

HIGH TEMPERATURE ASYMPTOTICS OF ORTHOGONAL MEAN-FIELD SPIN GLASSES

BHASWAR B. BHATTACHARYA AND SUBHABRATA SEN

ABSTRACT. We evaluate the high temperature limit of the free energy of spin glasses on the hypercube with Hamiltonian $H_N(\underline{\sigma}) = \underline{\sigma}^T J \underline{\sigma}$, where the coupling matrix J is drawn from certain symmetric orthogonally invariant ensembles. Our derivation relates the *annealed free energy* of these models to a spherical integral, and expresses the limit of the free energy in terms of the limiting spectral measure of the coupling matrix J . As an application, we derive the limiting free energy of the Random Orthogonal Model (ROM) at high temperatures, which confirms non-rigorous calculations of Marinari et al. [20]. Our methods also apply to other well-known models of disordered systems, including the SK and Gaussian Hopfield models.

1. INTRODUCTION

Consider a (random) function on the hypercube $H_N : S_N = \{-1, +1\}^N \rightarrow \mathbb{R}$ defined as

$$H_N(\underline{\sigma}) = \underline{\sigma}^T J \underline{\sigma} \tag{1.1}$$

with coupling matrix $J = ODO^T$, where O is Haar distributed over the orthogonal group $O(N)$ and $D = \text{diag}(d_1, \dots, d_N)$ is a diagonal matrix independent of O . This defines a probability distribution over S_N as follows: for $\tau \in S_N$ and $\beta \geq 0$,

$$\mathbb{P}(\underline{\sigma} = \tau) = \frac{1}{2^N} \cdot \frac{e^{\beta H_N(\tau)}}{Z_N(\beta, O, D)}, \tag{1.2}$$

where the *partition function* $Z_N(\beta, O, D) = \frac{1}{2^N} \sum_{\underline{\sigma} \in S_N} \exp(\beta H_N(\underline{\sigma}))$. These distributions arise frequently in the analysis of disordered systems in statistical physics. In this context, $H_N(\underline{\sigma})$ describes the energy of the configuration $\underline{\sigma}$, and is usually referred to as the Hamiltonian of the system. The parameter β denotes the inverse temperature, so the *high temperature* regime corresponds to small values of β . We seek to evaluate the large N limit of the *free energy*

$$\Phi_N(\beta, O, D) = \frac{1}{N} \log Z_N(\beta, O, D) \tag{1.3}$$

in these models.

Models of the form (1.2) will be referred to as *orthogonal mean-field spin glasses*— they include many well-known physical models of disordered systems:

- (a) *Sherrington-Kirkpatrick (SK) Model*: In the SK model of spin glasses the coupling matrix $J = \frac{1}{\sqrt{N}}W$, where W is a symmetric matrix drawn from the *Gaussian Orthogonal Ensemble*. It is well known that $W = ODO^T$, where $O \sim O(N)$ is Haar distributed and $D = \text{diag}(d_1, d_2, \dots, d_N)$ is a diagonal matrix independent of O , such that the empirical measure $\frac{1}{N} \sum_{i=1}^N \delta_{d_i}$ converges to the semi-circle law [2]. The limit of the free energy for all temperatures

Date: March 13, 2018.

2010 *Mathematics Subject Classification*. 60F10, 15B10, 82B44.

Key words and phrases. Large deviations, Random orthogonal matrices, Spherical integrals, Spin glasses.

was conjectured by Parisi using deep ideas of replica symmetry breaking, and was rigorously established by Talagrand [24] (refer to [22] for an introduction to this subject). Carmona and Hu [5] (see also Chatterjee [7]) proved that the Parisi formula continues to hold even if the entries of the coupling matrix $J = ((J_{ij}))$ are independent mean zero random variables, subject to some conditions on the higher moments.

- (b) *Random Orthogonal Model (ROM)*: Marinari et al. [20] introduced the ROM to model a deterministic system which exhibits glassy behavior. In this model the coupling matrix $J = ODO^T$, where $D = \text{diag}(d_1, \dots, d_N)$ is a deterministic sequence of $\{\pm 1\}$ such that the empirical measure

$$\mu_N(D) = \frac{1}{N} \sum_{i=1}^N \delta_{d_i} \xrightarrow{D} p\delta_1 + (1-p)\delta_{-1}, \quad (1.4)$$

for some $p \in (0, 1)$. The case $p = 1/2$ has received a lot of attention in the physics literature (see [3, 12, 19] and the references therein). The limiting free energy of this model is not known rigorously even in the high temperature regime. The coupling matrix J has dependent entries and non-rigorous calculations based on the replica method predict different behavior compared to the SK model [8, 19, 20]. This suggests that comparison/universality techniques like [5, 7] cannot be directly used to compute the free energy.

- (c) *Gaussian Hopfield Model*: Cherrier et al. [8] considered the Gaussian Hopfield Model where the coupling matrix $J = \frac{1}{p}XX^T$, where $X = ((X_{ij}))$ is a $N \times p$ matrix with i.i.d. $\mathcal{N}(0, 1)$. The coupling matrix of the usual Hopfield model has the same structure, but the matrix X consists of i.i.d. Rademacher $\{\pm 1\}$ random variables. Bovier et al. [4] studied the Gaussian Hopfield model with 2-patterns and this “simple” case already shows highly complicated behavior. It is generally believed that a Hopfield model with p parameters where $p \sim \lambda N$ is significantly more complicated compared to the one with a finite number of patterns.

This paper gives a general method for computing the limit of the free energy in orthogonal mean-field spin glass models at sufficiently high temperatures (see Theorem 1.2). Exploiting a connection with spherical integrals [6, 16] and using techniques from large deviations and random matrix theory, we rigorously justify certain heuristics employed in the traditional analyses of these systems. In particular, we derive:

1. the limiting free energy of the SK model in the entire high temperature phase (Corollary 2.1), re-deriving the classical result of Aizenman et al. [1],
2. the limiting free energy of ROM for β sufficiently small (Corollary 1.3), which verifies predictions of Marinari et al. [20], and
3. the limiting free energy of the Gaussian Hopfield model with $p/N \rightarrow \lambda \in (1, \infty)$ for high temperatures, confirming non-rigorous calculations of Cherrier et al. [8]. We remark that our techniques should also apply to the case $\lambda \in (0, 1)$ but we restrict ourselves to the first case for clarity.

1.1. Main Results. To state our main results we need to introduce some notations. The Haar measure on the orthogonal group $O(N)$ will be denoted by dO , and the expectation of a function f will be denoted by $\mathbb{E}_O f(O) := \int_O f(O) dO$.

For any probability measure μ , denote by supp the support of μ . We will always consider probability measures with bounded support so that $\text{supp}(\mu) \subseteq [\lambda_{\min}, \lambda_{\max}]$. To describe our results we need to introduce the Hilbert transform and the R -transform of a probability measure with bounded support:

Definition 1.1. The Hilbert transform H_μ of a measure μ is $H_\mu : \mathbb{R} \setminus \text{supp}(\mu) \rightarrow \mathbb{R}$

$$z \mapsto \int \frac{1}{z - \lambda} d\mu(\lambda). \quad (1.5)$$

It is easy to show that H_μ is a bijective map from $\mathbb{R} \setminus \text{supp}(\mu)$ to $(H_{\min}, H_{\max}) \setminus \{0\}$ (see [16]), where

$$H_{\max} = \lim_{z \downarrow \lambda_{\max}} H_\mu(z) \quad H_{\min} = \lim_{z \uparrow \lambda_{\min}} H_\mu(z). \quad (1.6)$$

Thus, setting $x_{\min} = \lambda_{\min} - 1/H_{\min}$, $x_{\max} = \lambda_{\max} - 1/H_{\max}$, and $m = \int \lambda d\mu_\lambda$, for $z \in H_\mu(\mathbb{R} \setminus \text{supp}(\mu))$, define the R -transform $R_\mu : (H_{\min}, H_{\max}) \setminus \{0\} \rightarrow (x_{\min}, x_{\max}) \setminus \{m\}$ as

$$H_\mu \left(R_\mu(z) + \frac{1}{z} \right) = z. \quad (1.7)$$

It is easy to see that R_μ is bijective and we denote its inverse by Q_μ . Let

$$I_\mu(\beta) = \frac{1}{2} \int_0^{2\beta} R_\mu(v) dv. \quad (1.8)$$

Finally, for any $\beta > 0$, define

$$U_L = 2\beta(1 - \tanh \beta(x_{\max} - x_{\min})), \quad U_R = 2\beta(1 + \tanh \beta(x_{\max} - x_{\min})). \quad (1.9)$$

We will restrict ourselves to models where the sequence of random empirical measures $\mu_N(D) = \frac{1}{N} \sum_{i=1}^N \delta_{d_i}$ corresponding to the matrix D in (1.2) satisfy certain ‘‘rigidity’’ properties. This allows us to neglect the fluctuations of the spectrum in the calculation of the free energy limit. We impose the following property on the law of the matrix D .

Hypothesis 1. Let $D = \text{diag}(d_1, d_2, \dots, d_N)$ be a (random) diagonal matrix with empirical measure $\mu_N(D) = \frac{1}{N} \sum_{i=1}^N \delta_{d_i}$. Assume that

- (a) there exists a sequence of numbers $M_N = o(\sqrt{N})$ such that

$$\lim_{N \rightarrow \infty} \mathbb{P}(\|D\|_\infty > M_N) = 0,$$

where $\|D\|_\infty = \max_{1 \leq i \leq N} |d_i|$;

- (b) there exists a deterministic measure ν_N supported on N points in \mathbb{R} such that for any $c > 0$,

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(W_2(\mu_N(D), \nu_N) > \frac{c}{\sqrt{N}} \right) \rightarrow 0 \quad (1.10)$$

where $W_2(\cdot, \cdot)$ is the 2-Wasserstein distance between two probability measures.

In most of our applications, it suffices to take $\nu_N = \frac{1}{N} \sum_{i=1}^N \delta_{\mathbb{E}(d_i)}$. It can be easily checked that all our results continue to hold with any sequence of probability measures ν_N satisfying Hypothesis 1. However, we state our results with $\nu_N = \sum_{i=1}^N \delta_{\mathbb{E}(d_i)}$ for clarity. We define, for any deterministic diagonal matrix $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$,

$$\Gamma_N(\beta, \Lambda) = \frac{1}{N} \mathbb{E}_0(\log Z_N(\beta, O, \Lambda)). \quad (1.11)$$

Note that due to the invariance of the Haar measure on $O(N)$, $\Gamma_N(\beta, \Lambda)$ is only a function of the empirical distribution $\mu_N(\Lambda) = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i}$. The following proposition establishes that we may neglect the fluctuations of the spectrum for the calculation of the free energy.

Proposition 1.1. *Consider an orthogonal mean field spin glass model (1.2) with a (random) diagonal matrix $D = \text{diag}(d_1, d_2, \dots, d_N)$. If the sequence of measures $\mu_N(D) = \frac{1}{N} \sum_{i=1}^N \delta_{d_i}$ satisfies Hypothesis 1, then*

$$|\Phi_N(\beta, O, D) - \Gamma_N(\beta, \mathbb{E}(D))| \xrightarrow{P} 0. \quad (1.12)$$

The proof of Proposition 1.1 is outlined in Section 3.1. Given this result, to compute the limit of the free energy $\lim_{N \rightarrow \infty} \Phi_N(\beta, O, D)$ it suffices to compute the limit of $\Gamma_N(\beta, \mathbb{E}(D))$.

A crucial ingredient in the analysis of the asymptotics of $\Gamma_N(\beta, \Lambda)$ is a connection with a spherical integral. Guionnet and Maida [16] derived the asymptotics of these integrals in terms of the R -transform of the limit μ of the empirical measure $\mu_N(\Lambda) = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i}$ (refer to Section 1.2 for details). They assume the following conditions on the measure $\mu_N(\Lambda)$:

Hypothesis 2. For a deterministic diagonal matrix $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$, denote by $\mu_N(\Lambda) = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i}$ the empirical measure of Λ . Assume that

- (a) the sequence of measures $\{\mu_N(\Lambda)\}_{N \geq 1}$ converges weakly to a compactly supported measure μ , and
- (b) $\lambda_{\min}(\Lambda) := \min_{1 \leq i \leq N} \lambda_i$ and $\lambda_{\max}(\Lambda) := \max_{1 \leq i \leq N} \lambda_i$ converge to λ_{\min} and λ_{\max} which are finite.

We will also assume Hypothesis 2 to determine the limit of the partition function $Z_N(\beta, O, D)$. We have the following general result for the limiting free energy at high temperature.

Theorem 1.2. *Consider an orthogonal mean field spin glass model (1.2) with a (random) diagonal matrix $D = \text{diag}(d_1, d_2, \dots, d_N)$. Assume that*

- (a) the sequence of (random) measures $\mu_N(D) = \frac{1}{N} \sum_{i=1}^N \delta_{d_i}$ satisfies Hypothesis 1, and
- (b) the sequence of deterministic measures $\mu_N(\mathbb{E}(D)) = \frac{1}{N} \sum_{i=1}^N \delta_{\mathbb{E}(d_i)} \xrightarrow{D} \mu$ and satisfies Hypothesis 2.
- (c) $\limsup_{\beta \rightarrow 0} \beta^2 \sup_{\beta_0 \in [U_L, U_R]} |R'(\beta_0)| < \frac{1}{4}$, where U_L and U_R are as defined in (1.9).

Then for β sufficiently small (depending on μ),

$$\Phi_N(\beta, O, D) \xrightarrow{P} I_\mu(\beta), \quad (1.13)$$

with I_μ defined in (1.8).

As a consequence of the above theorem, we obtain the limiting free energy for many well-known models of disordered systems. Most importantly, we derive the limiting free energy of ROM (1.4) for β sufficiently small (Corollary 1.3), which matches the predictions of Marinari et al. [20] obtained by non-rigorous methods. The limiting free energy for the case $p = 1/2$ is given in the following corollary. Refer to Proposition 2.2 for the expression for any $p \in (0, 1)$.

Corollary 1.3. *For the random orthogonal model (ROM) with $p = 1/2$, for β sufficiently small,*

$$\frac{1}{N} \log Z_N(\beta, O, D) \xrightarrow{P} \frac{1}{4} \left(\sqrt{16\beta^2 + 1} + \log \left(\frac{\sqrt{16\beta^2 + 1} - 1}{8\beta^2} \right) - 1 \right). \quad (1.14)$$

Using Theorem 1.2 we can also obtain the limiting free energy of the SK model in the entire high temperature phase (Corollary 2.1), re-deriving the classical result of Aizenman et al. [1]. Our calculations also give the limiting free energy for the Gaussian Hopfield model at high temperatures, verifying non-rigorous calculations of Cherrier et al. [8].

1.2. Proof Outline and Connections to Spherical Integrals. Spherical integrals over the orthogonal group $O(N)$ (also known as Harish Chandra-Itzykson-Zuber (HCIZ) integrals [6]) are integrals of the form

$$\int_{O(N)} \exp(N \operatorname{tr}(OD_N O^T E_N)) dO, \quad (1.15)$$

where D_N and E_N are $N \times N$ diagonal matrices. HCIZ integrals have been studied due to their connection to matrix models and the enumeration of planar maps (refer [17] and the references therein). Asymptotics of spherical integrals was studied by Guionnet and Maida [16] in the regime where the rank of D_N is small compared to N . An alternative simpler proof was provided in [9].

To see the connection of such integrals to mean-field orthogonal spin glass models consider the *annealed free energy* of the model (1.2): $\phi_N(\beta, \Lambda) = \frac{1}{N} \log \mathbb{E}_0 Z_N(\beta, O, \Lambda)$, where $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$ is a deterministic diagonal matrix. Note that

$$Z_N(\beta, O, \Lambda) = \frac{1}{2^N} \sum_{\underline{\sigma} \in S_N} \exp(\beta \underline{\sigma}^T O \Lambda O^T \underline{\sigma}) = \frac{1}{2^N} \sum_{\underline{\sigma} \in S_N} \exp\left(N\beta \operatorname{tr}\left\{O \Lambda O^T \left(\frac{\underline{\sigma} \underline{\sigma}^T}{N}\right)\right\}\right). \quad (1.16)$$

By the spectral decomposition $\frac{\underline{\sigma} \underline{\sigma}^T}{N} = P_{\underline{\sigma}} E_{11} P_{\underline{\sigma}}$ where $E_{11} = \operatorname{diag}(1, 0, \dots, 0)$. Using (1.16) and the invariance of the Haar distribution,

$$\mathbb{E}_0(Z_N(\beta, O, \Lambda)) = \mathbb{E}_0 \exp\left(N\beta \operatorname{tr}\left\{O \Lambda O^T E_{11}\right\}\right). \quad (1.17)$$

This is exactly of the form (1.15) with $D_N = \Lambda$ and $E_N = E_{11} = \operatorname{diag}(1, 0, \dots, 0)$. Therefore, the annealed free energy $\phi_N(\beta, \Lambda)$ for any deterministic diagonal matrix Λ , is given by a spherical integral. The limit of $\phi_N(\beta, \Lambda)$ was derived by Guionnet and Maida [16], when Hypothesis 2 holds:

Theorem 1.4 (Guionnet and Maida [16]). *Consider an orthogonal mean field spin glass model (1.2) with a deterministic diagonal matrix $\Lambda = \operatorname{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$. If the sequence of empirical measures $\mu_N(\Lambda) = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i} \xrightarrow{D} \mu$ and Hypothesis 2 holds, then for β sufficiently small (depending on μ)*

$$\lim_{N \rightarrow \infty} \phi_N(\beta, \Lambda) = I_{\mu}(\beta). \quad (1.18)$$

The proof of Theorem 1.2 proceeds as follows: when D is random in (1.2), then under Hypothesis 1 we can replace the random matrix D by the deterministic matrix $\mathbb{E}(D)$. Theorem 1.2 then involves computing the limit of the annealed free energy $\phi_N(\beta, \mathbb{E}(D))$ using the above theorem, and the corresponding second moment. This together with results about concentration of measure gives the desired result.

Remark 1.1. When D is random, another natural approach is to compute the *total annealed free energy* $\phi_N^{\text{ann}}(\beta) = \frac{1}{N} \log \mathbb{E}(Z_N(\beta, O, D))$, where the expectation is respect to the joint distribution of (O, D) . From (1.16) it is easy to see that

$$\phi_N^{\text{ann}}(\beta) = \frac{1}{N} \log \mathbb{E}\left(N\beta \frac{\sum_{i=1}^N d_i X_i^2}{\sum_{i=1}^N X_i^2}\right), \quad (1.19)$$

where the X_i are i.i.d. $\mathcal{N}(0, 1)$ random variables.

It is expected that for β sufficiently small, this also gives the correct limit for the free energy. To this end, consider the random measure $\nu_N = \sum_{i=1}^N \frac{X_i^2}{\sum_{i=1}^N X_i^2} \delta_{d_i}$, i.e., ν_N is a random discrete measure which assigns random weights $\frac{X_i^2}{\sum_{i=1}^N X_i^2}$ to the random positions d_i . Gamboa and Rouault [14] derived

a large deviation principle for the random measure ν_N , under certain technical assumptions on the sequence $\mu_N(D) = \frac{1}{N} \sum_{i=1}^N \delta_{d_i}$. We believe that under these assumptions a second moment argument can be done to derive the high temperature limit of the free energy $\Phi_N(\beta, O, D)$. However, this requires the full large deviation principle for the sequence $\{\mu_N(D)\}_{N \geq 1}$. On the other hand, we only need control on the tails of $\mu_N(D)$ in terms of the 2-Wasserstein distance, which is generally much easier to verify.

1.3. Organization. The rest of the paper is organized as follows: The proof of Corollary 1.3 and the application of Theorem 1.2 to various other examples are given in Section 2. The proofs of Proposition 1.1 and Theorem 1.2 are given in Section 3.1 and Section 3.2, respectively.

2. EXAMPLES

In this section, we apply Theorem 1.2 to evaluate the limit of the free energy in various orthogonal mean-field spin glass models.

2.1. The SK Model. Recall the definition of the SK-model introduced in Section 1. In this case, the coupling matrix $J = W/\sqrt{N}$, where W is a GOE matrix of order N . Thus $J = ODO^T$, where O is Haar distributed and independent of D . It is a classical result in random matrix theory that $\mu_N(D) = \frac{1}{N} \sum_{i=1}^N \mu_N(D)$ converges almost surely to the Wigner semicircle law [2]

$$\rho(x) = \frac{\sqrt{4-x^2}}{2\pi} \cdot \mathbf{1}\{x \in [-2, 2]\}. \quad (2.1)$$

Further, the edge of the empirical distribution converges to the edge of the semicircle law.

An application of Theorem 1.2 yields the following corollary about the high temperature limit of the free-energy. It is well known that the SK model has a phase transition at $\beta = 1/2$. Our approach covers the whole high temperature region of the SK model, thus re-deriving the classical result of Aizenman et. al. [1].

Corollary 2.1. *For the SK model with $\beta < 1/2$, $\lim_{N \rightarrow \infty} \frac{1}{N} \log Z_N(\beta, O, D) \xrightarrow{P} \beta^2$.*

Proof. In this case $\lambda_{\min} = -2$, $\lambda_{\max} = 2$. Using the density of the semi-circle law (2.1), the Hilbert transform can be easily computed to be $H_\rho(z) = \frac{1}{2}(z - \sqrt{z^2 - 4})$ for $z \in \mathbb{R} \setminus [-2, 2]$. This implies $H_{\max} = 1$, $H_{\min} = -1$, and $x_{\max} = 1$, $x_{\min} = -1$. Thus, using Definition 1.7, $R_\rho(z) = z$ on $(-1, 1) \setminus \{0\}$, $I_\rho(z) = z^2$, and condition (c) in Theorem 1.2 holds trivially. This gives the desired conclusion subject to the verification of the other conditions of Theorem 1.2.

It is well known that the measure $\mu_N(\mathbb{E}(D)) := \frac{1}{N} \sum_{i=1}^N \delta_{\mathbb{E}(d_i)}$ satisfies Hypothesis 2 [2]. Further, by [10, Corollary 4] there exists $C > 0$ such that

$$\mathbb{E}\{W_2(\mu_N(D), \mu_N(\mathbb{E}(D)))\} \leq C \frac{\sqrt{\log N}}{N}, \quad (2.2)$$

Hypothesis 1 then follows using Markov's inequality.

To see that the second moment method employed in our proof works up to $\beta < 1/2$, see Remark 3.1. \square

2.2. The Random Orthogonal Model. In the random orthogonal model (ROM) introduced in Section 1 the coupling matrix $J = ODO^T$, where $D = \text{diag}(d_1, \dots, d_N)$ is a deterministic sequence of $\{\pm 1\}$ such that the empirical measure $\mu_N(D)$ converges weakly to $\mu_p := p\delta_1 + (1-p)\delta_{-1}$.

Proposition 2.2. *For the random orthogonal model (ROM), there exists a $\beta_m > 0$ such that for $\beta < \beta_m$,*

$$\frac{1}{N} \log Z_N(\beta, O, D) \xrightarrow{P} \frac{1}{2} \int_0^{2\beta} \frac{\sqrt{1 + 4z(m+z)} - 1}{2z} dz. \quad (2.3)$$

where $m = 2p - 1$.

Proof. In this case, the diagonal matrix D is deterministic. Thus, Hypothesis 1 holds trivially. Also, since the limiting measure μ_p is supported on two points, Hypotheses 2 is satisfied.

In this case, $\lambda_{\max} = 1$ and $\lambda_{\min} = -1$. Moreover, by direct calculations $H_{\mu_p}(z) = \frac{z+m}{z^2-1}$ on $(-\infty, -1) \cup (1, \infty)$. Thus, $H_{\max} = \infty$ and $H_{\min} = -\infty$ which implies that

$$R_{\mu_p}(z) = \frac{1}{2z} \left(\sqrt{1 + 4z(m+z)} - 1 \right)$$

is bijective from $\mathbb{R} \setminus \{0\}$ to $(-1, 1) \setminus \{m\}$.

To verify condition (c) in Theorem 1.2, recall the definition of U_L and U_R from (1.9). For $\beta_0 \in [U_L, U_R]$, there exists $\varepsilon(\beta_0)$ such that $|\varepsilon(\beta_0)| \leq |\tanh \beta(x_{\max} - x_{\min})|$ and $\beta_0 = 2\beta(1 + \varepsilon(\beta_0))$. Thus,

$$R'_{\mu_p}(\beta_0) = \frac{16\beta^2(1 + \varepsilon(\beta_0))^2 - 1 + \sqrt{1 + 4\beta_0(m + \beta_0)}}{16\beta^2(1 + \varepsilon(\beta_0))^2 \sqrt{1 + 4\beta_0(m + \beta_0)}}. \quad (2.4)$$

From the above expression, it is easy to check that $\beta^2 \sup_{\beta_0 \in [U_L, U_R]} R'_{\mu_p}(\beta_0) \rightarrow 0$, as $\beta \rightarrow 0$. This verifies condition (c) and the result follows. \square

The integral in (2.3) has a closed form expression, which can be easily computed. We refrain from writing this explicitly for notational clarity. However, for $p = 1/2$, in which case $m = 0$, (2.3) simplifies to the expression in Corollary 1.3.

Remark 2.1. Marinari et al. [20] predicted that replica symmetry is broken in ROM with $p = 1/2$ for $\beta \geq 3.84$. The exact location of symmetry breaking is, however, unclear. Corollary 1.3 shows that there exists a β_0 up to which the limit of free energy is given by the annealed limit. The value of β_0 can be calculated as follows: Let $F(x, y) = \beta(x + y) + \log \cosh \beta(x + y)$, and

$$(x^*(\beta), y^*(\beta)) := \arg \sup_{x, y \in \mathbb{R}} (F(x, y) - T_\mu(x) - T_\mu(y)), \quad (2.5)$$

where $T_\mu(z) := -\frac{1}{4} \log(1 - z^2)$, for $z \in [-1, 1]$. It follows from the proof of Theorem 1.2 (see (3.29)) that β_0 is largest $\beta \geq 0$ such that the $x^*(\beta) = y^*(\beta)$. Numerically solving the optimization problem (2.5) approximately gives $\beta_0 \leq 2.7$, proving that replica symmetry is preserved for $\beta \leq 2.7$.

2.3. Gaussian Hopfield Model. In the Gaussian Hopfield model the coupling matrix $J = \frac{1}{p} X X^T$, where $X = ((X_{ij}))$ is a $N \times p$ matrix with i.i.d. $\mathcal{N}(0, 1)$. For simplicity, we assume $0 < c_1 < N/p < c_2 < 1$. In this case, spectral distribution of J converges weakly almost surely to the Marchenko-Pastur law with density

$$f(x) = \frac{\sqrt{4\lambda - (x - 1 - \lambda)^2}}{2\pi x}, \quad x \in ((1 - \sqrt{\lambda})^2, (1 + \sqrt{\lambda})^2), \quad (2.6)$$

where $p/N \rightarrow \lambda$. Thus, $\lambda_{\min} = (1 - \sqrt{\lambda})^2$ and $\lambda_{\max} = (1 + \sqrt{\lambda})^2$ in this example.

Using the above density and Theorem 1.2 the limit of the free energy can be derived for high temperatures.

Proposition 2.3. *In the Gaussian Hopfield model, for β sufficiently small,*

$$\frac{1}{N} \log Z_N(\beta) \xrightarrow{P} I_f(\beta) = \frac{\lambda}{2} \log \left(\frac{1}{1-2\beta} \right). \quad (2.7)$$

Proof. The Hilbert Transform of the Marchenko-Pastur law ([18, Example 3.3.5]) is known to be

$$H_f(x) = \frac{x+1-\lambda - \sqrt{(x-1-\lambda)^2 - 4\lambda}}{2x}. \quad (2.8)$$

Thus, in this example, $H_{\max} = 1/(1+\sqrt{\lambda})$ and $H_{\min} = 1/(1-\sqrt{\lambda})$, which implies that $x_{\max} = \lambda + \sqrt{\lambda}$ and $x_{\min} = \lambda - \sqrt{\lambda}$. The R -transform (1.7) is $R_f(z) = \lambda/(1-z)$. Hence, I_f (1.8) can be computed easily, which gives the formula in (2.7). Finally, to check condition (c) in Theorem 1 note that

$$R'(z) = \frac{\lambda}{(1-z)^2}.$$

The above representation implies that $\zeta(\beta) \rightarrow 0$ as $\beta \rightarrow 0$, thus verifying the required condition.

The result now follows if the spectrum of the coupling matrix J satisfies Hypothesis 1. To this end, note that simple modifications of the arguments in [11, Corollary 2] yield the following: there exists a constant $c > 0$ such that

$$\mathbb{E}(d_2(\mu_N(D), \mathbb{E}(\mu_N(D)))) \leq \frac{(\log N)^{c \log \log N}}{N}. \quad (2.9)$$

Hypothesis 1 follows by an application of Markov's inequality. \square

3. PROOFS

3.1. Proof of Proposition 1.1. In this section the proof of Proposition 1.1 is presented. Fix $\delta > 0$ and recall that $\Phi_N(\beta, O, D) = \frac{1}{N} \log Z_N(\beta, O, D)$. Therefore, by triangle inequality,

$$\mathbb{P}(|\Phi_N(\beta) - \Gamma_N(\beta, \mathbb{E}(D))| > \delta) \leq T_1 + T_2, \quad (3.1)$$

where

$$T_1 = \mathbb{P} \left(\left| \frac{1}{N} \log Z_N(\beta, O, D) - \frac{1}{N} \mathbb{E}_0 \log Z_N(\beta, O, D) \right| > \frac{\delta}{2} \right), \quad (3.2)$$

and

$$T_2 = \mathbb{P} \left(\left| \frac{1}{N} \mathbb{E}_0 (\log Z_N(\beta, O, D)) - \frac{1}{N} \mathbb{E}_0 \log Z_N(\beta, O, \mathbb{E}(D)) \right| > \frac{\delta}{2} \right). \quad (3.3)$$

We first control T_2 . By the rotational invariance of $O(N)$, $\frac{1}{N} \mathbb{E}_0 \log Z_N(\beta, O, D)$ is actually a function of only the empirical distribution $\mu_N(D) := \frac{1}{N} \sum_{i=1}^N \delta_{d_i}$, where $D = \text{diag}(d_1, \dots, d_N)$. Thus, without loss of generality assume $d_1 \geq d_2 \geq \dots \geq d_N$. Let $O = [o_1 : o_2 : \dots : o_N]$ be the columns of the matrix O . By the Cauchy-Schwarz inequality,

$$|\underline{\sigma}^T O D O^T \underline{\sigma} - \underline{\sigma}^T O \mathbb{E}(D) O^T \underline{\sigma}| = \left| \sum_{i=1}^N (d_i - \mathbb{E}(d_i)) (\underline{\sigma}^T o_i)^2 \right| \leq N \sqrt{\sum_i (d_i - \mathbb{E}(d_i))^2}, \quad (3.4)$$

since $(\underline{\sigma}^T o_i)^2 = N$, for all $i \in [N]$. This implies that

$$\begin{aligned} \left| \frac{1}{N} \mathbb{E}_0 \left(\log \frac{Z_N(\beta, O, D)}{Z_N(\beta, O, \mathbb{E}(D))} \right) \right| &\leq \beta \sqrt{\sum_i (d_i - \mathbb{E}(d_i))^2} \\ &= \beta \sqrt{N} W_2(\mu_N(D), \mu_N(\mathbb{E}(D))), \end{aligned} \quad (3.5)$$

where the last step uses $W_2^2(\mu_N(D), \mu_N(\mathbb{E}(D))) = \frac{1}{N} \sum_{i=1}^N (d_i - \mathbb{E}(d_i))^2$. Therefore,

$$T_2 \leq \mathbb{P} \left(W_2(\mu_N(D), \mathbb{E}(\mu_N(D))) > \frac{\delta}{2\beta\sqrt{N}} \right) \rightarrow 0 \quad (3.6)$$

by Hypothesis 1, as $N \rightarrow \infty$.

It remains to control the first term T_1 . For $O \in O(N)$ and any fixed diagonal matrix $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_N)$ define,

$$F_\Lambda(O) = \frac{1}{N} \log \sum_{\underline{\sigma} \in \mathcal{S}_N} \exp(\beta \underline{\sigma}^T O \Lambda O^T \underline{\sigma}). \quad (3.7)$$

Let $\|\Lambda\|_\infty = \max_{1 \leq i \leq N} |\lambda_i|$. Moreover, for any $N \times N$ symmetric matrix A , denote the spectral norm by $\|A\|_2 = \sup_{\underline{x} \in \mathbb{R}^N} \frac{\|A\underline{x}\|_2}{\|\underline{x}\|_2}$ and the Frobenius norm by $\|A\|_F = (\text{tr}(A^2))^{\frac{1}{2}}$. It is easy to see that for $O_1, O_2 \in O(N)$ and a unit vector \underline{x} (that is $\|\underline{x}\|_2 = 1$),

$$\begin{aligned} |\underline{x}^T O_1 \Lambda O_1^T \underline{x} - \underline{x}^T O_2 \Lambda O_2^T \underline{x}| &\leq |\underline{x}^T O_1 \Lambda (O_1 - O_2)^T \underline{x}| + |\underline{x}^T (O_1 - O_2) \Lambda O_2^T \underline{x}| \\ &\leq 2\|\Lambda\|_\infty \|O_1 - O_2\|_2 \\ &\leq 2\|\Lambda\|_\infty \|O_1 - O_2\|_F. \end{aligned} \quad (3.8)$$

Thus, using (3.8),

$$|F(O_1) - F(O_2)| = \frac{1}{N} \left| \log \frac{Z_N(\beta, O_1, \Lambda)}{Z_N(\beta, O_2, \Lambda)} \right| \leq 2\|\Lambda\|_\infty \beta \|O_1 - O_2\|_F.$$

This implies F is Lipschitz with respect to the Frobenius norm.

Sub-gaussian tail inequalities are known for Lipschitz functions on $SO(N)$ (see Gromov and Milman [15]). This can be used to complete the proof as follows: Now, let T be the operator which takes $O \in SO(N)$ and changes the sign of the first column of O . Clearly, for $O \in SO(N)$, $F(O) = F(TO)$. Let \mathbb{P}_1 and \mathbb{E}_1 be Haar measure and the expectation with respect to it on $SO(N)$, respectively. Thus, $\mathbb{E}_0(F_D(O)) = \mathbb{E}_1(F_D(O))$, and recalling (3.2) and (3.7) it follows that

$$\begin{aligned} T_1 &\leq \mathbb{E} \mathbb{P}_1 \left(|F_D(O) - \mathbb{E}_1(F_D(O))| > \frac{\delta}{2}, \|D\|_\infty \leq M_N \right) + \mathbb{P}(\|D\|_\infty > M_N) \\ &\leq \exp \left(-\frac{CN\delta^2}{\beta^2 M_N^2} \right) + \mathbb{P}(\|D\|_\infty > M_N), \end{aligned} \quad (3.9)$$

where $C > 0$ is a universal constant. By Hypothesis 1, the RHS above goes to zero as $N \rightarrow \infty$.

Combining (3.6) and (3.9) with (3.1) the result follows.

3.2. Proof of Theorem 1.2. By concentration arguments identical to those used in controlling the term T_1 in Proposition 1.1, the following lemma can be proved.

Lemma 3.1. For any $\beta > 0$, there exists a universal constant c , independent of N , such that

$$\mathbb{P}(|\Phi_N(\beta, O, \mathbb{E}(D)) - \mathbb{E}_0 \Phi_N(\beta, O, \mathbb{E}(D))| > \delta) \leq \exp(-cN\delta^2/\beta^2). \quad (3.10)$$

The proof of Theorem 1.2 also requires computing the first and second annealed moments of $Z_N(\beta, O, \mathbb{E}(D))$.

Proposition 3.1. Under the assumptions of Theorem 1.2, for β sufficiently small (possibly depending on the limiting measure μ),

$$\lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}_0(Z_N(\beta, O, \mathbb{E}(D))) = \lim_{N \rightarrow \infty} \frac{1}{2N} \log \mathbb{E}_0(Z_N(\beta, O, \mathbb{E}(D))^2). \quad (3.11)$$

The above lemma is the most challenging part of our argument and the proof is deferred to Section 3.2.1.

The proof of Theorem 1.2 can be completed easily by combining Lemma 3.1 and Proposition 3.1 with Theorem 1.4. To this end, set $\gamma_0 = \frac{4\mathbb{E}_0 Z_N(\beta, O, \mathbb{E}(D))^2}{(\mathbb{E}_0 Z_N(\beta, O, \mathbb{E}(D)))^2}$. Recall the definition of the *annealed free energy*

$$\phi_N(\beta, \Lambda) = \frac{1}{N} \log \mathbb{E}_0 Z_N(\beta, O, \Lambda).$$

Then by [21, Lemma 4.1.1]

$$\mathbb{P} \left(|\Phi_N(\beta, O, \mathbb{E}(D)) - \phi_N(\beta, \mathbb{E}(D))| < \frac{1}{N} \log \gamma_0 \right) \geq \frac{1}{\gamma_0}. \quad (3.12)$$

Also, note that $\Gamma_N(\beta, \mathbb{E}(D)) = \mathbb{E}_0 \Phi_N(\beta, O, \mathbb{E}(D))$. Thus, inequality (3.12) combined with Lemma 3.1 gives

$$\lim_{N \rightarrow \infty} \Gamma_N(\beta, \mathbb{E}(D)) = \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}_0(Z_N(\beta, O, \mathbb{E}(D))) = I_\mu(\beta), \quad (3.13)$$

where the last step uses Theorem 1.4. Finally, using Proposition 3.1, Theorem 1.2 follows.

3.2.1. Proof of Proposition 3.1. For any function $f : S_N \times S_N \mapsto \mathbb{R}$, denote by $\mathbb{E}_1 f(\underline{\sigma}, \underline{\tau}) = \frac{1}{2^{2N}} \sum_{\underline{\sigma}, \underline{\tau} \in S_N} f(\underline{\sigma}, \underline{\tau})$, the expectation over the uniform measure over $S_N \times S_N$. Let $\Lambda = \mathbb{E}(D) = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N)$. Therefore,

$$\begin{aligned} \mathbb{E}_0(Z_N(\beta, O, \mathbb{E}(D)))^2 &= \mathbb{E}_0(Z_N(\beta, O, \Lambda))^2 \\ &= \mathbb{E}_1 \mathbb{E}_0 \exp \left(N\beta \text{tr} \left\{ O\Lambda O^T \left(\frac{\underline{\sigma}\underline{\sigma}^T}{N} + \frac{\underline{\tau}\underline{\tau}^T}{N} \right) \right\} \right) \\ &= \mathbb{E}_1 \mathbb{E}_0 \exp \left(N\beta \left\{ \left(1 + \frac{\underline{\sigma}^T \underline{\tau}}{N} \right) (O\Lambda O^T)_{11} + \left(1 - \frac{\underline{\sigma}^T \underline{\tau}}{N} \right) (O\Lambda O^T)_{22} \right\} \right), \end{aligned}$$

where we use the observation that the non-zero eigenvalues of $(\underline{\sigma}\underline{\sigma}^T + \underline{\tau}\underline{\tau}^T)/N$ are $(1 + \underline{\sigma}^T \underline{\tau}/N)$ and $(1 - \underline{\sigma}^T \underline{\tau}/N)$ respectively. Let $V_1 = (O\Lambda O^T)_{11}$ and $V_2 = (O\Lambda O^T)_{22}$. By interchanging the order of the expectation and observing that $\mathbb{E}_1 e^{\lambda \underline{\sigma}^T \underline{\tau}} = (\cosh \lambda)^N$, for any $\lambda \in \mathbb{R}$, it follows that

$$\mathbb{E}_0(Z_N(\beta, O, \Lambda))^2 = \mathbb{E}_0(\exp(NF(V_1, V_2))) \quad (3.14)$$

where $F(x, y) = \beta(x + y) + \log \cosh \beta(x - y)$.

The non-negativity of the log cosh function trivially implies that $F(x, y) \geq \beta(x + y)$. Then by [16, Theorem 1.7], for β sufficiently small, we have

$$\begin{aligned} \liminf_{N \rightarrow \infty} \frac{1}{2N} \log \mathbb{E}_0(Z_N(\beta, O, \Lambda))^2 &\geq \lim_{N \rightarrow \infty} \frac{1}{2N} \log \mathbb{E}_0 \exp(N\beta(V_1 + V_2)) \\ &= \lim_{N \rightarrow \infty} \frac{1}{2N} \log \mathbb{E}_0 \exp(N\beta(V_1 + V_2)) = I_\mu(\beta), \end{aligned} \quad (3.15)$$

where μ is the limit of the empirical measure $\mu_N(\mathbb{E}(D)) := \frac{1}{N} \sum_{i=1}^N \delta_{\mathbb{E}(d_i)}$.

For the upper bound, let $\mathbf{X} = (X_1, X_2, \dots, X_N)'$ and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_N)'$ be i.i.d. $\mathcal{N}(0, I)$. If

$$\mathbf{Z} = \mathbf{Y} - \frac{\langle \mathbf{X}, \mathbf{Y} \rangle}{\langle \mathbf{X}, \mathbf{X} \rangle} \mathbf{X} = (Z_1, Z_2, \dots, Z_N)',$$

then

$$(V_1, V_2) \stackrel{\mathcal{D}}{=} \left(\frac{\sum_{i=1}^N \lambda_i X_i^2}{\sum_{i=1}^N X_i^2}, \frac{\sum_{i=1}^N \lambda_i Z_i^2}{\sum_{i=1}^N Z_i^2} \right). \quad (3.16)$$

Let $V_2' = \frac{\sum_{i=1}^N \lambda_i Y_i^2}{\sum_{i=1}^N Y_i^2}$. Note that V_1 and V_2' are independent, but V_1 and V_2 are not. The following lemma shows that we can replace V_2 by V_2' to get an upper bound:

Lemma 3.2. Under the assumptions of Theorem 1.2, for any $\beta > 0$,

$$\limsup_{N \rightarrow \infty} \frac{1}{2N} \log \mathbb{E}_0(Z_N(\beta, O, \Lambda)^2) \leq \lim_{N \rightarrow \infty} \frac{1}{2N} \log \mathbb{E}_0 \exp(NF(V_1, V_2')), \quad (3.17)$$

where $F(x, y) = \beta(x + y) + \log \cosh \beta(x - y)$.

Proof. The lemma will be established using a “localization” argument similar to the one used in [16]. Fix $\kappa < 1/2$ and

$$B_N(\kappa) = \left\{ \left| \frac{1}{N} \sum_{i=1}^N X_i^2 - 1 \right| \leq N^{-\kappa}, \left| \frac{1}{N} \sum_{i=1}^N Y_i^2 - 1 \right| \leq N^{-\kappa}, \left| \frac{1}{N} \sum_{i=1}^N X_i Y_i \right| \leq N^{-\kappa} \right\}. \quad (3.18)$$

We adopt the following system of coordinates in \mathbb{R}^{2N} : $r, \alpha_1^{(1)}, \dots, \alpha_{N-1}^{(1)}$ are the polar coordinates of \mathbf{X} , $r_2 = \|\mathbf{Y}\|$, β_2 is the angle between \mathbf{X} and \mathbf{Y} , and $\alpha_1^{(2)}, \dots, \alpha_{N-2}^{(2)}$ are the angles needed to spot \mathbf{Y} on a cone of angle β_2 around \mathbf{X} . It is easy to see that (V_1, V_2) is a function of the α 's while the event $B_N(\kappa)$ is determined by r and the β 's. So (V_1, V_2) and $B_N(\kappa)$ are independent.

Let $I_N = \mathbb{E}_0 \exp(NF(V_1, V_2))$. By (3.14), $\frac{1}{2N} \log \mathbb{E}_0(Z_N(\beta, O, \Lambda)^2) = \frac{1}{2N} \log I_N$. Therefore, to prove (3.17) it suffices to show that

$$I_N \leq \varepsilon(N, \kappa) \mathbb{E}_0(\mathbf{1}_{B_N(\kappa)} \exp(NF(V_1, V_2'))) \quad (3.19)$$

where $\varepsilon(N, \kappa) \leq C(\kappa) \exp(N^{1-2\kappa})$ for some constant $C(\kappa)$ and N sufficiently large.

By bounding the moment generating functions of X_1^2 and $X_1 Y_1$ suitably in a neighborhood of zero, we get

$$\begin{aligned} & \mathbb{P}(B_N(\kappa)^c) \\ & \leq \mathbb{P} \left(\left| \frac{1}{N} \sum_{i=1}^N X_i^2 - 1 \right| \geq N^{-\kappa} \right) + \mathbb{P} \left(\left| \frac{1}{N} \sum_{i=1}^N Y_i^2 - 1 \right| \geq N^{-\kappa} \right) + \mathbb{P} \left(\left| \frac{1}{N} \sum_{i=1}^N X_i Y_i \right| > N^{-\kappa} \right) \\ & \leq C'(\kappa) \exp(-cN^{1-2\kappa}), \end{aligned} \quad (3.20)$$

for some constants $C'(\kappa)$, $c > 0$ and N sufficiently large. Now, using the independence of (V_1, V_2) and $B_N(\kappa)$,

$$I_N \leq \frac{1}{\mathbb{P}(B_N(\kappa))} \mathbb{E}_0(\mathbf{1}_{B_N(\kappa)} \exp(NF(V_1, V_2))) \leq \varepsilon(N, \kappa) \mathbb{E}_0(\mathbf{1}_{B_N(\kappa)} \exp(NF(V_1, V_2))). \quad (3.21)$$

By the Lipschitz property of the log cosh function $|F(x, y) - F(x, z)| \leq 2\beta|y - z|$. Therefore,

$$I_N \leq \varepsilon(N, \kappa) \mathbb{E}_0(\mathbf{1}_{B_N(\kappa)} \exp(NF(V_1, V_2') + 2N\beta|V_2 - V_2'|)). \quad (3.22)$$

The upper bound in (3.19) follows if, on the set $B_N(\kappa)$, $|V_2 - V_2'| \lesssim N^{-\kappa}$. To this end, note that on $B_N(\kappa)$, $\frac{1}{N} \|\mathbf{Y} - \mathbf{Z}\|^2 \lesssim N^{-\kappa}$. Further, on $B_N(\kappa)$,

$$\frac{1}{N} |\mathbf{Z}^T \Lambda \mathbf{Z} - \mathbf{Y}^T \Lambda \mathbf{Y}| \leq \frac{2}{N} \|\Lambda\|_\infty \|\mathbf{Z}\| \|\mathbf{Z} - \mathbf{Y}\| \lesssim N^{-\kappa}, \quad (3.23)$$

since $\|\Lambda\|_\infty$ is finite by Hypothesis 2(b).

From this it is easy to see that on the set $B_N(\kappa)$, $|V_2 - V_2'| \lesssim N^{-\kappa}$, and the proof is complete. \square

Lemma 3.3. Under the assumptions of Theorem 1.2, for $\beta \geq 0$ sufficiently small,

$$\lim_{N \rightarrow \infty} \frac{1}{2N} \log \mathbb{E}_0 \exp(NF(V_1, V_2')) = I_\mu(\beta). \quad (3.24)$$

The proof of the above lemma is given below in Section 3.2.1. Note that the Lemma 3.3 together with (3.15) and (3.17) gives

$$\lim_{N \rightarrow \infty} \frac{1}{2N} \log \mathbb{E}_0 (Z_N(\beta, O, \mathbb{E}(D)))^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \log \mathbb{E}_0 (Z_N(\beta, O, \mathbb{E}(D))) = I_\mu(\beta), \quad (3.25)$$

where the last equality uses Theorem 1.4. This completes the proof of Proposition 3.1.

Proof of Lemma 3.3: The proof of this lemma follows from a large deviation result established in [16]. Recall the Hilbert transform and the R -transform of a probability measure ν defined in (1.5), and (1.7), respectively. Denote the inverse of H_ν by K_ν , and that of R_ν by Q_ν . Refer to [16] for further details about the Hilbert and the R -transforms.

Also, recall (1.6)

$$x_{\max} = \lambda_{\max} - \frac{1}{H_{\max}} \quad \text{and} \quad x_{\min} = \lambda_{\min} - \frac{1}{H_{\min}}, \quad (3.26)$$

where $H_{\max} = \lim_{z \downarrow \lambda_{\max}} H_\nu(z)$ and $H_{\min} = \lim_{z \uparrow \lambda_{\min}} H_\nu(z)$. Finally, for $\kappa \in (\lambda_{\min}, \lambda_{\max})^c$, define

$$h_x(\kappa) = \int \log \frac{\kappa - \lambda}{\kappa - x} d\nu(\lambda), \quad (3.27)$$

and $h_x^{\min} = \lim_{\kappa \uparrow \lambda_{\min}} h_x(\kappa)$ and $h_x^{\max} = \lim_{\kappa \downarrow \lambda_{\max}} h_x(\kappa)$.

The following proposition, proved in [16], gives the large deviations rate function for the random variable V_1 .

Proposition 3.4. ([16, Proposition 5.1]) If the sequence of non-random empirical measures $\mu_N(\Lambda) = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i} \rightarrow \mu$ and satisfies Hypothesis 2, then the law of the random variables V_1 defined in (3.16) satisfies a large deviation principle with scale N and good rate function

$$T_\mu(x) = \begin{cases} \frac{1}{2} h_x(K_\mu(Q_\mu(x))) & \text{if } x \in [x_{\min}, x_{\max}], \\ \frac{1}{2} h_x^{\max} & \text{if } x \in]x_{\max}, \lambda_{\max}[, \\ \frac{1}{2} h_x^{\min} & \text{if } x \in]\lambda_{\min}, x_{\min}[, \\ \infty & \text{otherwise.} \end{cases} \quad (3.28)$$

Since the empirical measures $\mu_N(\Lambda)$ satisfies Hypothesis 2, by Varadhan's lemma [13] we get

$$\lim_{N \rightarrow \infty} \frac{1}{2N} \log \mathbb{E}_0 \exp(NF(V_1, V_2')) = \frac{1}{2} \sup_{x, y \in \mathbb{R}} (F(x, y) - T_\mu(x) - T_\mu(y)), \quad (3.29)$$

where $T_\mu(\cdot)$ is the good rate function of V_1 and $F(x, y) = \beta(x + y) + \log \cosh \beta(x - y)$.

Set $\psi(x, y) = F(x, y) - T_\mu(x) - T_\mu(y)$. To prove Lemma 3.3 it suffices to establish that for β sufficiently small,

$$\sup_{x, y \in \mathbb{R}} \psi(x, y) = \sup_{(x, y) \in [x_{\min}, x_{\max}]^2} \psi(x, y) = 2I_\mu(\beta), \quad (3.30)$$

where x_{\max} and x_{\min} are as defined in (3.26).

To this end note that,

$$\frac{\partial \psi}{\partial x} = \beta + \beta \tanh \beta(x - y) - T'_\mu(x), \quad (3.31)$$

$$\frac{\partial \psi}{\partial y} = \beta - \beta \tanh \beta(x - y) - T'_\mu(y). \quad (3.32)$$

We begin by showing that the maxima of ψ on the set $[\lambda_{\min}, \lambda_{\max}]^2 \setminus [x_{\min}, x_{\max}]^2$ is attained on the boundary of the set $[x_{\min}, x_{\max}]^2$, for $\beta < H_{\max}/4$: For instance, for $(x, y) \in (x_{\max}, \lambda_{\max}] \times [x_{\min}, x_{\max}]$ (as in Figure 1), by (3.28) $T'_\mu(x) = \frac{1}{2} \frac{d}{dx} h_x^{\max} = \frac{1}{2(\lambda_{\max} - x)}$. Then for $\beta < H_{\max}/4$, using $\tanh \beta(x - y) \leq 1$ and $x \geq x_{\max}$,

$$\frac{\partial \psi}{\partial x}(x, y) = \beta(1 + \tanh \beta(x - y)) - \frac{1}{2(\lambda_{\max} - x)} < 0. \quad (3.33)$$

Therefore, $\psi(x, y) \leq \psi(x_{\max}, y)$ for $(x, y) \in (x_{\max}, \lambda_{\max}] \times [x_{\min}, x_{\max}]$. Other points in the region $[\lambda_{\min}, \lambda_{\max}]^2 \setminus [x_{\min}, x_{\max}]^2$ can be dealt with similarly.

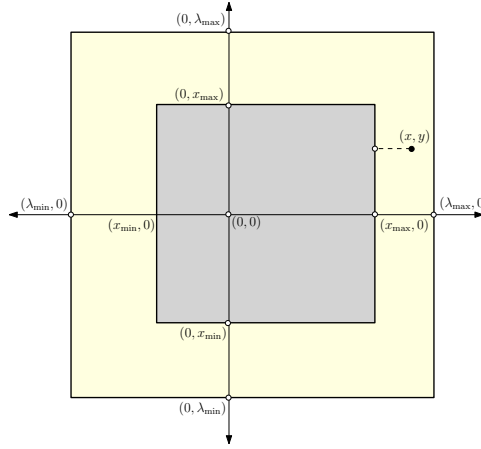


FIGURE 1. The domain of the function ψ .

Next, we establish that for β small enough, the maxima of ψ on $[x_{\min}, x_{\max}]^2$ cannot be attained on the boundary of $[x_{\min}, x_{\max}]^2$: For $y \in (x_{\min}, x_{\max})$,

$$\frac{\partial}{\partial x} \psi(x_{\max}^-, y) = \beta(1 + \tanh \beta(x_{\max} - y)) - \frac{1}{2} H_{\max} < 0.$$

if $\beta < 4H_{\max}$ as in (3.33). This ensures that the maxima cannot be attained for $x = x_{\max}$. The same analysis implies that the maxima is not attained for $y = x_{\max}$. This establishes the required assertion.

It remains to analyze the function ψ on $(x_{\min}, x_{\max})^2$. Note that for $x \in (x_{\min}, x_{\max})$, $T'_\mu(x) = \frac{1}{2} Q_\mu(x)$. Therefore, from (3.31)-(3.32) any stationary point of ψ in $(x_{\min}, x_{\max})^2$ is a solution of the system of equations

$$x = R_\mu(2\beta(1 + \tanh \beta(x - y))), \quad (3.34)$$

$$y = R_\mu(2\beta(1 - \tanh \beta(x - y))), \quad (3.35)$$

since R_μ is the inverse of Q_μ .

Let $a(\beta)$ be the solution of $\beta = T'_\mu(\cdot) = \frac{1}{2} Q_\mu(\cdot)$. It is easy to verify that $x^*(\beta) = y^*(\beta) = a(\beta)$ is a critical point of ψ in $(x_{\min}, x_{\max})^2$. It remains to show that for β sufficiently small the maximum

in (3.29) is attained at $x^*(\beta) = y^*(\beta) = a(\beta)$. This implies Lemma 3.3, since by [16, Lemma 5.7], $\{\beta a(\beta) - T_\mu(a(\beta))\} = I_\mu(\beta)$.

To show that the maximum in (3.29) is attained at $x^*(\beta) = y^*(\beta) = a(\beta)$ we will show (in Lemma 3.5 below) that for β sufficiently small, the system of equations (3.34)-(3.35) has a unique solution $(a(\beta), a(\beta))$ which is a local maxima. This establishes that $(a(\beta), a(\beta))$ maximizes ψ in $(x_{\min}, x_{\max})^2$ and (3.30) follows.

Lemma 3.5. For β sufficiently small and R_μ satisfying condition (c) in Theorem 1.2, the system of equations (3.34)-(3.35) has a unique solution which is a local maximum.

Proof. For $(x, y) \in [x_{\min}, x_{\max}]^2$, define

$$G(x, y) = \begin{pmatrix} R_\mu(2\beta(1 + \tanh \beta(x - y))) \\ R_\mu(2\beta(1 - \tanh \beta(x - y))) \end{pmatrix}. \quad (3.36)$$

For $\beta < H_{\max}/4$, G maps $[x_{\min}, x_{\max}]^2$ to $[x_{\min}, x_{\max}]^2$. Consider the set $M = [x_{\min}, x_{\max}]^2$ equipped with the L_1 metric, that is, $d_1((x_1, y_1), (x_2, y_2)) = |x_1 - x_2| + |y_1 - y_2|$. We note that (M, d_1) is a complete metric space. Now, by the mean value theorem and using $\text{sech}^2 \leq 1$,

$$|R_\mu(2\beta(1 + \tanh \beta(x_1 - y_1))) - R_\mu(2\beta(1 + \tanh \beta(x_2 - y_2)))| \leq 2\beta^2 \zeta(\beta) d_1((x_1, y_1), (x_2, y_2)).$$

where $\zeta(\beta) = \beta^2 \sup_{\beta_0 \in [U_L, U_R]} |R'(\beta_0)|$ and U_L and U_R are as in (1.9). This implies, by condition (c) of Theorem 1.2, that G is a contraction. Thus, using the Banach Fixed Point Theorem [23], G has a unique fixed point. This implies $(a(\beta), a(\beta))$ is the unique fixed point of G , and the only stationary point of ψ in $(x_{\min}, x_{\max})^2$.

Finally, we establish that the critical point $(a(\beta), a(\beta))$ is a local maxima of the function G . To see this note that the Hessian

$$\nabla^2 \psi(x, y) = -\text{diag}(T_\mu''(x), T_\mu''(y)) + \beta^2 \text{sech}^2 \beta(x - y) \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}. \quad (3.37)$$

Using $T_\mu''(x) = \frac{1}{2R_\mu(Q(x))}$, we note that $\nabla^2 \psi(a(\beta), a(\beta))$ is negative definite provided $\beta^2 R'(2\beta) \leq 1/4$. This condition is satisfied for β sufficiently small as condition (c) holds, thus completing the proof. \square

Remark 3.1. In the SK model, $R_\rho(z) = z$ (defining $R_\rho(0) = 0$), $x_{\max} = 1$, $x_{\min} = -1$, and ρ is the semi-circle law (2.1). Moreover,

$$T_\rho(x) = \begin{cases} \frac{1}{4} - \log(2 + x) & x \in [-2, -1], \\ \frac{x^2}{4} & x \in [-1, 1], \\ \frac{1}{4} - \log(2 - x) & x \in [1, 2]. \end{cases} \quad (3.38)$$

Note that $T_\rho(x) > x^2/4$ in $[-2, -1] \cup [1, 2]$. Therefore, $\psi(x, y) \leq F(x, y) - \frac{x^2}{4} - \frac{y^2}{4} := \tilde{\psi}(x, y)$ for $(x, y) \in [-2, 2]^2$. It is easy to see that $(2\beta, 2\beta) \in [-1, 1]^2$, for $\beta < 1/2$, is a stationary point of $\tilde{\psi}$. The Hessian $\nabla^2 \tilde{\psi}(x, y)$ is negative definitive for $(x, y) \in [-2, 2]^2$ if

- (a) $(\nabla^2 \tilde{\psi}(x, y))_{11} = \frac{1}{2} + \beta^2 \text{sech}^2 \beta(x - y) < 0$, which holds whenever $\beta < \frac{1}{\sqrt{2}}$; and
- (b) $\det(\nabla^2 \tilde{\psi}(x, y)) = \frac{1}{4} + \beta^2 \text{sech}^2 \beta(x - y) < 0$, whenever $\beta < \frac{1}{2}$.

Therefore, the maximum of $\tilde{\psi}$ is attained at $(2\beta, 2\beta) \in [-1, 1]^2$, for $\beta < 1/2$, where the two functions $\tilde{\psi}$ and ψ agree. Thus, for $\beta < 1/2$, which is the entire replica symmetric phase, the limit in Corollary 2.1 holds.

Acknowledgements The authors thank Amir Dembo, Andrea Montanari and Sourav Chatterjee for helpful discussions. S.S. thanks Zhou Fan for help with results about random matrices.

REFERENCES

- [1] M. Aizenman, J.L. Lebowitz and D. Ruelle, Some rigorous results on the Sherrington- Kirkpatrick spin glass model , *Comm. Math. Phys.*, Vol. 112 (1) , 3 – 20, 1987.
- [2] G.W. Anderson, A. Guionnet and O. Zeitouni, *An introduction to random matrices*, Vol. 118, Cambridge University Press, 2010.
- [3] J. Bernasconi, Low autocorrelation binary sequences: statistical mechanics and configuration space analysis, *J. Physique*, Vol. 48, 559, 1987.
- [4] A. Bovier, ACD van Enter and B. Niederhauser. Stochastic symmetry-breaking in a Gaussian Hopfield model, *Journal of Statistical Physics* Vol. 95 (1-2), 181–213, 1999.
- [5] P. Carmona and Y. Hu, Universality in Sherrington-Kirkpatrick’s spin glass model, *Annales de l’Institut Henri Poincare (B)*, Vol. 42 (2), 215–222, 2006.
- [6] H. Chandra, Differential operators on a semisimple Lie algebra, *Amer. J. Math.*, Vol. 79, 87–120. 1957.
- [7] S. Chatterjee, A simple invariance theorem, [arXiv:math/0508213](https://arxiv.org/abs/math/0508213), 2005.
- [8] R. Cherrier, D. S. Dean, A. Lefèvre, The role of the interaction matrix in mean-field spin glasses, *Phys. Rev. E*, Vol. 67, 046112, 2003.
- [9] B. Collins, and P. Śniady, New scaling of Itzykson-Zuber integrals, *Annales de l’Institut Henri Poincare (B)*, Vol. 43 (2), 139–146. 2007.
- [10] S. Dallaporta, Eigenvalue variance bounds for Wigner and covariance random matrices, *Random Matrices: Theory and Applications*, Vol. 1 (3), 1250007, 2012.
- [11] S. Dallaporta, Eigenvalue variance bounds for covariance matrices, [arXiv:1309.6265](https://arxiv.org/abs/1309.6265), 2013.
- [12] M. Degli Esposti, C. Giardiná, and S. Graffi, Energy landscape statistics of the random orthogonal model, *J. Phys. A: Math. Gen.*, Vol. 36, 2983–2994, 2003.
- [13] A. Dembo and O. Zeitouni, *Large deviations techniques and applications*, Second Ed., Vol. 38, Applications of Mathematics. Springer-Verlag, New York, 1998.
- [14] F. Gamboa and A. Rouault, Canonical moments and random spectral measures, *Journal of Theoretical Probability*, Vol. 23 (4), 1015–1038, 2010.
- [15] M. Gromov and V. D. Milman, A topological application of the isoperimetric inequality, *Amer. J. Math.*, Vol. 105 (4), 843–854, 1983.
- [16] A. Guionnet, M. Maida, Fourier view on the R -transform and related asymptotics of spherical integrals, *Journal of Functional Analysis*, Vol. 222, 435–490, 2005.
- [17] A. Guionnet, O. Zeitouni, Large Deviations Asymptotics for Spherical Integrals, *Journal of Functional Analysis*, Vol. 188(2), 461- 515, 2002.
- [18] F. Hiai, D. Petz, *The Semicircle Law, Free Random Variables and Entropy*, American Mathematical Society, 2000.
- [19] E. Marinari, G. Parisi and F. Ritort, Replica field theory for deterministic models: binary sequences with low autocorrelation, *J. Phys. A: Math. Gen.*, Vol. 27, 7615, 1994.
- [20] E. Marinari, G. Parisi, F. Ritort, Replica field theory for deterministic models. II. A non-random spin glass with glassy behavior, *J. Phys. A: Math. Gen.* Vol. 27, 7647, 1994.
- [21] A. Montanari, Statistical mechanics and algorithms on sparse and random graphs, St. Flour School of Probability, 2013. <http://web.stanford.edu/~montanar/OTHER/STATMECH/stflour.pdf>
- [22] D. Panchenko, *The Sherrington-Kirkpatrick model*, Springer Science & Business Media, 2013.
- [23] G. F. Simmons, *Introduction to Topology and Modern Analysis*, R.E. Krieger Pub. Co., 1983.
- [24] M. Talagrand, *Spin Glasses, A Challenge for Mathematicians*, Springer, 2003.

DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY, CALIFORNIA, USA, bhaswar@stanford.edu

DEPARTMENT OF STATISTICS, STANFORD UNIVERSITY, CALIFORNIA, ssen90@stanford.edu