



Free Energy Fluctuations of the Bipartite Spherical SK Model at Critical Temperature

Elizabeth W. Collins-Woodfin and Han Gia Le 

Abstract. The spherical Sherrington–Kirkpatrick (SSK) model and its bipartite analog both exhibit the phenomenon that their free energy fluctuations are asymptotically Gaussian at high temperature but asymptotically Tracy–Widom at low temperature. This was proved in two papers by Baik and Lee, for all non-critical temperatures. The case of the critical temperature was recently computed for the SSK model in two separate papers, one by Landon and the other by Johnstone, Klochkov, Onatski, Pavlyshyn. In the current paper, we derive the critical temperature result for the bipartite SSK model. In particular, we find that the free energy fluctuations exhibit a transition when the temperature is in a window of size $n^{-1/3}\sqrt{\log n}$ around the critical temperature, the same window as for the SSK model. Within this transitional window, the asymptotic fluctuations of the free energy are the sum of independent Gaussian and Tracy–Widom random variables.

1. Introduction

The Sherrington–Kirkpatrick (SK) and spherical Sherrington–Kirkpatrick (SSK) models devised in the 1970s are two classical examples of mean-field spin models in which the magnetic behavior of N particles, encoded in a spin vector σ , is governed by their identically distributed random pairwise interactions. The SK model has Ising spins $\sigma \in \{-1, 1\}^N$, and SSK is the continuous analog with $\sigma \in \{\mathbb{R}^N : \|\sigma\|^2 = N\}$. For a detailed exposition on these models, we refer readers to the book by Panchenko [44]. One limitation of these models is that of their mean-field structure, meaning that all pairs of particles interact according to the same rule. With the aim of reflecting inhomogeneities and community structures (e.g., in theoretical biology, social and neural networks), scholars have developed various extensions beyond mean-field models.

One extension is the multi-species model, in which the set of N spins is partitioned into a fixed number of disjoint subsets or “species” [15]. The

random interactions between spins are not identically distributed as in the SK and SSK models, but rather have variances depending on the species structure. For a k -species model, the covariance structure can be encoded in a $k \times k$ matrix Δ^2 , where $\Delta_{s,t}^2$ denotes the variance of the random interaction between a spin in species s and a spin in species t . In bipartite models, $k = 2$ and $\Delta_{s,s}^2 = 0$ and $\Delta_{s,t}^2 > 0$ for all $s \neq t$, meaning that interactions are only between spins of different species. Bipartite models have important applications in biology and neural networks [1, 17, 19]. Another multi-species model (with applications in artificial intelligence) is the deep Boltzmann machine, where the species or “layers” are ordered, and interactions are only between spins in adjacent layers [4–6, 32, 49].

Another direction of generalizing the SK and SSK models is to allow interactions, not only between pairs, but among groups of spins. A p -spin model has interactions among groups of p spins. Likewise, a (p, q) -spin bipartite model has interactions between a group of p spins from one species and a group of q spins from the other species. The case of spherical spins for this model was studied by Auffinger and Chen [9], where they obtained a minimization formula for the limiting free energy at sufficiently high temperature.

The current paper focuses on the bipartite $(1, 1)$ -spin SSK model. The setup for this model is as follows. Given two positive integers n, m , we define spin variables

$$\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \dots, \sigma_n) \in S_{n-1}, \quad \boldsymbol{\tau} = (\tau_1, \tau_2, \dots, \tau_m) \in S_{m-1},$$

where

$$S_{n-1} = \{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u}\|^2 = n\}.$$

The Hamiltonian for the model is given by

$$H(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \frac{1}{\sqrt{n+m}} \sum_{i=1}^n \sum_{j=1}^m J_{ij} \sigma_i \tau_j$$

where J_{ij} are independent, standard Gaussian random variables. The Gibbs measure and the free energy for this model at inverse temperature $\beta > 0$ are

$$p(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \frac{1}{Z_{n,m}} e^{\beta H(\boldsymbol{\sigma}, \boldsymbol{\tau})}, \quad F_{n,m}(\beta) = \frac{1}{n+m} \log Z_{n,m}, \quad (1.1)$$

respectively, where $Z_{n,m}$ is a normalization factor (i.e. partition function),

$$Z_{n,m} = \int_{S_{m-1}} \int_{S_{n-1}} e^{\beta H(\boldsymbol{\sigma}, \boldsymbol{\tau})} d\omega_n(\boldsymbol{\sigma}) d\omega_m(\boldsymbol{\tau}), \quad (1.2)$$

and $d\omega_n$ is the uniform probability measure on S_{n-1} .

1.1. Background and Related Literature

The free energy of SK and SSK has been well studied, although more is known in the spherical setting. The limiting free energy was first conjectured by Parisi for SK [47] and by Crisanti and Sommers for SSK [26]. Both conjectures were rigorously proved by Talagrand [54, 55]. The fluctuations of the SK model are

only known at high temperature [2, 14, 25, 31], but more is known for the spherical model, where additional analytic techniques are available. In 2016, Baik and Lee analyzed the fluctuations of the SSK free energy at non-critical temperature and found that the fluctuations at high temperature are asymptotically Gaussian while those at low temperature are asymptotically Tracy–Widom [12]. The fluctuations at the critical temperature were left open.

The fluctuations at the critical temperature of the SSK free energy were studied by Landon [39] and by Johnstone, Klochkov, Onatski, and Pavlyshyn [36], independently. Both papers showed that the critical scaling for the inverse temperature is $\beta = \beta_c + bn^{-1/3}\sqrt{\log n}$. Landon proved that, for fixed $b \leq 0$ and for $b \rightarrow 0$, the fluctuations are Gaussian while, for $b \rightarrow +\infty$ at any rate, the fluctuations are Tracy–Widom. For fixed $b > 0$, Landon showed tightness but did not obtain the limiting distribution. On the other hand, Johnstone et al. were able to compute fluctuations for all fixed b . Their result for $b \leq 0$ agrees with that of Landon and, for $b > 0$, they showed that the fluctuations are a sum of independent Gaussian and Tracy–Widom random variables.

Departure from the mean-field structure generally leads to a more challenging analysis. While the problem of the limiting free energy is solved for general one-species mixed p -spin SK and SSK models [22, 45, 47, 54, 55], limiting results remain incomplete for the multi-species and (p, q) -spin bipartite models. For the multi-species SK model, the limiting free energy is only verified under the assumption of positive definite Δ^2 (Barra et al. [15] proposed a Parisi-type formula and proved an upper bound, and Panchenko [46] proved a matching lower bound). For general Δ^2 , we only have a lower bound [46]. The bipartite model, one of the most natural multi-species examples, belongs to the indefinite Δ^2 case and is still open in the case of Ising spins (a conjecture on the limiting free energy was made [16, 18]). When it comes to fluctuations, a central limit theorem (CLT) for the free energy of the two-species SK model for general Δ^2 was obtained at high temperature by [41].

For the bipartite SSK model, more is known. Baik and Lee [13] obtained both the limit and the asymptotic fluctuations of the free energy, at all non-critical temperatures. More specifically, assuming $n, m \rightarrow \infty$ with $n/m = \lambda + O(n^{-1-\delta})$ for some $\lambda, \delta > 0$, they provided explicit formulas for the first two terms in the asymptotic expansion of the free energy for $\beta \neq \beta_c$, where the critical inverse temperature β_c is equal to $\sqrt{1 + \lambda}/\lambda^{1/4}$. The formulas imply that fluctuation is Gaussian with order n^{-1} for $\beta < \beta_c$ (high temperature) and is GOE Tracy–Widom of order $n^{-2/3}$ for $\beta > \beta_c$ (low temperature).

See [9, 20, 27, 52, 53] for high-temperature results for more general Ising or spherical spin models.

1.2. Main Theorem

The goal of this paper is to compute the fluctuations of the free energy in a transitional window around the critical temperature for the bipartite (1,1)-spin SSK model. In particular, this includes detailed knowledge of the free energy at the critical temperature, providing another result on critical temperature

among spin glass models, in addition to the independent results of Landon [39] and of Johnstone et al. [36].

We state our main result in the following theorem.

Theorem 1.1. *Let $F_{n,m}(\beta)$ denote the free energy of a bipartite SSK spin glass, given by (1.1), where the species sizes n, m satisfy $n/m = \lambda + O(n^{-1})$, for some constant $\lambda \in (0, 1]$, as $n, m \rightarrow \infty$. When the inverse temperature is at the critical scaling, namely $\beta = \beta_c + bn^{-1/3}\sqrt{\log n}$ for fixed b and $\beta_c := \sqrt{1 + \lambda}/\lambda^{1/4}$, the limiting distribution of the free energy is given by the formula below and this convergence holds in distribution.*

$$\frac{n + m}{\sqrt{\frac{1}{6} \log n}} \left(F_{n,m}(\beta) - F(\beta) + \frac{1}{12} \frac{\log n}{n + m} \right) \rightarrow \mathcal{N}(0, 1) + \frac{\sqrt{6}(1 + \lambda)^{\frac{1}{2}} b_+}{\lambda^{\frac{3}{4}}(1 + \sqrt{\lambda})^{\frac{2}{3}}} \text{TW}_1 \tag{1.3}$$

where TW_1 denotes the real Tracy–Widom distribution that is independent from the standard normal $\mathcal{N}(0, 1)$ and b_+ denotes the positive part of b . The limiting free energy is given by

$$F(\beta) = \begin{cases} \frac{\beta^2}{2\beta_c^4} & \text{for } \beta < \beta_c \\ f_\lambda + \frac{\lambda}{1+\lambda} A \left((1 + \sqrt{\lambda})^2, \frac{\beta}{\sqrt{\lambda(1+\lambda)}} \right) - \frac{1}{2} \log \beta - \frac{\lambda}{2(1+\lambda)} C_\lambda & \text{for } \beta \geq \beta_c \end{cases} \tag{1.4}$$

where

$$\begin{aligned} f_\lambda &= -\frac{1}{2} + \frac{\lambda - 1}{2(\lambda + 1)} \log 2 + \frac{\lambda - 1}{4(\lambda + 1)} \log \lambda + \frac{1}{4} \log(1 + \lambda), \\ A(x, B) &= \sqrt{\alpha^2 + xB^2} - \alpha \log \left(\frac{\alpha + \sqrt{\alpha^2 + xB^2}}{2B} \right), \\ C_\lambda &= (1 - \lambda^{-1}) \log(1 + \lambda^{1/2}) + \log(\lambda^{1/2}) + \lambda^{-1/2}. \end{aligned} \tag{1.5}$$

1.3. Overview of the Proof Methods

One valuable tool in the analysis of the free energy for SSK and bipartite SSK models is a contour integral representation for the partition function ($Z_{n,m}$ in our model). A priori, the partition function of SSK is a surface integral on a high-dimensional sphere (or two spheres in the bipartite case). However, this can be rewritten in terms of contour integrals in the complex plane, which are significantly easier to analyze. The contour integral representation for the SSK partition function was first observed by Kosterlitz, Thouless, and Jones [37]. The analogous representation for the spherical bipartite model, which we use in the current paper, was derived by Baik and Lee [13].

Armed with this contour integral representation, our analysis can be broken into two broad stages: (1) use steepest descent analysis to obtain an asymptotic expansion for the free energy and (2) analyze the limiting fluctuations using tools from random matrix theory. This general procedure has been followed in several recent papers on spherical spin glasses, including [36, 39] in

their analysis of SSK at critical temperature. While much of our analysis is inspired by the methods in these two papers, the bipartite setting introduces certain technical challenges beyond those that arise for unipartite SSK.

One challenge in the bipartite setting is that the representation for $Z_{n,m}$ is a double contour integral, rather than the single integral that arises for SSK. This makes the process of contour deformation and steepest descent analysis more delicate, particularly on the low-temperature side of the critical threshold, where the contour passes very close to the (random) singularities of the integrand. Another challenge in the bipartite setting is that the underlying random matrix is a Laguerre Orthogonal Ensemble (LOE) rather than the Gaussian Orthogonal Ensemble (GOE) that appears for SSK (more background on random matrices is in Sect. 2). While these ensembles have many similarities, certain analyses are more complicated for LOE.

From the steepest descent analysis, we obtain an asymptotic expansion for the free energy near the critical temperature, which depends on a sum of the form $\sum_{i=1}^n \log(\gamma - \mu_i)$. This is a logarithmic linear statistic of the eigenvalues $\{\mu_i\}_{i=1}^n$ of LOE. The CLT for this quantity is well known in random matrix theory in the case where $\gamma - d_+ > c$ for some constant c and d_+ being the upper edge of the matrix spectrum (see, e.g., [10, 11, 42]). However, this standard CLT for linear eigenvalue statistics does not address the case where γ approaches d_+ as $n \rightarrow \infty$, which is precisely the scenario that arises when analyzing the free energy at critical temperature. Thus, we need an “edge CLT” to treat the case where $\gamma \rightarrow d_+$. A similar challenge arises for the SSK model at critical temperature, where the log linear statistic depends on eigenvalues of GOE. The edge CLT for this statistic in the GOE case can be found in [35, 38], and these works provide a necessary ingredient for the analysis of SSK free energy at critical temperature.

When we began the current project, an analogous edge CLT for LOE did not exist in the literature. To fill this gap, we proved the following theorem in a separate paper [24].

Theorem 1.2. (Collins-Woodfin, Le [24]) *Let $M_{n,m}$ be an LOE matrix with $n, m, \lambda, C_\lambda, d_+$ as above. Let $\gamma = d_+ + \sigma_n n^{-2/3}$ with $-\tau < \sigma_n \ll (\log n)^2$ for some $\tau > 0$. Then,*

$$\frac{\sum_{i=1}^n \log|\gamma - \mu_i| - C_\lambda n - \frac{1}{\lambda^{1/2}(1+\lambda^{1/2})} \sigma_n n^{1/3} + \frac{2}{3\lambda^{3/4}(1+\lambda^{1/2})^2} \sigma_n^3 + \frac{1}{6} \log n}{\sqrt{\frac{2}{3} \log n}} \rightarrow \mathcal{N}(0, 1). \quad (1.6)$$

The above result is essential in proving Theorem 1.1 as it is the source of the Gaussian term in the limiting distribution.

The last step of our proof is to show the asymptotic independence of the Gaussian and Tracy–Widom terms in the limiting distribution. This involves a recurrence on the entries of the tridiagonal representation of LOE. In the course of this analysis, we prove a result that may be of independent interest, namely that the largest eigenvalue of an $n \times n$ LOE matrix depends (asymptotically)

on a minor of size $n^{1/3} \log^3 n$. This result is well known numerically (e.g., [29]), but we have not found an explicit proof of it in the literature.

1.4. Organization

In Sect. 2, we provide a more detailed setup of the problem along with various probability, spin glass, and random matrix theory results that will be used throughout the paper. Sections 3 and 4 contain our analysis of the free energy for $\beta = \beta_c + bn^{-1/3} \sqrt{\log n}$ in the cases of $b < 0$ (high critical temperature) and $b > 0$ (low critical temperature), respectively. The case of $b = 0$ is also addressed in Sect. 4. Finally, in Sect. 5, we prove the asymptotic independence of the Gaussian and Tracy–Widom terms in the main theorem. Appendices A and B provide proofs of some technical lemmas from Sects. 2 and 5, respectively.

2. Setup and Preliminaries

2.1. Preliminaries for Bipartite SSK Model

Double contour integral representation of free energy. One of the key tools that enable us to precisely calculate the free energy and its fluctuations is a contour integral representation of the partition function. A priori, $Z_{n,m}$ is given by the surface integral in (1.2). The contour integral representation of $Z_{n,m}$ was derived by Baik and Lee [13]. For the bipartite model, we assume, without loss of generality, that $n \leq m$. We use S^{n-1} to denote the unit n -sphere (as opposed to S_{n-1} , which denotes the n -sphere of radius \sqrt{n}). Then the partition function can be written as [13]

$$Z_{n,m}(\beta) = \frac{2^n}{|S^{m-1}||S^{n-1}|} \left(\frac{\pi^2(n+m)}{m^2 n \beta^2} \right)^{\frac{n+m-4}{4}} Q(n, \alpha_n, B_n) \quad (2.1)$$

where

$$Q_n := Q(n, \alpha_n, B_n) = - \int_{\gamma_1 - i\infty}^{\gamma_1 + i\infty} \int_{\gamma_2 - i\infty}^{\gamma_2 + i\infty} e^{nG(z_1, z_2)} dz_2 dz_1 \quad (2.2)$$

and $G(z_1, z_2)$ is a random function depending on the eigenvalues $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$ of $\frac{1}{m} J J^T$. The parameters γ_1, γ_2 can be any positive real numbers satisfying $4\gamma_1 \gamma_2 > \mu_1$. The function G is defined as

$$G(z_1, z_2) := B_n(z_1 + z_2) - \frac{1}{2n} \sum_{i=1}^n \log(4z_1 z_2 - \mu_i) - \alpha_n \log z_1 \quad (2.3)$$

where

$$\alpha_n := \frac{m-n}{2n}, \quad B_n := \frac{m}{\sqrt{n(n+m)}} \beta \quad (2.4)$$

Using this contour integral representation of $Z_{n,m}(\beta)$, the free energy of the bipartite SSK is

$$\begin{aligned}
 F_{n,m}(\beta) &= \frac{1}{n+m} \log Q(n, \alpha_n, B_n) \\
 &+ \frac{1}{n+m} \log \left(\frac{2^n}{|S^{n-1}| |S^{m-1}|} \left(\frac{\pi^2(n+m)}{m^2 n \beta^2} \right)^{\frac{n+m}{4}-1} \right). \quad (2.5)
 \end{aligned}$$

By direct computation, the second term of the right-hand side is $f_\lambda - \frac{1}{2} \log \beta + \frac{\lambda}{1+\lambda} \frac{\log n}{n} + O(n^{-1})$ as $n \rightarrow \infty$, where f_λ is as defined in (1.5). We obtain

$$F_{n,m}(\beta) = \frac{1}{n+m} \log Q(n, \alpha_n, B_n) + f_\lambda - \frac{1}{2} \log \beta + \frac{\lambda}{1+\lambda} \frac{\log n}{n} + O(n^{-1}) \quad (2.6)$$

so the computation of the free energy boils down to computing the integral Q_n . In order to compute this integral via steepest descent analysis, one needs to find a critical point of $G(z_1, z_2)$. Baik and Lee show that there exists a critical point (z_1, z_2) such that both coordinates are positive real and $4z_1 z_2 > \mu_1$. We can choose the contours of the double integral to pass through this critical point, which has coordinates

$$(\gamma_1, \gamma_2) = \left(\frac{\alpha_n + \sqrt{\alpha_n^2 + \gamma B_n^2}}{2B_n}, \frac{-\alpha_n + \sqrt{\alpha_n^2 + \gamma B_n^2}}{2B_n} \right) \quad (2.7)$$

where γ is the unique real number greater than μ_1 satisfying

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{\gamma - \mu_i} = \frac{B_n^2}{\alpha_n + \sqrt{\alpha_n^2 + \gamma B_n^2}}. \quad (2.8)$$

We see that γ is implicitly a function of the eigenvalues of $\frac{1}{m} J J^T$, which is a normalized Laguerre Orthogonal Ensemble (i.e., real Wishart matrix). Later in this section we recount some important properties of this matrix ensemble that will be used throughout the paper.

Critical inverse temperature β_c and critical window. As stated above, the critical inverse temperature of the bipartite SSK model is $\beta_c = \sqrt{1 + \lambda} / \lambda^{1/4}$. At this value of β , one sees a transition in the behavior of the critical point γ . We give a brief, heuristic description of the transition here and provide more details in the next two sections.

Equation (2.8), which is random and n -dependent, can be approximated by its deterministic, n -independent analog

$$\int_{\mathbb{R}} \frac{1}{z-x} p_{\text{MP}}(x) d(x) = \frac{B^2}{\alpha + \sqrt{\alpha^2 + z B^2}} \quad (2.9)$$

where p_{MP} denotes the Marčenko–Pastur measure (see definition in Eq. (2.12)) and α, B are given by

$$\alpha := \frac{1-\lambda}{2\lambda}, \quad B := \frac{\beta}{\sqrt{\lambda(1+\lambda)}}. \quad (2.10)$$

If Eq. (2.9) is to be of any use, then it should provide a solution $z \in (d_+, \infty)$ that is close to the solution γ of (2.8) (with high probability and for all sufficiently large n). Labeling the left and right sides of (2.9) as $L_\infty(z)$ and $R_\infty(z)$, respectively, Baik and Lee [13] observe that $\frac{L_\infty(z)}{R_\infty(z)}$ is a decreasing function of $z \in (d_+, \infty)$ with

$$\lim_{z \rightarrow \infty} \frac{L_\infty(z)}{R_\infty(z)} = 0, \quad \lim_{z \downarrow d_+} \frac{L_\infty(z)}{R_\infty(z)} = \frac{L_\infty(d_+)}{R_\infty(d_+)}. \quad (2.11)$$

Hence, (2.9) has a solution $z \in (d_+, \infty)$ if and only if $L_\infty(d_+) > R_\infty(d_+)$. We call this solution $\tilde{\gamma}$. By setting $L_\infty(d_+) = R_\infty(d_+)$ and solving for β , one obtains the critical inverse temperature. The implication of this is that, for $\beta < \beta_c$ (high temperature), γ can be approximated by $\tilde{\gamma}$, and this deterministic approximation turns out to be very accurate. However, for $\beta > \beta_c$ (low temperature), (2.9) can't be used to approximate γ , since it has no solution in (d_+, ∞) . Intuitively, this is due to the fact that, at low temperature, γ is very close to the eigenvalue μ_1 and may be above or below d_+ , depending on the value of μ_1 . A detailed analysis of γ in these two cases is provided in Sects. 3 and 4.

Finally, we comment on the scaling of the critical temperature window, $\beta = \beta_c + O(n^{-1/3} \sqrt{\log n})$. One can conjecture this critical scaling from the theorem of Baik and Lee by matching the order of the variance of the free energy at high and low temperature. For fixed $\beta < \beta_c$, the free energy has variance of order $n^{-2} \log(\beta_c - \beta)$ while, for fixed $\beta > \beta_c$, the free energy has variance of order $n^{-4/3}(\beta - \beta_c)^2$. By formally equating these, we find that their order matches when $\beta - \beta_c = \Theta(n^{-1/3} \sqrt{\log n})$ and we conjecture that the variance of the free energy in this critical scaling should be of order $n^{-2} \log n$. This conjecture turns out to be correct, as we will see in the subsequent sections.

2.2. Probability and Random Matrix Preliminaries

Notational conventions (probability and asymptotics). Below are several asymptotic notations that we use along with the definitions that we follow. For any sequence $\{a_n\}$ and positive sequence $\{b_n\}$, we write

- $a_n = O(b_n)$ if there exists some constant C such that $|a_n| \leq Cb_n$ for all n ,
- $a_n = \Omega(b_n)$ if there exists some constant C such that $|a_n| \geq Cb_n$ for all n ,
- $a_n = \Theta(b_n)$ if there exist constants C_1, C_2 such that $C_1b_n \leq |a_n| \leq C_2b_n$ for all n (or, equivalently, $a_n = O(b_n)$ and $a_n = \Omega(b_n)$),
- $a_n \ll b_n$ if $\lim_{n \rightarrow \infty} a_n/b_n = 0$,
- $a_n \gg b_n$ if $\lim_{n \rightarrow \infty} b_n/a_n = 0$.

In addition, we sometimes need to make asymptotic statements about the probability of events in a sequence $\{E_n\}$. We say that E_n occurs “asymptotically almost surely” if $\mathbb{P}(E_n) \rightarrow 1$ as $n \rightarrow \infty$. We say E_n occurs “with overwhelming probability” if, for all $D > 0$, there exists n_0 such that $\mathbb{P}(E_n) > 1 - n^{-D}$ for all $n > n_0$.

Laguerre Orthogonal Ensemble and Marčenko–Pastur measure. As we saw in the previous subsection, the eigenvalues of the matrix $\frac{1}{m}JJ^T$ will play an important role in our analysis. This is a normalized Laguerre Orthogonal Ensemble, and we provide an overview of some of its key properties here. Marčenko and Pastur [43] showed that the empirical spectral measure of LOE has the following convergence, as $n, m \rightarrow \infty$ with $n/m \rightarrow \lambda \leq 1$,

$$\frac{1}{n} \sum_{i=1}^n \delta_{\mu_i}(x) \rightarrow p_{\text{MP}}(x) dx := \frac{\sqrt{(d_+ - x)(x - d_-)}}{2\pi\lambda x} \mathbb{1}_{[d_-, d_+]}(x) dx. \quad (2.12)$$

The convergence is weakly in distribution and $d_{\pm} = (1 \pm \lambda^{1/2})^2$, and $p_{\text{MP}}(x)$ is referred to as the Marčenko–Pastur measure. In working with p_{MP} , we sometimes need to use its Stieltjes transform

$$s_{\text{MP}}(z) := \int_{\mathbb{R}} \frac{1}{z - x} p_{\text{MP}}(x) dx. \quad (2.13)$$

We note that it is common to define the Stieltjes transform as the negative of what we use here. However, our definition is consistent with that of [13] and is more logical in this context, since it results in a positive value of s_{MP} for our setting.

Tracy–Widom distribution. The location of the largest eigenvalue is particularly important in our analysis. The following result is well known in random matrix theory. See, for example, [34, 50] and Corollary 1.2 of [48].

Lemma 2.1. *Let μ_1 be the largest eigenvalue of $\frac{1}{m}M_{n,m}$, where $M_{n,m}$ is an $n \times n$ matrix from the Laguerre Orthogonal Ensemble. Then the following convergence in distribution holds.*

$$\frac{m\mu_1 - (\sqrt{n} + \sqrt{m})^2}{(\sqrt{n} + \sqrt{m}) \left((1/\sqrt{n}) + (1/\sqrt{m}) \right)^{1/3}} \rightarrow \text{TW}_1.$$

Under the condition $n/m \rightarrow \lambda \in (0, 1]$, the following form of Lemma 2.1 is useful in our paper.

$$\frac{n^{2/3}(\mu_1 - d_+)}{\lambda^{1/2}(1 + \lambda^{1/2})^{4/3}} \rightarrow \text{TW}_1. \quad (2.14)$$

Classical eigenvalue locations and rigidity. A key tool in our analysis is to approximate the eigenvalues by their “classical locations” (i.e., the quantiles of the Marčenko–Pastur measure). The classical locations $\{g_i\}$ are defined by the relation

$$\frac{i}{n} = \int_{g_i}^{d_+} p_{\text{MP}}(x) dx. \quad (2.15)$$

Using this definition, one can show that

$$g_i = d_+ - \left(\frac{3\pi\lambda^{3/4}d_+i}{2n} \right)^{2/3} + O\left(\frac{i^{4/3}}{n^{4/3}} \right), \quad i \leq n/2. \quad (2.16)$$

Thus, we expect that, for $i \ll n$, we will have $\mu_i \approx d_+ - \left(\frac{3\pi\lambda^{3/4}d_+i}{2n}\right)^{2/3}$. The concept of “eigenvalue rigidity” means that eigenvalues are close to their classical locations with high probability. More precisely, we define **eigenvalue rigidity** to be the event

$$\bigcap_{1 \leq i \leq n} \left\{ |\mu_i - g_i| \leq \frac{n^\delta}{n^{2/3} \min\{i^{1/3}, (n+1-i)^{1/3}\}} \right\},$$

which holds with overwhelming probability. This is proved in [48](Theorem 3.3) in the case $\lambda \in (0, 1)$. For $\lambda = 1$, the result follows from Corollary 1.3 of [3] and the relation $p_{\text{MP}}(x) = p_{\text{SC}}(\sqrt{x})$ between the Marčenko–Pastur and semicircle distributions.

In addition to eigenvalue rigidity, we sometimes need more precise control of the larger eigenvalues. For this purpose, we introduce the following lemma, which is proved in Appendix A. This lemma is inspired by a similar one proved in [40] for GOE matrices and used by Landon in his analysis of SSK at critical temperature [39].

Lemma 2.2. *Let $\{\mu_j\}_{j=1}^n$ be the eigenvalues of $\frac{1}{m}M_{n,m}$. For each j , define*

$$A_j = \left(\frac{3\pi\lambda^{3/4}d_+j}{2}\right)^{2/3} - n^{2/3}(d_+ - \mu_j). \tag{2.17}$$

Given $\varepsilon > 0$, there exists K such that for sufficiently large n ,

$$\mathbb{P} \left(\bigcap_{K \leq j \leq n^{2/5}} \left\{ |A_j| \leq \lambda j^{2/3} \right\} \right) \geq 1 - \varepsilon. \tag{2.18}$$

Furthermore, there exists $C, c > 0$ such that

$$\mathbb{E} \left[\mathbb{1}_{\{n^{2/3}(\mu_j - d_+) \leq -C\}} |A_j| \right] \leq \frac{c \log j}{j^{1/3}}, \quad \text{for } K \leq j \leq n^{2/5}. \tag{2.19}$$

Tridiagonal representation of LOE. In Sect. 5, when proving the asymptotic independence of the Gaussian and Tracy–Widom variables, we will need the tridiagonal representation of LOE. Dumitriu and Edelman [28] show that the eigenvalue distribution of the unnormalized LOE matrix $M_{n,m}$ is the same as that of the $n \times n$ matrix $T_n = BB^T$ where B is a bi-diagonal matrix of dimension $n \times n$. In particular,

$$B = \begin{bmatrix} a_1 & & & & & & & \\ b_1 & a_2 & & & & & & \\ & b_2 & a_3 & & & & & \\ & & \ddots & \ddots & & & & \\ & & & \ddots & \ddots & & & \\ & & & & b_{n-1} & a_n & & \end{bmatrix} \quad \text{so}$$

$$BB^T = \begin{bmatrix} a_1^2 & a_1 b_1 & & & & \\ a_1 b_1 & a_2^2 + b_1^2 & a_2 b_2 & & & \\ & a_2 b_2 & a_3^2 + b_2^2 & & & \\ & & & \ddots & & \\ & & & & a_{n-1} b_{n-1} & \\ & & & & a_{n-1} b_{n-1} & a_n^2 + b_{n-1}^2 \end{bmatrix} \quad (2.20)$$

where $\{a_i\}, \{b_i\}$ are all independent random variables with distributions satisfying

$$a_i^2 \sim \chi^2(m - n + i), \quad b_i^2 \sim \chi^2(i). \quad (2.21)$$

2.3. Defining the Event on Which Our Results Hold

Our arguments throughout this paper rely upon certain conditions on the eigenvalues, which hold with probability close to 1. To streamline the later proofs, we collect in this section various events involving the eigenvalues $\{\mu_i\}$ and provide probability bounds for each event. Finally, we define \mathcal{E}_ε to be the intersection of these events, which holds with probability $1 - \varepsilon$ for arbitrarily small choice of ε .

Definition 2.3. Let δ, s, t, r, R be positive numbers where $s < t, r < R$, and let K be a positive integer. We define the events $\mathcal{F}_\delta^{(1)}, \mathcal{F}_K^{(2)}, \mathcal{F}_{s,t}^{(3)}, \mathcal{F}_{r,R}^{(4)}$ as follows.

$$\mathcal{F}_\delta^{(1)} = \bigcap_{1 \leq i \leq n} \left\{ |\mu_i - g_i| \leq \frac{n^\delta}{n^{2/3} \min\{i^{1/3}, (n+1-i)^{1/3}\}} \right\}, \quad (2.22)$$

$$\mathcal{F}_K^{(2)} = \bigcap_{K \leq j \leq n^{2/5}} \left\{ \left| n^{2/3}(\mu_j - d_+) + \left(\frac{3\pi\lambda^{3/4}d_+}{2} j \right)^{2/3} \right| \leq \frac{j^{2/3}}{10} \right\}, \quad (2.23)$$

$$\mathcal{F}_{s,t}^{(3)} = \left\{ n^{2/3} |d_+ - \mu_1| \in [s, t] \right\}, \quad 0 < s < t, \quad (2.24)$$

$$\mathcal{F}_{r,R}^{(4)} = \left\{ r < n^{2/3}(\mu_1 - \mu_2) < R \right\}. \quad (2.25)$$

Remark 2.4. The event $\mathcal{F}_\delta^{(1)}$ is the eigenvalue rigidity condition with respect to the ‘‘classical location,’’ and $\mathcal{F}_K^{(2)}$ is inspired by a similar event used in the context of Gaussian ensembles by Landon and Sosoe [40].

Lemma 2.5. (Event probability bounds) *The following statements hold.*

- For any fixed $\delta > 0$, the event $\mathcal{F}_\delta^{(1)}$ holds with overwhelming probability.
- For any $\varepsilon > 0$, there exist positive constants K, s, t, r, R depending on ε but not on n such that, for sufficiently large n ,

$$\mathbb{P}[\mathcal{F}_K^{(2)}] \geq 1 - \frac{\varepsilon}{4}, \quad \mathbb{P}[\mathcal{F}_{s,t}^{(3)}] \geq 1 - \frac{\varepsilon}{4}, \quad \mathbb{P}[\mathcal{F}_{r,R}^{(4)}] \geq 1 - \frac{\varepsilon}{4}.$$

Proof. The bounds on the first three events are clear. The eigenvalue rigidity condition $\mathcal{F}_\delta^{(1)}$ holds with overwhelming probability (see explanation in Sect. 2.2). The bound on event $\mathcal{F}_K^{(2)}$ follows directly from Lemma 2.2, where we can take larger value of K to replace ε in the bound by $\varepsilon/4$. Result on $\mathcal{F}_{s,t}^{(3)}$ is a consequence of the Tracy–Widom convergence in Lemma 2.1.

Finally, we consider $\mathcal{F}_{r,R}^{(4)}$. The upper bound $n^{2/3}(\mu_1 - \mu_2) \leq R$ holds with probability $1 - \varepsilon/8$ for some $R > 0$ via a union bound (where $|\mu_1 - d_+|$ is controlled using $\mathcal{F}_{s,t}^{(3)}$ and $|\mu_2 - d_+|$ is bounded similarly using Tracy–Widom convergence of μ_2). For the lower bound on $n^{2/3}(\mu_1 - \mu_2)$, note that the joint distribution of μ_1 and μ_2 (each rescaled as in (2.14)) converges to the distribution given by the Tracy–Widom law (see, for example, [48, 50]). This law describes the joint distribution of the largest two eigenvalues of an operator \mathbf{H}_1 whose spectrum is simple with probability one (see, for example, (4.5.9) and Theorem 4.5.42 of [7]), implying an $r > 0$ such that $\mathbb{P}(n^{2/3}(\mu_1 - \mu_2) > r) \geq 1 - \varepsilon/8$ does exist for sufficiently large n .

Definition 2.6. Given $\varepsilon > 0$, we define \mathcal{E}_ε to be an event

$$\mathcal{E}_\varepsilon := \mathcal{F}_\delta^{(1)} \cap \mathcal{F}_K^{(2)} \cap \mathcal{F}_{s,t}^{(3)} \cap \mathcal{F}_{r,R}^{(4)}$$

where the parameters δ, K, s, t, r, R are chosen to satisfy the probability bounds in Lemma 2.5. Note that K, s, t, r, R depend on ε , but δ does not. The choice of these constants is not unique. However, for any given $\varepsilon > 0$, we fix these values and define \mathcal{E}_ε accordingly.

The following corollary follows directly from the above definition and Lemma 2.5.

Corollary 2.7. For any $\varepsilon > 0$, $\mathbb{P}[\mathcal{E}_\varepsilon] \geq 1 - \varepsilon$.

Computing the free energy in both the high- and low-temperature regimes involves analyzing linear statistics of eigenvalues of the form $\sum_{i=1}^n \frac{1}{(z - \mu_i)^k}$, on the event defined above. The key lemma that we use for handling these sums is the following.

Lemma 2.8. Let $z \in \mathbb{C}$ with $\text{Re}(z) \geq d_+$. Let $\{\mu_i\}$ be the eigenvalues of $\frac{1}{m}M_{m,n}$. Then, for any $\varepsilon > 0$ and any positive integer l ,

$$\begin{aligned} & \mathbb{E} \left[\mathbb{1}_{\mathcal{E}_\varepsilon} \left| \frac{1}{n} \sum_{j=K}^n \frac{1}{(z - \mu_j)^l} - \int_{d_-}^{g_K} \frac{1}{(z - y)^l} p_{\text{MP}}(y) dy \right| \right] \\ &= O \left(n^{\frac{2}{3}l-1} \cdot \min \left\{ \left| \frac{\log(n^{2/3}|z - d_+|)}{(n^{2/3}|z - d_+|)^l} \right|, 1 \right\} \right). \end{aligned} \tag{2.26}$$

Here, K is the constant depending on ε in $\mathcal{F}_K^{(2)}$ and \mathcal{E}_ε .

A proof of this lemma is included in Appendix A. The general approach is inspired by the method that Landon and Sosoe used in [40] to bound similar eigenvalue statistics in the case of Gaussian orthogonal ensembles. We prove a series of supporting lemmas, first for LUE, which allows us to make use of the determinantal properties. We then extend our final result to LOE by way of the interrelationship between eigenvalues of unitary and orthogonal ensembles provided in [30].

3. High Temperature

As mentioned in the previous section, the computation of the free energy reduces to the computation of the integral

$$Q_n = - \int_{\gamma_1 - i\infty}^{\gamma_1 + i\infty} \int_{\gamma_2 - i\infty}^{\gamma_2 + i\infty} e^{nG(z_1, z_2)} dz_2 dz_1 \quad (3.1)$$

where $G(z_1, z_2)$ is defined in (2.3). The general idea is that we should be able to compute this integral via steepest descent analysis by deforming the contours such that they pass through the critical point (γ_1, γ_2) , which is a function of γ as defined in (2.7)–(2.8). Baik and Lee [13] show that at fixed high temperature (i.e., constant $\beta < \beta_c$), the random variable γ is well approximated by $\tilde{\gamma}$, the solution to (2.9). Furthermore, $|\gamma - \tilde{\gamma}|$ is small enough that the integral computations can be carried out with $\tilde{\gamma}$ and the error remains sufficiently small.

In the high-temperature side of the critical window, we do not have fixed $\beta < \beta_c$ as in [13], but rather $\beta = \beta_c + bn^{-1/3}\sqrt{\log n}$ for $b < 0$. The first task of this section is to show that, even in this scaling, $\tilde{\gamma}$ remains a good approximation of γ . Namely, we need to compute the asymptotics of $\tilde{\gamma}$ and obtain an upper bound on $|\gamma - \tilde{\gamma}|$.

3.1. Bounds on G , Its Derivatives, and Its Critical Point

We begin with an asymptotic expansion for $\tilde{\gamma}$.

Lemma 3.1. *For fixed $b < 0$, the solution $\tilde{\gamma}$ to (2.9) satisfies*

$$\tilde{\gamma} = d_+ + \frac{4\lambda b^2}{1 + \lambda} n^{-2/3} \log n + O(n^{-1}(\log n)^{3/2}).$$

Proof. From [13] (see (6.17)), we obtain the closed-form expression

$$\tilde{\gamma} = (1 + \lambda)\beta^{-2} + 1 + \lambda + \frac{\lambda}{1 + \lambda}\beta^2. \quad (3.2)$$

Observe that the right-hand side, as a function of β , is equal to d_+ at β_c . Thus, by expanding the function around $\beta = \beta_c + bn^{-1/3}\sqrt{\log n}$, we obtain

$$\tilde{\gamma} - d_+ = \frac{4\lambda}{1 + \lambda}(\beta - \beta_c)^2 - \frac{4\lambda^{5/4}}{(1 + \lambda)^{3/2}}(\beta - \beta_c)^3 + O((\beta - \beta_c)^4), \quad (3.3)$$

and the lemma follows. \square

In order to obtain a sufficiently tight bound for $|\gamma - \tilde{\gamma}|$, we need bounds on various eigenvalue statistics and, in particular, we need to bound differences of the form

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{(z - \mu_i)^k} - \int \frac{p_{\text{MP}}(y) dy}{(z - y)^k}, \quad k \geq 1 \quad (3.4)$$

when z is close to μ_1 . Given the precision needed for computations in the critical window, the bound obtained using eigenvalue rigidity is not tight enough. Instead, we make use of the following lemma.

Lemma 3.2. *Let $z \in \mathbb{C}$ with $\operatorname{Re}(z) \geq d_+$ and $|z - d_+| > cn^{-2/3} \log n$ for some $c > 0$. Let $\{\mu_i\}$ be the eigenvalues of $\frac{1}{m}M_{m,n}$. Then, for any $\varepsilon > 0$ and any positive integer l ,*

$$\mathbb{E} \left[\mathbb{1}_{\mathcal{E}_\varepsilon} \left| \frac{1}{n} \sum_{j=1}^n \frac{1}{(z - \mu_j)^l} - \int \frac{p_{\text{MP}}(y) dy}{(z - y)^l} \right| \right] = O \left(n^{\frac{2}{3}l-1} \frac{\log(n^{2/3}|z - d_+|)}{(n^{2/3}|z - d_+|)^l} \right). \quad (3.5)$$

Proof of Lemma 3.2. Given $\varepsilon > 0$, let K be the integer in the events $\mathcal{F}_K^{(2)}$ and \mathcal{E}_ε . Recall the classical locations g_i , $i = 0, \dots, n$ of the Marčenko–Pastur measure. We start by writing $\frac{1}{n} \sum_{i=1}^n \frac{1}{(z - \mu_i)^l} - \int \frac{1}{(z - y)^l} p_{\text{MP}}(y) dy$ as the sum

$$S_1 + S_2 = \left(\frac{1}{n} \sum_{i=1}^K \frac{1}{(z - \mu_i)^l} - \int_{g_K}^{d_+} \frac{p_{\text{MP}}(y) dy}{(z - y)^l} \right) + \left(\frac{1}{n} \sum_{i=K+1}^n \frac{1}{(z - \mu_j)^l} - \int_{d_-}^{g_K} \frac{p_{\text{MP}}(y) dy}{(z - y)^l} \right). \quad (3.6)$$

For $i \leq K$, we observe that:

- On the event \mathcal{E}_ε , $n^{2/3}(d_+ - \mu_i)$ is uniformly bounded in i . Thus, $|z - \mu_i| \geq |z - d_+| - |d_+ - \mu_i| > \frac{1}{2}|z - d_+|$ by the assumption on z .
- As $\operatorname{Re}(z) > d_+$, we have $|z - y| \geq |z - d_+|$ for all real $y < d_+$.

Therefore,

$$\mathbb{1}_{\mathcal{E}_\varepsilon} |S_1| \leq \frac{1}{n} \sum_{i=1}^K \frac{1}{|z - \mu_i|^l} + \int_{g_K}^{d_+} \frac{1}{|z - y|^l} p_{\text{MP}}(y) dy \leq \frac{3K}{n|z - d_+|^l}. \quad (3.7)$$

We then bound $\mathbb{1}_{\mathcal{E}_\varepsilon} |S_2|$ using Lemma 2.8 to complete the proof of Lemma 3.2. \square

We obtain an upper bound for $\gamma - \tilde{\gamma}$ in the following lemma. Together with Lemma 3.1, it verifies that the order of $\gamma - \tilde{\gamma}$ is strictly less than that of $\tilde{\gamma} - d_+$.

Lemma 3.3. *If $b < 0$, then, on the event \mathcal{E}_ε for any given $\varepsilon > 0$,*

$$|\gamma - \tilde{\gamma}| = O \left(\frac{(\log \log n)^2}{n^{2/3} \sqrt{\log n}} \right).$$

Proof. Recall that γ and $\tilde{\gamma}$ are solutions to the equations $L(x) = R(x)$ and $L_\infty(x) = R_\infty(x)$, respectively, where

$$L(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{x - \mu_i(n)}, \quad R(x) = \frac{B_n^2}{\alpha_n + \sqrt{\alpha_n^2 + xB_n^2}}$$

and

$$L_\infty(x) = \int_{\mathbb{R}} \frac{p_{\text{MP}}(y) dy}{x - y}, \quad R_\infty(x) = \frac{B^2}{\alpha + \sqrt{\alpha^2 + xB^2}}.$$

Define $F(x) = R(x)/L(x)$ and let $F_\infty(x)$ be given similarly. Setting $\varepsilon_n = \frac{(\log \log n)^2}{n^{2/3} \sqrt{\log n}}$, we follow the method in [13] to prove $|\gamma - \tilde{\gamma}| = O(\varepsilon_n)$ by showing $F'(\tilde{\gamma} - \varepsilon_n) < 1 < F'(\tilde{\gamma} + \varepsilon_n)$. Since $F_\infty(\tilde{\gamma}) = 1$ and $F_\infty(\tilde{\gamma} - \varepsilon_n) < 1 < F_\infty(\tilde{\gamma} + \varepsilon_n)$, it suffices to show

$$|F(x) - F_\infty(x)| \ll |F'_\infty(\tilde{\gamma})| \varepsilon_n, \quad \text{for } x \in [\tilde{\gamma} - \varepsilon_n, \tilde{\gamma} + \varepsilon_n]. \quad (3.8)$$

Thus, we need a lower bound for $|F'_\infty(\tilde{\gamma})|$ and an upper bound for $|F(x) - F_\infty(x)|$. For the lower bound, begin with

$$F'_\infty(\tilde{\gamma}) = \frac{R'_\infty(\tilde{\gamma})L_\infty(\tilde{\gamma}) - L'_\infty(\tilde{\gamma})R_\infty(\tilde{\gamma})}{(L_\infty(\tilde{\gamma}))^2}.$$

Note that $L_\infty(\tilde{\gamma})$ and $R_\infty(\tilde{\gamma})$ are of order 1, and $R'_\infty(\tilde{\gamma}) = O(1)$ using the fact that $\alpha, B, \tilde{\gamma}$ are all of order 1.

We now show $|L'_\infty(\tilde{\gamma})|$ is of order at least $n^{1/3}(\log n)^{-1/2}$, which implies $|F'_\infty(\tilde{\gamma})|$ is as well.

Since we are interested in $L'_\infty(x)$ at $\tilde{\gamma}$, where $\tilde{\gamma} - d_+ = \Theta(n^{-2/3} \log n)$ by Lemma 3.1, we consider $L_\infty(d_+ + s)$ as a function of s and its derivative, and later set s to take value of order $n^{-2/3} \log n$. We have

$$\begin{aligned} L_\infty(d_+ + s) &= C \int_{d_-}^{d_+} \frac{\sqrt{(d_+ - y)(y - d_-)}}{(d_+ + s - y)y} dy \\ &= \int_0^{d_+ - d_-} \frac{C \sqrt{z(d_+ - d_- - z)}}{(z + s)(d_+ - z)} dz, \end{aligned}$$

where $C = \frac{2}{\pi}(\sqrt{d_+} - \sqrt{d_-})^{-2}$ and $z = d_+ - y$. We then write

$$\begin{aligned} L'_\infty(d_+ + s) &= \frac{d}{ds} \int_0^{\frac{d_+ - d_-}{2}} \frac{C \sqrt{z(d_+ - d_- - z)}}{(z + s)(d_+ - z)} dz \\ &\quad + \frac{d}{ds} \int_{\frac{d_+ - d_-}{2}}^{d_+ - d_-} \frac{C \sqrt{z(d_+ - d_- - z)}}{(z + s)(d_+ - z)} dz. \end{aligned} \quad (3.9)$$

First, we consider the derivative of a simplified version of the first integral:

$$\begin{aligned} \frac{d}{ds} \int_0^{\frac{d_+ - d_-}{2}} \frac{\sqrt{z}}{z + s} dz &= -s^{-1/2} \arctan \sqrt{\frac{(d_+ - d_-)/2}{s}} \\ &\quad + \sqrt{s} \frac{\sqrt{(d_+ - d_-)/2}}{s^{3/2}} \frac{1}{1 + \frac{(d_+ - d_-)/2}{s}} \\ &= -s^{-1/2} + O(1). \end{aligned} \quad (3.10)$$

Now that we have the derivative of this simplified integral, recall that the actual integrand is $\frac{C \sqrt{z(d_+ - d_- - z)}}{(z + s)(d_+ - z)}$ and make the following observations:

- For $z \in [0, \frac{d_+ - d_-}{2}]$, there exist positive constants C_1, C_2 such that $C_1 < \frac{C \sqrt{d_+ - d_- - z}}{d_+ - z} < C_2$.

• For any $z > 0$, the quantity $\frac{\sqrt{z}}{z+s}$ is a decreasing function of s when $s > 0$. From these two facts and the above computation, we conclude that, for small s ,

$$-C_2 s^{-1/2} \leq \frac{d}{ds} \int_0^{d_+ - d_-} \frac{C \sqrt{z(d_+ - d_- - z)}}{(z+s)(d_+ - z)} dz \leq -C_1 s^{-1/2}. \quad (3.11)$$

Finally, the second bullet point implies the second integral on the right side of (3.9) must be negative. Thus, $L'_\infty(d_+ + s) < -C_1 s^{-1/2}$, which implies $|L'_\infty(\tilde{\gamma})|$ is of order at least $n^{1/3}(\log n)^{-1/2}$. We obtain the lower bound

$$|F'_\infty(\tilde{\gamma})| = \Omega(n^{1/3}(\log n)^{-1/2}). \quad (3.12)$$

We now show an upper bound of $|F(x) - F_\infty(x)|$ for $x \in [\tilde{\gamma} - \varepsilon_n, \tilde{\gamma} + \varepsilon_n]$. For such x ,

$$F(x) - F_\infty(x) = \frac{(R(x) - R_\infty(x))L_\infty(x) + (L_\infty(x) - L(x))R_\infty(x)}{L(x)L_\infty(x)}$$

satisfies that the denominator, $L_\infty(x)$, and $R_\infty(x)$ all have order 1. Thus, it remains to bound the terms $R(x) - R_\infty(x)$ and $L_\infty(x) - L(x)$. As $\alpha_n - \alpha = O(n^{-1-\delta})$ and $B_n - B = O(n^{-1-\delta})$, we have

$$R(x) - R_\infty(x) = \frac{\sqrt{\alpha_n^2 + xB_n^2} - \alpha_n}{x} - \frac{\sqrt{\alpha^2 + xB^2} - \alpha}{x} = O(n^{-1-\delta}).$$

Lastly, Lemma 3.2 yields that

$$L(x) - L_\infty(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{x - \mu_i} - \int \frac{p_{\text{MP}}(y)dy}{x - y} = O(n^{-1/3}(\log \log n)(\log n)^{-1}).$$

Thus, we have shown that for $x \in [\tilde{\gamma} - \frac{(\log \log n)^2}{n^{2/3} \log n}, \tilde{\gamma} + \frac{(\log \log n)^2}{n^{2/3} \log n}]$,

$$\begin{aligned} |F(x) - F_\infty(x)| &= O\left(\frac{\log \log n}{n^{1/3} \log n}\right), \\ |F'_\infty(\tilde{\gamma})| &= \Omega\left(n^{1/3}(\log n)^{-1/2}\right). \end{aligned} \quad (3.13)$$

This verifies the inequality (3.8), and the lemma follows. \square

We now introduce a deterministic approximation G_∞ of the function G , given by

$$G_\infty(z_1, z_2) = B(z_1 + z_2) - \alpha \log z_1 - \frac{1}{2} \int \log(4z_1 z_2 - x) p_{\text{MP}}(x) dx. \quad (3.14)$$

We observe that $(\tilde{\gamma}_1, \tilde{\gamma}_2)$ is the unique critical point of G_∞ satisfying $4\tilde{\gamma}_1 \tilde{\gamma}_2 \in (d_+, \infty)$. This follows from the similar reasoning to what we used for (γ_1, γ_2) . We obtain the following asymptotic expressions for the functions G , G_∞ and their partial derivatives.

Lemma 3.4. *Let (z_1, z_2) satisfy $\text{Re}(4z_1 z_2) \geq d_+$ and $|4z_1 z_2 - d_+| \geq cn^{-2/3} \log n$ for some fixed $c > 0$. Then, on the event \mathcal{E}_ε , the following hold and are uniform in any compact region satisfying the constraints on (z_1, z_2) :*

(i) For every multi-index $k = (k_1, k_2)$ (with $|k| := k_1 + k_2 \geq 1$),

$$\partial^k G(z_1, z_2) - \partial^k G_\infty(z_1, z_2) = O\left(n^{\frac{2}{3}|k|-1} \frac{\log \log n}{(\log n)^{|k|}}\right). \quad (3.15)$$

(ii) For every multi-index k with $|k| \geq 1$,

$$\begin{aligned} \partial^k G_\infty(z_1, z_2) &= O(n^{\frac{2}{3}|k|-1} (\log n)^{-|k|+\frac{3}{2}}) \\ \partial^k G(z_1, z_2) &= O(n^{\frac{2}{3}|k|-1} (\log n)^{-|k|+\frac{3}{2}}). \end{aligned} \quad (3.16)$$

Proof. We recall

$$\begin{aligned} G(z_1, z_2) &= B_n(z_1 + z_2) - \alpha_n \log z_1 - \frac{1}{2n} \sum_{j=1}^n \log(4z_1 z_2 - \mu_j), \\ G_\infty(z_1, z_2) &= B(z_1 + z_2) - \alpha \log z_1 - \frac{1}{2} \int \log(4z_1 z_2 - x) p_{\text{MP}}(x) dx. \end{aligned}$$

Observe that over any fixed compact region of \mathbb{C}^2 , for every $|k| \geq 1$,

- $\partial^k G_\infty(z_1, z_2) = O\left(\int (4z_1 z_2 - x)^{-k} p_{\text{MP}}(x) dx\right)$, and
- the differences in the partials of G and G_∞ satisfy

$$\partial^k G(z_1, z_2) - \partial^k G_\infty(z_1, z_2) = O\left(\frac{1}{n} \sum_{i=1}^n \frac{1}{(4z_1 z_2 - \mu_i)^{|k|}} - \int \frac{p_{\text{MP}}(x) dx}{(4z_1 z_2 - x)^{|k|}}\right). \quad (3.17)$$

Applying Lemma 3.2 to (3.17) gives us part (i) of the lemma. For part (ii), we first obtain the bound for $\partial^k G_\infty$ by noting that

$$\begin{aligned} &\left| \int (4z_1 z_2 - x)^{-|k|} p_{\text{MP}}(x) dx \right| \\ &\leq \int \frac{1}{\max\{|4z_1 z_2 - x|, d_+ - x\}^{|k|}} p_{\text{MP}}(x) dx \\ &= O\left(\int_{n^{-2/3} \log n}^{\infty} \frac{\sqrt{y - n^{-2/3} \log n}}{y^{|k|}} dy\right) \\ &= O\left(\int_{n^{-2/3} \log n}^{\infty} y^{-|k|+\frac{1}{2}} dy\right) = O\left((n^{-2/3} \log n)^{-|k|+3/2}\right). \end{aligned} \quad (3.18)$$

Then, the bound for $\partial^k G$ as in (ii) follows by part (i) of the lemma and the bound obtained for $\partial^k G_\infty$. \square

We prove some further properties of G and G_∞ in the following lemma.

Lemma 3.5. *For the critical points (γ_1, γ_2) and $(\tilde{\gamma}_1, \tilde{\gamma}_2)$ of G and G_∞ , respectively, the following hold on event \mathcal{E}_ε .*

(i) We have

$$\begin{aligned} |\gamma_1 - \tilde{\gamma}_1| &= O(n^{-2/3} (\log \log n)^2 (\log n)^{-1/2}), \\ |\gamma_2 - \tilde{\gamma}_2| &= O(n^{-2/3} (\log \log n)^2 (\log n)^{-1/2}). \end{aligned}$$

(ii) *There is a positive constant c , independent of n , such that*

$$4\gamma_1\gamma_2 - \mu_1 > cn^{-2/3} \log n \quad 4\tilde{\gamma}_1\tilde{\gamma}_2 - d_+ > cn^{-2/3} \log n.$$

(iii) *We have*

$$G(\gamma_1, \gamma_2) = G(\tilde{\gamma}_1, \tilde{\gamma}_2) + O(n^{-1}(\log n)^{-3/2}(\log \log n)^4)$$

and for and multi-index $k = (k_1, k_2)$ satisfying $|k| > 0$,

$$\partial^k G(\gamma_1, \gamma_2) = \partial^k G(\tilde{\gamma}_1, \tilde{\gamma}_2) + O\left(n^{\frac{2}{3}|k|-1}(\log n)^{-|k|}(\log \log n)^2\right).$$

Proof. Part (i) follows from the equations for $\gamma_1, \gamma_2, \tilde{\gamma}_1, \tilde{\gamma}_2$ along with the bound on $|\gamma - \tilde{\gamma}|$.

Part (ii) follows from part (i) along with the computation of $\tilde{\gamma} - d_+$ and the fact that $|d_+ - \mu_1| = O(n^{-2/3})$.

For Part (iii), using the bounds from Lemma 3.4(ii) and Lemma 3.5(i), we get the Taylor expansion

$$\begin{aligned} G(\tilde{\gamma}_1, \tilde{\gamma}_2) &= G(\gamma_1, \gamma_2) + \partial_1 G(\gamma_1, \gamma_2)(\tilde{\gamma}_1 - \gamma_1) + \partial_2 G(\gamma_1, \gamma_2)(\tilde{\gamma}_2 - \gamma_2) \\ &\quad + O\left(\frac{n^{1/3}}{(\log n)^{1/2}} \cdot \left(\frac{(\log \log n)^2}{n^{2/3}(\log n)^{1/2}}\right)^2\right) \\ &= G(\tilde{\gamma}_1, \tilde{\gamma}_2) + O(n^{-1}(\log n)^{-3/2}(\log \log n)^4). \end{aligned}$$

Similarly, for the partials, we get

$$\begin{aligned} \partial^k G(\tilde{\gamma}_1, \tilde{\gamma}_2) &= \partial^k G(\gamma_1, \gamma_2) + O\left(n^{\frac{2}{3}(|k|+1)-1}(\log n)^{-(|k|+1)+\frac{3}{2}} \cdot \frac{(\log \log n)^2}{n^{2/3}(\log n)^{1/2}}\right) \\ &= \partial^k G(\gamma_1, \gamma_2) + O\left(n^{\frac{2}{3}|k|-1}(\log n)^{-|k|}(\log \log n)^2\right). \end{aligned}$$

□

3.2. Steepest Descent Analysis

We now perform steepest analysis to compute the contour integral in the high-temperature case. The method relies on the observation that the dominant contribution to the integral comes from within a small radius around the critical point of G . In this case, the radius is $r = n^{-2/3}(\log n)^{\frac{1}{4}+\varepsilon}$ for some $\varepsilon > 0$.

The intuition behind this choice of truncation radius is as follows: Consider a Taylor expansion of G_∞ where $z_1 = \tilde{\gamma}_1 + it_1/r_n$ and $z_2 = \tilde{\gamma}_2 + it_2/r_n$ with r_n to be determined. Let m denote a multi-index for the derivative, and let $|m|$ denote the length the multi-index. We want to choose r_n such that

$$\partial^{(m)} G_\infty(\tilde{\gamma}_1, \tilde{\gamma}_2) \cdot r_n^{|m|} = \begin{cases} \Theta\left(\frac{1}{n}\right) & |m| = 2 \\ o\left(\frac{1}{n}\right) & |m| \geq 3. \end{cases} \quad (3.19)$$

Using the previous lemmas, this is satisfied exactly when $r = \Theta(n^{-2/3}(\log n)^{1/4})$.

Lemma 3.6. *Let $\gamma_1 = \gamma_1(n)$ and $\gamma_2 = \gamma_2(n)$ be such that (γ_1, γ_2) is the critical point of $G(z_1, z_2)$ satisfying $\gamma = 4\gamma_1\gamma_2 > \mu_1(n)$. Then, for any $0 < \varepsilon < 1/4$*

and any $\Omega \subset \{(y_1, y_2) \in \mathbb{R}^2 : y_1^2 + y_2^2 \geq n^{-4/3}(\log n)^{1/2+2\varepsilon}\}$, on the event \mathcal{E}_ε , there exists some $C > 0$ such that

$$\int_{\Omega} \exp[n \operatorname{Re}(G(\gamma_1 + iy_1, \gamma_2 + iy_2) - G(\gamma_1, \gamma_2))] dy_2 dy_1 = O(e^{-C(\log n)^\varepsilon}).$$

Proof. Since $\gamma - \mu_n$ is bounded in n , Lemma 3.9 of [13] implies that with high probability, the portion of the above integral over $\Omega \cap \{(y_1, y_2) \in \mathbb{R}^2 : y_1^2 + y_2^2 \geq n^{-1+2\varepsilon}\}$ is $O(e^{-n^\varepsilon})$. Thus, it remains to consider the subset of Ω where $y_1^2 + y_2^2$ is between $n^{-4/3}(\log n)^{1/2+2\varepsilon}$ and $n^{-1+2\varepsilon}$. We denote this subset by $\tilde{\Omega}$.

The proof of Lemma 3.9 of [13] also shows that, for some constant $c_0 > 0$ and for any integer $K \geq 1$,

$$\begin{aligned} & \operatorname{Re}(G(\gamma_1 + iy_1, \gamma_2 + iy_2) - G(\gamma_1, \gamma_2)) \\ & \leq -\frac{1}{4n} \sum_{j=K}^n \log \left(1 + \frac{c_0}{(\gamma - \mu_j)^2} (y_1^2 + y_2^2) \right), \end{aligned} \quad (3.20)$$

for all $y_1, y_2 \in \mathbb{R}$. By Lemma 2.2, for every $\varepsilon > 0$, there exists $c, K > 0$ such that, with probability at least $1 - \varepsilon$,

$$\gamma - d_+ \leq cn^{-2/3} \log n \quad \text{and} \quad d_+ - \mu_j \leq \begin{cases} ci^{2/3}n^{-2/3}, & K \leq j \leq n^{2/5}, \\ c, & j > n^{2/5}. \end{cases}$$

Thus, with probability at least $1 - \varepsilon$,

$$\gamma - \mu_j \leq \begin{cases} cn^{-2/3} \log n, & K \leq j \leq (\log n)^{3/2}, \\ cj^{2/3}n^{-2/3}, & (\log n)^{3/2} \leq j \leq n^{2/5}, \\ c, & j > n^{2/5}. \end{cases}$$

Write $r^2 = y_1^2 + y_2^2$ using polar coordinates, then for $r \in [n^{-2/3} \log^{1/4+\varepsilon} n, n^{-1/2+\varepsilon}]$ and the above choice of K , the right-hand side of (3.20) has upper bound

$$-\frac{1}{4n} \left[(\log n)^{3/2} \log \left(1 + \frac{c'n^{4/3}}{\log^2 n} r^2 \right) + \sum_{j=(\log n)^{3/2}}^{n^{2/5}} \log \left(1 + \frac{c'r^2}{(j/n)^{4/3}} \right) + \frac{n}{2} \log(1 + c'r^2) \right]. \quad (3.21)$$

We then use $r \geq n^{-2/3}(\log n)^{1/4+\varepsilon}$ for the first and last terms inside the brackets, and the fact $\log(1+x) \geq x/2$ for small x to obtain a new bound

$$-\frac{c'}{4n} \left[(\log n)^{2\varepsilon} + (\log n)^{-3/2+2\varepsilon} \sum_{j=\log^{3/2} n}^{n^{2/5}} j^{-4/3} + \frac{n}{4} r^2 \right] \leq -\frac{c'r^2}{16} - \frac{c'(\log n)^{2\varepsilon}}{8n}, \quad (3.22)$$

noting that the sum over j is $O((\log n)^{-1/2})$. Therefore, the integral over $\tilde{\Omega}$ is bounded by

$$e^{-\frac{c'}{8}(\log n)^{2\varepsilon}} \int_{n^{-2/3} \log^{1/4+\varepsilon} n}^{n^{-1/2+\varepsilon}} e^{-\frac{c'}{16}r^2} r dr = O(e^{-C(\log n)^{2\varepsilon}}), \quad (3.23)$$

for some $C > 0$. This completes our proof. \square

Lemma 3.7. *If $\beta = \beta_c + bn^{-1/3}\sqrt{\log n}$ for fixed $b < 0$, then the integral Q_n in (3.1) satisfies*

$$Q_n = e^{nG(\gamma_1, \gamma_2)} \frac{\pi}{n\sqrt{D(\gamma_1, \gamma_2)}} \left(1 + O((\log n)^{-\frac{3}{2}+6\varepsilon}) \right),$$

where $\varepsilon > 0$ is arbitrarily small and $D(\gamma_1, \gamma_2)$ is the discriminant

$$D(\gamma_1, \gamma_2) := \partial_1^2 G(\gamma_1, \gamma_2) \cdot \partial_2^2 G(\gamma_1, \gamma_2) - (\partial_1 \partial_2 G(\gamma_1, \gamma_2))^2. \quad (3.24)$$

Proof. We make the change of variables

$$z_1 = \gamma_1 + ir_n t_1, \quad z_2 = \gamma_2 + ir_n t_2, \quad (3.25)$$

where the scaling $r_n := n^{-2/3}(\log n)^{1/4}$ is chosen such that the quadratic term in the Taylor expansion of G near (γ_1, γ_2) will be of order 1. With this change of variable, we have

$$Q_n = r_n^2 e^{nG(\gamma_1, \gamma_2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \left(n \left(G(\gamma_1 + ir_n t_1, \gamma_2 + ir_n t_2) - G(\gamma_1, \gamma_2) \right) \right) dt_2 dt_1. \quad (3.26)$$

Fix $0 < \varepsilon < 1/4$. We have shown in Lemma 3.6 that this integral outside a region of radius $(\log n)^\varepsilon$ around the critical point is $O(e^{-c(\log n)^\varepsilon})$ for some constant $c > 0$. We now consider the region where $|t_1|, |t_2| \leq (\log n)^\varepsilon$. In this region,

$$\begin{aligned} & G(\gamma_1 + ir_n t_1, \gamma_2 + ir_n t_2) - G(\gamma_1, \gamma_2) \\ &= -\frac{1}{2} r_n^2 \left(\partial_1^2 G(\gamma_1, \gamma_2) t_1^2 + 2\partial_1 \partial_2 G(\gamma_1, \gamma_2) t_1 t_2 + \partial_2^2 G(\gamma_1, \gamma_2) t_2^2 \right) \\ &\quad - \frac{i}{6} r_n^3 \left(\partial_1^3 G(\gamma_1, \gamma_2) t_1^3 + 3\partial_1^2 \partial_2 G(\gamma_1, \gamma_2) t_1^2 t_2 + 3\partial_1 \partial_2^2 G(\gamma_1, \gamma_2) t_1 t_2^2 \right. \\ &\quad \left. + \partial_2^3 G(\gamma_1, \gamma_2) t_2^3 \right) + O(\text{Taylor remainder}) \\ &=: -r_n^2 X_2(t_1, t_2) - ir_n^3 X_3(t_1, t_2) + O(n^{-1}(\log n)^{-\frac{3}{2}+4\varepsilon}). \end{aligned} \quad (3.27)$$

Thus, the integral on the central region becomes

$$\begin{aligned} & \int_{-(\log n)^\varepsilon}^{(\log n)^\varepsilon} \int_{-(\log n)^\varepsilon}^{(\log n)^\varepsilon} \exp \left(n \left(G(\gamma_1 + ir_n t_1, \gamma_2 + ir_n t_2) - G(\gamma_1, \gamma_2) \right) \right) dt_2 dt_1 \\ &= \int \int e^{-nr_n^2 X_2(t_1, t_2)} dt_2 dt_1 - i \int \int nr_n^3 X_3(t_1, t_2) e^{-nr_n^2 X_2(t_1, t_2)} dt_2 dt_1 \\ &\quad + O \left((\log n)^{-\frac{3}{2}+6\varepsilon} \right), \end{aligned} \quad (3.28)$$

where the second integral vanishes due to the fact that

$$X_3(-t_1, -t_2) e^{-nr_n^2 X_2(-t_1, -t_2)} = -X_3(t_1, t_2) e^{-nr_n^2 X_2(t_1, t_2)}.$$

It remains to compute $\int_{-(\log n)^\varepsilon}^{(\log n)^\varepsilon} \int_{-(\log n)^\varepsilon}^{(\log n)^\varepsilon} e^{-nr_n^2 X_2(t_1, t_2)} dt_2 dt_1$, which we replace by the integral over \mathbb{R}^2 , incurring an error on the order of

$$\int_{(\log n)^\varepsilon}^{\infty} e^{-x^2} dx < e^{-(\log n)^{2\varepsilon}} \ll (\log n)^{-3/2}. \quad (3.29)$$

Finally, applying Gaussian integration, we obtain the lemma. □

We observe from the lemma above that the integral Q_n depends on $G(\gamma_1, \gamma_2)$ and $D(\gamma_1, \gamma_2)$, which we compute in the following lemma.

Lemma 3.8. *If $\beta = \beta_c + bn^{-1/3}\sqrt{\log n}$ for some fixed $b < 0$, then*

$$\begin{aligned} G(\gamma_1, \gamma_2) &= A(\tilde{\gamma}, B) - \frac{1}{2n} \sum_{i=1}^n \log(\tilde{\gamma} - \mu_i) + O(n^{-1}) \\ D(\gamma_1, \gamma_2) &= \frac{\beta_c}{\lambda^2 b} n^{1/3} (\log n)^{-1/2} \left(1 + O\left((\log \log n)^2 (\log n)^{-3/2} \right) \right) \end{aligned}$$

where

$$A(x, B) := \sqrt{\alpha^2 + xB^2} - \alpha \log \left(\frac{\alpha + \sqrt{\alpha^2 + xB^2}}{2B} \right). \quad (3.30)$$

Proof. The computation of $G(\gamma_1, \gamma_2)$ relies upon $G_\infty(\tilde{\gamma}_1, \tilde{\gamma}_2)$, which we write as

$$G_\infty(\tilde{\gamma}_1, \tilde{\gamma}_2) = A(\tilde{\gamma}, B) - \frac{1}{2} H_{\text{MP}}(\tilde{\gamma}), \quad H_{\text{MP}}(z) := \int_{\mathbb{R}} \log(z - x) p_{\text{MP}}(x) dx. \quad (3.31)$$

Then, by Lemma 3.5(iii),

$$\begin{aligned} G(\gamma_1, \gamma_2) &= G_\infty(\tilde{\gamma}_1, \tilde{\gamma}_2) + [G(\tilde{\gamma}_1, \tilde{\gamma}_2) - G_\infty(\tilde{\gamma}_1, \tilde{\gamma}_2)] + O\left(n^{-1} \frac{(\log \log n)^4}{(\log n)^{3/2}} \right) \\ &= G_\infty(\tilde{\gamma}_1, \tilde{\gamma}_2) - \frac{1}{2n} \left[\sum_{i=1}^n \log(\tilde{\gamma} - \mu_i) - n H_{\text{MP}}(\tilde{\gamma}) \right] + O(n^{-1}) \\ &= A(\tilde{\gamma}, B) - \frac{1}{2n} \sum_{i=1}^n \log(\tilde{\gamma} - \mu_i) + O(n^{-1}). \end{aligned} \quad (3.32)$$

The same lemma and Lemma 3.4(ii) together yield

$$D(\gamma_1, \gamma_2) = D_\infty(\tilde{\gamma}_1, \tilde{\gamma}_2) + O\left(n^{1/3} \frac{(\log \log n)^2}{(\log n)^2} \right).$$

Recall from (3.2) that $\tilde{\gamma} = \frac{1+\beta^2+\beta_c^{-4}\beta^4}{(1+\lambda)^{-1}\beta^2}$, and $\beta_c = \lambda^{-\frac{1}{4}}(1+\lambda)^{1/2}$. We arrive at

$$\begin{aligned} D_\infty(\tilde{\gamma}_1, \tilde{\gamma}_2) &:= \partial_1^2 G_\infty(\tilde{\gamma}_1, \tilde{\gamma}_2) \cdot \partial_2^2 G_\infty(\tilde{\gamma}_1, \tilde{\gamma}_2) - (\partial_1 \partial_2 G_\infty(\tilde{\gamma}_1, \tilde{\gamma}_2))^2 \\ &= \frac{4\beta^4}{\lambda^2(\beta_c^4 - \beta^4)}. \end{aligned} \quad (3.33)$$

Apply this to the expression $D(\gamma_1, \gamma_2)$ and perform Taylor expansion around β_c ; we obtain the lemma. \square

3.3. High-Temperature Free Energy

Finally, using the contour integral computations from the previous section, we obtain the following lemma for the limiting fluctuations of the free energy on the high-temperature side of the critical temperature window.

Lemma 3.9. *Suppose $\beta = \beta_c + bn^{-1/3}\sqrt{\log n}$ for some fixed $b < 0$. We define $F(\beta) = \frac{\beta^2}{2\beta_c^4}$. Then the free energy satisfies*

$$\frac{m+n}{\sqrt{\frac{1}{6}\log n}} \left(F_{n,m}(\beta) - F(\beta) + \frac{1}{12} \frac{\log n}{n+m} \right) \rightarrow \mathcal{N}(0, 1). \quad (3.34)$$

Proof. We will show that

$$F_{n,m}(\beta) - \frac{\beta^2}{2\beta_c^4} + \frac{1}{12} \frac{\log n}{n+m} - \frac{\sqrt{\frac{1}{6}\log n}}{m+n} T_{0n} = O\left(\frac{\log \log n}{n}\right), \quad (3.35)$$

where

$$-T_{0n} := \frac{\sum_{i=1}^n \log(\tilde{\gamma} - \mu_i) - C_\lambda n - \frac{1}{\sqrt{\lambda(1+\sqrt{\lambda})}} n(\tilde{\gamma} - d_+) + \frac{2}{3\lambda^{3/4}(1+\sqrt{\lambda})^2} n(\tilde{\gamma} - d_+)^{3/2} + \frac{1}{6} \log n}{\sqrt{\frac{2}{3}\log n}} \quad (3.36)$$

with $C_\lambda := (1-\lambda^{-1})\log(1+\lambda^{1/2}) + \log(\lambda^{1/2}) + \lambda^{-1/2}$ and, by [24], T_{0n} converges in distribution to a standard normal. We now compute the left-hand side of (3.35) in terms of the parameters β and λ . From (2.6), we start by computing

$$\begin{aligned} \frac{1}{n+m} \log Q(n, \alpha_n, B_n) &= \frac{n}{n+m} G(\gamma_1, \gamma_2) + \frac{1}{2(n+m)} \log \left(\frac{\pi^2}{D(\gamma_1, \gamma_2)} \right) \\ &\quad - \frac{\log n}{n+m} + o(n^{-1}), \end{aligned} \quad (3.37)$$

using Lemma 3.7. By Lemma 3.8, the second term satisfies

$$\frac{1}{2(n+m)} \log \left(\frac{\pi^2}{D(\gamma_1, \gamma_2)} \right) = -\frac{1}{6} \frac{\log n}{n+m} + O\left(\frac{\log \log n}{n}\right). \quad (3.38)$$

Thus, using the computation of $G(\gamma_1, \gamma_2)$ from (3.32), (3.37) simplifies to

$$\begin{aligned} \frac{1}{n+m} \log Q(n, \alpha_n, B_n) &= \frac{n}{n+m} A(\tilde{\gamma}, B) - \frac{1}{2(n+m)} \sum_{i=1}^n \log(\tilde{\gamma} - \mu_i) \\ &\quad - \frac{7}{6} \frac{\log n}{n+m} + O\left(\frac{\log \log n}{n}\right). \end{aligned} \quad (3.39)$$

Recall that $\alpha = \frac{1}{2}(\lambda^{-1} - 1)$ and $B = \frac{1}{\sqrt{\lambda(1+\lambda)}}\beta$ for the bipartite SSK model, and $\tilde{\gamma}$ is given in (3.2). This implies $\sqrt{\alpha^2 + \tilde{\gamma}B^2} = \frac{\lambda+1}{2\lambda} + \frac{\beta^2}{1+\lambda}$, and

$$\frac{n}{n+m} A(\tilde{\gamma}, B) = \frac{1}{2} + \frac{\lambda\beta^2}{(1+\lambda)^2} + \frac{1-\lambda}{2(1+\lambda)} \log \left(\frac{2\beta\sqrt{\lambda(1+\lambda)}}{1+\lambda+\beta^2\lambda} \right) + O(n^{-1}). \quad (3.40)$$

Combining (2.6), (3.39), and (3.40), we have

$$\begin{aligned}
 F_{n,m}(\beta) = & -\frac{1}{2(n+m)} \sum_{i=1}^n \log(\tilde{\gamma} - \mu_i) + \frac{\lambda\beta^2}{(1+\lambda)^2} - \frac{1-\lambda}{2(1+\lambda)} \log(1+\lambda + \beta^2\lambda) \\
 & - \frac{\lambda}{1+\lambda} \log\beta + \frac{1}{2(\lambda+1)} \log(1+\lambda) - \frac{1}{6} \frac{\log n}{n+m} + O\left(\frac{\log\log n}{n}\right).
 \end{aligned} \tag{3.41}$$

In order to prove Eq. (3.35), we need express each β -dependent term as a Taylor expansion around β_c . More specifically, we define

$$\Delta_\beta := \beta_c - \beta = O(n^{-1/3} \sqrt{\log n}). \tag{3.42}$$

Using this and the fact that $\beta_c = \frac{\sqrt{1+\lambda}}{\lambda^{1/4}}$, we get

$$\begin{aligned}
 \beta^2 &= \frac{1+\lambda}{\sqrt{\lambda}} - 2\beta_c\Delta_\beta + \Delta_\beta^2 \\
 \log\beta &= \frac{1}{2} \log(1+\lambda) - \frac{1}{4} \log\lambda - \frac{1}{\beta_c} \Delta_\beta - \frac{1}{2\beta_c^2} \Delta_\beta^2 \\
 &\quad - \frac{1}{3\beta_c^3} \Delta_\beta^3 + O(\Delta_\beta^4) \\
 \log(1+\lambda + \beta^2\lambda) &= \log((1+\lambda)(1+\sqrt{\lambda})) - \frac{2\beta_c\lambda}{(1+\lambda)(1+\sqrt{\lambda})} \Delta_\beta \\
 &\quad + \frac{\lambda(1+\lambda + \beta_c^2\lambda - 2\beta_c^2)}{(1+\lambda)^2(1+\sqrt{\lambda})^2} \Delta_\beta^2 \\
 &\quad + \frac{2\beta_c\lambda^2(1+\lambda - \frac{1}{3}\beta_c^2\lambda)}{(1+\lambda)^3(1+\sqrt{\lambda})^3} \Delta_\beta^3 + O(\Delta_\beta^4)
 \end{aligned} \tag{3.43}$$

Furthermore, using Eq. (3.3) we have

$$\tilde{\gamma} - d_+ = \frac{4(1+\lambda)}{\beta_c^4} \Delta_\beta^2 + \frac{4(1+\lambda)}{\beta_c^5} \Delta_\beta^3 + O(\Delta_\beta^4). \tag{3.44}$$

Plugging these asymptotics into Eqs. (3.36) and (3.41), we verify (3.35), and the lemma follows. \square

4. Low Temperature

We now determine the asymptotics of the random double integral $Q_n = -\int_{\gamma_1-i\infty}^{\gamma_1+i\infty} \int_{\gamma_2-i\infty}^{\gamma_2+i\infty} e^{nG(z_1, z_2)} dz_2 dz_1$ when $\beta = \beta_c + bn^{-\frac{1}{3}} \sqrt{\log n}$ for fixed $b \geq 0$.

Recall that in the regime $\beta < \beta_c$, both for fixed β as in [13] and for β in Sect. 3, the critical point (γ_1, γ_2) of the function G is approximated by $(\tilde{\gamma}_1, \tilde{\gamma}_2)$, the critical point satisfying $4\tilde{\gamma}_1\tilde{\gamma}_2 > d_+$ of a deterministic approximation G_∞ of G . In the case $\beta > \beta_c$, a critical point of G_∞ satisfying this inequality does not exist, and we cannot approximate the product $\gamma = 4\gamma_1\gamma_2$ by a deterministic

number. In fact, the product γ gets close to the branch point μ_1 from above, which requires more delicate analysis.

We address this issue by focusing on G near the point $(\mu_1^{(1)}, \mu_1^{(2)})$, given by

$$\mu_1^{(1)} = \frac{\alpha_n + \sqrt{\alpha_n^2 + \mu_1 B_n^2}}{2B_n}, \quad \mu_1^{(2)} = \frac{-\alpha_n + \sqrt{\alpha_n^2 + \mu_1 B_n^2}}{2B_n}, \quad (4.1)$$

instead of (γ_1, γ_2) . We see that $4\mu_1^{(1)}\mu_1^{(2)} = \mu_1$, and $G(z_1, z_2)$ at $(\mu_1^{(1)}, \mu_1^{(2)})$ is undefined due to the term $\frac{1}{n} \log(4z_1 z_2 - \mu_1)$. However, the non-singular part given below will play an important role.

$$\widehat{G} := B_n(\mu_1^{(1)} + \mu_1^{(2)}) - \alpha_n \log \mu_1^{(1)} - \frac{1}{2n} \sum_{j=2}^n \log(\mu_1 - \mu_j) \quad (4.2)$$

In our computation of \widehat{G} as well as the contour integral, we need to work with sums of the form $\frac{1}{n} \sum_{i=2}^n \frac{1}{(\mu_1 - \mu_i)^l}$ for $l \geq 1$. More specifically, we need the following lemma.

Lemma 4.1. *For LOE eigenvalues, on the event \mathcal{E}_ε , we have*

$$\begin{aligned} \frac{1}{n} \sum_{i=2}^n \frac{1}{\mu_1 - \mu_i} - \frac{1}{\lambda^{1/2}(1 + \lambda^{1/2})} &= O(n^{-1/3}) \quad \text{and} \\ \frac{1}{n} \sum_{i=2}^n \frac{1}{(\mu_1 - \mu_i)^l} &= O(n^{\frac{2}{3}l-1}), \quad \text{for } l \geq 2. \end{aligned}$$

Proof. It suffices to prove the following statements:

(i) For any $l \geq 1$, on the event \mathcal{E}_ε ,

$$\left| \frac{1}{n} \sum_{i=k}^n \frac{1}{(\mu_1 - \mu_i)^l} - \int_{d_-}^{g_k} \frac{p_{\text{MP}}(x)}{(d_+ - x)^l} dx \right| = O(n^{\frac{2}{3}l-1}). \quad (4.3)$$

(ii) For any $l \geq 1$ and any fixed k , on the event \mathcal{E}_ε ,

$$\frac{1}{n} \sum_{i=2}^k \frac{1}{(\mu_1 - \mu_i)^l} = O(n^{\frac{2}{3}l-1}). \quad (4.4)$$

(iii) For the $l = 1$ case,

$$\int_{g_k}^{d_+} \frac{p_{\text{MP}}(x)}{d_+ - x} dx = O(n^{-\frac{1}{3}}). \quad (4.5)$$

(iii) For the $l \geq 2$ case,

$$\int_{d_-}^{g_k} \frac{p_{\text{MP}}(x)}{(d_+ - x)^l} dx = O(n^{\frac{2}{3}l-1}) \quad (4.6)$$

Verifying (ii) is straightforward after imposing the assumption $\mu_1 - \mu_i > cn^{-2/3}$ for some $c > 0$, which follows from event $\mathcal{F}_{r,R}^{(4)}$. Statements (iii) and (iv) follow from the definitions of p_{MP} and g_k .

We now turn to (i). It follows from Lemma 2.8 that, on the event $\mathcal{F}_K^{(2)}$,

$$\frac{1}{n} \sum_{i=K}^n \frac{1}{(d_+ - \mu_i)^l} - \int_{d_-}^{g_K} \frac{p_{\text{MP}}(x)}{(d_+ - x)^l} dx = O(n^{\frac{2}{3}l-1}).$$

Thus, it remains only to show that

$$\frac{1}{n} \sum_{i=K}^n \left(\frac{1}{(\mu_1 - \mu_i)^l} - \frac{1}{(d_+ - \mu_i)^l} \right) = O(n^{\frac{2}{3}l-1}). \quad (4.7)$$

This bound holds on the event $\mathcal{F}_K^{(2)} \cap \mathcal{F}_{s,t}^{(3)}$, which can be seen by observing that

$$\begin{aligned} \left| \frac{1}{(\mu_1 - \mu_i)^l} - \frac{1}{(d_+ - \mu_i)^l} \right| &= \left| \frac{(d_+ - \mu_1) \sum_{j=0}^{l-1} (d_+ - \mu_i)^j (\mu_1 - \mu_i)^{l-j-1}}{(\mu_1 - \mu_i)^l (d_+ - \mu_i)^l} \right| \\ &\leq \frac{l|d_+ - \mu_1|}{\min\{|d_+ - \mu_i|, |\mu_1 - \mu_i|\}^{l+1}}, \end{aligned}$$

and thus

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=K}^n \left(\frac{1}{(\mu_1 - \mu_i)^l} - \frac{1}{(d_+ - \mu_i)^l} \right) \right| &= O \left(\frac{1}{n} \sum_{i=K}^n \frac{l \cdot n^{-2/3}}{(d_+ - \mu_i)^{l+1}} \right) \\ &= O \left(n^{-5/3} \int_K^n l \left(\frac{x}{n} \right)^{-\frac{2}{3}(l+1)} dx \right) \\ &= O(n^{\frac{2}{3}l-1}). \end{aligned}$$

□

4.1. Computation of $\widehat{G}(\mu_1^{(1)}, \mu_1^{(2)})$

Lemma 4.2.

$$\begin{aligned} \widehat{G} &= A(d_+, B) - \frac{\log n}{3n} - \frac{1}{2n} \sum_{i=1}^n \log |d_+ - \mu_i| \\ &\quad + \frac{bn^{-\frac{1}{3}} \sqrt{\log n}}{\lambda^{\frac{1}{4}}(1+\lambda)^{\frac{1}{2}} d_+} (\mu_1 - d_+) + O(n^{-1}), \end{aligned}$$

where

$$A(x, B) := \sqrt{\alpha^2 + xB^2} - \alpha \log \left(\frac{\alpha + \sqrt{\alpha^2 + xB^2}}{B} \right). \quad (4.8)$$

Remark 4.3. The expression of \widehat{G} given by Lemma 4.2 contains two distinct random variables, $\sum_{i=1}^n \log |d_+ - \mu_i|$ and $\mu_1 - d_+$. Under appropriate translation and scaling, they are the quantities that give rise to the Gaussian and Tracy–Widom terms, respectively, in the convergence of free energy as stated in Theorem 1.1. The translation and scaling needed for these two random variables are, respectively, T_{1n} and T_{2n} , given by

$$T_{1n} = \frac{C_\lambda n - \frac{1}{6} \log n - \sum_{i=1}^n \log |d_+ - \mu_i|}{\sqrt{\frac{2}{3} \log n}}, \quad T_{2n} = \frac{n^{2/3}(\mu_1 - d_+)}{\sqrt{\lambda}(1 + \sqrt{\lambda})^{4/3}}, \quad (4.9)$$

where C_λ is as in (1.5). The expression of \widehat{G} then reads

$$\begin{aligned} \widehat{G} &= A(d_+, B) - \frac{1}{2}C_\lambda - \frac{\log n}{4n} + \left(\frac{1}{\sqrt{6}}T_{1n} + \frac{\lambda^{\frac{1}{4}}b}{(1 + \lambda^{\frac{1}{2}})^{\frac{2}{3}}(1 + \lambda)^{\frac{1}{2}}}T_{2n} \right) \frac{\sqrt{\log n}}{n} \\ &\quad + O(n^{-1}). \end{aligned} \quad (4.10)$$

Proof of Lemma 4.2. By definition,

$$\begin{aligned} \widehat{G} &= B_n(\mu_1^{(1)} + \mu_1^{(2)}) - \alpha_n \log(\mu_1^{(1)}) - \frac{1}{2n} \sum_{i=2}^n \log(\mu_1 - \mu_i) \\ &= \sqrt{\alpha_n^2 + \mu_1 B_n^2} - \alpha_n \log \left(\frac{\alpha_n + \sqrt{\alpha_n^2 + \mu_1 B_n^2}}{2B_n} \right) - \frac{1}{2n} \sum_{i=2}^n \log(\mu_1 - \mu_i). \end{aligned} \quad (4.11)$$

Replacing α_n, B_n by α, B , respectively (incurring an error of $n^{-1-\delta}$), and applying Taylor expansion with respect to μ_1 near d_+ , we obtain

$$\begin{aligned} &\sqrt{\alpha_n^2 + \mu_1 B_n^2} - \alpha_n \log \left(\frac{\alpha_n + \sqrt{\alpha_n^2 + \mu_1 B_n^2}}{2B_n} \right) \\ &= A(d_+, B) + \frac{B^2(\mu_1 - d_+)}{2(\alpha + \sqrt{\alpha^2 + d_+ B^2})} + O(n^{-1-\delta}). \end{aligned} \quad (4.12)$$

Note we have dropped the quadratic term in the Taylor expansion, which is $O(n^{-4/3})$. It remains to compute the summation in (4.11), which can be rewritten as

$$\sum_{i=2}^n \log(\mu_1 - \mu_i) = \sum_{i=2}^n \log |d_+ - \mu_i| - \frac{n(d_+ - \mu_1)}{\lambda^{\frac{1}{2}}(1 + \lambda^{\frac{1}{2}})} + E_1 + E_2 \quad (4.13)$$

where we define

$$\begin{aligned} E_1 &= n(d_+ - \mu_1) \left(\frac{1}{\lambda^{\frac{1}{2}}(1 + \lambda^{\frac{1}{2}})} - \frac{1}{n} \sum_{i=2}^n \frac{1}{\mu_1 - \mu_i} \right), \\ E_2 &= \sum_{i=2}^n \left(\frac{d_+ - \mu_1}{\mu_1 - \mu_i} - \log \left| 1 + \frac{d_+ - \mu_1}{\mu_1 - \mu_i} \right| \right). \end{aligned} \quad (4.14)$$

We now show $E_1 + E_2 = O(1)$, following an argument similar to that of Johnstone et al in [36]. The bound $E_1 = O(1)$ follows from Lemma 4.1. To bound E_2 , observe that, on the event we are considering, there exist k, C such that $\mu_1 \leq d_+ + Cn^{-2/3}$ and $\mu_k \leq d_+ - Cn^{-2/3}$. For any fixed i , we also have $d_+ - \mu_1 = \Theta(n^{-2/3})$ and $\mu_1 - \mu_i = \Theta(n^{-2/3})$. This implies

$$\sum_{i=2}^{k-1} \left(\frac{d_+ - \mu_1}{\mu_1 - \mu_i} - \log \left| 1 + \frac{d_+ - \mu_1}{\mu_1 - \mu_i} \right| \right) = O(1).$$

To bound the sum over the indices above k , we observe that, for $i \geq k$, we have $\frac{d_+ - \mu_1}{\mu_1 - \mu_i} \geq -\frac{1}{2}$ and, for any $x \geq -\frac{1}{2}$, there is C_1 such that $|\log(1+x) - x| \leq C_1 x^2$. This gives us

$$\sum_{i=k}^n \left(\frac{d_+ - \mu_1}{\mu_1 - \mu_i} - \log \left| 1 + \frac{d_+ - \mu_1}{\mu_1 - \mu_i} \right| \right) = O(1).$$

Finally, combining the results above and observing that $\frac{1}{2n} \log |d_+ - \mu_1| = -\frac{\log n}{3n} + O(n^{-1})$, we get

$$\begin{aligned} \widehat{G}(\mu_1^{(1)}, \mu_1^{(2)}) &= A(d_+, B) - \frac{\log n}{3n} - \frac{1}{2n} \sum_{i=1}^n \log |d_+ - \mu_i| + c_2(B)(\mu_1 - d_+) \\ &\quad + O(n^{-1}), \end{aligned} \quad (4.15)$$

where

$$c_2(B) = \frac{B^2}{2(\alpha + \sqrt{\alpha^2 + d_+ B^2})} - \frac{1}{2\lambda^{1/2}(1 + \lambda^{1/2})}. \quad (4.16)$$

Recall that B_c is defined to be the quantity satisfying

$$\frac{\sqrt{\alpha^2 + d_+ B_c^2} - \alpha}{d_+} = \int \frac{p_{\text{MP}}(x)}{d_+ - x} dx = \frac{1}{\lambda^{1/2}(1 + \lambda^{1/2})}. \quad (4.17)$$

Using this definition along with a Taylor expansion of c_2 near $B = B_c = \lambda^{-\frac{3}{4}}$, we get

$$\begin{aligned} c_2(B) &= \frac{B_c}{2\sqrt{\alpha^2 + d_+ B_c^2}} (B - B_c) + O((B - B_c)^2) \\ &= \frac{bn^{-\frac{1}{3}} \sqrt{\log n}}{\lambda^{\frac{1}{4}}(1 + \lambda)^{\frac{1}{2}} d_+} + O(n^{-2/3} \log n). \end{aligned}$$

Apply this to (4.15), we obtain the lemma. \square

4.2. Contour Integral Analysis

We now derive the asymptotics of the rescaled double integral

$$S_n := \exp(-n\widehat{G})Q_n = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[n(G(\gamma_1 + iy_1, \gamma_2 + iy_2) - \widehat{G})] dy_2 dy_1. \quad (4.18)$$

The analysis holds on the following probability event \mathcal{F}_ε for arbitrarily small $\varepsilon > 0$.

Lemma 4.4. *For each $\varepsilon > 0$, there exist positive numbers r, s, t , and C , depending on ε , such that the event \mathcal{F}_ε given by*

$$\mathcal{F}_\varepsilon = \left\{ \left| \sum_{j=2}^n \frac{1}{n^{\frac{2}{3}}(\mu_1 - \mu_j)} - s_{\text{MP}}(d_+) \right| \leq C \right\}$$

$$\cap \left\{ \sum_{j=2}^n \frac{1}{n^{\frac{4}{3}} (\mu_1 - \mu_j)^2} \leq C \right\} \mathcal{F}_{s,t}^{(3)} \cap \mathcal{F}_{r,R}^{(4)}$$

satisfies $\mathbb{P}(\mathcal{F}_\varepsilon) > 1 - \varepsilon$.

We note that the definition of \mathcal{F}_ε is not unique as it depends on the choice of s, t, r, R , and C . For any given $\varepsilon > 0$, we fix the values s, t, r, R, C and define \mathcal{F}_ε accordingly.

Proof. First, for some $C > 0$, each of the two events that involve $\frac{1}{n^{2/3}(\mu_1 - \mu_j)}$, with this C as upper bound, holds with probability at least $1 - \varepsilon/4$ by Lemma 4.1. Meanwhile, by Lemma 2.5, we can find $0 < s < t$ and $0 < r < R$ such that each of the events $\mathcal{F}_{s,t}^{(3)}$ and $\mathcal{F}_{r,R}^{(4)}$ holds with probability at least $1 - \varepsilon/4$. \square

Since the integral representation of the partition function only requires $\gamma_1, \gamma_2 > 0$ such that $4\gamma_1\gamma_2 > \mu_1$, we set $\gamma_1 = \mu_1^{(1)}$ and $\gamma_2 = \mu_1^{(2)} + n^{-1}$ in the low-temperature case. The shift n^{-1} in γ_2 is due to the deformation $\hat{\gamma}_2$, given in (4.20), that we later apply to the integral in the y_2 variable. The order n^{-1} is needed to cancel out a term of order n of the function in the exponent (see, for example, (4.23)). Thus,

$$S_n = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp[n(G(\mu_1^{(1)} + iy_1, \mu_1^{(2)} + n^{-1} + iy_2) - \hat{G})] dy_2 dy_1. \tag{4.19}$$

In the remainder of the subsection, we prove the following lemma, for fixed $\varepsilon > 0$ sufficiently small (e.g., $0 < \varepsilon < \frac{1}{100}$).

Lemma 4.5. *On the event \mathcal{F}_ε ,*

$$S_n = \begin{cases} e^{O(1)} n^{-\frac{5}{6}} (b\sqrt{\log n})^{-\frac{1}{2}}, & b > 0, \\ e^{O(\log \log n)} n^{-\frac{5}{6}}, & b = 0. \end{cases}$$

By Lemma 3.9 of [13], the part of the double integral S_n with $|y_1| > n^{-\frac{1}{2} + \varepsilon}$ is $O(e^{-n^\varepsilon})$ with high probability. For $|y_1| < n^{-\frac{1}{2} + \varepsilon}$, we modify the z_2 -integral by replacing the vertical contour $z_2 = \gamma_2 + iy_2$, $y_2 \in \mathbb{R}$ with the contour $z_2 = \hat{\gamma}_2 + iy_2$, $y_2 \in \mathbb{R}$, where $\hat{\gamma}_2$ is defined for each y_1 by

$$\hat{\gamma}_2(y_1) = \frac{\mu_1^{(1)}(\mu_1^{(2)} + n^{-1})}{\mu_1^{(1)} + iy_1}. \tag{4.20}$$

The new contour is a modification of the one introduced by Baik and Lee in [13]. Similar to the case in [13], we observe that the change in product $z_1 z_2$ for (z_1, z_2) near $(\mu_1^{(1)}, \mu_1^{(2)})$, but not the individual changes in z_1, z_2 with $z_1 z_2$ being fixed, greatly impacts the change in $G(z_1, z_2)$, since the main contribution for the latter comes from the term $\frac{1}{4z_1 z_2 - \mu_1}$. This suggests behavior of $G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2)$ should be similar to that of $G(\mu_1^{(1)}, \mu_1^{(2)} + n^{-1})$ for the current range of y_1 .

Note that this deformation for each $z_1 = \mu_1^{(1)} + iy_1$ is valid. Indeed, if (z_1, z_2) is a point on the branch cut of the logarithmic function in G , then $4z_1z_2 - \mu_1$ is real and non-positive. That is, for some $r \geq 0$,

$$\operatorname{Re} z_2 = \operatorname{Re} \left(\frac{\mu_1 - r}{4(\mu_1^{(1)} + iy_1)} \right) = \operatorname{Re} \left(\frac{\mu_1 - r}{4\mu_1^{(1)}(\mu_1^{(2)} + n^{-1})} \hat{\gamma}_2 \right) < \operatorname{Re} \hat{\gamma}_2.$$

This implies that the deformed contour does not cross the branch cut. Thus, the part of S_n with $|y_1| < n^{-1/2+\varepsilon}$ is equal to

$$\int_{-n^{-1/2+\varepsilon}}^{n^{-1/2+\varepsilon}} \int_{-\infty}^{\infty} \exp[n(G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \hat{G})] dy_2 dy_1.$$

We now carry out the analysis of this double integral, first by truncating the y_2 -integral. For given $y_1, y_2 \in \mathbb{R}$,

$$\begin{aligned} & G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \hat{G} \\ &= B_n \left(i(y_1 + y_2) + \frac{\mu_1^{(2)} + n^{-1}}{1 + i\frac{y_1}{\mu_1^{(1)}}} - \mu_1^{(2)} \right) - \alpha_n \log \left(1 + \frac{iy_1}{\mu_1^{(1)}} \right) \\ &\quad - \frac{1}{2n} \sum_{j=2}^n \log \left(1 + \frac{4\mu_1^{(1)}n^{-1} - 4y_1y_2}{\mu_1 - \mu_j} + i\frac{4\mu_1^{(1)}y_2}{\mu_1 - \mu_j} \right) \\ &\quad - \frac{1}{2n} \log(4\mu_1^{(1)}n^{-1} - 4y_1y_2 + i4\mu_1^{(1)}y_2). \end{aligned} \tag{4.21}$$

Our truncation procedure, which relies on bounding $|G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \hat{G}|$, aligns rather closely with the arguments in [13], where the difference $|G(\gamma_1 + iy_1, \hat{\gamma}_2 + iy_2) - G(\gamma_1, \gamma_2)|$ is the focus there. After truncating in the y_1 variable, the contribution from the part $|y_2| > n^{-\frac{1}{2}+\varepsilon}$ is as follows.

Lemma 4.6. *The following bound holds for the truncated integral.*

$$\int_{|y_1| \leq n^{-\frac{1}{2}+\varepsilon}} \int_{|y_2| > n^{-\frac{1}{2}+\varepsilon}} \exp[n(G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \hat{G})] dy_2 dy_1 = O(n^{-1}). \tag{4.22}$$

Proof. From (4.21),

$$\begin{aligned}
 & \operatorname{Re} \left[n \left(G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \widehat{G} \right) \right] \\
 &= \frac{B_n(\mu_1^{(1)})^2 - nB_n\mu_1^{(2)}y_1^2}{(\mu_1^{(1)})^2 + y_1^2} - \frac{\alpha_n n}{2} \log \left(1 + \left(\frac{y_1}{\mu_1^{(1)}} \right)^2 \right) \\
 &\quad - \frac{1}{4} \log \left(\left(\frac{4\mu_1^{(1)}}{n} - 4y_1y_2 \right)^2 + (4\mu_1^{(1)}y_2)^2 \right) \\
 &\quad - \frac{1}{4} \sum_{j=2}^n \log \left(\left(1 + \frac{4\mu_1^{(1)}n^{-1} - 4y_1y_2}{\mu_1 - \mu_j} \right)^2 + \frac{(4\mu_1^{(1)}y_2)^2}{(\mu_1 - \mu_j)^2} \right). \quad (4.23)
 \end{aligned}$$

Applying Taylor expansion in terms of y_1 around 0 to the first two terms on the right-hand side of (4.23), then for some $c > 0$, the first line has upper bound

$$c_0 - cn y_1^2 - \frac{1}{2} \log(4\mu_1^{(1)}|y_2|), \quad \text{uniformly in } |y_1| \leq n^{-\frac{1}{2}+\varepsilon}.$$

For the sum of log, by consider the cases $y_1y_2 > 0$ and $y_1y_2 < 0$ as in [13], there exists $c' > 0$ such that for all $j \in \{2, 3, \dots, n\}$, for all $|y_1| < n^{-\frac{1}{2}+\varepsilon}$ and $|y_2| > n^{-\frac{1}{2}+\varepsilon}$,

$$\left(1 + \frac{4\mu_1^{(1)}n^{-1} - 4y_1y_2}{\mu_1 - \mu_j} \right)^2 + \frac{(4\mu_1^{(1)}y_2)^2}{(\mu_1 - \mu_j)^2} \geq 1 + c'y_2^2.$$

Therefore,

$$\begin{aligned}
 \operatorname{Re} \left[n \left(G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \widehat{G} \right) \right] &\leq c_0 - cn y_1^2 - \frac{1}{2} \log(4\mu_1^{(1)}|y_2|) \\
 &\quad - \frac{n}{4} \log(1 + c'y_2^2),
 \end{aligned}$$

and the left-hand side of (4.22) has upper bound

$$\int_{|y_1| \leq n^{-\frac{1}{2}+\varepsilon}} \int_{|y_2| > n^{-\frac{1}{2}+\varepsilon}} e^{c_0 - cn y_1^2} e^{-\frac{n}{4} \log(1 + c'y_2^2)} (4\mu_1^{(1)}|y_2|)^{-\frac{1}{2}} dy_2 dy_1,$$

which is a product of a y_1 -integral and a y_2 -integral. Each individual integral is $O(n^{-\frac{1}{2}})$, so we obtain the lemma. \square

The computation of S_n is now reduced to that of the same integral, over the subset $|y_1| \leq n^{-\frac{1}{2}+\varepsilon}$ and $|y_2| < n^{-\frac{1}{2}+\varepsilon}$. However, we need to truncate the y_2 -integral further.

Lemma 4.7. *For this further truncation, we have the following bound.*

$$\begin{aligned}
 & \int_{|y_1| \leq n^{-\frac{1}{2}+\varepsilon}} \int_{n^{-\frac{2}{3}+2\varepsilon} < |y_2| < n^{-\frac{1}{2}+\varepsilon}} \exp[n(G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \widehat{G})] dy_2 dy_1 \\
 &= O(e^{-n^{4\varepsilon}}).
 \end{aligned}$$

Proof. Computations similar to the proof of Lemma 4.6 give

$$\begin{aligned} & \operatorname{Re} \left[n \left(G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \widehat{G} \right) \right] \\ & \leq c_0 - \frac{1}{4} \sum_{j=2}^n \log \left(\left(1 + \frac{4\mu_1^{(1)}n^{-1} - 4y_1y_2}{\mu_1 - \mu_j} \right)^2 + \frac{(4\mu_1^{(1)}y_2)^2}{(\mu_1 - \mu_j)^2} \right). \end{aligned} \quad (4.24)$$

Observe that $n^{-\frac{2}{3}} \ll \mu_1 - \mu_{n^{4\varepsilon}} \ll n^{-\frac{2}{3}+2\varepsilon}$. Thus, for $2 \leq j \leq n^{4\varepsilon}$, $\left(\frac{4\mu_1^{(1)}y_2}{\mu_1 - \mu_j} \right)^2 \geq (4\mu_1^{(1)})^2$, and we obtain

$$-\frac{1}{4} \sum_{j=1}^{n^{4\varepsilon}} \log \left(\left(1 + \frac{4\mu_1^{(1)}n^{-1} - 4y_1y_2}{\mu_1 - \mu_j} \right)^2 + \frac{(4\mu_1^{(1)}y_2)^2}{(\mu_1 - \mu_j)^2} \right) \leq -\frac{1}{2} \log(\mu_1^{(1)})n^{4\varepsilon}.$$

For $j > n^{4\varepsilon}$, $\mu_1 - \mu_j \geq \mu_1 - \mu_{n^{4\varepsilon}} \gg n^{-\frac{2}{3}}$. Since $|y_1|, |y_2| \leq n^{-\frac{1}{2}+\varepsilon}$, we have $\mu_1 - \mu_j \gg |4\mu_1^{(1)}n^{-1} - 4y_1y_2|$. Using $\log(1-x) \geq -2x$ for $x \in (0, 1)$, then for some constant $C, C' > 0$, the sum with indices $j > n^{4\varepsilon}$ on the right-hand side of (4.24) has upper bound

$$\begin{aligned} -\frac{1}{2} \sum_{j=n^{4\varepsilon}+1}^n \log \left(1 + \frac{4\mu_1^{(1)}n^{-1} - 4y_1y_2}{\mu_1 - \mu_j} \right) & \leq 4|\mu_1^{(1)}n^{-1} - y_1y_2| \sum_{j=n^{4\varepsilon}+1}^n \frac{1}{\mu_1 - \mu_j} \\ & \leq Cn|\mu_1^{(1)}n^{-1} - y_1y_2| \leq C'n^{2\varepsilon}. \end{aligned}$$

Here, the second inequality holds with probability at least $1 - \varepsilon$ by Lemma 4.1. Thus, we obtain the uniform bound

$$\operatorname{Re} \left[n \left(G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \widehat{G} \right) \right] \leq c_0 - Cn^{4\varepsilon}$$

for some constant $C > 0$. This implies the lemma. \square

Therefore, we have shown that

$$S_n = \int_{-n^{-\frac{1}{2}+\varepsilon}}^{n^{-\frac{1}{2}+\varepsilon}} \int_{-n^{-\frac{2}{3}+2\varepsilon}}^{n^{-\frac{2}{3}+2\varepsilon}} \exp[n(G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \widehat{G})] dy_2 dy_1 + O(n^{-1}). \quad (4.25)$$

We proceed to compute the double integral in (4.25). For $|y_1| \leq n^{-\frac{1}{2}+\varepsilon}$ and $|y_2| < n^{-\frac{2}{3}+2\varepsilon}$, by Taylor series and the definitions of $\mu_1^{(1)}$ and $\mu_1^{(2)}$ in (4.1), the second line of (4.21) for $G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \widehat{G}$ is

$$B_n(n^{-1} + iy_2) - i \frac{B_n n^{-1}}{\mu_1^{(1)}} y_1 - \frac{B_n \left(\frac{\mu_1^{(1)} + \mu_1^{(2)}}{2} + n^{-1} \right)}{(\mu_1^{(1)})^2} y_1^2 + O(y_1^3),$$

while the last line, after factorizing the arguments of logarithm functions, becomes

$$-\frac{1}{2n} \sum_{j=2}^n \log \left(1 + \frac{4\mu_1^{(1)}n^{-1} + 4i\mu_1^{(1)}y_2}{\mu_1 - \mu_j} \right) - \frac{1}{2n} \log(4\mu_1^{(1)}n^{-1} + 4i\mu_1^{(1)}y_2)$$

$$\begin{aligned}
 & -\frac{1}{2n} \sum_{j=2}^n \log \left(1 - \frac{4y_1 y_2}{\mu_1 - \mu_j + 4\mu_1^{(1)} n^{-1} + 4i\mu_1^{(1)} y_2} \right) \\
 & -\frac{1}{2n} \log \left(1 - \frac{4y_1 y_2}{4\mu_1^{(1)} n^{-1} + 4i\mu_1^{(1)} y_2} \right). \tag{4.26}
 \end{aligned}$$

Combining the above two displays, we obtain

$$\begin{aligned}
 & \exp[n(G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \widehat{G})] \\
 & = \exp \left[-\frac{iB_n}{\mu_1^{(1)}} y_1 - \frac{B_n(\frac{\mu_1^{(1)} + \mu_1^{(2)}}{2} + n^{-1})}{(\mu_1^{(1)})^2} n y_1^2 \right] \\
 & \cdot \exp \left[B_n n(n^{-1} + iy_2) - \frac{1}{2} \log(4\mu_1^{(1)} n^{-1} + 4i\mu_1^{(1)} y_2) \right. \\
 & \left. - \frac{1}{2} \sum_{j=2}^n \log \left(1 + \frac{4\mu_1^{(1)} n^{-1} + 4i\mu_1^{(1)} y_2}{\mu_1 - \mu_j} \right) \right] \\
 & \cdot \exp \left[-\frac{1}{2} \sum_{j=1}^n \log \left(1 - \frac{4y_1 y_2}{\mu_1 - \mu_j + 4\mu_1^{(1)} n^{-1} + 4i\mu_1^{(1)} y_2} \right) + O(ny_1^3) \right]. \tag{4.27}
 \end{aligned}$$

Let $H(y_1, y_2)$ denote the product of the first two exponential factors on the right-hand side of (4.27) and $L(y_1, y_2)$ be the last factor. That is,

$$\exp[n(G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \widehat{G})] = H(y_1, y_2)L(y_1, y_2).$$

There is a constant $c > 0$ such that $\frac{B_n(\frac{\mu_1^{(1)} + \mu_1^{(2)}}{2} + n^{-1})}{(\mu_1^{(1)})^2} > c$, so

$$\begin{aligned}
 |H(y_1, y_2)| & \leq \exp \left[B_n - \frac{B_n(\frac{\mu_1^{(1)} + \mu_1^{(2)}}{2} + n^{-1})}{(\mu_1^{(1)})^2} n y_1^2 \right. \\
 & \quad \left. - \frac{1}{2} \operatorname{Re} \sum_{j=2}^n \log \left(1 + \frac{4\mu_1^{(1)} n^{-1} + 4i\mu_1^{(1)} y_2}{\mu_1 - \mu_j} \right) \right] \\
 & \leq \exp \left[B_n - c n y_1^2 - \frac{1}{2} \log \left(\frac{4\mu_1^{(1)} |y_2|}{\mu_1 - \mu_2} \right) \right] \\
 & \leq C(\mu_1 - \mu_2)^{\frac{1}{2}} |y_2|^{-\frac{1}{2}} e^{-c n y_1^2}, \tag{4.28}
 \end{aligned}$$

for some constant $C > 0$. On the other hand, by Lemma 4.1, there exists constant $C > 0$ such that

$$\sum_{j=2}^n \frac{1}{|\mu_1 - \mu_j + 4\mu_1^{(1)} n^{-1} + 4i\mu_1^{(1)} y_2|^\ell} \leq \sum_{j=2}^n \frac{1}{(\mu_1 - \mu_j)^\ell}$$

$$\leq \begin{cases} Cn^{1+\varepsilon}, & \ell = 1, \\ Cn^{\frac{2\ell}{3}+\varepsilon}, & \ell = 2, 3, \dots \end{cases} \quad (4.29)$$

At the same time,

$$\left| \frac{4y_1y_2}{4\mu_1^{(1)}n^{-1} + 4i\mu_1^{(1)}y_2} \right| \leq \frac{|y_1|}{\mu_1^{(1)}} = O(n^{-\frac{1}{2}+\varepsilon}).$$

Thus, applying Taylor series, we have

$$L(y_1, y_2) = 1 + \sum_{k=1}^n \frac{2y_1y_2}{\mu_1 - \mu_j + 4\mu_1^{(1)}n^{-1} + 4i\mu_1^{(1)}y_2} + O(n^{-\frac{1}{2}+3\varepsilon}). \quad (4.30)$$

Observe that

$$|\mu_1 - \mu_j + 4\mu_1^{(1)}n^{-1} + 4i\mu_1^{(1)}y_2| \geq \begin{cases} \mu_1 - \mu_j, & j = 2, 3, \dots, n, \\ 4\mu_1^{(1)}|y_2|, & j = 1. \end{cases}$$

Applying (4.29) with $\ell = 1$, we obtain

$$|L(y_1, y_2) - 1| \leq Cn|y_1y_2| + C'n^{-\frac{1}{2}+3\varepsilon}. \quad (4.31)$$

We now write

$$\int_{-n^{-\frac{1}{2}+\varepsilon}}^{n^{-\frac{1}{2}+\varepsilon}} \int_{-n^{-\frac{2}{3}+2\varepsilon}}^{n^{-\frac{2}{3}+2\varepsilon}} \exp[n(G(\mu_1^{(1)} + iy_1, \hat{\gamma}_2 + iy_2) - \hat{G})] dy_2 dy_1 = I_1 + I_2, \quad (4.32)$$

where I_2 is given by

$$I_2 = \int_{-n^{-\frac{1}{2}+\varepsilon}}^{n^{-\frac{1}{2}+\varepsilon}} \int_{-n^{-\frac{2}{3}+2\varepsilon}}^{n^{-\frac{2}{3}+2\varepsilon}} H(y_1, y_2)(L(y_1, y_2) - 1) dy_2 dy_1. \quad (4.33)$$

By (4.28) and (4.31), there is constant $C_j > 0, j = 1, 2, 3$ such that

$$\begin{aligned} |I_2| &\leq C_1 \int_{-n^{-\frac{1}{2}+\varepsilon}}^{n^{-\frac{1}{2}+\varepsilon}} \int_{-n^{-\frac{2}{3}+2\varepsilon}}^{n^{-\frac{2}{3}+2\varepsilon}} |H(y_1, y_2)| \left(n|y_1y_2| + n^{-\frac{1}{2}+3\varepsilon} \right) dy_2 dy_1 \\ &\leq C_2 n(\mu_1 - \mu_2)^{\frac{1}{2}} \int_{-n^{-\frac{1}{2}+\varepsilon}}^{n^{-\frac{1}{2}+\varepsilon}} \int_{-n^{-\frac{2}{3}+2\varepsilon}}^{n^{-\frac{2}{3}+2\varepsilon}} e^{-cny_1^2} \left(|y_1||y_2|^{\frac{1}{2}} + n^{-\frac{3}{2}+3\varepsilon}|y_2|^{-\frac{1}{2}} \right) dy_2 dy_1 \\ &\leq C_3 n^{-1+4\varepsilon} (\mu_1 - \mu_2)^{\frac{1}{2}}. \end{aligned} \quad (4.34)$$

Together with (4.25) and (4.32), this implies that on the event $\mu_1 - \mu_2 \leq n^{-2/3+\varepsilon}$,

$$S_n = I_1 + O(n^{-1}).$$

Note that

$$I_1 = \int_{-n^{-\frac{1}{2}+\varepsilon}}^{n^{-\frac{1}{2}+\varepsilon}} \int_{-n^{-\frac{2}{3}+2\varepsilon}}^{n^{-\frac{2}{3}+2\varepsilon}} H(y_1, y_2) dy_2 dy_1 \quad (4.35)$$

is equal to the product of two single integrals I_{11} and I_{12} as follows. First,

$$\begin{aligned} I_{11} &= \int_{-n^{-\frac{1}{2}+\varepsilon}}^{n^{-\frac{1}{2}+\varepsilon}} \exp \left[-\frac{iB_n}{\mu_1^{(1)}} y_1 - \frac{B_n(\frac{\mu_1^{(1)}+\mu_1^{(2)}}{2} + n^{-1})}{(\mu_1^{(1)})^2} n y_1^2 \right] dy_1 \\ &= n^{-\frac{1}{2}} \int_{-n^\varepsilon}^{n^\varepsilon} e^{-c_1 x^2} \cos \left(\frac{c_2}{\sqrt{n}} x \right) dx, \\ (c_1, c_2) &:= \left(\frac{B_n(\frac{\mu_1^{(1)}+\mu_1^{(2)}}{2} + n^{-1})}{(\mu_1^{(1)})^2}, \frac{B_n}{\mu_1^{(1)}} \right). \end{aligned}$$

Using Taylor's series of cosine, we obtain that for some $C > 0$,

$$I_{11} = C n^{-\frac{1}{2}} (1 + O(n^{-1+2\varepsilon})). \quad (4.36)$$

Second, we have

$$I_{12} = \int_{-n^{-\frac{2}{3}+2\varepsilon}}^{n^{-\frac{2}{3}+2\varepsilon}} \exp \left[n(G(\mu_1^{(1)}, \mu_1^{(2)} + n^{-1} + iy) - \widehat{G}) \right] dy. \quad (4.37)$$

We first check that I_{12} is close to the integral over the whole real line

$$K_n := \int_{-\infty}^{\infty} \exp \left[n(G(\mu_1^{(1)}, \mu_1^{(2)} + n^{-1} + iy) - \widehat{G}) \right] dy. \quad (4.38)$$

By (4.27), for all $y \in \mathbb{R}$,

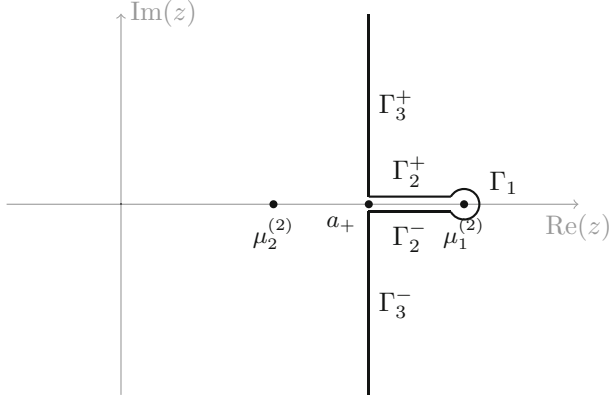
$$\begin{aligned} &\operatorname{Re} \left[n \left(G(\mu_1^{(1)}, \mu_1^{(2)} + \frac{1}{n} + iy) - \widehat{G} \right) \right] \\ &\leq c_0 - \frac{1}{4} \log \left((4\mu_1^{(1)} n^{-1})^2 + (4\mu_1^{(1)} y_2)^2 \right) \\ &\quad - \frac{1}{4} \sum_{j=2}^n \log \left(\left(1 + \frac{4\mu_1^{(1)}/n}{\mu_1 - \mu_j} \right)^2 + \left(\frac{4\mu_1^{(1)} y}{\mu_1 - \mu_j} \right)^2 \right). \end{aligned}$$

In the case $n^{-\frac{2}{3}+2\varepsilon} < |y| < n$, we use $-\frac{1}{4} \log \left((4\mu_1^{(1)} n^{-1})^2 + (4\mu_1^{(1)} y_2)^2 \right) \leq \frac{1}{2} \log n$ and bound

$$\sum_{j=2}^n \log \left(\left(1 + \frac{4\mu_1^{(1)}/n}{\mu_1 - \mu_j} \right)^2 + \left(\frac{4\mu_1^{(1)} y}{\mu_1 - \mu_j} \right)^2 \right) \geq 2 \sum_{j=2}^{n^{2\varepsilon}} \log \left(\frac{4\mu_1^{(1)} |y|}{\mu_1 - \mu_j} \right) \geq C n^{2\varepsilon}$$

using the fact that $\mu_1 - \mu_{n^{2\varepsilon}} \ll n^{-\frac{2}{3}+2\varepsilon}$ with high probability. For $|y| > n$, we drop the negative term $-\frac{1}{4} \log \left((4\mu_1^{(1)} n^{-1})^2 + (4\mu_1^{(1)} y_2)^2 \right)$, while, for some $c > 0$,

$$\sum_{j=2}^n \log \left(\left(1 + \frac{4\mu_1^{(1)}/n}{\mu_1 - \mu_j} \right)^2 + \left(\frac{4\mu_1^{(1)} y}{\mu_1 - \mu_j} \right)^2 \right) \geq n \log(1 + cy^2) \geq n \log(c|y|).$$


 FIGURE 1. Keyhole-like contour of integration Γ

Therefore, for some $C', C'' > 0$, it holds with high probability that

$$|K_n - I_{12}| \leq C' \left(n^{1/2} e^{-Cn^{-2\epsilon}} + \int_0^\infty (cy)^{-\frac{n}{4}} \right) dy \leq C'' e^{-c'n^{2\epsilon}}. \quad (4.39)$$

We determine in Sect. 4.2.1 that, on the event \mathcal{F}_ϵ ,

$$K_n = \begin{cases} e^{O(1)} n^{-\frac{1}{3}} (b\sqrt{\log n})^{-\frac{1}{2}}, & b > 0, \\ e^{O(\log \log n)} n^{-\frac{1}{3}}, & b = 0. \end{cases} \quad (4.40)$$

Assuming (4.40) is true, then using (4.39) and the fact that $S_n = I_{11} \cdot I_{12} + O(n^{-1})$, we obtain Lemma 4.5.

4.2.1. Proof of (4.40) When $b > 0$. For brevity, we introduce the following two notations to be used throughout the subsection:

$$a_+ = \frac{\mu_1^{(2)} + \mu_2^{(2)}}{2} = \frac{\mu_1}{8\mu_1^{(1)}} + \frac{\mu_2}{8\mu_2^{(1)}}, \quad (4.41)$$

where $\mu_2^{(1)} := \frac{\alpha_n + \sqrt{\alpha_n^2 + \mu_2 B_n^2}}{2B_n}$ and $\mu_2^{(2)} := \frac{-\alpha_n + \sqrt{\alpha_n^2 + \mu_2 B_n^2}}{2B_n}$.

We now show that the integral K_n , on the event \mathcal{F}_ϵ , satisfies (4.40), first under the assumption $b > 0$. By Cauchy theorem, for every $r \in (0, n^{-1}]$,

$$iK_n = \int_\Gamma \exp \left[n(G(\mu_1^{(1)}, z) - \widehat{G}) \right] dz,$$

where $\Gamma = \Gamma_1 \cup \Gamma_2^\pm \cup \Gamma_3^\pm$ is the vertical keyhole-like contour as in Fig. 1. In particular, given a function $\phi_r : \mathbb{R}_+ \rightarrow [0, \pi]$ of r such that $\phi_r \rightarrow 0$ as $r \downarrow 0$, we let Γ_1 be the arc $\{\mu_1^{(2)} + re^{i\theta} : \theta \in [-\pi + \phi_r, \pi - \phi_r]\}$, $\Gamma_2^\pm = \{x \pm r \sin \phi_r : x \in [a_+, \mu_1^{(2)} - r \cos \phi_r]\}$, and Γ_3^\pm be the rays $\{a_+ \pm iy : y \in [r \sin \phi_r, \infty)\}$. Then, for fixed n ,

$$iK_n = \lim_{r \downarrow 0} \int_\Gamma \exp \left[n(G(\mu_1^{(1)}, z) - \widehat{G}) \right] dz. \quad (4.42)$$

For Γ_1 , using the fact that $\log(x+it) \rightarrow \log|x|+i\pi$ as $t \downarrow 0$ for $x < 0$, and $dz = ire^{i\theta}d\theta$ where θ takes values in $[-\pi - \phi_r, \pi + \phi_r]$ as described above, one can verify using Fubini's that for each fixed n , the integral over Γ_1 converges to 0 as $r \rightarrow 0$.

We show in Lemma 4.8 that, in the limit $r \downarrow 0$, the contribution from $\Gamma_2^+ \cup \Gamma_2^-$ part of the contour satisfies the asymptotics (4.40) in both cases $b > 0$ and $b = 0$. In Lemma 4.9, we confirm that for any keyhole radius $r \in (0, 1/n]$, with probability arbitrarily close to 1, the contribution from $\Gamma_3^+ \cup \Gamma_3^-$ is little-o of that of $\Gamma_2^+ \cup \Gamma_2^-$ when $b > 0$. Together, the lemmas establish (4.40) when $b > 0$.

Lemma 4.8. *On the event \mathcal{F}_ε , it holds that*

$$\lim_{r \downarrow 0} \int_{\Gamma_2^+ \cup \Gamma_2^-} \exp \left[n(G(\mu_1^{(1)}, z) - \widehat{G}) \right] dz = \begin{cases} ie^{O(1)} n^{-\frac{1}{3}} (b\sqrt{\log n})^{-\frac{1}{2}}, & b > 0, \\ ie^{O(1)} n^{-\frac{1}{3}}, & b = 0. \end{cases} \tag{4.43}$$

Proof. Recall that, if $z \in \Gamma_2^\pm$, then $z = x \pm ir \sin \phi_r$ where $x \in [a_+, \mu_1^{(2)} - r \cos \phi_r]$. Set $s = \mu_1^{(2)} - x$, we have

$$\begin{aligned} n(G(\mu_1^{(1)}, z) - \widehat{G}) &= -nB_n(\mu_1^{(2)} - x) \pm inB_n r \sin \phi_r \\ &\quad - \frac{1}{2} \sum_{j=2}^n \log \left(1 - \frac{4\mu_1^{(1)}(\mu_1^{(2)} - x) \mp i4\mu_1^{(1)} r \sin \phi_r}{\mu_1 - \mu_j} \right) \\ &\quad - \frac{1}{2} \log \left(-4\mu_1^{(1)}(\mu_1^{(2)} - x) \pm i4\mu_1^{(1)} r \sin \phi_r \right) \\ &\stackrel{r \downarrow 0}{\rightarrow} -B_n n s - \frac{1}{2} \sum_{j=2}^n \log \left(1 - \frac{4\mu_1^{(1)}(\mu_1^{(2)} - x)}{\mu_1 - \mu_j} \right) \\ &\quad - \frac{1}{2} \log(4\mu_1^{(1)}(\mu_1^{(2)} - x)) \mp i \frac{\pi}{2}. \end{aligned}$$

Let A be the left-hand side of (4.43). We then obtain

$$A = \frac{2i}{\sqrt{4\mu_1^{(1)}}} \int_0^{\mu_1^{(2)} - a_+} \exp \left(-B_n n s - \frac{1}{2} \sum_{j=2}^n \log \left(1 - \frac{4\mu_1^{(1)} s}{\mu_1 - \mu_j} \right) \right) \frac{ds}{\sqrt{s}}. \tag{4.44}$$

Observe that $\frac{4\mu_1^{(1)} s}{\mu_1 - \mu_j} \in [0, \frac{1}{2}]$ for all $s \in [0, \mu_1^{(2)} - a_+]$ and all j . As $0 < -\log(1-x) - x \leq x^2$ for $x \in [0, \frac{1}{2}]$, there exists $\zeta \in [0, 1]$ such that

$$\begin{aligned} &-B_n n s - \frac{1}{2} \sum_{j=2}^n \log \left(1 - \frac{4\mu_1^{(1)} s}{\mu_1 - \mu_j} \right) \\ &= -B_n n s + 2\mu_1^{(1)} s \sum_{j=2}^n \frac{1}{\mu_1 - \mu_j} + \frac{\zeta(4\mu_1^{(1)} s)^2}{2} \sum_{j=2}^n \frac{1}{(\mu_1 - \mu_j)^2}. \end{aligned} \tag{4.45}$$

Define $y := n^{2/3}s \in [0, n^{2/3}(\mu_1^{(2)} - a_+)]$, and let ω_{1n}, ω_{2n} be random variables given by

$$\sum_{j=2}^n \frac{1}{n^{\frac{2}{3}}(\mu_1 - \mu_j)} = s_{\text{MP}}(d_+)n^{\frac{1}{3}} + \omega_{1n}, \quad \sum_{j=2}^n \frac{1}{\left(n^{\frac{2}{3}}(\mu_1 - \mu_j)\right)^2} = \omega_{2n}. \quad (4.46)$$

Then, (4.45) simplifies to

$$\begin{aligned} & -B_n n s - \frac{1}{2} \sum_{j=2}^n \log \left(1 - \frac{4\mu_1^{(1)}s}{\mu_1 - \mu_j} \right) \\ & = n^{\frac{1}{3}}y \left(-B_n + 2\mu_1^{(1)}s_{\text{MP}}(d_+) \right) + \left[y(2\mu_1^{(1)}\omega_{1n} + 2(\mu_1^{(1)})^2\zeta\omega_{2n}y^2) \right], \end{aligned} \quad (4.47)$$

where the term inside the square brackets is $O(1)$, uniformly for $y \in [0, n^{2/3}(\mu_1^{(2)} - a_+)]$. Observe also

$$-B_n + 2\mu_1^{(1)}s_{\text{MP}}(d_+) = -B_n + \frac{\alpha_n + \sqrt{\alpha_n^2 + \mu_1 B_n^2}}{B_n} s_{\text{MP}}(d_+),$$

where $B_n - B_c = \Theta(\beta - \beta_c)$ and B_c satisfies $\sqrt{\alpha^2 + d_+ B_c^2} = \alpha + d_+ s_{\text{MP}}(d_+)$. Therefore, applying Taylor expansion to the above expression with respect to B_n near B_c and μ_1 near d_+ , using $\mu_1 - d_+ = O(n^{-2/3})$ on the event \mathcal{F}_ε , we obtain

$$-B_n + 2\mu_1^{(1)}s_{\text{MP}}(d_+) = -\frac{2s_{\text{MP}}(d_+)\lambda^{\frac{1}{2}}}{\sqrt{1+\lambda}}(\beta - \beta_c) + O((\beta - \beta_c)^2). \quad (4.48)$$

Thus, on the event \mathcal{F}_ε ,

$$-B_n n s - \frac{1}{2} \sum_{j=2}^n \log \left(1 - \frac{4\mu_1^{(1)}s}{\mu_1 - \mu_j} \right) = -\frac{2s_{\text{MP}}(d_+)\lambda^{\frac{1}{2}}b\sqrt{\log n}}{\sqrt{1+\lambda}}y + O(1), \quad (4.49)$$

and we arrive at

$$\begin{aligned} A & = \frac{\text{i}e^{O(1)}}{n^{\frac{1}{3}}} \int_0^{n^{\frac{2}{3}}(\mu_1^{(2)} - a_+)} \exp \left(-\frac{2s_{\text{MP}}(d_+)\lambda^{\frac{1}{2}}b\sqrt{\log n}}{\sqrt{1+\lambda}}y \right) \frac{dy}{\sqrt{y}} \\ & = \begin{cases} \text{i}e^{O(1)}n^{-\frac{1}{3}}b^{-\frac{1}{2}}(\log n)^{-\frac{1}{4}}, & b > 0, \\ \text{i}e^{O(1)}n^{-\frac{1}{3}}, & b = 0. \end{cases} \end{aligned}$$

This completes the proof of the lemma. \square

Lemma 4.9. Let $\theta_n = \frac{n^{2/3}(\mu_1^{(2)} - \mu_2^{(2)})}{2}$. For $b \geq 0$ and for every $0 < r < n^{-1}$, on the event \mathcal{F}_ε ,

$$\left| \int_{\Gamma_3^+ \cup \Gamma_3^-} \exp \left[n(G(\mu_1^{(1)}, z) - \widehat{G}) \right] dz \right|$$

$$\leq n^{-\frac{1}{3}} \exp \left(-\frac{2s_{\text{MP}}(d_+) \sqrt{\lambda} \theta_n}{\sqrt{1+\lambda}} b \sqrt{\log n} + O(1) \right).$$

Proof. Since $G(\mu_1^{(1)}, \bar{z}) = \overline{G(\mu_1^{(1)}, z)}$ for all $z \in \mathbb{C}$, it suffices to bound the integral over Γ_3^+ . We define

$$G_+(\mu_1^{(1)}, a_+) = \lim_{t \downarrow 0} G(\mu_1^{(1)}, a_+ + it), \quad \tilde{G}(t) = G(\mu_1^{(1)}, a_+ + it) - G_+(\mu_1^{(1)}, a_+).$$

Then, for $z \in \Gamma_3^+$,

$$n(G(\mu_1^{(1)}, z) - \widehat{G}) = n(G_+(\mu_1^{(1)}, a_+) - \widehat{G}) + n\tilde{G}(t),$$

and we have

$$\left| \int_{\Gamma_3^+} \exp \left[n(G(\mu_1^{(1)}, z) - \widehat{G}) \right] dz \right| \leq \left| e^{n(G_+(\mu_1^{(1)}, a_+) - \widehat{G})} \right| \int_0^\infty e^{n \operatorname{Re} \tilde{G}(t)} dt. \quad (4.50)$$

For fixed $k > 2$,

$$\begin{aligned} n \operatorname{Re} \tilde{G}(t) &= -\frac{1}{4} \sum_{j=1}^n \log \left(1 + \left(\frac{4\mu_1^{(1)} t}{4\mu_1^{(1)} a_+ - \mu_j} \right)^2 \right) \\ &\leq -\frac{1}{4} \sum_{j=2}^n \log \left(1 + \left(\frac{4\mu_1^{(1)} t}{\mu_1 - \mu_j} \right)^2 \right) \leq -\frac{k}{4} \log \left(1 + \xi^{-2} n^{\frac{4}{3}} t^2 \right), \end{aligned}$$

where $\xi := \frac{n^{2/3}}{4\mu_1^{(1)}} |\mu_1 - \mu_{k+1}|$ is $O(1)$ on the event \mathcal{F}_ε . Thus,

$$\int_0^\infty e^{n \operatorname{Re} \tilde{G}(t)} dt \leq \int_0^\infty (1 + \xi^{-2} n^{\frac{4}{3}} t^2)^{-\frac{k}{4}} dt = \exp \left(-\frac{2}{3} \log n + O(1) \right). \quad (4.51)$$

At the same time, on the event \mathcal{F}_ε , $\theta_n = \Theta(n^{2/3}(\mu_1 - \mu_2)) = \Theta(1)$. Thus, similar to the proof of Lemma 4.8, we obtain that

$$\begin{aligned} n(G_+(\mu_1^{(1)}, a_+) - \widehat{G}) &= -B_n n^{\frac{1}{3}} \theta_n - \frac{1}{2} \sum_{j=2}^n \log \left(1 - \frac{4\mu_1^{(1)} \theta_n}{n^{\frac{2}{3}} (\mu_1 - \mu_j)} \right) \\ &\quad - \frac{1}{2} \log(4\mu_1^{(1)} n^{-2/3} \theta_n) - \frac{i\pi}{2} \\ &= -n^{\frac{1}{3}} \theta_n \left(B_n - 2\mu_1^{(1)} s_{\text{MP}}(d_+) - \omega_{1n} n^{-\frac{1}{3}} \right) \\ &\quad + (\zeta \theta_n)^2 \omega_{2n} + \frac{\log n}{3} + O(1) \\ &= -\frac{2\sqrt{\lambda} s_{\text{MP}}(d_+) \theta_n}{\sqrt{1+\lambda}} b \sqrt{\log n} + \frac{\log n}{3} + O(1), \end{aligned} \quad (4.52)$$

on the event \mathcal{F}_ε . Applying the above two displays to (4.50), we obtain the lemma. \square

4.2.2. Proof of (4.40) When $b = 0$. Observe that when $b = 0$, Lemmas 4.8 and 4.9 using keyhole contour show that, with probability $1 - \varepsilon$ for arbitrary small $\varepsilon > 0$, the contribution from the vertical and horizontal parts of the contour is both $n^{-\frac{1}{3}}e^{O(1)}$. This provides the upper bound for K_n . As some cancelation between the two contributions can occur, further analysis is required for the lower bound. In this section, we use the steepest descent contour of $G(\mu_1^{(1)}, z)$ crossing the real line above $\mu_1^{(2)}$ to obtain the needed lower bound

$$K_n \geq n^{-\frac{1}{3}}e^{O(\log \log n)}.$$

The argument is inspired by the one provided by Johnstone et al in [36].

Lemma 4.10. *There exists a unique saddle point of $G(\mu_1^{(1)}, z)$ on $z \in (\mu_1^{(2)}, \infty)$.*

Proof. Observe that

$$\partial_2 G(\mu_1^{(1)}, z) = B_n - \frac{1}{2n} \sum_{j=1}^n \frac{4\mu_1^{(1)}}{4\mu_1^{(1)}z - \mu_j}$$

is an increasing function of z on the interval $(\mu_1^{(2)}, \infty)$ and that

$$\lim_{z \downarrow \mu_1^{(2)}} \partial_2 G(\mu_1^{(1)}, z) = -\infty, \quad \lim_{z \rightarrow \infty} \partial_2 G(\mu_1^{(1)}, z) = B_n > 0.$$

Thus, there is a unique solution $z_c \in (\mu_1^{(2)}, \infty)$ to the equation $\partial_2 G(\mu_1^{(1)}, z) = 0$. Moreover, $\partial_2^2 G(\mu_1^{(1)}, z) > 0$ for all $z > \mu_1^{(2)}$. Thus, z_c is a saddle point of $\text{Re}[G(\mu_1^{(1)}, z)]$. \square

Let Γ_s be the steepest descent contour of $G(\mu_1^{(1)}, z)$ crossing z_c . For $z = x + iy \in \Gamma_s$,

$$0 = \text{Im}[G(\mu_1^{(1)}, z)] = B_n y - \frac{1}{2n} \sum_{j=1}^n \arg(4\mu_1^{(1)}x - \mu_j + i4\mu_1^{(1)}y),$$

which implies Γ_s is symmetric with respect to the x -axis. Moreover, for fixed $y > 0$, $\arg(4\mu_1^{(1)}x - \mu_j + i4\mu_1^{(1)}y)$ is strictly decreasing in x . This suggests there is at most one solution x to $\text{Im}[G(\mu_1^{(1)}, x + iy)] = 0$ for any $y > 0$. The same applies to $y < 0$ by symmetry. We then parameterize $\Gamma_s = \{\Gamma_s(t) : 0 < t < 1\}$ such that $\text{Im} \Gamma_s(t)$ is increasing in t .

As $B_n|y| \uparrow \frac{\pi}{2}$, $x \rightarrow -\infty$ so $\Gamma_s(0^+) = -\infty - i\frac{\pi}{2B_n}$ and $\Gamma_s(1^-) = -\infty + i\frac{\pi}{2B_n}$. We obtain $\text{Re} \Gamma_s(t)$ is bounded above, and K_n as in (4.42) satisfies

$$iK_n = \int_{\Gamma_s} \exp \left[n(G(\mu_1^{(1)}, z) - \widehat{G}) \right] dz.$$

We now consider points on the contour Γ_s with real part $\mu_1^{(2)}$.

Lemma 4.11. *The function*

$$f(y) := \text{Im}[G(\mu_1^{(1)}, \mu_1^{(2)} + iy)] = B_n y - \frac{\pi}{4n} - \frac{1}{2n} \sum_{j=2}^n \arctan \left(\frac{4\mu_1^{(1)}y}{\mu_1 - \mu_j} \right)$$

has a unique positive root y_0 . Furthermore, for any sequence $a_n \rightarrow \infty$, $a_n = O(n^\delta)$ for any $\delta > 0$,

$$n^{-2/3}a_n^{-1} \leq y_0 \leq n^{-2/3}a_n, \quad \text{asymptotically almost surely.} \quad (4.53)$$

Proof. Existence and uniqueness of $y_0 > 0$ follow from the fact that $f(y)$ is continuous, convex function on $[0, \infty)$ with $f(0) = -\frac{\pi}{4n}$ and $\lim_{y \rightarrow \infty} f(y) = \infty$.

Let y_-, y_+ denote the bounds $a_n^{-1}n^{-2/3}$ and $a_n n^{-2/3}$, respectively. We now verify (4.53) by showing that a.s., $f(y_-) < 0 < f(y_+)$. First, using $\arctan(x) \geq x - x^2/4$ for $x \geq 0$ and Lemma 4.1, then with probability $1 - \varepsilon$ for arbitrary $\varepsilon > 0$,

$$\begin{aligned} f(y_-) &\leq y_- \left(B_n - \frac{2\mu_1^{(1)}}{n} \sum_{j=2}^n \frac{1}{\mu_1 - \mu_j} \right) - \frac{\pi}{4n} + \frac{(4\mu_1^{(1)}y_-)^2}{8n} \sum_{j=2}^n \frac{1}{(\mu_1 - \mu_j)^2} \\ &= y_- \left(B_n - 2\mu_1^{(1)}s_{\text{MP}}(d_+) + O(n^{-1/3}) \right) - \frac{\pi}{4n} + y_-^2 \cdot O(n^{1/3}) \\ &= -\frac{\pi}{4n} + o(n^{-1}) < 0. \end{aligned}$$

In the last equality, $B_n - 2\mu_1^{(1)}s_{\text{MP}}(d_+) = O(n^{-1-\tau})$ due to rigidity of μ_1 and the fact $B_n = B_c + O(n^{-1-\tau})$ for any $\tau > 0$. The second part of the proof relies on the following statistics regarding the eigenvalues of a matrix from the Laguerre Orthogonal Ensemble. Let

$$\begin{aligned} j_0 &= \#\left\{ j : \mu_j > d_+ - \frac{1}{3}a_n n^{-2/3} \right\}, \\ j^* &= \#\left\{ j : \mu_j > \mu_1 - \left(1 + \frac{\pi}{2}\right)^{-1} a_n n^{-2/3} \right\}. \end{aligned}$$

By Chebyshev's inequality and (A.2), for some $c > 0$, it holds a.s. that $j_0 \geq ca_n^{3/2}$. Combine with the observation that a.s., $\mu_1 - d_+ = \Theta(n^{-2/3}) \ll a_n n^{-2/3}$, we obtain

$$j^* \geq j_0 \geq ca_n^{3/2} \quad \text{a.s.} \quad (4.54)$$

Since $\arctan(x) \leq x - 1$ for $x > 1 + \frac{\pi}{2}$ and $j^* = \max\{j : \frac{y_+}{\mu_1 - \mu_j} > 1 + \frac{\pi}{2}\}$, we have

$$\arctan\left(\frac{4\mu_1^{(1)}y_+}{\mu_1 - \mu_j}\right) \leq \frac{4\mu_1^{(1)}y_+}{\mu_1 - \mu_j} - \mathbb{1}_{\{j \leq j^*\}}.$$

Lemma 4.1 and the above display imply that a.s.,

$$\begin{aligned} f(y_+) &= B_n y_+ - \frac{1}{2n} \sum_{j=2}^n \arctan\left(\frac{4\mu_1^{(1)}y_+}{\mu_1 - \mu_j}\right) - \frac{\pi}{4n} \\ &\geq B_n y_+ - \frac{1}{2n} \sum_{j=2}^n \frac{4\mu_1^{(1)}y_+}{\mu_1 - \mu_j} + \frac{j^*}{2n} - \frac{\pi}{4n} \\ &\geq y_+ \cdot O(n^{-1/3}) + \frac{ca_n^{3/2}}{2n} - \frac{\pi}{4n}, \end{aligned}$$

which is strictly positive as $y_+ = a_n n^{-2/3}$. We obtain the lemma. \square

Let $z_0 = \mu_1^{(2)} + iy_0$, and consider the subset

$$\Gamma_0 = \{z \in \Gamma_s : |\operatorname{Im} z| \leq y_0\},$$

which is a connected curve with endpoints z_0, \bar{z}_0 by the parameterization. We have now obtained the needed tools to bound K_n as follows.

Observe that $G(\mu_1^{(1)}, z) - \widehat{G}$ is real on Γ_s and is monotone decreasing as z moves away from the point z_c along Γ_s . Also, $\frac{dy}{dt} > 0$ from the parameterization. Therefore,

$$\begin{aligned} K_n &= \frac{1}{i} \int_{\Gamma_s} \exp \left[n(G(\mu_1^{(1)}, z) - \widehat{G}) \right] dz \\ &\geq \int_{-y_0}^{y_0} \exp \left[n \operatorname{Re}(G(\mu_1^{(1)}, z(y)) - \widehat{G}) \right] dy \\ &\geq 2y_0 \exp \left[n \operatorname{Re}(G(\mu_1^{(1)}, z_0) - \widehat{G}) \right]. \end{aligned} \quad (4.55)$$

Here,

$$\begin{aligned} \log y_0 + n \operatorname{Re}(G(\mu_1^{(1)}, z_0) - \widehat{G}) &= \log y_0 - \frac{1}{2} \log(4\mu_1^{(1)} y_0) \\ &\quad - \frac{1}{4} \sum_{j=2}^n \log \left(1 + \frac{(4\mu_1^{(1)} y_0)^2}{(\mu_1 - \mu_j)^2} \right) \\ &\geq \frac{1}{2} \log y_0 - \frac{(4\mu_1^{(1)} y_0)^2}{4} \sum_{j=2}^n \frac{1}{(\mu_1 - \mu_j)^2} \\ &\geq -\frac{1}{3} \log n + O(\log \log n). \end{aligned} \quad (4.56)$$

The last inequality holds a.a.s., using Lemma 4.11 with $a_n^2 = \log \log n$ and the fact $\sum_{j=2}^n \frac{1}{(\mu_1 - \mu_j)^2}$ is $O(n^{4/3})$ under the event \mathcal{F}_ε . This completes the proof of the lower bound of K_n .

4.3. Low-Temperature Free Energy

Finally, using the contour integral computations from the previous section, we obtain the following lemma for the limiting fluctuations of the free energy on the low-temperature side of the critical temperature window.

Lemma 4.12. *If $\beta = \beta_c + bn^{-1/3}\sqrt{\log n}$ for some fixed $b \geq 0$, then the free energy satisfies*

$$\frac{m+n}{\sqrt{\frac{1}{6} \log n}} \left(F_{n,m}(\beta) - F(\beta) + \frac{1}{12} \frac{\log n}{n+m} \right) \rightarrow \mathcal{N}(0, 1) + \frac{\sqrt{6} \lambda^{\frac{1}{4}} b}{(1+\lambda)^{\frac{1}{2}} (1+\lambda^{\frac{1}{2}})^{\frac{2}{3}}} \operatorname{TW}_1,$$

where

$$F(\beta) = f_\lambda + \frac{\lambda}{1+\lambda} A(d_+, B) - \frac{1}{2} \log \beta - \frac{\lambda}{2(1+\lambda)} C_\lambda. \quad (4.57)$$

Proof. By (4.18),

$$\frac{1}{n+m} \log Q_n = \frac{n}{n+m} \widehat{G} + \frac{1}{n+m} \log S_n.$$

Note that $\frac{1}{n+m} \log S_n = -\frac{5}{6} \frac{\log n}{n+m} + O(n^{-1} \log \log n)$ by Lemma 4.5, while the quantity \widehat{G} is computed in Lemma 4.2. Combining them, we get

$$\begin{aligned} \frac{1}{n+m} \log Q_n &= \frac{\lambda}{1+\lambda} A(d_+, B) - \frac{7}{6} \frac{\log n}{n+m} \\ &\quad - \frac{1}{2(n+m)} \sum_{i=1}^n \log |d_+ - \mu_i| + \frac{\lambda^{\frac{3}{4}} b n^{-\frac{1}{3}} \sqrt{\log n}}{(1+\lambda)^{\frac{3}{2}} d_+} (\mu_1 - d_+) \\ &\quad + O\left(\frac{\log \log n}{n}\right). \end{aligned} \tag{4.58}$$

Apply this to (2.6), we obtain

$$\begin{aligned} F_{m,n}(\beta) &= f_\lambda + \frac{\lambda}{1+\lambda} A(d_+, B) - \frac{1}{2} \log \beta - \frac{1}{6} \frac{\log n}{n+m} \\ &\quad - \frac{1}{2(n+m)} \sum_{i=1}^n \log |d_+ - \mu_i| + \frac{\lambda^{\frac{3}{4}} b n^{-\frac{1}{3}} \sqrt{\log n}}{(1+\lambda)^{\frac{3}{2}} d_+} (\mu_1 - d_+) \\ &\quad + O\left(\frac{\log \log n}{n}\right). \end{aligned}$$

In terms of variables T_{1n} and T_{2n} as in (4.9), we get

$$\begin{aligned} F_{m,n}(\beta) &= f_\lambda + \frac{\lambda}{1+\lambda} A(d_+, B) - \frac{1}{2} \log \beta - \frac{\lambda}{2(1+\lambda)} C_\lambda - \frac{1}{12} \frac{\log n}{n+m} \\ &\quad + \frac{\sqrt{\frac{1}{6} \log n}}{n+m} \left(T_{1n} + \frac{\sqrt{6} \lambda^{\frac{1}{4}} b}{(1+\lambda)^{\frac{1}{2}} (1+\lambda^{\frac{1}{2}})^{\frac{2}{3}}} T_{2n} \right) + O\left(\frac{\log \log n}{n}\right). \end{aligned} \tag{4.59}$$

The theorem then follows since $T_{1n} \xrightarrow{d} \mathcal{N}(0, 1)$ by Theorem 1.2, and $T_{2n} \xrightarrow{d} \text{TW}_1$ by Lemma 2.1. \square

The fact that the Gaussian and Tracy–Widom limits are independent is shown in the next section.

5. Independence of Gaussian and Tracy–Widom Variables (Low Temperature)

Recall the quantities

$$T_{1n} := \frac{C_\lambda n - \frac{1}{6} \log n - \sum_{i=1}^n \log |d_+ - \mu_i|}{\sqrt{\frac{2}{3} \log n}}, \quad T_{2n} := \frac{n^{2/3} (\mu_1 - d_+)}{\sqrt{\lambda} (1 + \sqrt{\lambda})^{4/3}}, \tag{5.1}$$

$$C_\lambda = (1 - \lambda^{-1}) \log(1 + \lambda^{\frac{1}{2}}) + \log(\lambda^{\frac{1}{2}}) + \lambda^{-\frac{1}{2}}$$

The goal of this section is to show that, given an LOE matrix $M_{n,m}$ (which we assume without loss of generality to be in tridiagonal form), with probability arbitrarily close to one,

- $T_{1n} = \frac{Z_n}{\sqrt{\frac{2}{3} \log n}} + o(1)$ for Z_n depending only on the upper left minor of size $n - 2n^{1/3}(\log n)^3$ of the matrix $M_{n,m}$, and
- $T_{2n} = Y_n + o(1)$ for Y_n depending only on the lower right minor of size $2n^{1/3}(\log n)^3$ of the matrix.

Our proofs draw on ideas from the paper [36], which proves a similar result in the case of Wigner ensembles. We also make use of results from [24], which studies the asymptotics of the quantity $\sum_{i=1}^n \log |\gamma - \mu_i|$ for $\gamma \geq d_+$ by analyzing a recurrence on the determinants of the minors of $M_{n,m}$. In order to demonstrate the asymptotic independence of T_{1n} and T_{2n} , we need not only the main theorem of [24], but also many of the intermediate lemmas which involve recurrences on the matrix entries. For this purpose, we briefly summarize the setup from that paper along with the key notations that are used.

Recall from (2.20) that the tridiagonal representation of $M_{n,m}$ depends on χ -squared random variables $\{a_i^2\}$, $\{b_i^2\}$. Paper [24] works with centered and rescaled versions of these, denoted by α_i and β_i , respectively, which are defined as

$$\alpha_i = \frac{a_i^2 - (m - n + i)}{|\rho_i^+|}, \quad \beta_i = \frac{b_{i-1}^2 - (i - 1)}{|\rho_i^+|}. \quad (5.2)$$

Here, the scaling factor ρ_i^+ is one of the characteristic roots of the recurrence on determinants of the minors of $M_{n,m}$. This turns out to be a convenient rescaling since it prevents the iterates from blowing up. More precisely,

$$\rho_i^\pm := -\frac{1}{2}(\gamma m - (m - n + 2i - 1) \pm \sqrt{(\gamma m - (m - n + 2i - 1))^2 - 4(m - n + i - 1)(i - 1)}). \quad (5.3)$$

Throughout the proofs, we will also use the notations

$$\tau_i = \frac{m - n + i}{|\rho_i^+|}, \quad \delta_i = \frac{i - 1}{|\rho_i^+|}. \quad (5.4)$$

5.1. Proof for T_{1n}

Lemma 5.1. *There exists a random random Z_n , depending only on the upper left minor of size $n - 2n^{1/3}(\log n)^3$ of the matrix $M_{n,m}$ such that*

$$T_{1n} = \frac{Z_n}{\sqrt{\frac{2}{3} \log n}} + o(1).$$

Proof. We begin our analysis of T_{1n} by remarking that it is tricky to analyze the distribution of $\sum_{i=1}^n \log |d_+ - \mu_i|$ directly because of how close d_+ is to the eigenvalues $\{\mu_i\}$. For this reason, [24] uses the technique of first analyzing the sum $\sum_{i=1}^n \log |\gamma - \mu_i|$ for

$$\gamma = d_+ + \sigma_n n^{-2/3}, \quad (5.5)$$

then analyzing the original sum by comparison to the shifted one. We employ a similar technique here. More precisely, we take

$$\sigma_n = \bar{\sigma}_n := (\log \log n)^3. \quad (5.6)$$

From line (7.3) of [24], we have

$$\sum_{i=1}^n \log |d_+ - \mu_i| = \sum_{i=1}^n \log |d_+ + \bar{\sigma}_n n^{-2/3} - \mu_i| - C_1 \bar{\sigma}_n n^{1/3} + C_2 \bar{\sigma}_n^{3/2} + o(\sqrt{\log n}). \tag{5.7}$$

where

$$C_1 = \frac{1}{\lambda^{1/2}(1 + \lambda^{1/2})}, \quad C_2 = \frac{2}{3\lambda^{3/4}(1 + \lambda^{1/2})^2}. \tag{5.8}$$

Furthermore, from Lemma 3.1 and Section 4 of [24], we can rewrite the sum on the right-hand side of (5.7) as

$$\begin{aligned} \sum_{i=1}^n \log |d_+ + \bar{\sigma}_n n^{-2/3} - \mu_i| &= C_\lambda n - \sum_{i=3}^n L_i - \frac{1}{6} \log n + C_1 \bar{\sigma}_n n^{1/3} - C_2 \bar{\sigma}_n^{3/2} \\ &\quad 0 + o(\sqrt{\log n}) \end{aligned} \tag{5.9}$$

where C_1, C_2 are the same constants from (5.7) and L_i is given by the recursive formula

$$L_i := \xi_i + \omega_i L_{i-1} \text{ for } i \geq 4, \quad L_3 := \xi_3. \tag{5.10}$$

with

$$\xi_i := \alpha_i + \beta_i(1 + \tau_{i-1}) + \alpha_{i-1} \delta_i, \quad \omega_i := \tau_{i-1} \delta_i. \tag{5.11}$$

Thus, combining (5.7) and (5.9) with the definition of T_{1n} , we get

$$T_{1n} = \frac{\sum_{i=3}^n L_i}{\sqrt{\frac{2}{3} \log n}} + o(1). \tag{5.12}$$

It remains to show that $\sum_{i=3}^n L_i = Z_n + o(\sqrt{\log n})$ for some Z_n depending only on the upper left minor of $M_{n,m}$ of size $n - 2n^{1/3}(\log n)^3$. From the recursive definition of L_i , we have, for any $j \geq 4$,

$$\sum_{i=3}^n L_i = \sum_{i=3}^n \xi_i + \omega_i \xi_{i-1} + \cdots + \omega_i \cdots \omega_4 \xi_3 = \sum_{i=3}^n g_{i+1} \xi_i$$

where $g_i = 1 + \omega_i + \omega_i \omega_{i+1} + \cdots + \omega_i \cdots \omega_n$ for $3 \leq i \leq n$. Now we would like to compare this sum to a similar sum, truncated at index $i = n - 2n^{1/3}(\log n)^3$ and show that their difference is small, with probability arbitrarily close to 1. As this will involve computing the variance of the difference between the sums, we would like to eliminate the dependence between consecutive terms in the sum by rewriting

$$\sum_{i=3}^n L_i = \sum_{i=3}^n g_{i+1} X_i + \sum_{i=3}^n \alpha_i - g_3 \alpha_2.$$

where

$$X_i = (1 + \tau_{i-1})(\delta_i \alpha_{i-1} + \beta_i), \quad 3 \leq i \leq n. \tag{5.13}$$

Now we define

$$Z_n = \sum_{i=3}^{\lfloor n-2n^{1/3}(\log n)^3 \rfloor} g_{i+1} X_i. \quad (5.14)$$

This gives us

$$\sum_{i=3}^n L_i - Z_n = \sum_{i=\lceil n-2n^{1/3}(\log n)^3 \rceil}^n g_{i+1} X_i + \sum_{i=3}^n \alpha_i - g_3 \alpha_2. \quad (5.15)$$

It follows from line (5.21) of [24] that $\sum_{i=3}^n \alpha_i - g_3 \alpha_2 = o(\sqrt{\log n})$ with probability $1 - n^{-1/2}$. Finally, we bound the variance of the remaining sum on the right-hand side of (5.15). Since $\{X_i\}$ are pairwise independent and $\{g_i\}$ are deterministic, we have

$$\mathbb{E} \left[\left(\sum_{i=\lceil n-2n^{1/3}(\log n)^3 \rceil}^n g_{i+1} X_i \right)^2 \right] = \sum_{i=\lceil n-2n^{1/3}(\log n)^3 \rceil}^n g_{i+1}^2 \mathbb{E} X_i^2$$

From (4.42) of [24], we have $\mathbb{E} X_i^2 = O(n^{-1})$ uniformly in i . Combining Lemma 5.1 and Corollary 2.9 of [24], we have

$$g_i = \begin{cases} O(n^{1/2}(n-i)^{-1/2}) & i \leq n - n^{1/3} \sigma_n \\ O(n^{1/3} \sigma_n^{-1/2}) & i \geq n - n^{1/3} \sigma_n. \end{cases} \quad (5.16)$$

Thus, we can bound the sum as follows:

$$\begin{aligned} \sum_{i=\lceil n-2n^{1/3}(\log n)^3 \rceil}^n g_{i+1}^2 \mathbb{E} X_i^2 &\leq \sum_{i=\lceil n-2n^{1/3} \sigma_n(\log n)^3 \rceil}^{\lfloor n-n^{1/3} \sigma_n \rfloor} \frac{n}{n-i} \cdot \frac{C}{n} \\ &\quad + \sum_{i=\lceil n-n^{1/3} \sigma_n \rceil}^n \frac{n^{2/3}}{\sigma_n} \cdot \frac{C}{n} \\ &= O(\log \log n) + O(1). \end{aligned}$$

This completes the proof of the lemma concerning T_{1n} . \square

5.2. Proof for T_{2n}

We now verify that, $T_{2n} = Y_n + o(1)$, for some random variable Y_n depending only on the bottom-right minor of size $2n^{\frac{1}{3}}(\log n)^3$ of the matrix $M_{n,m}$ (in fact, we get a much tighter tail bound than $o(1)$). Recall that T_{2n} is a shifted rescaling of the largest eigenvalue μ_1 , and it converges to the Tracy–Widom distribution. Thus, Y_n , if it exists, must converge to the same limit, while only depending on the bottom corner of $M_{n,m}$. The following lemma shows that the largest eigenvalue of the minor described above, with the same transformation as in T_{2n} , is a good choice for Y_n .

Lemma 5.2. *Let $\tilde{\mu}_1$ be the largest eigenvalue of the bottom-right minor of $M_{n,m}$ of size $p > 2n^{\frac{1}{3}}(\log n)^3$. Then, for any $D > 0$ and $\varepsilon > 0$, with probability at*

least $1 - \varepsilon$,

$$|\mu_1 - \tilde{\mu}_1| = O(n^{-D}).$$

Furthermore, by setting $Y_n = \frac{n^{2/3}(\tilde{\mu}_1 - d_+)}{\sqrt{\lambda}(1+\sqrt{\lambda})^{4/3}}$ and taking $D > \frac{2}{3}$ arbitrarily large, we have

$$T_{2n} = Y_n + O(n^{-D+2/3}).$$

The key ingredient to bounding the difference $\mu_1 - \tilde{\mu}_1$ lies in controlling the first $n - 2n^{1/3}(\log n)^3$ components of an eigenvector corresponding to μ_1 . In particular, we need the following result.

Lemma 5.3. *If $\mathbf{v} = (v_1, \dots, v_n)^T$ is a principal eigenvector of $M_{n,m}$, then for any $\varepsilon > 0$ and $d > 0$, with probability at least $1 - \varepsilon$, we have*

$$\max_{j \leq n - 2n^{\frac{1}{3}}(\log n)^3} \frac{|v_j|}{\|\mathbf{v}\|} < n^{-d}.$$

Lemma 5.3 itself relies on the following two auxiliary Lemmas 5.4 and 5.5, both of which depend on the random entries in the tridiagonal matrix form. We include their proofs in Appendix B.

Lemma 5.4. *Let μ_1 be the largest eigenvalue of $M_{n,m}$. Let $\{F_j\}_{j=1}^{n-1}$ be the sequence given by*

$$\begin{aligned} F_1 &= -1 + \frac{\mu_1 m - a_1^2}{|\rho_1^+|}, & F_j &= -1 + \frac{\mu_1 m - (a_j^2 + b_{j-1}^2)}{|\rho_j^+|} \\ &+ \frac{(a_{j-1} b_{j-1})^2}{|\rho_j^+| |\rho_{j-1}^+|} \cdot \frac{1}{1 + F_{j-1}} \text{ for } j = 2, \dots, n-1. \end{aligned}$$

Here, ρ_j^+ is given by (5.3) with $\gamma = d_+$. Then, for every $\varepsilon > 0$, with probability at least $1 - \varepsilon$,

$$\max_{j \leq n - n^{\frac{1}{3}}(\log n)^3} |F_j| = o(n^{-\frac{1}{3}}). \quad (5.17)$$

Lemma 5.5. *Given $\varepsilon > 0$, then for sufficiently large n and a_i, b_i as defined in (2.21), we have*

$$\mathbb{P} \left(\left| \max_{j \leq n - n^{1/3}(\log n)^3} a_j b_j - \sqrt{mn} \right| \leq (\varepsilon \log n)^2 n^{1/2} \right) \geq 1 - \varepsilon. \quad (5.18)$$

Proof of Lemma 5.3. From the tridiagonal representation (2.20) and the notations presented at the beginning of Sect. 5, we obtain the system of linear equations

$$\begin{cases} \left(\frac{a_1^2}{m} - \mu_1 \right) v_1 + \frac{a_1 b_1}{m} v_2 = 0, \\ \frac{a_{j-1} b_{j-1}}{m} v_{j-1} + \left(\frac{a_j^2 + b_{j-1}^2}{m} - \mu_1 \right) v_j + \frac{a_j b_j}{m} v_{j+1} = 0, & j = 2, \dots, n-1. \end{cases}$$

With probability 1, $a_j > 0$ and $b_j > 0$ for $j = 1, \dots, n-1$. This implies $v_1 \neq 0$ (otherwise, \mathbf{v} is the zero vector). In fact, as functions of positive, continuous

random variables $a_1, \dots, a_{j-1}, b_1, \dots, b_{j-1}$, it holds with probability 1 that $v_j \neq 0$ for each j . Thus, we rescale \mathbf{v} to have $v_1 = 1$ and obtain

$$v_2 = \frac{\mu_1 m - a_1^2}{a_1 b_1}, \quad v_{j+1} = \frac{\mu_1 m - (a_j^2 + b_{j-1}^2)}{a_j b_j} v_j - \frac{a_{j-1} b_{j-1}}{a_j b_j} v_{j-1},$$

$$j = 2, \dots, n-1. \quad (5.19)$$

We introduce the following quantity

$$F_j = \frac{v_{j+1}}{v_j} \cdot \frac{a_j b_j}{|\rho_j^+|} - 1, \quad \text{for } j = 1, \dots, n-1. \quad (5.20)$$

Here, ρ_j^+ is given in (5.3) with $\gamma = d_+$. Set $k = \lceil n^{\frac{1}{3}} \rceil$, and let $j \leq n - 2n^{1/3}(\log n)^3$. Observe that

$$\frac{|v_j|}{\|\mathbf{v}\|} \leq \left| \frac{v_j}{v_{j+k}} \right| = \prod_{l=j}^{j+k-1} (1 + F_l)^{-1} \prod_{l=j}^{j+k-1} \frac{(a_l b_l)/m}{|\rho_l^+|/m}. \quad (5.21)$$

Since $\{F_l\}_{l=1}^{n-1}$ satisfies the hypothesis of Lemma 5.4 and each $l \in [j, j+k-1]$ satisfies $l \leq n - n^{\frac{1}{3}}(\log n)^3$, it follows that, with probability $1 - \varepsilon/2$, we have $\prod_{l=j}^{j+k-1} (1 + F_l)^{-1} = 1 + o(1)$.

We then consider the product $\prod_{l=j}^{j+k-1} \frac{(a_l b_l)/m}{|\rho_l^+|/m}$. As $|\rho_l^+|$ is decreasing in l by (5.3),

$$\begin{aligned} \frac{|\rho_l^+|}{m} &\geq \frac{|\rho_{n-n^{1/3}(\log n)^3}^+|}{m} \\ &= \frac{d_+ m - (m + n - 2n^{1/3}(\log n)^3 - 1)}{2m} \\ &\quad \left(1 + \sqrt{1 - \frac{4(m - n^{1/3}(\log n)^3 - 1)(n - n^{1/3}(\log n)^3 - 1)}{(d_+ m - (m + n - 2n^{1/3}(\log n)^3 - 1))^2}} \right). \end{aligned} \quad (5.22)$$

Using $d_+ m = m + n + 2\sqrt{mn}$, the first factor on the right-hand side of (5.22) is $\sqrt{\lambda}(1 + O(n^{-\frac{2}{3}}(\log n)^3))$, while the expression under the square root is $\Theta(n^{-\frac{2}{3}}(\log n)^3)$. Therefore, there is a constant $c > 0$ such that

$$\frac{|\rho_l^+|}{m} \geq \sqrt{\lambda} + cn^{-\frac{1}{3}}(\log n)^{\frac{3}{2}} \quad \text{for all } l \leq n - n^{\frac{1}{3}}(\log n)^3.$$

Combining this with Lemma 5.5, we obtain that, for some $c' > 0$, with probability $1 - \varepsilon/2$,

$$\prod_{l=j}^{j+k-1} \frac{(a_l b_l)/m}{|\rho_l^+|/m} \leq (1 - c'n^{-\frac{1}{3}}(\log n)^{\frac{3}{2}} + o(n^{-\frac{1}{3}}))^k. \quad (5.23)$$

Therefore, with probability $1 - \varepsilon$,

$$\max_{j \leq n - 2n^{\frac{1}{3}}(\log n)^3} \frac{|v_j|}{\|\mathbf{v}\|} \leq \max_{j \leq n - 2n^{\frac{1}{3}}(\log n)^3} \left| \frac{v_j}{v_{j+k}} \right| = \exp\left(-c'(\log n)^{3/2} + o(1)\right).$$

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

6. Appendix: Section 2 Proofs

In appendix, we provide a proof for Lemma 2.2, and then apply it to prove Lemma 2.8.

6.1. Proof of Lemma 2.2

Before beginning the main proof, we need the following preliminary results.

Lemma A.1. *Let $\{\mu_j\}_{j=1}^n$ be eigenvalues of a scaled LUE or LOE matrix $\frac{1}{m}M_{n,m}$. Assume s is such that $s > C$ for some $C > 0$ and $s = o(n^{2/3})$ as $n \rightarrow \infty$. The following statements hold for $\mathcal{N}_s := \#\{i : \mu_i \in [d_+ - sn^{-2/3}, \infty)\}$.*

$$\mathbb{E}\mathcal{N}_s = \frac{2}{3\pi\lambda^{3/4}d_+}s^{3/2} + O(s^{5/2}n^{-2/3}). \quad (\text{A.1})$$

$$\text{Var}(\mathcal{N}_s) = \frac{3}{4\pi^2} \log(s)(1 + o(1)). \quad (\text{A.2})$$

The lemma is the analog of Proposition 6.5 from [40], which bounds the expectation and variance of the counting function in the case of GOE matrices. There, the result was obtained by applying the corresponding result for GUE matrices by Gustavsson [33], and the relation between eigenvalues of Gaussian orthogonal and unitary ensembles in [30]. The proof of [40] works in our case, up to translating from Gaussian to Laguerre Ensembles. For completeness, we reproduce it here, first proving for LUE matrices using a result in [51], and then extend to LOE matrices using the following result.

Theorem A.2 (Theorem 5.2 of [30]). *For independent eigenvalue point processes $LOE_{n,m}$, $LOE_{n+1,m+1}$,*

$$\text{even}(LOE_{n,m} \cup LOE_{n+1,m+1}) = LUE_{n,m},$$

where the notation $\text{even}(\cdot)$ denotes the set containing only the even numbered elements among the ordered list of elements in the original set.

Proof of Lemma A.1. In the case of LUE matrix, the lemma follows from the results of Su in [51]. Namely, the first inequality holds by Lemma 1 of [51], which states that

$$\mathbb{E}\#\{j : \mu_j \in [t_n, \infty)\} = n \int_{t_n}^{\beta_{n,m}} p_{\text{MP}}(x) dx$$

$$= \frac{\sqrt{\beta_{n,m} - \alpha_{n,m}}}{3\pi\beta_{n,m}} n(\beta_{n,m} - t_n)^{3/2} + O(n(\beta_{n,m} - t_n)^{5/2}).$$

As the matrix in [51] is scaled by $1/n$ instead of $1/m$ as in this paper, our interval of interest $[d_+ - sn^{-2/3}, \infty)$ corresponds to $t_n = \beta_{n,m} - \frac{s}{\lambda}n^{-2/3}$ in [51]. Meanwhile, the inequality for variance directly follows from Lemma 4 there.

We now consider the case of LOE matrix. Let $M_{n,m}^{(1)}, M_{n+1,m+1}^{(1)}$ be independent LOE matrices of size $n \times m$ and $(n + 1) \times (m + 1)$, respectively, and let $M_{n,m}^{(2)}$ be a LUE matrix of size $n \times m$. Set $X_{n,m}^{(2)}$ to be the number of eigenvalues of $M_{n,m}^{(2)}$ that are at least $m(d_+ - sn^{-\frac{2}{3}})$. We define $X_{n,m}^{(1)}$ and $X_{n+1,m+1}^{(1)}$ similarly, for the two LOE matrices. Theorem A.2 implies that there is a random variable Y and a random variable $Z \in [0, 1]$ such that

$$X_{n,m}^{(2)} \stackrel{d}{=} Y, \quad Y - Z = \frac{1}{2} \left(X_{n,m}^{(1)} + X_{n+1,m+1}^{(1)} \right).$$

The estimates (A.1) and (A.2) hold for Y by the previous paragraph. The estimate (A.2) for Y , together with boundedness of Z and the fact $X_{n,m}^{(1)}$ and $X_{n+1,m+1}^{(1)}$ are independent implies that (A.2) holds for the $X^{(1)}$'s as well. Now,

$$\mathbb{E}[X_{n,m}^{(2)}] = \frac{1}{2} \left(\mathbb{E}[X_{n,m}^{(1)}] + \mathbb{E}[X_{n+1,m+1}^{(1)}] \right) + c, \quad \text{for some } c \in [0, 1]. \quad (\text{A.3})$$

From the tridiagonal form of Laguerre Ensembles, the top left $n \times n$ minor of $M_{n+1,m+1}^{(1)}$ has the same distribution as $M_{n,m}^{(1)}$. The eigenvalues of this minor interlace those of $M_{n+1,m+1}^{(1)}$, which implies there is a random variable $\tilde{X}_{n,m}^{(1)}$ with the same distribution as $X_{n,m}^{(1)}$ and satisfies

$$|\tilde{X}_{n,m}^{(1)} - X_{n+1,m+1}^{(1)}| \leq 1.$$

We then obtain (A.1) for $X_{n,m}^{(1)}$ and $X_{n+1,m+1}^{(1)}$, using (A.1) for $X_{n,m}^{(2)}$, (A.3) and the above inequality. \square

We now have the needed tools to prove Lemma 2.2.

Proof of Lemma 2.2. For $j = 1, \dots, n^{2/5}$ and $t > 0$, by definition,

$$\mathbb{P}(A_j \geq t) = \mathbb{P}\left(\mu_j \geq d_+ - \left((C^*j)^{2/3} - t\right)n^{-2/3}\right) = \mathbb{P}(\mathcal{N}_T \geq j), \quad (\text{A.4})$$

where $C^* = \frac{3}{2}\pi\lambda^{3/4}d_+$ and $T = T(j, t) := (C^*j)^{2/3} - t$. If $\mathcal{N}_T < j$, then

$$\mathbb{P}(\mathcal{N}_T \geq j) \leq \mathbb{P}(|\mathcal{N}_T - \mathbb{E}\mathcal{N}_T| \geq j - \mathbb{E}\mathcal{N}_T) \leq \frac{\text{Var } \mathcal{N}_T}{(j - \mathbb{E}\mathcal{N}_T)^2}. \quad (\text{A.5})$$

In order to make use of this inequality, we need to know what values of t (depending on j) satisfy $\mathbb{E}\mathcal{N}_T < j$. By Lemma A.1, there exist $K, c_0 > 0$

such that, for any $c_1 > 0$ and any sufficiently large n , if $K \leq j \leq n^{2/5}$ and $0 < t < (C^*j)^{2/3} - c_1$, then

$$\begin{aligned} j - \mathbb{E}\mathcal{N}_T &\geq j - \frac{1}{C^*}((C^*j)^{2/3} - t)^{3/2} - c_0j^{5/3}n^{-2/3} \\ &\geq j - j \left(1 - \frac{t}{(C^*j)^{2/3}}\right)^{3/2} - c_0 \geq \frac{tj^{1/3}}{(C^*)^{2/3}} - c_0. \end{aligned} \quad (\text{A.6})$$

In particular, this means that $\mathbb{E}\mathcal{N}_T < j$ is satisfied (along with the conditions of Lemma A.1) when $c_0(C^*)^{2/3}j^{-1/3} < t < (C^*j)^{2/3} - c_1$ and $K \leq j \leq n^{2/5}$ (note that one should choose $K > c_0$). Thus, for t, j satisfying these conditions, we combine (A.4)–(A.6) with the variance bound from Lemma A.1 to conclude that, for some $c_2 > 0$ and sufficiently large n ,

$$\mathbb{P}(A_j \geq t) \leq \frac{c_2 \log j}{((C^*)^{-2/3}tj^{1/3} - c_0)^2}. \quad (\text{A.7})$$

Next, taking $T' = (C^*j)^{2/3} + t$ we can follow the same argument to bound $\mathbb{P}(A_j < -t)$. This time, we find that $\mathbb{E}\mathcal{N}_{T'} \geq j$ is satisfied (along with the conditions of Lemma A.1) when $c_0(C^*)^{2/3}j^{-1/3} < t \ll n^{2/3}$ and $K \leq j \leq n^{2/5}$. Then, for t, j satisfying these conditions, and for some $c_3 > 0$ with sufficiently large n ,

$$\begin{aligned} \mathbb{P}(A_j \leq -t) &= \mathbb{P}(\mu_j < d_+ - T'n^{-2/3}) \leq \mathbb{P}(|\mathcal{N}_{T'} - \mathbb{E}\mathcal{N}_{T'}| > \mathbb{E}\mathcal{N}_{T'} - j) \\ &\leq c_3 \frac{\log j + \log(1+t)}{((C^*)^{-2/3}tj^{1/3} - c_0)^2}. \end{aligned} \quad (\text{A.8})$$

Thus, for j, t satisfying $K \leq j \leq n^{2/5}$ and $c_0(C^*)^{2/3}j^{-1/3} < t < (C^*j)^{2/3} - c_1$, we have

$$\mathbb{P}(|A_j| \geq t) = O\left(\frac{\log j + \log(1+t)}{((C^*)^{-2/3}tj^{1/3} - c_0)^2}\right).$$

Taking $t = \lambda j^{2/3}$, then for all $k \geq K$,

$$\mathbb{P}\left(\bigcup_{k \leq j \leq n^{2/5}} \{|A_j| \geq \lambda j^{2/3}\}\right) = O\left(\sum_{j=k}^{n^{2/5}} \frac{\log j}{j^2}\right) = O\left(\frac{\log k}{k}\right).$$

This bound holds uniformly for $K \leq k \leq n^{2/5}$. Taking $k \rightarrow \infty$ (for example $k = n^{1/5}$), we obtain (2.18).

It remains to prove the second part of the lemma. Set $t^* = c_0(C^*)^{2/3}j^{-1/3}$. For $K \leq j \leq n^{2/5}$, we have

$$\begin{aligned} \mathbb{E}\left[\mathbb{1}_{\{n^{2/3}(\mu_j - d_+) \leq -C\}} |A_j|\right] &\leq \int_0^\infty \mathbb{P}(A_j \geq t) dt + \int_0^\infty \mathbb{P}(-A_j \leq -t) dt \\ &\leq \left(t^* + \int_{t^*}^{(C^*j)^{2/3} - C} \mathbb{P}(A_j > t) dt + 0\right) \\ &\quad + \left(t^* + \int_{t^*}^{n^{\frac{2}{3}-\delta}} \mathbb{P}(-A_j \leq -t) dt + o(n^{-1})\right) \end{aligned}$$

$$\begin{aligned} &\leq 2t^* + C' \int_{t^*}^{\infty} \frac{\log j + \log(1+t)}{((C^*)^{-2/3} t j^{1/3} - c_0)^2} dt \\ &\leq 2t^* + C'' \frac{\log j}{j^{1/3}} = O\left(\frac{\log j}{j^{1/3}}\right), \end{aligned}$$

where, in the second line, we obtained $\int_{(C^* j)^{2/3} - C}^{\infty} \mathbb{P}(A_j \geq t) dt = 0$ from the indicator in the expectation, and $\int_{n^{\frac{2}{3}-\delta}}^{\infty} \mathbb{P}(-A_j \leq -t) dt = o(n^{-1})$ from eigenvalue rigidity. \square

6.2. Proof of Lemma 2.8

We observe that

$$\begin{aligned} S_2 &:= \frac{1}{n} \sum_{i=K+1}^n \frac{1}{(z - \mu_j)^l} - \int_{d_-}^{g_K} \frac{1}{(z - y)^l} p_{\text{MP}}(y) dy \\ &= \sum_{i>K} \int_{g_i}^{g_{i-1}} \frac{(z - y)^l - (z - \mu_i)^l}{(z - \mu_i)^l (z - y)^l} p_{\text{MP}}(y) dy. \end{aligned} \tag{A.9}$$

The modulus of this sum satisfies

$$\begin{aligned} |S_2| &\leq \sum_{i>K} \int_{g_i}^{g_{i-1}} \frac{l \max\{|z - y|, |z - \mu_i|\}^{l-1} |\mu_i - y|}{|z - \mu_i|^l |z - y|^l} p_{\text{MP}}(y) dy \\ &\leq \sum_{i>K} \int_{g_i}^{g_{i-1}} \frac{l \cdot |(\mu_i - g_i) + (g_i - y)|}{\min\{|z - \mu_i|, |z - y|\}^{l+1}} p_{\text{MP}}(y) dy. \end{aligned}$$

We now split the sum as $S_{21} + S_{22}$, summing over $K \leq i \leq n^{2/5}$ and $i > n^{2/5}$, respectively. First, consider $K \leq i \leq n^{2/5}$. By Lemma 2.2, given $\varepsilon > 0$, on the event \mathcal{E}_ε , there exists $c > 0$ such that, for sufficiently large n , $n^{2/3}(d_+ - \mu_i) \geq ci^{2/3}$ uniformly for all i in this range. Combining with the facts that $\text{Re } z \geq d_+$ and $d_+ \geq \mu_i$ for $i \geq K$ on \mathcal{E}_ε , we have

$$n^{2/3}|z - \mu_i| \geq \max\{n^{2/3}|z - d_+|, ci^{2/3}\}. \tag{A.10}$$

Meanwhile, there exists $C > 0$, independent of n , such that $C^{-1}i^{2/3} \leq n^{2/3}(d_+ - g_i) \leq Ci^{2/3}$ for all i (see, for example, [12]). Thus, (A.10) also holds for $n^{2/3}|z - y|$, uniformly for $y \in (g_i, g_{i-1})$. For the numerator, we have $n^{2/3}(y - g_i) \leq n^{2/3}(g_{i-1} - g_i) \leq ci^{-1/3}$, using

$$\frac{1}{n} = \int_{g_i}^{g_{i-1}} p_{\text{MP}}(y) dy \geq c\sqrt{d_+ - g_i}(g_{i-1} - g_i).$$

By (2.16), $n^{2/3}(\mu_i - g_i) = A_i + O\left(\frac{i^{4/3}}{n^{2/3}}\right)$, where A_i is given in (2.17). The term $i^{-1/3}$ is of larger order than $n^{-2/3}i^{4/3}$ when $K \leq i \leq n^{2/5}$, and they have the same order when $i = \Theta(n^{2/5})$. Thus,

$$\begin{aligned} \mathbb{1}_{\mathcal{E}_\varepsilon} \frac{l \cdot |(\mu_i - g_i) + (g_i - y)|}{\min\{|z - \mu_i|, |z - y|\}^{l+1}} &\leq Cln^{\frac{2}{3}}l \frac{i^{-1/3} + |A_i|}{i^{\frac{2}{3}(l+1)} + (n^{\frac{2}{3}}|z - d_+|)^{l+1}}, \\ &K \leq i \leq n^{2/5}. \end{aligned} \tag{A.11}$$

By Lemma 2.2,

$$\begin{aligned} \mathbb{E} [\mathbb{1}_{\mathcal{E}_\varepsilon} |S_{21}|] &\leq C l n^{\frac{2}{3}l-1} \sum_{K \leq i \leq n^{2/5}} \frac{i^{-1/3} + \mathbb{E} [\mathbb{1}_{\mathcal{E}_\varepsilon} |A_i|]}{i^{\frac{2}{3}(l+1)} + (n^{\frac{2}{3}}|z-d_+|)^{l+1}} \\ &\leq C' l n^{\frac{2}{3}l-1} \sum_{K \leq i \leq n^{2/5}} \frac{i^{-1/3} \log i}{i^{\frac{2}{3}(l+1)} + (n^{\frac{2}{3}}|z-d_+|)^{l+1}}. \end{aligned} \quad (\text{A.12})$$

Next, we consider two separate cases and conclude that, for some $C'' > 0$,

$$\mathbb{E} [\mathbb{1}_{\mathcal{E}_\varepsilon} |S_{21}|] \leq \begin{cases} C'' n^{\frac{2}{3}l-1} \frac{\log(n^{2/3}|z-d_+|)}{(n^{2/3}|z-d_+|)^l} & K^{2/3} < n^{2/3}|z-d_+|, \\ C'' n^{\frac{2}{3}l-1} & K^{2/3} \geq n^{2/3}|z-d_+|. \end{cases}$$

The bound in the first case is obtained by evaluating the right-hand side of (A.12) separately for $i^{2/3} < n^{2/3}|z-d_+|$ and $i^{2/3} > n^{2/3}|z-d_+|$. The bound in the second case follows from the convergence of $\sum_{i=K}^{\infty} i^{-\frac{2}{3}l-1} \log i$ for all $l \geq 1$. Thus, we obtain

$$\mathbb{E} [\mathbb{1}_{\mathcal{E}_\varepsilon} |S_{21}|] = O \left(n^{\frac{2}{3}l-1} \cdot \min \left\{ \left| \frac{\log(n^{2/3}|z-d_+|)}{(n^{2/3}|z-d_+|)^l} \right|, 1 \right\} \right). \quad (\text{A.13})$$

Lastly, for S_{22} , we bound the numerator (which is $l \cdot |\mu_i - y|$) using rigidity and bound $n^{2/3}|z-y| \geq ci^{2/3}$ by (A.10), while $|z-\mu_i| \geq \max\{|z-d_+|, d_+-\mu_i\}$, where, with high probability,

$$d_+ - \mu_i \geq \begin{cases} c > 0, & i > n/2, \\ ci^{2/3}n^{-2/3}, & n^{2/5} < i < n/2, \text{ using rigidity with } \delta < \frac{2}{15} \text{ and} \\ (2.16). \end{cases}$$

We obtain

$$\begin{aligned} \mathbb{1}_{\mathcal{E}_\varepsilon} |S_{22}| &\leq C l n^{\frac{2}{3}l-1+\delta} \sum_{i > n^{2/5}} \frac{1}{i^{\frac{2}{3}(l+1)} \min\{i^{1/3}, (n+1-i)^{1/3}\}} \\ &\leq C' l n^{\frac{2}{3}l-1+\delta} \left(\sum_{i=n^{2/5}}^{n/2} i^{-\frac{2}{3}l-1} + \sum_{i > n/2} \frac{1}{n^{\frac{2}{3}(l+1)}(n+1-i)^{1/3}} \right) \\ &= O(n^{\frac{2}{3}l-1} \cdot n^{-4l/15+\delta}), \end{aligned} \quad (\text{A.14})$$

which is $o(\mathbb{E} [\mathbb{1}_{\mathcal{E}_\varepsilon} |S_{21}|])$, provided $\delta < 4l/15$. This completes our proof of Lemma 2.8.

7. Appendix: Section 5 Proofs

In this section, we provide our proofs of Lemmas 5.4 and 5.5. The proof of Lemma 5.4 requires asymptotic bounds on $|\rho_j^\pm|$ when $\gamma = d_+$ and a few related quantities, which we state in the following two lemmas. Similar results were developed for the case $\gamma > d_+$ in Lemmas 2.7 and 2.8 in [24].

Lemma B.1. *The following asymptotic bounds hold, uniformly in $i \geq 2$ (where i can be fixed or n -dependent):*

- (i) $|\rho_i^+| = \Theta(n)$, $|\rho_i^-| = O(n)$,
- (ii) $|\rho_i^+| - |\rho_i^-| = \Theta(n^{1/2}(n-i+1)^{1/2})$,
- (iii) $|\rho_i^-| - |\rho_{i-1}^-| = O((\frac{n}{n-i+1})^{1/2})$ and $|\rho_{i-1}^+| - |\rho_i^+| = O((\frac{n}{n-i+1})^{1/2})$.

Proof. To show (i) for $|\rho_i^-|$, observe that $|\rho_i^-|$ is increasing in i , and

$$|\rho_n^-| = \frac{1}{2} \left(2\sqrt{mn} + 1 - 2\sqrt{\sqrt{mn} + n + m - \frac{3}{4}} + O(1) \right) = O(n).$$

Similarly, part (i) for $|\rho_i^+|$ holds since $|\rho_i^+|$ is decreasing in i , $|\rho_2^+| < 2\sqrt{mn} + 2n = \Theta(n)$, and

$$|\rho_n^+| > \frac{1}{2} (d_+m - (m+n-1)) = \frac{1}{2} (2\sqrt{mn} + 1) = \Theta(n).$$

For part (ii), we have

$$\begin{aligned} & |\rho_i^+| - |\rho_i^-| \\ &= \sqrt{(2\sqrt{mn} - 1 + 2(n-i+1))^2 - 4(m - (n-i+1))(n - (n-i+1))} \\ &= 2\sqrt{\sqrt{mn} + 1 + (m+n+2\sqrt{mn}-1)(n-i+1)} = \Theta(n^{1/2}(n-i+1)^{1/2}). \end{aligned}$$

Next, we verify (iii) by showing that $|\rho_i^-| - |\rho_{i-1}^-| + |\rho_{i-1}^+| - |\rho_i^+| = O((\frac{n}{n-i+1})^{1/2})$. Indeed, the left-hand side can be written as

$$(|\rho_{i-1}^+| - |\rho_{i-1}^-|) - (|\rho_i^+| - |\rho_i^-|) = \frac{(|\rho_{i-1}^+| - |\rho_{i-1}^-|)^2 - (|\rho_i^+| - |\rho_i^-|)^2}{|\rho_{i-1}^+| - |\rho_{i-1}^-| + |\rho_i^+| - |\rho_i^-|},$$

where numerator of the last ratio simplifies to $4d_+m - 4 = \Theta(n)$ and the denominator is $\Theta(n^{1/2}(n-i+1)^{1/2})$ by part (ii). \square

Lemma B.2. *There exist constants $0 < C_1 < C_2$ such that, for sufficiently large n , and $2 \leq i \leq n$,*

$$C_1 \left(\frac{n-i+1}{n} \right)^{1/2} \leq 1 - \omega_i \leq C_2 \left(\frac{n-i+1}{n} \right)^{1/2}.$$

Proof. We recall that $\omega_i = \frac{|\rho_i^-|}{|\rho_{i-1}^+|}$. Using the bounds $\frac{|\rho_{i-1}^-|}{|\rho_{i-1}^+|} < \omega_i < \frac{|\rho_i^-|}{|\rho_i^+|}$ we obtain

$$\frac{|\rho_i^+| - |\rho_i^-|}{|\rho_i^+|} < 1 - \omega_i < \frac{|\rho_{i-1}^+| - |\rho_{i-1}^-|}{|\rho_{i-1}^+|}.$$

Using Lemma B.1, the left and right sides of this inequality are both $\Theta((\frac{n-i+1}{n})^{1/2})$, uniformly in i , which gives the desired bounds. \square

7.1. Proof of Lemma 5.4

Using $\frac{1}{1+F_{i-1}} = 1 - \frac{F_{i-1}}{1+F_{i-1}} = 1 - F_{i-1} + \frac{F_{i-1}^2}{1+F_{i-1}}$ and the notations in (5.2) and (5.4), we have $F_1 = \frac{\mu_1 - d_+}{|\rho_1^+|/m} - \alpha_1$, and for $j = 2, \dots, n-1$,

$$F_j = -1 + \frac{\mu_1}{|\rho_j^+|/m} - (\alpha_j + \beta_j + \tau_j + \delta_j) - (\alpha_{j-1} + \tau_{j-1})(\beta_j + \delta_j) \left(1 - F_{j-1} + \frac{F_{j-1}^2}{1+F_{j-1}}\right).$$

As $1 + \tau_j + \delta_j = \frac{d_+ + m}{|\rho_j^+|} - \frac{|\rho_j^-|}{|\rho_j^+|}$, we re-arrange the terms to have

$$F_j = \eta_j - \xi_j + \omega_j F_{j-1} + \phi_j, \quad (\text{B.1})$$

where we define

$$\eta_j = \frac{\mu_1 - d_+}{|\rho_j^+|/m}, \quad (\text{B.2})$$

$$\phi_j = -\omega_j + \frac{|\rho_j^-|}{|\rho_j^+|} - \alpha_{j-1}\beta_j + (\alpha_{j-1}\beta_j + \alpha_{j-1}\delta_j + \tau_{j-1}\beta_j) \frac{F_{j-1}}{1+F_{j-1}} - \omega_j \frac{F_{j-1}^2}{1+F_{j-1}}, \quad (\text{B.3})$$

and ξ_j is given in (5.11). Note that, by Lemma B.1,

$$0 < \omega_j - \frac{|\rho_j^-|}{|\rho_j^+|} = \frac{|\rho_j^-|}{|\rho_j^+|} \frac{|\rho_{j-1}^+| - |\rho_j^+|}{|\rho_{j-1}^+|} = O(n^{-\frac{1}{2}}(n-j+1)^{-\frac{1}{2}}). \quad (\text{B.4})$$

Expanding the recurrence iteratively, we get

$$\begin{aligned} F_j &= \omega_j \dots \omega_2 F_1 + (\eta_j + \omega_j \eta_{j-1} + \dots + \omega_j \dots \omega_3 \eta_2) \\ &\quad - (\xi_j + \omega_j \xi_{j-1} + \dots + \omega_j \dots \omega_3 \xi_2) \\ &\quad + (\phi_j + \omega_j \phi_{j-1} + \dots + \omega_j \dots \omega_3 \phi_2). \end{aligned} \quad (\text{B.5})$$

On the event $\mathcal{F}_{s,t}^{(3)}$, which holds with probability $1 - \varepsilon/6$ for some s, t depending on ε , $|\mu_1 - d_+| \leq tn^{-\frac{2}{3}}$. As $|\rho_i^+|/m = \Theta(1)$ for all $i \leq n$, we obtain

$$\max_{j \leq n} |\eta_j| = O\left(n^{-\frac{2}{3}}\right).$$

We recall that α_j, β_j are the centered and scaled version of χ -squared random variables a_j^2, b_{j-1}^2 , respectively, and as such, they can be bounded using concentration of sub-gamma random variables (see, e.g., Theorem 2.3 of [21]). In particular, there exists some constant c such that, for all $j \leq n$ and for all $t > 0$,

$$\mathbb{P}\left(|\alpha_j| > c\left(\sqrt{\frac{t}{n}} + \frac{t}{n}\right)\right) \leq 2e^{-t} \quad (\text{B.6})$$

and likewise for each β_j , so we conclude that, for any ε , with probability at least $1 - \varepsilon/6$,

$$\max\{|\alpha_j|, |\beta_j| : j \leq n\} \leq cn^{-\frac{1}{2}}\sqrt{\log n}. \quad (\text{B.7})$$

Thus, for some constant $C_1 > 0$, with probability $1 - \varepsilon/3$,

$$|\omega_j \dots \omega_2 F_1| \leq |F_1| = |\eta_1 - \alpha_1| \leq C_1 n^{-\frac{1}{2}}. \quad (\text{B.8})$$

As ω_j is increasing in j ,

$$1 + \omega_j + \omega_j \omega_{j-1} + \dots + \omega_j \dots \omega_3 \leq 1 + \omega_j + \omega_j^2 + \dots = \frac{1}{1 - \omega_j}.$$

By Lemma B.2, $1 - \omega_j = \Theta\left(\left(\frac{n-j}{n}\right)^{\frac{1}{2}}\right)$. Thus, setting $j_0 := \lfloor n - n^{\frac{1}{3}}(\log n)^3 \rfloor$, we observe that, for some constant C_2 , with probability $1 - \varepsilon/3$,

$$\max_{j \leq j_0} |\eta_j + \omega_j \eta_{j-1} + \dots + \omega_j \dots \omega_3 \eta_2| \leq \max_{j \leq j_0} \left(|\eta_j| \frac{1}{1 - \omega_j} \right) \leq C_2 n^{-\frac{1}{3}} (\log n)^{-\frac{3}{2}}. \quad (\text{B.9})$$

Having bounded the first line of (B.5), we turn to the second line and recall the definition of L_j in (5.10). We have

$$\xi_j + \omega_j \xi_{j-1} + \dots + \omega_j \dots \omega_3 \xi_2 = L_j + \omega_j \dots \omega_3 \xi_2.$$

Note that $\max_{j \leq n} |\xi_j| = O(n^{-\frac{1}{2}} \sqrt{\log n})$ on the event (B.7). We also have, for some constant $C_3 > 0$, with probability $1 - O(n^{-1})$,

$$\max_{j \leq j_0} |L_j| = O(n^{-\frac{1}{3}} (\log n)^{-\frac{1}{4}}). \quad (\text{B.10})$$

The details for this bound can be obtained using a similar argument to the one found in Section 6.2 of [24]. In particular, the bound (B.10) follows from line (6.17) of that paper (where the notations α and Y_i can be translated as $\alpha = 2$ and $Y_i = L_i + O(n^{-\frac{1}{2}})$ in our context). Thus, for some constant $C_3 > 0$, with probability $1 - \varepsilon/3$,

$$\max_{j \leq j_0} |\xi_j + \omega_j \xi_{j-1} + \dots + \omega_j \dots \omega_3 \xi_2| \leq C_3 n^{-\frac{1}{3}} (\log n)^{-\frac{1}{4}}. \quad (\text{B.11})$$

Consider the event

$$\mathcal{G} := \{(\text{B.8}), (\text{B.9}), \text{ and } (\text{B.11}) \text{ hold}\}, \quad (\text{B.12})$$

which holds with probability $1 - \varepsilon$, for sufficiently large n . We now show that on this event, the third line of (B.5) is $o(n^{-\frac{1}{3}})$. Since this quantity depends on F_l 's up to F_{j-1} , we can control it in the process of using induction to show

$$\max_{j \leq j_0} |F_j| = o(n^{-\frac{1}{3}}) \quad \text{on the event } \mathcal{G}. \quad (\text{B.13})$$

More specifically, we will show that $\max_{j \leq j_0} |F_j| < 2C_3 n^{-\frac{1}{3}} (\log n)^{-\frac{1}{4}}$ where C_3 is the constant from (B.11). The base case holds by (B.8). Assume $\max_{l \leq j-1} |F_l| < 2C_3 n^{-\frac{1}{3}} (\log n)^{-\frac{1}{4}}$. Then, by (B.3), (B.4) and (B.7),

$$\max_{l \leq j} |\phi_l| = o(n^{-\frac{2}{3}}).$$

Note that the above maximum also includes ϕ_j . Thus, for some constant $C_4 > 0$,

$$|\phi_j + \omega_j \phi_{j-1} + \cdots + \omega_j \dots \omega_3 \phi_2| \leq \max_{l \leq j} |\phi_l| \frac{1}{1 - \omega_j} \leq C_4 n^{-\frac{1}{3}} (\log n)^{-3/2}.$$

Finally, by (B.5), we have that on \mathcal{G} ,

$$\begin{aligned} |F_j| &\leq C_1 n^{-\frac{2}{3}} (\log n) + C_2 n^{-\frac{1}{3}} (\log n)^{-\frac{1}{2}} + C_3 n^{-\frac{1}{3}} (\log n)^{-\frac{1}{4}} + C_4 n^{-\frac{1}{3}} (\log n)^{-3/2} \\ &< 2C_3 n^{-\frac{1}{3}} (\log n)^{-\frac{1}{4}}. \end{aligned}$$

This completes the induction step, and we obtain the lemma.

7.2. Proof of Lemma 5.5

Fix $\varepsilon > 0$. For $j_0 = \lfloor n - n^{1/3} (\log n)^3 \rfloor$ and $t = (e \log n)^2$, it suffices to show that, for sufficiently large n , each of the probabilities

$$p_1 := \mathbb{P} \left(\max_{j \leq j_0} a_j b_j < \sqrt{mn} - tn^{1/2} \right) \leq \mathbb{P} \left(a_{j_0} b_{j_0} < \sqrt{mn} - tn^{1/2} \right) \quad \text{and} \quad (\text{B.14})$$

$$p_2 := \mathbb{P} \left(\max_{j \leq j_0} a_j b_j > \sqrt{mn} + tn^{1/2} \right) = 1 - \prod_{j=1}^{j_0} \mathbb{P} \left(a_j b_j < \sqrt{mn} + tn^{1/2} \right) \quad (\text{B.15})$$

is less than $\varepsilon/2$. For any $j = 1, 2, \dots, j_0$, observe that

$$a_j^2 b_j^2 \stackrel{(d)}{=} \left(\sum_{i=1}^{m-n+j} g_i^2 \right) \left(\sum_{k=1}^j (g'_k)^2 \right), \quad (\text{B.16})$$

where $\stackrel{(d)}{=}$ denotes equality in distribution, and $g_1, \dots, g_{m-n+j}, g'_1, \dots, g'_j$ are independent standard gaussian variables. This implies that $\mathbb{E} a_j^2 b_j^2 = j(m - n + j)$ and $\text{Var}(a_j^2 b_j^2) = 2j(m - n + j)(m - n + 2j + 2)$. Viewing $a_j^2 b_j^2$ as a gaussian polynomial of degree 4 in $m - n + 2j$ variables g_i 's and g'_k 's, we have the following concentration result from [8] (see Corollary 5.49): For any $s \geq (2e)^2$,

$$\begin{aligned} &\mathbb{P} \left(|a_j^2 b_j^2 - j(m - n + j)| \right. \\ &\quad \left. \geq s \sqrt{2j(m - n + j)(m - n + 2j + 2)} \right) \leq \exp(-2\sqrt{s}/e). \quad (\text{B.17}) \end{aligned}$$

Apply this result to $(a_{j_0} b_{j_0})^2$ with $s = (e \log n)^2$, we obtain $p_1 \leq n^{-2}$. At the same time, (B.17) implies $\mathbb{P} \left(a_j b_j < \sqrt{mn} + tn^{1/2} \right) \geq 1 - n^{-2}$ for all $1 \leq j \leq j_0$, which yields $p_2 \leq 1 - e^{-c/n}$ for some $c > 0$. This completes the proof of the lemma.

References

- [1] Agliari, E., Barra, A., Bartolucci, S., Galluzzi, A., Guerra, F., Moauro, F.: Parallel processing in immune networks. *Phys. Rev. E* **87**, 042701 (2013)
- [2] Aizenman, M., Lebowitz, J.L., Ruelle, D.: Some rigorous results on the Sherrington–Kirkpatrick spin glass model. *Commun. Math. Phys.* **112**(1), 3–20 (1987)
- [3] Ajanki, O., Erdős, L., Krüger, T.: Local semicircle law with imprimitive variance matrix. *Electron. Commun. Probab.* **19**(33) (2014)
- [4] Alberici, D., Barra, A., Contucci, P., Mingione, E.: Annealing and replica-symmetry in deep Boltzmann machines. *J. Stat. Phys.* **180**, 665–677 (2020)
- [5] Alberici, D., Camilli, F., Contucci, P., Mingione, E.: The solution of the deep Boltzmann machine on the Nishimori line. *Commun. Math. Phys.* **387**(2), 1191–1214 (2021)
- [6] Alberici, D., Contucci, P., Mingione, E.: Deep Boltzmann machines: rigorous results at arbitrary depth. *Ann. Henri Poincaré* **22**(8), 2619–2642 (2021)
- [7] Anderson, G.W., Guionnet, A., Zeitouni, O.: *An Introduction to Random Matrices*. Cambridge Studies in Advanced Mathematics, vol. 118. Cambridge University Press, Cambridge (2010)
- [8] Aubrun, G., Szarek, S.J.: *Alice and Bob Meet Banach: The Interface of Asymptotic Geometric Analysis and Quantum Information Theory*, Mathematical Surveys and Monographs, vol. 223. American Mathematical Society, Providence, RI (2017)
- [9] Auffinger, A., Chen, W.-K.: Free energy and complexity of spherical bipartite models. *J. Stat. Phys.* **157**(1), 40–59 (2014)
- [10] Bai, Z.-D., Silverstein, J.W.: CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.* **32**(1A), 553–605 (2004)
- [11] Bai, Z.-D., Wang, X., Zhou, W.: Functional CLT for sample covariance matrices. *Bernoulli* **16**(4), 1086–1113 (2010)
- [12] Baik, J., Lee, J.-O.: Fluctuations of the free energy of the spherical Sherrington–Kirkpatrick model. *J. Stat. Phys.* **165**(2), 185–224 (2016)
- [13] Baik, J., Lee, J.-O.: Free energy of bipartite spherical Sherrington–Kirkpatrick model. *Ann. Inst. Henri Poincaré Probab. Stat.* **56**(4), 2897–2934 (2020)
- [14] Banerjee, D.: Fluctuation of the free energy of Sherrington–Kirkpatrick model with Curie–Weiss interaction: the paramagnetic regime. *J. Stat. Phys.* **178**(1), 211–246 (2019)
- [15] Barra, A., Contucci, P., Mingione, E., Tantari, D.: Multi-species mean field spin glasses. Rigorous results. *Ann. Henri Poincaré* **16**(3), 691–708 (2015)
- [16] Barra, A., Galluzzi, A., Guerra, F., Pizzoferrato, A., Tantari, D.: Mean field bipartite spin models treated with mechanical techniques. *Eur. Phys. J. B* **87**(3), Art. 74 (2014)
- [17] Barra, A., Genovese, G., Guerra, F.: The replica symmetric approximation of the analogical neural network. *J. Stat. Phys.* **140**(4), 784–796 (2010)
- [18] Barra, A., Genovese, G., Guerra, F.: Equilibrium statistical mechanics of bipartite spin systems. *J. Phys. A* **44**(24), 245002 (2011)

- [19] Barra, A., Genovese, G., Sollich, P., Tantari, D.: Phase diagram of restricted Boltzmann machines and generalized Hopfield networks with arbitrary priors. *Phys. Rev. E* **97**, 022310 (2018)
- [20] Bates, E., Sohn, Y.: Free energy in multi-species mixed p -spin spherical models. *Electron. J. Probab.* **27**, Art. 52 (2022)
- [21] Boucheron, S., Lugosi, G., Massart, P.: *Concentration Inequalities: A Nonasymptotic Theory of Independence*. With a foreword by Michel Ledoux, Oxford University Press, Oxford (2013)
- [22] Chen, W.-K.: The Aizenman–Sims–Starr scheme and Parisi formula for mixed p -spin spherical models. *Electron. J. Probab.* **18**(94), 14 (2013)
- [23] Chen, W.-K., Lam, W.-K.: Order of fluctuations of the free energy in the SK model at critical temperature. *ALEA Lat. Am. J. Probab. Math. Stat.* **16**(1), 809–816 (2019)
- [24] Collins-Woodfin, E. and Le, H.-G.: An edge CLT for the log determinant of Laguerre beta ensembles. *Ann. Inst. H. Poincaré Probab. Statist.* (2023) (to appear). [arXiv:2209.03271](https://arxiv.org/abs/2209.03271)
- [25] Comets, F., Neveu, J.: The Sherrington–Kirkpatrick model of spin glasses and stochastic calculus: the high temperature case. *Commun. Math. Phys.* **166**(3), 549–564 (1995)
- [26] Crisanti, A., Sommers, H.-J.: The spherical p -spin interaction spin glass model: the statics. *Zeitschrift für Physik B Condensed Matter* **87**(3), 341–354 (1992)
- [27] Dey, P.S., Wu, Q.: Fluctuation results for multi-species Sherrington–Kirkpatrick model in the replica symmetric regime. *J. Stat. Phys.* **185**(3), nos. 22, 40 (2021)
- [28] Dumitriu, I., Edelman, A.: Matrix models for beta ensembles. *J. Math. Phys.* **43**(11), 5830–5847 (2002)
- [29] Edelman, A., Wang, Y.: *Random matrix theory and its innovative applications*. In: Melnik, R., Kotsireas, I. (eds.) *Advances in Applied Mathematics, Modeling, and Computational Science*. Fields Institute Communications, vol. 66, pp. 91–116. Springer, Boston (2013)
- [30] Forrester, P.J., Rains, E.M.: Interrelationships between orthogonal, unitary and symplectic matrix ensembles. In: *Random Matrix Models and Their Applications*. *Math. Sci. Res. Inst. Publ.*, vol. 40, pp. 171–207. Cambridge Univ Press, Cambridge (2001)
- [31] Fröhlich, J., Zegarliński, B.: Some comments on the Sherrington–Kirkpatrick model of spin glasses. *Commun. Math. Phys.* **112**(4), 553–566 (1987)
- [32] Genovese, G.: Minimax formula for the replica symmetric free energy of deep restricted Boltzmann machines. *Ann. Appl. Probab.* **33**(3), 2324–2341 (2023)
- [33] Gustavsson, J.: Gaussian fluctuations of eigenvalues in the GUE. *Ann. Inst. H. Poincaré Probab. Statist.* **41**(2), 151–178 (2005)
- [34] Johnstone, I.M.: On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29**(2), 295–327 (2001)
- [35] Johnstone, I.M., Klochov, Y., Onatski, A., Pavlyshyn, D.: An edge CLT for the log determinant of Wigner ensembles (2020). [arXiv preprint arXiv:2011.13723](https://arxiv.org/abs/2011.13723)
- [36] Johnstone, I.M., Klochov, Y., Onatski, A., Pavlyshyn, D.: Spin glass to paramagnetic transition in spherical Sherrington–Kirkpatrick model with ferromagnetic interaction (2021). [arXiv preprint arXiv:2104.07629](https://arxiv.org/abs/2104.07629)

- [37] Kosterlitz, J.M., Thouless, D.J., Jones, R.C.: Spherical model of a spin glass. *Phys. Rev. Lett.* **36**(20), 1217 (1976)
- [38] Lambert, G., Paquette, E.: Strong approximation of Gaussian β -ensemble characteristic polynomials: the edge regime and the stochastic Airy function (2021). arXiv preprint [arXiv:2009.05003](https://arxiv.org/abs/2009.05003)
- [39] Landon, B.: Free energy fluctuations of the two-spin spherical SK model at critical temperature. *J. Math. Phys.* **63**(3), nos. 033301, 22 (2022)
- [40] Landon, B., Sosoe, P.: Fluctuations of the overlap at low temperature in the 2-spin spherical SK model. *Ann. Inst. Henri Poincaré Probab. Stat.* **58**(3), 1426–1459 (2022)
- [41] Liu, Q.: Fluctuations for the bipartite Sherrington–Kirkpatrick model. *J. Stat. Phys.* **184**(1), nos. 12, 21 (2021)
- [42] Lytova, A., Pastur, L.: Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.* **37**(5), 1778–1840 (2009)
- [43] Marčenko, V.A., Pastur, L.A.: Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)* **72**(114), 507–536 (1967)
- [44] Panchenko, D.: *The Sherrington–Kirkpatrick Model*. Springer, New York (2013)
- [45] Panchenko, D.: The Parisi formula for mixed p -spin models. *Ann. Probab.* **42**(3), 946–958 (2014)
- [46] Panchenko, D.: The free energy in a multi-species Sherrington–Kirkpatrick model. *Ann. Probab.* **43**(6), 3494–3513 (2015)
- [47] Parisi, G.: A sequence of approximated solutions to the S–K model for spin glasses. *J. Phys. A Math. Gen.* **13**(4), L115 (1980)
- [48] Pillai, N.S., Yin, J.: Universality of covariance matrices. *Ann. Appl. Probab.* **24**(3), 935–1001 (2014)
- [49] Salakhutdinov, R., Hinton, G.: An efficient learning procedure for deep Boltzmann machines. *Neural Comput.* **24**(8), 1967–2006 (2012)
- [50] Soshnikov, A.: A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Stat. Phys.* **108**, 1033–1056 (2002)
- [51] Su, Z.: Gaussian fluctuations in complex sample covariance matrices. *Electron. J. Probab.* **11**(48), 1284–1320 (2006)
- [52] Subag, E.: On the second moment method and RS phase of multi-species spherical spin glasses. *Electron. J. Probab.* **28**(50), 21 (2023)
- [53] Subag, E.: TAP approach for multispecies spherical spin glasses II: the free energy of the pure models. *Ann. Probab.* **51**(3), 1004–1024 (2023)
- [54] Talagrand, M.: Free energy of the spherical mean field model. *Probab. Theory Related Fields* **134**(3), 339–382 (2006)
- [55] Talagrand, M.: The Parisi formula. *Ann. Math. (2)* **163**(1), 221–263 (2006)

Elizabeth W. Collins-Woodfin
 Department of Mathematics and Statistics
 McGill University
 Montreal QC H3A 0G4
 Canada
 e-mail: elizabeth.collins-woodfin@mail.mcgill.ca

Free Energy Fluctuations of the Bipartite Spherical

Han Gia Le
Department of Mathematics
University of Michigan
Ann Arbor MI 48109
USA
e-mail: hanle@umich.edu

Communicated by Christian Maes.

Received: December 3, 2023.

Accepted: May 4, 2024.