Free energy of bipartite spherical Sherrington–Kirkpatrick model

JINHO BAIK*AND JI OON LEE†

November 17, 2017

Abstract

We consider the free energy of the bipartite spherical Sherrington–Kirkpatrick model. We find the critical temperature and prove the limiting free energy for all non-critical temperature. We also show that the law of the fluctuation of the free energy converges to the Gaussian distribution when the temperature is above the critical temperature, and to the GOE Tracy–Widom distribution when the temperature is below the critical temperature. The result is universal, and the analysis is applicable to a more general setting including the case where the disorders are non-identically distributed.

1 Introduction

1.1 Bipartite SSK

The Sherrington–Kirkpatrick (SK) model and the spherical Sherrington–Kirkpatrick (SSK) model are disordered systems in which the spin variables are subject to Gibbs probability measures defined by random Hamiltonians. They can be thought of as finite-temperature versions of the problem of finding the maximum of a random function on either a hypercube (SK model) or a sphere (SSK model). As such, there are significant interests in these models and their generalizations in probability and statistical physics, as well as computer science and social science. There is a long history to the subject with many important results. We refer to [24] and references therein.

A natural variation is the case when the spins are divided into two (or more) groups such that the spins of different group interact but the spins of same group do not interact. When there are two groups, we are lead to the bipartite system.

The bipartite spherical Sherrington–Kirkpatrick model (SSK) model is defined as follows. Let

$$S_{n-1} = \{\boldsymbol{u} \in \mathbb{R}^n : \|\boldsymbol{u}\| = \sqrt{n}\} \tag{1.1}$$

be a sphere in $\mathbb{R}^n$. Let N_1 and N_2 be two positive integers and consider two types of spin variables $\boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_{N_1})$ and $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_{N_2})$ on two different spheres,

$$\boldsymbol{\sigma} \in S_{N_1-1}, \qquad \boldsymbol{\tau} \in S_{N_2-1}. \tag{1.2}$$

*Department of Mathematics, University of Michigan, Ann Arbor, MI, 48109, USA
email: `baik@umich.edu`

†Department of Mathematical Sciences, KAIST, Daejeon, 34141, Korea
email: `jioon.lee@kaist.edu`

Define the Hamiltonian

$$H(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} J_{ij} \sigma_i \tau_j, \qquad N := N_1 + N_2, \tag{1.3}$$

where J_{ij} are independent random variables of mean 0 and variance 1. The bipartite SSK model is defined, for each $\beta > 0$, by the Gibbs probability measure

$$P(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \frac{1}{Z_{N_1,N_2}} e^{\beta H(\boldsymbol{\sigma}, \boldsymbol{\tau})}, \qquad (\boldsymbol{\sigma}, \boldsymbol{\tau}) \in S_{N_1-1} \times S_{N_2-1} \tag{1.4}$$

where β is called the inverse temperature and Z_{N_1,N_2} is the normalization constant, which is also known as the partition function. Note that the probability measure depends on the random variables J_{ij}.

The goal of this paper is to study the free energy $F_{N_1,N_2}(\beta) = N^{-1} \log Z_{N_1,N_2}(\beta)$ as $N_1, N_2 \to \infty$. For small enough β, Auffinger and Chen obtained a minimization formula for the limiting free energy in [3]. We mention that their work applies to more general mixed (p, q)-spin Hamiltonians with external fields. One of the contributions of this paper is the computation of the limiting free energy for the Hamiltonian (1.3) for all β other than a critical β_c. We also find the critical inverse temperature β_c explicitly. When β is small, our formula agrees with the result of Auffinger and Chen. Another contribution of this paper is the evaluation of the next order term. We obtain the limiting law of the fluctuations, again for all $\beta \neq \beta_c$. We show that the fluctuations are Gaussian for $\beta < \beta_c$, and are given by the Tracy–Widom distribution of random matrix theory for $\beta > \beta_c$. In this paper, the disorder parameters J_{ij} are not restricted to Gaussian variables.

For the usual SK and SSK models, the limiting free energy is given by the Parisi formula [34] and Crisanti–Sommers formula [18], which were rigorously proved by Talagrand in [40, 39]. The fluctuations were obtained for β below a critical value by first Aizenman, Lebowitz, and Ruelle in [1] and subsequently in [23, 17, 14]. There are several recent results for large β and and also for the case with the presence of the external field in [7, 38, 15, 16, 8].

1.2 Multi-species SK

The bipartite Sherrington–Kirkpatrick (SK) model is defined by the same Hamiltonian (1.3) but the spins are now assumed to be on a hypercube,

$$(\boldsymbol{\sigma}, \boldsymbol{\tau}) \in \{-1, 1\}^{N_1} \times \{-1, 1\}^{N_2} = \{-1, 1\}^{N_1+N_2}. \tag{1.5}$$

Note that for the spheres, $S_{N_1-1} \times S_{N_2-1}$ is not equal to $S_{N_1+N_2-1}$.

The bipartite SK model is a special case of the multi-species Sherrington–Kirkpatrick model. The multi-species SK model was introduced in [10], and it is defined as follows. Let

$$H_N^{\mathrm{MS}}(\sigma) = \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} g_{ij} \sigma_i \sigma_j, \qquad \boldsymbol{\sigma} = (\sigma_1, \cdots, \sigma_N) \in \{-1, 1\}^N. \tag{1.6}$$

be the usual SK Hamiltonian. The disorder parameters g_{ij} are independent centered random variables. However, we assume that the variances of g_{ij} are not uniform but they depend on the "species" of the index i and j. Let $\mathcal{S}$ be a finite set independent of N and call the elements of $\mathcal{S}$ species. Fix a map

$$s : \{1, \cdots, N\} \to \mathcal{S}. \tag{1.7}$$

The value $s(i)$ assigns a species to the index i. Now we assume that the variance of g_{ij} depends only on the species of i and j: Let

$$\Delta^2 = (\Delta^2_{st})_{s,t\in\mathcal{S}}, \qquad \Delta^2_{st} = \Delta^2_{ts} \tag{1.8}$$

be a symmetric matrix with non-negative entries and we assume that

$$\mathbb{E}[g^2_{ij}] = \Delta^2_{s(i),s(j)}. \tag{1.9}$$

Setting $N_s = |\{i : s(i) = s\}|$, the interesting case is when $\frac{N_s}{N} \to r_s \in (0, 1)$ as $N \to \infty$ for each $s \in \mathcal{S}$.

The bipartite SK model is the multi-species SK model when $|\mathcal{S}| = 2$ and $\Delta^2 = \frac{1}{4}\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$. Note that in this case Δ^2 is not positive-semidefinite.

The liming free energy of the multi-species model was studied in [10] and [33]. In [10], Barra, Contucci, Mingione, and Tantari obtained a lower bound of the limiting free energy assuming that Δ^2 is positive-semidefinite. On the other hand, Panchenko obtained an upper bound in [33] for general Δ^2. When Δ^2 is positive-semidefinite, the upper bound matches with the lower bound, and hence one obtains the limiting free energy. The general case, including the bipartite case, remains an open question; see [12, 11] for some conjectural formulas for the bipartite SK model.

1.3 Two multi-species SSK models

Let us consider a spherical version of multi-species SK model. We take the same Hamiltonian as (1.6) with same disorder parameters g_{ij} satisfying (1.9). Note that if $\boldsymbol{\sigma} \in \{-1, 1\}^N$, then $\|\boldsymbol{\sigma}\| = \sqrt{N}$. There are two different natural ways of embedding the hypercube. One way is that

$$\boldsymbol{\sigma} \in S_{N-1}. \tag{1.10}$$

The other way is that

$$\boldsymbol{\sigma} \in S_{N_{s_1}-1} \times \cdots \times S_{N_{s_m}-1} \tag{1.11}$$

where m is the number of species, the set of species is denoted by $\mathcal{S} = \{s_1, \cdots, s_m\}$, and N_{s_k} is the number of indices corresponding to the species s_k satisfying $\sum_{k=1}^m N_{s_k} = N$. In both cases, $\|\boldsymbol{\sigma}\| = \sqrt{N}$. Therefore, we have two different multi-species spherical Sherrington–Kirkpatrick models, one with spins on one sphere and the other with spins on a product space of spheres.

The bipartite SSK model we introduced earlier corresponds to a special case of (1.11). In this paper, we focus only on this model. However, using a method similar to this paper, one can study the model with (1.10) for bipartite case and also some multi-species cases (possibly not positive-semidefinite Δ^2). This "one-sphere multi-species SSK" model will be considered in a separate paper.

1.4 Connection to random matrices

We use a special structure of the Hamiltonian (1.3) to study its limiting free energy and the fluctuations. Setting the matrix $J = (J_{ij})$ and considering $\boldsymbol{\sigma}$ and $\boldsymbol{\tau}$ as column vectors, the critical points of the function $f(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \boldsymbol{\sigma}^T J \boldsymbol{\tau}$ (which is a constant multiple of the Hamiltonian) subject to the constraints $\|\boldsymbol{\sigma}\|^2 = N_1$ and $\|\boldsymbol{\tau}\|^2 = N_2$ satisfy the equations

$$J\boldsymbol{\tau} = \lambda_1 \boldsymbol{\sigma}, \qquad J^T \boldsymbol{\sigma} = \lambda_2 \boldsymbol{\tau}\,, \tag{1.12}$$

where λ_1 and λ_2 are Lagrange multipliers. These equations imply that $\boldsymbol{\sigma}$ is an eigenvector of the matrix JJ^T, $\boldsymbol{\tau}$ is an eigenvector of J^TJ, and $\lambda_1\lambda_2$ is an eigenvalue of J^TJ (and also JJ^T).

The matrix J is a random matrix with independent and identically distributed entries. The matrix J^TJ is called a (constant multiple of) sample covariance matrix (with null covariance) in statistics and also is said to belong to the Laguerre orthogonal ensemble in random matrix theory [31, 22, 5]. It is one of the fundamental matrices in random matrix theory. The behavior of the eigenvalues of J^TJ (the squared singular values of J) in the large dimension limit is well-studied.

There is a more direct connection between the random matrices and the free energy. In [27, 7], it was shown that the partition function of the usual SSK model can be expressed as a random single integral. In this paper, we obtain a similar result for the bipartite SSK model, but this time the random integral is a double integral; see Lemma 2.5. This random double integral involves the eigenvalues of J^TJ. We analyze the double integral asymptotically using the method of steepest-descent. The reason that we can apply the method of steepest-descent to the random integral is that even though the eigenvalues are random, their fluctuations about their classical locations are small. Precise estimates for the locations of the eigenvalues were obtained recently in random matrix theory. The "rigidity" estimates for the eigenvalues of J^TJ were proved by Pillai and Yin in 2014 [35] for the case $N_1 \neq N_2$. The estimates for the case $N_1 = N_2$ follow from [2]. The rigidity estimates are also crucial in proving the universality in random matrix. They are proved for several other random matrix ensembles [21, 20, 13]. Our analysis is applicable to a large class of random double integrals under certain general conditions (including the rigidity condition) on a sequence of random variables. We obtain the results for the bipartite SSK model as a special case of a more general asymptotic result for random double integrals.

The strategy above is an extension to our previous works [7, 8] for the SSK model. A similar idea was also used in an earlier physics paper [27] for a non-rigorous analysis for the limiting free energy. An important change from our previous work is that the random integral is a double integral this time. This change adds significant technical difficulties. Even in [7, 8], the asymptotic analysis for large β (the low temperature regime) was subtle due to the fact that the critical point in the method of steepest-descent is close to a branch point. While we could use a certain symmetry to simplify the situation in the SSK model, we lose the symmetry for the double integral in this paper. This leads us to a more involved analysis; see Section 5 for more discussions.

1.5 Organization of the paper

The paper is organized as follows. We state the precise definition of the model and state the main results in Section 2. The main fluctuation results are Theorem 2.2 and Theorem 2.4. We also state the double integral formula of the partition function. The asymptotic analysis of the double integral can be carried out under certain general conditions. In Section 3, we state these conditions and discuss the critical point for the steepest-descent analysis. The asymptotic analysis of the general random double integrals is performed for the high temperature regime in Section 4 and for the low temperature regime in Section 5. Section 5 is the most technical part of the paper. In Section 6, we prove Theorem 2.4 using the results of Sections 4 and 5. In Section 7, we derive Theorem 2.2 from Theorem 2.4 using results from random matrix theory. In Section 8, we briefly discuss the case where the disorders are non-identically distributed.

Acknowledgments

We would like to thank Antonio Auffinger, Wei-Kuo Chen, Dmitry Panchenko, and Eric Rains for useful conversations. The work of Jinho Baik was supported in part by NSF grants DMS-1361782 and DMS-1664692, and the Simons Fellows program. The work of Ji Oon Lee was supported in part by Samsung Science and Technology Foundation project number SSTF-BA1402-04.

2 Results

In this section, we define the model precisely and state the results.

2.1 Definitions

Let

$$S_{n-1} = \{\boldsymbol{u} \in \mathbb{R}^n : \|\boldsymbol{u}\| = \sqrt{n}\} \tag{2.1}$$

be a sphere of radius $\sqrt{n}$ in $\mathbb{R}^n$. Let N_1 and N_2 be positive integers and set

$$N = N_1 + N_2. \tag{2.2}$$

Let $J = (J_{ij})_{i=1,\cdots,N_1, j=1,\cdots,N_2}$ be an $N_1 \times N_2$ matrix with i.i.d. entries of mean 0 and variance 1. Define the Hamiltonian

$$H(\boldsymbol{\sigma}, \boldsymbol{\tau}) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} J_{ij} \sigma_i \tau_j = \frac{1}{\sqrt{N}} \langle \boldsymbol{\sigma}, J \boldsymbol{\tau} \rangle, \qquad (\boldsymbol{\sigma}, \boldsymbol{\tau}) \in S_{N_1 - 1} \times S_{N_2 - 1}. \tag{2.3}$$

The free energy of the bipartite SSK model at inverse temperature β is defined by

$$F_{N_1,N_2}(\beta) = \frac{1}{N} \log Z_{N_1,N_2}(\beta)\,, \tag{2.4}$$

where the partition function Z_{N_1,N_2} is defined by

$$Z_{N_1,N_2}(\beta) = \int_{S_{N_1-1}} \int_{S_{N_2-1}} e^{\beta H(\boldsymbol{\sigma}, \boldsymbol{\tau})} \mathrm{d}\omega_{N_2}(\boldsymbol{\tau}) \mathrm{d}\omega_{N_1}(\boldsymbol{\sigma}). \tag{2.5}$$

Here, $\mathrm{d}\omega_n(\boldsymbol{u})$ is the uniform probability measure on the sphere S_{n-1}.

We assume the following for J. Let J_{ij} be independent random variables such that:

- The entries are centered with unit variance, i.e., $\mathbb{E}[J_{ij}] = 0$ and $\mathbb{E}[J_{ij}^2] = 1$.
- For any i, j, $\mathbb{E}[J_{ij}^3] = W_3$ and $\mathbb{E}[J_{ij}^4] = W_4$ for some constants W_3, W_4.
- All moments of J_{ij} are finite.

We consider the limit as $N, N_1, N_2 \to \infty$. Assume that there is $\delta > 0$ such that

$$\frac{N_1}{N} = r_1 + O(N^{-1-\delta}), \qquad \frac{N_2}{N} = r_2 + O(N^{-1-\delta}). \tag{2.6}$$

for some

$$r_1, r_2 > 0 \text{ satisfying } r_1 + r_2 = 1. \tag{2.7}$$

2.2 Limiting free energy

We first state the limiting free energy.

Theorem 2.1. *Set*

$$\beta_c := (r_1 r_2)^{-\frac{1}{4}}. \tag{2.8}$$

Define, for $0 < \beta < \beta_c$,

$$F(\beta) = \frac{r_1 r_2 \beta^2}{2} \tag{2.9}$$

and for $\beta > \beta_c$,

$$\begin{aligned} F(\beta) = & \frac{(\sqrt{r_1} + \sqrt{r_2})\sqrt{\mathsf{S}} - \sqrt{r_1 r_2} - 1}{2} \\ & - \frac{r_1 - r_2}{4} \log \left(\frac{\sqrt{\mathsf{S}} + \sqrt{r_1} - \sqrt{r_2}}{\sqrt{\mathsf{S}} - \sqrt{r_1} + \sqrt{r_2}} \right) - \frac{r_2}{4} \log r_1 - \frac{r_1}{4} \log r_2 - \frac{1}{2} \log \beta \end{aligned} \tag{2.10}$$

where

$$\mathsf{S} = \mathsf{S}(\beta, r_1, r_2) := (\sqrt{r_1} - \sqrt{r_2})^2 + 4 r_1 r_2 \beta^2. \tag{2.11}$$

Then,

$$F_{N_1,N_2}(\beta) \to F(\beta) \tag{2.12}$$

as $N \to \infty$ *in probability for every* $\beta \neq \beta_c$.

Proof. This result is a simple consequence of Theorem 2.2 below for the fluctuations. □

Auffinger and Chen obtained the limiting free energy when β is small enough in [3] in terms of a minimization problem. Their result applies to general mixed (p, q)-spin Hamiltonians with the presence of the external field. The specialization to the $(p, q) = (1, 1)$ case (we also set $h_1 = h_2 = 0$ and $\beta_{1,1} = \sqrt{r_1 r_2}\beta$ in Theorem 1 of [3]) is the following: There is a small constant $\beta_0 > 0$ (which is not explicitly determined) such that for $\beta < \beta_0$,

$$\lim_{N\to\infty} F_{N_1,N_2}(\beta) = \min_{a,b\in[0,1)} P(a, b) \tag{2.13}$$

where

$$P(a, b) = \frac{r_1}{2}\left(\frac{a}{1-a} + \log(1-a)\right) + \frac{r_2}{2}\left(\frac{b}{1-b} + \log(1-b)\right) + \frac{r_1 r_2 \beta^2}{2}(1-ab). \tag{2.14}$$

It is easy to find the minimum explicitly. It is straightforward to check that the minimum occurs on the boundary of domain $[0,1) \times [0,1)$ when $\beta \le (r_1 r_2)^{-1/4}$ and inside the domain $[0,1) \times [0,1)$ when $\beta > (r_1 r_2)^{-1/4}$. The minimizers are $(a,b) = (0,0)$ when $\beta \le (r_1 r_2)^{-1/4}$ and

$$(a,b) = \left(1 - \frac{\sqrt{\mathsf{S}} - \sqrt{r_1} + \sqrt{r_2}}{2\sqrt{r_1} r_2 \beta^2}, 1 - \frac{\sqrt{\mathsf{S}} + \sqrt{r_1} - \sqrt{r_2}}{2 r_1 \sqrt{r_2} \beta^2}\right) \tag{2.15}$$

when $\beta > (r_1 r_2)^{-1/4}$, where S is (2.11). From this, we find that the minimum is equal to $F(\beta)$ in Theorem 2.1 for all $\beta \neq \beta_c$. Hence, Theorem 2.1 implies that the result (2.13) of Auffinger and Chen actually holds for all $\beta \neq \beta_c$ for the (1,1)-spin Hamiltonian.

2.3 Fluctuations of the free energy

Next result is about the fluctuations of the free energy.

Theorem 2.2. *We have the following convergence in distribution.*

(i) In the high temperature regime $0 < \beta < (r_1 r_2)^{-\frac{1}{4}}$,

$$N(F_{N_1,N_2} - F(\beta)) \Rightarrow \mathcal{N}(\mu, \sigma^2), \tag{2.16}$$

where $\mathcal{N}(\mu, \sigma^2)$ *is the Gaussian distribution with mean*

$$\mu = \frac{1}{4}\log(1 - r_1 r_2 \beta^4) - \log 2 - (W_4 - 3)\frac{r_1 r_2 \beta^4}{4} \tag{2.17}$$

and variance

$$\sigma^2 = -\frac{1}{2}\log\left(1 - r_1 r_2 \beta^4\right) + (W_4 - 3)\frac{r_1 r_2 \beta^4}{4}. \tag{2.18}$$

(ii) In the low temperature regime $\beta > (r_1 r_2)^{-\frac{1}{4}}$,

$$\frac{N^{\frac{2}{3}}}{\mathsf{A}}(F_{N_1,N_2} - F(\beta)) \Rightarrow \mathrm{TW} \tag{2.19}$$

where

$$\mathsf{A} = \mathsf{A}(\beta, r_1, r_2) = \frac{(\sqrt{r_1} + \sqrt{r_2})^{\frac{1}{3}}(\sqrt{\mathsf{S}} - \sqrt{r_1} - \sqrt{r_2})}{4(r_1 r_2)^{\frac{1}{6}}} \tag{2.20}$$

with $\mathsf{S} = (\sqrt{r_1} - \sqrt{r_2})^2 + 4 r_1 r_2 \beta^2$ *defined in* (2.11) *and* TW *denotes the GOE Tracy–Widom distribution.*

We remark that the limiting free energy $F(\beta)$ and the constants μ, σ^2, and A are all symmetric in r_1 and r_2.

The above change from the Gaussian distribution for high temperature to the Tracy–Widom distribution for low temperature also occurs in the usual SSK model [7].

2.4 Free energy and eigenvalues

Assume, without loss of generality, that

$$N_1 \geq N_2. \tag{2.21}$$

The matrix of the disorder parameters $J = (J_{ij})$ is an $N_1 \times N_2$ matrix. We consider the $N_2 \times N_2$ square random matrix

$$S = \frac{1}{N_1} J^T J. \tag{2.22}$$

In statistics, S is known as a sample covariance matrix (with null covariance). In random matrix theory, S is also known to belong to the Laguerre orthogonal ensemble [31, 22, 5]. Let

$$\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{N_2} \geq 0 \tag{2.23}$$

be the eigenvalues of S. We note that $\sqrt{\mu_i}$ are the singular values of $\frac{1}{\sqrt{N_1}} J$.

The eigenvalues of S are well studied in the random matrix theory. For example, the empirical spectral distribution (ESD) of S converges to the Marchenko-Pastur distribution [30]:

$$\frac{1}{N_2} \sum_{i=1}^{N_2} \delta_{\mu_i}(x) \mathrm{d}x \to \mathrm{d}\mu_{\mathrm{MP}}(x) \tag{2.24}$$

weakly in probability as $N_1, N_2 \to \infty$ with $\frac{N_2}{N_1} \to \frac{r_1}{r_2} \in (0, 1]$, where

$$\mathrm{d}\mu_{\mathrm{MP}}(x) := \frac{2\sqrt{(d_+ - x)(x - d_-)}}{\pi(\sqrt{d_+} - \sqrt{d_-})^2 x} \mathbb{1}_{(d_-, d_+)}(x)\, \mathrm{d}x \tag{2.25}$$

with

$$d_- = \frac{(\sqrt{r_1} - \sqrt{r_2})^2}{r_1}, \qquad d_+ = \frac{(\sqrt{r_1} + \sqrt{r_2})^2}{r_1}. \tag{2.26}$$

The next theorem relates the second leading term of the free energy with the eigenvalues of S. We begin by introducing a suitable notion for the estimates.

Definition 2.3 (High probability event)**.** We say that an N-dependent event Ω_N holds with high probability if, for any given $D > 0$, there exists $N_0 > 0$ such that

$$\mathbb{P}(\Omega_N^c) \leq N^{-D}$$

for all $N > N_0$.

Theorem 2.4. *Without loss of generality, assume that $r_1 \geq r_2$. The following hold with high probability for any fixed $0 < \epsilon < \frac{1}{100}$.*

(i) In the high temperature regime $0 < \beta < (r_1 r_2)^{-\frac{1}{4}}$,

$$\begin{aligned} F_{N_1,N_2}(\beta) = F(\beta) - \frac{1}{2N} \left[\sum_{i=1}^{N_2} \log(z_c - \mu_i) - N_2 \int \log(z_c - x) \mathrm{d}\mu_{\mathrm{MP}}(x) \right] \\ + \frac{1}{N} \left[\frac{1}{2} \log\left(1 - r_1 r_2 \beta^4\right) - \log 2 \right] + O(N^{-2+\epsilon}) \end{aligned} \tag{2.27}$$

where

$$z_c = \frac{1 + \beta^2 + r_1 r_2 \beta^4}{r_1 \beta^2}. \tag{2.28}$$

(ii) In the low temperature regime $\beta > (r_1 r_2)^{-\frac{1}{4}}$,

$$F_{N_1,N_2}(\beta) = F(\beta) + \left(\mu_1 - \frac{(\sqrt{r_1}+\sqrt{r_2})^2}{r_1}\right) \frac{r_1(\sqrt{\mathsf{S}} - \sqrt{r_1} - \sqrt{r_2})}{4(\sqrt{r_1}+\sqrt{r_2})} + O(N^{-1+\epsilon}). \tag{2.29}$$

with high probability where $\mathsf{S} = (\sqrt{r_1} - \sqrt{r_2})^2 + 4r_1r_2\beta^2$ as in (2.11).

Theorem 2.4 shows that the difference $F_{N_1,N_2}(\beta) - F(\beta)$ is governed by the top eigenvalue μ_1 when $\beta > \beta_c$ and by a certain combination of all eigenvalues when $\beta < \beta_c$. The behaviors of the top eigenvalue and the special combination of all the eigenvalues appearing in the theorem are well-known in random matrix theory. In Section 7, we prove Theorem 2.2 by combining Theorem 2.4 and the results from random matrix theory.

2.5 Special case

When $r_1 = r_2 = \frac{1}{2}$, the formulas are particularly simple. We will compare the formulas with the usual SSK model:

$$F^{\mathrm{SSK}}(\beta) = \begin{cases} \beta^2 & \text{for } \beta < \frac{1}{2}, \\ 2\beta - \frac{3}{4} - \frac{1}{2}\log(2\beta) & \text{for } \beta > \frac{1}{2}. \end{cases} \tag{2.30}$$

When $r_1 = r_2 = \frac{1}{2}$, we find that the limiting free energy of the bipartite SSK models satisfies

$$F^{\mathrm{BSSK}}(2\sqrt{2}\beta) = F^{\mathrm{SSK}}(\beta), \qquad \beta \neq \frac{1}{2}.$$

For general $r_1 \neq r_2$, we have

$$F^{\mathrm{BSSK}}\left(\beta\sqrt{\frac{2}{r_1 r_2}}\right) = F^{\mathrm{SSK}}(\beta), \qquad \beta < \frac{1}{2},$$

but this relationship is not true in low temperature regime $\beta > \frac{1}{2}$.

For the fluctuations, when $r_1 = r_2$ and $\beta > \frac{1}{2}$, $\mathsf{A}(2\sqrt{2}\beta) = \beta - \frac{1}{2}$. This is the same constant appearing for the low temperature fluctuations of the usual SSK model [7] (see (iv) of Section 3.1). However, when $r_1 = r_2$ and $\beta < \frac{1}{2}$, the constants $\mu(2\sqrt{2}\beta)$ and $\sigma^2(2\sqrt{2}\beta)$ are not same as the constants for the high temperature fluctuations of the usual SSK model ((3.12) and (3.13) of [7]).

We note that the limiting distribution of the eigenvalues associated to the bipartite SSK and the usual SSK are related when $r_1 = r_2$. When $r_1 = r_2$, then the Marchenko-Pastur distribution (2.25) is

$$\mu_{\mathrm{MP}}(x) = \frac{\sqrt{x(4-x)}}{2\pi x} \mathbb{1}_{(0,4)}(x)\,\mathrm{d}x\,. \tag{2.31}$$

After a simple change of variables $x = y^2$, this distribution is equal to the semicircle distribution,

$$\mu_{\mathrm{SC}}(y) = \frac{\sqrt{4-y^2}}{2\pi} \mathbb{1}_{(-2,2)}(y)\,\mathrm{d}y\,, \tag{2.32}$$

which is the limiting distribution for the random symmetric matrix associated to the usual SSK model.

2.6 Double integral representation

As mentioned in Introduction, the starting point of our analysis for Theorem 2.4 is an explicit double integral formula for the partition function. In this subsection, we state and prove the formula. Recall that we assume, without loss of generality, that $N_1 \geq N_2$. Let

$$S^{n-1} = \{\boldsymbol{u} \in \mathbb{R}^n : \|\boldsymbol{u}\| = 1\} \tag{2.33}$$

and $\Omega_n(\boldsymbol{u})$ is the surface measure (which is not normalized) on the unit sphere S^{n-1}. After setting $\boldsymbol{\sigma} = \sqrt{N_1}\boldsymbol{x}$ and $\boldsymbol{\tau} = \sqrt{N_2}\boldsymbol{y}$, the partition function (2.5) satisfies

$$Z_{N_1,N_2}(\beta) = \frac{\hat{Z}_{N_1,N_2}(N_1 N_2^{\frac{1}{2}} N^{-\frac{1}{2}}\beta)}{|S^{N_1-1}||S^{N_2-1}|} \tag{2.34}$$

where

$$\hat{Z}_{N_1,N_2}(b) = \int_{S^{N_1-1}} \int_{S^{N_2-1}} e^{b\langle \boldsymbol{x}, M\boldsymbol{y}\rangle} \mathrm{d}\Omega_{N_2}(\boldsymbol{y})\, \mathrm{d}\Omega_{N_1}(\boldsymbol{x}), \qquad M := \frac{J}{\sqrt{N_1}}. \tag{2.35}$$

Let $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_{N_2} \geq 0$ be the eigenvalues of the $N_2 \times N_2$ matrix $S = M^T M = \frac{1}{N_1} J^T J$. The following formula is a variation of a result in [7].

Lemma 2.5. *For $N_1 \geq N_2$, we have*

$$\hat{Z}_{N_1,N_2}(b) = -2^{N_2} \left(\frac{\pi}{b}\right)^{\frac{N_1+N_2}{2}-2} \int_{\gamma_1 - \mathrm{i}\infty}^{\gamma_1+\mathrm{i}\infty} \int_{\gamma_2-\mathrm{i}\infty}^{\gamma_2+\mathrm{i}\infty} \frac{e^{b(z_1+z_2)}}{z_1^{(N_1-N_2)/2} \prod_{i=1}^{N_2} \sqrt{4z_1 z_2 - \mu_i}} \mathrm{d}z_2\, \mathrm{d}z_1 \tag{2.36}$$

where γ_1 and γ_2 are any real positive constants satisfying $4\gamma_1\gamma_2 > \mu_1$.

Proof. From the singular value decomposition, $M = UDV$ where U and V are orthogonal matrices (of size N_1 and N_2, respectively) and $D = (D_{ij})$ is an $N_1 \times N_2$ matrix with $D_{ii} = \sqrt{\mu_i}$ and $D_{ij} = 0$ for $i \neq j$. Hence, after changing the variables $\boldsymbol{x}$ and $\boldsymbol{y}$ to $U\boldsymbol{x}$ and $V^T\boldsymbol{y}$, respectively, we have

$$\hat{Z}_{N_1,N_2}(b) = \int_{S^{N_1-1}} \int_{S^{N_2-1}} e^{b\sum_{i=1}^{N_2} \sqrt{\mu_i} x_i y_i} \,\mathrm{d}\Omega_{N_2}(\boldsymbol{y})\mathrm{d}\Omega_{N_1}(\boldsymbol{x}). \tag{2.37}$$

Consier

$$I(z_1, z_2) := \int_{\mathbb{R}^{N_1}} \int_{\mathbb{R}^{N_2}} e^{\sum_{i=1}^{N_2} \sqrt{\mu_i} X_i Y_i} e^{-z_1|\boldsymbol{X}|^2 - z_2|\boldsymbol{Y}|^2} \,\mathrm{d}^N\boldsymbol{Y}\, \mathrm{d}^M\boldsymbol{X}. \tag{2.38}$$

We evaluate this integral in two ways. First, by computing the Gaussian integrals,

$$I(z_1, z_2) = \frac{2^{N_2}\pi^{\frac{N_1+N_2}{2}}}{z_1^{\frac{N_1-N_2}{2}}} \prod_{i=1}^{N_2} \frac{1}{\sqrt{4z_1z_2 - \mu_i}} \tag{2.39}$$

for z_1 and z_2 satisfying $\operatorname{Re} z_1 > 0$, $\operatorname{Re} z_2 > 0$ and $\operatorname{Re}(4z_1z_2) > \mu_1$. Second, using polar coordinates $\boldsymbol{X} = \sqrt{u}\boldsymbol{x}$, $Y = \sqrt{v}\boldsymbol{y}$ with $u, v > 0$ and $\boldsymbol{x} \in S^{N_1-1}$, $\boldsymbol{y} \in S^{N_2-1}$, we find that

$$I(z_1, z_2) = \int_0^\infty \int_0^\infty \frac{1}{4} u^{\frac{N_1}{2}-1} v^{\frac{N_2}{2}-1} \hat{Z}_{N_1,N_2}(\sqrt{uv}) e^{-z_1u - z_2v} \,\mathrm{d}v\, \mathrm{d}u. \tag{2.40}$$

By taking the inverse Laplace transform twice, we find

$$u^{\frac{N_1}{2}-1} v^{\frac{N_2}{2}-1} \hat{Z}_{N_1,N_2}(\sqrt{uv}) = -\frac{1}{\pi^2} \int_{\gamma_1-\mathrm{i}\infty}^{\gamma_1+\mathrm{i}\infty} \int_{\gamma_2-\mathrm{i}\infty}^{\gamma_2+\mathrm{i}\infty} I(z_1, z_2) e^{z_1u+z_2v} \mathrm{d}z_2\, \mathrm{d}z_1 \tag{2.41}$$

for any $\gamma_1, \gamma_2 > 0$ satisfying $4\gamma_1\gamma_2 > \mu_1$. Setting $u = v = b$, we obtain the result. □

3 Random double integral

The main technical part of this paper is the asymptotic analysis of the double integral in Lemma 2.5. The integrand contains the random eigenvalues $\mu_i = \mu_i(N_2)$, $1 \le i \le N_2$. We use the method of steepest-descent to evaluate the double integral asymptotically. This is possible since the eigenvalues satisfy certain rigidity estimates [35, 2] with high probability. Since the analysis depends only on the rigidity estimates and a few other properties of μ_i, we present the analysis for a general sequence of random double integrals. In this section, we define general random double integrals and state the conditions for the parameters and random variables of the integrals. The asymptotic analysis is carried out in the next two sections, Sections 4 and 5. Section 5 is the most technical part of the analysis. We then discuss in Section 6 that the eigenvalues of the matrix $\frac{1}{N_1}J^TJ$ for the bipartite SSK model satisfy the conditions (with high probability) and derive Theorem 2.4 from the general asymptotic results, Proposition 4.4 and 5.8 for the double integrals.

3.1 General conditions for random double integrals

Let us define a sequence of general random double integrals.

Definition 3.1. Suppose that for each positive integer n, there are n non-negative numbers $\mu_1(n) \ge \cdots \ge \mu_n(n) \ge 0$. Let $\alpha_n \ge 0$ and $B_n > 0$ be real numbers. For each positive integer n, define

$$\mathbf{Q}_n = \mathbf{Q}(n, \alpha_n, B_n) := -\int_{\gamma_1 - \mathrm{i}\infty}^{\gamma_1 + \mathrm{i}\infty} \int_{\gamma_2 - \mathrm{i}\infty}^{\gamma_2 + \mathrm{i}\infty} \frac{e^{nB_n(z_1+z_2)}}{z_1^{n\alpha_n} \prod_{i=1}^n \sqrt{4z_1z_2 - \mu_i(n)}} \, \mathrm{d}z_2 \, \mathrm{d}z_1 \tag{3.1}$$

where γ_1 and γ_2 are any real positive constants satisfying $4\gamma_1\gamma_2 > \mu_1(n)$.

We consider large n asymptotics of $\mathbf{Q}_n$ under the following three conditions.

Condition 3.2. *There is $\delta > 0$ such that*

$$\alpha_n = \alpha + O(n^{-1-\delta}), \qquad B_n = B + O(n^{-1-\delta}) \tag{3.2}$$

for $\alpha \ge 0$ and $B > 0$.

Condition 3.3 (Regularity of measure)**.** *The empirical spectral distribution converges weakly in probability to a probability measure $\widehat{\mu}$, i.e.,*

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\mu_i(n)}(x)\mathrm{d}x \to \mathrm{d}\widehat{\mu}(x), \tag{3.3}$$

and $\widehat{\mu}$ satisfies the following properties:

- *$\widehat{\mu}$ is supported on a closed interval $[d_-, d_+]$.*
- *$\widehat{\mu}$ has a density that is positive on (d_-, d_+).*
- *The density of $\widehat{\mu}$ exhibits square-root decay at the upper edge, i.e., for some $c_{\widehat{\mu}} > 0$,*

$$\frac{\mathrm{d}\widehat{\mu}}{\mathrm{d}x}(x) = c_{\widehat{\mu}}\sqrt{d_+ - x}(1 + o(1)) \qquad \text{as } x \uparrow d_+. \tag{3.4}$$

Condition 3.4 (Rigidity)**.** *For a positive integer $k \in [1, n]$, let $\hat{k} := \min\{k, n+1-k\}$. Let g_k denote the "classical location" defined by the quantiles,*

$$\int_{g_k}^{\infty} \mathrm{d}\widehat{\mu} = \frac{1}{n}\left(k - \frac{1}{2}\right). \tag{3.5}$$

Then, for any $\epsilon > 0$,

$$|\mu_k(n) - g_k| \le \hat{k}^{-1/3} n^{-2/3+\epsilon} \tag{3.6}$$

for all $1 \le k \le n$ and for all n.

Note that the last two conditions imply that

$$\mu_1(n) \to d_+. \tag{3.7}$$

Remark 3.5 (Notational Remark 1)*.* Throughout the paper we use C or c in order to denote a constant that is independent of n. Even if the constant is different from one place to another, we may use the same notation C or c as long as it does not depend on n for the convenience of the presentation.

Remark 3.6 (Notational Remark 2)*.* We use standard notations $O(\cdot)$, $o(\cdot)$, $\ll$, and $\gg$ as $n \to \infty$.

In terms of the above notation, the partition function is given by (see Lemma 2.5)

$$Z_{N_1,N_2}(\beta) = \frac{\mathbf{Q}\left(N_2, \frac{N_1-N_2}{2N_2}, \frac{N_1}{\sqrt{N_2 N}}\beta\right)}{|S^{N_1-1}||S^{N_2-1}|} 2^{N_2} \left(\frac{\pi^2 N}{N_1^2 N_2 \beta^2}\right)^{(N-4)/4} \tag{3.8}$$

for $N_1 \ge N_2$, where $N = N_1 + N_2$ and $\mu_1 \ge \cdots \ge \mu_{N_2}$ are the eigenvalues of $\frac{1}{N_1} J^T J$. The eigenvalues satisfy Condition 3.3 and 3.4 with high probability; see Section 6.

3.2 Critical point

We write

$$\mathbf{Q}_n = -\int_{\gamma_1 - \mathrm{i}\infty}^{\gamma_1 + \mathrm{i}\infty} \int_{\gamma_2 - \mathrm{i}\infty}^{\gamma_2 + \mathrm{i}\infty} e^{nG(z_1,z_2)} \, \mathrm{d}z_2 \, \mathrm{d}z_1 \tag{3.9}$$

with

$$G(z_1, z_2) = B_n(z_1 + z_2) - \frac{1}{2n} \sum_{i=1}^{n} \log\left(4z_1 z_2 - \mu_i(n)\right) - \alpha_n \log z_1. \tag{3.10}$$

To evaluate the integral in (3.9) using the method of steepest-descent, we find the critical points of $G(z_1, z_2)$. We have

$$\partial_1 G = B_n - \frac{2z_2}{n} \sum_{i=1}^{n} \frac{1}{4z_1 z_2 - \mu_i(n)} - \frac{\alpha_n}{z_1}, \qquad \partial_2 G = B_n - \frac{2z_1}{n} \sum_{i=1}^{n} \frac{1}{4z_1 z_2 - \mu_i(n)}. \tag{3.11}$$

Hence the critical points satisfy the equations

$$z_1 - z_2 = \frac{\alpha_n}{B_n}, \qquad \frac{z_1}{n} \sum_{i=1}^{n} \frac{1}{4z_1 z_2 - \mu_i(n)} = \frac{B_n}{2}. \tag{3.12}$$

Taking the imaginary parts, we find that at the critical points,

$$\operatorname{Im} z_1 = \operatorname{Im} z_2, \qquad \frac{1}{n}\sum_{i=1}^{n} \frac{4|z_1|^2 \operatorname{Im} z_2 + \mu_i(n) \operatorname{Im} z_1}{|4z_1z_2 - \mu_i(n)|^2} = 0.$$

Since $\mu_i(n) \geq 0$, $\operatorname{Im} z_1 = \operatorname{Im} z_2 = 0$ at the critical points. Hence, all critical points, if exist, are real-valued.

We now look for real critical points. Due to the branch cut of G, we look only for real critical points (γ_1, γ_2) satisfying $4\gamma_1\gamma_2 > \mu_1(n)$, $\gamma_1 > 0$, and $\gamma_2 > 0$. We set $4\gamma_1\gamma_2 = \gamma$ and express the equations in terms of γ_1 and γ instead of γ_1 and γ_2:

$$\gamma_1 - \frac{\gamma}{4\gamma_1} = \frac{\alpha_n}{B_n}, \qquad \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\gamma - \mu_i(n)} = \frac{B_n}{2\gamma_1} \tag{3.13}$$

where $\gamma > \mu_1(n)$. The first equation is a quadratic equation of γ_1 for given γ, and hence there are two solutions. Only one of them is positive given by

$$\gamma_1 = \frac{\alpha_n + \sqrt{\alpha_n^2 + \gamma B_n^2}}{2B_n}. \tag{3.14}$$

This implies that

$$\gamma_2 = \frac{-\alpha_n + \sqrt{\alpha_n^2 + \gamma B_n^2}}{2B_n}. \tag{3.15}$$

Inserting (3.14) into the second equation of (3.13), we obtain an equation for γ given by (3.16) below. The next lemma proves the existence and the uniqueness of the solution.

Lemma 3.7. *The equation*

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1}{\gamma - \mu_i(n)} = \frac{B_n^2}{\alpha_n + \sqrt{\alpha_n^2 + \gamma B^2}} \tag{3.16}$$

has a unique solution in the interval $(\mu_1(n), \infty)$*.*

Proof. Let $L(\gamma)$ and $R(\gamma)$ be the left-hand side and right-hand side of (3.16), respectively. We observe that the function

$$f(\gamma) = \frac{c_1 + \sqrt{c_2 + \gamma}}{\gamma - \mu}$$

has the derivative

$$f'(\gamma) = \frac{-2c_1\sqrt{c_2 + \gamma} - 2c_2 - \gamma - \mu}{2\sqrt{c_2 + \gamma}(\gamma - \mu)^2}.$$

Hence, if $c_1, c_2 > 0$ and $\mu > 0$, then $h(\gamma)$ is a decreasing function of $\gamma \in (\mu, \infty)$. This shows that $\frac{L(\gamma)}{R(\gamma)}$ is a decreasing function of $\gamma \in (\mu_1(n), \infty)$. Since the equation (3.16) is equivalent to $\frac{L(\gamma)}{R(\gamma)} = 1$, if the solution exists in the interval $(\mu_1(n), \infty)$, it is unique in the same interval.

We now prove the existence. We first notice that $L(\gamma) \to \infty$ as $\gamma \downarrow \mu_1(n)$ and $R(\gamma)$ is bounded above. Furthermore, $L(\gamma) = O(\gamma^{-1})$ as $\gamma \to \infty$ and $R(\gamma) \geq C\gamma^{-\frac{1}{2}}$ for some $C > 0$ independent of γ. Thus,

$$\lim_{\gamma \downarrow \mu_1} \frac{L(\gamma)}{R(\gamma)} = +\infty, \qquad \lim_{\gamma \to \infty} \frac{L(\gamma)}{R(\gamma)} = 0. \tag{3.17}$$

Therefore, $\frac{L(\gamma)}{R(\gamma)} = 1$ has a unique solution in the interval $(\mu_1(n), \infty)$. □

In conclusion,

(i) there are no critical values of G with $\operatorname{Im} z_1 \neq 0$ or $\operatorname{Im} z_2 \neq 0$,

(ii) there is a unique critical value (γ_1, γ_2) such that γ_1 and γ_2 are real and positive, and $4\gamma_1\gamma_2 > \mu_1(n)$,

(iii) the critical value (γ_1, γ_2) is given by the formulas (3.14) and (3.15) where $\gamma \in (\mu_1(n), \infty)$ satisfies the equation (3.16).

Note that $(\gamma_1, \gamma_2) = (\gamma_1(n), \gamma_2(n))$ depends on n since G depends on n.

3.3 Critical temperature

We discuss how we find the critical temperature formally from the critical point.

Recall Condition 3.2 and Condition 3.3. Recall that d_+ denotes the rightmost point of the support of $\widehat{\mu}$. If γ in (3.16) is $O(1)$ distance to the right of d_+, then we may approximate the equation (3.16) by the n-independent equation

$$\int_{\mathbb{R}} \frac{1}{z-x} \mathrm{d}\widehat{\mu}(x) = \frac{B^2}{\alpha + \sqrt{\alpha^2 + zB^2}}. \tag{3.18}$$

Call the left-hand side and right-hand side by $L_\infty(z)$ and $R_\infty(z)$, respectively. Note that $L_\infty(z)$ is well-defined for all real-valued $z \geq d_+$ (and also non-real z). In particular, the integral converges when $z = d_+$ due to Condition 3.3. By the same calculation of the proof of Lemma 3.7, $\frac{L_\infty(z)}{R_\infty(z)}$ is a decreasing function of $z \in (d_+, \infty)$. As before, $\frac{L_\infty(z)}{R_\infty(z)} \to 0$ as $z \to \infty$. However, unlike the previous lemma, the limit

$$\lim_{z \downarrow d_+} \frac{L_\infty(z)}{R_\infty(z)} = \frac{L_\infty(d_+)}{R_\infty(d_+)}. \tag{3.19}$$

is finite. Hence the solution z to the equation (3.18) exists in (d_+, ∞) only if $L_\infty(d_+) > R_\infty(d_+)$, i.e., if

$$\int_{\mathbb{R}} \frac{1}{d_+ - x} \mathrm{d}\widehat{\mu}(x) > \frac{\sqrt{\alpha^2 + d_+ B^2} - \alpha}{d_+}. \tag{3.20}$$

Note that the left integral is a finite positive number due to the square-root vanishing assumption in Condition 3.3. Considered as a function of B, the right-hand side $f(B)$ is an increasing function of B, $f(0) = 0$, and $f(B) \to +\infty$ as $B \to \infty$. Hence the above inequality holds for all $B < B_c$ where B_c is defined by the equation

$$\int_{\mathbb{R}} \frac{1}{d_+ - x} \mathrm{d}\widehat{\mu}(x) = \frac{\sqrt{\alpha^2 + d_+ B_c^2} - \alpha}{d_+}. \tag{3.21}$$

Thus, we define the following critical value of B.

Definition 3.8. Define

$$B_c = \sqrt{d_+(\mathsf{s}(d_+))^2 + 2\alpha \mathsf{s}(d_+)} \quad \text{where } \mathsf{s}(z) := \int_{\mathbb{R}} \frac{1}{z-x} \mathrm{d}\widehat{\mu}(x). \tag{3.22}$$

The above discussion implies the following:

(a) For $0 < B < B_c$, there is a unique solution z_c in (d_+, ∞) to the equation (3.18).

(b) For $B > B_c$, there are no solutions to the equation (3.18) in (d_+, ∞), and

We will show in Section 4 that for the case (a), γ in Lemma 3.7 is indeed close to z_c. On the other hand, we will see in Section 5 that for the case (b), the assumption that the point γ in Lemma 3.7 is $O(1)$ away from d_+ is not true. This means that (3.18) is not a good approximation to the equation (3.16).

3.4 Truncation of the double integral

The following lemma gives an estimate on the double integral (3.9) outside a small disk of radius $N^{-\frac{1}{2}+\epsilon}$ about the point (γ_1, γ_2). This result is used in later sections. The lemma does not require that (γ_1, γ_2) is the critical point.

Lemma 3.9. *Let $\gamma_1 = \gamma_1(n)$, $\gamma_2 = \gamma_2(n)$ be any positive real numbers such that $4\gamma_1(n)\gamma_2(n) > \mu_1(n)$ for all n. Suppose that there is a constant $C' > 0$ such that $4\gamma_1(n)\gamma_2(n) - \mu_n(n) \le C'$ for all n. Then, for any $\epsilon > 0$ and any $\Omega \subset \{(y_1, y_2) \in \mathbb{R}^2 : y_1^2 + y_2^2 \ge n^{-1+2\epsilon}\}$,*

$$\int_\Omega \exp\left[n \operatorname{Re}\left(G(\gamma_1 + \mathrm{i}y_1, \gamma_2 + \mathrm{i}y_2) - G(\gamma_1, \gamma_2)\right)\right] \mathrm{d}y_2\, \mathrm{d}y_1 \le Ce^{-n^\epsilon} \tag{3.23}$$

with high probability.

Proof. We write $\mu_i(n) = \mu_i$ in this proof for a notational convenience. For $y_1, y_2 \in \mathbb{R}$, from the definition of G,

$$\begin{aligned}
&\operatorname{Re}\left[G(\gamma_1 + \mathrm{i}y_1, \gamma_2 + \mathrm{i}y_2) - G(\gamma_1, \gamma_2)\right] \\
&= -\frac{1}{4n}\sum_{i=1}^n \log\left[\left(1 - \frac{4y_1y_2}{4\gamma_1\gamma_2 - \mu_i}\right)^2 + 16\left(\frac{\gamma_2 y_1 + \gamma_1 y_2}{4\gamma_1\gamma_2 - \mu_i}\right)^2\right] - \frac{\alpha_n}{2}\log\left(1 + \frac{y_1^2}{\gamma_1^2}\right).
\end{aligned}$$

Consider the case $y_1y_2 \ge 0$. Then

$$\begin{aligned}
\left(1 - \frac{4y_1y_2}{4\gamma_1\gamma_2 - \mu_i}\right)^2 + 16\left(\frac{\gamma_2 y_1 + \gamma_1 y_2}{4\gamma_1\gamma_2 - \mu_i}\right)^2 &\ge 1 - \frac{8y_1y_2}{4\gamma_1\gamma_2 - \mu_i} + 16\left(\frac{\gamma_2 y_1 + \gamma_1 y_2}{4\gamma_1\gamma_2 - \mu_i}\right)^2 \\
&\ge 1 + 8\left(\frac{\gamma_2 y_1 + \gamma_1 y_2}{4\gamma_1\gamma_2 - \mu_i}\right)^2 \ge 1 + c(y_1^2 + y_2^2),
\end{aligned}$$

where we used the fact that

$$8(\gamma_2 y_1 + \gamma_1 y_2)^2 \ge 32\gamma_1\gamma_2 y_1 y_2 \ge 8y_1y_2(4\gamma_1\gamma_2 - \mu_i)$$

for the second inequality and that $|4\gamma_1\gamma_2 - \mu_i| < C$ uniformly for all i in the third inequality.

For the case $y_1y_2 < 0$, we consider the following sub-cases:

(i) If $\gamma_2|y_1| > 2\gamma_1|y_2|$, then $2|\gamma_2 y_1 + \gamma_1 y_2| \ge |\gamma_2 y_1|$, and hence

$$\left(1 - \frac{4y_1y_2}{4\gamma_1\gamma_2 - \mu_i}\right)^2 + 16\left(\frac{\gamma_2 y_1 + \gamma_1 y_2}{4\gamma_1\gamma_2 - \mu_i}\right)^2 \ge 1 + 4\left(\frac{\gamma_2 y_1}{4\gamma_1\gamma_2 - \mu_i}\right)^2 \ge 1 + c(y_1^2 + y_2^2).$$

(ii) If $\gamma_2|y_1| < \frac{1}{2}\gamma_1|y_2|$, then $2|\gamma_2 y_1 + \gamma_1 y_2| \geq |\gamma_1 y_2|$, and hence

$$\left(1 - \frac{4y_1y_2}{4\gamma_1\gamma_2 - \mu_i}\right)^2 + 16\left(\frac{\gamma_2 y_1 + \gamma_1 y_2}{4\gamma_1\gamma_2 - \mu_i}\right)^2 \geq 1 + 4\left(\frac{\gamma_1 y_2}{4\gamma_1\gamma_2 - \mu_i}\right)^2 \geq 1 + c(y_1^2 + y_2^2).$$

(iii) If $\frac{1}{2}\gamma_1|y_2| \leq \gamma_2|y_1| \leq 2\gamma_1|y_2|$, then

$$\left(1 - \frac{4y_1y_2}{4\gamma_1\gamma_2 - \mu_i}\right)^2 + 16\left(\frac{\gamma_2 y_1 + \gamma_1 y_2}{4\gamma_1\gamma_2 - \mu_i}\right)^2 \geq 1 - \frac{8y_1y_2}{4\gamma_1\gamma_2 - \mu_i} \geq 1 + c(y_1^2 + y_2^2),$$

since $-y_1y_2 = |y_1y_2| \geq c'(y_1^2 + y_2^2)$ for some $c' > 0$.

Thus, for all $y_1, y_2 \in \mathbb{R}$,

$$\mathrm{Re}\left[G(\gamma_1 + \mathrm{i}y_1, \gamma_2 + \mathrm{i}y_2) - G(\gamma_1, \gamma_2)\right] \leq -\frac{1}{4}\log\left(1 + c(y_1^2 + y_2^2)\right). \tag{3.24}$$

Now note that

$$\log\left(1 + c(y_1^2 + y_2^2)\right) \geq \log\left(1 + cn^{-1+2\epsilon}\right) \geq \frac{1}{2}cn^{-1+2\epsilon} \quad \text{for } y_1^2 + y_2^2 \in [n^{-1+2\epsilon}, n] \tag{3.25}$$

and

$$\log\left(1 + c(y_1^2 + y_2^2)\right) \geq \log\left(c(y_1^2 + y_2^2)\right) \quad \text{for } y_1^2 + y_2^2 > n. \tag{3.26}$$

Hence,

$$\begin{aligned}&\int_\Omega \exp\left[n\,\mathrm{Re}\left(G(\gamma_1 + \mathrm{i}y_1, \gamma_2 + \mathrm{i}y_2) - G(\gamma_1, \gamma_2)\right)\right] \mathrm{d}y_2\,\mathrm{d}y_1 \\ &\leq Ce^{-\frac{c}{8}n^{2\epsilon}} \int_{n^{-1/2+\epsilon}}^{n^{1/2}} r\mathrm{d}r + C\int_{n^{1/2}}^{\infty} (cr^2)^{-n/4} r\mathrm{d}r = O(e^{-c'n^{2\epsilon}}) + O(n^{-n/4}).\end{aligned} \tag{3.27}$$

This proves the lemma. □

4 High temperature

We consider the asymptotics of the double integral $\mathbf{Q}_n$ in (3.9) when $B < B_c$, where B_c is defined in (3.22). We assume Conditions 3.2, 3.3 and 3.4 throughout this section. Recall that

$$G(z_1, z_2) = B_n(z_1 + z_2) - \frac{1}{2n}\sum_{i=1}^{n} \log\left(4z_1z_2 - \mu_i(n)\right) - \alpha_n \log z_1. \tag{4.1}$$

As in our previous works [7, 8], we show that when $B < B_c$, the critical point of the random function G is close to the critical point of a deterministic function.

Define

$$G_\infty(z_1, z_2) = B(z_1 + z_2) - \frac{1}{2}\int_{\mathbb{R}} \log\left(4z_1z_2 - x\right) \mathrm{d}\widehat{\mu}(x) - \alpha \log z_1\,. \tag{4.2}$$

Then

$$\partial_1 G_\infty = B - \int_{\mathbb{R}} \frac{2z_2}{4z_1z_2 - x}\mathrm{d}\widehat{\mu}(x) - \frac{\alpha}{z_1}, \qquad \partial_2 G_\infty = B - \int_{\mathbb{R}} \frac{2z_1}{4z_1z_2 - x}\mathrm{d}\widehat{\mu}(x). \tag{4.3}$$

When $B < B_c$, the critical point of G_∞ is given by

$$z_1^c = \frac{\alpha + \sqrt{\alpha^2 + z_c B^2}}{2B}, \qquad z_2^c = \frac{-\alpha + \sqrt{\alpha^2 + z_c B^2}}{2B}, \tag{4.4}$$

where z_c is the solution to the equation

$$\int_{\mathbb{R}} \frac{1}{z - x} \mathrm{d}\widehat{\mu}(x) = \frac{B^2}{\alpha + \sqrt{\alpha^2 + zB^2}} \tag{4.5}$$

satisfying $z_c \in (d_+, \infty)$. We discussed in Subsection 3.3 that when $B < B_c$, there is unique z_c.

We start with the following lemma on the differences between the derivatives of G and G_∞, which is analogous to Lemma 5.1 of [7].

Lemma 4.1. *Fix $\theta > 0$ and set $B_\theta = \{(z_1, z_2) : -\theta < \mathrm{Re}(4z_1 z_2) < d_+ + \theta,\ -\theta < \mathrm{Im}(z_1 z_2) < \theta\}$. Then the following hold.*

(i) For every $\epsilon > 0$ and each multi-index $m = (m_1, m_2)$,

$$\partial^m G(z_1, z_2) - \partial^m G_\infty(z_1, z_2) = O(n^{-1+\epsilon}) \tag{4.6}$$

uniformly on any compact subset of the region $\mathbb{C}^2 \setminus B_\theta$.

(ii) For every multi-index m, $\partial^m G(z_1, z_2) = O(1)$ uniformly on any compact subset of $\mathbb{C}^2 \setminus B_\theta$.

Proof. (i) Let

$$\widetilde{G}(z_1, z_2) = B(z_1 + z_2) - \frac{1}{2n} \sum_{i=1}^{n} \log(4z_1 z_2 - g_i) - \alpha \log z_1 \tag{4.7}$$

where g_i is the classical location of the i-th eigenvalue defined in (3.5). Then, from Condition 3.2,

$$\begin{aligned} |G(z_1, z_2) - \widetilde{G}(z_1, z_2)| &= \left| (B_n - B)(z_1 + z_2) - \frac{1}{2n} \sum_{i=1}^{n} \log\left(\frac{4z_1 z_2 - \mu_i(n)}{4z_1 z_2 - g_i} \right) - (\alpha_n - \alpha) \log z \right| \\ &\le \frac{1}{2n} \sum_{i=1}^{n} \log(1 + C|\mu_i - g_i|) + C' n^{-1-\delta} \end{aligned}$$

uniformly on a compact subset of $\mathbb{C}^2 \setminus B_\theta$ since $|4z_1 z_2 - g_i| \ge c$. Hence, from the rigidity, Condition 3.4,

$$|G(z_1, z_2) - \widetilde{G}(z_1, z_2)| \le \frac{C}{2n} \sum_{i=1}^{n} |\mu_i - g_i| + \frac{C'}{n^{1+\delta}} \le \frac{Cn^\epsilon}{n}$$

in any compact subset of $\mathbb{C}^2 \setminus B_\theta$. We now compare $\widetilde{G}(z_1, z_2)$ with $G_\infty(z_1, z_2)$. For $2 \le i \le n-1$,

$$\int_{g_i}^{g_{i-1}} \log(4z_1 z_2 - x)\, \mathrm{d}\widehat{\mu}(x) \le \frac{1}{n} \log(4z_1 z_2 - g_i) \le \int_{g_{i+1}}^{g_i} \log(4z_1 z_2 - x)\, \mathrm{d}\widehat{\mu}(x).$$

Summing over the index i and using the trivial estimates

$$\int_{g_2}^{d_+} \log(4z_1 z_2 - x) \mathrm{d}\widehat{\mu}(x) = O(n^{-1}), \qquad \int_0^{g_{n-2}} \log(4z_1 z_2 - x) \mathrm{d}\widehat{\mu}(x) = O(n^{-1}),$$

and $\log(4z_1z_2 - g_i) = O(1)$ for any compact subset of $\mathbb{C}^2 \setminus B_\theta$, we find that $\widetilde{G}(z_1, z_2) - G_\infty(z_1, z_2) = O(n^{-1})$. Hence, $G(z_1, z_2) - G_\infty(z_1, z_2) = O(n^{-1+\epsilon})$. For the derivatives, the function $\log(4z_1z_2 - x)$ is replaced by $\frac{1}{(4z_1z_2-x)^k}$ for positive integers k, and the proof is almost same.

(ii) can be proved in a similar manner since, for any compact subset of $\mathbb{C}^2 \setminus B_\theta$, $\log(4z_1z_2 - g_i) = O(1)$ and $\frac{1}{(4z_1z_2-x)^k} = O(1)$ for positive integers k. □

We now compare the critical point (γ_1, γ_2) of G and the critical point (z_1^c, z_2^c) of G_∞. Recall that (γ_1, γ_2) depends on n.

Lemma 4.2. *For any $\epsilon > 0$, the following hold.*

(i) We have

$$\gamma_1 - z_1^c = O(n^{-1+\epsilon}), \qquad \gamma_2 - z_2^c = O(n^{-1+\epsilon}). \tag{4.8}$$

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

$$4\gamma_1\gamma_2 - \mu_1(n) > c \text{ and } 4\gamma_1\gamma_2 - d_+ > c. \tag{4.9}$$

(iii) We have

$$G(\gamma_1, \gamma_2) = G(z_1^c, z_2^c) + O(n^{-2+\epsilon}) \tag{4.10}$$

and for any multi-index $m = (m_1, m_2)$ satisfying $|m| > 0$,

$$\partial^m G(\gamma_1, \gamma_2) = \partial^m G(z_1^c, z_2^c) + O(n^{-1+\epsilon}). \tag{4.11}$$

Proof. (i) We first show that $\gamma = 4\gamma_1\gamma_2$ and $z_c = 4z_1^cz_2^c$ satisfy $\gamma - z_c = O(n^{-1+\epsilon})$. The value γ is determined by the equation in (3.16), which can be written as $L(x) = R(x)$ where

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

Similarly, the point z_c is a solution of the equation $L_\infty(x) = R(x)$ where

$$L_\infty(x) := \int_{\mathbb{R}} \frac{\mathrm{d}\widehat{\mu}(y)}{x - y}, \qquad R_\infty(x) := \frac{B^2}{\alpha + \sqrt{\alpha^2 + xB^2}}.$$

We showed in the proof of Lemma 3.7 that $F(x) := \frac{R(x)}{L(x)}$ satisfies $F'(x) < 0$ for all $x > \mu_1(n)$. The same calculation shows that $F_\infty(x) := \frac{R_\infty(x)}{L_\infty(x)}$ satisfies $F_\infty'(x) < 0$ for all $x > d_+$. Since

$$L(x) = \frac{1}{2}(B_n - \partial_2 G(1, x)), \qquad L_\infty(x) = \frac{1}{2}(B - \partial_2 G_\infty(1, x)),$$

we find from Lemma 4.1 (i) that $F(x) = F_\infty(x) + O(n^{-1+\frac{\epsilon}{2}})$ uniformly for x in any compact subset of the interval (d_+, ∞). Note that we used $\epsilon/2$ when we apply Lemma 4.1. Recall that $z_c > d_+$. Hence, $F(z_c) = F_\infty(z_c) + O(N^{-1+\frac{\epsilon}{2}}) = 1 + O(n^{-1+\frac{\epsilon}{2}})$. By Taylor series,

$$F(z_c \pm n^{-1+\epsilon}) = F_\infty(z_c \pm n^{-1+\epsilon}) + O(n^{-1+\frac{\epsilon}{2}}) = 1 \pm F_\infty'(z_c)n^{-1+\epsilon} + O(n^{-2+2\epsilon}) + O(n^{-1+\frac{\epsilon}{2}}).$$

Since $F'_\infty(z_c) < 0$, we find that

$$F(z_c - n^{-1+\epsilon}) < 1, \qquad F(z_c - n^{-1+\epsilon}) > 1.$$

This implies that

$$\gamma \in (z_c - n^{-1+\epsilon}, z_c + n^{-1+\epsilon}).$$

We obtain (i) since γ_1 and γ_2 are given in terms of γ by (3.14) and (3.15), and respectively, and z_1^c and z_2^c are given by (4.4) in terms of z_c.

(ii) follows from (i).

(iii) From the Taylor expansion and the bounds in Lemma 4.1 (ii),

$$\begin{aligned} G(z_1^c, z_2^c) &= G(\gamma_1, \gamma_2) + \partial_1 G(\gamma_1, \gamma_2)(z_1^c - \gamma_1) + \partial_2 G(\gamma_1, \gamma_2)(z_2^c - \gamma_2) + O(n^{-2+\epsilon}) \\ &= G(\gamma_1, \gamma_2) + O(n^{-2+\epsilon}) \end{aligned}$$

and

$$\partial^m G(z_1^c, z_2^c) = \partial^m G(\gamma_1, \gamma_2) + O(n^{-1+\epsilon})$$

for any multi-index m satisfying $|m| > 0$. This completes the proof of the lemma. □

We evaluate the integral (3.9) using the method of steepest-descent.

Lemma 4.3. *Let $B < B_c$ for B_c defined in* (3.22). *Then for any $\epsilon > 0$,*

$$\mathbf{Q}_n = e^{nG(\gamma_1,\gamma_2)} \frac{\pi}{n\sqrt{D(\gamma_1, \gamma_2)}} \left(1 + O(n^{-1+\epsilon})\right) \tag{4.12}$$

where $D(\gamma_1, \gamma_2)$ is the discriminant

$$D(\gamma_1, \gamma_2) = \partial_1^2 G(\gamma_1, \gamma_2) \cdot \partial_2^2 G(\gamma_1, \gamma_2) - \left(\partial_1 \partial_2 G(\gamma_1, \gamma_2)\right)^2.$$

Proof. Changing the variables,

$$\mathbf{Q}_n = \frac{1}{n} e^{nG(\gamma_1,\gamma_2)} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left[n\left(G(\gamma_1 + \mathrm{i}\frac{t_1}{\sqrt{n}}, \gamma_2 + \mathrm{i}\frac{t_2}{\sqrt{n}}) - G(\gamma_1, \gamma_2)\right)\right] \mathrm{d}t_2 \, \mathrm{d}t_1.$$

Lemma 3.9 shows that, the part of the last double integral over the region $\mathbb{R}^2 \setminus [-n^\epsilon, n^\epsilon]^2$ is $O(e^{-n^\epsilon})$. On the other hand, for $|t_1|, |t_2| \le n^\epsilon$,

$$\begin{aligned} &G(\gamma_1 + \mathrm{i}\frac{t_1}{\sqrt{n}}, \gamma_2 + \mathrm{i}\frac{t_2}{\sqrt{n}}) - G(\gamma_1, \gamma_2) \\ &= -\frac{1}{2n}\left(\partial_1^2 G(\gamma_1, \gamma_2) t_1^2 + 2\partial_1\partial_2 G(\gamma_1, \gamma_2) t_1 t_2 + \partial_2^2 G(\gamma_1, \gamma_2) t_2^2\right) \\ &\quad - \frac{\mathrm{i}}{6n^{\frac{3}{2}}}\left(\partial_1^3 G(\gamma_1, \gamma_2) t_1^3 + 3\partial_1^2\partial_2 G(\gamma_1, \gamma_2) t_1^2 t_2 + 3\partial_1\partial_2^2 G(\gamma_1, \gamma_2) t_1 t_2^2 + \partial_2^3 G(\gamma_1, \gamma_2) t_2^3\right) + O(n^{-2+4\epsilon}) \\ &=: -\frac{X_2(t_1, t_2)}{n} - \frac{\mathrm{i} X_3(t_1, t_2)}{n^{\frac{3}{2}}} + O(n^{-2+4\epsilon}) \end{aligned}$$

where we used Lemma 4.2 (i) and Lemma 4.1 (ii) for the error estimate. Hence,

$$\int_{-n^\epsilon}^{n^\epsilon}\int_{-n^\epsilon}^{n^\epsilon} \exp\left[n\left(G(\gamma_1+\mathrm{i}\frac{t_1}{\sqrt{n}},\gamma_2+\mathrm{i}\frac{t_2}{\sqrt{n}})-G(\gamma_1,\gamma_2)\right)\right]\,\mathrm{d}t_2\,\mathrm{d}t_1$$
$$=\int_{-n^\epsilon}^{n^\epsilon}\int_{-n^\epsilon}^{n^\epsilon} e^{-X_2(t_1,t_2)}\mathrm{d}t_2\,\mathrm{d}t_1-\mathrm{i}\int_{-n^\epsilon}^{n^\epsilon}\int_{-n^\epsilon}^{n^\epsilon}\frac{X_3(t_1,t_2)}{\sqrt{n}}e^{-X_2(t_1,t_2)}\mathrm{d}t_2\,\mathrm{d}t_1+O(n^{-1+6\epsilon}).$$

Since

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

the integral in the middle vanishes. On the other hand, from the estimate $\int_{n^\epsilon}^\infty e^{-t^2}\mathrm{d}t=O(n^{-\epsilon}e^{-n^{2\epsilon}})$, we obtain that

$$\int_{-n^\epsilon}^{n^\epsilon}\int_{-n^\epsilon}^{n^\epsilon} e^{-X_2(t_1,t_2)}\mathrm{d}t_2\,\mathrm{d}t_1=\frac{\pi}{\sqrt{D(\gamma_1,\gamma_2)}}+O(n^{-1+6\epsilon}).$$

Thus, we obtain the lemma. □

The following is the main result for the double integral $\mathbf{Q}_n$ when $B<B_c$.

Proposition 4.4 (Random double integral for high temperature)**.** *Assume Conditions 3.2, 3.3 and 3.4. Define*

$$\widehat{H}(z):=\int_{\mathbb{R}}\log(z-x)\mathrm{d}\widehat{\mu}(x)\,. \tag{4.13}$$

Suppose that B in Condition 3.2 satisfies $0<B<B_c$ where B_c is defined in (3.22)*. Then, setting z_c be the unique solution of the equation* (3.18)*,*

$$\widehat{H}'(z_c)=\frac{B^2}{\alpha+\sqrt{\alpha^2+z_cB^2}},\qquad z_c\in(d_+,\infty), \tag{4.14}$$

we have for every $\epsilon>0$,

$$\frac{1}{n}\log\mathbf{Q}_n=\widehat{A}-\frac{1}{2n}\left[\sum_{i=1}^n\log\left(z_c-\mu_i\right)-n\widehat{H}(z_c)\right]-\frac{\log n}{n}+\frac{1}{2n}\log\left(\frac{\pi^2}{\widehat{D}}\right)+O(n^{-2+\epsilon}) \tag{4.15}$$

where

$$\begin{aligned}\widehat{A}&=\sqrt{\alpha^2+z_cB^2}-\alpha\log\left(\frac{\alpha+\sqrt{\alpha^2+z_cB^2}}{2B}\right)-\frac{1}{2}\widehat{H}(z_c),\\ \widehat{D}&=-8\alpha\widehat{H}''(z_c)-8z_c\widehat{H}'(z_c)\widehat{H}''(z_c)-4(\widehat{H}'(z_c))^2.\end{aligned} \tag{4.16}$$

Proof. From Lemma 4.3,

$$\frac{1}{n}\log\mathbf{Q}_n=G(\gamma_1,\gamma_2)-\frac{\log n}{n}+\frac{1}{2n}\log\left(\frac{\pi^2}{D(\gamma_1,\gamma_2)}\right)+O(n^{-2+\epsilon}).$$

Using Lemma 4.2 (iii), we write

$$G(\gamma_1,\gamma_2)=G(z_1^c,z_2^c)+O(n^{-2+\epsilon})=G_\infty(z_1^c,z_2^c)+[G(z_1^c,z_2^c)-G_\infty(z_1^c,z_2^c)]+O(n^{-2+\epsilon}).$$

We have

$$G(z_1^c, z_2^c) - G_\infty(z_1^c, z_2^c) = -\frac{1}{2n}\left[\sum_{i=1}^n \log{(z_c - \mu_i)} - n\int_{\mathbb{R}} \log{(z_c - x)}\, \mathrm{d}\widehat{\mu}(x)\right] + O(n^{-1-\delta}).$$

We also have

$$G_\infty(z_1^c, z_2^c) = \sqrt{\alpha^2 + z_c B^2} - \frac{1}{2}\int_{\mathbb{R}} \log{(z_c - x)}\, \mathrm{d}\widehat{\mu}(x) - \alpha \log\left(\frac{\alpha + \sqrt{\alpha^2 + z_c B^2}}{2B}\right) = \widehat{A}\,.$$

It remains to compare $D(\gamma_1, \gamma_2)$ with $\widehat{D}$. Using Lemma 4.2 (iii) and Lemma 4.1 (i),

$$D(\gamma_1, \gamma_2) = D_\infty(z_1^c, z_2^c) + O(n^{-1+\epsilon})$$

where

$$D_\infty(z_1^c, z_2^c) := \partial_1^2 G_\infty(z_1^c, z_2^c) \cdot \partial_2^2 G_\infty(z_1^c, z_2^c) - \big(\partial_1\partial_2 G_\infty(z_1^c, z_2^c)\big)^2.$$

From direct computation,

$$D_\infty(z_1^c, z_2^c) = -8\alpha \widehat{H}''(z_c) - 8z_c \widehat{H}'(z_c)\widehat{H}''(z_c) - 4(\widehat{H}'(z_c))^2 = \widehat{D}.$$

This completes the proof. □

5 Low temperature

In this section, we consider the asymptotics of

$$e^{-nG(\gamma_1,\gamma_2)}\mathbf{Q}_n = -\int_{\gamma_1 - \mathrm{i}\infty}^{\gamma_1 + \mathrm{i}\infty}\int_{\gamma_2 - \mathrm{i}\infty}^{\gamma_2 + \mathrm{i}\infty} \exp\left[n\left(G(z_1, z_2) - G(\gamma_1, \gamma_2)\right)\right]\, \mathrm{d}z_2\, \mathrm{d}z_1\,. \tag{5.1}$$

when $B > B_c$. We assume Conditions 3.2, 3.3 and 3.4 throughout this section.

Unlike the previous section, when $B > B_c$, the critical point (γ_1, γ_2) of G is not approximated by the critical point of G_∞. Indeed, we showed in Subsections 3.2 and 3.3 that G_∞ has no critical point when $B > B_c$, while (γ_1, γ_2) exists for all B. We show in Lemma 5.3 below that $\gamma = 4\gamma_1\gamma_2$ is actually close to the branch point $\mu_1(n)$. Due to this fact, the control of the double integral becomes subtle. We had a similar situation for a random single integral in [7] for the usual SSK model. In this paper, we have a double integral, and this brings an additional difficulty. In particular, the symmetry we used in [7], which simplified the analysis, is no longer valid. In the below, we will choose the integration contours in a certain explicit way and show that it is possible to reduce the double integral to the product of two single integrals plus an error. One of the single integral is trivial and the other single integral has a certain symmetry that can be used to simplify the method of steepest-descent in a manner similar to the analysis of [7].

In Subsections 5.1–5.4, we prove the following lemma. The conclusion of this section is given in Subsection 5.5.

Lemma 5.1. *Assume Conditions 3.2, 3.3 and 3.4. Suppose that B in Condition 3.2 satisfies $B > B_c$. Let (γ_1, γ_2) be the critical point of G given by* (3.14) *and* (3.15)*. Then, for every $\epsilon > 0$, there is a constant $C > 0$ such that*

$$Cn^{-\frac{3}{2}-\epsilon} \le e^{-nG(\gamma_1,\gamma_2)}\mathbf{Q}_n \le Cn^{-\frac{1}{2}}. \tag{5.2}$$

Remark 5.2 (Notational Remark). In order to lighten up the notations, we will write μ_i for $\mu_i(n)$ in the rest of this section. It should be understood that μ_i depends on n.

5.1 A priori estimate on γ

We begin by approximating $\gamma = 4\gamma_1\gamma_2$ and introducing a priori estimates that will be used in this section.

Lemma 5.3. *For any $0 < \epsilon < 1$, the solution γ in Lemma 3.7,*

$$\frac{1}{n}\sum_{i=1}^{n} \frac{1}{\gamma - \mu_i} = \frac{B_n^2}{\alpha_n + \sqrt{\alpha_n^2 + \gamma B_n^2}}, \qquad \gamma > \mu_1, \tag{5.3}$$

satisfies the inequality

$$\frac{d_+^{1/2}}{2Bn} \le \gamma - \mu_1 \le \frac{n^\epsilon}{n}. \tag{5.4}$$

Proof. We follow the proof of Lemma 6.1 in [7]. Define $L(\gamma)$ and $R(\gamma)$ to be the left-hand side and the right-hand side of (5.3), respectively. The equation (5.3) is equivalent to the equation $\frac{L(\gamma)}{R(\gamma)} = 1$. Since $\mu_1 \to d_+$ and $B_n \to B$,

$$\mu_1 + \frac{d_+^{1/2}}{2Bn} \ge \frac{d_+}{2}.$$

Since $L(\gamma) \ge \frac{1}{n(\gamma - \mu_1)}$ for $\gamma > \mu_1$, we find that

$$L\big(\mu_1 + \frac{d_+^{1/2}}{2Bn}\big) \ge \frac{2B}{d_+^{1/2}} \ge \frac{B_n^2}{\sqrt{\big(\mu_1 + \frac{d_+^{1/2}}{2Bn}\big)B_n^2}} \ge R\big(\mu_1 + \frac{d_+^{1/2}}{2Bn}\big). \tag{5.5}$$

Since $L(x)/R(x)$ is a decreasing function of x (see the proof of Lemma 3.7), this implies the lower bound of (5.4).

The upper bound is proved if we show that $L(\mu_1 + n^{-1+4\epsilon}) < R(\mu_1 + n^{-1+4\epsilon})$ for any $0 < \epsilon < \frac{1}{4}$. From Condition 3.4, $|\mu_i - g_i| \le n^{-2/3}$ for $n^{3\epsilon} \le i \le n - n^{3\epsilon}$. For such i, we note that $d_+ - g_i \ge cn^{-2/3+2\epsilon}$. Since $\mu_1 = d_+ + O(n^{-2/3+\epsilon})$,

$$\frac{1}{n}\sum_{i=n^{3\epsilon}}^{n-n^{3\epsilon}} \frac{1}{\mu_1 + n^{-1+4\epsilon} - \mu_i} = \frac{1}{n}\sum_{i=n^{3\epsilon}}^{n-n^{3\epsilon}} \frac{1}{d_+ - g_i}\left(1 + O(n^{-\epsilon})\right).$$

Approximating the last sum by an integral as in the proof of Lemma 4.1, we find that

$$\left| \frac{1}{n}\sum_{i=n^{3\epsilon}}^{n-n^{3\epsilon}} \frac{1}{\mu_1 + n^{-1+4\epsilon} - \mu_i} - \int_{d_-}^{d_+} \frac{\mathrm{d}\widehat{\mu}(x)}{d_+ - x} \right| = O(n^{-1/3+\epsilon}).$$

(See also Equations (6.6) and (6.7) in [7].) For $1 \le i < n^{3\epsilon}$, since $\mu_1 \ge \mu_i$,

$$\frac{1}{n}\sum_{i=1}^{n^{3\epsilon}-1} \frac{1}{\mu_1 + n^{-1+4\epsilon} - \mu_i} = O(n^{-\epsilon}).$$

Finally, for $n - n^{3\epsilon} < i \le n$, since $\mu_1 - \mu_i > c > 0$,

$$\frac{1}{n}\sum_{i=n-n^{3\epsilon}+1}^{n} \frac{1}{\mu_1 + n^{-1+4\epsilon} - \mu_i} = O(n^{-1+3\epsilon}).$$

Combining the estimates, we find that

$$L(\mu_1 + n^{-1+4\epsilon}) = \int_{d_-}^{d_+} \frac{\mathrm{d}\widehat{\mu}(x)}{d_+ - x} + O(n^{-\epsilon}). \tag{5.6}$$

On the other hand, since $\mu_1 + n^{-1+4\epsilon} \to d_+$,

$$R(\mu_1 + n^{-1+4\epsilon}) \to \frac{B^2}{\alpha + \sqrt{\alpha^2 + d_+ B^2}} = \frac{\sqrt{\alpha^2 + d_+ B^2} - \alpha}{d_+}. \tag{5.7}$$

From the definition of B_c in (3.21),

$$\frac{\sqrt{\alpha^2 + d_+ B^2} - \alpha}{d_+} > \frac{\sqrt{\alpha^2 + d_+ B_c^2} - \alpha}{d_+} = \int_{d_-}^{d_+} \frac{\mathrm{d}\widehat{\mu}(x)}{d_+ - x}.$$

Hence

$$R(\mu_1 + n^{-1+4\epsilon}) > L(\mu_1 + n^{-1+4\epsilon}) + c \tag{5.8}$$

for some $c > 0$ for all large enough n. This proves the lemma. □

Since γ is well approximaed by μ_1 and μ_1 is close to d_+, heuristically,

$$\frac{1}{n}\sum_{i=1}^{n} \log(\gamma - \mu_i) \approx \frac{1}{n}\sum_{i=1}^{n} \left[\log(d_+ - \mu_i) + \frac{\gamma - d_+}{d_+ - \mu_i} \right] \approx \widehat{H}(d_+) + (\mu_1 - d_+)\widehat{H}'(d_+). \tag{5.9}$$

In the following lemma, we describe the approximation above rigorously and also estimate $\sum_{i=1}^{n}(\gamma - \mu_i)^{-\ell}$ for $\ell = 2, 3, \dots$. Since the following lemma can be proved in a similar manner to the proof of Lemma 6.2 of [7], we omit the proof.

Lemma 5.4. *Recall the definition of $\widehat{H}(z)$ in Proposition 4.3. Then, for any $0 < \epsilon < 1$,*

$$\frac{1}{n}\sum_{i=1}^{n} \log(\gamma - \mu_i) = \widehat{H}(d_+) + (\mu_1 - d_+)\widehat{H}'(d_+) + O(n^{-1+\epsilon}). \tag{5.10}$$

Furthermore, for any $0 < \epsilon < 1$ there is a constant $C_0 > 0$ such that

$$n^{\ell(1-\epsilon)} \le \sum_{i=1}^{n} \frac{1}{(\gamma - \mu_i)^{\ell}} \le C_0^{\ell} n^{\ell+\epsilon} \tag{5.11}$$

for all $\ell = 2, 3, \dots$. Here, C_0 does not depend on ℓ.

Proof. See Lemma 6.2 of [7]. □

5.2 Truncation and deformation of the coutour

In Subsections 5.2–5.4, we fix $0 < \epsilon < \frac{1}{100}$ and prove Lemma 5.1.

Lemma 3.9 implies that the part of the double integral (5.1) with $|\operatorname{Im} z_1| \ge n^{-\frac{1}{2}+\epsilon}$ is $O(e^{-n^{\epsilon}})$.

For the part $|\operatorname{Im} z_1| < n^{-\frac{1}{2}+\epsilon}$, we deform the z_2-integral to a different vertical contour passing through a new point $\widetilde{\gamma}_2$ such that the difference $|G(\gamma_1, \gamma_2) - G(\gamma_1 + \mathrm{i}y_1, \widetilde{\gamma}_2)|$ is sufficiently small.

Intuitively, since the main contribution to the change of $G(z_1, z_2)$ near the critical point comes from the term $\frac{1}{4z_1z_2-\mu_1}$, it must be very sensitive to the change of the product z_1z_2 but not to the change of the individual variable z_1 or z_2 while z_1z_2 is fixed. Thus, for $y_1 \in \mathbb{R}$, we define

$$\widetilde{\gamma}_2 \equiv \widetilde{\gamma}_2(y_1) = \frac{\gamma_1\gamma_2}{\gamma_1 + \mathrm{i}y_1} \tag{5.12}$$

and analyze the double integral in (5.1) with the deformed contour that passes through $\widetilde{\gamma}_2$ for the z_2-integral.

Before we peform the analysis, we check that it is possible to deform the contour $\gamma_2 + \mathrm{i}\mathbb{R}$ to $\widetilde{\gamma}_2 + \mathrm{i}\mathbb{R}$ for given $z_1 \in \gamma_1 + \mathrm{i}\mathbb{R}$. For fixed $z_1 = \gamma_1 + \mathrm{i}y_1$, the branch cut Γ_c of the logarithmic function in $G(z_1, z_2)$ as a function of z_2 is

$$\Gamma_c = \{z_2 \in \mathbb{C} : 4z_1z_2 - \mu_1 \in \mathbb{R}^- \cup \{0\}\}.$$

If $z_2 \in \Gamma_c$, then there exists $r \geq 0$ such that

$$z_2 = \frac{\mu_1 - r}{4(\gamma_1 + \mathrm{i}y_1)} = \frac{\mu_1 - r}{4\gamma_1\gamma_2}\widetilde{\gamma}_2.$$

Since $4\gamma_1\gamma_2 > \mu_1$, this implies that $\operatorname{Re} z_2 < \operatorname{Re}\widetilde{\gamma}_2$, and hence Γ_c does not intersect the half plane $\{z \in \mathbb{C} : \operatorname{Re} z \geq \operatorname{Re}\widetilde{\gamma}_2\}$. Therefore, we can deform the z_2-contour, and hence

$$\begin{aligned} &-\int_{\gamma_1-\mathrm{i}n^{-\frac{1}{2}+\epsilon}}^{\gamma_1+\mathrm{i}n^{-\frac{1}{2}+\epsilon}} \int_{\gamma_2-\mathrm{i}\infty}^{\gamma_2+\mathrm{i}\infty} \exp\left[n\left(G(z_1,z_2) - G(\gamma_1,\gamma_2)\right)\right] \mathrm{d}z_2\,\mathrm{d}z_1 \\ &= \int_{-n^{-\frac{1}{2}+\epsilon}}^{n^{-\frac{1}{2}+\epsilon}} \int_{-\infty}^{\infty} \exp\left[n\left(G(\gamma_1+\mathrm{i}y_1, \widetilde{\gamma}_2+\mathrm{i}y_2) - G(\gamma_1,\gamma_2)\right)\right] \mathrm{d}y_2\,\mathrm{d}y_1\,. \end{aligned} \tag{5.13}$$

Recall that $\widetilde{\gamma}_2 \equiv \widetilde{\gamma}_2(y_1)$ depends on y_1.

We now truncate the y_2-integral. From the definition of $\widetilde{\gamma}_2$ and G, for all $y_1, y_2 \in \mathbb{R}$,

$$\begin{aligned} G(\gamma_1+\mathrm{i}y_1, \widetilde{\gamma}_2+\mathrm{i}y_2) - G(\gamma_1,\gamma_2) &= \mathrm{i}B_n(y_1+y_2) - \mathrm{i}B_n\frac{\gamma_2y_1}{\gamma_1+\mathrm{i}y_1} \\ &- \alpha_n \log\left(1 + \frac{\mathrm{i}y_1}{\gamma_1}\right) - \frac{1}{2n}\sum_{i=1}^n \log\left(1 - \frac{4y_1y_2}{4\gamma_1\gamma_2-\mu_i} + \frac{4\mathrm{i}\gamma_1y_2}{4\gamma_1\gamma_2-\mu_i}\right). \end{aligned} \tag{5.14}$$

Lemma 5.5. *Uniformly for $|y_1| \leq n^{-\frac{1}{2}+\epsilon}$,*

$$\left(\int_{-\infty}^{-n^{-\frac{1}{2}+2\epsilon}} + \int_{n^{-\frac{1}{2}+2\epsilon}}^{\infty}\right) \exp\left[n\left(G(\gamma_1+\mathrm{i}y_1, \widetilde{\gamma}_2+\mathrm{i}y_2) - G(\gamma_1,\gamma_2)\right)\right] \mathrm{d}y_2 = O(e^{-n^{\epsilon}}) \tag{5.15}$$

Proof. The proof is similar to Lemma 3.9, but easier. Taking the real part of (5.14),

$$\begin{aligned} &\operatorname{Re}\left(G(\gamma_1+\mathrm{i}y_1, \widetilde{\gamma}_2+\mathrm{i}y_2) - G(\gamma_1,\gamma_2)\right) \\ &= -\frac{B_n\gamma_2y_1^2}{\gamma_1^2+y_1^2} - \frac{\alpha_n}{2}\log\left(1+\frac{y_1^2}{\gamma_1^2}\right) - \frac{1}{4n}\sum_{i=1}^n \log\left[\left(1 - \frac{4y_1y_2}{4\gamma_1\gamma_2-\mu_1}\right)^2 + \left(\frac{4\gamma_1y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2\right]. \end{aligned} \tag{5.16}$$

If $y_1y_2 \le 0$, then

$$\left(1-\frac{4y_1y_2}{4\gamma_1\gamma_2-\mu_1}\right)^2+\left(\frac{4\gamma_1y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2 \ge 1+\left(\frac{4\gamma_1y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2 \ge 1+cy_2^2.$$

If $y_1y_2 \ge 0$, then

$$\left(1-\frac{4y_1y_2}{4\gamma_1\gamma_2-\mu_1}\right)^2+\left(\frac{4\gamma_1y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2 \ge 1-\frac{8y_1y_2}{4\gamma_1\gamma_2-\mu_1}+\frac{16\gamma_1^2y_2^2}{(4\gamma_1\gamma_2-\mu_i)^2} \ge 1+cy_2^2.$$

since $y_1y_2 \le y_2^2n^{-\epsilon}$ for $|y_1| \le n^{-\frac{1}{2}+\epsilon}$ and $|y_2| \ge n^{-1+2\epsilon}$. The above estimates imply that

$$\operatorname{Re}\left(G(\gamma_1+\mathrm{i}y_1,\widetilde{\gamma}_2+\mathrm{i}y_2)-G(\gamma_1,\gamma_2)\right) \ge -c'y_1^2-\frac{1}{4}\log(1+cy_2^2)$$

and hence, the left-hand side of (5.15) is bounded above by

$$2e^{-c'ny_1^2}\int_{n^{-1/2+2\epsilon}}^{\infty} e^{-\frac{n}{4}\log(1+cy_2^2)}\,\mathrm{d}y_2\,.$$

The integral is uniformly bounded and $e^{-c'ny_1^2} \le e^{-c'n^{2\epsilon}}$. Hence we obtain the lemma. □

The above truncation is not enough. The next lemma show that we can truncate further to the interval $|y_2| \le n^{-\frac{2}{3}+2\epsilon}$. Here we use the fact that (γ_1,γ_2) is the critical point. Note that this y_2-interval is smaller than the interval $|y_1| \le n^{-\frac{1}{2}+\epsilon}$.

Lemma 5.6. *Uniformly for* $|y_1| \le n^{-\frac{1}{2}+\epsilon}$,

$$\left(\int_{-n^{-\frac{1}{2}+\epsilon}}^{-n^{-\frac{2}{3}+2\epsilon}}+\int_{n^{-\frac{2}{3}+2\epsilon}}^{n^{-\frac{1}{2}+\epsilon}}\right)\exp\left[n\left(G(\gamma_1+\mathrm{i}y_1,\widetilde{\gamma}_2+\mathrm{i}y_2)-G(\gamma_1,\gamma_2)\right)\right]\mathrm{d}y_2 = O(e^{-n^{4\epsilon}}) \tag{5.17}$$

Proof. We start with (5.16). From the fact that (γ_1,γ_2) is a critical point, we showed in (3.12) that

$$\alpha_n = B_n(\gamma_1-\gamma_2). \tag{5.18}$$

Inserting this into (5.16) to remove α_n, and then expanding the terms involving B_n in terms of powers of y_1, we find that

$$\begin{aligned}&\operatorname{Re}\left(G(\gamma_1+\mathrm{i}y_1,\widetilde{\gamma}_2+\mathrm{i}y_2)-G(\gamma_1,\gamma_2)\right)\\&=-\frac{B_n(\gamma_1+\gamma_2)}{2\gamma_1^2}y_1^2-\frac{1}{4n}\sum_{i=1}^{n}\log\left[\left(1-\frac{4y_1y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2+\left(\frac{4\gamma_1y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2\right]+O(y_1^4).\end{aligned} \tag{5.19}$$

From the rigidity, Condition 3.4, it is easy to check that

$$n^{-\frac{2}{3}} \ll \mu_1-\mu_{n^{4\epsilon}} \ll n^{-\frac{2}{3}+2\epsilon}. \tag{5.20}$$

The upper bound implies that $4\gamma_1\gamma_2 - \mu_{n^{4\epsilon}} \ll n^{-\frac{2}{3}+2\epsilon}$. Hence, for $|y_2| \geq n^{-\frac{2}{3}+2\epsilon}$,

$$\begin{aligned}
&\frac{1}{4n}\sum_{i=1}^{n^{4\epsilon}} \log\left[\left(1-\frac{4y_1y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2+\left(\frac{4\gamma_1y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2\right] \\
&\geq \frac{1}{2n}\sum_{i=1}^{n^{4\epsilon}} \log\left(\frac{4\gamma_1|y_2|}{4\gamma_1\gamma_2-\mu_i}\right) \geq Cn^{-1+4\epsilon}.
\end{aligned} \tag{5.21}$$

The lower bound of (5.20) implies that $4\gamma_1\gamma_2 - \mu_{n^{\epsilon}} \gg |y_1y_2|$ for $|y_2| \leq n^{-\frac{1}{2}+\epsilon}$ and $|y_1| \leq n^{-\frac{1}{2}+\epsilon}$. Hence,

$$\begin{aligned}
&\frac{1}{4n}\sum_{i=n^{4\epsilon}+1}^{n} \log\left[\left(1-\frac{4y_1y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2+\left(\frac{4\gamma_1y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2\right] \geq \frac{1}{2n}\sum_{i=n^{4\epsilon}+1}^{n} \log\left(1-\frac{4y_1y_2}{4\gamma_1\gamma_2-\mu_i}\right) \\
&\geq -\frac{C}{n}\sum_{i=n^{4\epsilon}+1}^{n} \frac{|y_1y_2|}{4\gamma_1\gamma_2-\mu_i} \geq -C'|y_1y_2| \geq -C'n^{-1+3\epsilon}.
\end{aligned}$$

Note that the exponent $(-1+3\epsilon)$ is smaller than $(-1+4\epsilon)$ in (5.21). Therefore, we obtain for $n^{-\frac{2}{3}+2\epsilon} \leq |y_2| \leq n^{-\frac{1}{2}+\epsilon}$ that

$$\mathrm{Re}\left(G(\gamma_1+\mathrm{i}y_1, \widetilde{\gamma}_2+\mathrm{i}y_2) - G(\gamma_1,\gamma_2)\right) \leq -Cn^{-1+4\epsilon}.$$

This implies the lemma. □

5.3 Decomposition of the double integral

We consider the part of the double integral (5.13) with $|y_1| \leq n^{-\frac{1}{2}+\epsilon}$ and $|y_2| \leq n^{-\frac{2}{3}+2\epsilon}$. From (5.14) and (5.18), using the Taylor series,

$$\begin{aligned}
&G(\gamma_1+\mathrm{i}y_1, \widetilde{\gamma}_2+\mathrm{i}y_2) - G(\gamma_1,\gamma_2) \\
&= \mathrm{i}B_ny_2 - \frac{B_n(\gamma_1+\gamma_2)}{2\gamma_1^2}y_1^2 - \frac{1}{2n}\sum_{i=1}^{n}\log\left(1-\frac{4y_1y_2}{4\gamma_1\gamma_2-\mu_i}+\frac{4\mathrm{i}\gamma_1y_2}{4\gamma_1\gamma_2-\mu_i}\right) + O(y_1^3).
\end{aligned} \tag{5.22}$$

Hence,

$$\begin{aligned}
&\exp\left[n\left(G(\gamma_1+\mathrm{i}y_1, \widetilde{\gamma}_2+\mathrm{i}y_2) - G(\gamma_1,\gamma_2)\right)\right] \\
&= \exp\left[\mathrm{i}B_nny_2 - \frac{B_n(\gamma_1+\gamma_2)}{2\gamma_1^2}ny_1^2 - \frac{1}{2}\sum_{i=1}^{n}\log\left(1+\frac{4\mathrm{i}\gamma_1y_2}{4\gamma_1\gamma_2-\mu_i}\right)\right] \\
&\times \exp\left[-\frac{1}{2}\sum_{i=1}^{n}\log\left(1-\frac{4y_1y_2}{4\gamma_1\gamma_2-\mu_i+4\mathrm{i}\gamma_1y_2}\right)+O(y_1^3)\right].
\end{aligned} \tag{5.23}$$

Applying Lemma 5.4,

$$\sum_{i=1}^{n}\frac{1}{|4\gamma_1\gamma_2-\mu_i+4\mathrm{i}\gamma_1y_2|^{\ell}} \leq \sum_{i=1}^{n}\frac{1}{|4\gamma_1\gamma_2-\mu_i|^{\ell}} \leq C_0^{\ell}n^{\ell+\epsilon} \tag{5.24}$$

for $\ell = 2, 3, \dots$, where C_0 is the constant in Lemma 5.4. For $\ell = 1$, we use the bound

$$\sum_{i=1}^{n} \frac{1}{|4\gamma_1\gamma_2 - \mu_i + 4\mathrm{i}\gamma_1 y_2|} \le \sum_{i=1}^{n} \frac{1}{4\gamma_1\gamma_2 - \mu_i} = \frac{n}{2\gamma_1}\left(B_n - \partial_2 G(\gamma_1, \gamma_2)\right) = \frac{nB_n}{2\gamma_1} = O(n). \tag{5.25}$$

This implies that, from the conditions on y_1, y_2,

$$\sum_{i=1}^{n} \left(\frac{y_1 y_2}{4\gamma_1\gamma_2 - \mu_i + 4\mathrm{i}\gamma_1 y_2}\right)^{\ell} = O(n^{-\frac{1}{6}\ell + (3\ell+1)\epsilon}) \tag{5.26}$$

for $\ell = 2, 3, \dots$ and

$$\sum_{i=1}^{n} \frac{y_1 y_2}{4\gamma_1\gamma_2 - \mu_i + 4\mathrm{i}\gamma_1 y_2} = O(n^{-\frac{1}{6} + 3\epsilon}). \tag{5.27}$$

Thus, expanding the last exponential function in (5.23), we obtain

$$\begin{aligned}
&\exp\left[n\left(G(\gamma_1 + \mathrm{i}y_1, \widetilde{\gamma}_2 + \mathrm{i}y_2) - G(\gamma_1, \gamma_2)\right)\right] \\
&= \exp\left[\mathrm{i}B_n n y_2 - \frac{B_n(\gamma_1 + \gamma_2)}{2\gamma_1^2} n y_1^2 - \frac{1}{2}\sum_{i=1}^{n} \log\left(1 + \frac{4\mathrm{i}\gamma_1 y_2}{4\gamma_1\gamma_2 - \mu_i}\right)\right] \\
&\quad \times \left(1 + 2y_1 y_2 \sum_{i=1}^{n} \frac{1}{4\gamma_1\gamma_2 - \mu_i + 4\mathrm{i}\gamma_1 y_2} + O(n^{-\frac{1}{3} + 6\epsilon})\right).
\end{aligned} \tag{5.28}$$

We thus have

$$\int_{-n^{-\frac{1}{2}+\epsilon}}^{n^{-\frac{1}{2}+\epsilon}} \int_{-n^{-\frac{2}{3}+2\epsilon}}^{n^{-\frac{2}{3}+2\epsilon}} \exp\left[n\left(G(\gamma_1 + \mathrm{i}y_1, \widetilde{\gamma}_2 + \mathrm{i}y_2) - G(\gamma_1, \gamma_2)\right)\right] \mathrm{d}y_2\, \mathrm{d}y_1 =: I_1 + I_2 + I_3, \tag{5.29}$$

where I_1, I_2, and I_3 are given as follows: First,

$$I_1 = \int_{-n^{-\frac{1}{2}+\epsilon}}^{n^{-\frac{1}{2}+\epsilon}} \int_{-n^{-\frac{2}{3}+2\epsilon}}^{n^{-\frac{2}{3}+2\epsilon}} \exp\left[\mathrm{i}B_n n y_2 - \frac{B_n(\gamma_1 + \gamma_2)}{2\gamma_1^2} n y_1^2 - \frac{1}{2}\sum_{i=1}^{n} \log\left(1 + \frac{4\mathrm{i}\gamma_1 y_2}{4\gamma_1\gamma_2 - \mu_i}\right)\right] \mathrm{d}y_2 \mathrm{d}y_1. \tag{5.30}$$

This is equal to the product of two single integrals I_{11} and I_{12}. The y_1-integral is

$$I_{11} := \int_{-n^{-\frac{1}{2}+\epsilon}}^{n^{-\frac{1}{2}+\epsilon}} \exp\left[-\frac{B_n(\gamma_1 + \gamma_2)}{2\gamma_1^2} n y_1^2\right] \mathrm{d}y_1. \tag{5.31}$$

This is real-valued and we have

$$\frac{C}{\sqrt{n}} \le I_{11} \le \frac{C'}{\sqrt{n}}. \tag{5.32}$$

The y_2-integral is

$$I_{12} := \int_{-n^{-\frac{2}{3}+2\epsilon}}^{n^{-\frac{2}{3}+2\epsilon}} \exp\left[\mathrm{i}B_n n y_2 - \frac{1}{2}\sum_{i=1}^{n} \log\left(1 + \frac{4\mathrm{i}\gamma_1 y_2}{4\gamma_1\gamma_2 - \mu_i}\right)\right] \mathrm{d}y_2. \tag{5.33}$$

This is also real-valued since the imaginary part of the integrand is an odd function of y_2. We have

$$\begin{aligned} I_{12} \le |I_{12}| &\le \int_{-\infty}^{\infty} \exp\left[-\frac{1}{4}\sum_{i=1}^{n} \log\left(1+\left(\frac{4\gamma_1 y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2\right)\right] \mathrm{d}y_2 \\ &\le \int_{-\infty}^{\infty} \exp\left[-\frac{n}{4}\log\left(1+Cy_2^2\right)\right] \mathrm{d}y_2 \le C. \end{aligned} \tag{5.34}$$

On the other hand, we will show the following lower bound in Subsection 5.4:

$$I_{12} \ge Cn^{-1-5\epsilon}. \tag{5.35}$$

Assuming this is true and using (5.34), we find that

$$Cn^{-\frac{3}{2}-5\epsilon} \le I_1 \le Cn^{-\frac{1}{2}}. \tag{5.36}$$

Second,

$$\begin{aligned} I_2 = \int_{-n^{-\frac{1}{2}+\epsilon}}^{n^{-\frac{1}{2}+\epsilon}} \int_{-n^{-\frac{2}{3}+2\epsilon}}^{n^{-\frac{2}{3}+2\epsilon}} \exp\left[\mathrm{i}B_n n y_2 - \frac{B_n(\gamma_1+\gamma_2)}{2\gamma_1^2} n y_1^2 - \frac{1}{2}\sum_{i=1}^{n}\log\left(1+\frac{4\mathrm{i}\gamma_1 y_2}{4\gamma_1\gamma_2-\mu_i}\right)\right] \\ \times\left(2y_1y_2\sum_{i=1}^{n}\frac{1}{4\gamma_1\gamma_2-\mu_i+4\mathrm{i}\gamma_1 y_2}\right)\mathrm{d}y_2\mathrm{d}y_1\,. \end{aligned}$$

Since the integrand is an odd function of y_1, we find that $I_2 = 0$.

Finally, I_3 satisfies

$$\begin{aligned} |I_3| &\le Cn^{-\frac{1}{3}+6\epsilon} \int_{-n^{-\frac{1}{2}+\epsilon}}^{n^{-\frac{1}{2}+\epsilon}} \int_{-n^{-\frac{2}{3}+2\epsilon}}^{n^{-\frac{2}{3}+2\epsilon}} \exp\left[-\frac{B_n(\gamma_1+\gamma_2)}{2\gamma_1^2} n y_1^2 - \frac{1}{2}\operatorname{Re}\sum_{i=1}^{n}\log\left(1+\frac{4\mathrm{i}\gamma_1 y_2}{4\gamma_1\gamma_2-\mu_i}\right)\right]\mathrm{d}y_2\mathrm{d}y_1 \\ &\le C' n^{-\frac{5}{6}+6\epsilon}\int_{-n^{-\frac{2}{3}+2\epsilon}}^{n^{-\frac{2}{3}+2\epsilon}} \exp\left[-\frac{1}{2}\log\left(\frac{4\gamma_1 y_2}{4\gamma_1\gamma_2-\mu_1}\right)\right]\mathrm{d}y_2 \\ &\le C'' n^{-\frac{5}{6}+6\epsilon}\int_{-n^{-\frac{2}{3}+2\epsilon}}^{n^{-\frac{2}{3}+2\epsilon}} \frac{1}{\sqrt{n^{1-\epsilon}y_2}}\mathrm{d}y_2 \le C'' n^{-\frac{5}{3}+8\epsilon}. \end{aligned}$$

Note that this upper bound is smaller than the lower bound of (5.36) if $\epsilon < \frac{1}{78}$.

Combining all estimates of Subsections 5.2 and 5.3, we obtain

$$Cn^{-\frac{3}{2}-5\epsilon} \le -\int_{\gamma_1-\mathrm{i}\infty}^{\gamma_1+\mathrm{i}\infty}\int_{\gamma_2-\mathrm{i}\infty}^{\gamma_2+\mathrm{i}\infty} \exp\left[n\left(G(z_1,z_2)-G(\gamma_1,\gamma_2)\right)\right]\,\mathrm{d}z_2\,\mathrm{d}z_1 \le Cn^{-\frac{1}{2}}\,, \tag{5.37}$$

thus prove Lemma 5.1, assuming that (5.35) is true.

5.4 Analysis of I_{12}

To complete the proof of Lemma 5.1, it remains to show the lower bound $I_{12} \ge Cn^{-1-5\epsilon}$ in (5.35). We note by checking directly from the definition of G that,

$$I_{12} = \int_{-n^{-\frac{2}{3}+2\epsilon}}^{n^{-\frac{2}{3}+2\epsilon}} \exp\left[n\left(G(\gamma_1,\gamma_2+\mathrm{i}y_2)-G(\gamma_1,\gamma_2)\right)\right]\mathrm{d}y_2. \tag{5.38}$$

Define

$$K := -\mathrm{i}\int_{\gamma_2-\mathrm{i}\infty}^{\gamma_2+\mathrm{i}\infty} \exp\left[n\left(G(\gamma_1,z)-G(\gamma_1,\gamma_2)\right)\right]\mathrm{d}z. \tag{5.39}$$

Then I_{12} is the same integral as K where the contour is restricted to the part $|z-\gamma_2| \le n^{-\frac{2}{3}+2\epsilon}$. Note that K is real-valued since $G(\gamma_1,\overline{z_2}) = \overline{G(\gamma_1,z_2)}$. The lower bound (5.35) follows if we show that

(a) $|K - I_{12}| \le e^{-Cn^{4\epsilon}}$, and

(b) $K \ge Cn^{-1-5\epsilon}$.

Since K is real-valued,

$$|K - I_{12}| \le \left(\int_{-\infty}^{\infty} - \int_{-n^{-\frac{2}{3}+2\epsilon}}^{n^{-\frac{2}{3}+2\epsilon}}\right) \exp\left[n \operatorname{Re}\left(G(\gamma_1,\gamma_2+\mathrm{i}y_2)-G(\gamma_1,\gamma_2)\right)\right]\mathrm{d}y_2.$$

We have

$$n\operatorname{Re}\left(G(\gamma_1,\gamma_2+\mathrm{i}y_2)-G(\gamma_1,\gamma_2)\right) = -\frac{1}{4}\sum_{i=1}^{n}\log\left[1+\left(\frac{4\gamma_1 y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2\right].$$

Using (5.21), for $n^{-\frac{2}{3}+2\epsilon} \le |y_2| \le n$, we have the estimate

$$\frac{1}{4}\sum_{i=1}^{n}\log\left[1+\left(\frac{4\gamma_1 y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2\right] \ge \frac{1}{4}\sum_{i=1}^{n^{4\epsilon}}\log\left[1+\left(\frac{4\gamma_1 y_2}{4\gamma_1\gamma_2-\mu_i}\right]^2\right) \ge Cn^{4\epsilon}.$$

For $|y_2| \ge n$,

$$\frac{1}{4}\sum_{i=1}^{n}\log\left[1+\left(\frac{4\gamma_1 y_2}{4\gamma_1\gamma_2-\mu_i}\right)^2\right] \ge \frac{n}{4}\log(1+cy_2^2) \ge \frac{n}{4}\log(c|y_2|).$$

Hence

$$|K - I_{12}| \le 2ne^{-Cn^{4\epsilon}} + 2\int_n^{\infty}(cy_2)^{-\frac{n}{4}}\mathrm{d}y_2 \le e^{-C'n^{4\epsilon}}. \tag{5.40}$$

We thus obtained property (a).

We now prove the property (b), $K \ge Cn^{-1-5\epsilon}$. We follow the proof of Lemma 6.3 in [7] closely. Observe that γ_2 is a critical point of the function $G(\gamma_1,z)$. Let Γ be the curve of steepest-descent that passes through the point γ_2. It satisfies $\operatorname{Im} G(\gamma_1,z)=0$. It is straightforward to check from the formula of G that

(i) Γ is symmetric about the real axis,

(ii) $\Gamma \cap \mathbb{C}^+$ is a C^1 curve,

(iii) Γ lies in the half plane $\operatorname{Re} z \le \gamma_2$,

(iv) Γ intersects with $\mathbb{R}$ only at γ_2,

(v) the tangent line of Γ at γ_2 is parallel to the imaginary axis,

(vi) the asymptote of Γ is the negative real axis.

For example, the property (iii) can be checked by noting that for $z = x + \mathrm{i}y$ with $x > \gamma_2$,

$$F(y) := \operatorname{Im} G(\gamma_1, x + \mathrm{i}y) = B_n y - \frac{1}{2n} \sum_{i=1}^{n} \arctan \left(\frac{4\gamma_1 y}{(4\gamma_1 x - \mu_i)^2 + (4\gamma_1 y)^2} \right)$$

has the global minimum at $y = 0$ by computing its derivative.

Since $\operatorname{Re}(G(\gamma_1, z) - G(\gamma_1, \gamma_2)) \le B_n \gamma_2 - \frac{1}{2} \log(R/2)$ for $|z| = R$ with $\operatorname{Re}(z) \le \gamma_2$, we can deform the contour so that

$$K = -\mathrm{i} \int_\Gamma \exp \left[n \left(G(\gamma_1, z) - G(\gamma_1, \gamma_2) \right) \right] \mathrm{d}z.$$

For $z \in \Gamma$, we let $x = \operatorname{Re} z$ and $y = \operatorname{Im} z$. Then, $\mathrm{d}z = \mathrm{d}x + \mathrm{i}\mathrm{d}y$ and

$$K = -\mathrm{i} \int_\Gamma \exp \left[n \left(G(\gamma_1, z) - G(\gamma_1, \gamma_2) \right) \right] \mathrm{d}x + \int_\Gamma \exp \left[n \left(G(\gamma_1, z) - G(\gamma_1, \gamma_2) \right) \right] \mathrm{d}y.$$

Let $\Gamma^+ = \Gamma \cap \mathbb{C}^+$. By symmetry,

$$K = 2 \int_{\Gamma^+} \exp \left[n \left(G(\gamma_1, z) - G(\gamma_1, \gamma_2) \right) \right] \mathrm{d}y. \tag{5.41}$$

In [7], the lower bound of K was obtained by restricting the integral to a small ball of radius n^{-2}. In the current work, however, we need to refine the argument further to prove (5.39). We let D_1 be the disk of radius $n^{-1-\epsilon}$ centered at γ_2, and similarly, D_2 be the disk of radius $n^{-1-2\epsilon}$ centered at γ_2. The rest of the contour is controlled by the following lemma.

Lemma 5.7 (Lemma 6.4 of [7]). *Suppose that f is a real-valued function defined on Γ^+ and $f(z)$ is decreasing along the curve Γ^+ as z moves from the point γ_2 to the point $-\infty$. Then,*

$$\int_{\Gamma^+} e^{f(z)} \mathrm{d}y \ge 0. \tag{5.42}$$

Since $G(\gamma_1, z)$ is analytic for z_2 in D_1, the series expansion

$$G(\gamma_1, z) - G(\gamma_1, \gamma_2) = \sum_{j=2}^{\infty} \frac{1}{j!} \partial_2^j G(\gamma_1, \gamma_2)(z - \gamma_2)^j \tag{5.43}$$

converges for $z \in \Gamma^+ \cap D_1$. Set $X = \operatorname{Re}(z - \gamma_2)$ and $Y = \operatorname{Im} z = \operatorname{Im}(z - \gamma_2)$. Comparing the imaginary parts of the both sides of (5.43) by using Lemma 5.4, we find that

$$0 = \partial_2^2 G(\gamma_1, \gamma_2) XY + \frac{1}{2} \partial_2^3 G(\gamma_1, \gamma_2 X^2 Y - \frac{1}{6} \partial_2^3 G(\gamma_1, \gamma_2) Y^3 + \widetilde{\Omega} \tag{5.44}$$

with

$$\widetilde{\Omega} = \sum_{j=4}^{\infty} \frac{1}{j!} \partial_2^j G(\gamma_1, \gamma_2) \operatorname{Im} \left((X + \mathrm{i}Y)^j \right).$$

Note that $\operatorname{Im}\left((X+\mathrm{i}Y)^j\right)$ is a homogeneous polynomial of X and Y with degree j. In the polynomial, every term contains both X and Y, possibly except the term Y^j when j is odd. In any case,

$$\left|\frac{1}{j!}\partial_2^j G(\gamma_1,\gamma_2)\operatorname{Im}\left((X+\mathrm{i}Y)^j\right)\right| \le \frac{(8\gamma_1 C_0)^j}{j!} n^{j-1+\epsilon}|z-\gamma_2|^{j-2}|XY| + \frac{(4\gamma_1 C_0)^j}{j!} n^{j-1+\epsilon} Y^j,$$

hence

$$|\widetilde{\Omega}| \le C n^{1-\epsilon}(|XY|+Y^2). \tag{5.45}$$

Define

$$\tau = -\frac{\partial_2^3 G(\gamma_1,\gamma_2)}{\partial_2^2 G(\gamma_1,\gamma_2)} > 0.$$

From Lemma 5.4 (by putting $\epsilon/4$ instead of ϵ), we find that $\partial_2^2 G(\gamma_1,\gamma_2) \ge C n^{1-\frac{\epsilon}{2}}$ and

$$\tau \ll n^{1-\epsilon}. \tag{5.46}$$

Thus, dividing both sides of (5.44) by $\partial_2^2 G(\gamma_1,\gamma_2)Y$, we obtain that

$$X(1+o(1)) + \frac{\tau}{6}Y^2(1+o(1)) = 0,$$

and thus,

$$X = -\frac{\tau}{6}Y^2(1+o(1)). \tag{5.47}$$

We also see that $\Gamma^+ \cap D_1$ is a graph, $\mathrm{d}y = \mathrm{d}Y$ is positive on $\Gamma^+ \cap D_1$, and Γ^+ intersects ∂D_1 at exactly one point.

Let ζ_1 (resp. ζ_2) be the point where Γ^+ and ∂D_1 (resp. ∂D_2) intersect. Then,

$$\begin{aligned} G(\gamma_1,\gamma_2) - G(\gamma_1,\zeta_1) &\ge \frac{1}{4}\partial_2^2 G(\gamma_1,\gamma_2)|\zeta_1-\gamma_2|^2 - \sum_{j=3}^{\infty}\frac{1}{j!}C_0^j n^{j-1+\frac{\epsilon}{4}}|\zeta_1-\gamma_2|^j \\ &\ge C n^{-1-\frac{5\epsilon}{2}} \end{aligned} \tag{5.48}$$

and

$$\begin{aligned} G(\gamma_1,\gamma_2) - G(\gamma_1,\zeta_2) &\le \frac{1}{2}\partial_2^2 G(\gamma_1,\gamma_2)|\zeta_2-\gamma_2|^2 + \sum_{j=3}^{\infty}\frac{1}{j!}C_0^j n^{j-1+\frac{\epsilon}{4}}|\zeta_2-\gamma_2|^j \\ &\le C n^{-1-\frac{7\epsilon}{2}}. \end{aligned} \tag{5.49}$$

We introduce the function

$$f(z) = \begin{cases} nG(\gamma_1,\zeta_1) - G(\gamma_1,\gamma_2) & \text{if } z \in \Gamma^+ \cap D_1, \\ nG(\gamma_1,z) - G(\gamma_1,\gamma_2) & \text{if } z \in \Gamma^+ \cap D_1^c. \end{cases}$$

It is obvious that $f(z)$ is a decreasing function of z along the curve Γ^+ as z moves from γ_2 to $-\infty$. Thus, applying Lemma 5.7 to the function f,

$$\int_{\Gamma^+\cap D_1} \exp\left[n\left(G(\gamma_1,\zeta_1) - G(\gamma_1,\gamma_2)\right)\right]\mathrm{d}y + \int_{\Gamma^+\cap D_1^c} \exp\left[n\left(G(\gamma_1,z) - G(\gamma_1,\gamma_2)\right)\right]\mathrm{d}y \ge 0. \tag{5.50}$$

Since $\mathrm{d}y$ is positive on $\Gamma^+ \cap D_1$,

$$\begin{aligned}\frac{K}{2} &= \int_{\Gamma^+} \exp\left[n\left(G(\gamma_1, z) - G(\gamma_1, \gamma_2)\right)\right] \mathrm{d}y \\ &= \int_{\Gamma^+ \cap D_2} \exp\left[n\left(G(\gamma_1, z) - G(\gamma_1, \gamma_2)\right)\right] \mathrm{d}y + \int_{\Gamma^+ \cap D_2^c} \exp\left[n\left(G(\gamma_1, z) - G(\gamma_1, \gamma_2)\right)\right] \mathrm{d}y \\ &\geq \int_{\Gamma^+ \cap D_2} \exp\left[n\left(G(\gamma_1, \zeta_2) - G(\gamma_1, \gamma_2)\right)\right] \mathrm{d}y + \int_{\Gamma^+ \cap D_2^c} \exp\left[n\left(G(\gamma_1, z) - G(\gamma_1, \gamma_2)\right)\right] \mathrm{d}y.\end{aligned} \tag{5.51}$$

Subtracting (5.50) from (5.51), we find that

$$\begin{aligned}\frac{K}{2} &\geq \int_{\Gamma^+ \cap D_2} \Big(\exp\left[n\left(G(\gamma_1, \zeta_2) - G(\gamma_1, \gamma_2)\right)\right] - \exp\left[n\left(G(\gamma_1, \zeta_1) - G(\gamma_1, \gamma_2)\right)\right] \Big) \mathrm{d}y \\ &\quad + \int_{\Gamma^+ \cap D_1 \cap D_2^c} \Big(\exp\left[n\left(G(\gamma_1, z) - G(\gamma_1, \gamma_2)\right)\right] - \exp\left[n\left(G(\gamma_1, \zeta_1) - G(\gamma_1, \gamma_2)\right)\right] \Big) \mathrm{d}y \\ &\geq \int_{\Gamma^+ \cap D_2} \Big(\exp\left[n\left(G(\gamma_1, \zeta_2) - G(\gamma_1, \gamma_2)\right)\right] - \exp\left[n\left(G(\gamma_1, \zeta_1) - G(\gamma_1, \gamma_2)\right)\right] \Big) \mathrm{d}y.\end{aligned} \tag{5.52}$$

From (5.47), we find that $\operatorname{Im} \zeta_2 \gg \operatorname{Re} \zeta_2$, hence $\operatorname{Im} \zeta_2 \geq C n^{-1-2\epsilon}$. Since $\int_{\Gamma^+ \cap D_2} \mathrm{d}y = \operatorname{Im} \zeta_2 \geq C n^{-1-2\epsilon}$, we find from the estimates (5.48) and (5.49) that

$$\frac{K}{2} \geq C n^{-1-2\epsilon} \left[\exp(-C n^{-\frac{7\epsilon}{2}}) - \exp(-C n^{-\frac{5\epsilon}{2}}) \right] \geq C n^{-1-5\epsilon}. \tag{5.53}$$

This proves the desired lower bound of K. Thus, Lemma 5.1 is proved.

5.5 Double integral in the low temperature regime

The following is the main result for the random double integral in the low temperature regime.

Proposition 5.8 (Random double integral for low temperature)**.** *Assume Conditions 3.2 3.3 and 3.4. Suppose that B in Condition 3.2 satisfies $B > B_c$ where B_c is defined in* (3.22)*. Then, for every $\epsilon > 0$,*

$$\frac{1}{n} \log \mathbf{Q}_n = \widehat{E} + (\mu_1 - d_+) \widehat{L} + O(n^{-1+\epsilon}) \tag{5.54}$$

where

$$\begin{aligned}\widehat{E} &= \sqrt{\alpha^2 + d_+ B^2} - \alpha \log\left(\frac{\alpha + \sqrt{\alpha^2 + d_+ B^2}}{2B} \right) - \frac{1}{2} \widehat{H}(d_+), \\ \widehat{L} &= \frac{B^2}{2(\alpha + \sqrt{\alpha^2 + d_+ B^2})} - \frac{1}{2} \widehat{H}'(d_+).\end{aligned} \tag{5.55}$$

Proof. From Lemma 5.1,

$$\frac{1}{n} \log \mathbf{Q}_n = G(\gamma_1, \gamma_2) + O\left(\frac{\log n}{n} \right).$$

We have

$$G(\gamma_1,\gamma_2) = \sqrt{\alpha_n^2+\gamma B_n^2} - \frac{1}{2n}\sum_{i=1}^n \log(\gamma-\mu_i) - \alpha_n \log\left(\frac{\alpha_n+\sqrt{\alpha_n^2+\gamma B_n^2}}{2B_n}\right). \tag{5.56}$$

From Lemma 5.4,

$$\frac{1}{n}\sum_{i=1}^n \log(\gamma-\mu_i) = \widehat{H}(d_+) + (\mu_1-d_+)\widehat{H}'(d_+) + O(n^{-1+\epsilon}).$$

On the other hand, for the other two terms in (5.56), we first replace α_n and B_n by α and B and introduce an error term $O(n^{-1-\delta})$. We then replace γ by μ_1 and introduce an error term $O(n^{-1+\epsilon})$ due to Lemma 5.3. Writing $\mu_1 = d_+ + (\mu_1 - d_+)$ and using the Taylor expansion up to the first order using $(\mu_1-d_+)^2 = O(n^{-4/3+2\epsilon})$, we find that

$$\begin{aligned}&\sqrt{\alpha_n^2+\gamma B_n^2} - \alpha_n\log\left(\frac{\alpha_n+\sqrt{\alpha_n^2+\gamma B_n^2}}{2B_n}\right)\\ &= \sqrt{\alpha^2+d_+B^2} - \alpha\log\left(\frac{\alpha+\sqrt{\alpha^2+d_+B^2}}{2B}\right) + \frac{B^2}{2(\alpha+\sqrt{\alpha^2+d_+B^2})}(\mu_1-d_+) + O(n^{-4/3+2\epsilon}).\end{aligned}$$

This completes the proof. □

6 Proof of Theorem 2.4

Recall that we assume $N_1 \geq N_2$ and $N = N_1 + N_2$. We take the limit as $N, N_1, N_2 \to \infty$ satisfying

$$\frac{N_1}{N} = r_1 + O(N^{-1-\delta}), \qquad \frac{N_2}{N} = r_2 + O(N^{-1-\delta}) \tag{6.1}$$

for $r_1 \geq r_2 > 0$ and $r_1 + r_2 = 1$. From (3.8), we have

$$Z_{N_1,N_2}(\beta) = \mathbf{Q}\left(N_2, \frac{N_1-N_2}{2N_2}, \frac{N_1}{\sqrt{N_2N}}\beta\right)R(N_1,N_2,\beta) \tag{6.2}$$

where $\mathbf{Q}$ is defined with the eigenvalues μ_i, $1 \leq i \leq N_2$, of the random matrix $\frac{1}{N_1}J^TJ$. Here J is an $N_1 \times N_2$ matrix with independent and identically distributed entries of mean 0 and variance 1 satisfying the assumptions in Subsection 2.1. The constant

$$R(N_1,N_2,\beta) = \frac{1}{|S^{N_1-1}||S^{N_2-1}|}2^{N_2}\left(\frac{\pi^2 N}{N_1^2 N_2 \beta^2}\right)^{(N-4)/4}. \tag{6.3}$$

We use the results of the previous two sections on $\mathbf{Q}_n$. To do that, we need to check Conditions 3.2, 3.3 and 3.4:

- Conditions 3.2 holds with

$$n = N_2, \qquad \alpha = \frac{r_1-r_2}{2r_2}, \qquad B = \frac{r_1}{\sqrt{r_2}}\beta. \tag{6.4}$$

  where $\alpha_n = \frac{N_1-N_2}{2N_2}$ and $B_n = \frac{N_1}{\sqrt{N_2N}}\beta$.

- Condition 3.3 follows from the well-known Marchenko-Pastur law [30] in random matrix theory. The limiting empirical measure is given by (2.25),

$$\mathrm{d}\widehat{\mu}(x) = \mathrm{d}\mu_{\mathrm{MP}}(x) := \frac{2\sqrt{(d_+ - x)(x - d_-)}}{\pi(\sqrt{d_+} - \sqrt{d_-})^2 x} \mathbb{1}_{(d_-,d_+)}(x)\mathrm{d}x \tag{6.5}$$

  where

$$d_- = \frac{(\sqrt{r_1} - \sqrt{r_2})^2}{r_1}, \qquad d_+ = \frac{(\sqrt{r_1} + \sqrt{r_2})^2}{r_1}. \tag{6.6}$$

- Condition 3.4 holds with high probability for the eigenvalues. This is proved recently by Pillai and Yin in [35] for $r_1 > r_2$. For the case $r_1 = r_2$, Condition 3.4 with high probability follows from Corollary 1.3 of [2] and the fact that $\mathrm{d}\mu_{\mathrm{MP}}(x) = \mathrm{d}\mu_{\mathrm{SC}}(\sqrt{x})$ where μ_{SC} is the Wigner semicircle distribution. (See also Equation (1.12) of [2].) Similar rigidity results were proved for other classes of random matrices starting with the Wigner matrices [21] and also various random matrix models including invariant ensembles [13] and sparse random matrices [20]. Rigidity estimates are obtained from the local laws such as local semicircle law or local Marchenko-Pastur law, and they are also crucial a priori estimates for the proof of bulk and edge universality of random matrices.

Before we deduce the limit of $\mathbf{Q}$, we first state the asymptotics of $R(N_1, N_2, \beta)$.

Lemma 6.1. *We have*

$$\begin{aligned}\frac{1}{N}\log\left(R(N_1, N_2, \beta)\right) = &-\frac{1}{2} + r_2\log 2 - \frac{1}{2}\log(2r_1\sqrt{r_2}\beta) + \frac{r_1}{2}\log r_1 + \frac{r_2}{2}\log r_2 \\ &+ \frac{\log N}{N} - \frac{1}{N}\log\left(\frac{\pi}{r_1\sqrt{r_1 r_2}\beta^2}\right) + O(N^{-2}).\end{aligned} \tag{6.7}$$

Proof. The area of unit sphere satisfies

$$\log(|S^{n-1}|) = \log\left(\frac{2\pi^{n/2}}{\Gamma(n/2)}\right) = \frac{n}{2}\log\left(\frac{2\pi e}{n}\right) + \frac{1}{2}\log\left(\frac{n}{\pi}\right) + O(n^{-1}). \tag{6.8}$$

Hence,

$$\frac{1}{N}\log(|S^{N_1-1}||S^{N_2-1}|) = \frac{1}{2}\log\left(\frac{2\pi e}{N}\right) - \frac{r_1}{2}\log r_1 - \frac{r_2}{2}\log r_2 + \frac{1}{N}\log\left(\frac{\sqrt{r_1 r_2}N}{\pi}\right) + O(N^{-2}).$$

On the other hand,

$$\frac{1}{N}\log\left(2^{N_2}\left(\frac{\pi^2 N}{N_1 N_2^2\beta^2}\right)^{(N-4)/4}\right) = r_2\log 2 + \left(\frac{1}{2} - \frac{2}{N}\right)\log\left(\frac{\pi}{r_1\sqrt{r_2}\beta N}\right).$$

We thus obtain the lemma. □

6.1 Transforms of the Marchenko-Pastur distribution and critical temperature

We need the following formulas of the log transform and the Stieltjes transform of the Marchenko-Pastur distribution.

Lemma 6.2. *Set*

$$R(z) := \sqrt{(z-d_-)(z-d_+)}. \tag{6.9}$$

We have

$$\begin{aligned} H_{\rm MP}(z) &:= \int_{\mathbb{R}} \log(z-x) \mathrm{d}\mu_{\rm MP}(x) \\ &= \frac{2}{(\sqrt{d_+}-\sqrt{d_-})^2}\Big[z - \sqrt{d_+d_+}\log z - R(z) + (d_+ + d_-)\log(\sqrt{z-d_-}+\sqrt{z-d_+}) \\ &\qquad + \sqrt{d_+d_-}\log\left(\frac{\sqrt{d_+(z-d_-)}-\sqrt{d_-(z-d_+)}}{\sqrt{d_+(z-d_-)}+\sqrt{d_-(z-d_+)}}\right)\Big] \end{aligned}$$

for $z \notin (-\infty, d_+)$ *and*

$$s_{\rm MP}(z) := H'_{\rm MP}(z) = \int_{\mathbb{R}} \frac{1}{z-x}\mathrm{d}\mu_{\rm MP}(x) = \frac{2(z-R(z)-\sqrt{d_+d_+})}{(\sqrt{d_+}-\sqrt{d_-})^2 z} \tag{6.10}$$

for $z \notin (d_-, d_+)$.

Proof. The computation of the Stieltjes transform (6.10) is a standard exercise in complex analysis. The log transform $H_{\rm MP}(z)$ is obtained from (6.10) by taking anti-derivative. □

In terms of r_1 and r_2,

$$R(z) = \frac{\sqrt{(r_1 z-1)^2 - 4r_1r_2}}{r_1}, \tag{6.11}$$

$$\begin{aligned} H_{\rm MP}(z) = \frac{1}{2r_2}\Big[& r_1 z - 1 - r_1 R(z) - (r_1-r_2)\log z + \log\left(\frac{r_1z-1+r_1R(z)}{2r_1}\right) \\ & + (r_1-r_2)\log\left(\frac{r_1z-(r_1-r_2)r_1R(z)-(r_1-r_2)^2}{2r_1r_2z}\right)\Big], \end{aligned} \tag{6.12}$$

and

$$s_{\rm MP}(z) = H'_{\rm MP}(z) = \frac{r_1z - r_1R(z)-(r_1-r_2)}{2r_2z}. \tag{6.13}$$

Let us consider the critical inverse temperature β_c. From (6.4), $\beta_c = \frac{\sqrt{r_2}}{r_1}B_c$, and from Definition (3.8), $B_c = \sqrt{d_+(s_{\rm MP}(d_+))^2 + 2\alpha s_{\rm MP}(d_+)}$. From the above explicit formulas, we see that

$$s_{\rm MP}(d_+) = H'_{\rm MP}(d_+) = \frac{r_1}{\sqrt{r_2}(\sqrt{r_1}+\sqrt{r_2})}. \tag{6.14}$$

Hence we find that the critical inverse temperature is

$$\beta_c = (r_1r_2)^{-\frac{1}{4}}. \tag{6.15}$$

6.2 High temperature case

For $\beta < \beta_c$, we evaluate the terms in Proposition 4.4 explicitly. We first find z_c solving the equation (4.14). From (6.4) and (6.13), this equation is

$$\frac{r_1 z_c - r_1 R(z_c) - (r_1 - r_2)}{2r_2 z_c} = \frac{2r_1^2\beta^2}{r_1 - r_2 + \sqrt{W(z_c)}} \tag{6.16}$$

where $R(z)$ is given by (6.11) and

$$W(z) := (r_1 - r_2)^2 + 4r_1^2 r_2 \beta^2 z. \tag{6.17}$$

We claim that the solution of this equation in $z_c \in (d_+, \infty)$ is given by

$$z_c = \frac{(1 + r_1\beta^2)(1 + r_2\beta^2)}{r_1\beta^2} = \frac{1 + \beta^2 + r_1 r_2 \beta^4}{r_1\beta^2}. \tag{6.18}$$

Indeed, with this z_c, we find that

$$\sqrt{W(z_c)} = 1 + 2r_1 r_2 \beta^2, \qquad R(z_c) = \frac{1 - r_1 r_2 \beta^4}{r_1\beta^2} \tag{6.19}$$

where we used the condition that $\beta < (r_1 r_2)^{-1/4}$. Using this, we find that both sides of (6.16) are $\frac{r_1\beta^2}{1+r_2\beta^2}$. This verifies (6.18).

Now from Proposition 4.4 and recalling that $n = N_2$, we find that for any $\epsilon > 0$,

$$\frac{1}{N}\log \mathbf{Q} = \tilde{A} - \frac{1}{2N}\left[\sum_{i=1}^{N_2} \log\left(z_c - \mu_i\right) - N_2 H_{\mathrm{MP}}(z_c)\right] - \frac{\log(r_2 N)}{N} + \frac{1}{2N}\log\left(\frac{\pi^2}{\widehat{D}}\right) + O(N^{-1-\epsilon}) \tag{6.20}$$

with high probability where

$$\tilde{A} = r_2\widehat{A} = \frac{\sqrt{W(z_c)}}{2} - \frac{r_1 - r_2}{2}\log\left(\frac{r_1 - r_2 + \sqrt{W(z_c)}}{4r_1\sqrt{r_2}\beta}\right) - \frac{r_2}{2}H_{\mathrm{MP}}(z_c) \tag{6.21}$$

and

$$\widehat{D} = -\frac{4(r_1 - r_2)}{r_2}H''_{\mathrm{MP}}(z_c) - 8z_c H'_{\mathrm{MP}}(z_c)H''_{MP}(z_c) - 4(H'_{\mathrm{MP}}(z_c))^2. \tag{6.22}$$

It is direct to check that

$$H_{\mathrm{MP}}(z_c) = r_1\beta^2 + \frac{r_1 - r_2}{r_2}\log\left(\frac{1}{1 + r_2\beta^2}\right) - \log\left(r_1\beta^2\right) \tag{6.23}$$

and

$$H'_{\mathrm{MP}}(z_c) = \frac{r_1\beta^2}{1 + r_2\beta^2}, \qquad H''_{\mathrm{MP}}(z_c) = -\frac{r_1^2\beta^4}{(1 + r_2\beta^2)^2(1 - r_1 r_2\beta^4)}. \tag{6.24}$$

Thus, we obtain

$$\tilde{A} = \frac{1}{2}(1 + r_1 r_2\beta^2) - \frac{r_1 - r_2}{2}\log\left(\frac{1}{2\sqrt{r_2}\beta}\right) + \frac{r_2}{2}\log\left(r_1\beta^2\right) \tag{6.25}$$

and

$$\widehat{D} = \frac{4r_1^3\beta^4}{r_2(1-r_1r_2\beta^4)}. \tag{6.26}$$

Recalling (6.2) and using Lemma 6.1, we find that for $\beta < \beta_c$,

$$\begin{aligned}\frac{1}{N}\log Z_{N_1,N_2}(\beta) = \frac{r_1r_2\beta^2}{2} - \frac{1}{2N}\left[\sum_{i=1}^{N_2}\log\left(z_c-\mu_i\right) - N_2H_{\mathrm{MP}}(z_c)\right] \\ +\frac{1}{N}\left[\frac{1}{2}\log\left(1-r_1r_2\beta^4\right)-\log 2\right] + O(N^{-1-\epsilon})\end{aligned} \tag{6.27}$$

with high probability. This proves the part (i) of Theorem 2.4.

6.3 Low temperature case

For $\beta < \beta_c$, Proposition 5.8 implies that (recall that $n = N_2$ and $\frac{N_2}{N} \to r_2$)

$$\frac{1}{N}\log \mathbf{Q}(n,\alpha,B) = \tilde{E} + (\mu_1 - d_+)\tilde{L} + O(N^{-1+\epsilon}). \tag{6.28}$$

where

$$\begin{aligned}\tilde{E} = r_2\widehat{E} &= \frac{\sqrt{W(d_+)}}{2} - \frac{r_1-r_2}{2}\log\left(\frac{r_1-r_2+\sqrt{W(d_+)}}{4r_1\sqrt{r_2}\beta}\right) - \frac{r_2}{2}H_{\mathrm{MP}}(d_+), \\ \tilde{L} = r_2\widehat{L} &= \frac{r_1^2r_2\beta^2}{r_1-r_2+\sqrt{W(d_+)}} - \frac{r_2}{2}H'_{\mathrm{MP}}(d_+)\end{aligned} \tag{6.29}$$

with

$$W(d_+) = (\sqrt{r_1}+\sqrt{r_2})^2((\sqrt{r_1}-\sqrt{r_2})^2 + 4r_1r_2\beta^2). \tag{6.30}$$

From the explicit formulas,

$$H_{\mathrm{MP}}(d_+) = \sqrt{\frac{r_1}{r_2}} - \frac{r_1-r_2}{r_2}\log\left(\frac{\sqrt{r_1}+\sqrt{r_2}}{\sqrt{r_1}}\right) + \log\sqrt{\frac{r_2}{r_1}} \tag{6.31}$$

and

$$H'_{\mathrm{MP}}(d_+) = \frac{r_1}{\sqrt{r_2}(\sqrt{r_1}+\sqrt{r_2})}. \tag{6.32}$$

Setting

$$\mathsf{S} = \mathsf{S}(r_1,r_2) := (\sqrt{r_1}-\sqrt{r_2})^2 + 4r_1r_2\beta^2, \tag{6.33}$$

we have

$$\tilde{E} = \frac{(\sqrt{r_1}+\sqrt{r_2})\sqrt{\mathsf{S}} - \sqrt{r_1r_2}}{2} - \frac{r_1-r_2}{2}\log\left(\frac{\sqrt{\mathsf{S}}+\sqrt{r_1}-\sqrt{r_2}}{4\sqrt{r_1r_2}\beta}\right) - \frac{r_2}{4}\log\frac{r_2}{r_1} \tag{6.34}$$

and

$$\tilde{L} = \frac{r_1\sqrt{r_1r_2}}{2(\sqrt{r_1}+\sqrt{r_2})}\left(\frac{2\sqrt{r_1r_2}\beta^2}{\sqrt{S}+\sqrt{r_1}-\sqrt{r_2}} - \frac{1}{\sqrt{r_1}}\right) = \frac{r_1(\sqrt{\mathsf{S}}-\sqrt{r_1}-\sqrt{r_2})}{4(\sqrt{r_1}+\sqrt{r_2})}. \tag{6.35}$$

Using $\left(\frac{\sqrt{A^2+B^2}+A}{B}\right)^2 = \frac{\sqrt{A^2+B^2}+A}{\sqrt{A^2+B^2}-A}$, we write the log term in $\tilde{E}$ as

$$\log\left(\frac{\sqrt{\mathsf{S}}+\sqrt{r_1}-\sqrt{r_2}}{4\sqrt{r_1 r_2}\beta}\right) = \frac{1}{2}\log\left(\frac{\sqrt{\mathsf{S}}+\sqrt{r_1}-\sqrt{r_2}}{\sqrt{\mathsf{S}}-\sqrt{r_1}+\sqrt{r_2}}\right) - \log 2. \tag{6.36}$$

Hence, recalling (6.2) and using Lemma 6.1, we find that for $\beta > \beta_c$,

$$\begin{aligned}
&\frac{1}{N}\log Z_{N_1,N_2}(\beta) \\
&= \frac{(\sqrt{r_1}+\sqrt{r_2})\sqrt{\mathsf{S}}-\sqrt{r_1 r_2}-1}{2} - \frac{r_1-r_2}{4}\log\left(\frac{\sqrt{\mathsf{S}}+\sqrt{r_1}-\sqrt{r_2}}{\sqrt{\mathsf{S}}-\sqrt{r_1}+\sqrt{r_2}}\right) - \frac{r_2}{4}\log r_1 - \frac{r_1}{4}\log r_2 - \frac{1}{2}\log\beta \\
&\quad + \left(\mu_1 - \frac{(\sqrt{r_1}+\sqrt{r_2})^2}{r_1}\right)\frac{r_1(\sqrt{\mathsf{S}}-\sqrt{r_1}-\sqrt{r_2})}{4(\sqrt{r_1}+\sqrt{r_2})} + O(N^{-1+\epsilon})
\end{aligned}$$

with high probability. This completes the proof of the part (ii) of Theorem 2.4.

7 Proof of Theorem 2.2

We derive Theorem 2.2 from Theorem 2.4.

7.1 High temperature

When $\beta < \beta_c$, from Theorem 2.4, we need the behavior of the sum of $\log(z_c - \mu_i)$. It follows form the following result in random matrix theory.

Proposition 7.1 (Linear statistics of the eigenvalues)**.** *For any function $\varphi : \mathbb{R} \to \mathbb{R}$ that is analytic in an open neighborhood of the support of μ_{MP}, the random variable*

$$\sum_{i=1}^{N_2} \varphi(\mu_i) - N_2 \int \varphi(x)\mathrm{d}\mu_{\mathrm{MP}}(x) \tag{7.1}$$

converges in distribution as $N_2 \to \infty$ to a Gaussian random variable with mean $M(\varphi)$ and variance $V(\varphi)$ given as follows: setting

$$\Phi(x) = \varphi\left(\frac{d_+ - d_-}{4}x + \frac{d_+ + d_-}{2}\right), \tag{7.2}$$

we have

$$M(\varphi) = M_{\mathrm{GOE}}(\Phi) - (W_4 - 3)\tau_2(\Phi), \qquad V(\varphi) = V_{\mathrm{GOE}}(\Phi) + (W_4 - 3)\tau_1(\Phi)^2 \tag{7.3}$$

where

$$\begin{aligned}
M_{\mathrm{GOE}}(\Phi) &= \frac{\Phi(-2)+\Phi(2)}{4} - \frac{\tau_0(\Phi)}{2}, \\
V_{\mathrm{GOE}}(\Phi) &= \frac{1}{2\pi^2}\int_{d_-}^{d_+}\int_{d_-}^{d_+}\left(\frac{\Phi(x_1)-\Phi(x_2)}{x_1-x_2}\right)^2 \frac{4-x_1x_2}{\sqrt{4-x_1^2}\sqrt{4-x_2^2}}\mathrm{d}x_1\mathrm{d}x_2.
\end{aligned} \tag{7.4}$$

and for $\ell = 0, 1, 2, \cdots$,

$$\tau_\ell(\Phi) = \frac{1}{2\pi}\int_{-\pi}^{\pi}\Phi(2\cos\theta)\cos(\ell\theta)\mathrm{d}\theta. \tag{7.5}$$

This result was first obtained in [4] (equation (5.13)). It was also obtained in [29] (Theorem 4.5) and [6] (Theorem 1.1). In the above formulas, GOE stands for Gaussian orthogonal ensemble, another classical random matrix ensemble.

To complete the proof of the part (i) of Theorem 2.2, we need to evaluate $M(\varphi)$ and $V(\varphi)$ when $\varphi(x) = \log(z_c - x)$. In this case,

$$\Phi(y) = \log\left(\frac{d_+ - d_-}{4}\right) + \log(\tilde{z} - y), \qquad \tilde{z} = \frac{4}{d_+ - d_-}\left(z_c - \frac{d_+ + d_-}{2}\right). \tag{7.6}$$

Since $\tau_0(1) = 1$, we find that

$$\tau_\ell(\Phi) = \delta_{\ell=0}\log\left(\frac{d_+ - d_-}{4}\right) + \tau_\ell(\Phi_0), \qquad \Phi_0(x) := \log(\tilde{z} - x). \tag{7.7}$$

We computed in Appendix A of [7] that for $\Phi_0(x) = \log(\tilde{z} - x)$,

$$\tau_0(\Phi_0) = \log(\tilde{z} + \sqrt{\tilde{z}^2 - 4}) - \log 2, \quad \tau_1(\Phi_0) = \frac{\sqrt{\tilde{z}^2 - 4}}{2} - \frac{\tilde{z}}{2}, \quad \tau_2(\Phi_0) = \frac{\tilde{z}\sqrt{\tilde{z}^2 - 4}}{4} - \frac{\tilde{z}^2}{4} + \frac{1}{2}.$$

Note that

$$\sqrt{\tilde{z}^2 - 4} = \frac{4}{d_+ - d_-}\sqrt{(z_c - d_-)(z_c - d_+)} = \frac{4}{d_+ - d_-}R(z_c), \tag{7.8}$$

which can be checked from the definition (6.9). Hence,

$$\begin{aligned} M_{\mathrm{GOE}}(\Phi) &= \frac{\log R(z_c)}{2} - \frac{1}{2}\left[\log\left(z_c - \frac{d_+ + d_-}{2} + R(z_c)\right) - \log 2\right], \\ \tau_1(\Phi_0) &= \frac{2}{d_+ - d_-}\left[R(z_c) - \left(z_c - \frac{d_+ + d_-}{2}\right)\right], \\ \tau_2(\Phi_0) &= \frac{4\left(z_c - \frac{d_+ + d_-}{2}\right)}{(d_+ - d_-)^2}\left[R(z_c) - \left(z_c - \frac{d_+ + d_-}{2}\right)\right] + \frac{1}{2}. \end{aligned} \tag{7.9}$$

On the other hand, from Lemma A.1 of [7],

$$V_{\mathrm{GOE}}(\Phi) = 2\log\left(\frac{z_c - \frac{d_+ + d_-}{2} + R(z_c)}{2R(z_c)}\right). \tag{7.10}$$

Inserting (6.5) for d_+ and d_+, and recalling (6.19) for $R(z_c)$, we find that

$$\begin{aligned} &\tau_1(\Phi_0) = -\sqrt{r_1 r_2}\beta^2, \qquad \tau_2(\Phi_0) = -\frac{r_1 r_2 \beta^4}{2} \\ &M_{\mathrm{GOE}}(\Phi) = \frac{1}{2}\log\left(1 - r_1 r_2 \beta^4\right), \qquad V_{\mathrm{GOE}}(\Phi) = -2\log\left(1 - r_1 r_2 \beta^4\right). \end{aligned} \tag{7.11}$$

Hence, we have

$$\begin{aligned} M(\varphi) &= \frac{1}{2}\log\left(1 - r_1 r_2 \beta^4\right) + (W_4 - 3)\frac{r_1 r_2 \beta^4}{2}, \\ V(\varphi) &= -2\log\left(1 - r_1 r_2 \beta^4\right) + (W_4 - 3) r_1 r_2 \beta^4. \end{aligned} \tag{7.12}$$

for $\varphi(x) = \log(z_c - x)$.

Part (i) of Theorem 2.4, Prosition 7.1 and (7.12) prove the part (i) of Theorem 2.2 when $r_1 \geq r_2$. The case when $r_1 < r_2$ follows from the symmetry, noting that $F(\beta)$, μ and σ^2 are all symmetric in r_1 and r_2.

7.2 Low temperature

When $\beta < \beta_c$, we need the behavior of the top eigenvalue μ_1. The following result is well-known in random matrix theory. See, for example, [25, 37] and also Corollary 1.2 of [35].

Proposition 7.2 (Tracy–Widom limit of the largest eigenvalue). *We have*

$$\frac{N_1}{\sqrt{N_1}+\sqrt{N_2}}\left(\frac{1}{\sqrt{N_1}}+\frac{1}{\sqrt{N_2}}\right)^{-\frac{1}{3}}(\mu_1 - d_+) \Rightarrow \mathrm{TW} \tag{7.13}$$

in distribution.

In terms of r_1, r_2,

$$\frac{N_1}{\sqrt{N_1}+\sqrt{N_2}}\left(\frac{1}{\sqrt{N_1}}+\frac{1}{\sqrt{N_2}}\right)^{-\frac{1}{3}} = \frac{r_1(r_1r_2)^{\frac{1}{6}}}{(\sqrt{r_1}+\sqrt{r_2})^{\frac{4}{3}}}N^{2/3}(1+O(N^{-1-\delta})). \tag{7.14}$$

Then, combining with the constant in (2.29),

$$\frac{(\sqrt{r_1}+\sqrt{r_2})^{\frac{4}{3}}}{r_1(r_1r_2)^{\frac{1}{6}}}\frac{r_1(\sqrt{\mathsf{S}}-\sqrt{r_1}-\sqrt{r_2})}{4(\sqrt{r_1}+\sqrt{r_2})} = \frac{(\sqrt{r_1}+\sqrt{r_2})^{1/3}(\sqrt{\mathsf{S}}-\sqrt{r_1}-\sqrt{r_2})}{4(r_1r_2)^{1/6}}. \tag{7.15}$$

This proves the part (ii) of Theorem 2.2 when $r_1 \geq r_2$. The case when $r_1 < r_2$ again follows from the symmetry noting that $F(\beta)$ and A are symmetric in r_1 and r_2.

8 Non-identically distributed disorders

In this section, we briefly discuss the case where the disorders are non-identically distributed. Let Σ be a positive-definite matrix of size $N_1 \times N_1$. Let J be the matrix of i.i.d. entries as before. Consider the new Hamiltonian

$$H(\boldsymbol{\sigma},\boldsymbol{\tau}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}(\Sigma^{1/2}J)_{ij}\sigma_i\tau_j. \tag{8.1}$$

The new disorder parameters are $(\Sigma^{1/2}J)_{ij} = \sum_k \Sigma_{ik}^{1/2}J_{kj}$. In particular, when $\Sigma^{1/2}$ is a diagonal matrix, the variances of the disorder parameters depend on the index i but not on j.

The associated random matrix is

$$S = \frac{1}{N_1}J^T\Sigma J. \tag{8.2}$$

In statistics, S is known as a sample covariance matrix with general population covariance matrix Σ. Recall the ingredients of the analysis we have done in this paper. The double integral formula in Lemma 2.5, which was the starting point of our analysis, holds for any sample covariance matrices: we set $M = \frac{1}{\sqrt{N_1}}\Sigma^{1/2}J$ in the proof. The proofs of asymptotic formulas for $\mathbf{Q}_n$ in Proposition 4.4 and Proposition 5.8 require the regularity of measure in Condition 3.3 and the rigidity of the eigenvalues in Condition 3.4. Finally, the fluctuation of the free energy was obtained by applying the linear statistics of the eigenvalues in Proposition 7.1 and the Tracy–Widom limit of the largest eigenvalue in Proposition 7.2.

The limiting empirical spectral distribution (ESD) of S is well studied in random matrix theory. Under a very general assumption, it was proved in [36] that the limiting ESD of S is regular. The typical assumption is as follows: Let $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{N_1}$ be the eigenvalues of Σ and denote by $\widehat{\sigma}$ the empirical spectral distribution of Σ. Define ξ_+ as the unique solution in $(0, \sigma_1^{-1})$ to the equation

$$\int \left(\frac{t\xi_+}{1 - t\xi_+} \right)^2 \mathrm{d}\widehat{\sigma}(t) = \frac{r_2^2}{r_1^2}. \tag{8.3}$$

If

$$\limsup \sigma_1 < \infty, \quad \liminf \sigma_{N_1} > 0, \quad \limsup \sigma_1 \xi_+ < 1, \tag{8.4}$$

then the limiting ESD of S is regular, and hence satisfy Condition 3.3. (See Theorem 3.1 of [9].) This assumption basically means that the largest eigenvalue σ_1 of Σ is not too far away from the rest of the eigenvalues. Such an assumption was also used in [19, 28, 26]. Under the same assumption, it was proved by Knowles and Yin [26] that Condition 3.4 holds with high probability.

Proposition 7.1, the central limit theorem for the linear statistics of the eigenvalues was first proved by Bai and Silverstein [4] when W_4, the fourth moment of J_{ij}, matches that of the standard Gaussian, which is 3. It was later extended to a general case with any finite fourth moment by Najim and Yao [32].

Proposition 7.2 was first proved for the complex case with Gaussian disorder J_{ij} by El Karoui [19]. Bao, Pan, and Zhou [9] proved the edge universality for the model, which asserts that the rescaled distribution of the largest eigenvalue does not depend on the distribution of J_{ij}. The result in [9] also holds for the real case, but the Tracy–Widom limit for the case was not proved. Proposition 7.2 was later proved in [28, 26].

Therefore, the main theorems, Theorem 2.2 and Theorem 2.4, also hold for the Hamiltonian with non-identically distributed disorders under a general assumption on the spectrum of the matrix Σ, with suitable changes on the constants μ, σ^2, A, and S.

References

[1] M. Aizenman, J. L. Lebowitz, and D. Ruelle. Some rigorous results on the Sherrington-Kirkpatrick spin glass model. *Comm. Math. Phys.*, 112(1):3–20, 1987.

[2] O. Ajanki, L. Erdős, and T. Krüger. Local semicircle law with imprimitive variance matrix. *Electron. Commun. Probab.*, 19:no. 33, 9, 2014.

[3] A. Auffinger and W.-K. Chen. Free energy and complexity of spherical bipartite models. *J. Stat. Phys.*, 157(1):40–59, 2014.

[4] Z. Bai and J. W. Silverstein. CLT for linear spectral statistics of large-dimensional sample covariance matrices. *Ann. Probab.*, 32(1A):553–605, 2004.

[5] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices.* Springer Series in Statistics. Springer, New York, second edition, 2010.

[6] Z. Bai, X. Wang, and W. Zhou. Functional CLT for sample covariance matrices. *Bernoulli*, 16(4):1086–1113, 2010.

[7] J. Baik and J. O. Lee. Fluctuations of the free energy of the spherical Sherrington-Kirkpatrick model. *J. Stat. Phys.*, 165(2):185–224, 2016.

[8] J. Baik and J. O. Lee. Fluctuations of the free energy of the spherical Sherrington-Kirkpatrick model with ferromagnetic interaction. *Ann. Henri Poincaré*, 18(6):1867–1917, 2017.

[9] Z. Bao, G. Pan, and W. Zhou. Universality for the largest eigenvalue of sample covariance matrices with general population. *Ann. Statist.*, 43(1):382–421, 2015.

[10] A. Barra, P. Contucci, E. Mingione, and D. Tantari. Multi-species mean field spin glasses. Rigorous results. *Ann. Henri Poincaré*, 16(3):691–708, 2015.

[11] A. Barra, A. Galluzzi, F. Guerra, A. Pizzoferrato, and D. Tantari. Mean field bipartite spin models treated with mechanical techniques. *Eur. Phys. J. B*, 87(3):Art. 74, 13, 2014.

[12] A. Barra, G. Genovese, and F. Guerra. Equilibrium statistical mechanics of bipartite spin systems. *J. Phys. A*, 44(24):245002, 22, 2011.

[13] P. Bourgade, L. Erdös, and H.-T. Yau. Edge universality of beta ensembles. *Comm. Math. Phys.*, 332(1):261–353, 2014.

[14] A. Bovier, I. Kurkova, and M. Löwe. Fluctuations of the free energy in the REM and the p-spin SK models. *Ann. Probab.*, 30(2):605–651, 2002.

[15] W.-K. Chen, P. Dey, and D. Panchenko. Fluctuations of the free energy in the mixed p-spin models with external field. *Probab. Theory Related Fields*, 168(1-2):41–53, 2017.

[16] W.-K. Chen and A. Sen. Parisi formula, disorder chaos and fluctuation for the ground state energy in the spherical mixed p-spin models. *Comm. Math. Phys.*, 350(1):129–173, 2017.

[17] F. Comets and J. Neveu. The Sherrington-Kirkpatrick model of spin glasses and stochastic calculus: the high temperature case. *Comm. Math. Phys.*, 166(3):549–564, 1995.

[18] A. Crisanti and H. J. Sommers. The spherical p-spin interaction spin glass model: the statics. *Z. Phys. B. Condensed Matter*, 87(3):341–354, 1992.

[19] N. El Karoui. A rate of convergence result for the largest eigenvalue of complex white Wishart matrices. *Ann. Probab.*, 34(6):2077–2117, 2006.

[20] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin. Spectral statistics of Erdős-Rényi graphs I: Local semicircle law. *Ann. Probab.*, 41(3B):2279–2375, 2013.

[21] L. Erdős, H.-T. Yau, and J. Yin. Rigidity of eigenvalues of generalized Wigner matrices. *Adv. Math.*, 229(3):1435–1515, 2012.

[22] P. J. Forrester. *Log-gases and random matrices*, volume 34 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2010.

[23] J. Fröhlich and B. Zegarliński. Some comments on the Sherrington-Kirkpatrick model of spin glasses. *Comm. Math. Phys.*, 112(4):553–566, 1987.

[24] F. Guerra. Spontaneous replica symmetry breaking and interpolation methods for complex statistical mechanics systems. In *Correlated random systems: five different methods*, volume 2143 of *Lecture Notes in Math.*, pages 45–70. Springer, Cham, 2015.

[25] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327, 2001.

[26] A. Knowles and J. Yin. Anisotropic local laws for random matrices. *Probab. Theory Related Fields*, 169(1-2):257–352, 2017.

[27] J. Kosterlitz, D. Thouless, and R. Jones. Spherical model of a spin-glass. *Phys. Rev. Lett.*, 36(20):1217–1220, 1976.

[28] J. O. Lee and K. Schnelli. Tracy-Widom distribution for the largest eigenvalue of real sample covariance matrices with general population. *Ann. Appl. Probab.*, 26(6):3786–3839, 2016.

[29] A. Lytova and L. Pastur. Central limit theorem for linear eigenvalue statistics of random matrices with independent entries. *Ann. Probab.*, 37(5):1778–1840, 2009.

[30] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb. (N.S.)*, 72 (114):507–536, 1967.

[31] M. L. Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Amsterdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.

[32] J. Najim and J. Yao. Gaussian fluctuations for linear spectral statistics of large random covariance matrices. *Ann. Appl. Probab.*, 26(3):1837–1887, 2016.

[33] D. Panchenko. The free energy in a multi-species Sherrington-Kirkpatrick model. *Ann. Probab.*, 43(6):3494–3513, 2015.

[34] G. Parisi. A sequence of approximate solutions to the S-K model for spin glasses. *J. Phys. A*, 13(4):L115–L121, 1980.

[35] N. S. Pillai and J. Yin. Universality of covariance matrices. *Ann. Appl. Probab.*, 24(3):935–1001, 2014.

[36] J. W. Silverstein and S.-I. Choi. Analysis of the limiting spectral distribution of large-dimensional random matrices. *J. Multivariate Anal.*, 54(2):295–309, 1995.

[37] A. Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. *J. Statist. Phys.*, 108(5-6):1033–1056, 2002. Dedicated to David Ruelle and Yasha Sinai on the occasion of their 65th birthdays.

[38] E. Subag and O. Zeitouni. The extremal process of critical points of the pure p-spin spherical spin glass model. *Probab. Theory Related Fields*, 168(3-4):773–820, 2017.

[39] M. Talagrand. Free energy of the spherical mean field model. *Probab. Theory Related Fields*, 134(3):339–382, 2006.

[40] M. Talagrand. The Parisi formula. *Ann. of Math. (2)*, 163(1):221–263, 2006.