x)| \leq \frac{C}{N(4-x^2)}\,. \tag{3.9}$$

In addition, the one-point function decays at least exponentially at the edge. Specifically, as discussed in [JKOP21], for all $s > -\kappa$, for large enough N

$$\rho_N(2 + sN^{-2/3}) \leq C(\kappa)N^{-1/3}e^{-2s}\,. \tag{3.10}$$

A similar bound holds at the negative edge, by symmetry. Corresponding bounds also hold for ρ_{SC}.

Convergence at the edge. We will rely on the properties of the top eigenvalues of scaled GUE or GOE. The celebrated papers [TW94; TW96; Die05] showed that, for each fixed j, the scaled eigenvalues $N^{2/3}(\lambda_j - 2)$ converge in law to the jth Tracy-Widom distribution, $\mathrm{TW}_{2/\alpha,j}$. We need some further consequences of this convergence, along with the extension of these consequences to Wigner matrices. The particular results are summarized in the following lemma, which is proved in Section 3.5.2:

Lemma 3.3. *Let $\lambda_1 \geq \cdots \geq \lambda_N$ be the eigenvalues of subcritically-spiked Wigner matrix satisfying the conditions **W1**, **W3** and **W4** stated in Definition 1.2. Then*

(i) For any $\varepsilon > 0$, there are $C_\varepsilon, N_\varepsilon$ such that for $N \geq N_\varepsilon$, with probability at least $1-\varepsilon$,

$$\lambda_1 \geq 2 + C_\varepsilon N^{-2/3}.$$

(ii) For any fixed $x \in \mathbf{R}$, there exists a constant C_x, such that

$$\mathbf{E}\# \left\{ j : \ \lambda_j \geq 2 - xN^{-2/3} \right\} \leq C_x.$$

(iii) For some $c_\varepsilon, N_\varepsilon$ and any $N \geq N_\varepsilon$, with probability at least $1-\varepsilon$

$$\lambda_1 - \lambda_2 \geq c_\varepsilon N^{-2/3}. \tag{3.11}$$

In other words, $\lambda_1 - \lambda_2 = \Theta_{\mathbf{P}}(N^{-2/3})$.

(iv) Suppose $b_N \to \infty$ is such that $b_N = O(N^\varepsilon)$ for any $\varepsilon > 0$. We then have a.a.s. that

$$\#\{j : \lambda_j > 2 - b_N N^{-2/3}\} \gtrsim b_N^{3/2}.$$

3.2.1 Proof strategy

The main derivations in this chapter revolve around the analysis of the integral

$$\int_{\mathcal{K}_\alpha} \exp\left\{(N/\alpha)G(z)\right\} \mathrm{d}z.$$

The fluctuations of this term differ qualitatively for $b < 0$ and $b \geq 0$, and are considered in Sections 3.3 and 3.4 respectively. In both cases the proofs involve Laplace approximation, but on different contours.

Section 3.3: Negative critical case

We use the vertical contour of Eq. (3.6), and a deterministic choice for γ suffices. Indeed, use the Stieltjes transform of the semicircle law to make the approximation

$$G'(z) = \beta - \frac{1}{N}\sum_{j=1}^{N}\frac{1}{z-\lambda_j} \approx \beta - \frac{z-\sqrt{z^2-4}}{2}.$$

When $b < 0$, the critical point of the approximation is $\gamma = \hat{\gamma} + o(N^{-1+\varepsilon})$ for $\hat{\gamma} = 2 + b^2 N^{-2/3}\log N$ and any small positive ε. Laplace approximation of the integral

$$\int_{\mathcal{K}} \exp\{(N/\alpha)[G(z) - G(\hat{\gamma})]\}\,\mathrm{d}z$$

requires bounds on derivatives $G^{(l)}(\hat{\gamma})$, for $l = 1, 2, 3$, stated in the key Lemma 3.6. Its proof, in Section 3.5.4, relies on the asymptotic approximation of the one-point correlation functions of the GUE developed by Götze and Tikhomirov [GT05]. Having established that the fluctuations of $F_{\alpha,N}$ depend asymptotically only on $G(\hat{\gamma})$, it remains only to apply Proposition 3.2, conclude that $F_{\alpha,N}$ is asymptotically Gaussian, and compute the correct centering and scaling constants.

Section 3.4: Non-negative critical case

When $b \geq 0$, the deterministic approximation to G no longer has a critical point along the real axis, and the approximation $\hat{\gamma}$ fails. Indeed, Lemma 3.6 shows that $G'(\hat{\gamma})$ is of greater order than when $b < 0$, so $G(z)$ oscillates too rapidly along the vertical contour through $\hat{\gamma}$.

We consider first $b > 0$, and instead use the contour of Fig. 3.1, which has a vertical part $\mathcal{K}_3$ through $\mu = (\lambda_1 + \lambda_2)/2$ and a keyhole part $\mathcal{K}_1 \cup \mathcal{K}_2$ extending horizontally from μ and surrounding λ_1. The integral turns out to be dominated by the keyhole part, with

$$\frac{1}{2\pi\mathrm{i}}\int_{\mathcal{K}_1\cup\mathcal{K}_2} \exp\{(N/\alpha)G(z)\}\,\mathrm{d}z = \exp\Big\{(N/\alpha)\hat{G}(\lambda_1) - \frac{\alpha-1}{3}\log N + O_{\mathbf{P}}(\log\log N)\Big\}, \tag{3.12}$$

where

$$\hat{G}(\lambda_1) = \beta\lambda_1 - \frac{1}{N}\sum_{j=2}^{N}\log(\lambda_1 - \lambda_j). \tag{3.13}$$

The proof requires bounds on the derivatives $\hat{G}^{(l)}(\lambda_1), l = 1, 2$ given in Lemma 3.10, and proved in Section 3.5.4 by reduction to $\hat{G}^{(l)}(2)$, which was studied in [JKOP20].

In the boundary case $b = 0$, the contributions of $\mathcal{K}_3$ and $\mathcal{K}_1 \cup \mathcal{K}_2$ are of the same order of magnitude, so we consider instead the contour of the steepest descent. We establish upper and lower bounds on the integral and recover the right-hand side of Eq. (3.12) in this case also.

The analysis of $\hat{G}(\lambda_1)$ is based on the approximation

$$\sum_{j=2}^{N} \log(\lambda_1 - \lambda_j) = \sum_{j=1}^{N} \log|2 - \lambda_j| + N(\lambda_1 - 2) + O_{\mathbf{P}}(1). \tag{3.14}$$

The right-hand side sum can be handled by Proposition 3.2. The λ_1 terms in Eq. (3.13) and Eq. (3.14) both contribute to Tracy-Widom fluctuations. The last part of the argument hinges on the asymptotic independence of λ_1 and $N^{-1}\sum_{j=1}^{N} \log|2 - \lambda_j|$. This result is stated for Gaussian matrices in Proposition 3.4 and extended to Wigner matrices in Proposition 3.17.

Independence of the largest eigenvalue from the linear statistic

The following results, established by Alexei Onatski and Yegor Klochkov in [JKOP21, Proposition 5.1], is vital for understanding the limiting distribution in the $b > 0$ case. We state it here and extend it to the Wigner case in Proposition 3.17.

Proposition 3.4. *For GUE ($\alpha = 1$) and GOE ($\alpha = 2$) the random variables*

$$\xi_{1N} = \frac{N/2 - \frac{\alpha-1}{6}\log N - \sum_{j=1}^{N}\log|2-\lambda_j|}{\sqrt{\frac{\alpha}{3}\log N}} \quad \text{and} \quad \xi_{2N} = N^{2/3}(\lambda_1 - 2)$$

are asymptotically independent in the sense that $(\xi_{1N}, \xi_{2N}) = (X_N, Y_N) + o_{\mathbf{P}(1)}$, where X_N, and Y_N are independent $O_{\mathbf{P}}(1)$ random variables.

The proof of this, which appears in [JKOP21, Section 5] relies on the tridiagonal representation of GUE and GOE matrices — showing that Y_N asymptotically depends only on a minor of order $N^{1/3}\log^3 N$ while X_N asymptotically depends only on the remainder of the matrix.

3.3 Negative-critical regime

For the case $b < 0$, we deform $\mathcal{K}_\alpha$ so that it is the vertical line passing through $\hat{\gamma}$, a point in $\mathbf{R}$ that approximates the critical point γ of the function $G(z)$. Note that

$$G'(z) = \beta - \frac{1}{N}\sum_{j=1}^{N}\frac{1}{z-\lambda_j},$$

where $\frac{1}{N}\sum_{j=1}^{N}\frac{1}{z-\lambda_{\alpha,j}}$ is the negative of the Stieltjes transform of the spectral distribution of M_α. For $z > 2$, it must converge to the negative of the Stieltjes transform of the semi-circle law, that is to

$$-m_{sc}(z) = \frac{z-\sqrt{z^2-4}}{2}.$$

Solving $\beta + m_{sc}(z) = 0$ for z when β is given by

$$\beta = 1 + bN^{-1/3}\sqrt{\log N}, \tag{3.15}$$

we obtain $z = \hat{\gamma} + o(N^{-1+\varepsilon})$ for any $\varepsilon > 0$, where

$$\hat{\gamma} = 2 + b^2N^{-2/3}\log N.$$

Lemma 3.5. *Suppose that $b < 0$. Then*

$$\int_{\hat{\gamma}-\mathrm{i}\infty}^{\hat{\gamma}+\mathrm{i}\infty}\exp\left\{\frac{N}{\alpha}[G(z)-G(\hat{\gamma})]\right\}\mathrm{d}z = 2\sqrt{\pi\alpha|b|}\frac{\mathrm{i}\log^{1/4}N}{N^{2/3}}\left(1+o_{\mathbf{P}}(1)\right).$$

Proof. Changing variables $z \mapsto \hat{\gamma} + \mathrm{i}t\frac{\log^{1/4}N}{N^{2/3}}$, we represent the integral as

$$\frac{\mathrm{i}\log^{1/4}N}{N^{2/3}}\int_{-\infty}^{\infty}\exp\left\{\frac{N}{\alpha}\left[G\left(\hat{\gamma}+\mathrm{i}t\frac{\log^{1/4}N}{N^{2/3}}\right)-G\left(\hat{\gamma}\right)\right]\right\}\mathrm{d}t.$$

Using the Lagrange form of the remainder in the Taylor expansions of the real and imaginary parts of $G(z) - G(\hat{\gamma})$, we arrive at the following inequality

$$\left|G(z) - G\left(\hat{\gamma}\right) - G'\left(\hat{\gamma}\right)\left(z-\hat{\gamma}\right) - \frac{1}{2}G''\left(\hat{\gamma}\right)\left(z-\hat{\gamma}\right)^2\right| \le \frac{\left|z-\hat{\gamma}\right|^3}{3}\sup_{\zeta-\hat{\gamma}\in\mathrm{i}\mathbf{R}}\left|G'''\left(\zeta\right)\right|. \tag{3.16}$$

Note that the event $\mathcal{E}_0 = \{\hat{\gamma} - \lambda_1 > \frac{b^2}{2}N^{-2/3}\log N\}$ holds with probability arbitrarily close to one for all sufficiently large N. On $\mathcal{E}_0$, the latter supremum is no larger than $|G'''(\hat{\gamma})|$ because, for any $z \in (\hat{\gamma} - \mathrm{i}\infty, \hat{\gamma} + \mathrm{i}\infty)$, $|z - \lambda_j|^{-3} \leq (\hat{\gamma} - \lambda_j)^{-3}$. Hence, we have

$$\left| \tilde{G}(t) - G'(\hat{\gamma})\mathrm{i}t \frac{\log^{1/4} N}{N^{2/3}} + \frac{1}{2} G''(\hat{\gamma}) t^2 \frac{\log^{1/2} N}{N^{4/3}} \right| \leq \frac{|t|^3}{3} \frac{\log^{3/4} N}{N^2} |G'''(\hat{\gamma})|,$$

where

$$\tilde{G}(t) \equiv G\left(\hat{\gamma} + \mathrm{i}t \frac{\log^{1/4} N}{N^{2/3}}\right) - G(\hat{\gamma}).$$

The following lemma is established for Gaussian matrices in [JKOP21, Section 6.2] and extended to Wigner matrices in Section 3.5.4:

Lemma 3.6. *Denote the l-th derivative of $G(z)$ as $G^{(l)}(z)$. Then,*

$$G'(\hat{\gamma}) = \begin{cases} 2bN^{-1/3} \log^{1/2} N + o_{\mathbf{P}}\left(N^{-1/3} \log^{-1/4} N\right) & \text{for } b > 0, \\ o_{\mathbf{P}}\left(N^{-1/3} \log^{-1/4} N\right) & \text{for } b < 0, \end{cases}$$

$$G^{(l)}(\hat{\gamma}) = (-1)^l \frac{(2l-4)!}{(l-2)!} \left(\frac{N^{1/3}}{2\,|b| \log^{1/2} N}\right)^{2l-3} (1 + o_{\mathbf{P}}(1)), \qquad b \neq 0, l \geq 2. \quad \square$$

This lemma and the inequality that precedes it yield, for any fixed $C > 0$,

$$\int_{-C}^{C} \exp\left\{\frac{N}{\alpha} \tilde{G}(t)\right\} \mathrm{d}t = (1 + o_{\mathbf{P}}(1)) \int_{-C}^{C} \exp\left\{-\frac{t^2}{4\alpha |b|}\right\} \mathrm{d}t. \tag{3.17}$$

Further, for any $t \in \mathbf{R}$, by definition,

$$\begin{aligned} \operatorname{Re} \tilde{G}(t) &= -\frac{1}{N} \sum_{j=1}^{N} \log \left| 1 + \frac{\mathrm{i}t \log^{1/4} N}{N^{2/3} (\hat{\gamma} - \lambda_j)} \right| \\ &= -\frac{1}{2N} \sum_{j=1}^{N} \log \left(1 + \frac{t^2 \log^{1/2} N}{N^{4/3} (\hat{\gamma} - \lambda_j)^2} \right). \end{aligned}$$

We will use the elementary inequality $\log(1+\delta) \geq \delta/2$ for $\delta \in [0,1]$. Conditionally on $\mathcal{E}_0$, for all $|t| \leq t_N \equiv \frac{b^2}{2} \log^{3/4} N$, we have

$$N \operatorname{Re} \tilde{G}(t) < -\frac{t^2 \log^{1/2} N}{4N^{4/3}} \sum\nolimits_{j=1}^{N} (\hat{\gamma} - \lambda_j)^{-2} = -\frac{t^2 \log^{1/2} N}{4N^{1/3}} G''(\hat{\gamma}).$$

Using Lemma 3.6, we conclude that for all $|t| \le t_N \equiv \frac{b^2}{2}\log^{3/4} N$,

$$N\,\mathrm{Re}\,\tilde{G}(t) < -\frac{t^2}{8|b|}(1+o_{\mathbf{P}}(1)).$$

Therefore,

$$\int_{-t_N}^{-C}\left|\exp\left\{\frac{N}{\alpha}\tilde{G}(t)\right\}\right|\mathrm{d}t+\int_{C}^{t_N}\left|\exp\left\{\frac{N}{\alpha}\tilde{G}(t)\right\}\right|\mathrm{d}t < (1+o_{\mathbf{P}}(1))\frac{8|b|\alpha}{C}\exp\left\{-\frac{C^2}{8|b|\alpha}\right\}. \quad (3.18)$$

Since C can be chosen arbitrarily large, equations Eq. (3.17) and Eq. (3.18) yield

$$\int_{\hat{\gamma}-\mathrm{i}t_N N^{-2/3}\log^{1/4} N}^{\hat{\gamma}+\mathrm{i}t_N N^{-2/3}\log^{1/4} N}\exp\left\{\frac{N}{\alpha}[G(z)-G(\hat{\gamma})]\right\}\mathrm{d}z = 2\sqrt{\pi\alpha|b|}\frac{\mathrm{i}\log^{1/4} N}{N^{2/3}}\left(1+o_{\mathbf{P}}(1)\right). \quad (3.19)$$

It remains to show that the contribution of the remaining parts of the integral is negligible. Clearly, it is sufficient to prove that

$$\int_{t_N}^{\infty}\left|\exp\left\{\frac{N}{\alpha}\tilde{G}(t)\right\}\right|\mathrm{d}t = o_{\mathbf{P}}(N^{-k}) \quad (3.20)$$

for arbitrarily large fixed k. Note that $\mathrm{Re}\,\tilde{G}(t)$ is a strictly decreasing function of $t\in[t_N,\infty)$. Therefore,

$$\begin{aligned}\int_{t_N}^{N^2}\left|\exp\left\{\frac{N}{\alpha}\tilde{G}(t)\right\}\right|\mathrm{d}t &< N^2\exp\left\{\frac{N}{\alpha}\,\mathrm{Re}\,\tilde{G}(t_N)\right\}\\ &< N^2\exp\left\{-\frac{|b|^3\log^{3/2} N}{32\alpha}(1+o_{\mathbf{P}}(1))\right\}\\ &= o_{\mathbf{P}}(N^{-k})\end{aligned}$$

for arbitrarily large fixed k. For $t > N^2$, on the event $(\hat{\gamma}-\lambda_N)^2 < \bar{C}$ that holds with high probability for some constant $\bar{C}$, we have

$$\begin{aligned}N\,\mathrm{Re}\,\tilde{G}(t) &= -\frac{1}{2}\sum_{j=1}^{N}\log\left(1+\frac{t^2\log^{1/2} N}{N^{4/3}(\hat{\gamma}-\lambda_j)^2}\right)\\ &\le -\frac{N}{2}\log\left(\frac{t^2\log^{1/2} N}{N^{4/3}\bar{C}}\right).\end{aligned}$$

Therefore,

$$\int_{N^2}^{\infty} \left| \exp\left\{ \frac{N}{\alpha} \tilde{G}(t) \right\} \right| \mathrm{d}t < \int_{N^2}^{\infty} \left(\frac{t^2 \log^{1/2} N}{N^{4/3} \bar{C}} \right)^{-\frac{N}{2\alpha}(1+o_{\mathbf{P}}(1))} \mathrm{d}t = o_{\mathbf{P}}(N^{-k})$$

for arbitrarily large fixed k as well. Hence, Eq. (3.20) indeed holds.

Now we are ready to prove the following theorem. Recall that $F_{\alpha,N} = \frac{\alpha}{2N} \log Z_{\alpha,N}$, where $Z_{\alpha,N}$ is as defined in Eq. (3.6). We will omit subscript α from the notations, to simplify them.

Theorem 3.7 (Negative-critical regime)**.** *Suppose* $\beta = 1 + bN^{-1/3} \log^{1/2} N$ *with* $b < 0$. *Then,*

$$\frac{N}{\sqrt{\frac{\alpha}{12} \log N}} \left(F_N - \frac{1}{4}\beta^2 + \frac{\log N}{12N} \right) \xrightarrow{\mathrm{d}} \mathcal{N}(0, 1).$$

Proof. After rearranging Eq. (3.6), we have

$$2NF_N = \alpha \log C_N + NG(\hat{\gamma}) + \alpha \log \frac{1}{2\pi \mathrm{i}} \int_{\hat{\gamma}-\mathrm{i}\infty}^{\hat{\gamma}+\mathrm{i}\infty} \exp\left\{ \frac{N}{\alpha} [G(z) - G(\hat{\gamma})] \right\} \mathrm{d}z.$$

For the first term, using Stirling's formula,

$$\alpha \log C_N = \alpha \log \frac{\sqrt{2\pi} (N/\alpha)^{N/\alpha - 1/2} e^{-N/\alpha}}{(N\beta/\alpha)^{N/\alpha - 1}} + o(1) = -N(1 + \log\beta) + \frac{\alpha}{2} \log N + O(1). \quad (3.21)$$

For the second term we have

$$\begin{aligned} NG(\hat{\gamma}) &= N\beta\hat{\gamma} - \sum_{j=1}^{N} \log(\hat{\gamma} - \lambda_j) \\ &= 2\beta N + b^2 N^{1/3} \log N + b^3 \log^{3/2} N - \textstyle\sum_{j=1}^{N} \log(\hat{\gamma} - \lambda_j). \end{aligned}$$

For the third term, using Lemma 3.5,

$$\alpha \log \frac{1}{2\pi \mathrm{i}} \int_{\hat{\gamma}-\mathrm{i}\infty}^{\hat{\gamma}-\mathrm{i}\infty} \exp\left\{ \frac{N}{\alpha} [G(z) - G(\hat{\gamma})] \right\} \mathrm{d}z = -\frac{2\alpha}{3} \log N + O_{\mathbf{P}}(\log\log N).$$

Combining the three terms, we obtain

$$2NF_N = N(-1 - \log\beta + 2\beta) + b^2 N^{1/3} \log N + b^3 \log^{3/2} N - \frac{\alpha}{6} \log N$$

$$-\sum_{j=1}^{N}\log(\hat{\gamma}-\lambda_j)+O_{\mathbf{P}}(\log\log N).$$

Let

$$N\xi_N \equiv \sum_{j=1}^{N}\log(\hat{\gamma}-\lambda_j)-\frac{N}{2}-b^2N^{1/3}\log N+\frac{2}{3}|b|^3\log^{3/2}N+\frac{\alpha-1}{6}\log N.$$

On $\mathcal{E}_0$, we have that $\lambda_1 \le \hat{\gamma}$, and so $N\xi_N$ a.a.s. satisfies the conditions of Proposition 3.2 with $\sigma_N = b^2\log N$. This means that

$$N\xi_N\Big/\sqrt{\frac{\alpha}{3}\log N}\xrightarrow{\mathrm{d}}\mathcal{N}(0,1). \tag{3.22}$$

Combining the last two displays and noting that

$$b^3\frac{\log^{3/2}N}{N}=(\beta-1)^3,$$

we get (for $b<0$)

$$2NF_N = N\left(2\beta-\log\beta-\frac{3}{2}+\frac{1}{3}(\beta-1)^3-\frac{\log N}{6N}\right)-N\xi_N+O_{\mathbf{P}}\left(\log\log N\right).$$

Using the Taylor expansion

$$\log\beta=(\beta-1)-\frac{1}{2}(\beta-1)^2+\frac{1}{3}(\beta-1)^3+o(N^{-1})$$

in the previous display, we obtain

$$2NF_N=\frac{N}{2}\beta^2-\frac{\log N}{6}-N\xi_N+O_{\mathbf{P}}\left(\log\log N\right). \tag{3.23}$$

On the other hand, recalling Eq. (3.22) yields Theorem 3.7 and hence the negative critical part of Theorem 3.1. □

3.4 Positive-critical regime

The vertical contour passing through $\hat{\gamma}$ will not work when $b \ge 0$ because $G'(\hat{\gamma})$ becomes non-negligible. As a result, the function $G(z)$ oscillates quickly along the vertical contour

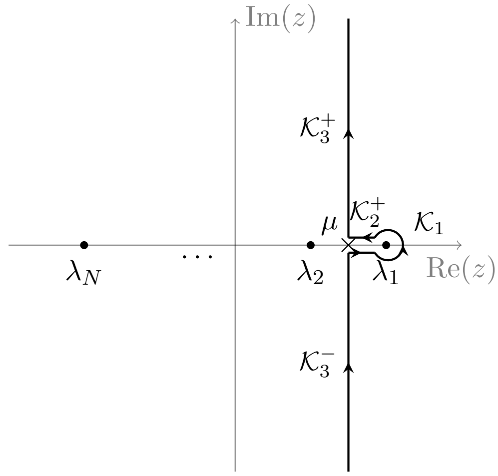


Figure 3.1: Contour of integration for positive b

near $\hat{\gamma}$. Instead we use contours crossing the real axis closer to λ_1. To this end, we consider the nonsingular part of G at λ_1 and define

$$\hat{G}(\lambda_1) = \beta\lambda_1 - \frac{1}{N}\sum_{j=2}^{N}\log(\lambda_1 - \lambda_j).$$

Proposition 3.8. *If $b \geq 0$, then for both $\alpha = 1, 2$,*

$$\frac{1}{2\pi \mathrm{i}}\int_{\mathcal{K}} \exp\{(N/\alpha)G(z)\}\,\mathrm{d}z = \exp\left\{(N/\alpha)\hat{G}(\lambda_1) - \frac{(\alpha-1)}{3}\log N + O_{\mathbf{P}}(\log\log N)\right\}.$$

For $b > 0$, we consider the vertical "keyhole contour" $\mathcal{K}$, Fig. 3.1, which is symmetric around the real axis and has the following form above the axis:

$$\mathcal{K}^+ = \mathcal{K}_1^+ \cup \mathcal{K}_2^+ \cup \mathcal{K}_3^+,$$

with $\mathcal{K}_1^+$ being a semi-circle with center at λ_1 and small radius ε, $\mathcal{K}_2^+$ being a horizontal segment connecting $\mu = \frac{\lambda_1+\lambda_2}{2}$ and $\lambda_1 - \varepsilon$, and $\mathcal{K}_3^+$ being a vertical ray starting from μ.

In the complex case, $\alpha = 1$, the integrand is analytic away from $\lambda_1, \ldots, \lambda_N$, and so the contributions of $\mathcal{K}_2^+$ and $\mathcal{K}_2^-$ cancel. On the other hand, in the real case, $\alpha = 2$, $\exp\{(N/\alpha)G(z)\}$ has a square-root-type singularity at $z = \lambda_1$. Hence, the contribution of

$\mathcal{K}_1$ to the integral $\int_{\mathcal{K}} \exp\{(N/\alpha)G(z)\}\,\mathrm{d}z$ converges to zero as $\varepsilon \to 0$. To summarize, let

$$I_N = \frac{1}{2\pi \mathrm{i}} \int_{\mathcal{K}} \exp\{(N/\alpha)G(z)\}\,\mathrm{d}z, \qquad I_{N,k} = \frac{1}{2\pi \mathrm{i}} \int_{\mathcal{K}_k} \exp\{(N/\alpha)G(z)\}\,\mathrm{d}z.$$

Thus, as $\varepsilon \to 0$ we have for both $\alpha = 1, 2$

$$I_N = I_{N,\alpha} + I_{N,3}.$$

Let

$$A_{N,\alpha} = \exp\{(N/\alpha)\hat{G}(\lambda_1) - \frac{\alpha - 1}{3} \log N\}.$$

When $b > 0$, we establish Proposition 3.8 in Section 3.4.1 by showing that

$$I_{N,\alpha} = A_{N,\alpha} \exp\{O_{\mathbf{P}}(\log\log N)\}, \qquad I_{N,3} = o_{\mathbf{P}}(I_{N,\alpha}).$$

The $b = 0$ case is more delicate. The keyhole contour yields both $I_{N,\alpha}, I_{N,3} = A_{N,\alpha} \exp\{O_{\mathbf{P}}(1)\}$ which suffices for the upper bound for I_N. Since the $O_{\mathbf{P}}(1)$ terms are in general complex, some cancellation between $I_{N,\alpha}$ and $I_{N,3}$ cannot be excluded, so further argument is needed for the lower bound. In Section 3.4.2 a separate argument using the steepest descent contour yields the required lower bound.

The following lemmas established for Gaussian matrices in [JKOP20, Proposition 4] and [JKOP21, Lemma 4.2] and extended to Wigner matrices in Sections 3.5.4 and 3.5.5 provides bounds for $\hat{G}'(\lambda_1)$ and $\hat{G}''(\lambda_1)$ used throughout the argument.

Lemma 3.9. *Let $\lambda_1 \geq \cdots \geq \lambda_N$ be the eigenvalues of a critically-spiked Gaussian matrix and let $C \in \mathbf{R}$. Then*

$$\frac{1}{N}\sum_{j=1}^{N} \frac{1}{2 + \lambda N^{-2/3} - \lambda_j} = 1 + O_{\mathbf{P}}(N^{-1/3}), \qquad \text{and} \qquad \frac{1}{N}\sum_{j=1}^{N} \frac{1}{(2 + \lambda N^{-2/3} - \lambda_j)^2} = O_{\mathbf{P}}(N^{1/3}).$$

Lemma 3.10. *Let $\lambda_1 \geq \cdots \geq \lambda_N$ be the eigenvalues of a subcritially spiked Wigner matrix. Then,*

$$\frac{1}{N}\sum_{j=2}^{N} \frac{1}{\lambda_1 - \lambda_j} = 1 + O_{\mathbf{P}}\left(N^{-1/3}\right), \qquad \text{and} \qquad \frac{1}{N}\sum_{j=2}^{N} \frac{1}{(\lambda_1 - \lambda_j)^2} = O_{\mathbf{P}}\left(N^{1/3}\right).$$

3.4.1 Proof of Proposition 3.8 for $b > 0$

Lemma 3.11. *Suppose that $b \geq 0$. Then for $\alpha = 1, 2$ we have*

$$I_{N,\alpha} = A_{N,\alpha} \exp\{O_{\mathbf{P}}(\log \log N)\}.$$

Proof. In the complex case, Cauchy's integral formula yields $I_{N,1} = \exp\{N\hat{G}(\lambda_1)\} \equiv A_{N,1}$, since $\mathcal{K}_1$ encircles only λ_1. The rest of this proof is devoted to the real case. First, consider

$$\frac{1}{2\pi \mathrm{i}} \int_{\mathcal{K}_2^+ \cup \mathcal{K}_2^-} \exp\{(N/2)G(z)\}\,\mathrm{d}z = \frac{1}{2\pi \mathrm{i}} \int_{\lambda_1}^{\mu} \frac{-\mathrm{i}}{\sqrt{\lambda_1 - y}} \exp\left\{\frac{N\beta y}{2} - \frac{1}{2}\sum_{j=2}^{N} \log(y - \lambda_j)\right\} \mathrm{d}y$$
$$+ \frac{1}{2\pi \mathrm{i}} \int_{\mu}^{\lambda_1} \frac{\mathrm{i}}{\sqrt{\lambda_1 - y}} \exp\left\{\frac{N\beta y}{2} - \frac{1}{2}\sum_{j=2}^{N} \log(y - \lambda_j)\right\} \mathrm{d}y.$$

Changing variables $y \mapsto x = \lambda_1 - y$, we obtain

$$\frac{1}{2\pi \mathrm{i}} \int_{\mathcal{K}_2^+ \cup \mathcal{K}_2^-} \exp\{(N/2)G(z)\}\,\mathrm{d}z = \frac{1}{\pi} \int_0^{\frac{\lambda_1 - \lambda_2}{2}} \frac{1}{\sqrt{x}} \exp\left\{\frac{N\beta(\lambda_1 - x)}{2} - \frac{1}{2}\sum_{j=2}^{N} \log(\lambda_1 - \lambda_j - x)\right\} \mathrm{d}x$$
$$= \exp\left\{(N/2)\hat{G}(\lambda_1)\right\} \mathcal{I},$$

where

$$\mathcal{I} = \frac{1}{\pi} \int_0^{\frac{\lambda_1 - \lambda_2}{2}} \frac{1}{\sqrt{x}} \exp\left\{-\frac{N}{2}\beta x - \frac{1}{2}\sum_{j=2}^{N} \log\left(1 - \frac{x}{\lambda_1 - \lambda_j}\right)\right\} \mathrm{d}x.$$

Since $0 \leq -\log(1-y) - y \leq y^2$ for $0 \leq y \leq \frac{1}{2}$, we have for some $\xi \in [0, 1]$,

$$-N\beta x - \sum_{j=2}^{N} \log\left(1 - \frac{x}{\lambda_1 - \lambda_j}\right) = -N\beta x + \sum_{j=2}^{N} \frac{x}{\lambda_1 - \lambda_j} + \xi \sum_{j=2}^{N} \frac{x^2}{(\lambda_1 - \lambda_j)^2}$$
$$= N^{2/3} x(-b\sqrt{\log N} + \omega_{1N}) + \xi N^{4/3} x^2 \omega_{2N}^+,$$

with random variables ω_{1N} and $\omega_{2N}^+ > 0$ both being $O_{\mathbf{P}}(1)$ from Lemma 3.10. Setting $y = N^{2/3}x$ and $\theta_3 = N^{2/3}(\lambda_1 - \lambda_2)/2 = \Theta_{\mathbf{P}}(1)$ (by Eq. (3.11)) and noting also that

$\omega_{1N} y + \xi \omega_{2N}^{+} y^2$ is uniformly $O_{\mathbf{P}}(1)$ for $0 \le y \le \theta_3$, we arrive at

$$\mathcal{I} = \frac{e^{O_{\mathbf{P}}(1)}}{N^{1/3}} \int_0^{\theta_3} \exp\{-\frac{1}{2} b y \sqrt{\log N}\} \frac{\mathrm{d}y}{\sqrt{y}} = \begin{cases} \dfrac{e^{O_{\mathbf{P}}(1)}}{N^{1/3}} \dfrac{1}{b^{1/2} \log^{1/4} N} & b > 0 \\ \dfrac{e^{O_{\mathbf{P}}(1)}}{N^{1/3}} & b = 0. \end{cases} \qquad \square$$

In what follows, we define $G(\mu) = \lim_{t \to +0} G(\mu + \mathrm{i}t)$, i.e., as a continuation from the upper-half plane, so that we have $\log(\lambda_1 - \mu) = \log|\lambda_1 - \mu| + \pi \mathrm{i}$.

Lemma 3.12. *For $b \ge 0$, we have $|I_{N,3}| \le A_{N,\alpha} \exp\left\{-\theta_N \frac{b\sqrt{\log N}}{\alpha} + O_{\mathbf{P}}(1)\right\}$, where $\theta_N = N^{2/3}(\lambda_1 - \lambda_2)/2$ is a non-negative $\Theta_{\mathbf{P}}(1)$ variable.*

Proof. It suffices to bound

$$I_{N,3}^{+} = \frac{1}{2\pi \mathrm{i}} \int_{\mathcal{K}_3^{+}} \exp\{(N/\alpha) G(z)\} \, \mathrm{d}z = \frac{1}{2\pi} \int_0^{\infty} \exp\{(N/\alpha) G(\mu + \mathrm{i}t)\} \, \mathrm{d}z,$$

as the analysis for $\mathcal{K}_3^{-}$ is analogous using $G(\bar{z}) = \overline{G(z)}$. Let $\tilde{G}(t) = G(\mu + \mathrm{i}t) - G(\mu)$. We have

$$|I_{N,3}^{+}| \le \frac{1}{2\pi} \exp\{(N/\alpha) \hat{G}(\lambda_1)\} |J_N| K_N, \tag{3.24}$$

with

$$J_N = \exp\{-(N/\alpha)[\hat{G}(\lambda_1) - G(\mu)]\}, \qquad K_N = \int_0^{\infty} \exp\{(N/\alpha) \operatorname{Re}[\tilde{G}(t)]\} \, \mathrm{d}t.$$

First we compare $\hat{G}(\lambda_1)$ and $G(\mu)$. Since $\log(\mu + \mathrm{i}0 - \lambda_1) = \log[(\lambda_1 - \lambda_2)/2] + \mathrm{i}\pi$, we have

$$N[\hat{G}(\lambda_1) - G(\mu)] = N\beta(\lambda_1 - \mu) + \log \frac{\lambda_1 - \lambda_2}{2} + \pi \mathrm{i} + \sum_{j=2}^{N} \log\left(1 - \frac{1}{2} \frac{\lambda_1 - \lambda_2}{\lambda_1 - \lambda_j}\right).$$

For $0 \le t \le 1$ we have that $|\log(1 - t/2) + t/2| \le t^2$. From Lemma 3.10 we then have

$$\sum_{j=2}^{N} \log\left(1 - \frac{1}{2} \frac{\lambda_1 - \lambda_2}{\lambda_1 - \lambda_j}\right) = -\frac{N(\lambda_1 - \lambda_2)}{2}[1 + O_{\mathbf{P}}(N^{-1/3})] + N(\lambda_1 - \lambda_2)^2 O_{\mathbf{P}}(N^{1/3}).$$

In addition, $2\theta_N := N^{2/3}(\lambda_1 - \lambda_2)$ is a $\Theta_{\mathbf{P}}(1)$ variable, and so $\log(\lambda_1 - \lambda_2) = -\frac{2}{3} \log N +$

$O_{\mathbf{P}}(1)$, and

$$\begin{aligned} N[\hat{G}(\lambda_1) - G(\mu)] &= \frac{N\beta}{2}(\lambda_1 - \lambda_2) - \frac{2}{3}\log N - \frac{N}{2}(\lambda_1 - \lambda_2) + O_{\mathbf{P}}(1) \\ &= \frac{N(\lambda_1 - \lambda_2)}{2}\frac{b\sqrt{\log N}}{N^{1/3}} - \frac{2}{3}\log N + O_{\mathbf{P}}(1) \\ &= -\frac{2}{3}\log N + \theta_N b\sqrt{\log N} + O_{\mathbf{P}}(1). \end{aligned} \tag{3.25}$$

Thus

$$|J_N| = \exp\{(2/3\alpha)\log N - \theta_N \frac{b\sqrt{\log N}}{\alpha} + O_{\mathbf{P}}(1)\}. \tag{3.26}$$

We turn to K_N in Eq. (3.24). Fix $k > \alpha$ and let $W_N = \max_{j\le k} N^{2/3}|\lambda_j - \mu|$. Neglecting negative terms, we have

$$N\,\mathrm{Re}\,\tilde{G}(t) = -\frac{1}{2}\sum_{j=1}^{N}\log\left(1 + \frac{t^2}{(\mu - \lambda_j)^2}\right) \le -\frac{k}{2}\log\left(1 + \frac{t^2}{W_N^2 N^{-4/3}}\right).$$

Since W_N is a $\Theta_{\mathbf{P}}(1)$ variable from TW convergence and Lemma 3.3 (iii), we have

$$\begin{aligned} K_N &\le \int_0^{+\infty}\left(1 + W_N^{-2}N^{4/3}t^2\right)^{-k/2\alpha}\mathrm{d}t \qquad (3.27) \\ &= W_N N^{-2/3}\int_0^{\infty}(1+s^2)^{-k/2\alpha}\,\mathrm{d}t = \exp\left\{-\frac{2}{3}\log N + O_{\mathbf{P}}(1)\right\}. \qquad (3.28) \end{aligned}$$

Combining Eq. (3.26) and Eq. (3.27), for each case $\alpha = 1, 2$ we arrive at

$$|J_N|K_N \le \exp\{-\frac{(\alpha-1)}{3}\log N - \theta_N\frac{b\sqrt{\log N}}{\alpha} + O_{\mathbf{P}}(1)\}.$$

Together with Eq. (3.24), this yields the lemma. □

3.4.2 The case $b = 0$

In this section we use the steepest descent contour to show that $I_N \ge A_{N,\alpha}\exp\{O_{\mathbf{P}}(\log\log N)\}$. When combined with the upper bound already established in Lemma 3.11, this yields Proposition 3.8 for $b = 0$.

Let Γ denote the contour of steepest descent of $G(z)$. For $z = x + \mathrm{i}y \in \Gamma$,

$$0 = \mathrm{Im}[G(z)] = y - \frac{1}{N}\sum_{j=1}^{N}\arg((x - \lambda_j) + \mathrm{i}y).$$

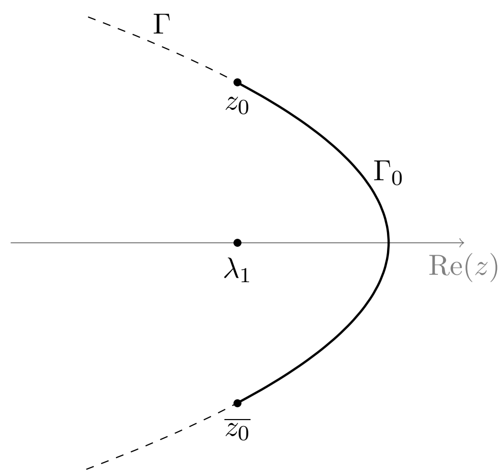


Figure 3.2: Curve of steepest descent of $G(z)$ near λ_1.

From this equation, we observe that Γ is symmetric around the real axis.

Next, observe that for a fixed imaginary part $y > 0$, $\arg((x-\lambda_j)+\mathrm{i}y)$ is strictly decreasing with x. Hence, $G(x+\mathrm{i}y)$ can have at most one root for any positive y. By symmetry around the real axis, this also holds for $y < 0$. This means that it is possible to parameterise

$$\Gamma = \{\Gamma(t) : 0 < t < 1\}$$

so that $\operatorname{Im}\Gamma(t)$ is increasing in t.

Moreover, we can see that $\Gamma(0^+) = -\infty - \mathrm{i}\pi$ and $\Gamma(1^-) = -\infty + \mathrm{i}\pi$. Therefore, Γ must have upper-bounded real part, and so

$$\int_{\mathcal{K}} \exp\{(N/\alpha)G(z)\}\,\mathrm{d}z = \int_{\Gamma} \exp\{(N/\alpha)G(z)\}\,\mathrm{d}z.$$

To continue, we need one last result about Γ, which formalizes the notion that Γ passes above λ_1 at a distance of roughly $N^{-2/3}$:

Lemma 3.13. *The function*

$$f(y) = \operatorname{Im}[G(\lambda_1 + \mathrm{i}y)] = y - \frac{\pi}{2N} - \frac{1}{N}\sum_{j=2}^{N} \arctan\Big(\frac{y}{\lambda_1 - \lambda_j}\Big)$$

has a unique positive root y_0. If $a_N \to \infty$ such that $a_N = o(N^{2/3})$, then a.a.s.

$$\frac{N^{-2/3}}{a_N} \le y_0 \le N^{-2/3} a_N. \tag{3.29}$$

Proof. Notice that, over $[0, \infty)$, f is convex with $f(0) = -\pi/(2N)$ and $\lim_{y\to\infty} f(y) = \infty$. In particular, this means that it has a unique positive root, which we will call y_0.

We will show that $f(N^{-2/3}a_N^{-1}) < 0 < f(N^{-2/3}a_N)$ a.a.s., which implies Eq. (3.29).

Let $y_- = N^{-2/3}a_N^{-1}$. Using $\arctan(x) \ge x - x^2/4$ for $x \ge 0$ and then Lemma 3.10, we have

$$\begin{aligned} f(y_-) &\le y_-\Big(1 - \frac{1}{N}\sum_{j=2}^{N}\frac{1}{\lambda_1 - \lambda_j}\Big) + \frac{y_-^2}{4N}\sum_{j=1}^{N}\frac{1}{(\lambda_1 - \lambda_j)^2} - \frac{\pi}{2N} \\ &= N^{-2/3}a_N^{-1}O_{\mathbf{P}}(N^{-1/3}) + \frac{1}{4}N^{-4/3}a_N^{-2}O_{\mathbf{P}}(N^{1/3}) - \frac{\pi}{2N} \\ &= -\frac{\pi}{2N} + o_{\mathbf{P}}(N^{-1}) \end{aligned}$$

Thus $f(y_-) < 0$ a.a.s.

Next, set $y_+ = N^{-2/3}a_N$. Now, some $y_+/(\lambda_1 - \lambda_j)$ terms will diverge to ∞ and so the linear approximation to arctan is not helpful. To handle these cases, we define

$$\begin{aligned} j^* &= \max\Big\{j : \frac{y_+}{\lambda_1 - \lambda_j} > 1 + \frac{\pi}{2}\Big\} \\ &= \#\Big\{j : \lambda_j > \lambda_1 - \Big(1 + \frac{\pi}{2}\Big)^{-1} a_N N^{-2/3}\Big\}. \end{aligned}$$

The significance of the $1 + \pi/2$ term is that $x - \arctan x \ge 1$ for x exceeding $1 + \pi/2$, and hence

$$\arctan\Big(\frac{y_+}{\lambda_1 - \lambda_j}\Big) \le \frac{y_+}{\lambda_1 - \lambda_j} - \mathbf{1}_{j \le j^*}.$$

We observe that, since $\lambda_1 - 2 \sim N^{-2/3} \ll a_N N^{-2/3}$, we have a.a.s.

$$j^* \ge j_0 = \#\Big\{j : \lambda_j > 2 - \frac{1}{3}a_N N^{-2/3}\Big\}.$$

Using part (iv) of Lemma 3.3 we have a.a.s. that $j^* \ge j_0 > C a_n^{3/2}$ for some $C > 0$.

Consequently

$$\begin{aligned} f(y_+) &= y_+ - \frac{1}{N}\sum_{j=2}^{N}\arctan\Big(\frac{y_+}{\lambda_1-\lambda_j}\Big) - \frac{\pi}{2N} \\ &\geq y_+ - \frac{1}{N}\sum_{j=2}^{N}\frac{y_+}{\lambda_1-\lambda_j} + \frac{j^*}{N} - \frac{\pi}{2N} \\ &\geq O_{\mathbf{P}}\Big(\frac{a_N}{N}\Big) + \frac{j^*}{N} - \frac{\pi}{2N}. \end{aligned}$$

Since, a.a.s., $j^* > Ca_N^{3/2}$, this means that $f(y_+) > 0$ for large enough N.

We have thus shown that for large N, with high probability, $f(y_-) < 0 < f(y_+)$, and the result follows. □

Having established necessary results about Γ, we also define $z_0 = \lambda_1 + \mathrm{i}y_0$ and

$$\Gamma_0 = \{z \in \Gamma : |\mathrm{Im}(z)| \leq y_0\}.$$

Since Γ can be parameterized with increasing imaginary part, this curve is connected.

Using the fact that $G(z)$ is purely real on Γ together with the parameterisation of Γ with increasing imaginary part, we have that

$$\begin{aligned} \frac{1}{2\pi\mathrm{i}}\int_\Gamma e^{(N/\alpha)[G(z)-\hat{G}(\lambda_1)]}\,\mathrm{d}z &\geq \frac{1}{2\pi}\int_{\Gamma_0} e^{(N/\alpha)\,\mathrm{Re}[G(z)-\hat{G}(\lambda_1)]}\,\mathrm{d}y \\ &\geq \frac{y_0}{\pi}e^{(N/\alpha)\,\mathrm{Re}[G(z_0)-\hat{G}(\lambda_1)]}, \end{aligned} \tag{3.30}$$

since the integrand is minimized on Γ_0 at the endpoints $z_0, \bar{z}_0 = \lambda_1 \pm \mathrm{i}y_0$.

Appealing again to Lemma 3.10, we obtain for $\alpha = 1, 2$

$$\begin{aligned} \log y_0 + (N/\alpha)\,\mathrm{Re}[G(z_0) - \hat{G}(\lambda_1)] &= \log y_0 - \frac{1}{\alpha}\log y_0 - \frac{1}{2\alpha}\sum_{j=2}^{N}\log\Big(1+\frac{y_0^2}{(\lambda_1-\lambda_j)^2}\Big) \\ &\geq \Big(1-\frac{1}{\alpha}\Big)\log y_0 - \frac{y_0^2}{2\alpha}\sum_{j=2}^{N}\frac{1}{(\lambda_1-\lambda_j)^2} \\ &= \Big(\frac{\alpha-1}{2}\Big)\log y_0 - \frac{y_0^2}{2\alpha}O_{\mathbf{P}}(N^{4/3}). \qquad (3.31) \\ &\geq -\Big(\frac{\alpha-1}{3}\Big)\log N + O_{\mathbf{P}}(\log\log N), \end{aligned}$$

since $N^{-2/3}/\log N \le y_0 \le N^{-2/3}\sqrt{\log\log N}$ a.a.s. according to Lemma 3.13.

Inserting this bound into Eq. (3.30), we obtain

$$\frac{1}{2\pi \mathrm{i}}\int_\Gamma e^{(N/\alpha)G(z)}\,\mathrm{d}z \ge \exp\Big\{(N/\alpha)\hat{G}(\lambda_1) - \Big(\frac{\alpha-1}{3}\Big)\log N + O_{\mathbf{P}}(\log\log N)\Big\},$$

which is the lower bound required to complete the proof.

3.4.3 Limiting law in the positive-critical regime

Theorem 3.14. *Suppose that $\beta = 1 + bN^{-1/3}\log^{1/2} N$ with $b \ge 0$. Then*

$$\frac{N}{\sqrt{\frac{\alpha}{12}\log N}}\left(F_N - \beta + \frac{1}{2}\log\beta + \frac{3}{4} + \frac{\log N}{12N}\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1) + \sqrt{\frac{3}{\alpha}}b\,\mathrm{TW}_{2/\alpha}$$

with independent $\mathcal{N}(0,1)$ and $\mathrm{TW}_{2/\alpha}$.

Proof. From Eq. (3.6) and Proposition 3.8 we have

$$2NF_N = \alpha\log C_N + N\hat{G}(\lambda_1) - \frac{\alpha(\alpha-1)}{3}\log N + O_{\mathbf{P}}(\log\log N). \tag{3.32}$$

The behavior of $N\hat{G}(\lambda_1)$ is governed by the approximation

$$\sum_{j=2}^{N}\log(\lambda_1-\lambda_j) = \sum_{j=2}^{N}\log|2-\lambda_j| + N(\lambda_1 - 2) + O_{\mathbf{P}}(1). \tag{3.33}$$

To verify this, let Δ_N denote the difference between right- and left-hand sides, without the error term. We set

$$\Delta_N = S_N + N(2-\lambda_1)\left[\frac{1}{N}\sum_{2}^{N}\frac{1}{\lambda_1-\lambda_j} - 1\right]$$

$$S_N = \sum_{j=2}^{N} X_{N,j}, \qquad X_{N,j} = \log|2-\lambda_j| - \log(\lambda_1-\lambda_j) - \frac{2-\lambda_1}{\lambda_1-\lambda_j}.$$

The second term of Δ_N is $O_{\mathbf{P}}(1)$ from Lemma 3.10 and the Tracy-Widom limit. For each fixed j, $X_{N,j} = O_{\mathbf{P}}(1)$ since both $|2-\lambda_j|$ and $\lambda_1-\lambda_j$ are $\Theta_{\mathbf{P}}(N^{-2/3})$, the latter by Lemma 3.3, part (iii).

To show that S_N is $O_{\mathbf{P}}(1)$, we use convergence criterion **C2** of Section 1.8. First, we argue that for each $\varepsilon > 0$, there exist $k = k(\varepsilon), C = C(\varepsilon) > 0$ such that the event

$$\mathcal{E}_{\varepsilon,N} = \{\lambda_1 \leq 2 + CN^{-2/3}, \lambda_k \leq 2 - CN^{-2/3}\}$$

has $P(\mathcal{E}_{\varepsilon,N}) > 1 - \varepsilon$ for large enough N. Indeed, Tracy-Widom convergence provides C_ε such that $\lambda_1 \leq 2 + C_\varepsilon N^{-2/3}$ with probability at least $1 - \varepsilon/2$. Lemma 3.3 part (ii) and Markov's inequality show that $P(\lambda_k \geq 2 - xN^{-2/3}) \leq C_x/k$. With $x = C_\varepsilon$, this can be made at most $\varepsilon/2$ by choosing $k(\varepsilon) = \lceil 2C_x/\varepsilon \rceil$. Hence $P(\mathcal{E}_{\varepsilon,N}) \geq 1 - \varepsilon$.

Let $S_{N,1}(\varepsilon), S_{N,2}(\varepsilon)$ denote the sum in S_N restricted to $j < k(\varepsilon)$ and $j \geq k(\varepsilon)$ respectively. On $\mathcal{E}_{\varepsilon,N}$, the sum $S_{N,1}(\varepsilon)$ has a finite number of $O_{\mathbf{P}}(1)$ terms and so is itself $O_{\mathbf{P}}(1)$. Also on $\mathcal{E}_{\varepsilon,N}$, observe that $(2-\lambda_1)/(\lambda_1-\lambda_j) \geq -\frac{1}{2}$ for all $j \geq k$. Since $|\log(1+x)-x| \leq C_1 x^2$ for $x > -\frac{1}{2}$ and some $C_1 > 0$, we have the bound

$$|S_{N,2}(\varepsilon)| \leq C_1(\lambda_1 - 2)^2 \sum_{j=2}^{N} \frac{1}{(\lambda_1 - \lambda_j)^2} = O_{\mathbf{P}}(1),$$

from Lemma 3.10 and the Tracy-Widom law. This completes the proof of Eq. (3.33).

Returning to $N\hat{G}(\lambda_1)$, using Eq. (3.33) and $\beta - 1 = bN^{-1/3}\log^{1/2} N$, we obtain the key decomposition

$$\begin{aligned} N\hat{G}(\lambda_1) &= 2N\beta + N\beta(\lambda_1 - 2) - \sum_{j=2}^{N} \log(\lambda_1 - \lambda_j) && (3.34) \\ &= 2N\beta - \sum_{2}^{N} \log|2 - \lambda_j| + b\sqrt{\log N} N^{2/3}(\lambda_1 - 2) + O_{\mathbf{P}}(1) \\ &= 2N\beta - \sum_{1}^{N} \log|2 - \lambda_j| - \frac{2}{3}\log N + b\sqrt{\log N}\xi_{2N} + O_{\mathbf{P}}(1), && (3.35) \end{aligned}$$

after adding and subtracting $\log|2-\lambda_1| = -\frac{2}{3}\log N + O_{\mathbf{P}}(1)$ and setting $\xi_{2N} = N^{2/3}(\lambda_1 - 2)$.

Combining this with Eq. (3.32) and Eq. (3.21), we obtain

$$2NF_N = N(-1 - \log\beta + 2\beta) - \frac{\alpha}{6}\log N - \sum_{1}^{N} \log|2 - \lambda_j| + b\sqrt{\log N}\xi_{2N} + O_{\mathbf{P}}(\log\log N),$$

where we note that the coefficient of $\log N$, namely $\frac{1}{2}\alpha - \frac{2}{3} - \frac{1}{3}\alpha(\alpha - 1)$, reduces to $-\frac{\alpha}{6}$ when $\alpha = 1$ or 2.

Let

$$N\check{\xi}_N = \sum_{j=1}^{N} \log|2-\lambda_j| - \frac{N}{2} + \frac{\alpha-1}{6}\log N.$$

Combining the two previous displays we obtain (compare Eq. (3.23))

$$2NF_N = N\left(-\frac{3}{2} - \log\beta + 2\beta\right) - \frac{\log N}{6} - N\check{\xi}_N + b\sqrt{\log N}\xi_{2N} + O_{\mathbf{P}}(\log\log N).$$

Now rewrite this as

$$NF_N = N\left(-\frac{3}{4} - \frac{1}{2}\log\beta + \beta - \frac{\log N}{12N}\right) + \sqrt{\frac{\alpha}{12}\log N}\,\xi_{1N} + \frac{b}{2}\sqrt{\log N}\,\xi_{2N} + O_{\mathbf{P}}(\log\log N), \tag{3.36}$$

where we set $\xi_{1N} = -N\check{\xi}_N/\sqrt{\frac{\alpha}{3}\log N}$.

By Proposition 3.17, (ξ_{1N}, ξ_{2N}) are asymptotically independent and $\xi_{1N} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$, whereas $\xi_{2N} \xrightarrow{\mathrm{d}} \mathrm{TW}_{2/\alpha}$. More precisely, there exist independent variables ζ_{1N}, ζ_{2N} such that $\xi_{jN} = \zeta_{jN} + o_{\mathbf{P}}(1)$. In this case marginal convergence $\xi_{jN} \xrightarrow{\mathrm{d}} \xi_j$ suffices for convergence $\xi_{1N} + c\xi_{2N} \xrightarrow{\mathrm{d}} \xi_1 + c\xi_2$. This completes the proof of Theorem 3.14 and thus of Theorem 3.1 in the positive critical case. □

3.5 Extension to spiked Wigner matrices

Here we extend the results that were proven above for the G(U/O)E cases to the Wigner case with a subcritical spike. The general strategy is much the same as in Section 2.3. Indeed, in this section we will make extensive use of Proposition 2.14 generally, and Proposition 2.10 to conclude results about the joint convergence of the largest eigenvalue and log-determinant. Throughout the remainder of this chapter we denote by W_N a matrix from GUE ($\alpha = 1$) or GOE ($\alpha = 2$) and by W_N' the corresponding real ($\alpha = 1$) or complex ($\alpha = 2$) Wigner matrix, that satisfies a specified subset of the conditions **W1**–**W4**.

In order to complete the proofs we also need to add a subcritical spike. For fixed $J \in [0, 1)$, consider the matrix

$$M_N = W_N' + J\mathbf{v}\mathbf{v}^*, \tag{3.37}$$

where $\mathbf{v}$ is arbitrary unit vector from $\mathbf{C}^N$ (from $\mathbf{R}^N$ if $\alpha = 2$). Below we denote by $\lambda_1 \geq \cdots \geq \lambda_N$ the eigenvalues of W_N, by $\lambda_1' \geq \cdots \geq \lambda_N'$ the eigenvalues of W_N, and by

$\mu_1 \geq \cdots \geq \mu_N$ the eigenvalues of M_N. We transfer the properties of W'_N to M_N using the Cauchy interlacing theorem

$$\mu_1 \geq \lambda'_1 \geq \mu_2 \geq \lambda'_2 \geq \cdots \geq \mu_N \geq \lambda'_N .$$

In addition, we will rely on the *stickiness* of the top eigenvalues of W'_N to its deformed counterpart M_N.

Proposition 3.15 (Stickiness of top eigenvalues)**.** *Suppose W'_N satisfies **W1** and **W3** and fix arbitrary $\varepsilon \in (0, 1/6)$. Let $J \in (0,1)$ and $M_N = W'_N + J\mathbf{v}\mathbf{v}^*$ for a unit vector $\mathbf{v}$. Then, w.o.p.,*

$$\max_{j \leq N^\varepsilon} |\mu_j - \lambda'_j| = O(N^{-1+2\varepsilon}).$$

The above bound is stated in [KY13b] for a constant number of top eigenvalues and extended to up to N^ε eigenvalues in [JKOP21, Section A.2].

3.5.1 Distribution of largest eigenvalues

Proposition 3.16. *Let $\mu_1 \geq \cdots \geq \mu_N$ be the eigenvalues of a subcritically-spiked M_N as in Eq.* (3.37) *and the Wigner matrix W'_N satisfies **W1-4**. Then, for any $k \in \mathbf{Z}_{>0}$,*

$$(N^{\frac{2}{3}}(\mu_1 - 2), \ldots, N^{\frac{2}{3}}(\mu_k - 2)) \xrightarrow{\mathrm{d}} (\mathrm{TW}_{\frac{2}{\alpha},1}, \ldots, \mathrm{TW}_{\frac{2}{\alpha},k}),$$

where $(\mathrm{TW}_{\frac{2}{\alpha},j})_{1\leq j\leq k}$ is the joint limiting distribution of the k largest eigenvalues for a GUE ($\alpha = 1$) or GOE ($\alpha = 2$).

Proof. Let $\xi_{N,j}(W) = N^{2/3}(\lambda_j(W) - 2)$. The convergence in distribution of $\boldsymbol{\xi}_N = \boldsymbol{\xi}_N(W_N)$ to $\boldsymbol{\xi} = (\mathrm{TW}_{\frac{2}{\alpha},1}, \ldots, \mathrm{TW}_{\frac{2}{\alpha},k})$ is established, e.g., in [AGZ09, Theorem 4.5.42]. We use the swapping corollary Proposition 2.10 to carry this over to convergence of $\boldsymbol{\xi}'_N = \boldsymbol{\xi}_N(W'_N)$: the key step is approximation by the Stieltjes functional $g(W, E)$ below, and then use of the derivative bounds in Proposition 2.14.

Introducing the rescaling $E(s) = 2 + N^{-2/3}s$, we may write

$$\mathbf{1}\{\xi_{N,j}(W) \geq s\} = \mathbf{1}\{\mathcal{N}_W(E(s), \infty) \geq j\}, \tag{3.38}$$

where $\mathcal{N}_W(E, \infty)$ denotes the number of eigenvalues of W that fall into the inverval (E, ∞).

Let $E_\infty = 2 + 2N^{-2/3+\varepsilon}, \eta = N^{-2/3-9\varepsilon}$ and

$$g(W, E) = \frac{N}{\pi} \int_E^{E_\infty} \operatorname{Im} s_W(y + \mathrm{i}\eta)\, \mathrm{d}y.$$

Under our assumptions on $W = W_N, W'_N$, Corollary 17.3 of [EY17] says that for $|E-2| \leq N^{-2/3+\varepsilon}$ and $\ell = \frac{1}{2}N^{-2/3-\varepsilon}$, and with overwhelming probability, we have inequalities

$$\mathcal{N}_W(E+\ell, \infty) - N^{-\varepsilon} \leq g(W, E) \leq \mathcal{N}_W(E-\ell, \infty) + N^{-\varepsilon}. \tag{3.39}$$

Let G_j be a smooth increasing function such that

$$G_j(x) = \begin{cases} 1 & \text{if } x \geq j - 1/3, \\ 0 & \text{if } x \leq j - 2/3. \end{cases}$$

From Eq. (3.39) we have w.o.p. for $W = W_N$ and W'_N that

$$\mathbf{1}\{\mathcal{N}_W(E+\ell, \infty) \geq j\} \leq G_j(g(W, E)) \leq \mathbf{1}\{\mathcal{N}_W(E-\ell, \infty) \geq j\}.$$

Applying this with $E = E(s_j) + \ell$ along with Eq. (3.38), we obtain

$$\mathbf{1}\{\xi_{N,j}(W) \geq s_j\} \geq G_j(g(W, E(s_j)+\ell)) \geq \mathbf{1}\{\xi_{N,j}(W) \geq s_j + N^{-\varepsilon}\}. \tag{3.40}$$

Setting $Q_j^+(W, s) = 1 - G_j(g(W, E(s_j)+\ell))$, we obtain bounds Eq. (2.13). Bounds Eq. (2.14) follow analogously with $Q_j^-(W, s) = 1 - G_j(g(W, E(s_j) - \ell))$.

The functions $Q_j^\pm(\cdot, s)$ satisfy Proposition 2.14 (2) with $\delta_{j,N} = N^{-1/3+O(\varepsilon)}$ and hence also condition F. Consequently the joint convergence for $\boldsymbol{\xi}'_N = \boldsymbol{\xi}_N(W'_N)$ follows from Proposition 2.10.

The result follows for subcritically-spiked Wigner matrix $M_N = W'_N + J\mathbf{v}\mathbf{v}^*$ applying Proposition 3.15 so that

$$\big(N^{\frac{2}{3}}(\mu_1 - 2), \ldots, N^{\frac{2}{3}}(\mu_k - 2)\big) = \big(N^{\frac{2}{3}}(\lambda'_1 - 2), \ldots, N^{\frac{2}{3}}(\lambda'_N - 2)\big) + o_{\mathbf{P}}(N^{-1/3+2\varepsilon}). \quad \square$$

3.5.2 Proof of Lemma 3.3

Let W_N, W'_N, M_N be as described in the introduction to Section 3.5. This proof will proceed by proving each part first for W_N, and then extending the result to W'_N and last to M_N.

Part (i): For each of W_N, W'_N and M_N, this follows from the convergence of $N^{2/3}(\lambda_1 - 2)$ to $\mathrm{TW}_{2/\alpha}$ shown in Proposition 3.16.

Part (ii): For GUE, this follows from the one-point function decay bound Eq. (3.10) and Eq. (3.8) applied to the counting function statistic built from $f_N(\lambda) = \mathbf{1}\{\lambda \geq 2 - xN^{-1/3}\}$. The extension to GOE follows from the comparison bound Corollary 2.6.

To extend this result to W'_N, we re-use several definitions from the proof of Proposition 3.16. In particular, for $\varepsilon > 0$, $E_\infty = 2 + 2N^{-2/3+\varepsilon}$ and $\eta = N^{-2/3-9\varepsilon}$, define

$$g(W, E) = N \int_E^{E_\infty} \operatorname{Im} s_W(u + \in \eta)\, \mathrm{d}y.$$

Following with $E = 2 - xN^{-2/3} - \ell$, we have that, w.o.p.,

$$\begin{aligned} \mathcal{N}_{W'_N}(2 - xN^{-2/3}, \infty) &\leq g(W'_N, 2 - xN^{-2/3} - \ell), \\ g(W_N, 2 - xN^{-2/3} - \ell) &\leq \mathcal{N}_{W_N}(2 - xN^{-2/3} - 2\ell, \infty). \end{aligned}$$

Moreover, since each of $\mathcal{N}_{W'_N}(2 - xN^{-2/3})$ and $\mathcal{N}_{W_N}(2 - xN^{-2/3})$ are almost surely bounded by N, it follows that, for any $A > 0$,

$$\begin{aligned} \mathbf{E}\mathcal{N}_{W'_N}(2 - xN^{-2/3}, \infty) &\leq \mathbf{E}g(W'_N, 2 - xN^{-2/3}) + O(N^{-A}), \\ \mathbf{E}g(W_N, 2 - xN^{-2/3} - \ell) &\leq \mathbf{E}\mathcal{N}_{W_N}(2 - xN^{-2/3} - 2\ell, \infty) + O(N^{-A}). \end{aligned}$$

Last, we apply Proposition 2.14 part (3) to conclude that g satisfies condition $F(\delta_N)$ with $\delta_N = O(N^{-1/3} + O(\varepsilon))$ and so, by Proposition 2.9,

$$\mathbf{E}g(W'_N, 2 - xN^{-2/3}) \leq \mathbf{E}g(W_N, 2 - xN^{-2/3}) + O(N^{-1/3+O(\varepsilon)}).$$

This completes the inequalities and renders

$$\begin{aligned} \mathbf{E}\mathcal{N}_{W'_N}(2 - xN^{-2/3}, \infty) &\leq \mathbf{E}\mathcal{N}_{W_N}(2 - (x + N^{-\varepsilon})N^{-2/3}, \infty) + o(1) \\ &\lesssim \mathbf{E}\mathcal{N}_{W_N}(2 - (x+1)N^{-2/3}, \infty) \\ &\leq C_{x+1} \end{aligned}$$

By interlacing of the eigenvalues of W_N' and M_n, we have that

$$\mathbf{E}\mathcal{N}_{M_N}(2 - xN^{-2/3}, \infty) \leq \mathbf{E}[\mathcal{N}_{W_N'}(2 - xN^{-2/3}, \infty) + 1] \lesssim C_{x+1} + 1.$$

Part (iii): Consider the operator

$$\mathcal{H}_{2/\alpha} = -\frac{\mathrm{d}^2}{\mathrm{d}x^2} + x + \sqrt{2\alpha}B'(x),$$

where $B'(x)$ is the derivative of the Brownian motion on $(0, \infty)$, and the operator acts on some Hilbert space $\mathcal{L}_*$, which consists of continuous functions supported on $(0, \infty)$, see p. 308 in [AGZ09]. The following result is Theorem 4.5.42 in [AGZ09] for the special case of just the two top eigenvalues,

$$(N^{2/3}(\lambda_1 - 2), N^{2/3}(\lambda_2 - 2)) \xrightarrow{\mathrm{d}} (-\Lambda_0, -\Lambda_1), \tag{3.41}$$

where Λ_0, Λ_1 are the bottom two eigenvalues of random operator $\mathcal{H}_{2/\alpha}$. In addition, in [AGZ09, Lemma 4.5.47] it is shown that the operator has simple spectrum with probability one. This implies (iii) for W_N.

But we know from Proposition 3.16 that (λ_1', λ_2') and (μ_1, μ_2) also have the same limiting distribution, so the result also holds for W_N' and M_N.

Part (iv): Lemma 2.2 of [Gus05] yields that, in the GUE case,

$$\mathbf{E}\#\{j : \lambda_j > 2 - b_N N^{-2/3}\} \gtrsim b_N^{3/2},$$

while lemma 2.3 of the same paper shows that

$$\mathrm{Var}(\#\{j : \lambda_j > 2 - b_N N^{-2/3}\}) \lesssim \log b_N.$$

The function $f_N(\lambda) = \mathbf{1}\{\lambda \geq 2 - b_N N^{-2/3}\}$ has $\mathrm{TV}(f) = 1$, and Corollary 2.6 yield mean and variance bounds of the same order in the GOE case.

A Chebyshev bound completes the proof for W_N, yielding in either case that, for some constant C,

$$\mathbf{P}(\#\{j : \lambda_j > 2 - b_N N^{-2/3}\} \leq C b_N^{2/3}) \to 0.$$

To extend this result to W'_N, let C be such that a.a.s.,

$$\#\{j : \lambda_j - b_N N^{-2/3}\} \gtrsim b_N^{3/2}.$$

Assume without loss of generality that $Cb_N^{3/2} \in \mathbf{Z}$.

Let G be a smooth increasing function such that

$$G(x) = \begin{cases} 1 & \text{if } x \geq Cb_N^{3/2} + 1/2, \\ 0 & \text{if } x \leq Cb_N^{3/2}. \end{cases}$$

According to Proposition 2.14(3), this satisfies condition $F(\delta_N)$ with $\delta_N = O(N^{-1/3+O(\varepsilon)})$.

Since $|(2 - b_N N^{-2/3}) - 2| \leq 2N^{-2/3+\varepsilon}$, we have by the equivalent of Eq. (3.40) together with Proposition 2.9 that

$$\begin{aligned} &\mathbf{P}(\mathcal{N}_{W'_N}(2 - b_N N^{-2/3}, \infty) \geq Cb_N^{3/2}) \\ &\quad \geq \mathbf{E}G(g(W'_N, 2 - b_N N^{-2/3} + \ell)) + o(1) \\ &\quad \geq \mathbf{E}G(g(W_N, 2 - b_N N^{-2/3} + \ell)) + O(N^{-1/3+O(\varepsilon)}) \\ &\quad \gtrsim \mathbf{P}(\mathcal{N}_{W_N}(2 - b_N N^{-2/3} + 2\ell, \infty) \geq Cb_N^{3/2} + 1/2) \\ &\quad = \mathbf{P}(\mathcal{N}_{W_N}(2 - (b_N - 2N^{-\varepsilon})N^{-2/3}, \infty) \geq C(b_N - 2N^{-\varepsilon})N^{-2/3}) \\ &\quad \to 1, \end{aligned}$$

where the inequality in the second-last line follows from the fact that $Cb_N^{3/2} \in \mathbf{Z}$.

The spiked case follows from the fact that each $\mu_j \geq \lambda'_j$, and so

$$\#\{j : \mu_j - b_N N^{-2/3}\} \geq \#\{j : \lambda'_j - b_N N^{-2/3}\} \gtrsim b_N^{3/2}.$$

3.5.3 Asymptotic independence of log-determinant and top eigenvalue

Proposition 3.17 (Asymptotic independence)**.** *Let W'_N be a Wigner matrix satisfying **W1-4** and M_N is as in Eq.* (3.37). *Define*

$$\begin{aligned} \xi_{1N}(W) &= \tau_N^{-1}(-\log|\det(W-2)| + \mu_N), \\ \xi_{2N}(W) &= N^{\frac{2}{3}}(\lambda_1(W) - 2). \end{aligned}$$

Then, $\xi_{1N}(M_N)$ and $\xi_{2N}(M_N)$ are asymptotically independent with limiting distribution given by

$$(\xi_{1N}(M_N), \xi_{2N}(M_N)) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1) \times \mathrm{TW}_{2/\alpha}\,.$$

Proof. First we consider the case where $J = 0$, i.e. with W'_N in place of M_N. In this case, this is an immediate consequence of Proposition 2.10 and previous arguments for the log-determinant in Theorem 2.1 and the largest eigenvalue in Proposition 3.16. Indeed, in the proof of Proposition 2.17, we show that

$$\xi_{1N}(W'_N) = \tau_N^{-1}(g_0(W'_N) - \bar{\mu}_N) + o_{\mathbf{P}}(1),$$

where for $\gamma_N = N^{-2/3-2\varepsilon}$ with $\varepsilon > 0$ small

$$g_0(W) = \int_{\gamma_N}^{N^{100}} \operatorname{Im} s_W(2 + i\eta)\,\mathrm{d}\eta, \qquad \bar{\mu}_N = \mu_N + N\log(N^{100}).$$

It is enough to consider joint convergence of

$$\tilde{\xi}_{1N}(W) = \tau_N^{-1}(g_0(W) - \bar{\mu}_N) \qquad \text{and} \qquad \xi_{2N},$$

since $(\xi_{1N}(W'_N), \xi_{2N}(W'_N)) = (-\tilde{\xi}_{1N}(W'_N), \xi_{2N}(W'_N)) + o_{\mathbf{P}}(1)$.

In Proposition 2.10 we set $Q_1^{\pm}(W,s) = G_s^{\pm}(g_0(W))$, and $Q_2^{\pm}(W,s) = 1 - G_1(g(W, E(s) \pm \ell))$. We use Eq. (2.36) and Eq. (3.40) and their analogs for Q_j^- to establish Eq. (2.13), Eq. (2.14), and conditions $F(\delta_{j,N})$ follow from the respective previous arguments.

Convergence of the G(U/O)E versions $(\xi_{1N}, \xi_{2N}) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1) \times TW_{2/\alpha}$ is established in Proposition 3.4, and the conclusion for (ξ'_{1N}, ξ'_{2N}) follows now from Proposition 2.10.

As for the spiked case, eq. (95) of [JKOP20] shows that $\xi_{1N}(M_N) = \xi_{1N}(W'_N) + o_{\mathbf{P}}(1)$ for a fixed $J \in (0,1)$. Moreover, thanks to the stickiness property of Proposition 3.15, we also have $\xi_{2N}(M_N) = \xi_{2N}(W'_N) + o_{\mathbf{P}}(1)$. Therefore, the limiting distribution of $(\xi_{1N}(M_N), \xi_{2N}(M_N))$ does not change as long as the spike is subcritical.

3.5.4 Inverse moments

Proposition 3.18. *Let W'_N be a Wigner matrix satisfying **W1-4** and M_N be as in Eq.* (3.37).

(a) (Wigner extension of Lemma 3.6) Let $b \neq 0$, and define $\hat{\gamma} = 2 + b^2 N^{-\frac{2}{3}} \log N$. For

$\beta = 1 + bN^{-\frac{1}{3}}\sqrt{\log N}$, *define the function*

$$G_{M_N}(z) = \beta z - \frac{1}{N}\sum_{j=1}^{N}\log(z-\mu_j).$$

Then, for $b \neq 0$ *and* $l \geq 1$,

$$G^{(l)}_{M_N}(\hat{\gamma}) = \begin{cases} 2b_+ N^{-\frac{1}{3}}\log^{\frac{1}{2}} N + o_{\mathbf{P}}(N^{-\frac{1}{3}}\log^{-\frac{1}{4}} N) & \text{if } l = 1, \\ (-1)^l \frac{(2l-4)!}{(l-2)!}\Big(\frac{N^{\frac{1}{3}}}{2|b|\log^{\frac{1}{2}} N}\Big)^{2l-3}(1+o_{\mathbf{P}}(1)) & \text{if } l \geq 2. \end{cases} \tag{3.42}$$

(b) (Wigner extension of Lemma 3.9). Let $C \in \mathbf{R}$ *be fixed. Then*

$$\frac{1}{N}\sum_{j=1}^{N}\frac{1}{2+CN^{-\frac{2}{3}}-\mu_j} = 1 + O_{\mathbf{P}}(N^{-1/3}), \quad \textit{and} \quad \frac{1}{N}\sum_{j=1}^{N}\frac{1}{(2+CN^{-\frac{2}{3}}-\mu_j)^2} = O_{\mathbf{P}}(N^{1/3}). \tag{3.43}$$

First we rewrite Eq. (3.42) and Eq. (3.43) in terms of Stieltjes transforms. Since $G_W(z) = \beta z - N^{-1}\sum_1^N \log(z-\lambda_j(W))$, we have for $l \geq 1$,

$$G^{(l)}(z) = \beta\mathbf{1}(l=1) + s_W^{(l-1)}(z).$$

Suppose as usual that $|E-2| \lesssim \check{\sigma}_N N^{-2/3}$ and define

$$g_0(W) = N^{-2l/3+1}s_W^{(l-1)}(E) = (l-1)!N^{-2l/3}\sum_{j=1}^{N}(\lambda_j - E)^{-l}.$$

For appropriate constants μ_N, σ_N and random variables Z_N, Eq. (3.42) and Eq. (3.43) can then be written as

$$g_0(W_N) = \mu_N + \sigma_N Z_N, \qquad Z_N = o_{\mathbf{P}}(1) \text{ or } O_{\mathbf{P}}(1). \tag{3.44}$$

Indeed, in the case of Eq. (3.42), with $E = \hat{\gamma}$, in Eq. (3.44) we have

$$\mu_N = -N^{1/3}\mathbf{1}(l=1) + c_l|b|^{3-2l}\log^{-(l-3/2)} N, \qquad \sigma_N = \log^{-L} N, \qquad L = \begin{cases} 1/4 & l = 1 \\ l - 3/2 & l \geq 2 \end{cases}$$

and $Z_N = o_{\mathbf{P}}(1)$, with $c_1 = 1, c_2 = 1/2$ and with the general form of c_l visible in Eq. (3.42).

In the case of Eq. (3.43), we take $E = 2 + CN^{-2/3}$, and for $l = 1, 2$ in Eq. (3.44) we have

$$\mu_N = -N^{1/3}\mathbf{1}(l = 1), \qquad \sigma_N = 1, \qquad Z_N = O_{\mathbf{P}}(1).$$

Thus Lemma 3.6 and [JKOP20, Proposition 3] establish the validity of Eq. (3.44) for W_N drawn from G(U/O)E. We wish to carry this over to $Z'_N = (g_0(W'_N) - \mu_N)/\sigma_N$ for W'_N a Wigner matrix satsifying **W1-4**. To do this, we approximate $g_0(W)$ by the Stieltjes functional

$$g(W) = N^{-2l/3+1} \operatorname{Re} s_W^{(l-1)}(E + \mathrm{i}\eta).$$

Lemma 3.19 (Approximation step). *Let W be an $N \times N$ Wigner matrix satisfying **W1-3** and let $E \in \mathbf{R}$ be such that $|E - 2| \le N^{-\frac{2}{3}}\check{\sigma}_N$. Let $\varepsilon > 0$ and define $\eta = N^{-\frac{2}{3}-3\varepsilon}$.*

Then, for all $l \in \mathbf{Z}_{>0}$, we have with high probability that

$$N^{-2l/3+1} s_W^{(l-1)}(E) = N^{-2l/3+1} \operatorname{Re} s_W^{(l-1)}(E + \mathrm{i}\eta) + O(N^{-\varepsilon}). \tag{3.45}$$

Proof. Let $\varepsilon_0 = \varepsilon/(l + 1)$. By eigenvalue non-concentration, there then exists a constant $d > 0$ such that the event

$$E_N = \left\{\min_{1\le j\le N} |\lambda_j - E| \ge N^{-\frac{2}{3}-\varepsilon_0}\right\}$$

holds with probability at least $1 - N^{-d}$. The rest of the argument occurs on the event E_N.

Now, the function $\sum_{j=1}^N \frac{1}{(z-\lambda_j)^l}$ is holomorphic in the open disk $\{z : |z - E| < N^{-\frac{2}{3}-\varepsilon_0}\}$. Since $\varepsilon_0 < \varepsilon$, the vertical segment γ connecting E to $E + \mathrm{i}\eta$ lies entirely within this disk, so the fundamental theorem of calculus applies, rendering

$$\begin{aligned}\left|\operatorname{Re}\sum_{j=1}^N \frac{1}{(E - \lambda_j + \mathrm{i}\eta)^l} - \sum_{j=1}^N \frac{1}{(E - \lambda_j)^l}\right| &= \left|\operatorname{Re}\int_\gamma \sum_{j=1}^N -\frac{l}{(z - \lambda_j)^{l+1}}\,\mathrm{d}z\right| \\ &\le l\eta \sum_{j=1}^N \frac{1}{|E - \lambda_j|^{l+1}}.\end{aligned}$$

By Lemma 2.16, this is $O(N^{-\frac{2}{3}-3\varepsilon} \cdot N^{(\frac{2}{3}+\varepsilon_0)(l+1)+\varepsilon)}) = O(N^{\frac{2}{3}l-\varepsilon})$ w.o.p. on E_N, from which the result follows. □

Lemma 3.19 says that $g(W_N)$ satisfies Eq. (3.44) exactly when $g_0(W_N)$ does. So we carry out the Lindeberg swapping with $g(W)$.

Let $\kappa > 0$ and $H : \mathbf{R} \to [0,1]$ be a smooth cutoff function satisfying

$$H(x) = \begin{cases} 1 & \text{if } |x| \le \kappa \\ 0 & \text{if } |x| \ge 2\kappa. \end{cases}$$

Let $G(x) = H((x-\mu_N)/\sigma_N)$, so that $\|G^{(j)}\|_\infty \lesssim b_N^j = \sigma_N^{-j}$. We apply Proposition 2.14(4) to G with $b_N = \sigma_N^{-1}$ and to g with $a_N = N^{-1/3+O(\varepsilon)}$ and hence $\delta_N = N^{-1/3+O(\varepsilon)}$, so that Proposition 2.9 yields $\mathbf{E}G(g(W_N')) = \mathbf{E}G(g(W_N)) + O(\delta_N)$. Write $\check{Z}_N = (g(W_N) - \mu_N)/\sigma_N$ and similarly for $\check{Z}_N'$. We conclude that

$$\begin{aligned} \mathbf{P}(|\check{Z}_N'| > \kappa) &= \mathbf{P}(|g(W_N') - \mu_N| > \kappa\sigma_N) \\ &\le \mathbf{E}G(g(W_N')) \le \mathbf{E}G(g(W_N)) + O(\delta_N) \\ &\le \mathbf{P}(|\check{Z}_N| > 2\kappa) + O(\delta_N). \end{aligned}$$

A similar bound holds reversing the roles of W_N and W_N'. Consequently $\check{Z}_N'$ is $o_{\mathbf{P}}(1)$ or $O_{\mathbf{P}}(1)$ exactly when $\check{Z}_N$ is. From Lemma 3.19, both $\check{Z}_N - Z_N$ and $\check{Z}_N' - Z_N'$ are $O(N^{-\varepsilon})$ with high probability. Thus Eq. (3.44) carries over to W_N' and so Proposition 3.18 is established.

Inverse moments for the spiked case. Suppose that $\tilde{W}_N = W_N + J\mathbf{v}\mathbf{v}^*$ has a spike with a value $J \in (0,1)$. Let us show that Eq. (3.42)and Eq. (3.43) still hold for this case. Let μ_j denote the eigenvalues of $\tilde{W}_N$ in the descending order, and λ_i are the eigenvalues of W_N.

Let γ equals either $\hat{\gamma}$ from Eq. (3.42) or $2 + CN^{-2/3}$ from Eq. (3.43). In addition, let i^* denotes the index of the nearest to γ among eigenvalues λ_i. Due to the interlacing property, we have that $0 \le \gamma - \mu_i \le \gamma - \lambda_i$ for $i > i^*$ and $\gamma - \mu_i \le \gamma - \lambda_i \le 0$ for $i < i^*$.

Then,

$$\sum_{i=1}^N (\gamma - \mu_i)^{-l} \ge \sum_{i=1}^N (\gamma - \lambda_i)^{-l} - (\gamma - \lambda_{i^*})^{-l} + (\gamma - \mu_{i^*})^{-l}.$$

Using the classical eigenvalue rigidity results [see, e.g. BK18, Theorem 2.9] to count the number of eigenvalues with index at lest i^*, we find that for any $\varepsilon > 0$ w.o.p.

$$O(N^{-2/3}\log N) = \left(\frac{i^*}{N}\right)^{2/3} + O\left(N^{-2/3+\varepsilon}(i^*)^{-1/3}\right),$$

which implies $i^* = O(N^\varepsilon)$. Therefore, using Proposition 3.15 we have that w.o.p. $|\mu_{i^*} - \lambda_{i^*}| \lesssim N^{-1+3\varepsilon}$. Furthermore, by the non-concentration result from Proposition 2.15, with high probability, $|\gamma - \lambda_{i^*}|^{-l} \leq N^{2l/3+l\varepsilon}$. Hence we obtain that, with high probability,

$$
\begin{aligned}
(\gamma - \lambda_i^*)^{-l} - (\gamma - \mu_{i^*})^{-l} &= (\gamma - \lambda_i^*)^{-l}\left[1 - \left(1 + \frac{\lambda_{i^*} - \mu_{i^*}}{\gamma - \lambda_{i^*}}\right)^{-l}\right] \\
&= O(N^{2l/3+l\varepsilon})\left[1 - \left(1 + O(N^{-1/3+C\varepsilon})\right)^{-l}\right] \\
&= O(N^{(2l-1)/3+C\varepsilon}).
\end{aligned}
$$

Taking ε sufficiently small, we obtain that for any $L > 0$,

$$
\sum_{i=1}^{N}(\gamma - \mu_i)^{-l} \geq \sum_{i=1}^{N}(\gamma - \lambda_i)^{-l} + o_{\mathbf{P}}(N^{2l/3}\log^{-L} N)\,.
$$

The inequality in the opposite direction can be obtained similarly by using $\mu_i \leq \lambda_{i+1}$ and additionally observing that $(\gamma - \mu_1)^{-l} = O_{\mathbf{P}}(1)$.

It is easy to make sure that the difference of $o_{\mathbf{P}}(N^{2l/3}\log^{-L} N)$ between the spiked statistics and non-spiked one is sufficient for Eq. (3.42)and Eq. (3.43) to hold in the spiked case as well.

3.5.5 Proof of Lemma 3.10

Let us first consider the case of no spike, $J = 0$. To this end, let $\{\lambda_j\}$ be the eigenvalues of a Wigner matrix satisfying **W1-4**. We rely on Proposition 3.18(b): for any fixed $C \in \mathbf{R}$,

$$
\frac{1}{N}\sum_{j=1}^{N}(2 - \lambda_j)^{-1} = 1 + O_{\mathbf{P}}(N^{-1/3}) \qquad \text{and} \qquad \frac{1}{N}\sum_{j=1}^{N}(2 - CN^{-2/3} - \lambda_j)^{-2} = O_{\mathbf{P}}(N^{1/3}). \tag{3.46}
$$

By the Cauchy-Schwarz inequality,

$$
\begin{aligned}
\left|\frac{1}{N}\sum\nolimits_{j=2}^{N}\frac{1}{\lambda_1 - \lambda_j} - \frac{1}{N}\sum\nolimits_{j=2}^{N}\frac{1}{2 - \lambda_j}\right| &= \left|\frac{1}{N}\sum\nolimits_{j=2}^{N}\frac{2 - \lambda_1}{(2 - \lambda_j)(\lambda_1 - \lambda_j)}\right| \\
&\leq |2 - \lambda_1|\left(\frac{1}{N}\sum\nolimits_{j=2}^{N}\frac{1}{(2 - \lambda_j)^2}\right)^{1/2}\left(\frac{1}{N}\sum\nolimits_{j=2}^{N}\frac{1}{(\lambda_1 - \lambda_j)^2}\right)^{1/2}.
\end{aligned}
$$

The bounds in Eq. (3.46) with $C = 0$, along with Tracy-Widom convergence $|\lambda_1 - 2| = O_{\mathbf{P}}(N^{-2/3})$ show that to establish Lemma 3.10 for the case $J = 0$, it is sufficient to show that $\frac{1}{N}\sum_{j=2}^{N} \frac{1}{(\lambda_1-\lambda_j)^2} = O_{\mathbf{P}}(N^{1/3})$.

For this, we use convergence criterion **C2** of Section 1.8. For each $\varepsilon > 0$, Tracy-Widom convergence, Lemma 3.3 (i), yields a constant C such that event $\mathcal{E}_{N,\varepsilon} = \{\lambda_1 > 2 - CN^{-2/3}\}$ has probability at least $1 - \varepsilon$ for large N. On this event,

$$\lambda_1 - \lambda_j \geq \begin{cases} 2 - CN^{-2/3} - \lambda_j & \text{if } \lambda_j \leq 2 - CN^{-2/3}, \\ \lambda_1 - \lambda_2 & \text{if } \lambda_j > 2 - CN^{-2/3}. \end{cases}$$

The number variable $\chi_N(C) = \#\{j \,:\, \lambda_1 > 2 - CN^{-2/3}\} = O_{\mathbf{P}}(1)$ and $\lambda_1 - \lambda_2 = \Theta_{\mathbf{P}}(N^{-2/3})$ by Lemma 3.3 parts (ii) and (iii) respectively. Using also Eq. (3.46) we obtain, on $\mathcal{E}_{N,\varepsilon}$,

$$\frac{1}{N}\sum_{j=2}^{N} \frac{1}{(\lambda_1 - \lambda_j)^2} \leq \frac{1}{N}\sum_{j=1}^{N} \frac{1}{(2 - CN^{-2/3} - \lambda_j)^2} + \frac{\chi_N(C)}{N(\lambda_1 - \lambda_2)^2} = O_{\mathbf{P}}(N^{1/3}).$$

This completes the proof of Lemma 3.10 for $J = 0$.

For $J \in (0, 1)$, the lemma follows from the interlacing inequalities that link the eigenvalues of $\tilde{W}_N = W_N + J\mathbf{v}\mathbf{v}^*$ and W_N, and from the fact that any finite number of the largest eigenvalues of $\tilde{W}_N$ are asymptotically distributed according to the multivariate Tracy-Widom law (of type one for GOE and of type two for GUE), established in Proposition 3.16.

Chapter 4

Limiting likelihood ratio

4.1 Introduction

In [LP21], the authors introduced a function that they called the stochastic Airy function, and demonstrated that it captured the limiting behavior of the characteristic function of a Gaussian matrix at the scale $2 + O(N^{-2/3})$.

This function of a real parameter t and complex parameter λ was defined implicitly as the unique solution to the SDE

$$\mathrm{d}\phi_\lambda'(t) = (t+\lambda)\phi_\lambda(t)\,\mathrm{d}t + \phi_\lambda(t)\,\mathrm{d}B(t)$$

over $L^2([0,\infty))$, where B is a Brownian motion with $\mathbf{E}B(t)^2 = 2\alpha t$.

More precisely, we have the following pathwise definition:

Definition 4.1 (Stochastic Airy function)**.** Let B be a standard Brownian motion, $\lambda \in \mathbf{C}$ and $\alpha \geq 0$. Then, define the kernel

$$\mathcal{U}_\lambda(t,u) = \frac{t^2 - u^2}{2} + \sqrt{2\alpha}(B(t) - B(u)) + \lambda(t-u).$$

Let $T \in \mathbf{R}$ and $c_1, c_2 \in \mathbf{C}$. The stochastic Airy equations are then the integral equations

$$\Phi_\lambda(t) = c_2 + \int_T^t \mathcal{U}_\lambda(t,u)\Phi_\lambda(u)\,\mathrm{d}u + c_1\mathcal{U}_\lambda(t,T),$$
$$\phi_\lambda(t) = c_1 + \int_T^t \Phi_\lambda(u)\,\mathrm{d}u.$$

There is, up to a constant multiple, a unique choice of (c_1, c_2) such that the corresponding solutions $(\Phi_\lambda, \phi_\lambda)$ remain bounded as $t \to \infty$. The multiple c_1/c_2 is then chosen in such a way that as $\alpha \to 0$, the $t \to \infty$ asymptotics of $\mathrm{SAi}_\lambda(t)$ coincide with those of $\mathrm{Ai}(t + \lambda)$ as per [LP21, Eq. (1.7)].

The stochastic Airy function is then defined by

$$\mathrm{SAi}_\lambda(t) := \phi_\lambda(t)$$

for this choice of (c_1, c_2).

The main result of [LP21] as pertains to this chapter is then as follows:

Theorem 4.2 ([LP21, Theorem 1.1])**.** *Let W_N be a scaled G(U/O)E. Let $\alpha = 1$ in the GUE case and $\alpha = 2$ in the GOE case. Let φ_N be the characteristic polynomial of W_N and let w_N be the weight function defined by*

$$w_N(z) := \left((2\pi)^{1/4} e^{Nz^2} 2^{-N} (Nz^2)^{-1/12} \sqrt{\frac{N!}{N^N}}\right)^{-1}.$$

Define the scaled quantity

$$\Psi_N(\lambda) := 2^{-N} w_N(1 + \lambda N^{-2/3}/2) \varphi_N(2 + \lambda N^{-2/3}).$$

Then there exists a centered Gaussian random variable G_N with

$$\mathbf{E} G_N^2 = \frac{\alpha}{3} \log N + O(1)$$

such that as a random real-analytic function under the topology of locally uniform convergence of the function and all its derivatives,

$$\Big(\frac{\mathbf{E} e^{G_N}}{e^{G_N}} \Psi_N(\lambda) : \lambda \in \mathbf{R}\Big) \xrightarrow{\mathrm{d}} (\mathrm{SAi}_\lambda(0) : \lambda \in \mathbf{R}).$$

We will not directly use this result in this chapter, but we will use the machinery developed in [LP21] to establish similar results for Gaussian matrices with critical spikes.

For the characteristic function of a Gaussian matrix with a critical spike, the corresponding limiting object is no longer exactly the stochastic Airy function, but instead, the following modified version:

Definition 4.3. Let $\mathrm{SAi}_\lambda(t)$ be the stochastic Airy function.

For $b \in \mathbf{R}$, we then define

$$s_b(\lambda) = -b\,\mathrm{SAi}_\lambda(0) - \mathrm{SAi}'_\lambda(0).$$

Let $\alpha \geq 1$ and let μ^* be the largest root of $s_b(\lambda)$. We then define $s_b^{(\alpha)}$ to be the unique function that is analytic on $\mathbf{C} \setminus (-\infty, \mu^*]$ and satisfies

$$(s_b^{(\alpha)})^\alpha = s_b(\lambda)$$

and $s_b^{(\alpha)}(x) > 0$ for $x \in (\mu^*, \infty)$. We will refer to $s_b^{(\alpha)}$ as a *spiked stochastic Airy function.*

The existence and properties of these functions is discussed in Sections 4.3.2 and 4.3.4.

4.2 Main results

We begin by presenting a direct extension of Theorem 4.2 to the critically-spiked case. The joint convergence result of Theorem 4.4 is critical for inferring various results about the critically-spiked scaled characteristic polynomial $\Psi_N^{(b)}$ from the corresponding unspiked quantity Ψ_N.

Theorem 4.4. *In the setting of Theorem 4.2, let $b \in \mathbf{R}$ and let $W_N^{(b_0)} = W_N + (1 + b_0 N^{-1/3})e_1 e_1^*$ be a scaled spiked Gaussian matrix with scaled characteristic polynomial $\Psi_N^{(b_0)}$.*

Then, under the topology of locally uniform convergence of the function and all its derivatives,

$$\frac{\mathbf{E}e^{G_N}}{e^{G_N}}\Big((\Psi_N(\lambda), N^{1/3}\Psi_N^{(b_0)}(\lambda)) : \lambda \in \mathbf{R}\Big) \xrightarrow{\mathrm{d}} ((\mathrm{SAi}_\lambda(0), -\mathrm{SAi}'_\lambda(0) - b_0\,\mathrm{SAi}_\lambda(0)) : \lambda \in \mathbf{R}). \tag{4.1}$$

Having established this generic result, we move to the investigation of quantities specifically related to the likelihood ratio of two critically-spiked Gaussian matrices. To this end, let $b, b_0 \in \mathbf{R}$. Let W_N be a Gaussian matrix with spike $\beta_0 = 1 + b_0 N^{-1/3}$ and call its eigenvalues $\tilde{\lambda}_{1,N}^{b_0} > \cdots > \tilde{\lambda}_{N,N}^{b_0}$.

The main object of study in this chapter is the following object:

$$I_{b,b_0} := \int_{\mathcal{K}} \exp\Big\{\frac{N}{\alpha}\Big[\beta z - \frac{1}{N}\sum_{j=1}^{N}\log(z - \tilde{\lambda}^{b_0}_{j,N})\Big]\Big\},$$

which is the critically-spiked version of the integral $\int_{\mathcal{K}} \exp\{(N/\alpha)G(z)\,\mathrm{d}z\}$ that was studied in Chapter 3.

In Theorem 4.5, we will describe the limiting behaviour of this quantity in terms of the stochastic Airy function:

Theorem 4.5. *Let $b_0, b \in \mathbf{R}$ and $\alpha \in \{1,2\}$. There exist*

1. *a sequence $\{W_N\}_N$ of random matrices whose eigenvalues are equal in distribution to those of a GOE if $\alpha = 2$ or GUE if $\alpha = 1$ and spike $\beta_0 = 1 + b_0 N^{-1/3}$, and*
2. *a spiked stochastic Airy function $s^{(\alpha)}_{b_0}$ coupled with $\{W_N\}_N$*

such that, for any $b \in \mathbf{R}$,

$$N^{2/3}\exp\Big\{-\frac{2N}{\alpha} - \frac{2bN^{2/3}}{\alpha}\Big\}|\varphi^{(b_0)}_N(2)|^{1/\alpha} I_{b,b_0} \xrightarrow{\text{a.s.}} \int_{\mathcal{K}} e^{bw/\alpha} s^{(\alpha)}_{b_0}(w)^{-1}\,\mathrm{d}w, \tag{4.2}$$

where is a contour that runs from $-\mathrm{i}\infty$ to $+\mathrm{i}\infty$ and passes on the positive side of $\mu^{(b_0)}_{1,\infty}$ and where $\varphi^{(b_0)}_N$ is the characteristic polynomial of W_N.

This result will be the main tool used to find the limiting behaviour of the SSK model at the triple point, and then to describe the limiting distribution of the likelihood ratio for testing between two critically-spiked Gaussian matrices.

4.2.1 The SSK triple point

The results of Chapter 3 complete the paramagnetic-spin glass phase transition of the SSK phase diagram show in Fig. 1.1.

This leaves only the triple-point unaddressed. That is, in the notation of Eqs. (3.1) to (3.3), we would like to know the limiting distribution of $F_{\alpha,N}$ where $\beta, J \sim 1$.

Recall from Eqs. (3.3) and (3.6) that $F_{\alpha,N} = (\alpha/2N)\log Z_{\alpha,N}$, where

$$Z_{\alpha,N} = C_{\alpha,N}\frac{I_{b,b_0}}{2\pi\mathrm{i}}$$

for the constant $C_{\alpha,N}$ defined by

$$C_{\alpha,N} = \frac{\Gamma(N/\alpha)}{(\beta N/\alpha)^{N/\alpha-1}}.$$

Directly using the limiting description of I_{b,b_0} we can establish the following extension of Theorem 3.1 to a regime within the triple-point

Theorem 4.6. *Let $b, b_0 \in \mathbf{R}$ and $\alpha \in \{1,2\}$. Let $F_{\alpha,N}$ be defined as in Eq.* (3.1) *–* *Eq.* (3.3) *with $\beta = 1 + bN^{-1/3}$ and $J = 1 + b_0 N^{-1/3}$. Then*

$$\frac{N}{\sqrt{\frac{\alpha}{12}\log N}}\left(F_{\alpha,N} - F(\beta) - \frac{\log N}{12N}\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1), \tag{4.3}$$

where $F(\beta) = \beta - \frac{1}{2}\log\beta - \frac{3}{4}$.

Remark 4.7. In the above, we could also have chosen $F(\beta) = \beta^2/4$, which is the other piece in Eq. (1.12), since the difference between these terms is $O(N^{-1})$.

Comparing this with Theorem 3.1, we also see that the sign of the $\log N/12N$ term is flipped. In light of the differences between Theorem 2.1 and Lemma 4.26, we find that this is because the additional $O(N^{-1/3})$ separation of $\tilde{\lambda}^{b_0}_{1,N}$ from the bulk caused by the critical spike causes $|\varphi^{(b_0)}_N(2)|$ to be $O(N^{1/3})$ times smaller than the corresponding unspiked quantity. This then induces a $O(\log N/N)$ shift in $F_{\alpha,N}$.

The lack of dependence on b_0 and b in the limiting distribution on the right-hand side of Eq. (4.3) suggests that this in an incomplete description of the SSK triple point. Indeed, by using the results of this chapter, we can extend the reasoning of Chapter 3 to the triple point for $\beta = 1 + bN^{-1/3}\sqrt{\log N}$ with $b > 0$. Precisely, we have the following:

Theorem 4.8. *Consider $F_{\alpha,N}$ with $\alpha = 1$ or $\alpha = 2$, as defined in Eq.* (3.1) *–* *Eq.* (3.3)*. Let $\beta = 1 + bN^{-1/3}\sqrt{\log N}$ for a constant $b \geq 0$ and let $0 \leq J < 1$. Then*

$$\frac{N}{\sqrt{\frac{\alpha}{12}\log N}}\left(F_{\alpha,N} - F(\beta) - \frac{\log N}{12N}\right) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1) + \sqrt{\frac{3}{\alpha}}b\mathrm{BV}(-b_0)_{2/\alpha}, \tag{4.4}$$

where BV_2 and BV_1 are the complex and real Bloemendal-Viràg distributions, respectively, independent from the $\mathcal{N}(0,1)$, and where $F(\beta) = \beta - \frac{1}{2}\log\beta - \frac{3}{4}$.

The proof of these results is completed in Section 4.4.

4.2.2 Likelihood ratio of critically-spiked Gaussians

Let W_N be a scaled Gaussian matrix with critical spike h. We return to our initial testing problem of

$$H_0\colon h = \beta_0 := 1 + b_0 N^{-1/3} \qquad \text{vs} \qquad H_1\colon h = \beta := 1 + bN^{-1/3}.$$

Recall from Section 1.2.1 that the likelihood ratio for this test is based on $L(\Lambda) := p_N(\Lambda;\beta)/p_N(\Lambda;\beta_0)$, where $p_N(\,\cdot\,;h)$ is the density of the eigenvalues of W_N.

Now according to Eq. (1.3), we can express

$$p_N(\Lambda;h) = c(\Lambda)d(h)Z_{\alpha,N} = \frac{c(\Lambda)}{2\pi \mathrm{i}} d(h) C_{\alpha,N} I_{b,b_0}.$$

where $d(h) = \exp\{(2N/\alpha)\cdot h^2/4\}$.

Unfortunately, the limiting description of $Z_{\alpha,N}$ implied by Theorem 4.6 is not in itself sufficient to pin down the limiting distribution of $L(\Lambda)$. However, we can write

$$\begin{aligned}\log\frac{p_N(\Lambda;\beta)}{p_N(\Lambda;\beta_0)} &= \frac{2N}{\alpha}\frac{1}{4}(\beta^2-\beta_0^2) + \Big(\frac{N}{\alpha}-1\Big)\log\frac{\beta}{\beta_0} + \log\frac{I_{b,b_0}}{I_{b_0,b_0}} \\ &= \frac{2}{\alpha}N^{2/3}(b-b_0) + \log\frac{I_{b,b_0}}{I_{b_0,b_0}} + o(1). \end{aligned} \tag{4.5}$$

Now, the results of Theorem 4.5 can be used to properly analyze $\log(I_{b,b_0}/I_{b_0,b_0})$, and yield the following description of the likelihood ratio's limiting distribution:

Theorem 4.9. *Let $\alpha \in \{1,2\}$ and $N \in \mathbf{Z}_{>0}$. Let $b, b_0 \in \mathbf{R}$ and let $\beta = 1 + bN^{-1/3}, \beta_0 = 1 + b_0N^{-1/3}$.*

Let $p_N(\,\cdot\,;h)$ be the density of the eigenvalues of an $N\times N$ GUE if $\alpha = 1$ or GOE if $\alpha = 2$ with a spike of h. If $\Lambda \sim p_N(\,\cdot\,;\beta_0)$, then

$$\frac{p_N(\Lambda;\beta)}{p_N(\Lambda;\beta_0)} \xrightarrow{\mathrm{d}} \frac{\int_{\mathcal{K}} e^{bw/\alpha} s_{b_0}^{(\alpha)}(w)^{-1}\,\mathrm{d}w}{\int_{\mathcal{K}} e^{b_0 w/\alpha} s_{b_0}^{(\alpha)}(w)^{-1}\,\mathrm{d}w},$$

where $\mathcal{K}$ is a contour that runs from $-\mathrm{i}\infty$ to $+\mathrm{i}\infty$ and passes on the positive side of the largest root of $s_{b_0}^{(\alpha)}(w)$.

As with Theorem 4.6, this result is proved in Section 4.4.

Remark 4.10. In Theorem 4.9, β and β_0 are of the same form. This means that, under the alternative hypotheses, $\Lambda_1 \sim p_N(\,\cdot\,;\beta)$, it is also true that

$$\frac{p_N(\Lambda_1;\beta_0)}{p_N(\Lambda_1;\beta)} \xrightarrow{\mathrm{d}} \frac{\int_{\mathcal{K}} e^{b_0 w/\alpha} s_b^{(\alpha)}(w)^{-1}\,\mathrm{d}w}{\int_{\mathcal{K}} e^{bw/\alpha} s_b^{(\alpha)}(w)^{-1}\,\mathrm{d}w}.$$

Taking $\theta = b - b_0$, the distributional convergence under the corresponding nulls of both likelihood ratios to a non-zero limit implies, by Le Cam's first lemma, that for any $b_0 \in \mathbf{R}$, the experiment

$$H_0\colon h = h_0 := 1 + b_0 N^{-1/3} \qquad \text{vs.} \qquad H_1\colon h = h_0 + \theta N^{-1/3}$$

has mutually contiguous null and alternative hypotheses.

4.3 Preliminary results

Before proceeding to the proof of Theorem 4.5, we must establish several results about the convergence of the characteristic polynomial of a critically-spiked Gaussian.

To this end, the rest of this chapter is organized as follows:

1. In Section 4.3.2, we establish convergence analogous to theorem 1.6 of [LP21] in the $b \in \mathbf{R}$ case and define s_b.

2. In order to properly define a consistent contour of integration and complex roots, we must first establish that for any j, the scaled eigenvalues $|N^{2/3}(\tilde{\lambda}_{j,N}^{(b)} - 2)|$ are almost surely bounded. We address this in Section 4.3.3.

3. In Section 4.3.4, we properly define $s_b^{(\alpha)}$ as the αth root of s_b and show that we can infer limiting results about αth root a function from corresponding results from the original function. even though one is not a simple function of the other.

4. The proof of Theorem 4.5, which is mostly devoted to verifying the conditions for dominated convergence for I_{b,b_0}, is in Section 4.4. It is then joined there by the proofs of the other important theorems of this chapter.

4.3.1 Notation

Let $\mathbf{A}$ be the symmetric tridiagonal semi-infinite matrix described of [LP21, Eq. (1.10)] and let $2/\alpha$ be the corresponding Dyson parameter. That is,

$$\mathbf{A} = \begin{pmatrix} b_1 & a_1 & & \\ a_1 & b_2 & a_2 & \\ & a_2 & b_3 & \ddots \\ & & \ddots & \ddots \end{pmatrix},$$

where

$$b_i \sim \mathcal{N}(0,2), \qquad a_i \sim \chi_{(2/\alpha)i}$$

are independent random variables. We use the notation $[\mathbf{A}]_{n,n}$ to denote the principal $n \times n$ submatrix of $\mathbf{A}$.

For $b \in \mathbf{R}$ and $n \in \mathbf{Z}_{>0}$, define

$$[\mathbf{A}]^{(b)}_{n,n} = [\mathbf{A}]_{n,n} + \sqrt{n(2/\alpha)}h_n e_n e_n^{\mathrm{T}}$$

where $h_n = 1 + bn^{-1/3}$ and where e_n is the nth standard basis vector in $\mathbf{R}^n$. In this way, the eigenvalues of $[\mathbf{A}]^{(b)}_{n,n}$ have the same joint distribution as the eigenvalues of a G(U/O)E with spike h_n.

Now for $b \in \mathbf{R}$, let $\lambda^{(b)}_{1,N} > \lambda^{(b)}_{2,N} > \cdots > \lambda^{(b)}_{N,N}$ be the eigenvalues of $\mathbf{A}^{(b)}_N$ and define the scaled versions of these eigenvalues by

$$\tilde{\lambda}^{(b)}_{j,N} = \frac{1}{\sqrt{N\beta}}\lambda^{(b)}_{j,N},$$
$$\mu^{(b)}_{j,N} = N^{2/3}(\tilde{\lambda}^{(b)}_{j,N} - 2).$$

Using these quantities, define the functions:

$$\varphi^{(b)}_N(z) = \prod_{j=1}^{N}(z - \tilde{\lambda}^{(b)}_{j,N}),$$
$$f^{(\alpha,b)}_N(w) = \exp\Big\{\frac{1}{\alpha}\Big[-N^{1/3}w + \sum_{j=1}^{N}\log(2 + wN^{-2/3} - \tilde{\lambda}^{(b)}_{j,N})\Big]\Big\},$$

where $\alpha \in \mathbf{Z}_{>0}$.

Here, $\varphi_N^{(b)}$ is the characteristic polynomial of $[\mathbf{A}]_{N,N}^{(b)}/\sqrt{N\beta}$, whereas $f_N^{(\alpha,b)}$ is in some sense the "scaled αth root of $\varphi_N^{(b)}$." This notion is made precise in Section 4.3.4.

In any case, notice that $\mu_{j,N}^{(b)}$ are the zeroes of $f_N^{(1,b)}$.

4.3.2 Characteristic polynomial in the spiked case

Following the notation of [LP21], for a fixed N and $b \in \mathbf{R}$, define

$$\begin{aligned}\Phi_n^{(b)}(z) &= \det(zI - (4N\beta)^{-1/2}[\mathbf{A}]_{n,n}^{(b)}),\\ \Psi_n^{(b)}(\lambda) &= w_n\Big(1+\frac{\lambda}{2N^{2/3}}\Big)\Phi_n\Big(1+\frac{\lambda}{2N^{2/3}}\Big),\end{aligned}$$

where w_n is the weight function

$$w_n(z) = \left((2\pi)^{1/4}e^{Nz^2}2^{-n}(Nz^2)^{-1/12}\sqrt{\frac{n!}{N^n}}\right)^{-1}$$

given in eq. 1.2 of [LP21]. In terms of these quantities,

$$\varphi_N^{(b)}(2) = 2^N\Phi_N^{(b)}(1), \tag{4.6}$$

From the recurrence on the first display on p. 6 of [LP21], we have that

$$\begin{aligned}\Phi_N^{(b)}(z) &= \det(zI - (4N\beta)^{-1/2}[\mathbf{A}]_{N,N}^{(b)})\\ &= \Big(z - \frac{b_N}{2\sqrt{N\beta}} - \frac{h_N}{2}\Big)\Phi_{N-1}(z) - \frac{a_{N-1}^2}{4N\beta}\Phi_{N-2}(z),\\ &= \Phi_N(z) - \frac{h_N}{2}\Phi_{N-1}(z).\end{aligned}$$

It follows that

$$\begin{aligned}\Psi_N^{(b)}(\lambda) &= w_N(z)\Phi_N^{(b)}(z)\\ &= \Psi_N(\lambda) - \frac{h_N}{2}\frac{w_N(z)}{w_{N-1}(z)}\Psi_{N-1}(\lambda)\\ &= \Psi_N(\lambda) - h_N\Psi_{N-1}(\lambda).\end{aligned}$$

The last line follows from the fact that, for $m, n \in \mathbf{Z}_{\geq 0}$, we have that

$$\begin{aligned}\frac{w_n(z)}{w_m(z)} &= \frac{2^{-m}\sqrt{m!/N^m}}{2^{-n}\sqrt{n!/N^n}} \\ &= 2^{n-m}\sqrt{\frac{m!}{n!}}N^{\frac{n-m}{2}},\end{aligned}$$

and in particular $w_N(z)/w_{N-1}(z) = 2$.

As in eq. (9.2) of [LP21], define piecewise linear interpolations $P_\lambda\colon (-\infty, T] \to \mathbf{R}$ over the grid $N^{-1/3}\mathbf{Z}$ (where $T = N^{-1/3}\omega_N \sim \log^{1-\kappa} N$ for some $\kappa > 0$ and $\omega_N \in \mathbf{Z}_{>0}$) by

$$P'_\lambda(t) = N^{1/3}(\Psi_{n-1}(\lambda) - \Psi_n(\lambda))\mathbf{1}\{N - n = \lfloor tN^{1/3} \rfloor\} \qquad P_\lambda(0) = \Psi_\lambda(0).$$

With this notation, we can write

$$\begin{aligned}\Psi_N^{(b)}(\lambda) &= P_\lambda(0) - h_N P_\lambda(N^{-1/3}) \\ &= (1 - h_N)P_\lambda(0) - h_N N^{-1/3} P'_\lambda(0),\end{aligned}$$

where the above is an exact equality due to the fact that P_λ is piecewise linear.

Since $h_N = 1 + bN^{-1/3}$, the above becomes

$$\begin{aligned}\Psi_N^{(b)}(\lambda) &= -N^{-1/3}[bP_\lambda(0) + P'_\lambda(0)] - N^{-2/3}bP'_\lambda(0), \\ N^{1/3}\Psi_N^{(b)}(\lambda) &= -bP_\lambda(0) - P'_\lambda(0) + N^{-1/3}bP'_\lambda(0).\end{aligned}$$

By the eighth display on [LP21, p. 79], we have that

$$\begin{aligned}P_\lambda(t) &= C_N(\mathrm{SAi}_\lambda(t)e^{\varepsilon_{N,\lambda}} + \chi_{N,\lambda}(t)), \\ P'_\lambda(t) &= C_N(\mathrm{SAi}'_\lambda(t)e^{\varepsilon_{N,\lambda}} + \chi'_{N,\lambda}(t)),\end{aligned}$$

where

$$C_N = \frac{e^{G_N}}{\mathbf{E}e^{G_N}}, \qquad G_N \sim \mathcal{N}\Big(0, \frac{\alpha}{3}\log N + O(1)\Big)$$

is a random constant that doesn't depend on λ or t. Moreover, with probability $1 -$

$e^{-(\log N)^{1+\varepsilon}}$ for $k \in \{0,1\}$,

$$\sup_{\lambda\in K}|e^{\varepsilon_{N,\lambda}}-1| = O((\log N)^{-1/6+\varepsilon}),$$
$$\sup_{\lambda\in K, t\in[-e^T,T]}|\partial_t^k\chi_{N,\lambda}(t)| = O(N^{\kappa-1/6}).$$

The last display of [LP21, p. 79] only explicitly states the bound for $k=0$. However, the bounds in [LP21, Eqs. 9.14 and 9.17] also hold unchanged for $k=1$, and so this bound also holds for $k=1$.

To conveniently represent the various limiting results that hold uniformly in $\lambda \in K$, we will use $g_N(\lambda) = O_K(a_N)$ to mean $\sup_{\lambda\in K}|g_N(\lambda)| = O(a_N)$.

Hence, we have that

$$\begin{aligned}\frac{N^{1/3}}{C_N}\Psi_N^{(b)}(\lambda) &= -b[\mathrm{SAi}_\lambda(0)e^{\varepsilon_{N,\lambda}}+\chi_{N,\lambda}(0)] - [\mathrm{SAi}_\lambda'(0)e^{\varepsilon_{N,\lambda}}+\chi_{N,\lambda}'(0)](1-N^{-1/3}b)\\ &= s_b(\lambda)e^{\varepsilon_{N,\lambda}} - \chi_{N,\lambda}(0)+\chi_{N,\lambda}'(0)(1-N^{-1/3}b)\\ &= s_b(\lambda)(1+O_K((\log N)^{-1/6+\varepsilon}))+O_K(N^{\kappa-1/6}).\end{aligned}\tag{4.7}$$

Now, Eq. (4.1) follows immediately from the above display. In particular:

Proof of Theorem 4.4. The eighth display on [LP21, p. 79] has the equivalent of Eq. (4.7) in the unspiked case, namely that

$$\frac{1}{C_N}\Psi_N(\lambda) = \mathrm{SAi}_\lambda(0)(1+O_K((\log N)^{-1/6+\varepsilon}))+O_K(N^{\kappa-1/6}).$$

Since $s_b(\lambda) = -\,\mathrm{SAi}_\lambda'(0) - b\,\mathrm{SAi}_\lambda(0)$, the joint convergence of Theorem 4.4 follows. □

Equation (4.1) also allows us to analyze the integrand of I_{b,b_0}. To this end, define

$$f_N^{(\alpha,b)}(\lambda) = \exp\Big\{-N^{1/3}\lambda + \sum_{j=1}^N \log(2+\lambda N^{-2/3}-\tilde\lambda_{j,N}^{(b)})\Big\}.$$

Proposition 4.11. *Let $K \subseteq \mathbf{R}$ be compact. For any $b \in \mathbf{R}$, we have that*

$$\sup_{\lambda\in K}\left|\frac{f_N^{(1,b)}(\lambda)}{|\varphi_N^{(b)}(2)|} - \frac{s_b(\lambda)}{|s_b(0)|}\right| \xrightarrow{\text{a.s.}} 0 \tag{4.8}$$

Proof. Let $z = 1 + \lambda/(2N^{2/3})$ for $\lambda \in K$. We then have that

$$\begin{aligned} f_N^{(1,b)}(\lambda) &= \exp\Big\{-N^{1/3}\lambda + \sum_{j=1}^{N} \log(2 + \lambda N^{-2/3} - \tilde{\lambda}_{j,N}^{(b)})\Big\} \\ &= \exp\Big\{-N^{1/3}\lambda + N\log 2 + \sum_{j=1}^{N} \log\Big(1 + \frac{\lambda}{2N^{2/3}} - \frac{\lambda_{j,N}^{(b)}}{\sqrt{4N\beta}}\Big)\Big\} \\ &= 2^N e^{-N^{1/3}\lambda}\Phi_N^{(b)}(z), \end{aligned} \tag{4.9}$$

recalling that $\lambda_1 > \ldots > \lambda_N$ are the eigenvalues of $[\mathbf{A}]_{N,N}^{(b)}$.

Here, the last line follows from the fact that

$$\exp\Big\{\sum_{j=1}^{N} \log\Big(z - \frac{\lambda_{j,N}^{(b)}}{\sqrt{4N\beta}}\Big)\Big\} = \prod_{j=1}^{N}\Big(z - \frac{\lambda_{j,N}^{(b)}}{\sqrt{4N\beta}}\Big).$$

Notice, however, that for $\alpha \neq 1$, we have in general that

$$\exp\Big\{\frac{1}{\alpha}\sum_{j=1}^{N} \log\Big(z - \frac{\lambda_{j,N}^{(b)}}{\sqrt{4N\beta}}\Big)\Big\} \neq \prod_{j=1}^{N}\Big(z - \frac{\lambda_{j,N}^{(b)}}{\sqrt{4N\beta}}\Big)^{1/\alpha}.$$

This discrepancy is the reason that it is not the case that $\frac{f_N^{(\alpha,b)}(\lambda)}{|\varphi_N^{(b)}(2)|} \to \frac{s_b(\lambda)^{1/\alpha}}{|s_b(0)|}$ for any choice of branch cut in $z \mapsto z^{1/\alpha}$, and why we must be more careful when defining $s_N^{(\alpha)}(\lambda)$.

Further, notice that

$$\begin{aligned} \frac{w_N(z)}{w_N(1)} &= e^{N(1-z^2)}N^{1/6} \\ &= \exp\Big\{N\Big(-\frac{\lambda}{2N^{2/3}}\Big)\Big(2 + \frac{\lambda}{2N^{2/3}}\Big) + \frac{1}{6}\log\Big(1 + \frac{\lambda}{2N^{2/3}}\Big)\Big\} \\ &= e^{-N^{1/3}\lambda}(1 + O_K(N^{-2/3})). \end{aligned}$$

Combining the above with Eqs. (4.6) and (4.9), yields that

$$\frac{f_N^{(1,b)}(\lambda)}{|\varphi_N^{(b)}(2)|} = e^{-N^{1/3}\lambda}\frac{\Phi_N^{(b)}(z)}{|\Phi_N^{(b)}(1)|} = \frac{\Psi_N^{(b)}(\lambda)}{|\Psi_N^{(b)}(0)|}(1 + O_K(N^{-2/3})). \tag{4.10}$$

Now, it follows from the representation of $\Phi_N^{(b)}$ in Eq. (4.1) that

$$\frac{\Psi_N^{(b)}(\lambda)}{|\Psi_N^{(b)}(0)|} = \frac{s_b(\lambda)(1 + O_K((\log N)^{-1/6+\varepsilon})) + O_K(N^{\kappa-1/6})}{|s_b(0)(1 + O_K((\log N)^{-1/6+\varepsilon})) + O_K(N^{\kappa-1/6})|}.$$

Since $\mathbf{P}(s_b(0) = 0) = 0$, we can conclude that

$$\sup_{\lambda\in K}\left|\frac{\Psi_N^{(b)}(\lambda)}{|\Psi_N^{(b)}(0)|} - \frac{s_b(\lambda)}{|s_b(0)|}\right| \xrightarrow{\text{a.s.}} 0.$$

Combining this with Eq. (4.10) completes the proof. □

We would like to simply extend Proposition 4.11 directly to the complex plane using the results of [Ass22]. However, the function $f_N^{(1,b)}/|\varphi_N^{(b)}(2)|$ is not of the correct form due to the absolute value in the denominator. To this end, we must prove the following intermediate result:

Corollary 4.12. *Let $K \subseteq \mathbf{R}$ be compact. For any $b \in \mathbf{R}$, we have that*

$$\sup_{\lambda\in K}\left|\frac{f_N^{(1,b)}(\lambda)}{\varphi_N^{(b)}(2)} - \frac{s_b(\lambda)}{s_b(0)}\right| \xrightarrow{\text{a.s.}} 0. \tag{4.11}$$

Proof. Since $s_b(0) \neq 0$ almost surely, we have that

$$\operatorname{sign}\varphi_N^{(b)}(2) = \operatorname{sign} f_N^{1,b}(\lambda) \to \operatorname{sign} s_b(0),$$

from which we see that

$$\frac{f_N^{(1,b)}(\lambda)}{\varphi_N^{(b)}(2)} = \frac{f_N^{(1,b)}(\lambda)}{|\varphi_N^{(b)}(2)|\operatorname{sign}\varphi_N^{(b)}(2)} \xrightarrow{\text{a.s.}} \frac{s_b(\lambda)}{|s_b(0)|\operatorname{sign} s_b(0)} = \frac{s_b(\lambda)}{s_b(0)}.$$

uniformly in K. □

Corollary 4.13. *Let $b \in \mathbf{R}$. Let $K \subseteq \mathbf{C}$ be compact. For any $b \in \mathbf{R}$, we have that*

$$\sup_{\lambda\in K}\left|\frac{f_N^{(1,b)}(\lambda)}{\varphi_N^{(b)}(2)} - \frac{s_b(\lambda)}{s_b(0)}\right| \xrightarrow{\text{a.s.}} 0. \tag{4.12}$$

Proof. Now, we have that

$$\frac{f_N^{1,b}(\lambda)}{\varphi_N^{(b)}(2)} = e^{-N^{1/3}\lambda}\prod_{i=1}^{N}\Big(1+\frac{\lambda}{-\mu_{j,N}^{(b)}}\Big).$$

Together with Corollary 4.12, this shows that $f_N^{1,b}(\lambda)/\varphi_N^{(b)}(2)$ satisfies the hypothesis of Proposition 4.6 of [Ass22], and so we have that, almost surely,

$$\frac{f_N^{(1,b)}}{\varphi_N^{(b)}(2)} \to \frac{s_b}{s_b(0)}$$

uniformly in compact subsets of $\mathbf{C}$. The conclusion follows again from the observation that $\operatorname{sign}\varphi_N^{(b)}(2) \xrightarrow{\text{a.s.}} \operatorname{sign} s_b(0)$. ☐

4.3.3 Extreme values of top eigenvalues

This section is devoted to showing that, almost surely,

$$\sup_N |\mu_{j,N}^{(b)}| < \infty. \tag{4.13}$$

A critical result for this conclusion is the following description of the zeroes of s_b:

Lemma 4.14. *Let $b \in \mathbf{R}$, and let M_b be the zero set of s_b. It holds almost surely that*

1. *M_b is a countable subset of $\mathbf{R}$,*
2. *M_b is bounded above,*
3. *M_b has no accumulation points.*

Proof. This is a restatement of theorem 6.7 of [LP21] with the identification $\omega = -b$.

Definition 4.15. Given the convergence outlined in Proposition 4.11 of $f_N^{(1,b)}$ to s_b up to some scaling, it is evocative to think of the zeroes of s_b as the limiting zeroes of $f_N^{(1,b)}$.

Moreover, Lemma 4.14 confirms that these zeroes can be enumerated in decreasing order, so denote the zeroes of s_b by

$$\mu_{1,\infty}^{(b)} > \mu_{2,\infty}^{(b)} > \cdots .$$

Now, we demonstrate the positive and negative parts of the bound Eq. (4.13) separately in the following subsections:

Lower bound

We know from Lemma 4.14 that s_b has infinitely many zeroes. Hence, to demonstrate the lower bound of Eq. (4.13), we must show that these zeroes don't appear "at the last minute," as happens, for example in the limit of $g_N(x) = x^2 + \frac{1}{N}$

The relevant observation is that each function $f_N^{(1,b)}$ is a polynomial multiplied by an exponential, and Lemma 4.17 shows that for such functions, since the zeroes and stationary points are interlaced, it cannot happen that a zero is introduced to the limiting function without a nearby zero being present in the $f_N^{(1,b)}$.

This is enough to "anchor" the zeroes of $f_N^{(1,b)}$ near to those of s_b and so guarantee that $\liminf_{N\to\infty} \mu_{j,N}^{(b)}$ remains finite.

Lemma 4.16. *Let p be a real-rooted polynomial and let g be a log-concave function.*

If $f(x) = p(x)g(x)$, then all local maxima of f are non-negative and all local minima of f are non-positive.

Proof. Let x_0 be a local extremum of f such that $f(x_0) \neq 0$. Suppose that $f(x_0) > 0$. Since f is continuous, let $(a, b) \ni x_0$ be an interval such that f, and so p, is positive on (a, b).

Let $\lambda_1, \ldots, \lambda_N \in \mathbf{R}$ be the roots of p so that we have, for $x \in (a, b)$

$$\frac{\mathrm{d}}{\mathrm{d}x} \log p(x) = \frac{p'(x)}{p(x)} = \sum_{j=1}^{N} \frac{1}{x - \lambda_j},$$

$$\frac{\mathrm{d}^2}{\mathrm{d}x^2} \log p(x) = -\sum_{j=1}^{N} \frac{1}{(x - \lambda_j)^2} < 0.$$

That is, p, and so f, is log-concave. In particular, the extremum of f at x_0 must be a local maximum.

Similarly, if $f(x_0) < 0$, then f has a local minimum at x_0. □

Lemma 4.17. *Let $\{p_N\}$ be a sequence of real-rooted polynomials, $\{\kappa_N\}$ a sequence in $\mathbf{R}$, and f a function such that $f_N(x) := p_N(x)e^{\kappa_N x} \to f(x)$ pointwise.*

Let $x \in \mathbf{R}$ be an isolated root of f. Then there exists a sequence of real numbers $\{x_N\}$ such that

1. *$f_N(x_N) = 0$ for each N, and*

2. *$x_N \to x$.*

Proof. Let $\varepsilon > 0$ be small enough that f has no zeroes other than x in $[x-\varepsilon, x+\varepsilon]$.

First, consider the case where $f(x-\varepsilon)$ and $f(x+\varepsilon)$ have opposite signs. Without loss of generality, assume that

$$f(x-\varepsilon) < -\eta < \eta < f(x+\varepsilon)$$

for some $\eta > 0$,

Let N_0 be such that, for all $N > N_0$, $|f_N(y) - f(y)| < \varepsilon/2$ for $y \in \{x-\varepsilon, x+\varepsilon\}$. It follows that

$$f_N(x-\varepsilon) < -\frac{\eta}{2} < \frac{\eta}{2} < f_N(x+\varepsilon),$$

and so that each such f_N has a root in $(x-\varepsilon, x+\varepsilon)$.

On the other hand, consider the case where $f(x-\varepsilon)$ and $f(x+\varepsilon)$ have the same sign. Without loss of generality, assume that

$$f(x-\varepsilon), f(x+\varepsilon) > \eta$$

for some $\eta > 0$.

Let N_0 be such that, for all $N > N_0$, $|f_N(y) - f(y)| < \varepsilon/3$ for $y \in \{x-\varepsilon, x, x+\varepsilon\}$. It follows that

$$f_N(x-\varepsilon), f_N(x+\varepsilon) > \frac{2\eta}{3},$$
$$f_N(x) < \frac{\eta}{3},$$

and so that f_N has a local minimum at some $x_0 \in (x-\varepsilon, x+\varepsilon)$.

Since f_N of the form described in Lemma 4.16, this means that $f_N(x_0) < 0$, and so, since $f_N(x-\varepsilon) > 0$, that f_N has a root in $(x-\varepsilon, x+\varepsilon)$.

It follows that we can always choose a sequence $\{x_N\}$ of zeroes of $\{f_N\}$ such that $x_N \to x$. □

Proposition 4.18. *For any $b \in \mathbf{R}$ and for any $k \in \mathbf{Z}_{>0}$, it holds almost surely that*

$$\inf_N \mu_{j,N}^{(b)} > -\infty. \tag{4.14}$$

Proof. Let $\varepsilon > 0$. By Lemma 4.14, there exists a K_ε such that $\mu_{k,\infty}^{(b)} > C_\varepsilon$ on an event A_ε such that $\mathbf{P}(A_\varepsilon) > 1 - \varepsilon$.

Since, by Lemma 4.14(3), each $\mu_{j,\infty}^{(b)}$ is an isolated root of $s_b(w)/|s_b(0)|$, and since Proposition 4.11 yields that, almost surely $f_N^{(1,b)}/|\varphi_N^{(b)}(2)| \to s_b/|s_b(0)|$, these functions satisfy the hypotheses of Lemma 4.17.

Therefore, for each $j \le k$ there exists a sequence $\{\tilde{\mu}_{j,N}\}$ such that $\tilde{\mu}_{j,N}$ is a root of $f_N^{(1,b)}$ and $\tilde{\mu}_{j,N} \to \mu_{j,\infty}^{(b)}$ (it is not explicitly stated or required for this argument that $\tilde{\mu}_{j,N} = \mu_{j,N}^{(b)}$).

In particular, there is an N_0 such that, for $N > N_0$, $\tilde{\mu}_{j,N} > C_\varepsilon - 1$ for each $j \le k$, and so $f_N^{(1,b)}$ has at least k zeroes exceeding $C_\varepsilon - 1$. Therefore, on A_ε, $\liminf_{N\to\infty} \mu_{j,N}^{(b)} > C_\varepsilon - 1$, and so

$$\mathbf{P}\big(\inf_N \mu_{j,N}^{(b)} > -\infty\big) > 1 - \varepsilon.$$

Since this holds for any $\varepsilon > 0$, it follows that Eq. (4.14) holds almost surely. □

Upper bound

For the upper bound of Eq. (4.13), we again know from Lemma 4.14(2) that $\mu_{1,\infty}^{(b)} < \infty$.

If $\liminf_{N\to\infty} \mu_{1,N}^{(b)} = \infty$, we show that this conflicts with the convergence in distribution of $\mu_{1,N}^{(b)}$. On the other hand, we show how eigenvalue interlacing guarantees that it is impossible for $\mu_{1,N}^{(b)}$ to oscillate in such a way that only a subsequence diverges to infinity.

Proposition 4.19. *For any $b \in \mathbf{R}$, it holds almost surely that*

$$\sup_N \mu_{1,N} < \infty.$$

Proof. Since $\mu_{1,\infty}^{(b)}$ is the largest root of s_b, it must be that s_b is eventually either positive or negative. We deal with these cases separately.

Case 1: $\lim_{w\to\infty} \operatorname{sign} s_b(w) = -1$.

Let $x_0 = \mu_{1,\infty}^{(b)} + 1$ and let $x > x_0$ Since $s_b(x) = -\eta$ for some $\eta > 0$, there is an N_0 such that, for all $N > N_0$,

$$\left|\frac{f_N^{(1,b)}(x)}{|\varphi_N^{(b)}(2)|} - \frac{s_b(x)}{|s_b(0)|}\right| < \frac{\eta}{2},$$

and in particular, $f_N^{(1,b)}(x) < 0$.

But since $f_N^{(1,b)}$ is a monic polynomial multiplied by a positive function, this means that, for any $N > N_0$, $f_N^{(1,b)}$ must have an odd number of zeroes exceeding x.

In particular, $\liminf_{N\to\infty} \mu_{1,\infty}^{(b)} > x$. But since this holds for arbitrarily large x, it follows that $\liminf_{N\to\infty} \mu_{1,N}^{(b)} = \infty$.

Case 2: $\lim_{w\to\infty} \operatorname{sign} s_b(w) = +1$.

Again, let $x_0 = \mu_{1,\infty}^{(b)} + 1$ and let $x > x_0$. Following the reasoning of the previous case, let N_0 be such that $f_N^{(1,b)}(x) > 0$ for $N > N_0$.

Since $[\mathbf{A}]_{N,N}$ is a principal submatrix of $[\mathbf{A}]_{N+1,N+1}^{(b)}$, we have by Cauchy interlacing that

$$\lambda_{2,N+1}^{(b)} = \lambda_2([\mathbf{A}]_{N+1,N+1} + \sqrt{(N+1)\beta} h_{N+1} e_{N+1} e_{N+1}^{\mathrm{T}}) \le \lambda_1(\mathbf{A}_N).$$

Moreover, since $\sqrt{N\beta} h_N e_N e_N^{\mathrm{T}}$ is non-negative definite, we have that

$$\lambda_1([\mathbf{A}]_{N,N}) \le \lambda_1([\mathbf{A}]_{N,N} + \sqrt{N\beta} h_N e_N e_N^{\mathrm{T}}) = \lambda_{1,N}^{(b)}.$$

Let N_1 be large enough that the function $y \mapsto \sqrt{y}(2 + xy^{-2/3})$ is increasing on $[N_1, \infty)$.

Suppose that $\mu_{1,N}^{(b)} < x$ for some $N > N_0 \vee N_1$. We then have that

$$\begin{aligned} \frac{\lambda_{2,N+1}^{(b)}}{\sqrt{\beta}} &\le \frac{\lambda_{1,N}^{(b)}}{\sqrt{\beta}} \\ &\le \sqrt{N}(2 + xN^{-2/3}) \\ &\le \sqrt{N+1}(2 + x(N+1)^{-2/3}), \end{aligned}$$

and so $\mu_{2,N+1}^{(b)} < x$.

In particular, since $f_{N+1}^{(1,b)}$ is a monic polynomial multiplied by a positive function,

$$\begin{aligned} \operatorname{sign}[f_{N+1}^{(1,b)}(x)] &= \operatorname{sign}(x - \mu_{1,N+1}^{(b)}) \operatorname{sign}\Big[\prod_{j=2}^{N} (x - \mu_{2,N+1}^{(b)})\Big] \\ &= \operatorname{sign}(x - \mu_{1,N+1}^{(b)}). \end{aligned}$$

But since $N + 1 > N_0$, $f_{N+1}^{(1,b)}(x) > 0$, so it follows that $\mu_{1,N+1}^{(b)} < x$. By induction, either $\limsup_{N\to\infty} \mu_{1,N}^{(b)} < x$, or $\liminf_{N\to\infty} \mu_{1,N}^{(b)} > x$. Since this dichotomy holds for arbitrarily large x, if follows that if $\sup_N \mu_{1,N}^{(b)} = \infty$, then necessarily $\liminf_{N\to\infty} \mu_{1,N}^{(b)} = \infty$.

Conclusion: consequence of $\liminf_{N\to\infty} \mu_{1,N}^{(b)} = \infty$.

Let $L > 0$ and let $h_L \colon \mathbf{R} \to [0,1]$ be a continuous function such that

$$h_L(x) = \begin{cases} 0 & \text{if } x \leq L, \\ 1 & \text{if } x \geq L+1. \end{cases}$$

From Theorem 1.5 of [BV13], we have that $\mu_{1,N}^{(b)} \xrightarrow{\text{d}} \mu_{1,\infty}^{(b)}$. If follows from Fatou's lemma that

$$\begin{aligned} \mathbf{P}(\mu_{1,\infty}^{(b)} \geq L) &\geq \mathbf{E} h_L(\mu_{1,\infty}^{(b)}) \\ &= \liminf_{N\to\infty} \mathbf{E} h_L(\mu_{1,N}^{(b)}) \\ &\geq \mathbf{E} \liminf_{N\to\infty} h_L(\mu_{1,N}^{(b)}) \\ &\geq \mathbf{P}(\liminf_{N\to\infty} \mu_{1,N}^{(b)} = \infty) \\ &\geq \mathbf{P}(\sup_N \mu_{1,N}^{(b)} = \infty). \end{aligned}$$

Since $\mathbf{P}(\mu_{1,\infty}^{(b)} = \infty) = 0$, we take $L \to \infty$ to conclude from the above that

$$\mathbf{P}\big(\sup_N \mu_{1,N}^{(b)} = \infty\big) = 0. \quad \square$$

Remark 4.20. Notice that the above proof also demonstrates that

$$\mathbf{P}(\lim_{w\to\infty} s_b(w) = -1) \leq \mathbf{P}(\liminf_{N\to\infty} \mu_{1,N}^{(b)} = \infty) = 0.$$

4.3.4 Complex roots of the characteristic polynomial

Let h be an analytic function such that $h(x) > 0$ for all $x > x_0$ and such that h has no zeroes away from the real line.

For $\alpha \in \mathbf{Z}_{>0}$, define a functional G_α acting on such h so that $G_\alpha(h)$ is a function on $\mathbf{C} \setminus (-\infty, x_0]$ defined by

$$G_\alpha(h) = \exp\Big\{\frac{1}{\alpha}\Big[\int_{x_0}^{w} \frac{h'(w)}{h(w)}\,\mathrm{d}w + \log h(w)\Big\}, \tag{4.15}$$

where the integral it taken over a path in $\mathbf{C} \setminus (-\infty, x_0]$. Since h'/h is analytic over this simply-connected region, this is well-defined.

Since G_α is built from a standard construction of the complex logarithm [see, e.g. Rud87, Theorem 13.11], it is a standard exercise in complex analysis to verify that it has the following properties:

Lemma 4.21. *Let h be an analytic function such that $h(x) > 0$ for all $x > x_0$ and such that h has no zeroes away from the real line. Then $G_\alpha(h)$ is analytic in $\mathbf{C} \setminus (-\infty, x_0]$ and has the following properties:*

1. *$[G_\alpha(h)(w)]^\alpha = h(w)$ for all $w \in \mathbf{C} \setminus (-\infty, x_0]$,*
2. *$G_\alpha(h)(x) = h(x)^{1/\alpha}$ for all $x > x_0$, and*
3. *If $C > 0$ is a constant, then $G_\alpha(Ch) = C^{1/\alpha} G_\alpha(h)$.*

Lemma 4.22. *Let $\{h_N\}$ and h_∞ be holomorphic functions with no zeroes away from the real line such that, for all $N \in \overline{\mathbf{Z}}_{>0}$, $h_N(x) > 0$ for all $x > x_N$.*

If $h_N \to h$ uniformly in compact subsets of $\mathbf{C}$, and if $\sup_N x_N = X^* < \infty$then

$$G_\alpha(h_N) \to G_\alpha(h_\infty)$$

uniformly in compact subsets of $\mathbf{C} \setminus (-\infty, x^*]$.

Proof. First, take $x_0 = x^* + 1$ as a common integration basepoint in Eq. (4.15). The result follows from the fact that, on any compact K, h_N and h_N must be uniformly bounded away from 0, from which is follows that h'_N/h_N converges to h'_∞/h_∞ uniformly in K.

Moreover, there is an $L < \infty$ such that, for any $u \in K$, there is a path in K connecting x_0 to u whose length is less than L, allowing us to bound the difference in integrals. □

Proposition 4.23. *We have that*

$$f_N^{(\alpha,b)} = G_\alpha(f_N^{(1,b)}).$$

Proof. This follows from the fact that these are two functions analytic on $\mathbf{C} \setminus (-\infty, \mu_{1,N}^{(b)}]$ that coincide on the ray $(\mu_{1,N}^{(b)}, \infty)$.

Definition 4.24. Recall that in Remark 4.20, we noted that $s_b(x)$ is eventually positive for $x \to \infty$ along the real axis.

It follows that we can define

$$s_b^{(\alpha)} = G_\alpha(s_b).$$

Proposition 4.25. *Let $b \in \mathbf{R}$ and $\alpha \in \mathbf{Z}_{>0}$.*

Let $\mu^* = \sup_N \mu_{1,N}^{(b)}$. It holds almost surely that

$$\frac{f_N^{(\alpha,b)}}{|\varphi_N^{(b)}(2)|^{1/\alpha}} \to \frac{s_b^{(\alpha)}}{|s_b(0)|^{1/\alpha}}$$

uniformly in compact subsets of $\mathbf{C} \setminus (-\infty, \mu^*]$.

Proof. First, Lemma 4.21(3) together with Proposition 4.23 yields that

$$\frac{f_N^{(\alpha,b)}}{|\varphi_N^{(b)}(2)|^{1/\alpha}} = G_\alpha\Big[\frac{f_N^{(1,b)}}{|\varphi_N^{(b)}(2)|}\Big], \qquad \frac{s_b^{(\alpha)}}{|s_b(0)|^{1/\alpha}} = G_\alpha\Big[\frac{s_b}{|s_b(0)|}\Big]$$

Moreover, Proposition 4.19 guarantees that $\mu^* < \infty$ almost surely, and so we can apply Lemma 4.22 together with Corollary 4.13 to conclude that, almost surely,

$$G_\alpha\Big[\frac{f_N^{(1,b)}}{|\varphi_N^{(b)}(2)|}\Big] \to G_\alpha\Big[\frac{s_b}{|s_b(0)|}\Big]$$

uniformly in compact subsets of $\mathbf{C} \setminus (-\infty, \mu^*]$. Combining the previous two displays completes the proof. □

4.4 Proofs of main results

Fix a $b_0 \in \mathbf{R}$ and define

$$I_{b,b_0} = \int_\Gamma \exp\Big\{\frac{1}{\alpha}\Big[N(1 + bN^{-1/3})z - \sum_{j=1}^N \log(z - \tilde\lambda_{j,N}^{(b_0)})\Big]\Big\}\,\mathrm{d}z,$$

where Γ is a (random) contour with bounded real part that runs from $-\mathrm{i}\infty$ to $+\mathrm{i}\infty$ and passes on the positive side of $\tilde\lambda_{1,N}^{(b_0)}$.

Proof of Theorem 4.5

Let $\varepsilon > 0$. Using Propositions 4.18 and 4.19, there exists a $K_\varepsilon > 0$ and an event A_ε with $\mathbf{P}(A_\varepsilon > 0)$ such that

$$\max_{1 \le j \le 2\alpha} \sup_N N^{2/3} |\tilde\lambda_{j,N}^{(b)} - 2| < K_\varepsilon. \tag{4.16}$$

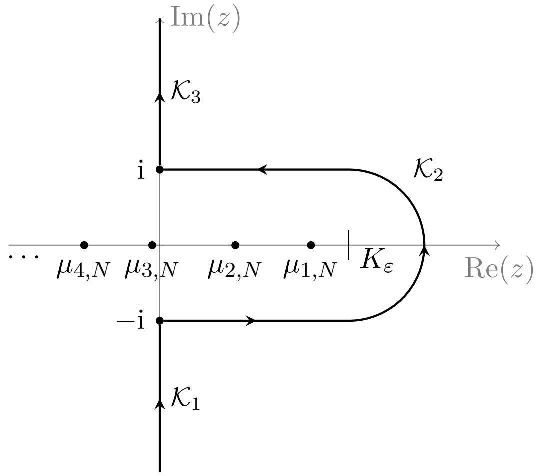


Figure 4.1: Contour of integration

Define a contour $\mathcal{K}$ that runs from $-\infty\mathrm{i}$ to $-\mathrm{i}$ along the imaginary axis, from $-\mathrm{i}$ to $+\mathrm{i}$ by crossing the real axis to the right of K_ε and then from $+\mathrm{i}$ to $+\infty$ along the imaginary axis. Call these three sections $\mathcal{K}_1, \mathcal{K}_2, \mathcal{K}_3$ respectively. The contour is shown in Fig. 4.1.

Based on this, define

$$\Gamma_N = \{2 + wN^{-\frac{2}{3}} : w \in \mathcal{K}\}.$$

Since Γ_N passes on the positive side of the zeroes of $f_N^{(1,b_0)}$ and has bounded real part, we can use it as the contour of integration for I_b. That is,

$$I_{b,b_0} = \int_{\Gamma_N} \exp\Big\{\frac{1}{\alpha}\Big[N(1 + bN^{-1/3})z - \sum_{j=1}^{N} \log(z - \tilde{\lambda}_{j,N}^{(b_0)})\Big]\Big\}\,\mathrm{d}z.$$

Making the substitution $z = 2 + wN^{-2/3}$, we have

$$\begin{aligned}
&\int_{\Gamma_N} \exp\Big\{\frac{1}{\alpha}\Big[N(1 + bN^{-1/3})z - \sum_{j=1}^{N} \log(z - \tilde{\lambda}_{j,N}^{(b_0)})\Big]\Big\}\,\mathrm{d}z \\
&\quad = \int_{\mathcal{K}} \exp\Big\{\frac{1}{\alpha}\Big[N(1 + bN^{-1/3})(2 + wN^{-2/3}) - \sum_{j=1}^{N} \log(2 + wN^{-2/3} - \tilde{\lambda}_{j,N}^{(b_0)})\Big]\Big\} N^{-2/3}\,\mathrm{d}w \\
&\quad = N^{-2/3} \exp\Big\{\frac{2N}{\alpha} + \frac{2bN^{2/3}}{\alpha}\Big\} \int_{\mathcal{K}} \exp\Big\{\frac{bw}{\alpha} + \frac{w}{\alpha}\Big[N^{1/3} - \sum_{j=1}^{N} \log(2 + wN^{-2/3} - \tilde{\lambda}_{j,N}^{(b_0)})\Big]\Big\}\,\mathrm{d}w
\end{aligned}$$

$$= N^{-2/3} \exp\Big\{\frac{2N}{\alpha} + \frac{2bN^{2/3}}{\alpha}\Big\} \int_{\mathcal{K}} \exp\Big\{\frac{bw}{\alpha}\Big\} f_N^{(\alpha,b_0)}(w)^{-1}\,\mathrm{d}w.$$

Thus, we have

$$N^{2/3} \exp\Big\{-\frac{2N}{\alpha} - \frac{2bN^{2/3}}{\alpha}\Big\} |\varphi_N^{(b_0)}(2)|^{1/\alpha} I_{b,b_0} = \int_{\mathcal{K}} e^{bw/\alpha} |\varphi_N^{(b_0)}(2)|^{1/\alpha} f_N^{(\alpha,b_0)}(w)^{-1}\,\mathrm{d}w \quad (4.17)$$

Now, Proposition 4.25 demonstrates the convergence of the integrand of I_b in the sense that that

$$e^{bw/\alpha} \frac{|\varphi_N^{(b_0)}(2)|^{1/\alpha}}{f_N^{(\alpha,b_0)}(w)} \to e^{bw/\alpha} \frac{|s_{b_0}(0)|^{1/\alpha}}{s_{b_0}^{(\alpha)}(w)} \quad (4.18)$$

uniformly in compact subsets of $\mathcal{K}$.

Consider $w = \mathrm{i}y$. We have

$$\begin{aligned} \Big| e^{bw/\alpha} \frac{|\varphi_N^{(b_0)}(2)|^{1/\alpha}}{f_N^{(\alpha,b_0)}(w)} \Big| &= \prod_{j=1}^{N} \Big| \frac{2 + \mathrm{i}yN^{-2/3} - \tilde{\lambda}_{j,N}^{(b_0)}}{2 - \tilde{\lambda}_{j,N}} \Big|^{-1/\alpha} \\ &= \prod_{j=1}^{N} \Big(1 + \frac{y^2}{N^{4/3}(2 - \tilde{\lambda}_{j,N}^{(b_0)})^2}\Big)^{-1/2\alpha} \\ &\le \prod_{j=1}^{2\alpha} \Big(1 + \frac{y^2}{N^{4/3}(2 - \tilde{\lambda}_{j,N}^{(b_0)})^2}\Big)^{-1/2\alpha} \\ &\le \Big(1 + \frac{y^2}{K_\varepsilon^2}\Big)^{-1}, \end{aligned}$$

where the last inequality follows from Eq. (4.16).

Since this is integrable over $y \in \mathbf{R}$, we have, by dominated convergence that

$$\int_{\mathcal{K}_1 \cup \mathcal{K}_3} e^{bw/\alpha} |\varphi_N^{(b_0)}(2)|^{1/\alpha} f_N^{(\alpha,b_0)}(w)^{-1}\,\mathrm{d}w \to \int_{\mathcal{K}_1 \cup \mathcal{K}_3} e^{bw/\alpha} |s_{b_0}(0)|^{1/\alpha} s_{b_0}^{(\alpha)}(w)^{-1}\,\mathrm{d}w.$$

Moreover, since $\mathcal{K}_2$ is compact, and since the convergence of Eq. (4.18) holds uniformly over it, and so the corresponding integral is also dominated and so converges. Hence,

$$\int_{\mathcal{K}} e^{bw/\alpha} |\varphi_N^{(b_0)}(2)|^{1/\alpha} f_N^{(\alpha,b_0)}(w)^{-1}\,\mathrm{d}w \to \int_{\mathcal{K}} e^{bw/\alpha} |s_{b_0}(0)|^{1/\alpha} s_{b_0}^{(\alpha)}(w)^{-1}\,\mathrm{d}w.$$

Combining this with Eq. (4.17), we conclude that Eq. (4.2) holds on the set A_ε, and in particular that that convergence holds with probability at least $1 - \varepsilon$.

Taking $\varepsilon \to 0$, it follows that Eq. (4.2) holds almost surely.

We are now nearly ready to complete the proof of the main results of Section 4.2. The last step is to address the $|\varphi_N^{(b)}(2)|$ term in Eq. (4.2), which we establish in the following lemma:

Lemma 4.26. *We have that*

$$\log|\varphi_N^{(b)}(2)| = \frac{N}{2} - \frac{1+\alpha}{6}\log N + \sqrt{\frac{\alpha}{3}\log N} \cdot Z_N + O_{\mathbf{P}}(1),$$

where $Z_N \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$.

Proof. We first expand

$$\begin{aligned}
\varphi_N^{(b)}(2) &= 2^N w_N(1)^{-1} \Psi_N^{(b)}(0) \\
&= e^N N^{-1/12} \sqrt{\frac{N!}{N^N}} \Psi_N^{(b)}(0) \\
&= (2\pi)^{1/4} e^{N/2} N^{1/6} (1 + o(1)) \cdot \Psi_N^{(b)}(0). \qquad (4.19)
\end{aligned}$$

Recalling Eq. (4.7), we have

$$\Psi_N^{(b)}(0) = N^{-1/3} C_N s_b(0)(1 + o(1)),$$

where G_N is a centered Gaussian with $\mathbf{E}G_N^2 = \frac{\alpha}{3}\log N + O(1)$.

Hence, we can write

$$\log|\Psi_N^{(b)}(0)| = -\frac{1}{3}\log N + G_N - \frac{\alpha}{6}\log N + O_{\mathbf{P}}(1). \qquad (4.20)$$

Combining Eqs. (4.20) and (4.19), we find that

$$\log|\varphi_N^{(b)}(2)| = \frac{N}{2} - \frac{1+\alpha}{6}\log N + G_N + O_{\mathbf{P}}(1), \qquad (4.21)$$

from which the result follows. □

Proof of Theorem 4.6

We have that

$$\begin{aligned} F_{\alpha,N} &= \frac{\alpha}{2N}\log Z_{\alpha,N} \\ &= \frac{\alpha}{2N}\Big[\log\Gamma\Big(\frac{N}{\alpha}\Big) - \Big(\frac{N}{\alpha}-1\Big)\log\Big(\frac{\beta N}{\alpha}\Big) + \log\Big(\frac{I_{b,b_0}}{2\pi \mathrm{i}}\Big)\Big]. \end{aligned}$$

On the one hand, we have by Stirling's formula that

$$\log\Gamma\Big(\frac{N}{\alpha}\Big) - \Big(\frac{N}{\alpha}-1\Big)\log\Big(\frac{\beta N}{\alpha}\Big) = -\frac{N}{\alpha} + \frac{1}{2}\log N - \frac{N}{\alpha}\log\beta + O(1).$$

On the other hand, Theorem 4.5 yields that

$$\log I_{b,b_0} = -\frac{2}{3}\log N + \frac{2N}{\alpha} + \frac{2bN^{2/3}}{\alpha} - \frac{1}{\alpha}\log|\varphi_N^{(b_0)}(2)| + O_{\mathbf{P}}(1).$$

Moreover, by Lemma 4.26, we write

$$\log|\varphi_N^{(b_0)}(2)| = \frac{N}{2} - \frac{1+\alpha}{6}\log N + \sqrt{\frac{\alpha}{3}\log N}\cdot Z_N + O_{\mathbf{P}}(1),$$

where $Z_N \xrightarrow{\mathrm{d}} \mathcal{N}(0,1)$.

Combining the above, we find that

$$F_{\alpha,N} = \frac{\log N}{12N} + \frac{1}{4} - \frac{1}{2}\log\beta + bN^{-1/3} + \frac{1}{N}\sqrt{\frac{\alpha}{12}\log N}\cdot Z_N + O_{\mathbf{P}}(N^{-1}).$$

Now, noting that

$$\begin{aligned} \frac{1}{4} - \frac{1}{2}\log\beta + bN^{-1/3} &= \frac{1}{4} + \frac{1}{2}bN^{-1/3} - \frac{1}{4}b^2N^{-2/3} + O(N^{-1}) \\ &= \frac{\beta^2}{4} + O(N^{-1}) \end{aligned}$$

yields

$$\frac{N}{\sqrt{\frac{\alpha}{12}\log N}}\Big(F_{\alpha,N} - \frac{\beta^2}{4} - \frac{\log N}{12N}\Big) = Z_N + o_{\mathbf{P}}(1). \quad \square$$

Proof of Theorem 4.8

This proof proceeds exactly as the proof of in the Theorem 3.1, except using a critically-spiked underlying matrix.

In this section, we will show how Theorem 4.4 can be used to generalize the necessary preliminary lemmas to the critically-spiked case, and so to conclude the result of Theorem 4.8. Specifically, we will generalize the conclusions of Lemmas 3.3, 3.9 and 3.10 and Propositions 3.2 and 3.4

First, the following result can be established from known limiting properties of the largest eigenvalues of a critically-spiked Gaussian matrix without using the stochastic Airy machinery. We need to generalize only parts (ii) and (iii) of Lemma 3.3, so we present only those results, with the numbering maintained to make the analogy clearer:

Lemma 4.27 (Equivalent of parts of Lemma 3.3)**.** *Let $\lambda_1^{(b)} \geq \cdots \geq \lambda_N^{(b)}$ be the eigenvalues of a critically-spiked Gaussian matrix. Then*

(ii) For any fixed $x \in \mathbf{R}$, there exists a constant C_x such that

$$\mathbf{E}\#\{j : \lambda_j^{(b)} \geq 2 - xN^{-2/3}\} \leq C_x.$$

(iii) For some $c_\varepsilon, N_\varepsilon$ and any $N \geq N_\varepsilon$, with probability at least $1-\varepsilon$,

$$\lambda_1^{(b)} - \lambda_2^{(b)} \geq c_\varepsilon N^{-2/3}.$$

Proof. Both of these follow from the joint limiting distribution of $(\lambda_1^{(b)}, \ldots, \lambda_k^{(b)})$ described in [BV13, Theorem 1.5].

Next, we use Eq. (4.1) to establish a limiting representation of the logarithmic derivatives of $\Psi_N^{(b)}$. In particular, for $k \in \mathbf{Z}_{\geq 0}$,

$$\begin{aligned} \partial_\lambda^k \log \Psi_N^{(b)}(\lambda) &= \mathbf{1}_{\{k=0\}}\Big(-\frac{2+\alpha}{6}\log N + G_N\Big) + \partial_\lambda^k s_b(\lambda) + O(1), \\ \partial_\lambda^k \log \varphi_N^{(b)}(2 + N^{-2/3}\lambda) &= \mathbf{1}_{\{k=0\}}\Big(-\frac{2+\alpha}{6}\log N + \sqrt{\frac{\alpha}{3}\log N}\cdot Z_N\Big) \\ &\quad - \log(2^{-N}w_N(1 + N^{-2/3}\lambda/2)) + \partial_\lambda^k \log s_b(\lambda) + O(1), \end{aligned} \tag{4.22}$$

where $Z_N = G_N/\sqrt{\frac{\alpha}{3}\log N} \xrightarrow{\text{d}} \mathcal{N}(0,1)$.

Taking $k = 0$, we establish the log-determinant central limit theorem:

Proposition 4.28 (Equivalent of Proposition 3.2). *Let $\lambda_1^{(b)} \geq \cdots \geq \lambda_N^{(b)}$ be the eigenvalues of a critically-spiked Gaussian matrix with Dyson parameter $2/\alpha$ for $\alpha \in \{1,2\}$, and let $\gamma = 2 + \lambda N^{-2/3}$ for some $\lambda \in \mathbf{R}$. Then*

$$\frac{\sum_{j=1}^N \log|\gamma - \lambda_j^{(b)}| - \frac{N}{2} - N^{1/3}C + \frac{\alpha+1}{6}\log N}{\sqrt{\frac{\alpha}{3}\log N}} \xrightarrow{\mathrm{d}} \mathcal{N}(0,1).$$

Proof. This follows from Eq. (4.22), which yields

$$\begin{aligned}\sum_{j=1}^N \log|\gamma - \lambda_j^{(b)}| &= \log|\varphi_N^{(b)}(2 + \lambda N^{-2/3})| \\ &= -\frac{2+\alpha}{6}\log N + \sqrt{\frac{\alpha}{3}\log N}\cdot Z_N - \log(2^{-N}w_N(z)) + \log|s_b(\lambda)| + O(1),\end{aligned}$$

where

$$\begin{aligned}\log(2^{-N}w_N(z)) &= -Nz^2 + \frac{1}{12}\log(Nz^2) - \frac{1}{2}\log\frac{N!}{N^N} + O(1) \\ &= -Nz^2 + \frac{1}{6}\log z + \frac{N}{2} - \frac{1}{6}\log N + O(1),\end{aligned}$$

from which we see

$$\begin{aligned}\log(2^{-N}w_N(1 + \lambda N^{-2/3}/2)) &= -\frac{N}{2} - \lambda N^{1/3} - \frac{1}{6}\log N + O(1), \\ \log|\varphi_N^{(b)}(2 + N^{-2/3}\lambda)| &= \frac{N}{2} + \lambda N^{1/3} - \frac{1+\alpha}{6}\log N + \sqrt{\frac{\alpha}{3}\log N}\cdot Z_N + O_{\mathbf{P}}(1),\end{aligned}$$

and so conclude that

$$\frac{1}{\sqrt{\frac{\alpha}{3}\log N}}\Big(\log|\varphi_N^{(b)}(2 + N^{-2/3}\lambda)| - \frac{N}{2} - \lambda N^{1/3} + \frac{1+\alpha}{6}\log N\Big) \xrightarrow{\mathrm{d}} \mathcal{N}(0,1). \quad \square$$

On the other hand, taking $k = 1, 2$ yields the relevant inverse moment bounds:

Lemma 4.29 (Equivalent of Lemma 3.9). *Let $\lambda_1^{(b)} \geq \cdots \geq \lambda_N^{(b)}$ be the eigenvalues of a* *critically-spiked Gaussian matrix and let $C \in \mathbf{R}$. Then*

$$\frac{1}{N}\sum_{j=1}^{N}\frac{1}{2+\lambda N^{-2/3}-\lambda_j^{(b)}} = 1 + O_{\mathbf{P}}(N^{-1/3}), \quad \text{and} \quad \frac{1}{N}\sum_{j=1}^{N}\frac{1}{(2+\lambda N^{-2/3}-\lambda_j^{(b)})^2} = O_{\mathbf{P}}(N^{1/3}).$$

Proof. We use Eq. (4.22), for which we expand

$$\begin{aligned}
\partial_z^k \log(2^{-N} w_N(z)) &= -2Nz^{2-k}\mathbf{1}_{\{k\le 2\}} + \frac{(-1)^k(k-1)!}{6}z^{-k}, \\
\partial_\lambda^k \log(2^{-N} w_N(1+\lambda N^{-2/3}/2)) &= (2N^{2/3})^{-k}[-2N\mathbf{1}_{\{k\le 2\}} - 2\lambda N^{1/3}\mathbf{1}_{\{k=1\}} + O(1)] \\
&= -N^{1/3}\mathbf{1}_{\{k=1\}} + o(1).
\end{aligned}$$

Therefore, as in the unspiked case

$$\begin{aligned}
\frac{1}{N}\sum_{j=1}^{N}\frac{1}{(2+\lambda N^{-2/3}-\lambda_j^{(b)})^k} &= N^{\frac{2}{3}k-1}(-1)^{k+1}\cdot \partial_\lambda^k \log \varphi_N(2+\lambda N^{-2/3}) \\
&= N^{\frac{2}{3}k-1}(-1)^{k+1}(N^{1/3}\mathbf{1}_{\{k=1\}} + O_{\mathbf{P}}(1))) \\
&= \mathbf{1}_{\{k=1\}} + O(N^{\frac{2}{3}k-1}). \quad \square
\end{aligned}$$

Lemma 4.30 (Equivalent of Lemma 3.10)**.** *Let $\lambda_1^{(b)} \ge \cdots \ge \lambda_N^{(b)}$ be the eigenvalues of a critically-spiked Gaussian matrix. Then*

$$\frac{1}{N}\sum_{j=2}^{N}\frac{1}{\lambda_1^{(b)}-\lambda_j^{(b)}} = 1 + O_{\mathbf{P}}(N^{-1/3}), \qquad \text{and} \qquad \frac{1}{N}\sum_{j=2}^{N}\frac{1}{(\lambda_1^{(b)}-\lambda_j^{(b)})^2} = O_{\mathbf{P}}(N^{1/3}).$$

Proof. The proof proceeds exactly as in [JKOP21, Section 8.1], making use of Lemmas 4.27 and 4.29 exactly as that result makes use of Lemmas 3.3 and 3.9

It remains to demonstrate the asymptotic independence of the largest eigenvalue from the log-determinant in the critically-spiked case:

Proposition 4.31 (Equivalent of Proposition 3.4)**.** *Define*

$$\begin{aligned}
\xi_{1N}^{(b)} &= \Big(\frac{\alpha}{3}\log N\Big)^{-1/2}\Big[\frac{N}{2} - \frac{1+\alpha}{6}\log N - \sum_{j=1}^{N}\log|2-\lambda_j^{(b)}|\Big], \\
\xi_{2N}^{(b)} &= N^{2/3}(\lambda_1^{(b)} - 2).
\end{aligned}$$

Then $(\xi_{1N}^{(b)}, \xi_{2N}^{(b)}) = (X_N^{(b)}, Y_N^{(b)}) + o_{\mathbf{P}}(1)$*, where* $(X_N^{(b)}, Y_N^{(b)})$ *are independent* $O_{\mathbf{P}}(1)$ *random variables.*

Proof. This proof relies on results in section 5 of [JKOP21], which are not presented in this thesis. The relevant part of that argument is that it shows that $(\xi_{1N}, \xi_{2N}) = (X_N, Y_N) + o_{\mathbf{P}}(1)$, where X_N and Y_N depend on $[\mathbf{A}]_{N,N}$ through a_i and b_i for $i < N - 2N^{1/3}\log^3 N$ and $i > N - 2N^{1/3}\log^3 N$ respectively, and so are independent.

In the critically-spiked case, we apply the joint convergence of Theorem 4.4. That is,

$$(\Psi_N(\lambda), N^{1/3}\Psi_N^{(b)}(\lambda))\frac{\mathbf{E}e^{G_N}}{e^{G_N}} \xrightarrow{\text{d}} (\mathrm{SAi}_\lambda(0), -b\,\mathrm{SAi}_\lambda(0) - \mathrm{SAi}'_\lambda(0)),$$

from which we see that

$$\begin{aligned}
\log|\Psi_N^{(b)}(0)| &= -\frac{1}{3}\log N + \log|\Psi_N(0)| + O_{\mathbf{P}}(1), \\
\xi_{1N}^{(b)}\sqrt{\frac{\alpha}{3}\log N} &= \xi_{1N}\sqrt{\frac{\alpha}{3}\log N} + O_{\mathbf{P}}(1) \\
\xi_{1N}^{(b)} &= X_N + o_{\mathbf{P}}(1),
\end{aligned}$$

and thus can take $X_N^{(b)} = X_N$.

Next, let $Y_N^{(b)}$ be the largest eigenvalue of the bottom-right minor of $[\mathbf{A}]_{N,N}^{(b)}$ of size $l = 2N^{1/3}\log^3 N$. Following the proof of [JKOP21, Proposition 5.3], the difference $\lambda_1^{(b)} - Y_N^{(b)}$ is bounded in terms of the top-left $(N-l)\times(N-l)$ minor of $[\mathbf{A}]_{N,N}^{(b)}$.

But this minor is exactly $[\mathbf{A}]_{(N-1),(N-1)}$, and so it follows from the subsequent analysis of $[\mathbf{A}]_{(N-1),(N-1)}$ in [JKOP21, Proposition 5.3] that $|\lambda_1^{(b)} - Y_N^{(b)}| = O_{\mathbf{P}}(N^{-K})$ for any $K > 0$.

Noticing that $X_N^{(b)}$ and $Y_N^{(b)}$ depend on disjoint parts of the matrix $[\mathbf{A}]_{N,N}^{(b)}$ then completes the proof. □

To complete the proof of Theorem 4.8 we follow the argument of Section 3.4.3 until Eq. (3.36), where we reach

$$NF_N = N\left(-\frac{3}{4} - \frac{1}{2}\log\beta + \beta + \frac{\log N}{12N}\right) + \sqrt{\frac{\alpha}{12}\log N}\,\xi_{1N}^{(b)} + \frac{b}{2}\sqrt{\log N}\,\xi_{2N}^{(b)} + O_{\mathbf{P}}(\log\log N) \tag{4.23}$$

with $(\xi_{1N}^{(b)}, \xi_{2N}^{(b)})$ defined as in Proposition 4.31 Now, Eq. (4.4) follows from the limiting distribution of $(\xi_{1N}^{(b)}, \xi_{2N}^{(b)})$ established in Propositions 3.2 and 4.31.

Proof of Theorem 4.9

This is a straightforward application of Theorem 4.5 to Eq. (4.5), which renders

$$\begin{aligned}\log\frac{p_N(\Lambda;\beta)}{p_N(\Lambda;\beta_0)} &= \frac{2}{\alpha}N^{2/3}(b-b_0)+\log\frac{I_{b,b_0}}{I_{b_0,b_0}}+o(1)\\ &\xrightarrow{\text{a.s.}} \log\frac{\int_{\mathcal{K}} e^{bw/\alpha}s_{b_0}^{(\alpha)}(w)^{-1}\,\mathrm{d}w}{\int_{\mathcal{K}} e^{b_0 w/\alpha}s_{b_0}^{(\alpha)}(w)^{-1}\,\mathrm{d}w}.\end{aligned}$$

Chapter 5

Algorithms for simulating RMT quantities

Over the course of the investigations presented in this thesis, we performed many numerical experiments and simulations involving large random matrices. Often, our particular settings required novel algorithms for efficiently computing the specific quantities of interest to us.

This chapter contains descriptions and mathematical justifications for two such algorithms: the banded representation of multi-spiked Gaussian and Wishart matrices is discussed in Section 5.1 and the simulation of an approximation to the stochastic Airy function presented in Section 5.2.

5.1 Banded representation of multi-spiked ensembles

When investigating quantities computed from random matrices, it is often useful to sample from the distribution of the eigenvalues of a high-dimensional random matrix ensemble. This can be done naively by drawing a matrix from the ensemble and computing its eigenvalues. However, since the computation of eigenvalues of an unstructured $n \times n$ matrix requires $O(n^3)$ operations and $O(n^2)$ allocations, this becomes impractical even for moderate n

In [DE02] Dumitriu and Edelman described ensembles of tridiagonal matrices whose eigenvalue distributions match those of Gaussian and Wishart matrices. Since the eigendecomposition of an $n \times n$ tridiagonal matrix can be computed with $O(n^2)$ operations and $O(n)$ allocations, this allows us to sample eigenvalues of much larger Gaussian and Wishart matrices.

In this section, we generalize the results of [DE02] to Gaussian and Wishart matrices with k spikes. In particular, we define ensembles of banded matrices with bandwidth $2k-1$ whose eigenvalues distributions match those of k-spiked Gaussian and Wishart matrices.

All of these algorithms are implemented in my Julia library `RandomMatrixDistributions`, which can be found in the Julia package repository. The code is available at:
`github.com/damian-t-p/RandomMatrixDistributions.jl`.

5.1.1 Notation and definitions

Banded matrices: Let $l, u \leq n$. An $n \times n$ matrix M is (l,u)-*banded* if it has only zeroes below the lth subdiagonal or above that uth superdiagonal. That is,

$$M_{i,i+k} = 0 \text{ if } k > u \text{ or } k < l.$$

We also use $\mathsf{band}_k(M)$ to denote the kth band of M. That is, for k such that $-n \leq k \leq n$, we have the $(n-|k|)$-element vector given by

$$\mathsf{band}_k(M) = \begin{cases} (M_{1,1+k}, M_{2,2+k}, \dots, M_{n-k,n}) & \text{if } k \geq 0, \\ (M_{1-k,1}, M_{2-k,2}, \dots, M_{n,n+k}) & \text{if } k < 0. \end{cases}$$

Block matrices: Throughout this section, for a matrix M we will use the notation $[M]_{i:i',j:j'}$ to denote the $(i'-i+1) \times (j'-j+1)$ submatrix of M whose (l,m) entry is $([M]_{i:i',j:j'})_{l,m} = M_{i+l-1,j+m-1}$.

We will also sometimes record for clarity the dimensions of the block of a block matrix outside the matrix delimiters, for example, the following shows a decomposition of a $p \times d$ matrix M into $d \times d$, $d \times (n-d)$, $(p-d) \times d$ and $(p-d) \times (n-d)$ blocks:

$$M = \begin{array}{r} \\ {\scriptstyle d} \\ {\scriptstyle p-d} \end{array} \!\!\begin{array}{c} \begin{array}{cc} {\scriptstyle d} & {\scriptstyle n-d} \end{array} \\ \begin{pmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{pmatrix} \end{array}.$$

Complex Gaussian distribution: This section deals extensively with both real and complex Gaussian distributions. To unify this presentation, we will use the notation $\mathcal{N}_\beta(\mu, \sigma^2)$ indexed by the Dyson parameter β.

When $\beta = 1$, this denotes the usual Gaussian distribution. When $\beta = 2$, then for $\sigma \geq 0$

and $\mu \in \mathbf{C}$, we use $\mathcal{N}_2(\mu, \sigma^2)$ to denote the distribution of

$$\frac{\sigma}{\sqrt{2}}(Z_1 + \mathrm{i}Z_2) + \mu,$$

where Z_1, Z_2 are independent real standard Gaussians.

With these definitions, for $\beta = 1, 2$, if Z is an n-dimensional random vector with iid $\mathcal{N}_\beta(0, \sigma^2)$ entries, then

$$\|Z\| \sim \frac{\sigma}{\sqrt{\beta}} \chi_{\beta n}. \tag{5.1}$$

Remark 5.1. In this section, all proofs will only be done for the $\beta = 2$ case. The $\beta = 1$ case is always entirely analogous, with "GOE" and "orthogonal" matrices replacing "GUE" and "unitary" matrices throughout.

5.1.2 Preliminary results

Lemma 5.2. *Let $\beta = 2$ (resp. $\beta = 1$). Let X be a $p \times n$ matrix with iid $\mathcal{N}_\beta(0, 1)$ entries. Let Q be an $p \times p$ unitary (resp. orthogonal) random matrix.*

If Q is independent of X, then $QX \stackrel{\mathrm{d}}{=} X$, and moreover Q is independent of QX.

Proof. This follows from conditioning on Q, whereupon we use the deterministic version of this result that $(QX|Q) \stackrel{\mathrm{d}}{=} X$. Since this distribution has no dependence on Q, the result follows. □

Proposition 5.3 (QR decomposition of a matrix of iid Gaussians)**.** *Let $\beta \in \{1, 2\}$ and let X be a $p \times n$ matrix whose entries are iid with $\mathcal{N}_\beta(0, 1)$ distribution.*

Then, for any d such that $p \wedge n \le d \le p$, X can be decomposed as $X = QR$, where

1. *Q is a $p \times d$ matrix with orthonormal columns,*
2. *R is a $d \times n$ matrix whose entries strictly below the main diagonal are all 0,*
3. *The non-zero entries of R are independent with distributions given by*

$$R_{ij} \sim \begin{cases} \frac{1}{\sqrt{\beta}} \chi_{\beta(p-i+1)} & \text{if } i = j, \\ \mathcal{N}_\beta(0, 1) & \text{if } i < j. \end{cases} \tag{5.2}$$

Proof. For $d = p$, we construct Q from Householder reflections. Namely, if $x_1, \dots, x_n$ are the columns of X, define $Q_1 = I - 2vv^*/\|v\|^2$, where $v = x_1 - \|x_1\|e_1$ so that Q_1 is unitary and $Q_1 x_1 = \|x_1\|e_1$.

Moreover for $i \geq 2$, since Q_1 is unitary and independent of x_i, the $Q_1 x_i$ are iid $\mathcal{N}_\beta(0, I_p)$ random vectors. That is, the entries of $Q_1 X$ are independent with distributions given by

$$[Q_1 X]_{ij} \sim \begin{cases} \frac{1}{\sqrt{\beta}}\chi_{\beta p} & \text{if } i = j = 1, \\ 0 & \text{if } j = 1, i \geq 2, \\ \mathcal{N}_\beta(0,1) & \text{otherwise.} \end{cases}$$

We repeat this process for the lower-right submatrix $[Q_1 X]_{2:n,2:p}$, calling the corresponding Householder reflection $\tilde{Q}_2$ and defining the unitary matrix

$$Q_2 = \begin{pmatrix} 1 & 0 \\ 0 & \tilde{Q}_2 \end{pmatrix}.$$

Continuing in this way, we have that $Q := (Q_p \cdots Q_1)^*$ is a unitary matrix such that $X = QR$, where Q and R satisfy the conclusions of the proposition.

When $n < p$, since R has only zeroes below its main diagonal, $[R]_{(p+1):n,1:p} = 0$. Hence, for any d such that $n \leq d \leq p$ if we define $Q' = [Q]_{1:p,1:d}$ and $R' = [R]_{1:d,1:n}$, then we have

$$X = QR = Q'R',$$

which completes the proof. $\square$

5.1.3 Gaussian matrices

Let X is a G(U/O)E matrix and $d \in \mathbf{Z}_{>0}$. If H is a $d \times d$ diagonal matrix and V is a matrix in $\mathbf{R}^{n\times d}$ or $\mathbf{C}^{n\times d}$ in the GOE and GUE cases respectively with orthonormal columns, then we say that

$$X + \sqrt{n} V H V^*$$

is a *d-spiked G(U/O)E matrix* with spikes H.

By rotational invariance, if we are interested in the eigenvalues of such matrices, it is enough to consider V whose columns are given by $V_i = e_i$. That is,

$$X + \begin{array}{r} \\ {\scriptstyle d} \\ {\scriptstyle n-d} \end{array} \!\! \begin{array}{c} \begin{matrix} {\scriptstyle d} & \quad {\scriptstyle n-d} \end{matrix} \\ \begin{pmatrix} \sqrt{n}H & 0 \\ 0 & 0 \end{pmatrix} \end{array}. \tag{5.3}$$

This section is then devoted to finding banded matrices whose eigenvalues match those of matrices of the form Eq. (5.3).

Theorem 5.4. *Let X be an $n \times n$ G(U/O)E matrix with corresponding Dyson parameter β and let $1 \leq d \leq n$*

Then there exists an $n \times n$ Unitary or Orthogonal block-diagonal matrix P

$$P := \begin{array}{r} \\ {\scriptstyle d} \\ {\scriptstyle n-d} \end{array} \!\! \begin{array}{c} \begin{matrix} {\scriptstyle d} & \quad {\scriptstyle n-d} \end{matrix} \\ \begin{pmatrix} I & 0 \\ 0 & P_0 \end{pmatrix} \end{array}, \tag{5.4}$$

such that $\tilde{X}$ defined by $\tilde{X} := P^* X P$ is a Hermitian (d,d)-banded $n \times n$ with independent lower-triangular entries whose distributions are given by

$$\tilde{X}_{j,j-k} \sim \begin{cases} \mathcal{N}_1(0, 2/\beta) & \text{if } k = 0, \\ \mathcal{N}_\beta(0,1) & \text{if } 1 \leq k < d, \\ \frac{1}{\sqrt{\beta}} \chi_{\beta(n-d-1+j)} & \text{if } k = d. \end{cases} \tag{5.5}$$

Proof. We proceed by induction on d, with the base case $d \leq n$ understood to be trivial with $\tilde{X} = X$ and $P = I$.

Suppose that a banded decomposition as described in the theorem statement exists when $n \leq kd$ for some $k \in \mathbf{Z}_{>0}$ and consider n such that $kd < n \leq (k+1)d$. Then, decompose X as a block matrix

$$X = \begin{array}{r} \\ {\scriptstyle d} \\ {\scriptstyle n-d} \end{array} \!\! \begin{array}{c} \begin{matrix} {\scriptstyle d} & \quad {\scriptstyle n-d} \end{matrix} \\ \begin{pmatrix} X_{11} & X_{21}^* \\ X_{21} & X_{22} \end{pmatrix} \end{array},$$

where X_{11}, X_{21}, X_{22} are $d \times d$, $(n-d) \times d$ and $(n-d) \times (n-d)$ matrices respectively.

Let $X_{21} = QR$ be the QR decomposition as described in Proposition 5.3. Now, $Q^* X_{22} Q$

is an $(n-d) \times (n-d)$ GUE matrix. Since $n-d \leq kd$, under the induction hypothesis, let P_0 a be unitary block-diagonal $(n-d) \times (n-d)$ such that $[P_0]_{1:d,1:d} = I$ (if $n-d \leq d$, then P_0 is simply taken to be the identity) such that

$$\tilde{X}_{22} = P_0^*(Q^* X_{22} Q) P_0$$

satisfies Eq. (5.5). Notice that, since $[P_0]_{1:d,1:d} = I$ while $[R]_{(d+1):(n-d)} = 0$, we have that $P_0^* R = R$. Moreover, since X_{22} is a GUE independent of X_{21} and so of Q, by Lemma 5.2,$Q^* X_{22} Q$ and so $\tilde{X}_{22}$ is independent of (X_{11}, X_{21}), and so also of R.

We now have that

$$\begin{pmatrix} I_d & 0 \\ 0 & QP_0 \end{pmatrix}^* \begin{pmatrix} X_{11} & X_{21}^* \\ X_{21} & X_{22} \end{pmatrix} \begin{pmatrix} I_d & 0 \\ 0 & QP_0 \end{pmatrix} = \begin{pmatrix} X_{11} & (P_0^* Q^* X_{21})^* \\ P_0^* Q^* X_{21} & P_0^*(Q^* X_{22} Q) P_0 \end{pmatrix}$$
$$= \begin{pmatrix} X_{11} & R^* \\ R & \tilde{X}_{22} \end{pmatrix}.$$

By Proposition 5.3 and the induction hypothesis, the distributions of the entries of R and $\tilde{X}_{22}$ are such that this matrix satisfies Eq. (5.5), completing the proof. □

Corollary 5.5. *Let X be an $n \times n$ G(U/O)E matrix with corresponding Dyson parameter β and let $1 \leq d \leq n$*

Let $\tilde{X}$ be a Hermitian (d,d)-banded $n \times n$ matrix whose independent lower-triangular entries have distributions given by

$$\tilde{X}_{j,j-k} \sim \begin{cases} \mathcal{N}_1(0, 2/\beta) & \text{if } k = 0, \\ \mathcal{N}_\beta(0,1) & \text{if } 1 \leq k < d, \\ \frac{1}{\sqrt{\beta}} \chi_{\beta(n-d-j+1)} & \text{if } k = d. \end{cases}$$

Define the block matrix

$$H = \begin{array}{c} \\ d \\ n-d \end{array} \!\!\begin{array}{c} \begin{array}{cc} d & n-d \end{array} \\ \begin{pmatrix} A & 0 \\ 0 & 0 \end{pmatrix} \end{array},$$

where A is a $d \times d$ Hermitian matrix and let $1 \leq d \leq p$.

Then the eigenvalues of $\tilde{X} + H$ have the same distribution as those of $X + H$.

Proof. Let $\tilde{X}, P$ be as in Theorem 5.4. Notice that, since $[P]_{1:d,1:d} = I$, while $[H]_{1:d,1:d}$ is the only non-zero part of that matrix, $HP = P^*H = H$. Therefore,

$$\begin{aligned}\tilde{X} + H &= P^*XP + P^*HP \\ &= P^*(X+H)P.\end{aligned}$$

The result then follows from the fact that P is unitary or orthogonal. □

The above corollary gives a recipe for sampling banded matrices whose eigenvalue distributions coincide with Gaussian matrices with d spikes.

We detail this procedure in Algorithm 1. The algorithm requires the following functions, whose naming conventions follow those of the corresponding R functions so far as is practical:

- LENGTH(v): returns the length of the vector v.
- NEWBANDEDMATRIX(n, l, u): creates an empty $n \times n$ (l, u)-banded matrix.
- RNORM(n, dyson): samples a vector of n independent $\mathcal{N}_{\mathsf{dyson}}(0,1)$ random variables.
- RCHISQ(df): samples a chi-squared random variables with df degrees of freedom. If df is a vector, samples a vector whose components are independent $\chi^2_{\mathsf{df}_i}$ random variables.

5.1.4 Wishart matrices

We use the term d-spiked Wishart matrix to denote the sample covariance of Gaussian data whose covariance matrix differs from the identity by a rank-d perturbation.

Due to rotational invariance of the Gaussian distribution, it is enough to consider diagonal covariance matrices $\Sigma = I + H^*H$, where H is a rank-d diagonal matrix.

If X is a $p \times n$ matrix of iid $\mathcal{N}_\beta(0,1)$, then the eigenvalues of the matrix

$$\begin{array}{c@{}c} & \begin{array}{cc} p & p-d \end{array} \\ \begin{array}{c} p \\ p-d \end{array} & \begin{pmatrix} I+H & 0 \\ 0 & I \end{pmatrix} \end{array} XX^* \begin{array}{c@{}c} & \begin{array}{cc} p & p-d \end{array} \\ \begin{array}{c} p \\ p-d \end{array} & \begin{pmatrix} I+H^* & 0 \\ 0 & I \end{pmatrix} \end{array} \tag{5.6}$$

Algorithm 1 Sample a (d, d)-banded $n \times n$ matrix whose eigenvalues have the same distribution as those of a Gaussian matrix with spikes $h_1, \ldots, h_d$.

procedure $\textsc{rBandedGaussian}(n, h, \beta)$
  $d \leftarrow \textsc{length}(h)$
  $W \leftarrow \textsc{newBandedMatrix}(n, d, d)$
  $\mathsf{band}_0(W) \leftarrow \textsc{rnorm}(n, \mathsf{dyson} = 1) \cdot \sqrt{2/\beta}$
  for $j \in \{1, \ldots, d\}$ **do**
    $W_{ii} \leftarrow W_{ii} + h_i\sqrt{n}$
  end for
  for $k \in \{1, \ldots, d-1\}$ **do**
    $\mathsf{band}_k(W) \leftarrow \textsc{rnorm}(n-k, \mathsf{dyson} = \beta)$
    $\mathsf{band}_{-k}(W) \leftarrow \mathsf{band}_k(W)^*$
  end for
  $\mathsf{band}_d(W) \leftarrow \sqrt{\textsc{rchisq}(\mathsf{df} = \beta \cdot \textsc{seq}(n-d, 1, \mathsf{by} = -1)/\beta)}$
  $\mathsf{band}_{-d}(W) \leftarrow \mathsf{band}_d(W)$
  return W
end procedure

have the same joint distribution as the eigenvalues of a covariance matrix of $\mathcal{N}_\beta(0, I + H^*H)$ data. We will then investigate banded matrices whose eigenvalue distributions match those of Eq. (5.6).

Theorem 5.6. *Let X be an $n \times p$ matrix with iid $\mathcal{N}_\beta(0,1)$ entries. Suppose that $d \le p \le n$. Then there exist random matrices U, V and $\tilde{X}$ such that $\tilde{X} = U^*XV$ and*

- *U is a unitary or orthogonal and block-diagonal $p \times p$ matrix such $[U]_{1:d,1:d} = I$,*
- *V is an $n \times p$ matrix with orthonormal columns, and*
- *$\tilde{X}$ is a $(d, 0)$-banded $p \times p$ matrix with independent entries whose distribution is given by*

$$\tilde{X}_{j,j-k} \sim \begin{cases} \frac{1}{\sqrt{\beta}}\chi_{\beta(n-k)} & \text{if } k = 0, \\ \mathcal{N}_\beta(0,1) & \text{if } 1 \le k < d, \\ \frac{1}{\sqrt{\beta}}\chi_{\beta(p-d-1+j)} & \text{if } k = d. \end{cases} \tag{5.7}$$

Proof. We proceed by induction on p. For the base case, where $p \le d$, take $U = I_p$ and

$$V = \begin{array}{c} \\ p \\ n-p \end{array} \begin{pmatrix} \overset{p}{I} \\ 0 \end{pmatrix}.$$

Next, suppose that a decomposition described in the theorem statement is available for all matrices of Gaussians such that $p \le kd$ and let p be such that suppose that $kd < p \le (k+1)d$. Decompose X vertically into a $d \times n$ and $(p-d) \times n$ block:

$$X = \begin{array}{c} d \\ p-d \end{array} \overset{n}{\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}}.$$

Let $X_1^* = QR$ be a QR decomposition as described in Proposition 5.3, where Q and R are $n \times p$ and $p \times d$ matrices respectively. Denote the top $d \times d$ submatrix $[R]_{1:d,1:d}$ by R_0.

We then have that

$$XQ = \begin{array}{c} d \\ p-d \end{array} \overset{p}{\begin{pmatrix} R^* \\ X_2 Q \end{pmatrix}}.$$

Write $X_2' := X_2 Q$. Since Q has orthonormal columns and is independent of X_2, we have by Lemma 5.2 that X_2' is a $(p-d) \times n$ matrix of iid standard Gaussians independent of R.

Decomposing X_2' horizontally into a $(p-d) \times d$ and $(p-d) \times (n-d)$ block according to $X_2' = (X_{21}', X_{22}')$, we can write XQ as a block matrix:

$$XQ = \begin{array}{c} \\ d \\ p-d \end{array} \begin{array}{c} \begin{array}{cc} d & p-d \end{array} \\ \begin{pmatrix} R_0^* & 0 \\ X_{21}' & X_{22}' \end{pmatrix} \end{array}.$$

Now, let $X_{21}' = US$ be a QR decomposition as per Proposition 5.3 with U and S being $(p-d) \times (p-d)$ and $(p-d) \times d$ matrices respectively. We then have

$$\begin{pmatrix} I_d & 0 \\ 0 & U^* \end{pmatrix} XQ = \begin{pmatrix} R_0 & 0 \\ S & U^* X_{22}' \end{pmatrix}.$$

Once again, $U^* X'_{22}$ is a $(p-d) \times (p-d)$ matrix of iid standard Gaussians independent of R_0 and S. By the induction hypothesis, let $\tilde{U}, \tilde{V}$ be unitary $(p-d) \times (p-d)$ matrices such that $\tilde{X}_{22} = \tilde{U}^*(U^* X'_{22})\tilde{V}$, where $\tilde{X}_{22}$ satisfies Eq. (5.7).

We then have that

$$\begin{pmatrix} I_d & 0 \\ 0 & \tilde{U}U^* \end{pmatrix} XQ \begin{pmatrix} I_d & 0 \\ 0 & \tilde{V}^* \end{pmatrix} = \begin{pmatrix} R_0 & 0 \\ S & \tilde{X}_{22} \end{pmatrix}.$$

But by Lemma 5.2 and the induction hypothesis, S and $\tilde{X}_{22}$ have the right structure for these matrices to satisfy the conclusions of the theorem. □

Corollary 5.7. *Let X be a $p \times n$ matrix with iid $\mathcal{N}_\beta(0,1)$ entries.*

Let $\tilde{X}$ be a $(d,0)$-banded $p \times p$ matrix with independent entries whose distribution is given by

$$\tilde{X}_{j,j-k} \sim \begin{cases} \frac{1}{\sqrt{\beta}}\chi_{\beta(n-k)} & \text{if } k = 0, \\ \mathcal{N}_\beta(0,1) & \text{if } 1 \le k < d, \\ \frac{1}{\sqrt{\beta}}\chi_{\beta(p-d-1+j)} & \text{if } k = d. \end{cases}$$

Define the block matrix

$$H = \begin{pmatrix} A & 0 \\ 0 & I \end{pmatrix},$$

where A is a $d \times d$ matrix.

Then the eigenvalues of $H\tilde{X}\tilde{X}^*H^*$ have the same distribution as those of HXX^*H^*.

Proof. Let $\tilde{X}, U, V$ be as in Theorem 5.6. Since $[U]_{d\times d} = I$ and $[H]_{(p-d)\times(p-d)} = I$, we have that U commutes with H.

Therefore, we have

$$\begin{aligned} HXX^*H^* &= H(U\tilde{X}V^*)(U\tilde{X}V^*)^*H^* \\ &= UH\tilde{X}\tilde{X}^*H^*U^*, \end{aligned}$$

from which the result follows. □

As with Corollary 5.5, we use the above result to define a procedure for sampling (d, d)-banded matrices whose eigenvalues have the same distribution as those of a Wishart with d spikes.

The procedure is implemented in Algorithm 2 using the same functions as required for Algorithm 1.

Algorithm 2 Sample a (d, d)-banded $p \times p$ matrix whose eigenvalues have the same distribution as those of a $p \times p$ Wishart matrix with n degrees of freedom and spikes $h_1, \dots, h_d$.

```
procedure rBandedWishart(n, p, h, β)
    d ← length(h)
    W ← newBandedMatrix(n, d, 0)
    band_0(W) ← sqrt(rchisq(df = β · seq(n, n − p + 1, by = −1))/β)
    for k ∈ {1, ..., d − 1} do
        band_{−k}(W) ← rnorm(n − k, dyson = β)
    end for
    band_d(W) ← sqrt(rchisq(df = β · seq(p − d, 1, by = −1))/β)
    for i ∈ {1, ... d} do
        for j ∈ {1, ..., i} do
            W_ij ← W_ij · sqrt(1 + h_i)
        end for
    end for
    return WW*
end procedure
```

5.2 Stochastic Airy function

Following [LP21], eq. 1.5, fix an $[a, b] \subseteq \mathbf{R}$ and a realisation B of a Brownian motion. We recall from Definition 4.1 the definition of the stochastic Airy equations over $[a, b]$ with respect to the Brownian motion B:

Define the kernel

$$\begin{aligned} \mathcal{U}_\lambda(t, s) &= \frac{t^2 - s^2}{2} + \sqrt{2\gamma}(B(t) - B(s)) + \lambda(t - s) \\ &= \mathcal{U}_0(t, s) + \lambda(t - s). \end{aligned}$$

For $t \in [a, b]$, the system of integral equations

$$\Phi_\lambda(t) = c_2(\lambda) + c_1(\lambda)\mathcal{U}_\lambda(t, b) + \int_b^t \mathcal{U}_\lambda(t, u)\Phi_\lambda(u)\,\mathrm{d}u, \tag{5.8}$$

$$\varphi_\lambda(t) = c_1(\lambda) + \int_b^t \Phi_\lambda(u)\,\mathrm{d}u, \tag{5.9}$$

are the stochastic Airy equations.

In the above, it will typically be the case that $t < b$, so it is important to note that the integrals are signed.

This section describes algorithms for efficiently solving the stochastic Airy equations for a fixed realization of B. The code for the implementation of these algorithms is available at `github.com/damian-t-p/StochasticAiry.jl`.

5.2.1 Approximating the initial conditions

An important issue when simulating the stochastic Airy function is that the values $c_1(\lambda), c_2(\lambda)$ are defined only implicitly, as they are chosen so that $\mathrm{SAi}_\lambda(t)$ remains bounded as $t \to \infty$. Following the notation of [LP21], we instead find solutions $(\phi'_\lambda, \phi_\lambda)$ to the stochastic Airy equations with initial conditions given by

$$c_1(\lambda) = \mathrm{Ai}(\lambda + b), \quad c_2(\lambda) = \mathrm{Ai}'(\lambda + b). \tag{5.10}$$

We will refer to these as the deterministic Airy initial conditions. Now, [LP21, Theorem 9.5] states that, for $\varepsilon \in (0, 1/6)$ and N large, if $b = (\log N)^{1-\varepsilon}$, then there exists a Gaussian process $\mathcal{X}$ as well as functions Θ_λ and $\chi_\lambda(t)$ such that we can write

$$\phi_\lambda(t) = (\Theta_\lambda \,\mathrm{SAi}_\lambda(t) + \chi_\lambda(t)) \frac{\exp\{\int_0^b \mathcal{X}(u)\,\mathrm{d}u\}}{\mathbf{E}\exp\{\int_0^b \mathcal{X}(u)\,\mathrm{d}u\}},$$

where, for any compact $K \subseteq \mathbf{C}$, there exists a constant C such that with probability at least $1 - e^{-b}$, it holds that, for $\ell \in \{1, 2\}$,

$$\sup_{\lambda \in K} |\partial_\lambda^{\ell-1}\Theta_\lambda - 1| \leq C b^{-\varepsilon/6}, \qquad \sup_{\lambda \in K, t \in [-e^b, b/2]} |\partial_\lambda^{\ell-1}\partial_t^k \chi_\lambda(t)| \leq C N^{\ell\varepsilon} e^{-b^{3/2}/5}.$$

Notice that the time-dependent error term χ_λ is asymptotically much smaller than the space-dependent error term Θ_λ.

This suggests that for a fixed λ, $\phi_\lambda(t)$ is, up to a constant independent of t, a good approximation of $\mathrm{SAi}_\lambda(t)$. However, b must be very large for $\phi_\lambda(0)$ to be a good approximation of $\mathrm{SAi}_\lambda(0)$ up to a constant independent of λ.

In the remainder of this section, we detail how to numerically solve these $(\phi'_\lambda, \phi_\lambda)$.

5.2.2 Solving the stochastic Airy equation as a function of time

In this section, we fix $\lambda \in \mathbf{C}$ and a large b and discuss how to solve for ϕ solving the stochastic Airy equations with the deterministic Airy initial conditions given by Eq. (5.10).

In particular, we choose $a < b$ and fix B, a realized Brownian path over the interval $I = [a, b]$ with $B_b = 0$. For such a fixed B, we see that Eq. (5.8) is a Volterra equation of the second kind with a continuous kernel.

Hence, we can solve it with a standard method, by approximating the integral in Eq. (5.8) with a trapezoidal integration. To this end, we define a grid τ over $[a, b]$ as either a strictly increasing sequence $a = t_1 < \cdots < t_n = b$ or a strictly decreasing sequence $b = t_1 > \cdots > t_n = a$.

For this grid, the corresponding differences are then

$$\Delta t_k = \begin{cases} 0 & \text{if } k \in \{1, n+1\}, \\ t_k - t_{k-1} & \text{otherwise.} \end{cases}$$

Notice that these can be either positive or negative depending on whether τ is increasing or decreasing.

Last, we will say that a function $f\colon I \to \mathbf{C}$ satisfies a γ-Hölder condition with γ-Hölder modulus C if, for any s, t, we have

$$|f(s) - f(t)| \leq C|s - t|^\gamma.$$

With this, we establish the following control for errors in trapezoidal integration of γ-Hölder functions that we will need for later results.

Lemma 5.8 (Trapezoidal integration for γ-Hölder functions)**.** *Let $I \subseteq \mathbf{R}$ be an interval and $\{t_1, \ldots, t_n\}$ a grid covering I.*

Let $\{h_x\colon I \to \mathbf{C}\}_{x \in \mathcal{X}}$ be a collection of γ-Hölder functions with common γ-Hölder modulus C and let $\{n_x\}_x$ be a set of integers in $\{1, \ldots, n\}$.

If $\mathcal{H}_x$ is the trapezoidal approximation to $\int_{t_1}^{t_{n_x}} h_x(t)\,\mathrm{d}t$ given by

$$\mathcal{H}_x = \sum_{i=1}^{n_x} h_x(t_i) \frac{\Delta t_i + \Delta t_{i+1}\mathbf{1}_{\{i \leq n_x\}}}{2},$$

then, for any $x \in \mathcal{X}$,

$$\Big| \int_{t_1}^{t_{n_x}} h_x(t)\,\mathrm{d}t - \mathcal{H}_x \Big| \leq 2C \sum_{i=1}^{n_x} |\Delta t_i|^{1+\gamma}.$$

Proof. Let $\tilde{h}_x$ be the linear interpolation of h_x along the grid. We then have that, for any t falling between t_{i-1} and t_i that,

$$\begin{aligned} |h_x(t) - \tilde{h}_x(t)| &\leq |h_x(t) - h_x(t_{i-1})| + |\tilde{h}_x(t) - \tilde{h}_x(t_{i-1})| \\ &\leq |h_x(t) - h_x(t_{i-1})| + |h_x(t_i) - h_x(t_{i-1})| \\ &\leq 2C\Delta t_i^{\gamma}. \end{aligned}$$

But since $\mathcal{H}_x$ is the integral of $\tilde{h}_x$, we have that

$$\begin{aligned} \Big| \int_{t_1}^{t_{n_x}} h_x(t)\,\mathrm{d}t - \mathcal{H}_x \Big| &\leq \sum_{i=2}^{n_x} \int_{t_{i-1}}^{t_i} |h_x(t) - \tilde{h}_x(t)|\,\mathrm{d}t \\ &\leq 2C \sum_{i=1}^{n_x} \Delta t_i^{1+\gamma}. \quad \square \end{aligned}$$

We then use this result to establish the following theorem controlling the error in the solutions to a discretized version of a Volterra equation where the forcing function and kernel are γ-Hölder.

Theorem 5.9. *Let $I = [a, b]$ be an interval. Let $f\colon I \to \mathbf{C}, K\colon I \times I \to \mathbf{C}$ be bounded functions such that $K(t, t) = 0$ for all $t \in I$.*

Suppose that there exists a $C > 0$ and $\gamma \in [0, 1]$ such that, for all $s, t \in I$,

$$|f(s) - f(t)|, \quad \sup_{u \in I} |K(u, s) - K(u, t)| \leq C|s - t|^{\gamma}. \tag{5.11}$$

Let $b = t_1 > \cdots > t_n = a$ be a grid and suppose that $\varphi, \tilde{\varphi}\colon T \to \mathbf{C}$ are functions that *solve the following equations:*

$$\varphi(t_k) = f(t_k) + \int_{t_1}^{t_k} K(t_k, u)\varphi(u)\,\mathrm{d}u, \tag{5.12}$$
$$\tilde{\varphi}(t_k) = f(t_k) + \sum_{i=1}^{k-1} K(t_k, t_i)\tilde{\varphi}(t_i)\frac{\Delta t_i + \Delta t_{i+1}}{2}$$

for all $1 \leq k \leq n$. *If* $\max_{1\leq k\leq n}|\Delta t_k| \to 0$ *as* $n \to \infty$, *then*

$$\max_{1\leq k\leq n} |\varphi(t_k) - \tilde{\varphi}(t_k)| = O(n \max_{1\leq k\leq n} |\Delta t_k|^{1+\gamma}).$$

Proof. For functions of one or two variables, we will write $\|f\|_\infty = \sup_{t\in T}|f(t)|$ and $\|K\|_\infty = \sup_{s,t\in I}|K(s,t)|$. Notice that by Eq. (5.12), we have that, for all $t \in I$,

$$|\varphi(t)| \leq \|f\|_\infty + \|K\|_\infty \int_{t_1}^{t} \varphi(u)\,\mathrm{d}u,$$

from which we conclude that

$$\begin{aligned} |\varphi(t)| &\leq \|f\|_\infty \exp\{\|K\|_\infty (t - t_1)\} \\ &\leq \|f\|_\infty \exp\{\|K\|_\infty (b - a)\}, \end{aligned}$$

and so that $\|\varphi\|_\infty < \infty$.

Now, we have for any $t, s \in I$ with $t > s$ that

$$\begin{aligned} \varphi(t) - \varphi(s) &= f(t) - f(s) + \int_{t_1}^{s} (K(t,u) - K(s,u))\varphi(u)\,\mathrm{d}u + \int_{s}^{t} K(t,u)|\varphi(u)|\,\mathrm{d}u, \\ |\varphi(t) - \varphi(s)| &\leq C(1 + (t_n - t_1)\|\varphi\|_\infty)|t-s|^\gamma + \|K\|_\infty\|\varphi\|_\infty|t-s| \\ &\leq C'|t-s|^\gamma \end{aligned}$$

for a sufficiently large C' that is uniform in t, s. That is, φ is γ-Hölder.

Let $\varepsilon_k = \varphi(t_k) - \tilde{\varphi}(t_k)$ be the error in estimating φ by $\tilde{\varphi}$ at t_k. We then have that

$$\begin{aligned} \varepsilon_k &= \int_{t_1}^{t_k} K(t,u)\varphi(u)\,\mathrm{d}u - \sum_{i=1}^{k-1} K(t_k, t_i)\tilde{\varphi}(t_i)\frac{\Delta t_i + \Delta t_{i+1}}{2} \\ &= \sum_{i=1}^{k-1} K(t_k, t_i)\frac{\Delta t_i + \Delta t_{i+1}}{2}\varepsilon_i + \sum_{i=1}^{k-1}\Big[\int_{t_i}^{t_{i+1}} K(t_k, u)\varphi(u)\,\mathrm{d}u - (K(t_k, t_i)\varphi(t_i) + K(t_k, t_{i+1})\varphi(t_{i+1}))\frac{\Delta t_i}{2}\Big]. \end{aligned}$$

Now, by the γ-Hölder conditions for K and φ, we have that, for any $x \in I$,

$$\begin{aligned}|K(x,t)\varphi(t) - K(x,s)\varphi(s)| &\leq \|K\|_\infty|\varphi(t) - \varphi(s)| + \|\varphi\|_\infty|K(x,t) - K(x,s)| \\ &\leq (\|K\|_\infty C' + \|\varphi\|_\infty C)|t-s|^\gamma,\end{aligned}$$

so that as a function of t, $K(x,t)\varphi(t)$ is γ-Hölder uniformly in $x \in I$. But this means that, by Lemma 5.8, we can write

$$|\varepsilon_k| \leq \sum_{i=1}^{k-1}|K(t_k,t_i)|\frac{|\Delta t_i + \Delta t_{i+1}|}{2}|\varepsilon_i| + \sum_{i=1}^{k-1} C''|\Delta t_i|^{1+\gamma},$$

where the last line holds with some constant $C'' > 0$ from the fact that φ and K are γ-Hölder and bounded.

But now, writing $\Delta t_{\max} = \max_{1\leq k\leq n}|\Delta t_k|$, we see that

$$|\varepsilon_k| \leq \|K\|_\infty \Delta t_{\max} \sum_{i=1}^{k-1}\Big(|\varepsilon_i| + \frac{C''}{\|K\|_\infty}\Delta t_{\max}^\gamma\Big). \tag{5.13}$$

One can verify by induction that for any sequence $(a_k)_k$ of non-negative real numbers that satisfies $a_k \leq C\sum_{i=1}^{k-1}(a_k + d)$ for some constants $C, d > 0$, we have that

$$a_k \leq [(C+1)^{k-1} - 1]d \leq Cdk,$$

where the second inequality holds when $C < 1$. This fact, together with Eq. (5.13) yields that $\max_{1\leq k\leq n}|\varepsilon_k| = O(n\Delta t_{\max}^{1+\gamma})$. □

Solving the discretized equation

Notice that Eq. (5.8) is of the form of Eq. (5.12), with

$$K(t,s) = \mathcal{U}_\lambda(t,s), \tag{5.14}$$

$$f(t) = c_2(\lambda) + c_1(\lambda)\mathcal{U}_\lambda(t,b). \tag{5.15}$$

Moreover, for any $\gamma \in [0, 1/2)$ it is almost surely the case that B is γ-Hölder continuous. Hence$f$ and K satisfy Eq. (5.11) for any $\gamma \in [0, 1/2)$. Therefore, Theorem 5.9 states that to approximate ϕ'_λ, it suffices to solve the equation

$$\tilde{\phi}'_\lambda(t_k) = c_2(\lambda) + c_1(\lambda)\mathcal{U}_\lambda(t_k, t_1) + \sum_{i=1}^{k} \mathcal{U}_\lambda(t_k, t_i) \frac{\Delta t_i + \Delta t_{i+1}}{2} \tilde{\phi}'_\lambda(t_i). \tag{5.16}$$

Having done this, we can also approximate ϕ_λ with $\tilde{\phi}_\lambda$ defined by the following trapezoidal rule equivalent of Eq. (5.9):

$$\tilde{\phi}_\lambda(t_k) = c_1(\lambda) + \sum_{i=1}^{k} \frac{\Delta t_i + \Delta t_{i+1} \mathbf{1}_{\{i+1<k\}}}{2} \tilde{\phi}'_\lambda(t_i). \tag{5.17}$$

From now on, we will call Eqs. (5.16) and (5.17) the τ-discretized stochastic Airy equations.

To this end, define the vectors $\boldsymbol{\phi}'_\lambda, \boldsymbol{\phi}, c_\lambda \in \mathbf{C}^n$ and matrix $\mathbf{K}_\lambda \in \mathbf{C}^{n\times n}$ by

$$\begin{aligned}
\boldsymbol{\phi}'_{\lambda,i} &= \tilde{\phi}'_\lambda(t_i), \\
\mathbf{c}_{\lambda,i} &= c_2(\lambda) + c_1(\lambda)\mathcal{U}_\lambda(t_i, t_1), \\
\mathbf{K}_{\lambda,ij} &= \begin{cases} \mathcal{U}_\lambda(t_i, t_j)\frac{\Delta t_j + \Delta t_{j+1}}{2} & \text{if } i > j, \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

In this way, we can write Eq. (5.16) as

$$\begin{aligned}
\boldsymbol{\phi}'_\lambda &= \mathbf{c}_\lambda + \mathbf{K}_\lambda \boldsymbol{\phi}'_\lambda, \\
\boldsymbol{\phi}'_\lambda &= (I - \mathbf{K}_\lambda)^{-1}\mathbf{c}_\lambda,
\end{aligned} \tag{5.18}$$

and Eq. (5.17) as

$$\boldsymbol{\phi}_\lambda = c_{1,\lambda}\mathbf{1} + \mathbf{T}\boldsymbol{\phi}'_\lambda,$$

where $\mathbf{T}$ is the $n \times n$ matrix corresponding to the trapezoidal rule given by

$$\mathbf{T}_{ij} = \begin{cases} \Delta t_j/2 & \text{if } i = j, \\ (\Delta t_j + \Delta t_{j+1})/2 & \text{if } i > j, \\ 0 & \text{otherwise.} \end{cases}$$

We then have by Theorem 5.9 and Lemma 5.8 that

$$\max_{1\le k\le n}|\boldsymbol{\phi}_{\lambda,k}-\phi_\lambda(t_k)|,\ \max_{1\le k\le n}|\boldsymbol{\phi}'_{\lambda,k}-\phi'_\lambda(t_k)|\ = O(n\max_{1\le k\le n}|\Delta t_k|^{1+\gamma}).$$

An example

Figure 5.1 shows a simulation of $(\tilde{\phi}'_0(t),\tilde{\phi}_0(t))$ with the following parameter values and an evenly-spaced grid τ:

Parameter	Value
β	1
$[a,b]$	$[-20,10]$
n	3000

Since the grid is even, we have that $|\Delta t_j|=(b-a)/n$, and so the approximation has a uniform error of order $O(n^{-1/2+\varepsilon})$ for any $\varepsilon>0$.

We see that $\tilde{\phi}_0(t)$ is qualitatively similar to $\mathrm{Ai}(t)$ but with Brownian continuity properties.

5.2.3 Efficiently solving the stochastic Airy equation over a region in the complex plane

In the applications given in Chapter 4, the quantity of interest is the function $\lambda\mapsto\mathrm{SAi}_\lambda(0)$ — that is, the stochastic Airy function as a function of λ.

In order to simulate this, we would like to compute $(\tilde{\phi}'_\lambda(0),\tilde{\phi}'_\lambda(0))$ for a single realization B of a Brownian motion and for all $\lambda\in\Lambda$ for some large set Λ. According to Eq. (5.18), this requires the computation of $(I-\mathbf{K}_\lambda)^{-1}\mathbf{c}_\lambda$ for each λ. Naively, this would require $O(|\Lambda|n^2)$ operations, which becomes impractical when $|\Lambda|$ is large.

In this subsection, we will detail an approach that takes advantage of the structure of $\mathbf{K}_\lambda$ and $\mathbf{c}_\lambda$ to perform these computations in time $O(n^3+|\Lambda|n)$.

Proposition 5.10. *Let $I=[0,b]$ be an interval with corresponding grid $\tau=\{b=t_1>\cdots>t_n=0\}$ and $\Lambda\subseteq\mathbf{C}$.*

Let $(\tilde{\phi}'_\lambda,\tilde{\phi}_\lambda)$ be the solutions to the τ-discretized stochastic Airy equations given in Eqs. (5.16) and (5.17) with the deterministic Airy initial conditions Eq. (5.10). Then, for $m\in\{1,2,3\}$ there exist:

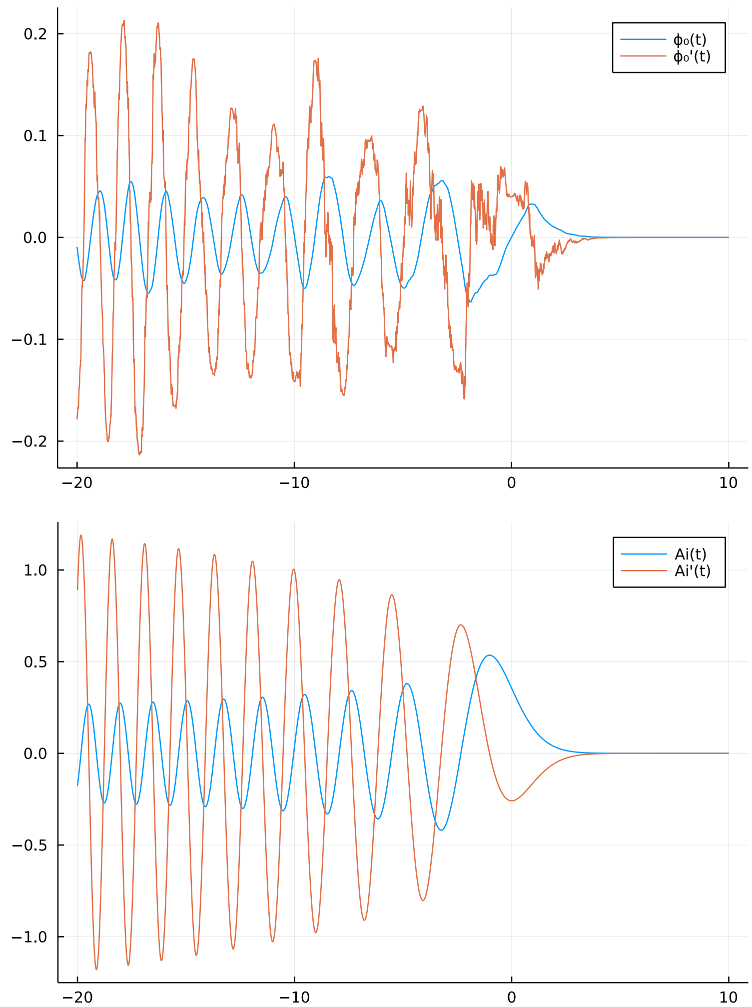


Figure 5.1: Realisation of $(\phi_0'(t), \phi_0(t))$ compared to the Airy function.

- *vectors* $\mathbf{u} \in \mathbf{C}^2$ *and* $\mathbf{v}_m \in \mathbf{C}^n$*,*
- *scalar functions* $C_m \colon \mathbf{C} \to \mathbf{C}$*,*
- *matrices* $\mathbf{U} \in \mathbf{C}^{n\times 2}$ *and* $\mathbf{N} \in \mathbf{C}^{n\times n}$*, where* $\mathbf{N}$ *is strictly lower-triangular;*

such that, for all $\lambda \in \Lambda$*, it holds that*

$$\begin{pmatrix} \tilde{\phi}'_\lambda(0) \\ \tilde{\phi}_\lambda(0) \end{pmatrix} = \mathbf{u} + \sum_{m=1}^{3} \sum_{k=0}^{n-1} C_m(\lambda)\lambda^k \cdot \mathbf{U}^{\mathrm{T}}\mathbf{N}^k\mathbf{v}_m. \tag{5.19}$$

Proof. From the results of the previous section, we can write

$$\begin{aligned} \begin{pmatrix} \tilde{\phi}'_\lambda(0) \\ \tilde{\phi}_\lambda(0) \end{pmatrix} &= \begin{pmatrix} 0 \\ c_1(\lambda) \end{pmatrix} + \begin{pmatrix} \mathbf{e}_n^{\mathrm{T}} \\ \mathbf{t}^{\mathrm{T}} \end{pmatrix} \boldsymbol{\phi}'_\lambda & (5.20) \\ &= \begin{pmatrix} 0 \\ c_1(\lambda) \end{pmatrix} + \begin{pmatrix} \mathbf{e}_n^{\mathrm{T}} \\ \mathbf{t}^{\mathrm{T}} \end{pmatrix} (I - \mathbf{K}_\lambda)^{-1}\mathbf{c}_\lambda, & (5.21) \end{aligned}$$

where $\mathbf{e}_n \in \mathbf{R}^n$ is the standard basis vector with a 1 in its nth entry, and where $\mathbf{t} \in \mathbf{R}^n$ is the vector corresponding to trapezoidal integration over the grid τ. That is,

$$\mathbf{t}_j = \frac{\Delta t_j + \Delta t_{j+1}}{2}.$$

We proceed by investigating the dependence on λ of $\mathbf{c}_\lambda$ and $\mathbf{K}_\lambda$.

Decomposition of $\mathbf{c}_\lambda$: Recall that

$$\begin{aligned} \mathbf{c}_{\lambda,i} &= c_2(\lambda) + c_1(\lambda)\mathcal{U}_\lambda(t_i, t_1) \\ &= c_2(\lambda) + c_1(\lambda)\mathcal{U}_0(t_i, t_1) + c_1(\lambda)\lambda(t_i - t_1), \end{aligned}$$

so that we can write

$$\mathbf{c}_\lambda = C_1(\lambda)\mathbf{v}_1 + C_2(\lambda)\mathbf{v}_2 + C_3(\lambda)\mathbf{v}_3, \tag{5.22}$$

for vectors $\mathbf{v}_m \in \mathbf{C}^n$ that do not depend on λ defined by

$$\begin{aligned} \mathbf{v}_{1,i} &= 1, \\ \mathbf{v}_{2,i} &= \mathcal{U}_0(t_i, t_1), \end{aligned}$$

$$\mathbf{v}_{3,i} = t_i - t_1$$

and λ-dependent scalar functions C_m defined by

$$\begin{aligned} C_1(\lambda) &= c_2(\lambda), \\ C_2(\lambda) &= c_1(\lambda), \\ C_3(\lambda) &= \lambda c_1(\lambda). \end{aligned}$$

Decomposition of $\mathbf{K}_\lambda$: Notice that $\mathbf{K}_\lambda$ can be written as

$$\mathbf{K}_\lambda = \mathbf{K}_0 + \lambda \mathbf{L},$$

where $\mathbf{L}$ is a strictly lower-triangular $n \times n$ matrix with

$$\mathbf{L}_{ij} = \begin{cases} (t_i - t_j)\frac{\Delta t_j + \Delta t_{j+1}}{2} & \text{if } i > j, \\ 0 & \text{otherwise.} \end{cases}$$

Using this identity, we can write the inverse as

$$\begin{aligned} (I - \mathbf{K}_\lambda)^{-1} &= (I - \mathbf{K}_0 - \lambda \mathbf{L})^{-1} \\ &= (I - \mathbf{K}_0)^{-1}(I - \lambda \mathbf{L}(I - \mathbf{K}_0)^{-1})^{-1}. \end{aligned}$$

Define $\mathbf{N} = \mathbf{L}(I - \mathbf{K}_0)^{-1}$. Since $\mathbf{L}$ is strictly lower-triangular and $(I - \mathbf{K}_0)^{-1}$ is lower-triangular, $\mathbf{N}$ is strictly lower triangular. We can thus take advantage of the fact that it is nilpotent, with $\mathbf{N}^n = 0$, concluding that

$$(I - \lambda \mathbf{N})^{-1} = \sum_{k=0}^{n-1} \lambda^k \mathbf{N}^k. \tag{5.23}$$

Combining: Putting together Eqs. (5.22) and (5.23), we find that

$$(I - \mathbf{K}_\lambda)^{-1}\mathbf{c}_\lambda = (I - \mathbf{K}_0)^{-1}\Big[\sum_{k=0}^{n-1} \lambda^k \mathbf{N}^k\Big]\Big[\sum_{m=1}^{3} C_m(\lambda)\mathbf{v}_m\Big].$$

Hence, if we define

$$\mathbf{U} = (I - \mathbf{K}_0^{\mathrm{T}})^{-1}(\mathbf{e}_n\ \mathbf{t}),$$

then indeed

$$\begin{pmatrix} \mathbf{e}_n^{\mathrm{T}} \\ \mathbf{t}^{\mathrm{T}} \end{pmatrix} (I - \mathbf{K}_\lambda)^{-1}\mathbf{c}_\lambda = \mathbf{U}^{\mathrm{T}}\Big[\sum_{k=0}^{n-1} \lambda^k \mathbf{N}^k\Big]\Big[\sum_{m=1}^{3} C_m(\lambda)\mathbf{v}_m\Big]$$
$$= \sum_{m=1}^{3}\sum_{k=0}^{n-1} C_m(\lambda)\lambda^k \mathbf{U}^{\mathrm{T}}\mathbf{N}^k\mathbf{v}_m. \quad \square$$

Remark 5.11. The above proposition gives a recipe for efficiently computing $(\tilde{\phi}_\lambda'(0), \tilde{\phi}_\lambda(0))_{\lambda\in\Lambda}$. That is,

1. Compute $\mathbf{U} = (I - \mathbf{K}_0^{\mathrm{T}})^{-1}(\mathbf{e}_n\ \mathbf{t})$ by back-substitution ($O(n^2)$ operations, $O(n^2)$ allocations).

2. Compute $\mathbf{N} = \mathbf{L}(I - \mathbf{K}_0^{\mathrm{T}})^{-1}$ by back-substitution ($O(n^3)$ operations, $O(n^2)$ allocations)

3. For each $k \in \{1, \ldots, n-1\}$ and $m \in \{1, 2, 3\}$ ($O(n)$ iterations):

   (a) Compute $\mathbf{N}^k\mathbf{v}_m$ by multiplying $\mathbf{N}\cdot\mathbf{N}^{k-1}\mathbf{v}_m$, where $\mathbf{N}^{k-1}\mathbf{v}_m$ is stored from the previous iteration ($O(n^2)$ operations, $O(n)$ allocations)

   (b) Compute $\mathbf{U}^{\mathrm{T}}\mathbf{N}^k\mathbf{v}_m$ ($O(n)$ operations, $O(1)$ allocations)

4. For each $k \in \{1, \ldots, n-1\}$, $m \in \{1, 2, 3\}$ and $\lambda \in \Lambda$ ($O(|\Lambda|n)$ iterations):

   (a) Compute $C_m(\lambda)\lambda^k\mathbf{U}^{\mathrm{T}}\mathbf{N}^k\mathbf{v}_m$ ($O(1)$ operations, $O(1)$ allocations). In this step, λ^k can be very large while $\mathbf{U}^{\mathrm{T}}\mathbf{N}^k\mathbf{v}_m$ is very small. Hence, in practice, it is convenient to compute $\exp\{k\log\lambda + \log(\mathbf{U}^{\mathrm{T}}\mathbf{N}^k\mathbf{v}_m)\}$.

   (b) Add $C_m(\lambda)\lambda^k\mathbf{U}^{\mathrm{T}}\mathbf{N}^k\mathbf{v}_m$ to a running total ($O(1)$ operations, $O(1)$ allocations).

Compiling the above steps, we find the algorithm requires $O(n^3 + |\Lambda|n)$ operations and $O(n^2 + |\Lambda|n)$ allocations.

When $|\Lambda|$ is larger than n, for example when Λ is a fine 2-dimensional grid in $\mathbf{C}$, this is faster then the naive algorithm, which runs in $O(|\Lambda|n^2)$ time.

I have implemented this algorithm in the Julia programming language, and the code is available at `https://github.com/damian-t-p/StochasticAiry.jl`.

An example

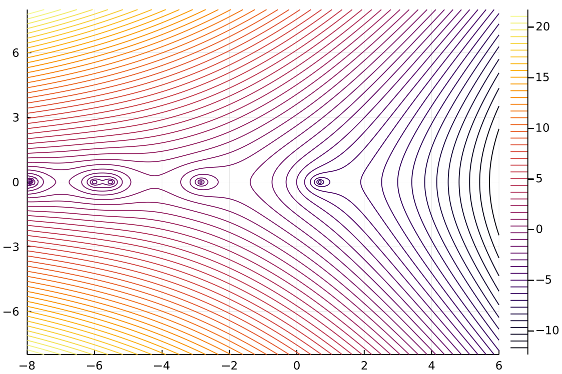
Figures 5.2 and 5.3 show a simulation of $\lambda \mapsto \tilde{\phi}_\lambda(0)$ with the following parameter values:

Parameter	Value
β	1
$[a, b]$	$[0, 10]$
n	2000

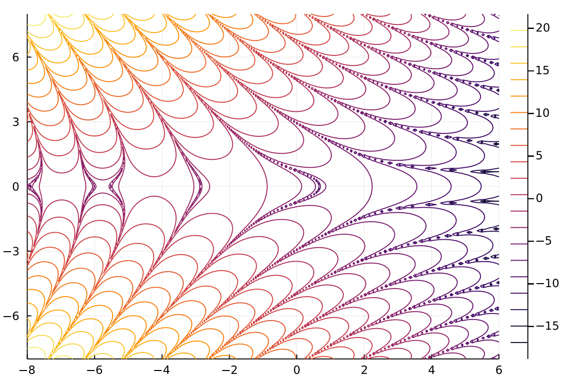
over a uniform grid τ and a 500×1000 grid of λ values ranging from $-8 - 8\mathrm{i}$ to $6 + 8\mathrm{i}$.

Note that in this case, $|\Lambda| \gg n$, so the efficient algorithm is required to compute $\tilde{\phi}_\lambda(0)$ at the desired resolution.

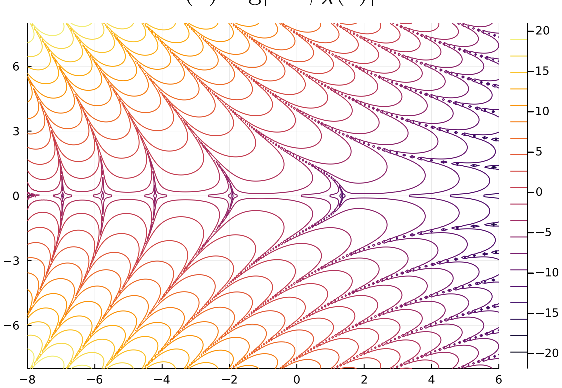
Figure 5.2a makes clear the zeroes of $\varphi_\lambda(0)$, which correspond to an Airy point field. In Fig. 5.3b, we observe steepest-descent behaviour closely mirroring that of $\mathrm{Ai}(z)$ - namely quadratic-like behaviour near the critical point with tapers off to rays emanating at angles of roughly $\pm 2\pi/3$ from the positive real axis.



(a) $\log|\tilde{\phi}_\lambda(0)|$

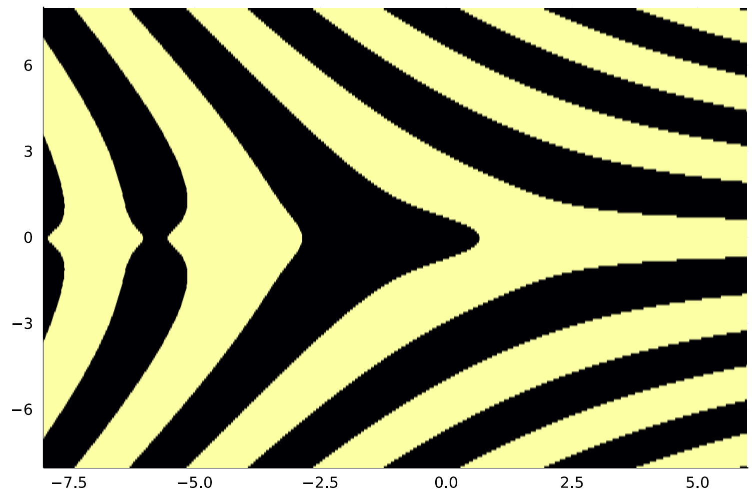


(b) $\log|\operatorname{Re}\tilde{\phi}_\lambda(0)|$

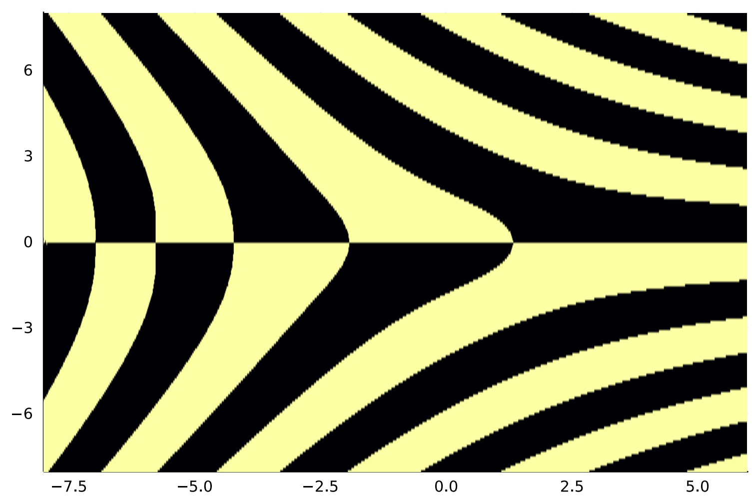


(c) $\log|\operatorname{Im}\tilde{\phi}_\lambda(0)|$

Figure 5.2: Contours of quantities associated with a realisation of $\lambda \mapsto \tilde{\phi}_\lambda(0)$



(a) $\mathrm{sign}(\mathrm{Re}\,\tilde{\phi}_\lambda(0))$



(b) $\mathrm{sign}(\mathrm{Im}\,\tilde{\phi}_\lambda(0))$

Figure 5.3: Signs of real and imaginary parts of a realisation of $\lambda \mapsto \tilde{\phi}_\lambda(0)$. Black regions are negative and yellow regions are positive

Bibliography

[AGZ09] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, 2009.

[Ass22] T. Assiotis. "Random Entire Functions from Random Polynomials with Real Zeros". Feb. 10, 2022. arXiv: 2202.03362 [math-ph].

[BBP05] J. Baik, G. Ben Arous, and S. Péché. "Phase Transition of the Largest Eigenvalue for Nonnull Complex Sample Covariance Matrices". In: *The Annals of Probability* 33.5 (Sept. 1, 2005). DOI: 10.1214/009117905000000233.

[BDG01] G. Ben Arous, A. Dembo, and A. Guionnet. "Aging of Spherical Spin Glasses". In: *Probability Theory and Related Fields* 120.1 (May 2001), pp. 1–67. DOI: 10.1007/PL00008774.

[BK18] F. Benaych-Georges and A. Knowles. "Lectures on the Local Semicircle Law for Wigner Matrices". Sept. 10, 2018. arXiv: 1601.04055 [math-ph].

[BL16] J. Baik and J. O. Lee. "Fluctuations of the Free Energy of the Spherical Sherrington-Kirkpatrick Model". In: *Journal of Statistical Physics* 165.2 (Oct. 2016), pp. 185–224. DOI: 10.1007/s10955-016-1610-0. arXiv: 1505.07349.

[BL17] J. Baik and J. O. Lee. "Fluctuations of the Free Energy of the Spherical Sherrington-Kirkpatrick Model with Ferromagnetic Interaction". In: *Annales Henri Poincaré* 18.6 (June 2017), pp. 1867–1917. DOI: 10.1007/s00023-017-0562-5. arXiv: 1607.01089.

[BLW18] J. Baik, J. O. Lee, and H. Wu. "Ferromagnetic to Paramagnetic Transition in Spherical Spin Glass". In: *Journal of Statistical Physics* 173.5 (Dec. 2018), pp. 1484–1522. DOI: 10.1007/s10955-018-2150-6. arXiv: 1805.05630.

[BV13] A. Bloemendal and B. Virág. “Limits of Spiked Random Matrices I”. In: *Probability Theory and Related Fields* 156.3-4 (Aug. 2013), pp. 795–825. DOI: 10.1007/s00440-012-0443-2.

[BY05] Z. Bai and J. Yao. “On the Convergence of the Spectral Empirical Process of Wigner Matrices”. In: *Bernoulli* 11.6 (Dec. 1, 2005). DOI: 10.3150/bj/1137421640.

[CDF09] 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”. In: *Annals of Probability* 37.1 (Jan. 2009), pp. 1–47. DOI: 10.1214/08-AOP394.

[DE02] I. Dumitriu and A. Edelman. “Matrix Models for Beta Ensembles”. In: *Journal of Mathematical Physics* 43.11 (Nov. 2002), pp. 5830–5847. DOI: 10.1063/1.1507823. arXiv: math-ph/0206043.

[Die05] M. Dieng. “Distribution Functions for Edge Eigenvalues in Orthogonal and Symplectic Ensembles: Painleve Representations”. In: *International Mathematics Research Notices* 37 (2005), pp. 2263–2287.

[DJO18] P. Dharmawansa, I. M. Johnstone, and A. Onatski. “Local Asymptotic Normality in High-Dimensional Spiked Models”. 2018. arXiv: 1411.3875 [stat].

[EKYY12] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. “Spectral Statistics of Erdős-Rényi Graphs II: Eigenvalue Spacing and the Extreme Eigenvalues”. In: *Communications in Mathematical Physics* 314.3 (Sept. 2012), pp. 587–640. DOI: 10.1007/s00220-012-1527-7. arXiv: 1103.3869.

[ES07] A. Edelman and B. D. Sutton. “From Random Matrices to Stochastic Operators”. In: *Journal of Statistical Physics* 127.6 (June 5, 2007), pp. 1121–1165. DOI: 10.1007/s10955-006-9226-4.

[EY17] L. Erdős and H.-T. Yau. *Dynamical Approach to Random Matrix Theory*. 2017.

[FR99] P. Forrester and E. Rains. “Interrelationships between Orthogonal, Unitary and Symplectic Matrix Ensembles”. In: 1999.

[Ful04] J. Fulman. “Stein’s Method, Jack Measure, and the Metropolis Algorithm”. In: *Journal of Combinatorial Theory, Series A* 108.2 (Nov. 1, 2004), pp. 275–296. DOI: 10.1016/j.jcta.2004.07.003.

[GT05] F. Götze and A. Tikhomirov. “The Rate of Convergence for Spectra of GUE and LUE Matrix Ensembles”. In: *Central European Journal of Mathematics* 3.4 (Dec. 2005), pp. 666–704. DOI: `10.2478/BF02475626`.

[Gus05] J. Gustavsson. “Gaussian Fluctuations of Eigenvalues in the GUE”. In: *Annales de l’Institut Henri Poincare (B) Probability and Statistics* 41.2 (Mar. 2005), pp. 151–178. DOI: `10.1016/j.anihpb.2004.04.002`. arXiv: `math/0401076`.

[Jam64] A. T. James. “Distributions of Matrix Variates and Latent Roots Derived from Normal Samples”. In: *The Annals of Mathematical Statistics* 35.2 (June 1964), pp. 475–501. DOI: `10.1214/aoms/1177703550`.

[JKOP20] I. M. Johnstone, Y. Klochkov, A. Onatski, and D. Pavlyshyn. “An Edge CLT for the Log Determinant of Gaussian Ensembles”. Nov. 27, 2020. arXiv: `2011.13723 [math]`.

[JKOP21] I. M. Johnstone, Y. Klochkov, A. Onatski, and D. Pavlyshyn. “Spin Glass to Paramagnetic Transition in Spherical Sherrington-Kirkpatrick Model with Ferromagnetic Interaction”. Apr. 30, 2021. arXiv: `2104.07629 [math]`.

[JO20] I. M. Johnstone and A. Onatski. “Testing in High-Dimensional Spiked Models”. In: *The Annals of Statistics* 48.3 (June 1, 2020). DOI: `10.1214/18-AOS1697`.

[Joh01] I. M. Johnstone. “On the Distribution of the Largest Eigenvalue in Principal Components Analysis”. In: *The Annals of Statistics* 29.2 (Apr. 1, 2001). DOI: `10.1214/aos/1009210544`.

[KTJ76] J. M. Kosterlitz, D. J. Thouless, and R. C. Jones. “Spherical Model of a Spin-Glass”. In: *Physical Review Letters* 36.20 (May 17, 1976), pp. 1217–1220. DOI: `10.1103/PhysRevLett.36.1217`.

[KY13a] A. Knowles and J. Yin. “Eigenvector Distribution of Wigner Matrices”. In: *Probability Theory and Related Fields* 155.3-4 (Apr. 2013), pp. 543–582. DOI: `10.1007/s00440-011-0407-y`.

[KY13b] A. Knowles and J. Yin. “The Isotropic Semicircle Law and Deformation of Wigner Matrices”. In: *Communications on Pure and Applied Mathematics* 66.11 (Nov. 2013), pp. 1663–1749. DOI: `10.1002/cpa.21450`.

[Le 60] L. Le Cam. “Locally Asymptotically Normal Families of Distributions”. In: *University of California Publications in Statistics* 3 (1960), pp. 37–98.

[LP21] G. Lambert and E. Paquette. “Strong Approximation of Gaussian β-Ensemble Characteristic Polynomials: The Edge Regime and the Stochastic Airy Function”. July 30, 2021. arXiv: 2009.05003 [math-ph].

[LS19] B. Landon and P. Sosoe. “Applications of Mesoscopic CLTs in Random Matrix Theory”. Nov. 27, 2019. arXiv: 1811.05915 [math].

[LSX20] Y. Li, K. Schnelli, and Y. Xu. “Central Limit Theorem for Mesoscopic Eigenvalue Statistics of Deformed Wigner Matrices and Sample Covariance Matrices”. Sept. 18, 2020. arXiv: 1909.12821 [math].

[Mai07] M. Maida. “Large Deviations for the Largest Eigenvalue of Rank One Deformations of Gaussian Ensembles”. In: *Electronic Journal of Probability* 12 (none Jan. 1, 2007). DOI: 10.1214/EJP.v12-438.

[Ona14] A. Onatski. “Detection of Weak Signals in High-Dimensional Complex-Valued Data”. In: *Random Matrices: Theory and Applications* 03.01 (Jan. 2014), p. 1450001. DOI: 10.1142/S2010326314500014.

[Péc06] S. Péché. “The Largest Eigenvalue of Small Rank Perturbations of Hermitian Random Matrices”. In: *Probability Theory and Related Fields* 134.1 (Jan. 2006), pp. 127–173. DOI: 10.1007/s00440-005-0466-z.

[PR29] M. Plancherel and W. Rotach. “Sur les valeurs asymptotiques des polynomes d’Hermite”. In: *Commentarii Mathematici Helvetici* 1.1 (Dec. 1, 1929), pp. 227–254. DOI: 10.1007/BF01208365.

[PT07] D. Panchenko and M. Talagrand. “On the Overlap in the Multiple Spherical SK Models”. In: *The Annals of Probability* 35.6 (Nov. 1, 2007). DOI: 10.1214/009117907000000015.

[RRV11] J. Ramírez, B. Rider, and B. Virág. “Beta Ensembles, Stochastic Airy Spectrum, and a Diffusion”. In: *Journal of the American Mathematical Society* 24.4 (May 6, 2011), pp. 919–944. DOI: 10.1090/S0894-0347-2011-00703-0.

[Rud87] W. Rudin. *Real and Complex Analysis*. 3rd ed. Singapore: McGraw-Hill Book Company, 1987.

[Tal06] M. Talagrand. “Free Energy of the Spherical Mean Field Model”. In: *Probability Theory and Related Fields* 134.3 (Mar. 2006), pp. 339–382. DOI: 10.1007/s00440-005-0433-8.

[TV10] T. Tao and V. Vu. "Random Matrices: Universality of Local Eigenvalue Statistics up to the Edge". In: *Communications in Mathematical Physics* 298.2 (Sept. 2010), pp. 549–572. DOI: 10.1007/s00220-010-1044-5.

[TV11a] T. Tao and V. Vu. "Random Matrices: The Four Moment Theorem for Wigner Ensembles". Dec. 8, 2011. arXiv: 1112.1976 [math].

[TV11b] T. Tao and V. Vu. "Random Matrices: Universality of Local Eigenvalue Statistics". In: *Acta Mathematica* 206.1 (2011), pp. 127–204. DOI: 10.1007/s11511-011-0061-3.

[TV12] T. Tao and V. Vu. "A Central Limit Theorem for the Determinant of a Wigner Matrix". Mar. 29, 2012. arXiv: 1111.6300 [math].

[TW94] C. A. Tracy and H. Widom. "Level-Spacing Distributions and the Airy Kernel". In: *Communications in Mathematical Physics* 159 (1994), pp. 151–174.

[TW96] C. A. Tracy and H. Widom. "On Orthogonal and Symplectic Matrix Ensembles". In: *Communications in Mathematical Physics* 177.3 (Jan. 1996), pp. 727–754.

[Wig58] E. P. Wigner. "On the Distribution of the Roots of Certain Symmetric Matrices". In: *Annals of Mathematics* 67.2 (1958), pp. 325–327. DOI: 10.2307/1970008. JSTOR: 1970008.