Low-rank Matrix Estimation with Inhomogeneous Noise

Alice Guionnet, Justin Ko, Florent Krzakala, Lenka Zdeborová

LOW-RANK MATRIX ESTIMATION WITH INHOMOGENEOUS NOISE

ALICE GUIONNET*, JUSTIN KO*, FLORENT KRZAKALA, LENKA ZDEBOROVÁ

Abstract. We study low-rank matrix estimation for a generic inhomogeneous output channel through which the matrix is observed. This generalizes the commonly considered spiked matrix model with homogeneous noise to include for instance the dense degree-corrected stochastic block model. We adapt techniques used to study multispecies spin glasses to derive and rigorously prove an expression for the free energy of the problem in the large size limit, providing a framework to study the signal detection thresholds. We discuss an application of this framework to the degree corrected stochastic block models.

1. Introduction

Heterogeneity is a fundamental part of real world problems, as opposed to simpler homogeneous modeling assumptions. In this paper, we investigate the effect of such inhomogeneity in spike models and community detection.

We focus on a mathematical formulation, common in statistics, inference and machine learning, where the aim is to reconstruct a rank κ vector x^0 in $\mathbb{R}^{\kappa\times N}$ from noisy measurements D of the inner product of the vector. A substantial amount of recent work focuses on this issue in high dimension, as $N\to\infty$, and under separable priors on x^0. Applications of this setting range from the Wigner and Wishart spiked models, to the stochastic block model, sparse PCA, or clustering mixtures of Gaussians (see, e.g. [5, 14, 20, 32, 30, 29, 6]). Here, we study this problem in the case of inhomogeneous noise. In this situation, both the law and the strength of the noisy measurements D can be different for any pairs i, j,

$$D_{ij}\sim P_{ij}\Big(D_{ij}\,\Big|\,\frac{x_i^0\cdot x_j^0}{\sqrt{N}}\Big). \tag{1.1}$$

One of the main motivation for this problem, on which we shall focus for the concrete application of our results, is a dense version of a well known model of community detection called the degree-corrected stochastic block model (DCSBM) [26].

1.1. Highlights of our main contributions.

We now summarize our main results that are then described precisely in Section 2.

(i) We prove a **Gaussian universality theorem** — that generalizes the "homogeneous" universality proven in [28] — that shows that for a matrix factorization problem with inhomogeneous noise distributions, a large class of noise models (including for instance sign flips, or additive non-Gaussian noise, and the DCSBM), one may transform the model into an equivalent Gaussian one. This means that we need only to consider the case where the distribution P_{ij} is Gaussian. This transformation amounts into working with the Fisher score matrix. This results is of crucial importance as it allows to study an entire, complex, and diverse family of statistical model just by focusing on an equivalent spike model problem [19] with a Gaussian noise. The universality at the level of free energy is stated in Theorem 2.7 and the stronger form at the level of the spectrum is stated in Theorem 2.9.

(ii) **Proof of the free energy** – Focusing on the spike model with Gaussian noise, we then study the value of the likelihood ratio with the corresponding null model, (or equivalently the free energy in statistical physics terminology, or the mutual information in information theory). We then prove a formula for the asymptotic value of the likelihood ratio in Theorem 2.10 and Theorem 2.13, generalizing the homogeneous results [14, 18, 29]. This is achieved by using methods from mathematical physics of spin glasses.

(iii) We study the free energy and determine the **phase boundary** (see Lemma 2.15) in terms of the signal-to-noise ratio, focusing in particular to the dense DCSBM model. The phase transition marks a transition from a phase where it is information theoretically impossible to reconstruct the community better than a random guess, from a phase where it is possible to do so. This phase boundary is compared

*Supported in part by ERC Project LDRAM: ERC-2019-ADG Project 884584.

to the separation of the largest eigenvalue of the matrix after a naive transformation to transfer the noise profile to the signal. We show that the regime where such a transformation leads to a extremal eigenvalue is contained in the information theoretically to detect the signal in Proposition 2.18.

1.2. **Relation with previous work.** The universality property has been used extensively in the homogeneous cases, see e.g. [31, 28, 39]. It is in particular at the roots of the identification of the dense stochastic block model with an equivalent spike model [31, 14]. Such spike models [19] have been the subject of many studies over the last few years, in particular in random matrix theory [5], in statistical inference (e.g. without pretension of exhaustivity [40, 15, 32, 39, 8, 12]) with many different applications [16, 17, 30].

In particular, the last decade has witnessed spectacular progress in the rigorous approach to the computation of the asymptotic mutual information for such problems [28, 18, 29, 21, 6, 22]. We shall use these techniques, in particular the one of [29], in our approach.

Community detection of one of the most fundamental problem of graph theory. The connection between the low rank factorization problem and the SBM was unveiled in [31, 14]. Here we shall be instead focusing on the inhomogenous Degree-Corrected SBM [26], a much more realistic model. On a side note, the inhomogenous setting that we shall be looking at has deep connection with the spatial coupling introduced in [25, 18].

To compute the limit of the free energy, we use a modification of the synchronization mechanisms for multi-species [35] and vector spin glasses [38, 37]. By adding some extra terms to the perturbations, we can regularize our posterior probability by introducing the Ghirlanda–Guerra identities while preserving the Nishimori identities and the concentration of the overlaps. The synchronization of spin glasses was also recently adapted in other contexts [27, 10, 9, 1, 3] to compute the free energy of various models. A different point of view to study multispecies models using the TAP approach was also applied recently in [41, 42].

2. Main results

2.1. **Setting.** We consider the following inference problem. Given a probability measure P_0 on $\mathbb{R}^\kappa$ consider a *signal* consisting of N independent copies $x^0 = (x_1^0, \dots, x_N^0) \in \mathbb{R}^{\kappa \times N}$ generated from the product measure

$$x^0 \sim \prod_{1 \le i \le N} P_0(x_i^0).$$

Given the inner products of the signal $w_{ij}^0 = \frac{x_i^0 \cdot x_j^0}{\sqrt{N}}$, we generate independently some *observed data* D_{ij} conditionally on w_{ij}^0 according to the probability measure

$$P_{ij}\Big(D_{ij} \,\Big|\, \frac{x_i^0 \cdot x_j^0}{\sqrt{N}}\Big). \tag{2.1}$$

A critical distinction in these inhomogeneous models is the fact that while the data D_{ij} are generated independently, the conditional distributions change depending on the indices i, j. The problem we are interested in is the estimation of the signal given an observation of the *observed data* $D = (D_{ij})_{i,j \le N}$.

The posterior probability of the signal given D can be expressed in the form of a inhomogeneous vector spin model with respect to an arbitrary function g_{ij} and probability measure P_X

$$dP(X \mid D) = \frac{1}{Z_X^g(D)} \prod_{1 \le i < j \le N} e^{g_{ij}(D_{ij}, \frac{x_i \cdot x_j}{\sqrt{N}})} \prod_{1 \le i \le N} dP_X(x_i)$$

where

$$Z_X^g(D) = \int \prod_{1 \le i < j \le N} e^{g_{ij}(D_{ij}, \frac{x_i \cdot x_j}{\sqrt{N}})} \prod_{1 \le i \le N} dP_X(x_i)\,. \tag{2.2}$$

Our study will be restricted to the so-called Bayes optimal setting which amounts to make the following hypothesis:

Hypothesis 2.1 (Bayes-optimality)**.** Suppose that $P_0 = P_X$ and if P_{ij} is the distribution of D_{ij} in (2.1), the function $g_{ij}(D, w)$ is the log-likelihood:

$$g_{ij}(D, w) = \ln P_{ij}\Big(D \,\Big|\, \frac{x_i^0 \cdot x_j^0}{\sqrt{N}} = w\Big)\,.$$

where we use in short P_{ij} to denote also the probability density function of P_{ij}.

Our goal is to compute the normalized free energy

$$F_N(g) = \frac{1}{N}\left(\mathbb{E}_D\Big(\ln Z_X^g(D) - \sum_{i<j} g(D_{ij},0)\Big)\right) \tag{2.3}$$

where $\mathbb{E}_D$ is the expectation under $\otimes P_{ij}(D_{ij}|\frac{x_i^0\cdot x_j^0}{\sqrt{N}})P_0^{\otimes N}(x_i^0)$ and the function Z_X^g was defined in (2.2). Notice that, with this definition, the free energy is nothing but the expected log-likelihood ratio (under data generated by the model) between the likelihood that data are generated by the present model with the likelihood that they are generated from the null model (where there is no signal at all):

$$F_N(g) = \frac{1}{N}\left(\mathbb{E}_D \log \frac{P_D}{P_{D|x^0=0}}\right). \tag{2.4}$$

The limit of the free energy will depend on the following "Fisher Information" matrix, defined as the expectation of the Fisher score:

$$\frac{1}{\Delta_{ij}} = \mathbb{E}_{P_{ij}(D|w=0)}(\partial_w g_{ij}(D,0))^2, \quad 1 \le i < j \le N. \tag{2.5}$$

We first assume that this matrix of variances is piece-wise constant.

Hypothesis 2.2 (Block-constant Noise Profile)**.** Given $n \ge 1$, there exists a partition of $[N]$

$$[N] = \bigsqcup_{s=1}^{n} I_s$$

such that the $\Delta_{i,j}$ are constant in the groups $I_s \times I_t$ for $s,t \in \{1,\dots,n\}$

$$\Delta_{ij} = \Delta_{st}, \quad \text{for } i \in I_s,\ j \in I_t \tag{2.6}$$

and $(\Delta_{st})_{s,t\le n}$ are independent of N. Furthermore, the proportions of configurations in each group converges in the limit

$$\frac{|I_s|}{N} \to \rho_s \in (0,1) \quad \text{for all} \quad s \le n. \tag{2.7}$$

We will also assume that $\Delta_{s,t}$ belongs to $(0,+\infty)$ for each s,t and the $n\times n$ symmetric matrix $\frac{1}{\Delta}$ with entries $(\frac{1}{\Delta_{s,t}})_{s,t\le n}$ is positive semidefinite.

We finally describe our technical hypotheses. We first need to assume that the signal is compactly supported.

Hypothesis 2.3 (Compact Support)**.** P_0 and P_X are compactly supported so that x and x^0 take values in $[-C,C]^\kappa$ for some finite C. We also assume that κ is independent of N.

This hypothesis implies that, uniformly, we have

$$|w_{ij}| = \Big|\frac{x_i\cdot x_j}{\sqrt{N}}\Big| \le \frac{C^2\kappa}{\sqrt{N}}. \tag{2.8}$$

This uniform bound will allow to expand the functions g_{ij} in the variables w_{ij}. To do so, we need to assume sufficient regularity of the g_{ij}, namely that, if $\|\cdot\|$ denotes the supremum norm:

Hypothesis 2.4 (Regularity of Log Likelihood)**.** $\|\partial_w g_{ij}(\cdot,0)\|, \|\partial_w^2 g_{ij}(\cdot,0)\|, \|\partial_w^3 g_{ij}\|$ are bounded, uniformly in $1<i<j\le N$ and $N\in\mathbb{N}$.

Remark 2.5. We can weaken the condition on the first derivative of $\partial_w g_{ij}$. For the proof of universality in disorder in Lemma 3.3, we require the third moment $\mathbb{E}_{P_{ij}(D\,|\,w^0)}[\partial_w g_{ij}(D,w^0)]^3$ to be bounded for all D and all w^0 in the support.

There assumptions will be used to prove a universality result which states that a class of statistical inference problems are equivalent to a low rank matrix factorization problem where the noise matrix has a variance profile. In particular, we shall use as an application the degree-corrected stochastic block model.

2.2. **Gaussian Estimation Problems with Covariance Profiles.** The key point in our approach is that the general inhomogeneous vector spin model can be reduced to a model spiked Gaussian matrix model with a variance profile.

2.2.1. *Effective data matrix.* Our first observation says that the derivatives up to the second order derivatives of the function g encode asymptotically all its information. This holds even without the Bayes optimality Hypothesis 2.1.

Lemma 2.6 (*Free Energy Universality 1*)

Suppose that Hypothesis 2.4 and Hypothesis 2.3 are satisfied. Let B be a σ algebra such that the D_{ij} are independent conditionally to B. Then

$$F_N(g) = F_N(\bar{g}) + O\Big(\frac{\kappa^2}{\sqrt{N}}\Big)$$

with

$$\bar{g}_{ij}(D,w) = g_{ij}(D_{ij},0) + \partial_w g_{ij}(D,0)w + \frac{1}{2}\mathbb{E}_D[\partial_w^2 g_{ij}(D,0) \mid B]w^2\,.$$

This is a first fundamental universality result that motivates our study. Informally, this means that in order to perform (in high-dimensions) estimation when we observed the data D given by the likelihood

$$g_{ij}(D,w) = \ln P_{ij}\Big(D_{ij} \,\Big|\, \frac{x_i^0 \cdot x_j^0}{\sqrt{N}} = w\Big)\,,$$

we can simply create an *effective* data matrix Y based on the Fisher score and Fisher information as

$$Y_{ij} = \Delta_{ij}\partial_w g_{ij}(D,0), \text{ with } \frac{1}{\Delta_{ij}} = \mathbb{E}_{P_{ij}(D|w=0)}(\partial_w g_{ij}(D,0))^2 \tag{2.9}$$

then the free energy (or the likelihood ratio) depends only on this new matrix.

2.2.2. *Equivalent spike model.* Under the Bayes optimality assumption Hypothesis 2.1, we can further simplify the free energy by connecting with a Gaussian spike model. i.e. a model where the effective matrix Y is indeed sampled from a Gaussian spike model with inhomogeneous noise.

For a $N \times N$ Gaussian matrix W, a deterministic $N \times N$ standard deviation matrix $\Delta^{\odot\frac{1}{2}}$, and a $N \times \kappa$ matrix x^0 with column vectors $x_i^0 \in \mathbb{R}^\kappa$ we want to estimate

$$Y = Y^\Delta = \Delta^{\odot\frac{1}{2}} \odot W + \sqrt{\frac{1}{N}}x^0(x^0)^\intercal \tag{2.10}$$

in other words, for $1 \le i < j \le N$

$$Y_{ij} = \sqrt{\Delta_{ij}}W_{ij} + \frac{1}{\sqrt{N}}x_i^0 \cdot x_j^0.$$

Note that the Δ_{ij} are non negative by (2.9). We also only care about the off-diagonal terms because the diagonals are negligible.

The main difference in this setting, in contrast to the standard spiked matrix models, is that the coordinates i, j play an important role in the behavior of our model. Observe that

$$\Delta_{ij}^{-\frac{1}{2}}\Big(Y_{ij} - \frac{1}{\sqrt{N}}x_i^0 \cdot x_j^0\Big)$$

follows a standard Gaussian law. Then the random posterior distribution of $X = (x_1, \dots, x_N)$ is

$$dP(X \mid Y) = \frac{1}{Z}\exp\Big(-\sum_{i<j}\frac{1}{2\Delta_{ij}}\Big(Y_{ij} - \frac{1}{\sqrt{N}}x_i \cdot x_j\Big)^2\Big)\,dP_0^{\otimes N}(X).$$

After absorbing the terms that do not depend on X into the normalization, the density is encoded by the Hamiltonian given by

$$\begin{aligned} H_N(x) &= \sum_{i<j}\frac{1}{\sqrt{\Delta_{ij}N}}Y_{ij}(x_i \cdot x_j) - \frac{1}{2\Delta_{ij}N}(x_i \cdot x_j)^2 \\ &= \sum_{i<j}\frac{1}{\sqrt{N\Delta_{ij}}}g_{ij}(x_i \cdot x_j) + \frac{1}{\Delta_{ij}N}(x_i^0 \cdot x_j^0)(x_i \cdot x_j) - \frac{1}{2\Delta_{ij}N}(x_i \cdot x_j)^2. \end{aligned} \tag{2.11}$$

We define

$$F_N(\Delta) = \frac{1}{N}\mathbb{E}_Y \log \int e^{H_N(x)}\,dP_0^{\otimes N}(x) \tag{2.12}$$

where $H_N(x)$ is the Hamiltonian defined in (2.11) and Δ is the variance profile.

We will prove that solving the free energy of Gaussian estimation problems are equivalent to solving the free energy of general inhomogeneous vector spin models with a specific choice of parameters in the Bayes optimal setting:

Theorem 2.7 (*Free Energy Universality 2*)

Suppose we are in the Bayes optimal setting of Hypothesis 2.1, g satisfies Hypothesis 2.4 and the signal space is compact as in Hypothesis 2.3. If we define

$$\frac{1}{\Delta_{ij}} = \mathbb{E}_{P_{ij}(Y|w=0)}(\partial_w g_{ij}(Y,0))^2 = \int (\partial_w g_{ij}(Y,0))^2 e^{g_{ij}(y,0)}\, dy,$$

then the free energy $F_N(g)$ of the inhomogeneous vector spin models (2.3) satisfies

$$\left|F_N(g) - F_N(\Delta)\right| = O(\kappa^3 N^{-1/2}).$$

Remark 2.8. We will need that the matrix $1/\Delta$ satisfy Hypothesis 2.2 or 2.11 to get the limit of the free energy. Similarly, κ could go to infinity with N provided $\kappa^3 N^{-1/2}$ goes to zero but then one would need to understand the asymptotics of $\sup_Q \varphi(Q)$: this will be the subject of a separate work.

We will also derive a stronger form of universality at the level of the spectrum of random matrices, instead of at the level of the free energy. Consider the transformed data matrix

$$\frac{\tilde{Y}_{ij}}{\sqrt{N}} = \frac{1}{\sqrt{N}}\Big(\partial_w g_{ij}(D,w)\Big)\Big|_{D=D_{ij},w=0} \qquad i,j \le N$$

where D_{ij} is the random variable with law given by (1.1) and the spiked matrix with variance profile Y^Δ

$$Y^\Delta = \Delta^{\odot\frac{1}{2}} \odot W + \frac{x^0(x^0)^\intercal}{\sqrt{N}}$$

defined in (2.10). Under some conditions on the smallest entries of the Fisher information matrix $\frac{1}{\Delta}$ (see Hypothesis 3.13), we have the following universality result for the spectrum.

Theorem 2.9 (*Universality of the Spectrum*)

If g satisfies Hypothesis 2.2 and the corresponding Fisher information matrix (2.5) satisfies Hypothesis 3.13, then

(1) Conditionally on x^0, the empirical distribution μ_1 of the eigenvalues of $\frac{\tilde{Y}_{ij}}{\sqrt{N}}$ and the empirical distribution μ_2 of the eigenvalues of $\frac{1}{\sqrt{N}\Delta} \odot Y^\Delta$ satisfy

$$\lim_{N\to\infty} d(\mu_1,\mu_2) \to 0$$

in probability, where d is a distance compatible with the weak topology.

(2) Conditionally on x^0, when the dimension goes to infinity, $\frac{1}{\sqrt{N}\Delta} \odot Y^\Delta$ has an extremal eigenvalue away from the bulk if and only if $\frac{\tilde{Y}_{ij}}{\sqrt{N}}$ does, for almost all Δ and ρ.

2.3. **The Limit of the Free Energy and Consequences.** Given a sequence $Q = (Q_s)_{s\le n}$ of symmetric $\kappa\times\kappa$ positive semidefinite matrices, the replica symmetric free energy is given by

$$\varphi(Q) = -\sum_{s,t=1}^n \frac{\rho_s\rho_t}{4\Delta_{st}} \operatorname{Tr}(Q_sQ_t) + \sum_{s=1}^n \rho_s \mathbb{E}_{z,x^0} \ln\left[\int e^{(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s} z)^\intercal x - \frac{x^\intercal \tilde{Q}_s x}{2}}\, dP_0(x)\right] \tag{2.13}$$

where

$$\tilde{Q}_s = \sum_{t=1}^n \frac{1}{\Delta_{st}} \rho_t Q_t,$$

and $x^0 \sim P_0$ and $z \sim N(0, I_r)$ are independent. We have the following limit of the free energy.

Theorem 2.10 (*Bayes Optimal Free Energy*)

Suppose that Hypotheses 2.1 and 2.2 are verified, as well as technical Hypotheses 2.3 and 2.4. Then,

$$\lim_{N\to\infty} F_N(g) = \sup_Q \varphi(Q).$$

We can generalize Theorem 2.10 to the case where Δ is not piecewise constant, but we then need to approximate it by piecewise constant matrices. We then replace Hypothesis 2.2 by the following:

Hypothesis 2.11. Assume that there exists a non-negative measurable function $\Delta(s,t)$ such that

$$\lim_{N\to\infty} \sup_{s,t\in[0,1]} \left| \frac{1}{\Delta_{\lfloor sN\rfloor,\lfloor tN\rfloor}} - \frac{1}{\Delta(s,t)} \right| = 0.$$

We will also assume that there exists $\varepsilon > 0$ such that $\Delta(s,t)$ belongs to $(\varepsilon, 1/\varepsilon)$ for each $s,t \in [0,1]$ and the symmetric operator $\frac{1}{\Delta}$ on $L^2([0,1])$ given by

$$\frac{1}{\Delta} f(t) = \int_0^1 \frac{1}{\Delta(t,s)} f(s) ds$$

is non-negative.

Remark 2.12. This hypothesis is satisfied if, for instance, Δ is C^0 and bounded below.

Under this hypothesis, Theorem 2.10 generalizes as follows. Let $Q_s : [0,1] \mapsto \mathbf{\Sigma}_\kappa^+$ be a measurable function with values in the set of $\kappa \times \kappa$ symmetric definite matrices, and define

$$\tilde{\varphi}(Q) := -\int_0^1 \frac{1}{4\Delta(s,t)} \operatorname{Tr}(Q_s Q_t) ds dt + \int_0^1 ds \mathbb{E}_{z,x^0} \ln \left[\int e^{(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s} z)^\intercal x - \frac{x^\intercal \tilde{Q}_s x}{2}} dP_0(x) \right]$$

where

$$\tilde{Q}_s = \int_0^1 \frac{1}{\Delta(s,t)} Q_t dt.$$

Then, we have:

Theorem 2.13 (*Bayes Optimal Free Energy with General Kernel*)

Suppose that Hypotheses 2.1 and 2.11 are verified, as well as the technical Hypotheses 2.3 and 2.4. Then,

$$\lim_{N\to\infty} F_N(g) = \sup_Q \tilde{\varphi}(Q).$$

When the variance profile is discrete and if the probability measure P_0 is centered, the maximizers of φ defined in (2.13) satisfy the following fixed point equation

$$\tilde{Q}_s = \sum_{t=1}^n \frac{\rho_t}{\Delta_{s,t}} \mathbb{E} \langle x \rangle_{\tilde{Q}_t} \langle x \rangle_{\tilde{Q}_t}^\intercal \tag{2.14}$$

where $\langle \cdot \rangle_{\tilde{Q}_t}$ denotes the average

$$\langle f(x) \rangle_{\tilde{Q}_t} = \frac{\int f(x) e^{(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s} z)^\intercal x - \frac{x^\intercal \tilde{Q}_s x}{2}} dP_0(x)}{\int e^{(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s} z)^\intercal x - \frac{x^\intercal \tilde{Q}_s x}{2}} dP_0(x)}.$$

This fact allows us to compute the informationally theoretical optimal thresholds of the inhomogeneous estimation problems.

We define the matrix minimal means square estimator of our signal $\boldsymbol{x}^0 (\boldsymbol{x}^0)^\intercal$ by

$$\mathrm{MMSE}(N) = \min_\theta \frac{2}{N(N-1)} \sum_{i<j} \mathbb{E}(x_i^0 \cdot x_j^0 - \theta_{i,j}(Y))^2$$

where the minimum is over all possible estimators θ that only depend on the data Y. We have the following result for the limit of the minimal mean squared error.

Corollary 2.14 (*Limiting MMSE*)

Suppose that Hypotheses 2.1 and 2.2 are verified, as well as technical Hypotheses 2.3 and 2.4. Then for almost all Δ and ρ, for any maximizer $(Q_1, \dots, Q_n)$ of (2.13) and

$$\lim_{N\to\infty} \mathrm{MMSE}(N) = \mathbb{E}_{P_0} \|x x^\intercal\|_2^2 - \sum_{s,t=1}^n \rho_s \rho_t \operatorname{Tr}(Q_s Q_t).$$

It follows that the maximizers of (2.13) and the fact that they vanish or not are essential to quantify the information theoretic thresholds for these inhomogenous factorization problems. Notice that we do not prove the uniqueness of the maximizers but the uniqueness of the values of $(\mathrm{Tr}(Q_s Q_t))_{s,t}$ regardless of the choice of the maximizer, which is enough to guarantee the above statement.

We now classify detectability phase transitions with respect to the size of the variance profiles. Consider the noise parameter matrices

$$\rho = \mathrm{diag}(\rho_1, \dots, \rho_n) \in \mathbb{R}^{n \times n} \qquad \frac{1}{\Delta} = \left(\frac{1}{\Delta_{s,t}}\right) \in \mathbb{R}^{n \times n}.$$

The size of the noise of the inhomogeneous models can be encoded by the largest eigenvalue of the matrix $\sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho}$. Let $\|\cdot\|_{op}$ denote the operator norm of the matrix, which is equivalent to the largest eigenvalue when the matrix is symmetric and positive semidefinite. We have the following thresholds on recovery.

Lemma 2.15 (*Recovery Transitions*)

Suppose that Hypotheses 2.1 and 2.2 are verified, as well as technical Hypotheses 2.3 and 2.4. Furthermore, suppose that P_0 is symmetric. We have that

(1) If

$$\left\| \sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho} \right\|_{op} < \frac{1}{9\kappa^4 C^6}$$

then $\lim_{N\to\infty} F_N(g) = 0$ and $\lim_{N\to\infty} \mathrm{MMSE}(N) = \mathbb{E}_{P_0}\|xx^\intercal\|_2^2$.

(2) If

$$\left\| \sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho} \right\|_{op} > \frac{1}{\|\mathrm{Cov}(x)\|_{op}^2}$$

then $\lim_{N\to\infty} F_N(g) > 0$ and $\lim_{N\to\infty} \mathrm{MMSE}(N) < \mathbb{E}_{P_0}\|xx^\intercal\|_2^2$.

Remark 2.16. If P_0 is a standard Gaussian, then we can improve the first bound to conclude that if

$$\left\| \sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho} \right\|_{op} < 1$$

then the signal is not recoverable, which gives us a sharp phase transition when the priors are Gaussian. Technically P_0 does not have compact support, so it will violate Hypothesis 2.3, so universality of models with Gaussian signals will require a bit more work to show. However, the computation of the free energy of the inhomogenous spiked matrix model (2.10) does not require the compact support Hypothesis, so we do have a rigorous proof of the sharp transition in this case.

There is a gap between the two thresholds in Lemma 2.15. To explore the phase transition, we can numerically solve for the maximizers of φ defined in (2.13) to test if the phase transition is tight.

We first check the behavior for $\kappa = 1$, $n = 2$, and a non-sparse symmetric prior

$$P_0(\pm 1) = \frac{1}{4} \quad P_0(0) = \frac{1}{2}.$$

Under these assumptions, it follows that $\|\mathrm{Cov}(x)\|_{op}^2 = \frac{1}{4}$, so the phase transition in part 2 of Lemma 2.15 happens at $\left\| \sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho} \right\|_{op} = 4$. For matrix paths $\rho(t)$ and matrix $\Delta(t)$ we plot the free energy $F_N(g)$ corresponding to the parameters $\Delta(t)$ and $\rho(t)$ versus the operator norm of $\sqrt{\rho(t)}\frac{1}{\Delta(t)}\sqrt{\rho(t)}$ where $t \in \mathbb{R}^+$. We consider the following noise profiles

$$\frac{1}{\Delta_1(t)} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & t \end{bmatrix} \qquad \frac{1}{\Delta_2(t)} = \begin{bmatrix} t & \frac{t}{2} \\ \frac{t}{2} & t \end{bmatrix} \qquad \frac{1}{\Delta_3(t)} = \begin{bmatrix} \frac{t}{3} & \frac{t}{4} \\ \frac{t}{4} & t \end{bmatrix}$$

with proportion $\rho = \mathrm{diag}(0.4, 0.6)$. Figure 1 shows the relationship between the free energy and the inhomogeneous noises at the 3 different choices of the matrices Δ and ρ. We see in the figure that the transition derived in part 2 of Lemma 2.15 appears tight in this case when the prior is not sparse.

Another example, where the transition in part 2 of Lemma 2.15 does not appear to be tight is when estimating a sparse signal. We thus now examine the behavior for $\kappa = 1$, $n = 2$, and symmetric prior

$$P_0(\pm 1) = 0.03 \quad P_0(0) = 0.94.$$

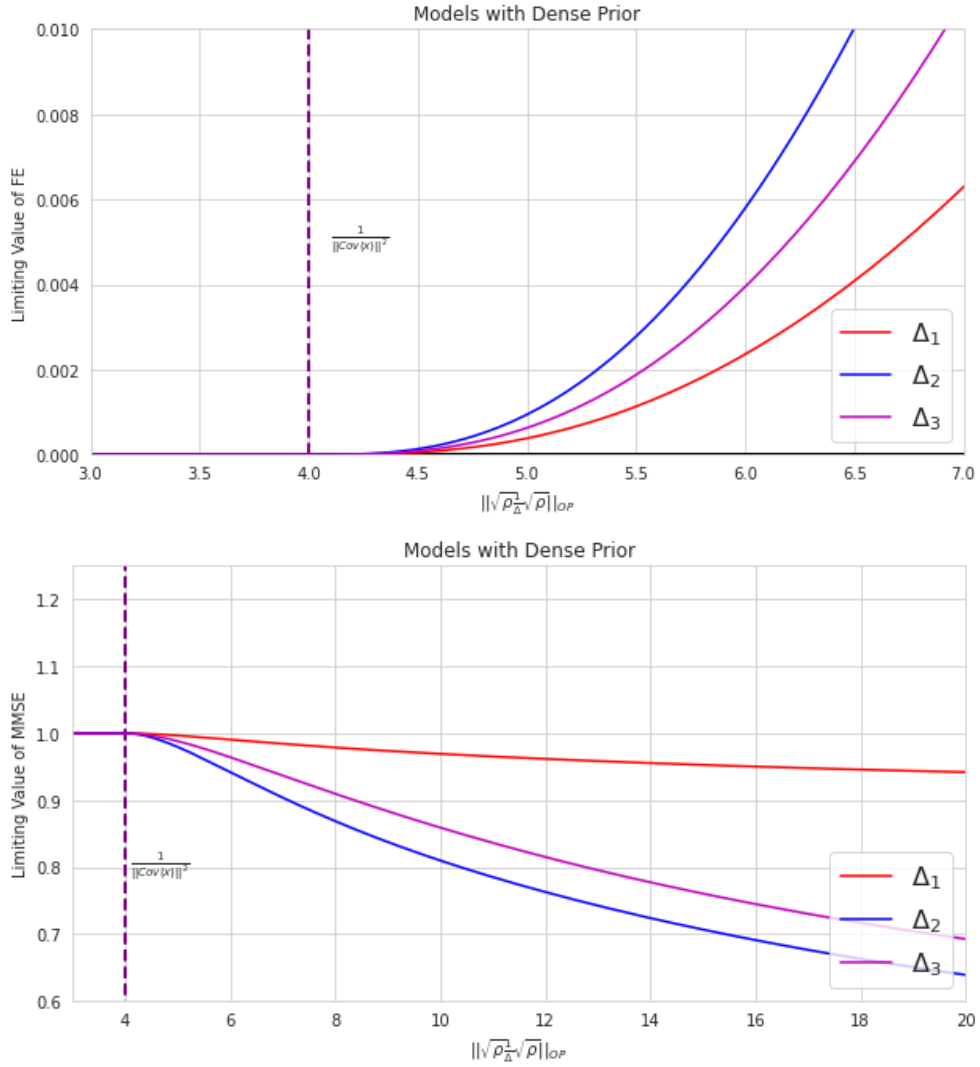

FIGURE 1. The free energy as a function of the noise. A case with a continuous phase transition (dashed line) separating the undetectable regime from one where detection of the signal is possible.

Under these assumptions, it follows that $\frac{1}{\|\operatorname{Cov}(x)\|_{op}^2} \approx 278$. In Figure 2 we plot the limit of the free energy of the matrix

$$\frac{1}{\Delta(t)} = \begin{bmatrix} t & \frac{t}{2} \\ \frac{t}{2} & 2t \end{bmatrix}$$

and proportion $\rho = \operatorname{diag}(0.4, 0.6)$ versus the operator norm of $\sqrt{\rho}\frac{1}{\Delta(t)}\sqrt{\rho}$ where $t \in \mathbb{R}^+$, In this scenario, there is a clear gap between the upper bound on the transition $\frac{1}{\|\operatorname{Cov}(x)\|^2}$ (denoted by the purple line) and the actual transition (denoted by the green line) when the free energy becomes positive. In this paper we do not focus on algorithms but in analogy with the homogeneous case [31, 30] we anticipate that the transition at $\frac{1}{\|\operatorname{Cov}(x)\|^2}$ in fact has an algorithmic meaning of a threshold beyond which a corresponding message passing algorithm is able to find a vector correlated with the signal and below which it is not. The region in between these two threshold is then conjectured to be algorithmically hard.

Next, we examine the behavior of the limiting free energy as the proportions of the two blocks vary in the same setting of the sparse Rademacher prior. We consider the case when $\rho = (p, 1-p)$. We consider fixed

$$\frac{1}{\Delta(t)} = \begin{bmatrix} t & 0 \\ 0 & 2t \end{bmatrix}.$$

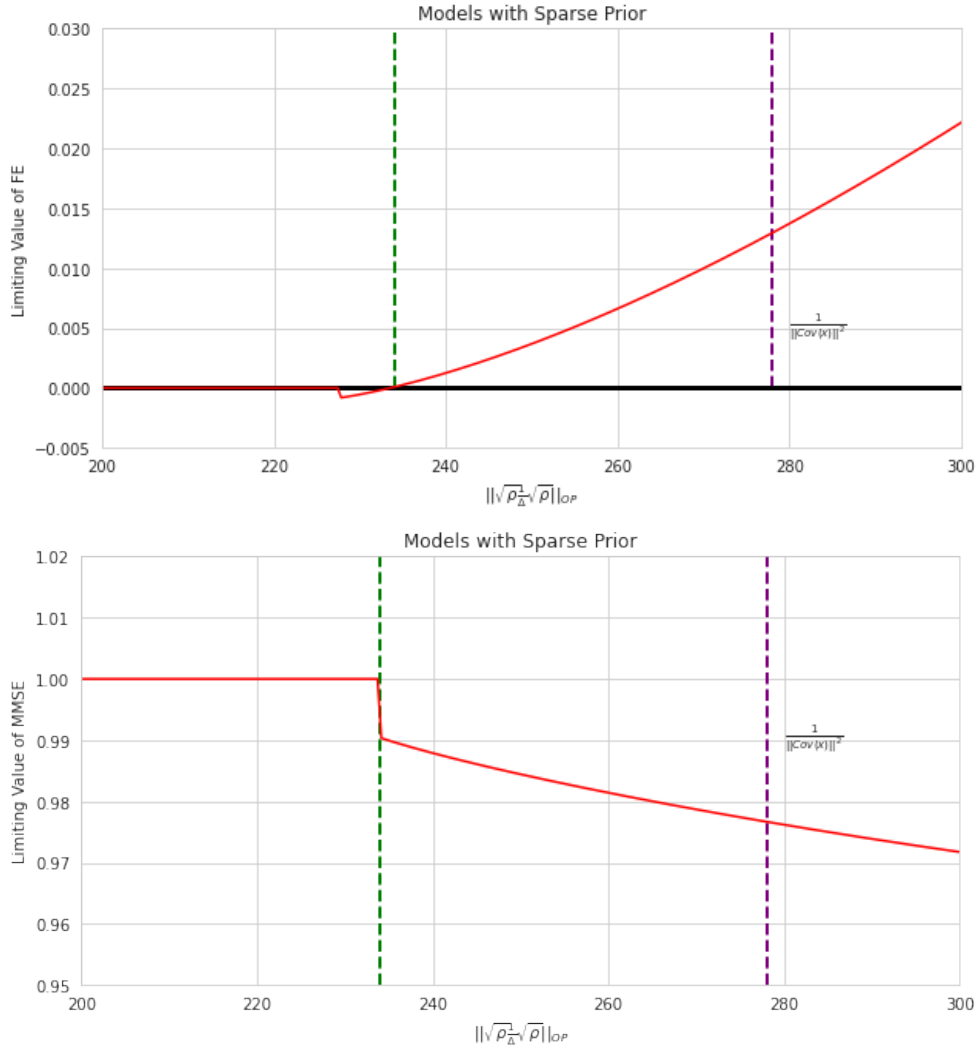

FIGURE 2. The free energy as a function of the noise. A case with a discontinuous phase transition (orange dashed line) separating the undetectable regime from one where detection of the signal is possible. We plots the free energy of two local maximizers (in blue and red), it is the larger one that provides the final result. The purple dashed line is a position of the threshold from part 2 of Lemma 2.15 that is not tight in this case.

For each choice of $p \in \{0.1, 0.2, 0.3, 0.4\}$, we again plot in Figure 3 the limit of the free energy of the matrix $\Delta(t)$ and proportion ρ versus the operator norm of $\sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho}$ where $t \in \mathbb{R}^+$. In this case, the phase transition is independent of the proportions p.

If we now consider a non-diagonal inhomogeneity matrix

$$\frac{1}{\Delta} = \begin{bmatrix} t & t \\ t & 4t \end{bmatrix}$$

the phase transition is dependent on the proportions p as depicted in Figure 4. It remains to study the dependence of the limiting free energies on the structure of the noise more rigorously. In particular, it appears that phase transition at $\frac{1}{\|\operatorname{Cov}(x)\|_{op}^2}$ can be improved for certain choices of sparse symmetric priors. However, this will likely be a difficult problem to solve, because it will depend on both the covariance structure Δ and the proportions ρ.

2.4. **The BBP Transition of a Transformed Homogeneous Noise Model.** Recovery transitions can also be studied from the point of view of the BBP transition [5] in random matrix theory. In the following, we

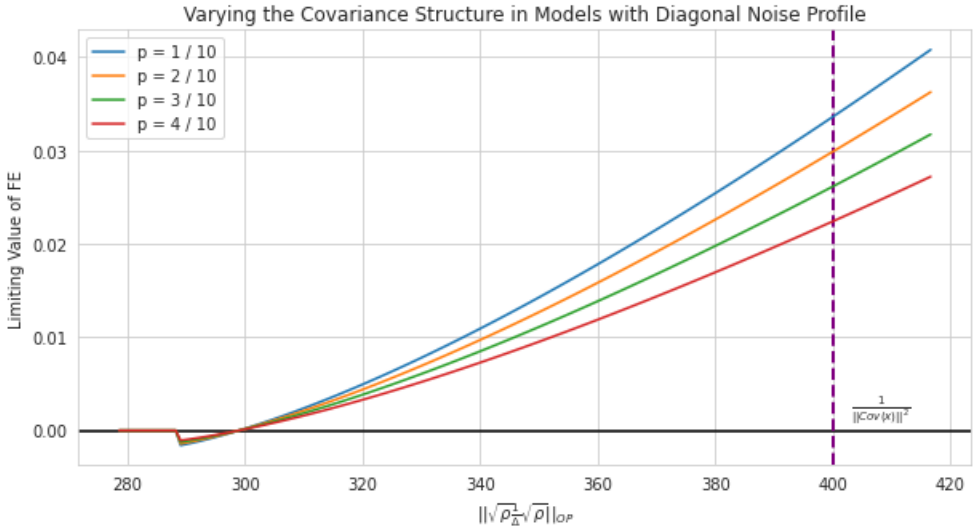

FIGURE 3. The free energy as a function of the noise parameter, for groups of different sizes p, and diagonal inhomogeneity Δ.

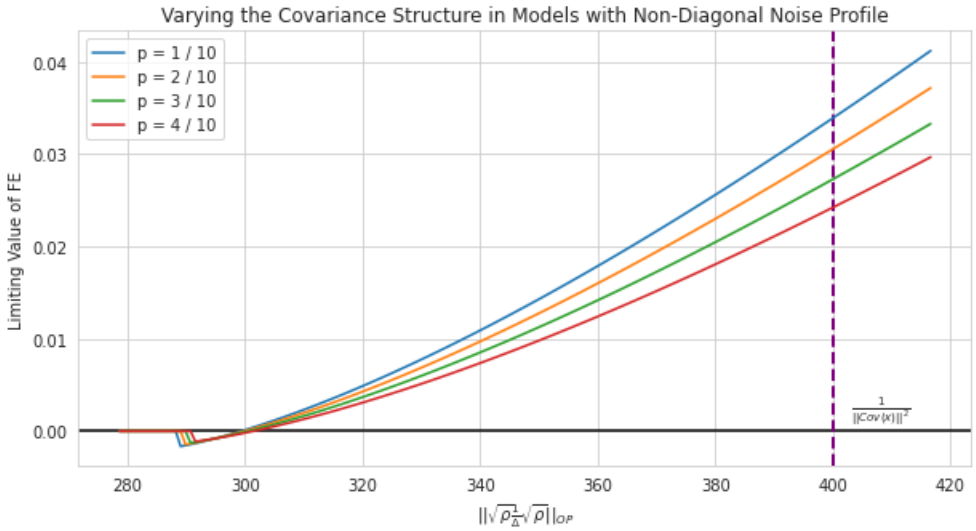

FIGURE 4. The free energy as a function of the noise parameter, for groups of different sizes p, and nondiagonal inhomogeneity Δ.

compute the transition for a normalization of the spiked matrix that will move the noise profile onto the signal,

$$\frac{1}{\sqrt{N}}\frac{1}{\Delta^{\odot\frac{1}{2}}}\odot Y^{\Delta}=\frac{1}{\sqrt{N}}W+\frac{1}{N}\frac{1}{\Delta^{\odot\frac{1}{2}}}\odot x^0(x^0)^{\intercal}.$$

The normailzation by $\frac{1}{\sqrt{N}}$ was to ensure the eigenvalue distribution is supported almost surely on a compact set, and the normalization by $\frac{1}{\Delta^{\odot\frac{1}{2}}}$ moved the noise profile onto the signal. Classical random matrix theory arguments can be used to compute when the largest eigenvalue separates from the bulk of the transformed matrix.

Lemma 2.17 (*BBP Transition*)

The largest eigenvalue of the matrix $\frac{1}{\sqrt{N}}Y \odot \frac{1}{\Delta^{\odot \frac{1}{2}}}$ *separates from the bulk when*

$$\left\| \sqrt{\rho} \frac{1}{\Delta^{\odot \frac{1}{2}}} \sqrt{\rho} \right\|_{op} > \frac{1}{\| \operatorname{Cov}(x) \|_{op}}.$$

When the noise profile Δ is not identically equal to one, the region of noise described in Lemma (2.17) where the top eigenvalue separates from the bulk is contained in the recovery regime Lemma 2.15. This means that a naive reduction of the inhomogeneous problem to a homogeneous problem does not yield the sharpest transition for these inhomogeneous noise models. This behavior is different from the classical homogeneous noise models.

Proposition 2.18 (*Gap in Thresholds*)

We have

$$\left\| \sqrt{\rho} \frac{1}{\Delta^{\odot \frac{1}{2}}} \sqrt{\rho} \right\|_{op}^2 \leq \left\| \sqrt{\rho} \frac{1}{\Delta} \sqrt{\rho} \right\|_{op},$$

with equality holding if and only if $\Delta = c$ *for some constant* c*. In particular, when* $\Delta \neq c$

$$\left\{ \left\| \sqrt{\rho} \frac{1}{\Delta^{\odot \frac{1}{2}}} \sqrt{\rho} \right\|_{op} \geq \frac{1}{\| \operatorname{Cov}(x) \|_{op}} \right\} \subset \left\{ \left\| \sqrt{\rho} \frac{1}{\Delta} \sqrt{\rho} \right\|_{op} \geq \frac{1}{\| \operatorname{Cov}(x) \|_{op}^2} \right\}$$

so the information theoretic detectability regime strictly contains the BBP transition.

2.5. **Application to the Degree Corrected Stochastic Block Model.** We now apply our results to the degree corrected stochastic block models (DCSBM) introduced in [26].

We consider a dense community detection problem with $\kappa \geq 1$ communities in a network of N nodes. If the vertex i belongs to the community $t_i \in \{1, \dots, \kappa\}$, then its membership can be encoded by $x_i = e_{t_i} \in \mathbb{R}^\kappa$ where e_{t_i} is the t_ith basis element in $\mathbb{R}^\kappa$. In particular, for any two nodes i and j, $x_i \cdot x_j = 1$ if they belong to the same community and $x_i \cdot x_j = 0$ otherwise. If we choose the prior P_0 to be uniform on all basis elements, then we are in the setting with equally sized groups.

We now build a site dependent adjacency matrix that encodes different probability of attaching an edge depending on if the vertices are in the same group or not. In contrast to the standard stochastic block model, we consider a vector of site specific parameters $(\theta_1, \dots, \theta_N) \in (0,1)^N$ which controls the global probability of any vertex being attached to node i. For each i, j we also define a $\kappa \times \kappa$ connectivity matrix

$$\boldsymbol{C}_{ij} = \begin{bmatrix} \theta_i\theta_j & \theta_i\theta_j & \dots & \theta_i\theta_j \\ \theta_i\theta_j & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \theta_i\theta_j \\ \theta_i\theta_j & \dots & \theta_i\theta_j & \theta_i\theta_j \end{bmatrix} + \frac{\lambda}{\sqrt{N}} \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & 1 \end{bmatrix}.$$

In particular, this implies that the probability to attach an edge between vertex i and vertex j is $\theta_i\theta_j$ if they are in different communities and if they are in the same community, then the probability is different by an additive factor of $\frac{\lambda}{\sqrt{N}}$. The precise output channel for the adjacency matrix of the entire graph is given by

$$P_{\text{out}}(D_{ij} = 1 \mid x^0, \boldsymbol{\theta}) = \theta_i\theta_j + \frac{\lambda}{\sqrt{N}} x_i^0 \cdot x_j^0. \tag{2.15}$$

We will generalize the framework slightly and consider parameters $(\theta_i)_{i \leq N}$ with each θ_i sampled independently from some distribution P_θ on $(0,1) \subset \mathbb{R}^+$ (the precise interval doesn't matter because we can scale the $p_{\text{out}} > 0$). As usual, we also generate a vector $x^0 \in \mathbb{R}^\kappa$ according to P_0. Finally, we generate the adjacency matrix D_{ij} according to (2.15). The parameter λ is the signal to noise ratio in these models.

Remark 2.19. If $P_\theta(\theta = 1) = 1$, then this model reduces to the classical stochastic block model.

2.5.1. *The Limit of the Free Energy.* We now state the limit of the free energies in these models. Suppose that P_θ is supported on finitely many points $(\theta^{(1)}, \dots, \theta^{(n)}) \in (0,1)$ with probabilities $\rho_1, \dots, \rho_n$. We define $\Delta = (\Delta_{s,t})_{1 \leq s,t \leq n}$ by

$$\frac{1}{\Delta_{st}} = \left(\frac{\lambda^2}{\theta^{(s)}\theta^{(t)}} + \frac{\lambda^2}{(1 - \theta^{(s)}\theta^{(t)})} \right) \tag{2.16}$$

and consider the conditional probabilities

$$dP(X|D,\Theta) = \frac{1}{Z_X(D,\Theta)} \prod_{1\le i<j\le N} e^{\ln P_{\rm out}(D_{ij}\,|\,x,\boldsymbol{\theta})} \prod_{1\le i\le N} dP_0(x_i) \tag{2.17}$$

where

$$P_{\rm out}(D_{ij}=1\,|\,x,\boldsymbol{\theta}) = \theta_i\theta_j + \frac{\lambda}{\sqrt{N}} x_i\cdot x_j \quad \text{and} \quad P_{\rm out}(D_{ij}=0\,|\,x^0,\boldsymbol{\theta}) = 1-\theta_i\theta_j - \frac{\lambda}{\sqrt{N}} x_i\cdot x_j. \tag{2.18}$$

We define the collection $Q=(Q_1,\dots,Q_n)$ of symmetric positive semidefinite $\kappa\times\kappa$ matrices and the functional

$$\varphi(Q) = -\sum_{s,t=1}^n \frac{\rho_s\rho_t}{4\Delta_{s,t}} \mathrm{Tr}(Q_sQ_t) + \sum_{s=1}^n \rho_s \mathbb{E}_{z,x^0} \ln\left[\int e^{(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s} z)^\intercal x - \frac{x^\intercal \tilde{Q}_s x}{2}}\, dP_0(x)\right] \tag{2.19}$$

where for $s\in\{1,\dots,n\}$,

$$\tilde{Q}_s = \sum_{t=1}^n \frac{1}{\Delta_{s,t}}\rho_t Q_t$$

$x^0\sim P_0$ and $z\sim N(0,I_\kappa)$ are independent. The limit of the free energy is given by maximizing this functional.

Corollary 2.20 (*Free Energy for Discrete Degree Corrected Stochastic Block Models*)

If $P_0=\sum_{i=1}^n \rho_i\delta_{\theta^{(i)}}$ with $\theta^{(1)},\dots,\theta^{(n)}\in(0,1)$ then the free energy of the degree corrected stochastic block model is

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}\left(\ln Z_X(D,\Theta) - \sum_{i<j}\ln P_{\rm out}(D_{ij}\,|\,x=0,\boldsymbol{\theta})\right) = \sup_Q \varphi(Q).$$

where the parameters of the model are given in (2.16), (2.17), (2.18) and (2.19).

Proof. We essentially check that conditional probabilities in (2.18) satisfy the conditions for the inhomogeneous vector spin models required in Theorem 2.10. Hypotheses 2.1 and 2.3 are clearly verified. We next check the two other assumptions.

Hypothesis 2.4 (Bounds on the derivatives): The corresponding function g_{ij} is the log likelihood expressed as a function of D_{ij} and $\frac{1}{\sqrt{N}}x_i^0\cdot x_j^0$,

$$g_{ij}(1,w) = \ln P_{\rm out}(D_{ij}=1\,|\,w,\boldsymbol{\theta}) = \ln(\theta_i\theta_j+\lambda w)$$

and

$$g_{ij}(0,w) = \ln P_{\rm out}(D_{ij}=0\,|\,w,\boldsymbol{\theta}) = \ln(1-\theta_i\theta_j-\lambda w).$$

Notice that our function g is smooth in w, and well defined for w such that $1-\theta_i\theta_j-\lambda w\in(0,1)$. Furthermore,

$$\partial_w g_{ij}(D,w) = D\frac{\lambda}{\theta_i\theta_j+\lambda w} - (1-D)\frac{\lambda}{1-\theta_i\theta_j-\lambda w} \tag{2.20}$$

$$\partial_w^2 g_{ij}(D,w) = D\frac{-\lambda^2}{(\theta_i\theta_j+\lambda w)^2} - (1-D)\frac{\lambda^2}{(1-\theta_i\theta_j-\lambda w)^2} \tag{2.21}$$

$$\partial_w^3 g_{ij}(D,w) = D\frac{2\lambda^3}{(\theta_i\theta_j+\lambda w)^3} - (1-D)\frac{2\lambda^3}{(1-\theta_i\theta_j-\lambda w)^3} \tag{2.22}$$

which are all uniformly bounded provided that $0<\theta_i\theta_j+\lambda w<1$ for all i and j and w in the domain. Since w goes to zero, for N large enough, it is enough that the $\theta^{(i)}$ belong to $(0,1)$ independently of N.

Hypothesis 2.2 (Positive Definiteness of Noise Profile): By definition, we have

$$\begin{aligned}\frac{1}{\Delta_{ij}} = \mathbb{E}_{P_{ij}(D|w=0)}\left[(\partial_w g_{ij}(D,0))^2\right] &= \left(P(D_{ij}=1\,|\,w_{ij}^0=0,\theta)\frac{\lambda^2}{\theta_i^2\theta_j^2} + P(D_{ij}=0\,|\,w_{ij}^0=0,\theta)\frac{\lambda^2}{(1-\theta_i\theta_j)^2}\right)\\ &= \left(\frac{\lambda^2}{\theta_i\theta_j} + \frac{\lambda^2}{(1-\theta_i\theta_j)}\right)\end{aligned} \tag{2.23}$$

because $P_{ij}(D_{ij} = 1 \mid 0, \theta) = \theta_i\theta_j = 1 - P(D_{ij} = 0 \mid 0, \theta)$. Observe that Δ_{ij} is piecewise constant since the θ_i's are piecewise constant. We denote as well Δ the $n \times n$ symmetric defined in Hypothesis 2.2. We next show that the matrix $\frac{1}{\Delta^2}$ is non-negative. To see this, notice that

$$\left[\frac{\lambda^2}{\theta^((s)\theta^(t)}\right]_{1\le s,t\le n} = \lambda^2\left[\frac{1}{\theta^{(s)}}\cdot\frac{1}{\theta^{(t)}}\right]_{1\le i,j\le n} \tag{2.24}$$

is positive semidefinite because it is a Gram matrix. Furthermore, for $|\theta^{(s)}\theta^{(t)}| < 1$, we have

$$\frac{\lambda^2}{(1-\theta^{(s)}\theta^{(t)})} = \lambda^2\sum_{k=0}^{\infty}(\theta^{(s)})^k(\theta^{(t)})^k,$$

by its Taylor expansion, so

$$\left[\frac{\lambda^2}{(1-\theta^{(s)}\theta^{(t)})}\right]_{1\le s,t\le n} = \lambda^2\sum_{k=0}^{\infty}\Theta^{\odot k}, \tag{2.25}$$

where $\Theta = [\theta^{(s)}\theta^{(t)}]_{1\le s,t\le n}$. The matrix Θ is positive semidefinite because it is a Gram matrix, and the Schur product theorem implies that the Hadamard powers are also positive semidefinite, so is also (2.25) positive semidefinite. This concludes the proof. □

We can also apply Theorem 2.13 to this setting. To this end we assume that

Hypothesis 2.21. Assume that $\theta_{\lfloor Ns\rfloor}$ converges uniformly to a measurable function $(\theta^{(s)})_{s\in[0,1]}$ with values in a compact subset $[a, b]$ of $(0, 1)$.

Let $Q(\theta) : [a, b] \to \mathbb{S}^+_\kappa$ be a measurable matrix valued function. Let θ and θ' be independent samples from $\tilde{P}_\theta$. We define

$$\frac{1}{\Delta(\theta,\theta')} = \left(\frac{\lambda^2}{\theta\theta'} + \frac{\lambda^2}{(1-\theta\theta')}\right) \tag{2.26}$$

We have

$$\varphi(Q) = -\int_0^1 \frac{\mathrm{Tr}(Q(\theta^{(s)}), Q(\theta^{(t)}))}{4\Delta(\theta^{(s)},\theta^{(t)})}dsdt + \int_0^1 ds \ln\left[\int e^{(\tilde{Q}(\theta^{(s)})x^0 + \sqrt{\tilde{Q}(\theta^{(s)})}z)^\mathsf{T}x - \frac{x^\mathsf{T}\tilde{Q}(\theta^{(s)})x}{2}}dP_0(x)\right] \tag{2.27}$$

where the expected value is with respect to θ, θ', x^0, z and

$$\tilde{Q}(\theta) = \int_0^1 \frac{1}{\Delta(\theta,\theta^{(s)})}Q(\theta^{(s)})ds\,. \tag{2.28}$$

It follows by approximation that Corollary 2.20 holds in the limit.

Corollary 2.22 (*Free Energy for Degree Corrected Stochastic Block Models*)

Under Hypothesis 2.21, the free energy of the degree corrected stochastic block model is given by

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}\ln\left(Z_X(Y,\tilde{\Theta}) - \sum_{i<j}\ln P_{\mathrm{out}}(Y_{ij} \mid x = 0, \boldsymbol{\theta})\right) = \sup_Q \varphi(Q).$$

where the parameters of the model are given in (2.17) (2.26) and (2.27) and (2.28).

2.5.2. *Rademacher Prior.* We now consider the case when P_0 takes values in $\{-1, 0, 1\}$ such that for $p \in [0, \frac{1}{2}]$

$$P_0(-1) = P_0(1) = p \qquad P_0(0) = 1 - 2p. \tag{2.29}$$

If $x \sim P_0$, then $\mathrm{Var}(x) = 2p$. We define the $n \times n$ matrices

$$\frac{1}{\Delta} = \left(\frac{1}{\Delta_{s,t}}\right)_{s,t\le n} \qquad \rho = \mathrm{diag}(\rho_1, \ldots, \rho_n).$$

In the degree corrected stochastic block model, the parameters were encoded by P_θ supported on finitely many points $(\theta^{(1)}, \ldots, \theta^{(n)}) \in (0, 1)$ with probabilities $\rho_1, \ldots, \rho_n$,

$$\frac{1}{\Delta_{st}} = \left(\frac{\lambda^2}{\theta^{(s)}\theta^{(t)}} + \frac{\lambda^2}{(1-\theta^{(s)}\theta^{(t)})}\right) \tag{2.30}$$

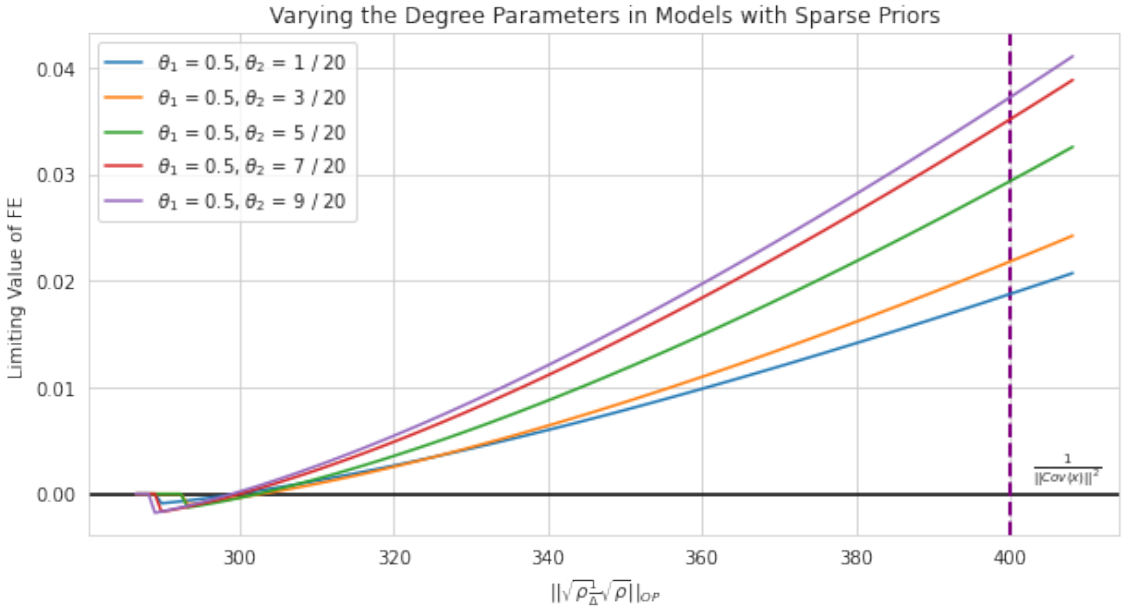

FIGURE 5. Free energy as the function of noise for the DCSBM with a small size of one of the groups. We see a discontinuous phase transition (purple line).

and for i.i.d. $\theta_i \sim P_\theta$

$$P_{\text{out}}(Y_{ij} = 1 \mid x, \boldsymbol{\theta}) = \theta_i\theta_j + \frac{\lambda}{\sqrt{N}} x_i \cdot x_j \quad \text{and} \quad P_{\text{out}}(Y_{ij} = 0 \mid x^0, \boldsymbol{\theta}) = 1 - \theta_i\theta_j - \frac{\lambda}{\sqrt{N}} x_i \cdot x_j. \tag{2.31}$$

Recall that the replica symmetric functional is

$$\varphi(Q) = -\sum_{s,t=1}^{n} \frac{\rho_s\rho_t}{4\Delta_{s,t}} \operatorname{Tr}(Q_s Q_t) + \sum_{s=1}^{n} \rho_s \mathbb{E}_{z,x^0} \ln \left[\int \exp\left(\left(\tilde{Q}^s x^0 + \sqrt{\tilde{Q}^s} z \right)^{\mathsf{T}} x - \frac{x^{\mathsf{T}} \tilde{Q}^s x}{2} \right) dP_0^\kappa(x) \right] \tag{2.32}$$

where

$$\tilde{Q}^s = \sum_{t=1}^{n} \frac{1}{\Delta_{s,t}} \rho_t Q_t.$$

The replica free energy functional is identical to the one computed by a spiked matrix with covariance profile given by (2.30). By Lemma 6.4 and Lemma 6.5, we can conclude that

(1) φ has a unique maximum at a sequence of matrices with all entries equal to 0 when

$$\left\| \sqrt{\rho} \frac{1}{\boldsymbol{\Delta}} \sqrt{\rho} \right\|_{op} < \frac{1}{9}$$

(2) φ has a maximum at a sequence of matrices where at least one matrix has a non-zero entry when

$$\left\| \sqrt{\rho} \frac{1}{\boldsymbol{\Delta}} \sqrt{\rho} \right\|_{op} > \frac{1}{4p^2}.$$

Since $p \leq \frac{1}{2}$, this transition is clearly not tight.

We can numerically solve the free energies to see how the phase transitions depends on the inhomogeneity. In the $n = 2$ case, a perfectly homogeneous model is when all entries of $\frac{1}{\Delta}$ are identical the phase transitions agree with known results [31]. We analyze the effects of changing the homogeneity on the locations of the maximizers. To do this, we can fix θ_1 and θ_2 and vary the parameter λ. If we consider the sparse case when $p = 0.025$ in (2.29), we plot in Figure 5 the limit of the free energy versus the operator norm of $\|\sqrt{\rho}\frac{1}{\boldsymbol{\Delta}}\sqrt{\rho}\|_{op}$. In contrast, we consider the dense case when $p = 0.25$ in (2.29) and plot again limit of the free energy versus the operator norm of $\|\sqrt{\rho}\frac{1}{\boldsymbol{\Delta}}\sqrt{\rho}\|_{op}$ in Figure 6. The second bound appears to be tight when the average degree encoded by p is high.

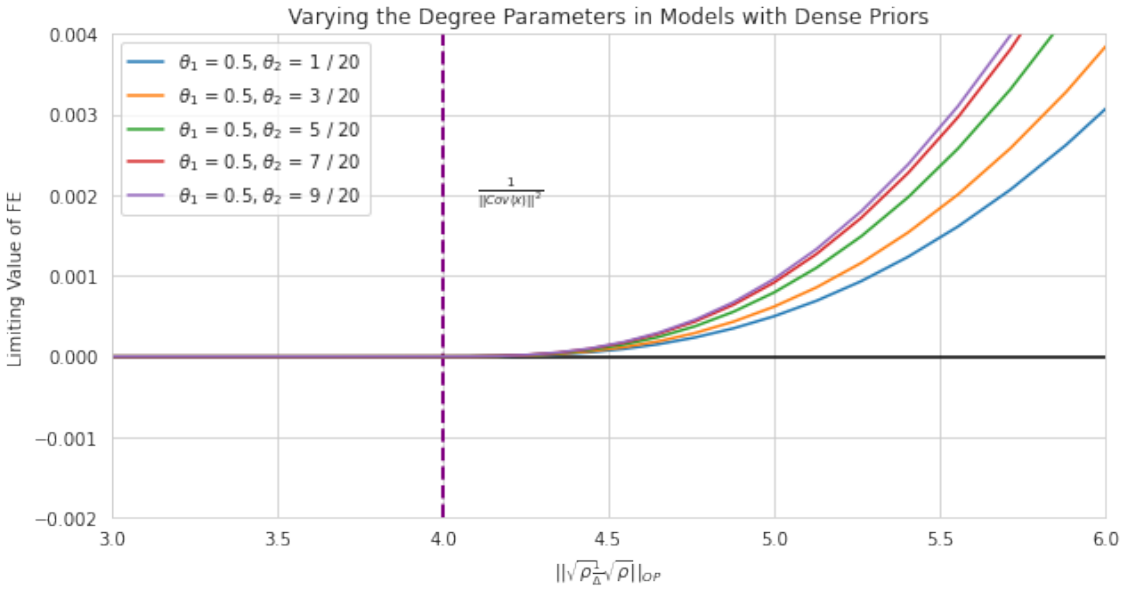

FIGURE 6. Free energy as the function of noise for the DCSBM with comparable sizes of one of the groups. We see a continuous phase transition (purple line).

2.6. **Outline of the Paper.** We first start with a proof of the universality result explained in Lemma 2.7. In Section 3, we move onto proving the limit of the free energy in Theorem 2.10 by proving the lower bound using interpolation in Section 4. We prove the matching upper bound using concentration of the overlaps and cavity computations in Section 5. In Section 6, we study the maximizers of the free energy, and explore its consequences on the information theoretic thresholds.

3. Universality in Inhomogenous Vector Spin Models

In this section, we proceed to several approximations to arrive to Theorem 2.7. We start by showing only the second order Taylor expansion of g matters in the computation of the free energy (recall the definition of $F_N(g)$ in (2.3)).

Lemma 3.1 (*Independence of Third Order Expansions*)

If $\sup_{i,j} \|\partial_w^3 g_{ij}\|_\infty < \infty$, *then*

$$F_N(g) = F_N(\tilde{g}) + O\Big(\frac{\kappa^3}{\sqrt{N}}\Big)$$

where

$$\tilde{g}_{ij}(D,w) = g_{ij}(D_{ij},0) + \partial_w g_{ij}(D,0)w + \frac{1}{2}\partial_w^2 g_{ij}(D,0)w^2.$$

Proof. By Taylor's theorem, for all i,j,

$$g_{ij}(D_{ij},w_{ij}) - g_{ij}(D_{ij},0) = \partial_w g_{ij}(D_{ij},0)w_{ij} + \frac{1}{2}\partial_w^2 g_{ij}(D_{ij},0)w_{ij}^2 + \frac{w_{ij}^3}{3!}\partial_w^3 g_{ij}(D_{ij},\theta_{ij}w_{ij})$$

for some $\theta_{ij} \in [0,1]$. Since our hypothesis implies that $|w_{ij}|_\infty \le C^2\kappa/\sqrt{N}$, our assumption that $\sup_{i,j}\|\partial_w^3 g_{ij}\|_\infty < \infty$ implies

$$\Big\| \frac{1}{N}\sum_{i<j} \frac{w_{ij}^3}{3!} g_w^{(3)}(Y_{ij},\theta_{ij}w_{ij}) \Big\| = O\Big(\frac{\kappa^3}{N^{1/2}}\Big)$$

from which the result follows. □

The next step in the reduction is to prove that the coefficient of the second derivative term can be replaced by its conditional average. We let B be a sub σ-algebra of $\mathbb{R}^{N^2} \times (\mathbb{R}^\kappa)^N$ and denote by $\mathbb{P}_B$ (resp. $\mathbb{E}_Y[.|B]$) its

associated conditional probability under P_Y (resp. its conditional expectation). Later on we will simply take B to be the σ-algebra generated by x^0, and therefore P_B will just be the distribution of Y knowing the x^0.

Lemma 3.2 (*Concentration of Second Order Terms*)

Assume $\sup_{i,j} \|\partial_w^2 g_{ij}(\cdot,0)\|_\infty < \infty$ *and* P_0 *compactly supported. Let* B *be a* σ *algebra such that the* Y_{ij} *are independent conditionally to* B*. Then*

$$F_N(\tilde{g}) = F_N(\bar{g}) + O\Big(\frac{\kappa^2}{\sqrt{N}}\Big)$$

with

$$\bar{g}_{ij}(D,w) = g_{ij}(D_{ij},0) + \partial_w g_{ij}(D,0)w + \frac{1}{2}\mathbb{E}_D[\partial_w^2 g_{ij}(D,0) \mid B]w^2\,.$$

Proof. Notice that

$$F_N(\tilde{g}) - F_N(\bar{g}) = \mathbb{E}_D \frac{1}{N} \ln \Big\langle e^{\frac{1}{2\sqrt{N}} \sum_{i<j} \frac{1}{\sqrt{N}} (\partial_w^2 g_{ij}(D_{ij},0) - \mathbb{E}_D[\partial_w^2 g_{ij}(D_{ij},0)|B])(x_i^\mathsf{T} x_j)^2)} \Big\rangle$$

where

$$\langle f \rangle = \frac{\int f e^{\sum_{i<j} \bar{g}_{ij}(D_{ij},w_{ij})} dP_0^{\otimes N}(x)}{\int e^{\sum_{i<j} \bar{g}_{ij}(D_{ij},w_{ij})} dP_0^{\otimes N}(x)}\,.$$

Let Z be the $N \times N$ symmetric matrix with entries so that $Z_{ii} = 0$ and for $i \neq j$

$$Z_{ij} = \frac{1}{4\sqrt{N}} (\partial_w^2 g_{ij}(\Sigma_{ij}^{-1} D_{ij}, 0) - \mathbb{E}_D[\partial_w^2 g_{ij}(\Sigma_{ij}^{-1} D_{ij}, 0)|B])\,.$$

As a consequence

$$\sum_{i<j} \frac{1}{2\sqrt{N}} (g_{ww}(D_{ij},0) - \mathbb{E}_D[g_{ww}(D_{ij},0)|B])(x_i^\mathsf{T} x_j)^2 = \mathrm{Tr}\left(Z(x^\mathsf{T} x)^2\right).$$

Z is a random symmetric matrix under P_B which has centered independent entries with covariance bounded by C/N and where $(x^\mathsf{T} x)^2$ is the matrix with entries $(x_i^\mathsf{T} x_j)^2$. Because $\sqrt{N}Z$ has bounded entries, we can use concentration inequalities due to Talagrand (see [4, Theorem 2.3.5] and [24, Lemma 5.6]) to see that there exists some finite L_0 such that, uniformly,

$$\mathbb{P}_B\left(\|Z\|_\infty \geq L\right) \leq e^{-N(L-L_0)}\,. \tag{3.1}$$

On $\{\|Z\|_\infty \leq L\}$, we have the bound

$$\left|\mathrm{Tr}\big(Z(x^\mathsf{T} x)^2\big)\right| = \left|\sum_{k,k'=1}^{\kappa} \sum_{i,j} Z_{ij} x_i(k) x_i(k') x_j(k) x_j(k')\right| \leq L \sum_{k,k'=1}^{\kappa} \sum_{i=1}^{N} (x_i(k) x_i(k'))^2 \leq CL\kappa^2 N$$

for some finite constant C depending only on the bound on the support of P_0. Hence, we deduce

$$F_N(\tilde{g}) - F_N(\bar{g}) = \mathbb{E}_D 1_{\|Z\|\geq L} \frac{1}{N} \ln \langle e^{\frac{1}{\sqrt{N}} \sum_{i\leq j} \frac{1}{2\sqrt{N}} (\partial_w^2 g_{ij}(Y_{ij},0) - \mathbb{E}_D[\partial_w^2 g_{ij}(D_{ij},0)|B])(x_i^\mathsf{T} x_j)^2)} \rangle + O\Big(\frac{\kappa^2}{\sqrt{N}}\Big).$$

Moreover as we assumed that $\partial_w^2 g_{ij}(D,0)$ is uniformly bounded, the term in the above expectation is uniformly bounded and therefore the first term is going to zero exponentially fast by (3.1) for L large enough. □

We finally compare our free energy to those of a spin glass model. It will depend on three matrices with entries:

$$\gamma_{ij} = \mathbb{E}_D[\partial_w^2 g_{ij}(D_{ij},0) \mid x^0], \quad \mu_{ij} = \mathbb{E}_D[\partial_w g_{ij}(D_{ij},0) \mid x^0], \quad \sigma_{ij}^2 = \mathbb{E}_D[(\partial_w g_{ij}(D_{ij},0) - \mu_{ij})^2 \mid x^0]. \tag{3.2}$$

By universality, we will prove that we can replace $\partial_w g_{ij}(D_{ij},0)$ by $\sigma_{ij} W_{ij} + \mu_{ij}$ where W_{ij} are i.i.d. standard Gaussian variables (under the assumption that $\sqrt{N}\mu_{ij} = O(1)$). Let

$$F_N(\sigma,\mu,\gamma) = \mathbb{E}_{W,x^0}\left[\frac{1}{N} \ln \mathbb{E}_x\left[\exp(H_N(x))\right]\right]$$

with

$$H_N(x) = \left(\sqrt{\frac{1}{N}}\sum_{i<j}\sigma_{ij}W_{ij}x_i^\mathsf{T}x_j + \mu_{ij}x_i^\mathsf{T}x_j\right) + \frac{1}{2N}\sum_{i<j}\gamma_{ij}(x_i^\mathsf{T}x_j)^2$$

$$= \sum_{k=1}^{\kappa}\left(\sqrt{\frac{1}{N}}\sum_{i<j}\sigma_{ij}W_{ij}x_i(k)x_j(k) + \mu_{ij}x_i(k)x_j(k)\right) + \frac{1}{2N}\sum_{k,\ell=1}^{\kappa}\sum_{i<j}\gamma_{ij}x_i(k)x_j(k)x_i(\ell)x_j(\ell). \tag{3.3}$$

Then we shall prove that :

Lemma 3.3 (*Universality in Disorder*)

Assume that

$$\sup_{i,j}\|\mu_{ij}\|_\infty = O(N^{-1/2}),\ \sup_{ij}\|\sigma_{ij}^2\|_\infty < \infty,\ \sup_{i,j}\left\|\frac{\mathbb{E}_D[|\partial_w g_{ij}(D_{ij},0)-\mu_{ij}|^3|x^0]}{\sigma_{ij}^3}\right\|_\infty < \infty$$

then

$$F_N(\bar{g}) = F_N(\sigma,\mu,\gamma) + O\Big(\frac{\kappa^3}{N^{1/2}}\Big)$$

where $F_N(\sigma,\mu,\gamma)$ is the free energy with respect to the Hamiltonian defined in (3.3).

The proof follows from the following approximate integration by parts lemma [34, Lemma 3,7].

Lemma 3.4

Suppose x is a random variable that satisfies $\mathbb{E}x = 0$, $\mathbb{E}|x|^3 < \infty$. If $f : \mathbb{R} \to \mathbb{R}$ is twice continuously differentiable and $\|f''\|_\infty < \infty$, then

$$|\mathbb{E}xf(x) - \mathbb{E}x^2\mathbb{E}f'(x)| \le \frac{3}{2}\|f''\|_\infty\mathbb{E}|x|^3.$$

Proof. We follow the proof of Carmona–Hu [13] presented in [34, Theorem 3.9]. To compare the free energies $F_N(\bar{g})$ and $F_N(\sigma,\mu,\kappa)$ we use an interpolation argument. Conditionally on x^0, consider the interpolating Hamiltonian defined for $t \in [0,1]$ by

$$H_N(x,t) = \frac{1}{\sqrt{N}}\sum_{i<j}\Big(\sqrt{t}(\partial_{ij}g(D_{ij},0)-\mu_{ij}) + \sqrt{1-t}\sigma_{ij}W_{ij}\Big)x_i^\mathsf{T}x_j + \frac{1}{\sqrt{N}}\sum_{i<j}\mu_{ij}x_i^\mathsf{T}x_j + \frac{1}{2N}\sum_{i<j}\gamma_{ij}(x_i^\mathsf{T}x_j)^2$$

$$= \frac{1}{\sqrt{N}}\sum_{i<j}\sigma_{ij}\Big(\sqrt{t}\tilde{W}_{ij} + \sqrt{1-t}W_{ij}\Big)x_i^\mathsf{T}x_j + \frac{1}{\sqrt{N}}\sum_{i<j}\mu_{ij}x_i^\mathsf{T}x_j + \frac{1}{2N}\sum_{i<j}\gamma_{ij}(x_i^\mathsf{T}x_j)^2$$

where we defined $\tilde{W}_{ij} = \sigma_{ij}^{-1}(\partial_w g_{ij}(D_{ij},0)-\mu_{ij})$ to simplify notation. Notice that

$$\mathbb{E}_D[\tilde{W}_{ij}^2 \mid x^0] = \sigma_{ij}^{-2}\mathbb{E}_D[(\partial_w g_{ij}(D_{ij},0)-\mu_{ij})^2|x^0] = 1$$

and

$$\mathbb{E}_D[\tilde{W}_{ij} \mid x^0] = \sigma_{ij}^{-1}\mathbb{E}_D[(\partial_w g_{ij}(D_{ij},0)-\mu_{ij}) \mid x^0] = 0$$

so both W and $\tilde{W}$ have mean zero and variance 1. Also $\mathbb{E}_D|\tilde{W}_{ij}^3| \le \frac{1}{\sigma_{ij}^3}(\|\partial_w g\|_\infty)^3$ is uniformly bounded.

We define the interpolating free energy

$$\varphi(t) = \frac{1}{N}\mathbb{E}_D[\ln\mathbb{E}_X\exp(H_N(x,t))|x^0], \langle f\rangle_t = \frac{\int f(x)\exp(H_N(x,t))dP_0(x)}{\int\exp(H_N(x,t))dP_0(x)},$$

and notice that

$$\varphi'(t) = \mathbb{E}_Y\left[\frac{1}{2\sqrt{t}N^{3/2}}\sum_{i<j}\sigma_{ij}\tilde{W}_{ij}\langle x_i^\mathsf{T}x_j\rangle_t - \frac{1}{2\sqrt{1-t}N^{3/2}}\sum_{i<j}\sigma_{ij}W_{ij}\langle x_i^\mathsf{T}x_j\rangle_t \,\middle|\, x^0\right].$$

Let $f(\tilde{W}_{ij}) = \langle x_i^\mathsf{T}x_j\rangle_t$ (the dependence on $\tilde{W}$ is in the numerator and denominator in the Gibbs measure). We find that

$$\frac{\partial f}{\partial\tilde{W}_{ij}} = \frac{\sqrt{t}\sigma_{ij}}{\sqrt{N}}\Big(\langle(x_i^1\cdot x_j^1)^2\rangle_t - \langle(x_i^1\cdot x_j^1)(x_i^2\cdot x_j^2)\rangle_t\Big)$$

and

$$\frac{\partial^2 f}{\partial \tilde{W}_{ij}^2} = \frac{t\sigma_{ij}^2}{N}\Big(\langle (x_i^1 \cdot x_j^1)^3\rangle_t - 2\langle (x_i^1 \cdot x_j^1)^2 (x_i^2 \cdot x_j^2)\rangle_t - \langle (x_i^1 \cdot x_j^1)(x_i^2 \cdot x_j^2)^2\rangle_t + 2\langle (x_i^1 \cdot x_j^1)(x_i^2 \cdot x_j^2)(x_i^3 \cdot x_j^3)\rangle_t\Big).$$

Therefore the second derivative is bounded by

$$\|\frac{\partial^2 f}{\partial \tilde{W}_{ij}^2}\| \le \frac{\sup_{ij} \|\mathbb{E}_D[|\partial_w g_{ij}(D_{ij},0) - \mu_{ij}|^2 | x^0]\|_\infty 6C^6\kappa^3}{N}$$

where C is such that $x \in [-C,C]^\kappa$ almost surely. Applying the approximate integration by parts lemma to $\tilde{W}$ stated in Lemma 3.4 applied conditionally on x^0 implies

$$\begin{aligned}&\Big|\mathbb{E}_D[\frac{1}{2\sqrt{t}N^{3/2}}\sigma_{ij}\tilde{W}_{ij}\langle x_i^\mathsf{T} x_j\rangle_t \mid x^0] - \frac{\sigma_{ij}^2}{2N^2}\Big(\mathbb{E}_D[\langle (x_i^1 \cdot x_j^1)^2\rangle_t \mid x^0] - \mathbb{E}[\langle (x_i^1 \cdot x_j^1)(x_i^2 \cdot x_j^2)\rangle_t \mid x^0]\Big)\Big| \\ &\le \frac{\sup_{ij} \|\mathbb{E}_D[|\partial_w g_{ij}(D_{ij},0) - \mu_{ij}|^2 | x^0]\|_\infty 6C^6\kappa^3}{N} \cdot \frac{3\sup_{i,j}\mathbb{E}[|\tilde{W}_{i,j}|^3 \mid x^0]}{4N^{3/2}} = O\Big(\frac{\kappa^3}{N^{5/2}}\Big)\end{aligned}$$

by our assumption on the uniform bounds on the conditional expectation of $\tilde{W}$. The classical integration by parts lemma for Gaussian variables implies

$$\mathbb{E}_W\Big[\frac{1}{2\sqrt{1-t}N^{3/2}}\sigma_{ij}W_{ij}\langle x_i^\mathsf{T} x_j\rangle_t\Big] = \frac{\sigma_{ij}^2}{2N^2}\mathbb{E}_W\Big[\Big(\langle (x_i^1 \cdot x_j^1)^2\rangle_t - \langle (x_i^1 \cdot x_j^1)(x_i^2 \cdot x_j^2)\rangle_t\Big)\Big].$$

Summing over $i<j$ gives us the bound

$$|\varphi'(t)| \le O\Big(\frac{\kappa^3}{N^{1/2}}\Big)$$

so that

$$|\varphi(1) - \varphi(0)| = |F_N(\bar{g}) - F_N(\sigma,\mu,\kappa)| \le O\Big(\frac{\kappa^3}{N^{1/2}}\Big).$$

□

Remark 3.5. We use the notation $x_i^1, x_i^2, \dots$ to denote the replicas. They are independent copies of x under the corresponding Gibbs measures.

We can now further reduce our problem to a Gaussian estimation problem under the Bayes optimal assumption of Hypothesis 2.1. Indeed, recalling the definition of the Fisher score matrix (2.5), it results in special relation between σ, μ and γ which state as follows.

Lemma 3.6

Assume Hypotheses 2.1 and 2.3. Then for all $i,j \in [N]$ the terms defined in (3.2) satisfy

$$\mu_{ij} = \frac{x_i^0 \cdot x_j^0}{\Delta_{ij}\sqrt{N}} + O(\kappa^2 N^{-1}), \quad \sigma_{ij}^2 = -\gamma_{ij} + O(\kappa N^{-1/2}), \quad \gamma_{ij} = -\frac{1}{\Delta_{ij}} + O(\kappa N^{-1/2})$$

Proof. The fact that $g_{ij}(D,w)$ is a log-likelihood is very important in this section, and the results in this final decomposition do not apply for general g. Firstly, using the fact that $P_{ij}(\cdot|w)$ is a probability measure for all w and $i<j$,

$$\int P_{ij}(y|w)\,dy = \int e^{g_{ij}(y,w)}dy = 1. \tag{3.4}$$

By differentiating (3.4), it follows that for all w,

$$\mathbb{E}_{P_{ij}(D|w)}\partial_w g_{ij}(D,w) = \partial_w \int e^{g_{ij}(y,w)}\,dy = 0 \tag{3.5}$$

and

$$\mathbb{E}_{P_{ij}(D|w)}\Big((\partial_w g_{ij}(D,w))^2 + \partial_w^2 g_{ij}(D,w)\Big) = \partial_w^2 \int e^{g_{ij}(y,w)}\,dy = 0. \tag{3.6}$$

The conditional laws of D_{ij} given the signal x^0 are independent, so we have the following reductions of the parameters μ_{ij} and σ_{ij}. By (3.5) we get

$$\begin{aligned}\mu_{ij} &= \int \partial_w g_{ij}(y,0) e^{g_{ij}(y,w_{ij}^0)} dy \\ &= \int (\partial_w g_{ij}(y,0) - \partial_w g_{ij}(y,w_{ij}^0)) e^{g_{ij}(y,w_{ij}^0)} dy \\ &= -w_{ij}^0 \int \partial_w^2 g_{ij}(y,0) e^{g_{ij}(y,0)} dy + O((w_{ij}^0)^2) \\ &= -\Big(\frac{x_i^0 \cdot x_j^0}{\sqrt{N}}\Big) \mathbb{E}_{P_{ij}(D|w=0)} \partial_w^2 g_{ij}(D,0) + O(\kappa^2 N^{-1})\end{aligned}$$

where we applied Taylor's theorem in the fourth equality. Applying (3.6) one more time implies that

$$\mu_{ij} = \frac{x_i^0 \cdot x_j^0}{\sqrt{N}} \mathbb{E}_{P_{ij}(D|w=0)} (\partial_w g_{ij}(D,0))^2 + O(\kappa^2 N^{-1}) = \frac{x_i^0 \cdot x_j^0}{\Delta_{ij}\sqrt{N}} + O(\kappa^2 N^{-1})$$

where we recall that (2.5)

$$\frac{1}{\Delta_{ij}} = \mathbb{E}_{P_{ij}(D|w=0)} (\partial_w g_{ij}(D,0))^2.$$

Similarly, (3.6) implies

$$\begin{aligned}\sigma_{ij}^2 &= \int (\partial_w g_{ij}(y,0))^2 e^{g(y,w_{ij}^0)} dy - \mu_{ij}^2 \\ &= \int (\partial_w g_{ij}(y,0))^2 - (\partial_w g_{ij}(y,w_{ij}^0))^2) e^{g_{ij}(y,w_{ij}^0)} dy - \int \partial_w^2 g_{ij}(y,w_{ij}^0) e^{g_{ij}(y,w_{ij}^0)} dy - \mu_{ij}^2 \\ &= -\gamma_{ij} + O(w_{ij}) = -\gamma_{ij} + O(\kappa N^{-1/2}).\end{aligned}$$

We can do this trick one more time and apply (3.6) to see that

$$\begin{aligned}\gamma_{ij} &= \int \partial_w^2 g_{ij}(y,0) e^{g_{ij}(y,w_{ij}^0)} dy \\ &= -w_{ij} \int \partial_w^3 g_{ij}(y,0) e^{g_{ij}(y,w_{ij}^0)} dy - \int ((\partial_w g_w(y,w))^2 - (\partial_w g_{ij}(y,0))^2) e^{g_{ij}(y,0)} dy - \frac{1}{\Delta_{ij}} + O(\kappa N^{-1/2}) \\ &= -\frac{1}{\Delta_{ij}} + O(\kappa N^{-1/2}).\end{aligned}$$

□

With this in mind, an interpolation argument and Gaussian integration by parts will prove that the Hamiltonian associated with the free energy $F(\sigma,\mu,\gamma)$ in the Bayes optimal case

$$H_{N,\sigma,\mu,\gamma}(x) = \sum_{i<j} \frac{\sigma_{ij} W_{ij}}{\sqrt{N}} (x_i \cdot x_j) + \frac{\mu_{ij}}{\sqrt{N}} (x_i \cdot x_j) + \frac{1}{2N} \gamma_{ij} (x_i \cdot x_j)^2$$

can be replaced with the following Hamiltonian

$$H_{N,\Delta}(x) = \sum_{i<j} \frac{W_{ij}}{\sqrt{\Delta_{ij} N}} (x_i \cdot x_j) + \frac{(x_i^0 \cdot x_j^0)}{\Delta_{ij} N} (x_i \cdot x_j) - \frac{1}{2\Delta_{ij} N} (x_i \cdot x_j)^2$$

without changing the limit of the free energy. We let

$$F_N(\Delta) = \frac{1}{N} \mathbb{E}_{W,x^0} \ln \mathbb{E}_X e^{H_{N,\Delta}(x)}. \tag{3.7}$$

The form of this Hamiltonian is identical to the free energy in the low rank matrix estimation with Hadamard covariance profile, which we introduced in Subsection 2.2.1.

Lemma 3.7 (*Reduction to Low Rank Hamiltonian*)

If (3.4) holds, then

$$F_N(\sigma,\mu,\kappa) = F_N(\Delta) + O(\kappa^3 N^{-1/2}).$$

Proof. Consider the interpolating Hamiltonian,

$$H_N(t,x) = \sum_{i<j} \frac{\sqrt{t}\sigma_{ij}W_{ij}}{\sqrt{N}}(x_i\cdot x_j) + \frac{t\mu_{ij}}{\sqrt{N}}(x_i\cdot x_j) + \frac{t}{2N}\kappa_{ij}(x_i\cdot x_j)^2 + \sum_{i<j}\frac{\sqrt{1-t}\tilde{W}_{ij}}{\sqrt{N\Delta_{ij}}}(x_i\cdot x_j) + \frac{t(x_i^0\cdot x_j^0)}{\Delta_{ij}N}(x_i\cdot x_j) - \frac{t}{2\Delta_{ij}N}(x_i\cdot x_j)^2.$$

where W and $\tilde{W}$ are independent standard Gaussians. If we define

$$\varphi(t) = \frac{1}{N}\mathbb{E}_{W,\tilde{W},x^0}\ln \mathbb{E}_x e^{H_N(t,x)}$$

then

$$\begin{aligned}\varphi'(t) &= \frac{1}{N}\mathbb{E}\Big\langle \partial_t H_N(t,x)\Big\rangle_t \\ &= \frac{1}{N}\mathbb{E}\Big(\sum_{i<j}\frac{\sigma_{ij}W_{ij}}{2\sqrt{t}\sqrt{N}}\langle x_i\cdot x_j\rangle_t + \frac{\mu_{ij}}{\sqrt{N}}\langle x_i\cdot x_j\rangle_t + \frac{1}{2N}\kappa_{ij}\langle (x_i\cdot x_j)^2\rangle_t\Big) \\ &\quad - \frac{1}{N}\mathbb{E}\Big(\sum_{i<j}\frac{\tilde{W}_{ij}}{2\sqrt{1-t}\sqrt{\Delta_{ij}N}}\langle x_i\cdot x_j\rangle_t + \frac{(x_i^0\cdot x_j^0)}{\Delta_{ij}N}\langle x_i\cdot x_j\rangle_t - \frac{1}{2\Delta_{ij}N}\langle (x_i\cdot x_j)^2\rangle_t\Big)\end{aligned}$$

where $\langle\cdot\rangle_t$ is the average with respect to the Gibbs measure $G_t \propto e^{H_N(t,x)}$. Recall that (3.5) and (3.6) imply

$$\mu_{ij} = \frac{x_i^0\cdot x_j^0}{\sqrt{\Delta_{ij}N}} + O(\kappa^2 N^{-1}) \quad \text{and} \quad \sigma_{ij}^2 = -\gamma_{ij} + O(\kappa N^{-1/2}) = \frac{1}{\Delta_{ij}} + O(\kappa N^{-1/2})$$

This implies that the terms without the Gaussian $W, \tilde{W}$ cancel each other. For the first terms of each expectation, we integrate by parts to find

$$\frac{1}{N}\mathbb{E}\sum_{i<j}\frac{\sigma_{ij}W_{ij}}{2\sqrt{t}\sqrt{N}}\langle x_i\cdot x_j\rangle_t = \frac{1}{2N^2}\sum_{i<j}\sigma_{ij}^2\big(\langle (x_i^1\cdot x_j^1)^2\rangle_t - \langle (x_i^1\cdot x_j^1)(x_i^2\cdot x_j^2)\rangle_t\big)$$

and

$$\frac{1}{N}\mathbb{E}\sum_{i<j}\frac{\tilde{W}_{ij}}{2\sqrt{1-t}\sqrt{\Delta_{ij}N}}\langle x_i\cdot x_j\rangle_t = \frac{1}{2\Delta_{ij}N^2}\sum_{i<j}\big(\langle (x_i^1\cdot x_j^1)^2\rangle_t - \langle (x_i^1\cdot x_j^1)(x_i^2\cdot x_j^2)\rangle_t\big)$$

so the difference of the Gaussian terms are also of order $O(\kappa N^{-1/2})$. Therefore,

$$|\varphi'(t)| = O(\kappa N^{-1/2}),$$

which completes the proof after integrating φ on $[0,1]$. □

Remark 3.8. Notice that $\mu_{ij} = O(\frac{1}{\sqrt{N}})$, so one of the hypothesis in Lemma 3.3 is automatically satisfied.

Remark 3.9. In the spin glass setting,

$$\int e^{g(D,w)}\,dy = \int e^{Dw}\,dy$$

does not equal to 1, so (3.4) is not satisfied and Lemma 3.7 doesn't apply.

The proof of the main universality theorem is now immediate.

Proof of Lemma 2.7. The result follows by combining Lemma 3.1, Lemma 3.2, Lemma 3.3, and Lemma 3.7. □

Lastly, the computations will be much simpler if $P_0(x^0)$ was supported on finitely many points. We will now prove that general P_0 with bounded support can be approximated by P_0 with finite support. Suppose that P_0 is supported on $[-C,C]^\kappa$. We can discretize the support in K blocks by defining the function $f : [-C,C]^\kappa \to [-C,C]^\kappa$ by

$$f(x_1,\dots,x_\kappa) = \Big(\frac{C}{K}\Big\lfloor\frac{x_1K}{C}\Big\rfloor,\dots,\frac{C}{K}\Big\lfloor\frac{x_\kappa K}{C}\Big\rfloor\Big).$$

Notice that each coordinate of $f(x)$ is supported on $2K$ points. We define $P_d = P\circ f^{-1}$, and set

$$F_N(\Delta) = \mathbb{E}_{W,x^0\sim P_0}\Big[\frac{1}{N}\ln\mathbb{E}_{x\sim P_0}\left[\exp\{H_{N,\Delta}(x)\}\right]\Big] \text{ and } F_N^d(\Delta) = \mathbb{E}_{W,x^0\sim P_d}\Big[\frac{1}{N}\ln\mathbb{E}_{x\sim P_d}\left[\exp\{H_{N,\Delta}(x)\}\right]\Big].$$

Lemma 3.10 (*Approximations of References Measures with Finite Support*)

For any $N \geq 1$, we have

$$|F_N(\Delta) - F_N^d(\Delta)| \leq \frac{2C^4\kappa^2}{\|\Delta\|_\infty K}.$$

Proof. To simplify notation, we define $\tilde{x}_i = f(x_i)$ to be the discretization of x_i and $F_N(\Delta) = F_N(P_0)$. Consider the interpolating Hamiltonian

$$\begin{aligned} H_N(x;t) = &\sum_{i<j} \frac{\sqrt{t}W_{ij}}{\sqrt{\Delta_{ij}N}}(x_i \cdot x_j) + \frac{t(x_i^0 \cdot x_j^0)}{\Delta_{ij}N}(x_i \cdot x_j) - \frac{t}{2\Delta_{ij}N}(x_i \cdot x_j)^2 \\ &+ \sum_{i<j} \frac{\sqrt{1-t}\tilde{W}_{ij}}{\Delta_{ij}\sqrt{N}}(\tilde{x}_i \cdot \tilde{x}_j) + \frac{(1-t)(\tilde{x}_i^0 \cdot \tilde{x}_j^0)}{\Delta_{ij}N}(\tilde{x}_i \cdot \tilde{x}_j) - \frac{(1-t)}{2\Delta_{ij}N}(\tilde{x}_i \cdot \tilde{x}_j)^2 \end{aligned}$$

where $\tilde{W}_{ij}$ and W_{ij} are independent standard Gaussians. We consider the usual interpolation Hamtiltonian

$$\varphi(t) = \frac{1}{N}\mathbb{E}\log \mathbb{E}_x e^{H_N(x;t)}.$$

From integration by parts and the Nishimori property (see equation 4.4) to simplify the Gaussian terms, we see that

$$\varphi'(t) = \sum_{i<j} \frac{1}{2\Delta_{ij}N^2}\Big(\mathbb{E}\langle (x_i^0 \cdot x_j^0)(x_i \cdot x_j)\rangle - \mathbb{E}\langle (\tilde{x}_i^0 \cdot \tilde{x}_j^0)(\tilde{x}_i \cdot \tilde{x}_j)\rangle\Big).$$

Since the $\frac{1}{\Delta_{ij}^2}$ are uniformly bounded, we have the upper bound

$$\begin{aligned} |\varphi'(t)| &= \frac{1}{4\|\Delta\|_\infty}\Big|\mathbb{E}\langle (x_i^0 \cdot x_j^0)(x_i \cdot x_j) - (\tilde{x}_i^0 \cdot \tilde{x}_j^0)(\tilde{x}_i \cdot \tilde{x}_j)\rangle\Big| \\ &\leq \frac{1}{4\|\Delta\|_\infty}\Big|\mathbb{E}\langle (x_i \cdot x_j - \tilde{x}_i \cdot \tilde{x}_j)(x_i^0 \cdot x_j^0)\rangle\Big| + \frac{1}{4\|\Delta\|_\infty}\Big|\mathbb{E}\langle (x_i^0 \cdot x_j^0 - \tilde{x}_i^0 \cdot \tilde{x}_j^0)(\tilde{x}_i \cdot \tilde{x}_j)\rangle\Big| \\ &\leq \frac{1}{4\|\Delta\|_\infty}\Big(\big(\mathbb{E}\langle (x_i \cdot x_j - \tilde{x}_i \cdot \tilde{x}_j)^2\rangle \mathbb{E}\langle (x_i^0 \cdot x_j^0)^2\rangle\big)^{1/2} + \big(\mathbb{E}\langle (x_i^0 \cdot x_j^0 - \tilde{x}_i^0 \cdot \tilde{x}_j^0)^2\rangle \mathbb{E}\langle (\tilde{x}_i \cdot \tilde{x}_j)^2\rangle\big)^{1/2}\Big) \\ &\leq \frac{C^2\kappa}{\|\Delta\|_\infty}\Big(\mathbb{E}\langle (x_i \cdot x_j - \tilde{x}_i \cdot \tilde{x}_j)^2\rangle\Big)^{1/2} \\ &\leq \frac{C^2\kappa}{\|\Delta\|_\infty}\Big(\mathbb{E}\langle (x_i \cdot (x_j - \tilde{x}_j) + \tilde{x}_j \cdot (x_i - \tilde{x}_i))^2\rangle\Big)^{1/2}. \end{aligned}$$

Since $\|x_i - \tilde{x}_i\|_\infty \leq \frac{C}{K}$ and $\|x_j - \tilde{x}_j\|_\infty \leq \frac{C}{K}$ and P_0 has compact support, we get the rough bound

$$|\varphi'(t)| \leq \frac{2C^4\kappa^2}{\|\Delta\|_\infty K} \implies |\varphi(1) - \varphi(0)| \leq \frac{2C^4\kappa^2}{\|\Delta\|_\infty K},$$

so the statement follows since $\varphi(1) = F_N(P_0)$ and $\varphi(0) = F_N(P_d)$. □

Remark 3.11. We can also modify this proof so that it holds as long the probability is compactly supported if we assume that the tails do not grow too much. We just have to be more careful since instead of using the uniform bound on the measure, we can get a bound in terms of $\mathbb{E}\|x\|_2^2$ like in the proof in [29]. The same argument (interpolating among the Δ's rather than the x) shows how to deduce Theorem 2.13 from Theorem 2.10.

Therefore, without loss of generality, we may assume that P_0 is supported on finitely many points, because we can always approximate general P_0 with compact support up to arbitrary accuracy with a probability measure supported on only finitely many points. Similarly we can assume that Δ is stepwise constant.

3.1. **Spectrum Universality.** In this section, we state a stronger form of universality than Lemma 2.7. Consider the transformed data matrix

$$\frac{\tilde{Y}_{ij}}{\sqrt{N}} = \frac{1}{\sqrt{N}}\partial_w g_{ij}(D_{ij}, 0) \qquad i, j \leq N$$

where $(D_{ij})_{i,j}$ are independent and distributed according to P_{ij} conditionally on $\boldsymbol{x}^0$ as in (2.1), and the normalized spiked matrix with variance profile $\frac{1}{\sqrt{N}}Y^\Delta$ defined in (2.10)

$$Y^\Delta = \Delta^{\odot\frac{1}{2}} \odot W + \frac{(x^0)(x^0)^T}{\sqrt{N}}.$$

In the Bayes optimal setting the first two moments of the matrix $\tilde{Y}$ are equivalent to those of $\frac{1}{\Delta} \odot Y^\Delta$ up to a $O(\frac{1}{\sqrt{N}})$ term. Define

$$\tilde{\mu}_{ij} = \mathbb{E}_Y[\tilde{Y}_{ij} \mid \boldsymbol{x}^0], \quad \tilde{\sigma}^2_{ij} = \mathbb{E}_Y[(\tilde{Y}_{ij} - \tilde{\mu}_{ij})^2 \mid \boldsymbol{x}^0].$$

and

$$\mu_{ij} = \mathbb{E}_Y\left[\frac{1}{\Delta} \odot Y^\Delta \,\middle|\, \boldsymbol{x}^0\right], \quad \sigma^2_{ij} = \mathbb{E}_Y\left[(\frac{1}{\Delta} \odot Y^\Delta - \mu_{ij})^2 \,\middle|\, \boldsymbol{x}^0\right]$$

By Lemma 3.6, it follows that

Corollary 3.12

Assume that $P_X \in \mathscr{P}(\mathbb{R}^k)$ has compact support. Then for all $i, j \in [N]$, we have

$$\tilde{\mu}_{ij} = \frac{x^0_i \cdot x^0_j}{\Delta_{ij}\sqrt{N}} + O(k^2 N^{-1}), \ \tilde{\sigma}^2_{ij} = \frac{1}{\Delta_{ij}} + O(kN^{-1/2})$$

and

$$\mu_{ij} = \frac{x^0_i \cdot x^0_j}{\Delta_{ij}\sqrt{N}}, \ \sigma^2_{ij} = \frac{1}{\Delta_{ij}}.$$

Furthermore, we assume that the Fisher information matrix (2.5) satisfies the following assumption

Hypothesis 3.13 (Quadratic Vector Equation Conditions)**.** Assume that there exists parameters $p, q, P > 0$ and $L \in \mathbb{N}$ such that

(1) For all N,

$$\frac{1}{\Delta_{ij}} \le q \qquad i, j \le N$$

(2) For all N,

$$\left(\frac{1}{N\Delta}\right)^L_{ij} \ge \frac{p}{N} \qquad i, j \le N.$$

(3) The unique solution $(m_i(z))_{i\le N}$ of vector of analytic functions on $\mathbb{C}^+ = \{\Im z > 0\}$ to the following quadratic vector equation,

$$-\frac{1}{m_i(z)} = z + \sum_{j=1}^N \frac{1}{\Delta_{ij}} m_j(z) \qquad \Im(z) > 0.$$

going to zero when $\Im z$ goes to infinity, satisfies

$$|m_i(z)| \le P, \qquad i, j \le N, \Im(z) > 0.$$

We have

Theorem 3.14 (*Universality of the Spectrum*)

If g satisfies Hypothesis 2.2 and the corresponding Fisher information matrix (2.5) satisfies Hypothesis 3.13, then

(1) Conditionally on x^0, the empirical distribution μ_1 of the eigenvalues of $\frac{\tilde{Y}_{ij}}{\sqrt{N}}$ and the empirical distribution μ_1 of the eigenvalues of $\frac{1}{\sqrt{N\Delta}} \odot Y^\Delta$ satisfy

$$\lim_{N\to\infty} d(\mu_1, \mu_2) \to 0 \tag{3.8}$$

in probability.

(2) Conditionally on x^0, when the dimension goes to infinity, $\frac{1}{\sqrt{N\Delta}} \odot Y^\Delta$ has an extremal eigenvalue away from the bulk iff $\frac{\tilde{Y}_{ij}}{\sqrt{N}}$ does, for almost all Δ and ρ.

Proof. We fix the realization of x^0 and assume that $x^0 \in \mathbb{R}^n$ to simplify notation (see Remark 3.15 for the generalization to higher rank). We first show that the spectrum of $\frac{1}{\sqrt{N}}\tilde{Y}^\Delta$ and the spectrum of

$$Z = \frac{1}{\sqrt{N}}\left(\tilde{Y} - \tilde{\mu} + \mu\right)$$

differ by a matrix with operator norm bounded by $O(1/\sqrt{N})$. Indeed, bounding the operator norm by the Hilbert-Schmidt norm, we get,

$$\|Z - \frac{1}{\sqrt{N}}\tilde{Y}\|_{op} \leq \left(\frac{1}{N}\sum_{ij}(\mu_{ij} - \tilde{\mu}_{ij})^2\right)^{1/2} \leq \sqrt{O\Big(\frac{k^4N^2}{N^3}\Big)} = O\Big(\frac{k^2}{\sqrt{N}}\Big)\,.$$

Next, we check that the empirical distribution of the eigenvalues of Z and of $\frac{1}{\sqrt{N}\Delta} \odot Y^\Delta$ are close, as well as the largest eigenvalue. We first consider the recentered matrices

$$\tilde{W} = \frac{1}{\sqrt{N}}\left(\tilde{Y} - \tilde{\mu}\right), \qquad W_{ij} = \frac{1}{\sqrt{N}}(\Delta_{ij}^{-1}Y_{ij}^\Delta - \mu_{ij})\,.$$

We use [2] and check that all the conditions of this paper are satisfied by these matrices W and $\tilde{W}$. We therefore can use [2, Theorem 1.7 and (1.22)], to conclude that the Stieltjes transform $G_{\tilde{W}}(z)$ of $\tilde{W}$ and the Stilejes transform $G_W(z)$ of W are close to their deterministic limits. Namely, under these hypotheses, if X is a matrix with centered independent entries with variance s_{ij} bounded by c/N then for any deterministic vector w so that $\|w\|_\infty \leq 1$

$$\left|\frac{1}{N}\sum_{i=1}^N w_i((z-X)^{-1}_{ii} - m_i^s(z))\right| \leq C\frac{1}{\sqrt{N\Im z}} \tag{3.9}$$

where $\mathbf{m}^s = \{m_i^s\}_{1\leq i\leq N}$ is the unique solution of the vector equation

$$-\frac{1}{m_i^s(z)} = z + \sum_{i=1}^N s_{ij}m_j^s(z).$$

We therefore only need to check that $m^{\tilde{\sigma}/N}$ and $m^{\sigma/N}$ are close and apply (3.9) with $w_i = 1$ for all i to conclude that the Stieltjes transform of both matrices are close to each other, yielding the conclusion by (3.9). This follows from [2, Corollary 3.4] which asserts that if $\Im z > \delta$,

$$\|\mathbf{m}^{\sigma/N}(z) - \mathbf{m}^{\tilde{\sigma}/N}(z)\|_\infty \leq \frac{1}{\delta}\max_i \frac{1}{N}\sum_j |\sigma_{ij}^2 - \tilde{\sigma}_{ij}^2| \leq O\Big(\frac{k}{\sqrt{N}\delta}\Big)\,. \tag{3.10}$$

Hence, combining with (3.9) (with $w_i = 1$), we conclude that the empirical measures μ_W and $\mu_{\tilde{W}}$ converge vaguely to the same deterministic limit. Since moreover $\frac{1}{N}\operatorname{Tr}(W^2)$ and $\frac{1}{N}\operatorname{Tr}(\tilde{W}^2)$ are uniformly bounded with overwhelming probability, we deduce that $d(\mu_1, \mu_2)$ goes to zero in probability.

Observe that μ has finite rank because $x^0(x^0)^T$ has finite rank and Δ is piecewise constant. Therefore, the empirical measure of the eigenvalues of $\frac{1}{\sqrt{N}}\tilde{Y}$ and $\frac{1}{\sqrt{N}\Delta} \odot Y^\Delta$ are approximately the same as those of W and $\tilde{W}$ by Weyl's interlacing property. This shows (3.8).

Moreover, [2, Corollary 1.10] show that the eigenvalues of W and $\tilde{W}$ stick to the bulk, namely the extreme eigenvalues converge towards the boundary of the support of the measure with Stieljes transform $\sum \rho_s m_s^{\sigma/N}(z)$ and $\sum \rho_s m_s^{\tilde{\sigma}/N}(z)$. Because of (3.10), these boundaries are very close to each other.

We next study the BBP transition and show that the top eigenvalues of the matrices

$$Z = \tilde{W} + \frac{1}{\sqrt{N}}\mu \text{ and } \frac{1}{\sqrt{N}\Delta} \odot Y^\Delta = W + \frac{1}{\sqrt{N}}\mu$$

have the same limits. Recall that in general, if X is a self-adjoint matrix and $R = \sum_{i=1}^r \theta_i v_i v_i^T, \theta_i \neq 0$, is a finite rank matrix then λ is an eigenvalue of $X + R$ iff $\det(\lambda - X - R)$ vanishes, and therefore if λ does not belong to the spectrum of X, this is also equivalent to:

$$0 = \det(I - (\lambda - X)^{-1}R) = \prod \theta_i \det(\operatorname{diag}(\theta_j^{-1}) - \big(\langle v_i, (\lambda - X)^{-1}v_j\rangle\big)_{1\leq i,j\leq r}). \tag{3.11}$$

Therefore, to prove that an eigenvalue pops out of the bulk, it is necessary and sufficient to prove that the above right hand side vanishes for some λ outside of the bulk of X. We will show that the matrix

$\big(\langle v_i, (\lambda - X)^{-1} v_j\rangle\big)_{1\le i,j\le r}$ converges for $X = Z$ and $\frac{1}{\sqrt{N}\Delta} \odot Y^{\Delta}$ and that the limiting equation for the outliers has a unique and stable solution. In our case,

$$R = \frac{1}{\sqrt{N}}\mu = \frac{1}{N}\frac{1}{\Delta} \odot xx^{\mathsf{T}}$$

and Δ is piecewise constant. Let $x(s) = (x_i \mathbb{1}(i \in I_s))_{i\le N}$. We have the following decomposition

$$R = \frac{1}{N}\sum_{s,t=1}^{n} \frac{\|x(s)\|\|x(t)\|}{\Delta_{s,t}} \frac{x(s)}{\|x(s)\|}\left(\frac{x(t)}{\|x(t)\|}\right)^{\mathsf{T}}. \tag{3.12}$$

For fixed x, s, t we denote $M_{st} = \frac{\|x(s)\|\|x(t)\|}{\Delta_{st}}$. M is symmetric and if (γ_i, w_i) are its eigenvectors and eigenvalues, we find that

$$R = \frac{1}{N}\sum_{i=1}^{n} \gamma_i v_i v_i^T$$

with $v_i = \sum_{s=1}^{n} \frac{x(s)}{\|x(s)\|} w_i(s)$ an orthonormal family of eigenvectors of R. [2, Theorem 1.13] implies that for any $\gamma > 0$

$$\left\| \big(\langle v_i, (\lambda + iN^{-1+\gamma} - Z)^{-1} v_j\rangle\big)_{1\le i,j\le n} - \left(\sum_{k=1}^{N} m_k^{\tilde\sigma}(\lambda + iN^{-1+\gamma}) v_i(k) v_j(k)\right)_{1\le i,j\le n} \right\|$$

and

$$\left\| \left(\langle v_i, (\lambda + iN^{-1+\gamma} - \frac{1}{\sqrt{N}\Delta^2} \odot Y^{\Delta})^{-1} v_j\right)_{1\le i,j\le n} - \left(\sum_{k=1}^{N} m_k^{\sigma}(\lambda + iN^{-1+\gamma}) v_i(k) v_j(k)\rangle\right)_{1\le i,j\le n} \right\|$$

go to zero with overwhelming probability. If λ is outside of the support of the limiting distribution then we can remove the small complex number $iN^{-1+\gamma}$, and since we have seen that the support of the eigenvalues of both centered matrices converge to the same limit this is fine for any λ at a positive distance of this limiting support. Moreover, by the stability property (3.10) we know that

$$\left\| \left(\sum_{k=1}^{N} m_k^{\tilde\sigma/N}(\lambda + iN^{-1+\gamma}) v_i(k) v_j(k)\right)_{1\le i,j\le n} - \left(\sum_{k=1}^{N} m_k^{\sigma/N}(\lambda + iN^{-1+\gamma}) v_i(k) v_j(k)\rangle\right)_{1\le i,j\le n} \right\|$$

goes to zero if λ is away from the support of the limiting measure. Hence the only thing to verify is that the largest solution λ to (3.11) does not change much under these small perturbations. To that end, first notice that because Δ is piecewise constant, so is σ and therefore $\mathbf{m}^{\sigma}$ is piecewise constant, and $m_i^{\sigma}(z)$ equals to $m_s^{\sigma/N}(z)$ for $i \in I_s$. We can therefore sum over the indices inside each I_s and find

$$\left(\sum_{k=1}^{N} m_k^{\sigma/N}(z) v_i(k) v_j(k)\right)_{1\le i,j\le n} = \left(\sum_{s=1}^{n} m_s^{\sigma/N}(z) w_i(s) w_j(s)\right)_{1\le i,j\le n} = w\mathrm{diag}(m(z))w^T$$

Therefore the outliers of $X = \frac{1}{\sqrt{N}\Delta} \odot Y^{\Delta}$ and $X = Z$ satisfy

$$\det\Big(\mathrm{diag}(\gamma_j^{-1}) - w\mathrm{diag}(m^{\sigma/N}(\lambda))w^T + \varepsilon(X)\Big) = 0$$

with $\varepsilon(X)$ a matrix with operator norm going to zero. The last thing to check is that the largest solution to this equation are arbitrarily close to each others when $\|\varepsilon(X)\|_{op}$ go to zero. But the above equations characterize the outliers as zeroes of the analytic function (outside of the support of the limiting measure)

$$F(M, \lambda) = \det\Big(M - w\mathrm{diag}(m^{\sigma/N}(\lambda))w^T\Big) = 0$$

where M belongs to a neighborhood of $\mathrm{diag}(\gamma_j^{-1})$. As long as the derivative of F in λ does not vanish, its solution is smooth. This is true for almost all γ_i's, namely almost all Δ and ρ.

Finally, observe that if two $N \times N$ matrices X and Y are such that their largest eigenvalues are close and their empirical measures are close (with atomless limits) then X and Y are close in operator norms in the sense that if the eigenvalues $\lambda_i(X)$ and $\lambda_i(Y)$ are increasing in i,

$$\limsup_{N\to\infty} \max_i |\lambda_i(X) - \lambda_i(Y)| = 0$$

in probability. This applies to $X = Z$ and $Y = \frac{1}{\sqrt{N}}\tilde{Y}$ by the previous arguments. Indeed, for any self-adjoint matrix Z

$$\hat{x}_Z^{i-1} \le \lambda_i(Z) \le \hat{x}_Z^i$$

where $\hat{x}_Z^i = \inf\{x : \hat{\mu}_Z([x, \lambda_{max}(Z)]) \ge (N-i)/N\}$ and $\hat{\mu}_Z$ is the empirical measure of the eigenvalues of Z. But, because the empirical measure of the eigenvalues converge towards the same limit and the limit correspond to an atomless measure, together with the convergence towards the same limit of $\lambda_{max}(X)$ and $\lambda_{max}(Y)$, we find that for each $\delta > 0$, for N large enough $\max_i |\hat{x}_X^i - \hat{x}_Y^i| \le \delta$. □

Remark 3.15. If $x^0 \in \mathbb{R}^{N\times\kappa}$, then we can write decompose

$$(x^0)(x^0)^\intercal = \sum_{j=1}^{\kappa} \theta^j u^j (u^j)^\intercal,$$

where $\theta^1 \ge \theta^2 \ge \cdots \ge \theta_\kappa$. Then repeating the computation following (3.12) with $u^j(s) = (\sqrt{\theta^j} u_i^j \, \mathbb{1}(i \in I_s))_{i\le N}$ and R of the form

$$R = \frac{1}{N}\sum_{j=1}^{\kappa}\sum_{s,t=1}^{n} \frac{\|v^j(s)\|\|v^j(t)\|}{\Delta_{s,t}} \frac{v^j(s)}{\|v^j(s)\|}\left(\frac{v^j(t)}{\|v^j(t)\|}\right)^\intercal.$$

The rest of the proof remains unchanged, since we only examine the behavior of v^1 .

4. Lower Bound - Gaussian Interpolation

Given a sequence of symmetric matrix $\kappa\times\kappa$ matrix Q_s for each $s \le n$, we want to derive the replica symmetric formula. Let

$$\tilde{Q}_s = \sum_{t\le n} \frac{1}{\Delta_{s,t}} \rho_t Q_t.$$

and define

$$\varphi(Q) = -\sum_{s,t=1}^{n} \frac{\rho_s\rho_t}{4\Delta_{s,t}} \operatorname{Tr}((Q_s)^\intercal Q_t) + \sum_{s=1}^{n} \rho_s \mathbb{E}\ln\left[\int \exp\left(\left(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s} z\right)^\intercal x - \frac{x^\intercal \tilde{Q}_s x}{2}\right) dP_X(x)\right]$$

where $x^0 \sim P_0$ and $z \sim N(0, I_r)$. Recall that, in the Bayes optimal case we defined in (2.5)

$$\frac{1}{\Delta_{ij}} = \mathbb{E}_{P_{\text{out}}(D|w=0)}(\partial_w g_{ij}(D, 0))^2$$

which takes n^2 different values by Hypothesis 2.4 on Δ. The goal of this section is to prove that φ is a lower bound for the free energy.

Theorem 4.1 (*Bayes Optimal Lower Bound of the Free Energy*)

Assume Hypotheses 2.3, 2.2, 2.1. Then, for any $Q = (Q_1, \dots, Q_n) \in (\mathbb{S}_\kappa^+)^n$,

$$F_N(\Delta) \ge \varphi(Q) - O(\kappa N^{-1/2}).$$

Proof. We follow the standard interpolation proof. Fix a sequence $(Q_s)_{s\le n}$ of positive semidefinite matrices and for each $i \in I_s \subset N$, we set $\tilde{Q}_i = \tilde{Q}_s$. Let z_i be i.i.d. standard Gaussians independent of all other sources of randomness, and consider the interpolating Hamiltonian

$$\begin{aligned} H_N(t, x) = &\sum_{i<j} \frac{\sqrt{t}W_{ij}}{\sqrt{\Delta_{ij}N}}(x_i \cdot x_j) + \frac{t}{\Delta_{ij}N}(x_i \cdot x_j)(x_i^0 \cdot x_j^0) - \frac{t}{2\Delta_{ij}N}(x_i \cdot x_j)^2 \\ &+ \sum_{i\le N} \sqrt{1-t}\big((\tilde{Q}_i^{1/2} z_i) \cdot x_i\big) + (1-t)\big((\tilde{Q}_i x_i^0) \cdot x_i\big) - \frac{(1-t)}{2} x_i^\intercal \tilde{Q}_i x_i. \end{aligned}$$

The corresponding interpolating free energy is given by

$$\varphi(t) = \frac{1}{N}\mathbb{E}\ln \int e^{H_N(t,x)} \, dP_0^{\otimes N}(x).$$

It follows that

$$\varphi(1) = F_N(\Delta) \quad \text{and} \quad \varphi(0) = \sum_{s\le n} \rho_s \mathbb{E}_{z,x^0} \ln\left[\int \exp\left(\left(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s} z\right)^{\mathsf{T}} x - \frac{x^{\mathsf{T}} \tilde{Q}_s x}{2}\right) dP_X(x)\right]. \tag{4.1}$$

We now control the derivative

$$\begin{aligned}\varphi'(t) = \frac{1}{N}\mathbb{E}\Big(\sum_{i<j} \frac{W_{ij}}{2\sqrt{t}\sqrt{\Delta_{ij}N}}\langle x_i\cdot x_j\rangle_t + \frac{1}{\Delta_{ij}N} x_i^0\cdot x_j^0 \langle x_i\cdot x_j\rangle_t - \frac{1}{2\Delta_{ij}N}\langle (x_i\cdot x_j)^2\rangle_t\Big)\\ - \frac{1}{N}\mathbb{E}\Big(\sum_{i=1}^N \frac{1}{2\sqrt{1-t}}\big\langle (\tilde{Q}_i^{1/2} z_i)\cdot x_i\big\rangle_t + \big\langle (\tilde{Q}_i x_i^0)\cdot x_i\big\rangle_t - \frac{1}{2}\langle x_i^{\mathsf{T}}\tilde{Q}_i x_i\rangle_t\Big)\end{aligned} \tag{4.2}$$

where the inner average is with respect to the Gibbs measure associated with $H_N(t,x)$,

$$\langle f\rangle_t = \frac{\int f e^{H_N(t,x)}\, dP_0^{\otimes N}(x)}{\int e^{H_N(t,x)}\, dP_0^{\otimes N}(x)}.$$

The Gaussian terms in (4.2) can be simplified by integrating by parts,

$$\mathbb{E}_W \sum_{i<j} \frac{W_{ij}}{2\sqrt{t}\sqrt{N\Delta_{ij}}}\langle x_i\cdot x_j\rangle_t = \mathbb{E}_W \sum_{i<j}\frac{1}{2\Delta_{ij}N}\langle (x_i^1\cdot x_j^1)^2\rangle_t - \mathbb{E}_W\sum_{i<j}\frac{1}{2\Delta_{ij}N}\langle (x_i^1\cdot x_j^1)(x_i^2\cdot x_j^2)\rangle_t$$

and similarly,

$$\mathbb{E}_z \sum_{i=1}^N \frac{1}{2\sqrt{1-t}}\big\langle (\tilde{Q}_i^{1/2} z_i)\cdot x_i\big\rangle_t = \mathbb{E}_z\sum_{i=1}^N \frac{1}{2}\big\langle (x_i^1)^{\mathsf{T}}\tilde{Q}_i(x_i^1)\big\rangle_t - \mathbb{E}_z\sum_{i=1}^N\frac{1}{2}\big\langle (x_i^1)^{\mathsf{T}}\tilde{Q}_i x_i^2\big\rangle_t$$

where $\boldsymbol{x}^2$ is an independent copy (replica) of $\boldsymbol{x}^1$. All the second order terms cancel with the self overlap terms in (4.2) leaving us with

$$\begin{aligned}\varphi'(t) = \frac{1}{N}\mathbb{E}\Big(\sum_{i<j} -\frac{1}{2\Delta_{ij}N}\langle (x_i^1\cdot x_j^1)(x_i^2\cdot x_j^2)\rangle_t + \frac{x_i^0\cdot x_j^0}{\Delta_{ij}N}\langle x_i\cdot x_j\rangle_t\Big)\\ - \frac{1}{N}\mathbb{E}\Big(\sum_{i=1}^N -\frac{1}{2}\big\langle (x_i^1)^{\mathsf{T}}\tilde{Q}_i x_i^2\big\rangle_t + \big\langle (x_i^0)^{\mathsf{T}}\tilde{Q}_i x_i\big\rangle_t\Big) + O(\kappa N^{-1/2})\end{aligned} \tag{4.3}$$

where the error comes from the diagonal terms of the overlap matrices, which are of order κ. We can now use the Nishimori property (see for example [29, Proposition 16])

$$\mathbb{E}\langle f(x^1, x^2, \dots, x^n)\rangle_t = \mathbb{E}\langle f(x^0, x^2, \dots, x^n)\rangle_t \tag{4.4}$$

to replace one replica under the average interpolating Gibbs measure with the signal. To prove (4.4), it is enough to show that the average $\langle\cdot\rangle_t$ can be interprated as a distribution of x conditionally on x^0, W and z. Indeed, if we let W, z to be Gaussian and set

$$Y_{ij} = \sqrt{t} x_i^0\cdot x_j^0 + \sqrt{\Delta_{ij}} W_{ij}.$$

Then the law of Y_{ij} has density proportional to $e^{-\frac{1}{2\Delta_{ij}}(Y_{ij}-\sqrt{t}x_i^0\cdot x_j^0)^2}$. Hence the law of x such that

$$Y_{ij} = \sqrt{t} x_i\cdot x_j + \sqrt{\Delta_{ij}} W_{ij}$$

conditionally to x^0 and W has density with respect to $P(x)$ proportional to

$$\exp\Big(-\frac{1}{2\Delta_{ij}}\Big(\sqrt{t}x_i\cdot x_j + \Delta_{ij}W_{ij} - \sqrt{t}x_i^0\cdot x_j^0\Big)^2\Big)$$

which is proportional to

$$\exp\Big(\sum_{i<j}\frac{\sqrt{t}W_{ij}}{\sqrt{\Delta_{ij}N}}(x_i\cdot x_j) + \frac{t}{\Delta_{ij}N}(x_i\cdot x_j)(x_i^0\cdot x_j^0) - \frac{t}{2\Delta_{ij}N}(x_i\cdot x_j)^2\Big).$$

The same is true for the second term if we write

$$y_i = z_i + \sqrt{1-t}\tilde{Q}_i^{1/2} x_i^0$$

with z a standard Gaussian vector, and condition by $y_i = z_i + \sqrt{1-t}\tilde{Q}_i^{1/2} x_i$. Hence $e^{H_N(t,x)}\, dP_0^{\otimes N}(x)$ is the distribution of x conditioned by W, z and x^0. This implies (4.4). Plugging the Nishimori equation (4.4) into (4.3) yields that

$$\begin{aligned}\varphi'(t) &= \mathbb{E}\bigg(\frac{1}{2N^2}\Big\langle \sum_{i<j} \langle \frac{1}{\Delta_{ij}}(x_i^1 \cdot x_j^1)(x_i^2 \cdot x_j^2)\Big\rangle_t\bigg) - \mathbb{E}\bigg(\frac{1}{2N}\Big\langle \sum_{i=1}^N \mathrm{Tr}(Q_i x_i^1 (x_i^2)^\intercal\Big\rangle_t\bigg) + O(\kappa N^{-1/2}) \\ &= \mathbb{E}\bigg(\frac{1}{4}\Big\langle \sum_{s,t=1}^n \frac{\rho_s\rho_t}{\Delta_{s,t}} \mathrm{Tr}\left((R_{1,2}^s)^\intercal R_{1,2}^t\right)\Big\rangle_t\bigg) - \mathbb{E}\bigg(\frac{1}{2}\Big\langle \sum_{s=1}^n \frac{\rho_s\rho_t}{\Delta_{s,t}} \mathrm{Tr}\left((\tilde{Q}_s)^\intercal R_{1,2}^t\right)\Big)\Big\rangle_t\bigg) + O(\kappa N^{-1/2})\end{aligned}$$

where we denoted by $R \in \mathbb{R}^{\kappa\times\kappa}$ the overlap matrix defined for $t \in \{1,\dots,n\}$ defined by:

$$R_{ab}^t = \frac{1}{|I_t|}\sum_{i\in I_t} x_i^a(\ell) x_i^b(k), \quad \ell, k \in [\kappa]\,.$$

Adding and subtracting $\sum_{s,t\le n} \frac{\rho_s\rho_t}{4\Delta_{s,t}^2}\mathrm{Tr}((Q_s)^\intercal Q_t)$, completes the square so the formula simplifies to

$$\varphi'(t) = \mathbb{E}\bigg(\frac{1}{4}\Big\langle \sum_{s,t=1}^N \frac{\rho_s\rho_t}{\Delta_{s,t}} \mathrm{Tr}\left(((R_{1,2}^s) - Q_s)^\intercal (R_{1,2}^t - Q_t)\right)\Big\rangle_t\bigg) - \sum_{s,t=1}^N \frac{\rho_s\rho_t}{4\Delta_{s,t}}\mathrm{Tr}((Q_s)^\intercal Q_t) + O(\kappa N^{-1/2})$$

Our assumption that $\frac{1}{\Delta}$ is a non-negative matrix by Hypotheses 2.2 implies that the first term is non-negative, so we arrive at the lower bound

$$\varphi'(t) \ge -\sum_{s,t=1}^n \frac{\rho_s\rho_t}{4\Delta_{s,t}}\mathrm{Tr}((Q_s)^\intercal Q_t) + O(\kappa N^{-1/2})$$

Integrating this bound implies that

$$\varphi(1) \ge \varphi(0) - \sum_{s,t=1}^n \frac{\rho_s\rho_t}{4\Delta_{s,t}}\mathrm{Tr}((Q_s)^\intercal Q_t) + O(\kappa N^{-1/2})$$

so the conclusion follows. □

5. The Upper Bound — Cavity Computations

5.1. **Concentration of the Overlaps.** We will introduce a perturbation of the Hamiltonian that will imply concentration of the Hadamard powers of the overlaps and the generalized Ghirlanda–Guerra identities [36, 38, 37] in each block of the inhomogeneous vector spin models. Given a vector $\lambda = (\lambda(1),\dots,\lambda(\kappa)) \in \mathbb{R}^\kappa$, consider the p-spin Gaussian estimation problem

$$Y_{i_1,\dots,i_p} = g_{i_1,\dots,i_p} + \frac{s}{N^{\frac{p-1}{2}}}\sum_{k\le\kappa}\lambda(k) x_{i_1}^0(k)\cdots x_{i_p}^0(k) \tag{5.1}$$

where $1 \le i_1,\dots,i_p \le N$ is some enumeration of the indices. Later on, we will restrict $i_1,\dots,i_p \in I_s$, but the pertubation Hamiltonian can be defined more generally. Since

$$Y_{i_1,\dots,i_p} - \frac{s}{N^{\frac{p-1}{2}}}\sum_{k\le\kappa}\lambda(k) x_{i_1}^0(k)\cdots x_{i_p}^0(k)$$

is a standard Gaussian variable, the maximum likelihood estimator of this Gaussian channel is proportional to

$$\begin{aligned} dP(x^0|Y_{i_1,\dots,i_p}) &= \frac{1}{Z}\exp\bigg(-\frac{1}{2}(Y_{i_1,\dots,i_p} - \frac{s}{N^{\frac{p-1}{2}}}\sum_{k\le\kappa}\lambda(k)x_{i_1}(k)\cdots x_{i_p}(k))^2\bigg) dP_0^{\otimes N}(x) \\ &= \frac{1}{Z'}\exp\bigg(\frac{s}{N^{\frac{p-1}{2}}}\sum_{k\le\kappa} Y_{i_1,\dots,i_p}\lambda(k)x_{i_1}(k)\cdots x_{i_p}(k) \\ &\quad - \frac{s^2}{2N^{p-1}}\sum_{k,k'\le\kappa}\lambda(k)x_{i_1}(k)\cdots x_{i_p}(k)\lambda(k')x_{i_1}(k')\cdots x_{i_p}(k')\bigg) dP_0^{\otimes N}(x)\end{aligned}$$

where Z and Z' are the partition functions or normalizing constants. We denote for $t \le n$

$$(R_{\ell,\ell'}^t)(k,k') = \frac{1}{N_t}\sum_{i\in I_t} x_i^\ell(k) x_i^{\ell'}(k') \text{ and } (R_{\ell,\ell'}^t)^{\odot p}(k,k') = (R_{\ell,\ell'}^t(k,k'))^p, k, k' \in [\kappa]$$

where N_t are the proportions of indices in the group with index t as defined in (2.7). We denote also in short $(\rho_t R^t_{\ell,\ell'})^{\odot p} = \rho_t^p (R^t_{\ell,\ell'})^{\odot p}$.

If we consider an independent copy for each $i_1, \dots, i_p$ conditionally on x^0, then the perturbation Hamiltonian (the log-likelihood) for $t \le n$ is given by

(5.2)
$$H^t_{N,p}(x,g,\lambda,s) = \sum_{i_1,\dots,i_p \in I_t} \frac{s}{N^{\frac{p-1}{2}}} \sum_{k\le\kappa} g_{i_1,\dots,i_p}\lambda(k)x_{i_1}(k)\cdots x_{i_p}(k) + s^2 N\big(\lambda^{\mathsf T}(\rho_t R^t_{1,0})^{\odot p}\lambda\big) - \frac{s^2}{2} N\big(\lambda^{\mathsf T}(\rho_t R^t_{1,1})^{\odot p}\lambda\big)$$

Notice that the covariance of the Gaussian term is

$$\begin{aligned}
&\frac{1}{N}\mathbb{E}\Big(\sum_{i_1,\dots,i_p\in I_t} g_{i_1,\dots,i_p}\frac{s}{N^{\frac{p-1}{2}}}\sum_{k\le\kappa}\lambda(k)x^1_{i_1}(k)\cdots x^1_{i_p}(k)\Big)\Big(\sum_{i_1,\dots,i_p\in I_t} g_{i_1,\dots,i_p}\frac{s}{N^{\frac{p-1}{2}}}\sum_{k\le\kappa}\lambda(k)x^2_{i_1}(k)\cdots x^2_{i_p}(k)\Big)\\
&= \frac{s^2}{N^p}\sum_{i_1,\dots,i_p\in I_t}\Big(\sum_{k\le\kappa}\lambda(k)x^1_{i_1}(k)\cdots x^1_{i_p}(k)\Big)\Big(\sum_{k\le\kappa}\lambda(k)x^2_{i_1}(k)\cdots x^2_{i_p}(k)\Big)\\
&= s^2\big(\lambda^{\mathsf T}(\rho_t R^t_{1,2})^{\odot p}\lambda\big).
\end{aligned}$$

For s sufficiently large, adding this perturbation to the Gibbs measure will imply concentration of the quadratic forms $(\lambda^{\mathsf T} R^{\odot p}_{1,2}\lambda)$. For applications, we will need concentration for all λ and all $p \ge 1$. For any $\lambda \in [-1,1]^\kappa$ and $p \ge 1$, the overlap is uniformly bounded

$$(\lambda^{\mathsf T} R^{\odot p}_{1,2}\lambda) \le \kappa^2 C^{2p}, \tag{5.3}$$

since the vector spin coordinates x are uniformly bounded by some $C \ge 1$. Let λ_m be a countable enumeration of elements of the dense set $([-1,1]\cap\mathbb{Q})^\kappa$ and consider the perturbed Hamiltonian

$$H^{\mathrm{pert}}_N(x,\lambda) = H_N(x) + \sum_{t=1}^n \sum_{p\ge1}\sum_{m\ge1} H^t_{N,p}\Big(x, g_{m,t}, \lambda_m, \frac{u_{m,p,t}\varepsilon_N}{2^{m+p}\kappa C^p}\Big).$$

The above sum is infinite but the constants will be chosen so that the covariance of the above Gaussian process is absolutely converging. The gaussian variables $g_{m,t}$ appearing in different $H^t_{N,p}$ are independent. We choose the scaling coefficient $C_{m,p} := 2^{m+p}\kappa C^p$ so that the covariance is of order $O(s^2 N)$

$$\begin{aligned}
&\frac{1}{N}\operatorname{Cov}\Big(\sum_{t=1}^n\sum_{p\ge1}\sum_{m\ge1} H^p_N\Big(x,g_{m,t},\lambda_m,\frac{u_{m,p,t}\varepsilon_N}{C_{m,p}}\Big), \sum_{t=1}^n\sum_{p\ge1}\sum_{m\ge1} H^p_N\Big(x,g_{m,t},\lambda_m,\frac{u_{m,p,t}\varepsilon_N}{C_{m,p}}\Big)\Big)\\
&= \varepsilon_N^2\sum_{t=1}^n\sum_{p\ge1}\sum_{m\ge1} u^2_{m,p,t}\frac{(\lambda_m^{\mathsf T}(\rho_t R^t_{1,2})^{\odot p}\lambda_m)}{C^2_{m,p}}\\
&\le \varepsilon_N^2 \max_{m,p,t}(u^2_{m,p,t}).
\end{aligned} \tag{5.4}$$

If $\max_{m,p,t} u_{m,p,t}\varepsilon_N \to 0$, then the covariance of the perturbation term will be of lower order than the Hamiltonian, so the limit of the free energy will not change because of the perturbation. To make this precise, consider the perturbed free energy as a function of the infinite sequence $u = (u_{m,p})$

$$F^{\mathrm{pert}}_N(u) = \frac{1}{N}\log\int e^{H_N(x)+\sum_{t=1}^n\sum_{p\ge1}\sum_{m\ge1} H^t_{N,p}(x,g_{m,t},\lambda_m,\frac{u_{m,p,t}\varepsilon_N}{2^{m+p}rC^p})}\,dP_0^{\otimes N}(x) \tag{5.5}$$

and denote in short $F_N = F^{\mathrm{pert}}_N(0)$. In this section, we also let $\langle\cdot\rangle_{\mathrm{pert}}$ denote the Gibbs average with respect the perturbed Hamiltonian,

$$\langle f(x)\rangle_{\mathrm{pert}} = \frac{\int f(x)e^{H^{\mathrm{pert}}_N(x)}\,dP_0^{\otimes N}(x)}{\int e^{H^{\mathrm{pert}}_N(x)}\,dP_0^{\otimes N}(x)}$$

which corresponds to averages with respect to the probability $P(x^0|Y,(Y_\theta)_{\theta\in\Theta})$ where Y_θ in an enumeration various p-spin Gaussian interference problems introduced in (5.1). In particular, the Nishimori property is valid for averages with respect to the random Gibbs measure because it corresponds to a conditional probability.

Lemma 5.1 (*Equivalence of the Perturbed Free Energy*)

Assume Hypothesis 2.3. Uniformly over all $u_{m,p,t} \in [1/2, 1]$

$$|\mathbb{E}F_N^{\mathrm{pert}}(u) - \mathbb{E}F_N| \le \varepsilon_N^2.$$

In particular, if $\varepsilon_N^2 \to 0$*, then the perturbation will not change the limit of the free energy.*

Observe that this estimate holds independently of the choice of the original Hamiltonian H_N.

Proof. Consider the interpolating free energy,

$$\varphi(\tau) = \mathbb{E}\left[\frac{1}{N}\log\int e^{H_N(x)+\sum_{p\ge1}\sum_{m\ge1} H^t_{N,p}(x,g_{m,t},\lambda_m,\frac{u_{m,p,t}\tau}{2^{m+p}rC^p})}\,dP_0(x)\right].$$

as a function of the ε_N parameter. Notice that $\varphi(0) = \mathbb{E}F_N$ and $\varphi(\varepsilon_N) = \mathbb{E}F_N^{\mathrm{pert}}(u)$. A straightforward integration by parts computation and the Nishimori identity (4.4) implies that

$$\begin{aligned}
\varphi'(\tau) &= \mathbb{E}\Big\langle \sum_{t=1}^{n}\sum_{p\ge1}\sum_{m\ge1}\sum_{i_1,\dots,i_p\in I_t}\frac{u_{m,p,t}}{C_{m,p}N^{\frac{p-1}{2}+1}}\sum_{k\le\kappa} g_{i_1,\dots,i_p}\lambda_{m,p}(k)x_{i_1}(k)\cdots x_{i_p}(k)\\
&\quad + 2\frac{\tau u_{m,p,t}^2}{C_{m,p}^2}\big(\lambda_{m,p}^{\mathsf{T}}(\rho_t R_{1,1}^t)^{\odot p}\lambda_{n,p}\big) - \frac{\tau u_{m,p,t}^2}{C_{m,p}^2}\big(\lambda_{m,p}^{\mathsf{T}}(\rho_t R_{1,1}^t)^{\odot p}\lambda_{n,p}\big)\Big\rangle_{\mathrm{pert}}\\
&= \mathbb{E}\Big\langle \sum_{t=1}^{n}\sum_{p\ge1}\sum_{m\ge1}\frac{tu_{m,p,t}^2}{C_{m,p}^2}\big(\lambda_{m,p}^{\mathsf{T}}(\rho_t R_{1,1}^t)^{\odot p}\lambda_{m,p}\big) - \frac{tu_{m,p,t}^2}{C_{m,p}^2}\big(\lambda_{m,p}^{\mathsf{T}}(\rho_t R_{1,2}^t)^{\odot p}\lambda_{m,p}\big)\\
&\quad + \frac{2\tau u_{m,p,t}^2}{C_{m,p}^2}\big(\lambda_{m,p}^{\mathsf{T}}(\rho_t R_{1,0}^t)^{\odot p}\lambda_{m,p}\big) - \frac{\tau u_{m,p,t}^2}{C_{m,p}^2}\big(\lambda_{m,p}^{\mathsf{T}}(\rho_t R_{1,1}^t)^{\odot p}\lambda_{m,p}\big)\Big\rangle_{\mathrm{pert}}\\
&\le \tau\max_{m,p}(u_{m,p,t}^2).
\end{aligned}$$

because the overlaps are bounded uniformly under Hypothesis 2.3. Therefore, for $t \in [0, \varepsilon_N]$, we have $\varphi'(\tau) \le \varepsilon_N$ since $\max_{m,p,t}(u_{m,p,t}^2) \le 1$, so the result follows. □

On the other hand, if we take ε_N going to zero sufficiently slowly, then we will be able to regularize the Gibbs measure using this perturbation. We will fix a $t \le n$, and show that the overlaps within $t \le n$ will concentrate in the limit.

We define

$$v_N = \sup_u \mathbb{E}(NF_N^{\mathrm{pert}}(u) - N\mathbb{E}F_N^{\mathrm{pert}}(u))^2 \tag{5.6}$$

where the supremum is taken over all $u_{n,p,t} \in [1/2, 1]$ and the expectation $\mathbb{E}$ is over the Gaussian variables $g_{m,t,i_1,\dots,i_p}$ and the $\boldsymbol{x}^0$. In our applications, v_N is usually of order N as we will see by using concentration of measure.

We consider the case when the parameters $u_{m,p,t}$ are random. For each $m, p \ge 1$, consider $u_{m,p,t}\varepsilon_N$ where $u_{m,p,t} \in [1/2, 1]$ are uniform and independent and $\varepsilon_N^2 \to 0$, so that the limit of the free energy is unchanged by Lemma 5.1. We will prove that if $\varepsilon_N = N^{-\gamma}$ for $\gamma < \frac{1}{4}$, then the perturbation can be large enough to regularize the Gibbs measure by implying concentration of the quadratic forms of overlaps and the uniform concentration of the log partition function (or the free energies not normalized by N) with respect to the perturbed as a function of $u_{m,p,t}$. Taking $u_{m,p,t} \sim U[1/2, 1]$ independent for all m, p, t, we get the following bound for the concentration of the overlaps on average.

Theorem 5.2 (*Concentration Bound of the Overlap*)

Assume Hypotheses 2.3 and 2.1. If $\frac{v_N}{N^2\varepsilon_N} \to 0$*, then for any* $m, p \ge 1$ *there exists a constant* $L_{m,p}$ *that only depends on* m *and* p *such that*

$$\max_t \mathbb{E}_u\mathbb{E}\langle((\lambda_m^{\mathsf{T}}(R_{1,2}^t)^{\odot p}\lambda_m) - \mathbb{E}\langle(\lambda_m^{\mathsf{T}}(R_{1,2}^t)^{\odot p}\lambda_m)\rangle_{\mathrm{pert}})^2\rangle_{\mathrm{pert}} \le L_{m,p}\left(\left(\frac{v_N}{N^2\varepsilon_N^4}\right)^{1/3} + \frac{1}{\varepsilon_N^2 N}\right). \tag{5.7}$$

where $\langle\cdot\rangle = \langle\cdot\rangle_{\mathrm{pert}}$ *is the average with respect to the perturbed Gibbs measure.* $\mathbb{E}$ *denotes the average with respect to the Gaussian random variables that appear in the Hamiltonian and the signal variable* $\boldsymbol{x}^0$*. The outer average* $\mathbb{E}_u$ *is with respect to the uniform random variables* $(u_{m',p',t'})_{m',p',t'}$*.*

Proof. This proof is a generalization of the case when $p = 1$ and $\kappa = 1$ found in [7]. To simplify notation, we will drop the subscript on the Gibbs average, $\langle\cdot\rangle := \langle\cdot\rangle_{\text{pert}}$ and the subscripts m, p, t because they are fixed throughout the proof. Furthermore, since the $(u_{m,p,t})_{m,p,t\geq 1}$ are independent, we can fix the $u_{m',p',t'}$ for $m' \neq n$, $p' \neq p$, and $t \neq t'$ and average with respect to the Gaussian g and $u_{m,p,t}$ first. This restriction will not affect the validity of the Nishimori property (4.4) because the Gibbs measure is a conditional probability corresponding to a Gaussian estimation problem for all $u_{m,p,t}$. To also simplify notation, we will abuse notation and define

$$R_{\ell,\ell'}^{\odot p} = (\rho R_{\ell,\ell'}^{t})^{\odot p}$$

in this proof because ρ, p, t are fixed, so it will not affect any computations.

Step 1: We first bound the moments and show that

$$\mathbb{E}\langle((\lambda_m^{\mathsf{T}} R_{1,0}^{\odot p}\lambda_m) - \mathbb{E}\langle(\lambda_m^{\mathsf{T}} R_{1,0}^{\odot p}\lambda_m)\rangle)^2\rangle \leq \frac{4C_{m,p}^2}{\varepsilon_N^2 N^2}\mathbb{E}\big\langle(H' - \mathbb{E}\langle H'\rangle)^2\big\rangle \tag{5.8}$$

where $H' = \partial_s H$ denotes the derivative with respect to the last coordinate of the Hamiltonian defined in (5.2) which we denote by

$$s := s_{m,p,t} = \frac{u_{m,p,t}\varepsilon_N}{2^{m+p}\kappa C^p} \in \Big[\frac{\varepsilon_N}{2\cdot 2^{m+p}\kappa C^p}, \frac{\varepsilon_N}{2^{m+p}\kappa C^p}\Big]. \tag{5.9}$$

We also recall that $C_{m,p} = 2^{m+p}\kappa C^p$. During the proof we write in short u for $u_{m,p,t}$. This inequality comes from an integration by parts argument and the Nishimori property. A crucial observation is that we can integrate by parts with respect to the Gaussian random variables conditionally on all other sources of randomness by independence. By independence, we first do the computation conditionally on s and denote in short λ for $\lambda_{m,p}$. We first prove that

$$\frac{1}{N}\mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,2}^{\odot p}\lambda)(H' - \mathbb{E}\langle H'\rangle)\rangle \leq s\mathbb{E}\langle((\lambda^{\mathsf{T}} R_{1,2}^{\odot p}\lambda) - \mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,2}^{\odot p}\lambda)\rangle)^2\rangle + s\mathbb{E}\langle((\lambda^{\mathsf{T}} R_{1,2}^{\odot p}\lambda) - \langle(\lambda^{\mathsf{T}} R_{1,2}^{\odot p}\lambda)\rangle)^2\rangle. \tag{5.10}$$

Notice that the left hand side simplifies to

$$\begin{aligned}
&\frac{1}{N}\mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)(H' - \mathbb{E}\langle H'\rangle)\rangle \\
&= \frac{1}{N^{\frac{p+1}{2}}}\sum_{i_1,\dots,i_p}\mathbb{E}\Big[\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)g_{i_1,\dots,i_p}\sum_k \lambda(k)x_{i_1}(k)\cdots x_{i_p}(k)\rangle + 2s\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle - s\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)(\lambda^{\mathsf{T}} R_{1,1}^{\odot p}\lambda)\rangle\Big] \\
&- \mathbb{E}\Big[\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle\Big]\Big(\frac{1}{N^{\frac{p+1}{2}}}\sum_{i_1,\dots,i_p}\mathbb{E}\Big[\langle g_{i_1,\dots,i_p}\sum_k \lambda(k)x_{i_1}(k)\cdots x_{i_p}(k)\rangle + 2s\mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle - s\mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,1}^{\odot p}\lambda)\rangle\Big)\Big].
\end{aligned}$$

Since $\boldsymbol{x}^0$ is independent from the Gaussian terms, we can integrate the Gaussian terms using integration by parts conditionally on the $\boldsymbol{x}^0$ (and recalling the extra s factor in the exponent) to conclude that the above right hand side equals

$$\begin{aligned}
&s\mathbb{E}\Big[\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)(\lambda^{\mathsf{T}} R_{1,1}^{\odot p}\lambda)\rangle - \langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)(\lambda^{\mathsf{T}} R_{1,2}^{\odot p}\lambda)\rangle + 2\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle - \langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)(\lambda^{\mathsf{T}} R_{1,1}^{\odot p}\lambda)\rangle\Big] \\
&- \mathbb{E}\Big[\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle\Big]\Big(s\mathbb{E}\Big[\langle(\lambda^{\mathsf{T}} R_{1,1}^{\odot p}\lambda)\rangle - s\mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,2}^{\odot p}\lambda)\rangle + 2s\mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle - s\mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,1}^{\odot p}\lambda)\rangle\Big)\Big].
\end{aligned} \tag{5.11}$$

Next, we integrate with respect to $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$ independently since they are independent conditionally on $\boldsymbol{x}^0$ and apply the Nishimori property to conclude that

$$\mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)(\lambda^{\mathsf{T}} R_{1,2}^{\odot p}\lambda)\rangle = \mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)(\lambda^{\mathsf{T}} R_{2,0}^{\odot p}\lambda)\rangle = \mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle^2.$$

which implies that (5.11) can be further simplified to

$$\begin{aligned}
&- s\mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle^2 + 2s\mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)^2\rangle - s(\mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle)^2 \\
&= s\Big(\mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)^2\rangle - (\mathbb{E}_{\langle}(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle)^2\Big) + s\Big(\mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)^2\rangle - \mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle^2\Big) \\
&= s\mathbb{E}\langle((\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda) - \mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle)^2\rangle + s\mathbb{E}\langle((\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda) - \langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle)^2\rangle.
\end{aligned}$$

We can now conclude that

$$\frac{1}{N}\mathbb{E}\langle((\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda) - \mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle)(H' - \mathbb{E}\langle H'\rangle)\rangle \geq s\mathbb{E}\langle((\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda) - \mathbb{E}\langle(\lambda^{\mathsf{T}} R_{1,0}^{\odot p}\lambda)\rangle)^2\rangle$$

so the Cauchy–Schwarz inequality implies

$$s\mathbb{E}\langle((\lambda^\intercal R_{1,0}^{\odot p}\lambda)-\mathbb{E}\langle(\lambda^\intercal R_{1,0}^{\odot p}\lambda)\rangle)^2\rangle \le \frac{1}{N}|\mathbb{E}\langle((\lambda^\intercal R_{1,0}^{\odot p}\lambda)-\mathbb{E}\langle(\lambda^\intercal R_{1,0}^{\odot p}\lambda)\rangle)(H'-\mathbb{E}\langle H'\rangle)\rangle| \\ \le \frac{1}{N}\Big(\mathbb{E}\langle((\lambda^\intercal R_{1,0}^{\odot p}\lambda)-\mathbb{E}\langle(\lambda^\intercal R_{1,0}^{\odot p}\lambda)\rangle)^2\rangle\mathbb{E}\langle(H'-\mathbb{E}\langle H'\rangle)^2\rangle\Big)^{1/2},$$

which simplifies to

$$s^2\mathbb{E}\langle((\lambda^\intercal R_{1,0}^{\odot p}\lambda)-\mathbb{E}\langle(\lambda^\intercal R_{1,0}^{\odot p}\lambda)\rangle)^2\rangle \le \frac{1}{N^2}\mathbb{E}\big\langle(H'-\mathbb{E}\langle H'\rangle)^2\big\rangle.$$

Using the fact that $s \ge \frac{\varepsilon_N}{2C_{m,p}}$ by (5.9) implies (5.8).

Step 2: We now have to control the variance of the derivative of the Hamiltonian. Again, this part of the proof is identical to the one dimensional case, because the variances of H can be expressed in terms of the derivatives of the free energy, which can be bounded using elementary facts about convex functions [7]. We recreate the details for completeness.

We begin by bounding the difference of derivatives in (5.8) with its thermal and quenched variances

$$\frac{4C_{m,p}^2}{\varepsilon_N^2N^2}\mathbb{E}\big\langle(H'-\mathbb{E}\langle H'\rangle)^2\big\rangle \le \frac{4C_{m,p}^2}{\varepsilon_N^2N^2}\mathbb{E}\big\langle(H'-\langle H'\rangle)^2\big\rangle + \frac{4C_{m,p}^2}{\varepsilon_N^2N^2}\mathbb{E}\big\langle(\langle H'\rangle-\mathbb{E}\langle H'\rangle)^2\big\rangle \tag{5.12}$$

Our goal is to use the convexity and boundedness of the perturbed free energy functions to bound the right hand side of (5.12) from above.

We first fix the rest of the uniform random variables in the Hamiltonian except for $u_{m,p}$. We begin by bounding the first term in the RHS of (5.12) using a bound on the first derivative of the free energy. We prove that

$$\frac{4C_{m,p}^2}{\varepsilon_N^2N^2}\mathbb{E}_u\mathbb{E}\big\langle(H'-\langle H'\rangle)^2\big\rangle \le \frac{12C_{n,p}^4 4^{n+p}}{\varepsilon_N^2N}. \tag{5.13}$$

To this end, first notice that

$$\frac{d^2}{ds^2}\mathbb{E}\log Z_N^{\mathrm{pert}}(s) = \mathbb{E}\langle(H'-\langle H'\rangle)^2\rangle + 2N\mathbb{E}\langle\lambda R_{1,0}^{\odot p}\lambda\rangle - N\mathbb{E}\langle\lambda R_{1,1}^{\odot p}\lambda\rangle \tag{5.14}$$

where the last term comes from $\partial_s^2 H$. If we rearrange terms, then

$$\int_{\frac{\varepsilon_N}{2C_{n,p}}}^{\frac{\varepsilon_N}{C_{m,p}}}\mathbb{E}\big\langle(H'-\langle H'\rangle)^2\big\rangle\,ds \\ = \int_{-\frac{\varepsilon_N}{2C_{m,p}}}^{\frac{\varepsilon_N}{C_{m,p}}}\frac{d^2}{ds^2}\mathbb{E}\log Z_N^{\mathrm{pert}}(s)\,ds - N\int_{\frac{\varepsilon_N}{2C_{m,p}}}^{\frac{\varepsilon_N}{C_{m,p}}}\mathbb{E}\Big\langle 2\lambda^\intercal R_{1,0}^{\odot p}\lambda - \lambda^\intercal R_{1,1}^{\odot p}\lambda\Big\rangle\,ds \\ \le \frac{d}{ds}\mathbb{E}\log Z_N^{\mathrm{pert}}(s)\Big|_{\frac{\varepsilon_N}{2C_{m,p}}}^{\frac{\varepsilon_N}{C_{m,p}}} + 3N\int_{\frac{\varepsilon_N}{2C_{m,p}}}^{\frac{\varepsilon_N}{C_{m,p}}}\kappa^2C^{2p}\,ds.$$

where we used (5.3). Next, to control the first term, notice that

$$\frac{d}{ds}\mathbb{E}\log Z_N^{\mathrm{pert}}(s) = \mathbb{E}\langle H'\rangle = 2sN\mathbb{E}\langle\lambda^\intercal R_{1,0}^{\odot p}\lambda\rangle - sN\mathbb{E}\langle\lambda^\intercal R_{1,2}^{\odot p}\lambda\rangle \le 3Ns\kappa^2C^{2p}$$

so we can conclude that

$$\int_{\varepsilon_N/2C_{m,p}}^{\varepsilon_N/C_{m,p}}\mathbb{E}\big\langle(H'-\langle H'\rangle)^2\big\rangle\,ds \le 3N\kappa^2C^{2p}\frac{\varepsilon_N}{C_{m,p}}$$

Since $s = \frac{u\varepsilon_N}{C_{n,p}}$ where $u \sim U[1/2,1]$, the above reads

$$\mathbb{E}_u\mathbb{E}\big\langle(H'-\langle H'\rangle)^2\big\rangle \le 6N\kappa^2C^{2p} \tag{5.15}$$

Eventhough this bound is not great, remember that we will multiply it by $\frac{4C_{n,p}^2}{\varepsilon_N^2N^2}$ which is very small, see (5.8).

Step 3: We now bound the second term in the RHS of (5.12) and show that

$$\frac{4C_{m,p}^2}{\varepsilon_N^2N^2}\mathbb{E}_u\mathbb{E}\big\langle(\langle H'\rangle-\mathbb{E}\langle H'\rangle)^2\big\rangle \le 1000C_{m,p}^{\frac{16}{3}}4^{m+p}\Big(\frac{v_N}{N^2\varepsilon_N^4}\Big)^{1/3}. \tag{5.16}$$

The proof uses the expectation over $u_{n,p}$ and the convexity of the modified free energy

$$\tilde{F}_N(s) = \frac{1}{N}\log Z_N^{\text{pert}}(s) + 3s^2 4^{m+p} C_{m,p}^2 .$$

$\tilde{F}_N$ is a convex function because the computation in (5.14) implies that the second derivative

$$\frac{d^2}{ds^2}\tilde{F}_N = \frac{1}{N}\langle (H' - \langle H'\rangle)^2\rangle + 2\langle \lambda R_{1,0}^{\odot p}\lambda\rangle - \langle \lambda R_{1,1}^{\odot p}\lambda\rangle + 6\cdot 4^{m+p}C_{m,p}^2$$

is non-negative because $C_{m,p} = 2^{m+p}\kappa C^p$ and $\langle \lambda R_{1,2}^{\odot p}\lambda\rangle$ is uniformly bounded by $\kappa^2 C^{2p}$. Our goal is to control the difference of the derivatives of this convex function with its expected value

$$\tilde{F}_N' - \mathbb{E}\tilde{F}_N' = \frac{1}{N}(\langle H'\rangle - \mathbb{E}\langle H'\rangle). \tag{5.17}$$

We can now use the bounds on the derivatives of convex functions given by the following lemma from [34, Lemma 3.2].

Lemma 5.3 (*A Bound for Convex Functions*)

Let G and g be convex differentiable functions. For $\delta > 0$ and nonnegative functions $C_\delta^-(x) = g'(x) - g'(x-\delta)$ and $C_\delta^+(x) = g'(x+\delta) - g'(x)$ then for all real numbers x

$$|G'(x) - g'(x)| = \frac{1}{\delta}\sum_{u\in\{x-\delta,x,x+\delta\}} |G(u) - g(u)| + C_\delta^+(x) + C_\delta^-(x)$$

To control (5.17), we apply this lemma to $G(s) = \tilde{F}_N(s)$ and $g(s) = \mathbb{E}\tilde{F}_N(s)$ to conclude that for any $s \in (\varepsilon_N/2C_{m,p}, \varepsilon_N/C_{m,p})$ and $\delta > 0$ sufficiently small so that $s - \delta > 0$,

$$\frac{1}{N}|\langle H'\rangle - \mathbb{E}\langle H'\rangle| \le \delta^{-1}\sum_{u\in\{x-\delta,x,x+\delta\}} |F_N(u) - \mathbb{E}F_N(u)| + |\mathbb{E}\tilde{F}_N'(x) - \mathbb{E}\tilde{F}_N'(x-\delta)| + |\mathbb{E}\tilde{F}_N'(x+\delta) - \mathbb{E}\tilde{F}_N'(x)|$$

We square both sides and apply Jensen's inequality $(\sum_{i=1}^m a_i)^2 \le m\sum_{i=1}^m a_i^2$ to arrive at the bound

$$\begin{aligned}\frac{1}{5N^2}\mathbb{E}_u\mathbb{E}(\langle H'\rangle - \mathbb{E}\langle H'\rangle)^2 \le \delta^{-2}&\sum_{u\in\{x-\delta,x,x+\delta\}} \mathbb{E}_u\mathbb{E}|F_N(u) - \mathbb{E}F_N(u)|^2 \\ &+ \mathbb{E}_u|\mathbb{E}\tilde{F}_N'(s) - \mathbb{E}\tilde{F}_N'(s-\delta)|^2 + \mathbb{E}_u|\mathbb{E}\tilde{F}_N'(s+\delta) - \mathbb{E}\tilde{F}_N'(s)|^2\end{aligned} \tag{5.18}$$

It now remains to control the two terms in the above upper bound.

(1) From the definition of v_N,

$$\sup_u \mathbb{E}(F_N(u) - \mathbb{E}F_N(u))^2 \le \frac{v_N}{N^2},$$

we get

$$\delta^{-2}\sum_{u\in\{x-\delta,x,x+\delta\}} \mathbb{E}_u\mathbb{E}|F_N(u) - \mathbb{E}F_N(u)|^2 \le \frac{3v_N}{N^2\delta^2}. \tag{5.19}$$

(2) For the two last terms, we begin by controlling

$$\begin{aligned}&2\int_{\frac12}^1 \left(\left(\mathbb{E}\tilde{F}_N'\Big(\frac{u\varepsilon_N}{C_{m,p}}\Big) - \mathbb{E}\tilde{F}_N'\Big(\frac{u\varepsilon_N}{C_{m,p}} - \delta\Big)\right)^2 + \left(\mathbb{E}\tilde{F}_N'\Big(\frac{u\varepsilon_N}{C_{m,p}}\Big) - \mathbb{E}\tilde{F}_N'\Big(\frac{u\varepsilon_N}{C_{m,p}} + \delta\Big)\right)^2\right) du \\ &= \frac{2C_{m,p}}{\varepsilon_N}\int_{\frac{\varepsilon_N}{2C_{m,p}}}^{\frac{\varepsilon_N}{C_{m,p}}} \left(|\mathbb{E}\tilde{F}_N'(s) - \mathbb{E}\tilde{F}_N'(s-\delta)|^2 + |\mathbb{E}\tilde{F}_N'(s+\delta) - \mathbb{E}\tilde{F}_N'(s)|^2\right) ds\end{aligned} \tag{5.20}$$

By the Nishimori property, we get the uniform bound

$$\mathbb{E}\tilde{F}_N'(s) = \frac{1}{N}\mathbb{E}\langle H'\rangle = 2s\mathbb{E}\langle \lambda^\intercal R_{1,0}^{\odot p}\lambda\rangle - s\mathbb{E}\langle \lambda^\intercal R_{1,2}^{\odot p}\lambda\rangle \le 3\varepsilon_N C_{m,p}^2 4^{m+p} \tag{5.21}$$

so

$$\max(|\mathbb{E}\tilde{F}_N'(s) - \mathbb{E}\tilde{F}_N'(s-\delta)|, |\mathbb{E}\tilde{F}_N'(s+\delta) - \mathbb{E}\tilde{F}_N'(s)|) \le 6\varepsilon_N C_{m,p}^2 4^{m+p}$$

Since F'_N is increasing, $\mathbb{E}\tilde{F}'_N(s) - \mathbb{E}F'_N(s-\delta)$ is nonnegative so we conclude that (5.20) is bounded by

$$12C_{m,p}^3 4^{m+p} \int_{\frac{\varepsilon_N}{2C_{m,p}}}^{\frac{\varepsilon_N}{C_{m,p}}} \mathbb{E}\tilde{F}'_N(s) - \mathbb{E}\tilde{F}'_N(s-\delta) + \mathbb{E}\tilde{F}'_N(s+\delta) - \mathbb{E}\tilde{F}'_N(s)\, ds$$
$$= 12C_{m,p}^3 4^{m+p}\left(\mathbb{E}\tilde{F}_N\Big(\frac{\varepsilon_N}{C_{m,p}}+\delta\Big) - \mathbb{E}\tilde{F}_N\Big(\frac{\varepsilon_N}{C_{m,p}}-\delta\Big) + \mathbb{E}\tilde{F}_N\Big(\frac{\varepsilon_N}{2C_{m,p}}-\delta\Big) - \mathbb{E}\tilde{F}_N\Big(\frac{\varepsilon_N}{2C_{m,p}}+\delta\Big)\right)$$
$$\leq 72\delta\varepsilon_N C_{m,p}^5 8^{m+p}$$

by the mean value theorem and the bound on the derivatives (5.21).

We deduce from (5.18) and (5.19) that

$$\frac{1}{5N^2}\mathbb{E}_u\mathbb{E}(\langle H'\rangle - \mathbb{E}\langle H'\rangle)^2 \leq \frac{3v_N}{N^2\delta^2} + 72\delta\varepsilon_N C_{n,p}^5 8^{n+p}.$$

We finally choose δ to be given by

$$\delta^3 = \frac{v_N}{24N^2\varepsilon_N C_{n,p}^5 8^{n+p}}$$

so

$$\frac{1}{N^2}\mathbb{E}_u\mathbb{E}(\langle H'\rangle - \mathbb{E}\langle H'\rangle)^2 \leq 2\left(\frac{3v_N}{N^2}\left(\frac{24N^2\varepsilon_N C_{m,p}^5 8^{m+p}}{v_N}\right)^{2/3} + 72\left(\frac{v_N}{24N^2\varepsilon_N C_{m,p}^5 8^{m+p}}\right)^{1/3}\varepsilon_N C_{m,p}^5 8^{m+p}\right)$$

which gives (5.16) .

Step 4: From steps 1 and 3, we conclude that conditionally on $u_{m',p'}$ for $(m',p') \neq (m,p)$ that

$$\mathbb{E}_{u_{m,p}}\mathbb{E}\langle((\lambda_{m,p}^{\mathsf{T}} R_{1,2}^{\odot p}\lambda_{m,p}) - \mathbb{E}\langle(\lambda_{n,p}^{\mathsf{T}} R_{1,2}^{\odot p}\lambda_{m,p})\rangle)^2\rangle \leq \left(1000C_{n,p}^{\frac{16}{3}} 4^{n+p}\Big(\frac{v_N}{N^2\varepsilon_N^4}\Big)^{1/3} + \frac{12C_{n,p}^4 4^{m+p}}{\varepsilon_N^2 N}\right).$$

Since this bound is uniform in $u_{m',p'}$, Theorem 5.2 follows.

□

Remark 5.4. In the proof above, we assumed that $C_{m,p} \geq 1$ and $\varepsilon_N \leq 1$. This is not a problem, because $\varepsilon_N = N^{-\gamma}$ for $0 < \gamma < 1/4$ in our applications, and we can assume $C_{m,p} \geq 1$ at the cost of a less sharp upper bound. The constant $L_{m,p}$ is also of order $C_{m,p}^{16/3}$.

Using Gaussian concentration to explicitly estimate v_N, we can conclude concentration of all quadratic forms associated to the Hadamard powers of the overlap at rational vectors.

Corollary 5.5 (*Overlap Quadratic Form Concentration*)

If $\varepsilon_N = N^{-\gamma}$ for $0 < \gamma < 1/4$, then for any $p \geq 1$ rational valued $\lambda \in [-1,1]^\kappa$, and $t \leq n$

$$\mathbb{E}_u\mathbb{E}\langle((\lambda^{\mathsf{T}}(R_{1,2}^t)^{\odot p}\lambda) - \mathbb{E}\langle(\lambda^{\mathsf{T}}(R_{1,2}^t)^{\odot p}\lambda)\rangle_{\text{pert}})^2\rangle_{\text{pert}} \to 0. \tag{5.22}$$

Proof. We explicitly compute the rate v_N defined in (5.6) for this model. We fix our parameter u, and let $\varphi = NF_N^{\text{pert}}(u)$. By independence, we can split the expected values into a statement about the concentration of $\boldsymbol{x}^0$ and the Gaussian terms,

$$\mathbb{E}|\varphi - \mathbb{E}\varphi| \leq \mathbb{E}|\varphi - \mathbb{E}_{W,g}\varphi| + \mathbb{E}_{\boldsymbol{x}^0}|\mathbb{E}_{W,g}\varphi - \mathbb{E}\varphi| \tag{5.23}$$

The average $\mathbb{E}_W$ is with respect to the 'W' Gaussian terms in H_N, the average $\mathbb{E}_{\boldsymbol{x}^0}$ is with respect to '$\boldsymbol{x}^0$' terms in the approximate indicator and $\mathbb{E}_g$ is with respect to the 'g' Gaussian terms $g_N(\boldsymbol{x})$, and $\mathbb{E}$ is the average with respect to all sources of randomness.

We start by proving the first term satisfies

$$\mathbb{E}|\varphi - \mathbb{E}_{W,g}\varphi| \leq O(\sqrt{N + N\varepsilon_N^2}). \tag{5.24}$$

By independence, we can compute this upper bound conditionally on $\boldsymbol{x}^0$. Since the Gaussian Hamiltonian has variance of order N, the classical Gaussian concentration inequality for Lipschitz functions [34, Theorem 1.2] implies that

$$\mathbb{E}|\varphi - \mathbb{E}_{W,g}\varphi|^2 \leq 8\gamma^2 C^{2p}N + N\varepsilon_N^2.$$

We now focus on the second term and prove it satisfies

$$\mathbb{E}_{\boldsymbol{x}^0}|\mathbb{E}_{W,g}\varphi - \mathbb{E}\varphi| \leq O(N^{\frac{1}{2}}). \tag{5.25}$$

We use the bounded difference property and a consequence of the Efron–Stein inequality. For N sufficiently large

$$|\mathbb{E}_{W,g}\partial_{x_i^0(k)}\varphi| \leq \frac{1}{\|\Delta_\infty\|}\left|\mathbb{E}_{W,g}\left\langle x_j^0(k)(x_i\cdot x_j) + \sum_{p\geq 1}\frac{p}{C_{m,p}}R_{1,0}^{p-1}x_i(k)\right\rangle_{\text{pert}}\right| \leq L$$

for some universal constant L that only depends on C and $\|\Delta\|_\infty$. Since the $\mathbb{E}_{W,g}\varphi$ has a bounded derivative for each coordinate $x_i^0(k)$ and $x_i^0(k)$ is almost surely bounded by C, it satisfies the bounded difference inequality,

$$|\mathbb{E}_{W,g}\varphi(x_1^0(1),\dots,x_i^0(k),\dots,x_N^0(\kappa)) - \mathbb{E}_{W,g}\varphi(x_1^0(1),\dots,\tilde{x}_i^0(k),\dots,x_N^0(\kappa))| \leq CL$$

so φ satisfies the bounded difference property for some universal constant L that only depends on the maximal value of the support. From the concentration inequality for functions of independent random variables satisfying the bounded differences property [11, Corollary 3.2] if follows that

$$\mathbb{E}_{x^0}(\mathbb{E}_{W,g}\varphi - \mathbb{E}\varphi)^2 \leq C^2L^2\kappa N.$$

Combining the estimates in (5.24) and (5.25) proves (5.25) after applying Jensen's inequality.

From the overlap concentration estimate Lemma 5.2, we deduce

$$\mathbb{E}_u\mathbb{E}\langle((\lambda_{m,p}^\intercal(R_{1,2}^t)^{\odot p}\lambda_{m,p}) - \mathbb{E}\langle(\lambda_{m,p}^\intercal(R_{1,2}^t)^{\odot p}\lambda_{m,p})\rangle_{\text{pert}})^2\rangle_{\text{pert}} \leq L_{m,p}\left(\left(\frac{N+N\varepsilon_N^2}{N^2\varepsilon_N^4}\right)^{1/3} + \frac{1}{\varepsilon_N^2 N}\right). \tag{5.26}$$

In particular, if $\varepsilon_N = N^{-\gamma}$ for $0<\gamma<1/4$, then $N\varepsilon_N^4\to\infty$ so (5.22) holds for any m,p and t. Since the $\lambda_{m,p}$ are dense in the sphere, the result follows. □

Remark 5.6. The condition that $\gamma>0$ implies that the limit of the free energy is unchanged by Lemma 5.1. The condition that $\gamma<1/4$ is required to ensure that the perturbation is large enough to regularize the Gibbs measure. The exponent $1/4$ is not expected to be optimal.

The concentration of the quadratic forms in Corollary 5.5 is insufficient to determine concentration of the overlaps (see Corollary 5.8) because the limiting arrays may not be a priori symmetric. Indeed, even though the convergence in Corollary 5.5 can be turned into an almost sure uniform convergence in λ

$$(\lambda^\intercal R_{1,2}^t\lambda) - \mathbb{E}\langle(\lambda^\intercal\mathbb{E}R_{1,2}^t\lambda)\rangle_{\text{pert}} \to 0 \qquad a.s.$$

this allows to conclude only that

$$((R_{1,2}^t)^{\odot p} - \mathbb{E}\langle(R_{1,2}^t)^{\odot p}\rangle_{\text{pert}} + ((R_{1,2}^t)^{\odot p} - \mathbb{E}\langle(R_{1,2}^t)^{\odot p}\rangle_{\text{pert}})^\intercal$$

goes to zero. It is not hard to see that $\mathbb{E}\langle(R_{1,2}^t)^{\odot p}\rangle_{\text{pert}}$ is symmetric, but at this point nothing guarantees that $(R_{1,2}^t)^{\odot p}$ is as well symmetric. To prove asymptotic symmetry of the overlaps, we invoke the synchronization property first observed for overlap matrix arrays in the vector spin glass models.

The entries of the overlap matrix are bounded, the matrix of overlaps is tight and there exists a subsequence such that $R_N^t = (R^t(x^\ell,x^{\ell'}))_{\ell,\ell'\geq 1}$ converges in distribution to some matrix $\bar{R}^t = (\bar{R}(x^\ell,x^{\ell'}))_{\ell,\ell'\geq 1}$. The concentration of the quadratic forms (5.22) proves that $(\lambda_k^\intercal(R_{\ell,\ell'}^t)^{\odot p}\lambda_k)$ is close to $\mathbb{E}\langle(\lambda_k^\intercal(R_{\ell,\ell'}^t)^{\odot p}\lambda_k)\rangle_{\text{pert}}$ for $\ell\neq\ell'$ with probability going to one in the limit. This means that the off diagonal entries of the array of quadratic forms are constant

$$((\lambda_k^\intercal(R_{\ell,\ell'}^t)^{\odot p}\lambda_k))_{\ell\neq\ell'} = L_k^{t,p} \tag{5.27}$$

for some function L that only depends on t, p and λ_k. The fact that the off diagonals of the quadratic forms are constant is sufficient to proving that the limiting array $\bar{R}$ must have symmetric entries. In particular, $\bar{R}$ satisfies the synchronization property proved for vector spin models in [37, Theorem 4].

Theorem 5.7 (*Synchronization*)

Any infinite array $(\bar{R}_{\ell,\ell'}^t)_{\ell,\ell'\geq 1}$ of $\kappa\times\kappa$ matrices that satisfy (5.22) implies that $\bar{R}_{\ell,\ell'}^t$ is almost surely symmetric for all $\ell,\ell'\geq 1$.

Proof. We adapt the general proof of synchronization to this simpler setting where we have concentation of the quadratic forms of the powers of the overlaps. The proof is essentially identical to the general case [38, Section 6], but avoids relying on the Ghirlanda–Guerra identities because they are trivially satisfied by the replica symmetric array.

Step 1: We first recover the values of the overlaps. By (5.27), we observe that there exists some constants $L^t_k, L^t_{k,k'}$ etc that depend only on $t \le n$ and entries k, k' such that :

(1) *First Order Diagonal Elements:* For all $1 \le k \le \kappa$, if we take $p = 1$ and $\lambda = e_k$ then
$$\bar{R}^t_{\ell,\ell'}(k,k) = L^t_{k,k}$$
for some constant $L^t_k = \mathbb{E}\bar{R}^t_{1,2}(k,k)$.

(2) *First Order Off-Diagonal Elements:* For all $1 \le k \ne k' \le \kappa$, if we take $p = 1$ and $\lambda_1 = e_k + e_{k'}$ and $\lambda_2 = e_k - e_{k'}$ then (5.27) implies
$$\bar{R}^t_{\ell,\ell'}(k,k) + \bar{R}^t_{\ell,\ell'}(k',k') + \bar{R}^t_{\ell,\ell'}(k,k') + \bar{R}^t_{\ell,\ell'}(k',k) = L^t_{e_k+e_{k'}}$$
$$\bar{R}^t_{\ell,\ell'}(k,k) + \bar{R}^t_{\ell,\ell'}(k',k') - \bar{R}^t_{\ell,\ell'}(k,k') - \bar{R}^t_{\ell,\ell'}(k',k) = L^t_{e_k-e_{k'}}$$
so
$$\bar{R}^t_{\ell,\ell'}(k,k') + \bar{R}^t_{\ell,\ell'}(k',k) = L^t_{k,k'} = \frac{1}{2}(L^t_{e_k+e_{k'}} - L^t_{e_k-e_{k'}}). \tag{5.28}$$

(3) *Second Order Off-Diagonal Elements:* For $1 \le k \ne k' \le k$, if we take $p = 2$ and $\lambda_1 = e_k + e_{k'}$ and $\lambda_2 = e_k - e_{k'}$ then (5.27) gives a $\tilde{L}^t_{k,k'}$ such that for all $\ell \ne \ell'$
$$(\bar{R}^t_{\ell,\ell'}(k,k'))^2 + (\bar{R}^t_{\ell,\ell'}(k',k))^2 = \tilde{L}^t_{k,k'}. \tag{5.29}$$

From the formulas (5.28) and (5.29) we can explicitly solve the system to conclude that $\bar{R}_{\ell,\ell'}(k,k')$ and $\bar{R}_{\ell,\ell'}(k,k')$ can take one of two possible values
$$\bar{R}^t_{\ell,\ell'}(k,k'), \bar{R}^t_{\ell,\ell'}(k',k) = \frac{L^t_{k,k'} \pm \sqrt{2\tilde{L}^t_{k,k'} - (L^t_{k,k'})^2}}{2}.$$
Note that all the concentration results above only apply for the offdiagonal elements of the overlap array, because Theorem 5.2 only holds when $\ell \ne \ell'$. Obviously L^t and $\tilde{L}^t$ are symmetric matrices. Moreover, by Corollary 5.5, the indices does not depend on the choice of ℓ, ℓ' as the distribution of any limit point should be symmetric in the replicas.

Step 2: It remains to show that
$$\bar{R}^t_{\ell,\ell'}(k,k') = \bar{R}^t_{\ell,\ell'}(k',k) = \frac{L^t_{k,k'}}{2}.$$
We proceed by contradiction by showing that if $\bar{R}^t_{\ell,\ell'}(k,k') \ne \bar{R}^t_{\ell,\ell'}(k',k)$ with positive probability, then we can always examine a large enough subarray such that the diagonals are identical. We consider a $2n \times 2n$ array of 2×2 blocks of the form
$$\begin{bmatrix} \bar{R}_{\ell,\ell'}(k,k) & \bar{R}_{\ell,\ell'}(k,k') \\ \bar{R}_{\ell,\ell'}(k',k) & \bar{R}_{\ell,\ell'}(k',k') \end{bmatrix}_{\ell,\ell'}$$
By the computations in step 1, for $\ell \ne \ell'$ the offdiagonal blocks must be of the form
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \text{ or } \begin{bmatrix} a & c \\ b & d \end{bmatrix}$$
where
$$a = L^t_{k,k} \quad d = L^t_{k',k'} \quad b = \frac{L^t_{k,k'} + \sqrt{2\tilde{L}^t_{k,k'} - (L^t_{k,k'})^2}}{2} \quad c = \frac{L^t_{k,k'} - \sqrt{2\tilde{L}^t_{k,k'} - (L^t_{k,k'})^2}}{2}.$$
We will show that $b \ne c$ is impossible, by examining the barycenters of the infinite arrays, namely using the fact that we have n, the number of replicas, as large as we wish. For large enough n, we can find arbitrary large disjoint set of indices $\ell \in V_1$ and $\ell' \in V_2$ each of cardinality m [23, Theorem 3] such that the corresponding matrix array is only of the first form
$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$
We now restrict ourselves to indices from $V = V_1 \sqcup V_2$. Because the arrays of each of the entries restricted to V is positive semidefinite, we can find vectors u_ℓ and w_ℓ in a Hilbert space such that
$$\begin{bmatrix} \langle u_\ell, u_{\ell'} \rangle & \langle u_\ell, w_{\ell'} \rangle \\ \langle w_\ell, u_{\ell'} \rangle & \langle w_\ell, w_{\ell'} \rangle \end{bmatrix}_{\ell,\ell' \in V} = \begin{bmatrix} \bar{R}_{\ell,\ell'}(k,k) & \bar{R}_{\ell,\ell'}(k,k') \\ \bar{R}_{\ell,\ell'}(k',k) & \bar{R}_{\ell,\ell'}(k',k') \end{bmatrix}_{\ell,\ell' \in V}. \tag{5.30}$$

Notice that for $\ell \neq \ell' \in V$ we have $\langle u_\ell, u_{\ell'}\rangle = a$ and $\langle w_\ell, w_{\ell'}\rangle = d$. Furthermore, by construction if $\ell \in V_1$ and $\ell' \in V_2$, we see that

$$\begin{bmatrix} \langle u_\ell, u_{\ell'}\rangle & \langle u_\ell, w_{\ell'}\rangle \\ \langle w_\ell, u_{\ell'}\rangle & \langle w_\ell, w_{\ell'}\rangle \end{bmatrix}_{\ell\in V_1, \ell'\in V_2} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

For $i = 1, 2$, if we examine the barycenters

$$U_i = \frac{1}{m}\sum_{\ell\in V_i} u_\ell \quad \text{and} \quad W_i = \frac{1}{m}\sum_{\ell\in V_i} w_\ell,$$

then (5.30) readily gives

$$\begin{bmatrix} \langle U_1, U_2\rangle & \langle U_1, W_2\rangle \\ \langle W_1, U_2\rangle & \langle W_1, W_2\rangle \end{bmatrix} = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

However, if we look at the differences between the barycenters, (5.30) implies that

$$\|U_1 - U_2\|^2 = \frac{1}{m^2}\Big\|\sum_{\ell\in V_1} u_\ell - \sum_{\ell\in V_2} u_\ell\Big\|^2 \leq \frac{2C^2 + 2a}{m}$$

and

$$\|W_1 - W_2\|^2 = \frac{1}{m^2}\Big\|\sum_{\ell\in V_1} w_\ell - \sum_{\ell\in V_2} w_\ell\Big\|^2 \leq \frac{2C^2 + 2d}{m}.$$

Indeed, the diagonal entries of the overlap arrays are bounded by some universal constant C^2 and the offdiagonals are fixed for $\ell \in V_1$ and $\ell' \in V_2$ by (5.30). We used the fact that $\langle u_\ell, u_{\ell'}\rangle = a$ and $\langle w_\ell, w_{\ell'}\rangle = d$ for any $\ell \neq \ell'$ to cancel off all the offdiagonal terms. If we take $m \to \infty$, then $U_1 \approx U_2$ and $W_1 \approx W_2$, so $\langle U_1, W_2\rangle = \langle W_1, U_2\rangle$ which implies that $b = c$, so the overlap array must be symmetric. □

In particular, the limiting matrix overlaps $\bar{R}^t$ are almost surely symmetric. This allows us to conclude the concentration of the overlap entries instead of its quadratic forms as previously found in Theorem 5.2, see (5.7).

Theorem 5.8 (*Concentration of the Overlaps*)

If $\varepsilon_N = N^{-\gamma}$ for some $0 < \gamma < 1/4$, then for all $t \leq n$

(1)

$$\mathbb{E}\langle\|R^t_{1,2} - \mathbb{E}\langle R^t_{1,2}\rangle_{\text{pert}}\|_2^2\rangle_{\text{pert}} \to 0.$$

(2)

$$\mathbb{E}\langle\|R^t_{1,0} - \mathbb{E}\langle R^t_{1,0}\rangle_{\text{pert}}\|_2^2\rangle_{\text{pert}} \to 0.$$

Furthermore,

$$Q_t := \mathbb{E}\langle R^t_{1,2}\rangle_{\text{pert}} = \mathbb{E}\langle R^t_{1,0}\rangle_{\text{pert}},$$

so the limit points are the same.

Proof. The first point is a direct consequence of Corollary 5.5 and Theorem 5.7. For the second point we use the Nishimori property to get concentration of the overlap with respect to the planted signal from the first case

$$\mathbb{E}\langle\|R_{1,0} - \mathbb{E}\langle R_{1,0}\rangle_{\text{pert}}\|_2^2\rangle_{\text{pert}} \to 0.$$

and

$$\mathbb{E}\langle R^t_{1,2}\rangle_{\text{pert}} = \mathbb{E}\langle R^t_{1,0}\rangle_{\text{pert}}.$$

□

5.2. **Cavity Computations via the Aizenman–Sims–Starr Scheme.** We will now apply the concentration of the overlaps to prove the matching upper bound of the free energy using the Aizenman–Sims–Starr scheme. We have for any $M \geq 1$

$$F(\Delta) := \limsup_{N\to\infty} F_N(\Delta) = \limsup_{N\to\infty} \frac{1}{M}((N+M)F_{N+M}(\Delta) - NF_N(\Delta)).$$

We can partition the cavity coordinates into groups such that

$$I^+ = \{N+1, \dots, N+M\} = \bigcup_{s\leq n} I^+_s,$$

and the proportions in each of these groups converge as $M \to \infty$ for every fixed N (at least on average)

$$\lim_{M\to\infty} \rho_s^M = \frac{|I_s^+|}{M} = \rho_s$$

We decompose the Hamiltonians into its common part and its cavity fields,

$$H_{N+M}(x,y) = H'_N(x) + \sum_{i=1}^M (y_i^\mathsf{T} z_i(x) + y_i^\mathsf{T} m_i(x) + y_i^\mathsf{T} s_i(x) y_i) + o(N)$$

and

$$H_N(x) \stackrel{d}{=} H'_N(x) + y(x)$$

Here, the common Hamiltonian is given by

$$H'_N(x) = \sum_{1\le i<j\le N} \frac{1}{\sqrt{\Delta_{ij}}\sqrt{N+M}} g_{ij} x_i\cdot x_j + \sum_{1\le i<j\le N} \frac{1}{\Delta_{ij}(N+M)} (x_i^0\cdot x_j^0)(x_i\cdot x_j) - \sum_{1\le i<j\le N} \frac{1}{2\Delta_{ij}(N+M)} (x_i\cdot x_j)^2$$

The cavity fields in this model are of the form :

$$\begin{aligned}
z_i(x) &= \sum_{j=1}^N g_{j,N+1} \frac{1}{\sqrt{\Delta_{ij}}\sqrt{N+M}} x_j \\
m_i(x) &= \sum_{j=1}^N \frac{1}{\Delta_{ij}(N+M)} x_j^0 \cdot x_{N+i}^0 x_j = \sum_{s\le n} \frac{\rho_t^M}{\Delta_{st}} R_{1,0}^t x_{N+i}^0 + O(N^{-1}) = \tilde R_{1,0}^i x_i^0 + O(N^{-1}) \qquad i\in I_s^+ \\
s_i(x) &= -\sum_{j=1}^N \frac{1}{2\Delta_{ij}(N+M)} x_j x_j^\mathsf{T} = -\frac12 \sum_{s\le n} \frac{\rho_t^M}{\Delta_{st}} R_{1,1}^t + O(N^{-1}) = -\frac12 \tilde R_{1,1}^i + O(N^{-1}) \qquad i\in I_s^+
\end{aligned}$$

where we used the notation

$$\tilde R_{1,1}^i = \sum_{s\le n} \frac{\rho_t^M}{\Delta_{st}} R_{1,1}^t = \sum_{s\le n} \frac{\rho_t^M}{\Delta_{st}} R^t(\boldsymbol{x},\boldsymbol{x}) \qquad \tilde R_{1,0}^i = \sum_{s\le n} \frac{\rho_t^M}{\Delta_{st}} R_{1,0}^t = \sum_{s\le n} \frac{\rho_t^M}{\Delta_{st}} R^t(\boldsymbol{x},\boldsymbol{x}^0).$$

Noting that we can write that the Gaussian variables with variance $1/N$ decompose as $g_{ij}/\sqrt{(N+M)} + g'_{ij}\sqrt{M/N(N+M)}$ with independent Gaussian variables g_{ij} and g'_{ij}

$$\begin{aligned}
y(x) &= \sum_{1\le i<j\le N} \frac{\sqrt{M}}{\sqrt{\Delta_{ij}}\sqrt{N(N+M)}} g'_{ij} x_i\cdot x_j + \sum_{i<j} \frac{M}{\Delta_{ij}N(N+M)} (x_i^0\cdot x_j^0)(x_i\cdot x_j) - \sum_{i<j} \frac{M}{2\Delta_{ij}N(N+M)} (x_i\cdot x_j)^2 \\
&= \sum_{i<j} \frac{\sqrt{M}}{\Delta_{ij}N} g'_{ij} x_i\cdot x_j + \frac{M}{2}\sum_{st} \frac{\rho_s^M\rho_t^M}{\Delta_{st}} \langle R_{1,0}^s R_{1,0}^t\rangle - \frac{M}{4}\sum_{st} \frac{\rho_s^M\rho_t^M}{\Delta_{st}} \langle R_{1,1}^s R_{1,1}^t\rangle + O(N^{-1})
\end{aligned}$$

where g'_{ij} is independent of g_{ij}. Notice that the covariance of the Gaussian fields are given by

$$\mathbb{E} z_i(x^1) z_i(x^2)^\mathsf{T} = \sum_t \frac{\rho_t^M}{\Delta_{s,t}^2} R_{1,2}^t := \tilde R_{1,2}^i \qquad i\in I_s^+ \tag{5.31}$$

and

$$\mathbb{E}\left(\left(\sum_{i<j} \frac{1}{\sqrt{\Delta_{ij}}\sqrt{(N+M)}} g_{ij} x_i^1\cdot x_j^1\right)\left(\sum_{i<j} \frac{1}{\sqrt{\Delta_{ij}}\sqrt{(N+M)}} g_{ij} x_i^2\cdot x_j^2\right)\right) = \frac{M}{2}\sum_{s,t} \frac{\rho_s^M\rho_t^M}{\Delta_{st}} \langle R_{1,2}^s, R_{1,2}^t\rangle.$$

By adding and subtracting the normalization terms H'_N, we need to compute

$$\Delta F_{N,M} := \frac{1}{M}\left(\mathbb{E}\log\left\langle \int \exp\left(\sum_{i=1}^M y_i^\mathsf{T} z_i(x) + y_i^\mathsf{T} m_i(x) + y_i^\mathsf{T} s_i(x) y_i\right) dP_0(y)\right\rangle' - \mathbb{E}\log\langle \exp(y(x))\rangle'\right)$$

where $\langle\cdot\rangle'$ denotes the average with respect to the Gibbs measure H'_N. Of course, this procedure works as well if we added the perturbation terms in Theorem 5.2, resulting in

$$\Delta F_{N,M}^{\text{pert}} = \frac{1}{M}\left(\mathbb{E}\log\left\langle \int \exp\left(\sum_{i=1}^M y_i^\mathsf{T} z_i(x) + y_i^\mathsf{T} m_i(x) + y_i^\mathsf{T} s_i(x) y_i\right) dP_0(y)\right\rangle_{\text{pert}} - \mathbb{E}\log\langle \exp(y(x))\rangle_{\text{pert}}\right) + o(1) \tag{5.32}$$

after a straightforward interpolation argument (see for example [34, Chapter 3.5]). It suffices to compute this quantity in the limit.

We now recall a modification of a general result [34, Theorem 1.4] that implies that the functional $\Delta F_{N,M}$ is continuous with respect to the distribution of the off-diagonal elements of the array $(R_{\ell,\ell'})_{\ell\neq\ell'}$. We now state precisely what we mean.

Lemma 5.9 (*Continuity with Respect to Overlap Distribution*)

For every $\varepsilon > 0$, there exists a function $F_\varepsilon(R^n_{\neq})$ of finitely many elements of $R^n_{\neq} = (R_{\ell,\ell'})_{1\leq\ell\neq\ell'\leq n}$ such that

$$|\Delta F^{\text{pert}}_{N,M} - F_\varepsilon(R^n_{\neq})| < \varepsilon.$$

Proof. Recall that without loss of generality, we can assume that P_0 is supported on finitely many points. The proof is essentially identical to the proof of in [34, Theorem 1.4]. The key difference is that the covariance structures of the cavity fields in these models imply that the diagonal self overlap terms cancel, leaving us with a functional of the off diagonal terms of the overlaps.

By Gaussian concentration, we will be able to restrict the logarithms in $\Delta F_{N,M}$ to a compact set without introducing a large error. We first define the truncation

$$f_a(x) = \begin{cases} -a & x < -a \\ x & -a \leq x \leq a \\ a & x > a \end{cases}$$

and the corresponding truncated Hamiltonians

$$H^a_Z(x,y) = f_a\Big(\sum_{i=1}^M y_i^\intercal z_i(x) + y_i^\intercal \tilde{R}^i_{1,0} x^0_i - \frac{1}{2} y_i^\intercal \tilde{R}^i_{1,1} y_i\Big)$$

$$H^a_Y(x) = f_a(\sqrt{M} y(x^\ell) + \frac{M}{2}\sum_{st}\frac{\rho^M_s\rho^M_t}{\Delta_{st}}\langle R^s_{1,0}R^t_{1,0}\rangle - \frac{M}{4}\sum_{st}\frac{\rho^M_s\rho^M_t}{\Delta_{st}}\langle R^s_{1,1}R^t_{1,1}\rangle).$$

Notice that H^∞_Z and H^∞_Y are the corresponding untruncated cavity field Hamiltonians appearing in $\Delta F_{N,M}$. To simplify notation, we use $\langle\cdot\rangle = \langle\cdot\rangle_{\text{pert}}$. By a standard Gaussian concentration argument, we will prove that there is a constant c such that

$$\Big|\frac{1}{M}\mathbb{E}\log\Big\langle\int e^{H^\infty_Z(x,y)}\,dP_0(y)\Big\rangle - \frac{1}{M}\mathbb{E}\log\Big\langle\int e^{H^a_Z(x,y)}\,dP_0(y)\Big\rangle\Big| \leq e^{-ca^2} \tag{5.33}$$

and

$$\Big|\frac{1}{M}\Big(\mathbb{E}\log\Big\langle\int\exp\Big(\sum_{i=1}^M y_i^\intercal z_i(x) + y_i^\intercal m_i(x) + y_i^\intercal s_i(x) y_i\Big)\,dP_0(y)\Big\rangle - \frac{1}{M}\mathbb{E}\log\langle e^{H^a_Y(x)}\rangle\Big| \leq e^{-ca^2} \tag{5.34}$$

for all a sufficiently large. We will defer the proof of this fact to the end of the proof, and first explain the rest of the logic.

The key observation is that by Weierstrass Theorem, we can approximate the logarithm by a polynomial uniformly on a compact set, so we can approximate $\Delta F_{N,M}$ as a linear combination of the moments of the overlaps. Since $H_Z(x)$ and $H_Y(x)$ is bounded, we can restrict the logarithm to the set $[e^{-a}, e^a]$ and approximate it by polynomials. The expectations of these polynomials are of sums of terms of the form

$$\mathbb{E}\Big\langle\int e^{H_Z(x,y)}\,dP_0(y)\Big\rangle^r \quad\text{and}\quad \mathbb{E}\langle e^{H_Y(x)}\rangle^r = \Big\langle\mathbb{E}\prod_{\ell\leq r} e^{H_Y(x)}\Big\rangle$$

We start by computing the first of these terms. These moments can be computed explicitly. Because of (5.33) and (5.33) it suffices to compute the moments with respect to the untruncated Hamiltonians. In particular, for

any $r > 0$, the moments simplify to

$$\begin{aligned}
&\mathbb{E}\Big\langle \int \exp\Big(\sum_{i=1}^M y_i^\mathsf{T} z_i(x) + y_i^\mathsf{T}\tilde{R}^i_{1,0}x_i^0 - \frac{1}{2}y_i^\mathsf{T}\tilde{R}^i_{1,1}y_i\Big)\,dP_0(y)\Big\rangle^r \\
&= \mathbb{E}\Big\langle \prod_{\ell\le r}\int \exp\Big(\sum_{i=1}^M (y_i^\ell)^\mathsf{T} z_i(x^\ell) + (y_i^\ell)^\mathsf{T}\tilde{R}_{\ell,0}x_i^0 - \frac{1}{2}(y_i^\ell)^\mathsf{T}\tilde{R}_{\ell,\ell}y_i^\ell\Big)\,dP_0(y^\ell)\Big\rangle \\
&= \Big\langle \mathbb{E}\prod_{\ell\le r}\int \exp\Big(\sum_{i=1}^M (y_i^\ell)^\mathsf{T} z_i(x^\ell) + (y_i^\ell)^\mathsf{T}\tilde{R}_{\ell,0}x_i^0 - \frac{1}{2}(y_i^\ell)^\mathsf{T}\tilde{R}_{\ell,\ell}y_i^\ell\Big)\,dP_0(y^\ell)\Big\rangle
\end{aligned}$$

where $\langle\cdot\rangle = \langle\cdot\rangle_{\mathrm{pert}}$ is the average on with respect to the perturbed Gibbs measure. Taking expectations with respect to z_i

$$\begin{aligned}
&\mathbb{E}\exp\Big(\sum_{\ell\le r}\sum_{i=1}^M\Big((y_i^\ell)^\mathsf{T} z_i(x^\ell) + (y_i^\ell)^\mathsf{T}\tilde{R}_{\ell,0}x_i^0 - \frac{1}{2}(y_i^\ell)^\mathsf{T}\tilde{R}_{\ell,\ell}y_i^\ell\Big)\Big) \\
&= \mathbb{E}\exp\Big(\sum_{\ell\neq\ell'}\sum_{i=1}^M\Big(\frac{1}{2}(y_i^\ell)^\mathsf{T}\tilde{R}^i_{\ell,\ell'}y_i^{\ell'} + (y_i^\ell)^\mathsf{T}\tilde{R}_{\ell,0}x_i^0\Big)\Big)
\end{aligned}$$

since

$$\mathbb{E}\Big(\sum_{\ell\le r}\sum_{i=1}^M (y_i^\ell)^\mathsf{T} z_i(x^\ell)\Big)\Big(\sum_{\ell\le r}\sum_{i=1}^M (y_i^{\ell'})^\mathsf{T} z_i(x^{\ell'})\Big) = \frac{1}{2}\sum_{\ell,\ell'=1}^r\sum_{i,j=1}^M (y_i^\ell)^\mathsf{T}\tilde{R}^i_{\ell,\ell'}y_j^{\ell'}.$$

In particular, these terms only depend on the offdiagonal elements of $R_{\ell,\ell'}$ and $R_{0,\ell}$, because the diagonal terms canceled off. A similar computation works for the second term in $\Delta F_{N,M}$ since again

$$\begin{aligned}
&\mathbb{E}\exp\Big(\sum_{\ell\le r}\sqrt{M}y(x^\ell) + \frac{M}{2}\sum_{st}\frac{\rho_s^M\rho_t^M}{\Delta_{st}}\langle R^s_{1,0}R^t_{1,0}\rangle - \frac{M}{4}\sum_{st}\frac{\rho_s^M\rho_t^M}{\Delta_{st}}\langle R^s_{1,1}R^t_{1,1}\rangle\Big) \\
&= \mathbb{E}\exp\Big(\sum_{\ell\neq\ell'} -\frac{M}{4}\sum_{st}\frac{\rho_s^M\rho_t^M}{\Delta_{st}}\langle R^s_{\ell,\ell'}R^t_{\ell,\ell'}\rangle + \frac{M}{2}\sum_{st}\frac{\rho_s^M\rho_t^M}{\Delta_{st}}\operatorname{Tr}(R^s_{\ell,0}R^t_{\ell,0}))\Big)
\end{aligned}$$

is a function of the off diagonal elements of the array $(R_{\ell,\ell'})_{\ell,\ell'\ge 0}$. Therefore, we can approximate $\Delta F_{N,M}$ with a continuous function of finitely many off diagonal elements of the array.

All that remains is to prove that the logarithm can be approximated by polynomials by proving the bounds (5.33) and (5.34). The proofs of both inequalities follow from the same argument, so we only show the first one. We prove this by showing that

$$\Big|\frac{1}{M}\mathbb{E}\log\Big\langle\int e^{H_Z^\infty(x,y)}\,dP(y)\Big\rangle - \frac{1}{M}\mathbb{E}\log_a\Big\langle\int e^{H_Z^\infty(x,y)}\,P_X(y)\Big\rangle\Big| \le e^{-ca^2} \tag{5.35}$$

and

$$\Big|\frac{1}{M}\mathbb{E}\log_a\Big\langle\int e^{H_Z^\infty(x,y)}\,dP(y)\Big\rangle - \frac{1}{M}\mathbb{E}\log_a\Big\langle\int e^{H_Z^a(x,y)}\,P_X(y)\Big\rangle\Big| \le e^{-ca^2} \tag{5.36}$$

where $\log_a(x) = f_a(\log(x))$ is the truncated logarithm. We begin by proving the first inequality. Because the support of P_X is bounded, the covariance

$$\mathbb{E}z_i(x^1)z_i(x^2)^\mathsf{T} = \tilde{R}^i_{1,2}$$

defined in (5.31) is almost surely bounded by some universal constant. Let

$$F_Z = \frac{1}{M}\log\Big\langle\int e^{H_Z^\infty(x,y)}\,dP(y)\Big\rangle = \frac{1}{M}\log\Big\langle\int\exp\Big(\sum_{i=1}^M y_i^\mathsf{T} z_i(x) + y_i^\mathsf{T} m_i(x) + y_i^\mathsf{T} s_i(x)y_i\Big)\,dP_0(y)\Big\rangle.$$

By Gaussian concentration [34, Theorem 1.2],

$$P(|F_Z - \mathbb{E}F_Z| \ge a) \le 2e^{-ca^2} \tag{5.37}$$

where c only depends on $\sup_{t\le n}\sup_{x\in\operatorname{supp} P_X}\|\tilde{R}^t_{1,2}\|$. By Jensen's inequality, and the concavity of the logarithm

$$0 \le \mathbb{E}F_Z \le \frac{1}{M}\log\left\langle \mathbb{E}\int \exp\left(\sum_{i=1}^M y_i^\intercal z_i(x) + y_i^\intercal m_i(x) + y_i^\intercal s_i(x) y_i\right) dP_0(y)\right\rangle \le C$$

for some constant that only depends on the upper bound of the covariance $\sup_{t\le n}\sup_{x\in\operatorname{supp} P_X}\|\tilde{R}^t_{1,2}\|$. This implies that $\{|F_Z|\ge a\}\subset\{|F_Z-\mathbb{E}F_Z|\ge a/2\}$ for $a\ge C$. The Gaussian concentration inequality (5.37) implies that there exists a constant c such that for all a sufficiently large

$$P(|F_Z|\ge a)\le e^{-ca^2}$$

and the bounds on the tails of a sub-Gaussian random variable [34, Equation 1.80] implies that

$$\mathbb{E}(|F_Z|\,\mathbb{1}(|F_Z|\ge a))\le e^{-ca^2}$$

for some universal constant c. This proves (5.35).

To prove (5.36), notice that $|\log_a(x)-\log_a(y)|\le e^a|x-y|$ so (5.36) is bounded by

$$e^a\left|\left\langle\int e^{H_Z^\infty(x,y)}P(y)-\int e^{H_Z^a(x,y)}P(y)\right\rangle\right|\le e^a\mathbb{E}\langle|H_Z^\infty(x,y)|\,\mathbb{1}(|H_Z^\infty(x,y)|\ge a)\rangle.$$

Since the H_Z is a Gaussian process, the exponential decay of the tails implies that we can also bound this by e^{-ca^2} for a sufficiently large. □

Remark 5.10. The fact that the functionals do not depend on the diagonal terms is essential to understanding the limiting behavior, because concentration Corollary 5.8 only applies to the offdiagonal terms of the array.

In the limit, we use synchronization of vector spins and multispecies models to study the limiting behavior of the overlaps under the asymptotic Gibbs measure. In our case, Theorem 5.8 implies that

$$R^s_{\ell,\ell'}=Q^s_{\ell,\ell'}=Q^s_{1,0}=Q^s \qquad \text{for all } s\le n, \ell\ne\ell'$$

where $Q_{\ell,\ell'}$ is the values of the limiting overlap matrix. One can check that the free energy functional (5.32) in the limit is equivalent to

$$\begin{aligned}&\frac{1}{M}\left(\mathbb{E}\log\int\exp\left(\sum_{i=1}^M y_i^\intercal\tilde{Q}_i^{1/2}z_i+y_i^\intercal\tilde{Q}_i x_i^0-\frac{1}{2}y_i^\intercal\tilde{Q}_i y_i\right)dP_0(y)\right.\\ &\left.-\mathbb{E}\log\exp\left(y+\frac{M}{4}\sum_{st}\frac{\rho_s^M\rho_t^M}{\Delta_{st}}\langle Q_s,Q_t\rangle\right)\right).\end{aligned}\tag{5.38}$$

where

$$\tilde{Q}_i=\tilde{Q}_s=\sum_{t\le n}\frac{1}{\Delta_{s,t}}\rho_t^M Q_t \qquad i\in I_s.$$

and the Gaussians z and y have covariance

$$\operatorname{Cov}(z_i,z_j)=\delta_{j=i}Q_i \quad \operatorname{Var}(y)=\frac{M}{2}\sum_{st}\frac{\rho_s^M\rho_t^M}{\Delta_{st}}\langle Q^s,Q^t\rangle$$

Lemma 5.11 (*Equivalence of the Functionals of the Overlaps*)

If the overlaps concentrate, then (5.32) and (5.38) can be approximated by the same functionals $F_\varepsilon(R^n_{\ne})$ of finitely many off diagonal entries of the overlap array.

Proof. In our setting, we can replace the off diagonal entries of the array $(R_{\ell\ne\ell'})$ and Q, since the elements of the off diagonal array take only one value by concentration. The original cavity fields appear to have some terms that depend on the diagonal terms $R_{\ell,\ell}$, but we will show that the moments of the cavity field are independent of this term by cancellation.

Since P_0 can be assumed to have finite support by Lemma 3.10, the computation in the proof of Lemma 5.9 implies that for fixed y and x^0 the moments of the first cavity field can be written as a weighted sum of terms of the form

$$\mathbb{E}\exp\left(\sum_{i=1}^M (y^\ell)_i^\intercal z_i(x)+(y^\ell)_i^\intercal m_i(x)+(y^\ell)_i^\intercal s_i(x)(y^\ell)_i\right)\right)=\exp\left(\sum_{\ell\ne\ell'}^r\sum_{i=1}^M\frac{1}{2}y_i^\intercal\tilde{R}^i_{\ell,\ell'}y_j+y_i^\intercal\tilde{R}^i_{0,\ell}x_i^0\right)$$

where

$$\tilde{R}^i_{\ell,\ell'} = \sum_t \frac{\rho^M_t}{\Delta_{s,t}} R^t_{1,2} \qquad \text{for } i \in I_s.$$

When compared to

$$\begin{aligned}
&\mathbb{E}\exp\Big(\sum_{\ell=1}^r\sum_{i=1}^M (y^\ell)_i^\intercal z_i + (y^\ell)_i^\intercal \tilde{Q}_i x_i^0 - \frac{1}{2}(y^\ell)_i^\intercal \tilde{Q}_i (y^\ell)_i\Big)\\
&= \exp\Big(\sum_{\ell,\ell'=1}^r\sum_{i=1}^M \frac{1}{2}(y^\ell)_i^\intercal \tilde{Q}_i (y^\ell)_i + y_i^\intercal \tilde{Q}_i x_i^0 - \frac{1}{2}(y^\ell)_i^\intercal \tilde{Q}_i (y^\ell)_i\Big)\\
&= \exp\Big(\sum_{1\le \ell\neq\ell'\le r}\sum_{i=1}^M \frac{1}{2}(y^\ell)_i^\intercal \tilde{Q}_i (y^\ell)_i + y_i^\intercal \tilde{Q}_i x_i^0 - \frac{1}{2}(y^\ell)_i^\intercal \tilde{Q}_i (y^\ell)_i\Big)
\end{aligned}$$

where

$$\tilde{Q}_i = \sum_{t\le n} \frac{\rho^M_t}{\Delta_{s,t}} Q^t \qquad i \in I_s$$

we can conclude that the moments are functionals of the same values of the overlap.

For the second cavity field, it is easy to see that all moments are the same of the overlaps. A direct computation of the moments using the moment generating function of a Gaussian gives

$$\begin{aligned}
\mathbb{E}\Big(\exp(y(x))\Big)^r &= \exp\Big(\sum_{\ell=1}^r y(x^\ell) + \frac{M}{2}\sum_{st}\frac{\rho^M_s\rho^M_t}{\Delta_{st}}\langle R^s_{\ell,0}, R^t_{\ell,0}\rangle - \frac{M}{4}\sum_{st}\frac{\rho^M_s\rho^M_t}{\Delta_{st}}\langle R^s_{\ell,\ell}, R^t_{\ell,\ell}\rangle\Big)\\
&= \exp\Big(\sum_{1\le\ell\neq\ell'\le r}\frac{M}{4}\sum_{st}\frac{\rho^M_s\rho^M_t}{\Delta_{st}}\langle R^s_{\ell,\ell'}, R^t_{\ell,\ell'}\rangle + \frac{M}{2}\sum_{st}\frac{\rho^M_s\rho^M_t}{\Delta_{st}}\langle R^s_{\ell,0}, R^t_{\ell,0}\rangle\Big).
\end{aligned}$$

and when compared to

$$\begin{aligned}
&\mathbb{E}\exp\Big(\sum_{\ell=1}^r y + \frac{M}{4}\sum_{st}\frac{\rho^M_s\rho^M_t}{\Delta_{st}}\langle Q^s, Q^t\rangle\Big)\\
&= \exp\Big(\sum_{1\le\ell\neq\ell'\le r}\frac{M}{8}\sum_{st}\frac{\rho^M_s\rho^M_t}{\Delta_{st}}\langle Q^s, Q^t\rangle + \frac{M}{4}\sum_{st}\frac{\rho^M_s\rho^M_t}{\Delta_{st}}\langle Q^s, Q^t\rangle\Big)
\end{aligned}$$

so both functionals are the same functions of the overlaps $(R_{\ell,\ell'})_{\ell\neq\ell'\ge 0}$ and Q.

□

By overlap concentration Theorem 5.8, we know $(R^t_{\ell,\ell'})_{\ell\neq\ell'} \to (Q^t)_{\ell\neq\ell'}$ for some constant matrix Q^t in the limit. Lemma 5.9 implies that the convergence of the offidagonal elements of the array is sufficient and Lemma 5.11 implies that the functional (5.38) characterizes the limiting behavior of $\Delta F_{N,M}$ defined in (5.32). That is,

$$\begin{aligned}
&\lim_{N\to\infty}\Delta F^{\text{pert}}_{N,M} = F_\varepsilon(R^n_{\neq}) + O(\varepsilon)\\
&= \frac{1}{M}\Big(\mathbb{E}\log\int\exp\Big(\sum_{i=1}^M y_i^\intercal \tilde{Q}_i^{1/2} z_i + y_i^\intercal \tilde{Q}_i x_i^0 - \frac{1}{2}y_i^\intercal \tilde{Q}_i y_i\Big)\,dP_0(y) - \mathbb{E}\log\exp\Big(y + \frac{M}{4}\sum_{st}\frac{\rho^M_s\rho^M_t}{\Delta^2_{st}}\langle Q^s, Q^t\rangle\Big) + O(\varepsilon)
\end{aligned}$$

(5.39)

$$= \frac{1}{M}\Big(\sum_{i\in I_s^+}\mathbb{E}\log\int\exp\Big(y_i^\intercal \tilde{Q}_s^{1/2} z_s + y_i^\intercal \tilde{Q}_s x_i^0 - \frac{1}{2}y_i^\intercal \tilde{Q}_s y_i\Big)\,dP_0(y) - \frac{M}{4}\sum_{st}\frac{\rho^M_s\rho^M_t}{\Delta^2_{st}}\langle Q^s, Q^t\rangle\Big) + O(\varepsilon)$$

The parameters $Q_s = Q^M_s$ appearing above depend on our choice of M.

The lower bound (5.39) holds for all M. By compactness, there exists a subsequence such that Q^M_s converges to a limiting object Q. We may take $M\to\infty$ along a subsequence such that the proportions converge $\frac{|I_s^+|}{M}\to\rho_s$ and Q^M_s converges, then $\varepsilon\to 0$ and use the continuity of our functional in Q to conclude that

$$\lim_{M\to\infty}\lim_{N\to\infty}\Delta F_{N,M} = \varphi(Q)$$

for some sequence $Q = (Q^s)_{s\le n}$. This gives the matching upper bound.

Theorem 5.12 (*Bayes Optimal Upper Bound of the Free Energy*)

We have

$$\limsup_{N\to\infty} \frac{1}{N}\mathbb{E}_Y \ln Z_X(Y) = \limsup_{N\to\infty} \frac{1}{N}\mathbb{E}_Y \ln Z_X^{\text{pert}}(Y) \leq \sup_Q \varphi(Q).$$

6. Solving the Variational Problem

In this section, we examine the stability of the critical points of the functional describing the limit of the free energy

$$\varphi(Q) = -\sum_{s,t\leq n} \frac{\rho_s\rho_t}{4\Delta_{s,t}} \operatorname{Tr}(Q_sQ_t) + \sum_{s\leq n} \rho_s \mathbb{E}_{z,x^0} \ln\left[\int \exp\left(\left(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s}z\right)^{\mathsf{T}} x - \frac{x^{\mathsf{T}}\tilde{Q}_s x}{2}\right) dP_0(x)\right] \tag{6.1}$$

where

$$\tilde{Q}_s = \sum_{t\leq n} \frac{1}{\Delta_{s,t}}\rho_t Q_t.$$

We begin with the case when the prior distribution is Gaussian, which will allow us to explicitly compute a closed form of the solution. Later, we will generalize this analysis to general bounded priors.

6.1. Standard Gaussian Prior.

Suppose that the prior is a standard Gaussian on $\mathbb{R}^\kappa$,

$$dP_0(x) = \frac{1}{(2\pi)^{\frac{\kappa}{2}}} e^{-\frac{x\cdot x}{2}}.$$

In this case, the second term in the functional (6.1) can be computed explicitly,

$$\begin{aligned}
&\sum_{s=1}^n \rho_s \mathbb{E}_{z,x^0} \ln\left[\int \exp\left(\left(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s}z\right)^{\mathsf{T}} x - \frac{x^{\mathsf{T}}\tilde{Q}_s x}{2}\right) dP_0(x)\right] \\
&= \sum_{s=1}^n \rho_s \mathbb{E}_{z,x^0} \ln\left[\frac{1}{(2\pi)^{\kappa/2}}\int \exp\left(\left(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s}z\right)^{\mathsf{T}} x - \frac{x^{\mathsf{T}}\tilde{Q}_s x}{2} - \frac{x^{\mathsf{T}} I x}{2}\right) dx\right] \\
&= \sum_{s=1}^n \rho_s \mathbb{E}_{z,x^0} \ln\left[\frac{1}{(2\pi)^{\kappa/2}}\int \exp\Big(-\frac{1}{2}\Big(x - (\tilde{Q}_s+I)^{-1}(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s}z)\Big)^{\mathsf{T}}(\tilde{Q}_s+I)\Big(x - (\tilde{Q}_s+I)^{-1}(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s}z)\Big)\right. \\
&\qquad \left. + \frac{1}{2}\Big(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s}z\Big)^{\mathsf{T}}(\tilde{Q}_s+I)^{-1}\Big(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s}z\Big)\Big) dx\right] \\
&= \sum_{s=1}^n \rho_s \Big(-\frac{1}{2}\ln\det(\tilde{Q}_s+I) + \frac{1}{2}\mathbb{E}_{z,x^0}\Big(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s}z\Big)^{\mathsf{T}}(\tilde{Q}_s+I)^{-1}\Big(\tilde{Q}_s x^0 + \sqrt{\tilde{Q}_s}z\Big)\Big) \\
&= \sum_{s=1}^n \rho_s \Big(-\frac{1}{2}\ln\det(\tilde{Q}_s+I) + \frac{1}{2}\operatorname{Tr}\Big(\tilde{Q}_s(\tilde{Q}_s+I)^{-1}\tilde{Q}_s\Big) + \frac{1}{2}\operatorname{Tr}\Big(\sqrt{\tilde{Q}_s}(\tilde{Q}_s+I)^{-1}\sqrt{\tilde{Q}_s}\Big)\Big).
\end{aligned}$$

In the last computation, we used the fact that x^0, z are independent centered standard Gaussians in $\mathbb{R}^\kappa$. By adding and subtracting the identity matrix and using the fact that the trace is invariant under cyclic permutations,

$$\begin{aligned}
&\frac{1}{2}\operatorname{Tr}\Big(\tilde{Q}_s(\tilde{Q}_s+I)^{-1}\tilde{Q}_s\Big) + \frac{1}{2}\operatorname{Tr}\Big(\sqrt{\tilde{Q}_s}(\tilde{Q}_s+I)^{-1}\sqrt{\tilde{Q}_s}\Big) \\
&= \frac{1}{2}\operatorname{Tr}\Big(\tilde{Q}_s(\tilde{Q}_s+I)^{-1}(\tilde{Q}_s+I-I)\Big) + \frac{1}{2}\operatorname{Tr}\Big(\tilde{Q}_s(\tilde{Q}_s+I)^{-1}\Big) \\
&= \frac{1}{2}\operatorname{Tr}(\tilde{Q}_s).
\end{aligned}$$

Therefore, the general replica symmetric functional φ_g in the standard Gaussian case is

$$\varphi_g(Q) = -\sum_{s,t=1}^n \frac{\rho_s\rho_t}{4\Delta_{s,t}} \operatorname{Tr}(Q_sQ_t) + \sum_{s=1}^n \frac{\rho_s}{2}\Big(\operatorname{Tr}(\tilde{Q}_s) - \ln\det(\tilde{Q}_s+I)\Big). \tag{6.2}$$

where

$$\tilde{Q}_s = \sum_{t=1}^n \frac{1}{\Delta_{s,t}}\rho_t Q_t$$

We next investigate the maximizers of φ_g on the set of symmetric positive semidefinite matrices. Observe that φ_g is a C^∞ function that goes to $-\infty$ when $\sup_s \|Q_s\|_{op} \to \infty$ hence the supremum is attained. We are interested in finding conditions on $\Delta_{s,t}$ and ρ_s to determine when $\varphi(Q)$ has a maximizer at the origin $Q = 0$. We define the following block matrices indexed by $s \le n$,

$$\tilde{Q} = (\tilde{Q}_s)_{s\le n} \quad \frac{1}{\Delta} = \Big(\frac{1}{\Delta_{s,t}}\Big)_{s,t\le n} \quad \rho = \mathrm{diag}(\rho_1, \dots, \rho_n). \tag{6.3}$$

Lemma 6.1 (*Phase Transition with Gaussian Prior*)

(1) The functional φ has a unique maximizer at $Q = 0$ if

$$\Big\|\sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho}\Big\|_{op} < 1$$

(2) The functional φ achieves its maximum value away from $Q = 0$ if

$$\Big\|\sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho}\Big\|_{op} > 1$$

Remark 6.2. In the case when $n = 1$, we have that the phase transition happens when $\frac{1}{\Delta} = 1$, which is precisely the phase transition in that model [29].

Proof.

First and Second Variations: Consider an arbitrary perturbation of $M = (M_1, \dots, M_n)$ of Q. The first and second variation is denoted by

$$\nabla_M \varphi_g(Q) = \partial_\varepsilon \varphi_g(Q + \varepsilon M)\Big|_{\varepsilon=0} \quad \text{and} \quad \nabla^2_M \varphi_g(Q) = \partial^2_\varepsilon \varphi_g(Q + \varepsilon M)\Big|_{\varepsilon=0}. \tag{6.4}$$

We can compute these directly to see that

$$\nabla_M \varphi_g(Q) = \Big(-\sum_{s,u} \frac{\rho_s\rho_u}{2\Delta_{s,u}} \mathrm{Tr}(Q_u M_s) + \sum_{s,u} \frac{\rho_s\rho_u}{2\Delta_{s,u}} \mathrm{Tr}((I - (\tilde{Q}_u + I)^{-1})M_s)\Big)$$

and

$$\nabla^2_M \varphi_g(Q) = -\sum_{s,t} \frac{\rho_s\rho_t}{2\Delta_{s,t}} \mathrm{Tr}(M_s M_t) + \sum_{s,t}\sum_u \frac{\rho_s\rho_t\rho_u}{2\Delta_{s,u}\Delta_{u,t}} \mathrm{Tr}((\tilde{Q}_u + I)^{-1} M_s (\tilde{Q}_u + I)^{-1} M_t).$$

Small Operator Norm: We first find a sufficient condition for φ_g to have a unique maximizer at the origin. At the critical point, the derivative is equal to 0 for all symmetric M, so we can conclude that

$$\tilde{Q}_s = \sum_u \frac{\rho_u}{\Delta_{s,u}} Q_u = \sum_u \frac{\rho_u}{\Delta_{s,u}} \big(I - (\tilde{Q}_u + I)^{-1}\big) \qquad \forall s. \tag{6.5}$$

Consider the following vectors of $\kappa \times \kappa$ matrices

$$\tilde{Q} = (\tilde{Q}_s)_{s\le n} \quad f(\tilde{Q}) = (I - (\tilde{Q}_s + I)^{-1})_{s\le n} = (\tilde{Q}_s(\tilde{Q}_s + I)^{-1})_{s\le n}.$$

The critical point condition can be simplified to

$$\tilde{Q} = \frac{1}{\Delta}\rho f(\tilde{Q}) \implies \sqrt{\rho}\tilde{Q} = \Big(\sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho}\Big)(\sqrt{\rho}f(\tilde{Q})) = A(\sqrt{\rho}f(\tilde{Q})). \tag{6.6}$$

where $A = \sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho}$. We therefore can compute the L^2 norm

$$\|\sqrt{\rho}\tilde{Q}\|_2^2 = \mathrm{Tr}\left((\sqrt{\rho}\tilde{Q})^{\mathsf{T}}(\sqrt{\rho}\tilde{Q})\right) = \mathrm{Tr}\left((A(\sqrt{\rho}f(\tilde{Q}))^{\mathsf{T}}(A(\sqrt{\rho}f(\tilde{Q}))\right) \le \|A\|_{op}\|\sqrt{\rho}f(\tilde{Q})\|_2^2. \tag{6.7}$$

Because $\|\sqrt{\rho}f(\tilde{Q})\|_2 \le \|\sqrt{\rho}\tilde{Q}\|_2\|(\tilde{Q} + I)^{-1}\|_2 \le \|\sqrt{\rho}\tilde{Q}\|_2$, we arrive at

$$\|\sqrt{\rho}\tilde{Q}\|_2^2 \le \|A\|_{op}\|\sqrt{\rho}\tilde{Q}\|_2^2.$$

In particular, when $\|A\|_{op} < 1$ there exists a unique solution to the critical point equation at $\tilde{Q} = 0$.

Large Operator Norm: We examine the Taylor expansion of φ around the origin. The first and second variation simplify greatly at the origin $Q = 0$ to give

$$\nabla_M \varphi_g(0) = 0 \quad \nabla^2_M \varphi_g(0) = \sum_{s,t} \Big(-\frac{\rho_s\rho_t}{2\Delta_{s,t}} \mathrm{Tr}(M_s M_t) + \sum_u \frac{\rho_s\rho_t\rho_u}{2\Delta_{s,u}\Delta_{u,t}} \mathrm{Tr}(M_s M_t)\Big). \tag{6.8}$$

The second variation will always be negative for any choice of M if and only if

$$\rho\frac{1}{\Delta}\rho\frac{1}{\Delta}\rho-\rho\frac{1}{\Delta}\rho=\rho\sqrt{\frac{1}{\Delta}}\left(\sqrt{\frac{1}{\Delta}}\rho\sqrt{\frac{1}{\Delta}}-I\right)\sqrt{\frac{1}{\Delta}}\rho$$

has only negative eigenvalues. By Sylvester's law of inertia, it suffices to study the eigenvalues of the matrix

$$\sqrt{\frac{1}{\Delta}}\rho\sqrt{\frac{1}{\Delta}}-I$$

and by the invariance of the operator norm of cyclic permutations of the matrices, we get the condition that the Hessian is negative semidefinite when

$$\|A\|_{op}=\left\|\sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho}\right\|_{op}<1$$

which is precisely the condition for a unique maximizer at 0. We now claim that when the Hessian has a strictly positive eigenvalue, then there exists a maximizer away from the origin.

Suppose now that $\|A\|_{op}>1$. Then there exists a unit eigenvector $v\in\mathbb{R}^n$ such that $Av=\|A\|_{op}v$. Furthermore, the entries of A are non-negative so the entries of the eigenvector v are also non-negative by the Perron–Frobenius Theorem. Therefore, if we define $M=(u_1I,\dots,u_nI)$ where $u_i=\frac{1}{\sqrt{\rho_i}}v_i\geq 0$ then the Hessian at 0 defined in (6.8) is given by

$$\nabla^2_M\varphi_g(0)=n\left(u^\intercal\rho\frac{1}{\Delta_{s,t}}\rho\frac{1}{\Delta_{s,t}}\rho u-u^\intercal\rho\frac{1}{\Delta_{s,t}}\rho u\right)=n(v^\intercal A^2v-v^\intercal Av)\geq n(\|A\|^2_{op}-\|A\|_{op})$$

which is strictly non-negative if $\|A\|_{op}>1$.

Therefore, the function $g(\varepsilon)=\varphi_g(\varepsilon M)$ satisfies $g''(0)>0$, so it is convex on $[0,\delta]$ for some $\delta>0$. Furthermore, $g'(0)=0$, so we can conclude that for δ small enough

$$\varphi_g(\delta M)=g(\delta)>g(0)=\varphi_g(0)$$

by convexity. Lastly, notice that $M\geq 0$ since the entries of u are non-negative, so M is in the domain of the optimization problem. Observe also that 0 is no longer a maximizer. □

6.2. **General Centered Prior.** Now suppose that we are in the scenario that P_X is a centered prior measure on $\mathbb{R}^\kappa$ with compact support. We want to study the maximizers of the functional (6.1). Let $\langle\cdot\rangle_Q$ denote the average with respect the Gibbs measure associated with the Hamiltonian $(Qx^0+\sqrt{Q}z)^\intercal x-\frac{x^\intercal Qx}{2}$. The partial derivatives of φ in the direction M_s can be computed using Gaussian integration by parts and the Nishimori property (4.4)

$$\begin{aligned}\partial_\varepsilon\varphi(Q+\varepsilon M_s)|_{\varepsilon=0}&=\left(-\sum_{t=1}^n\frac{\rho_s\rho_t}{2\Delta_{s,t}}\operatorname{Tr}(Q_tM_s)+\sum_{t=1}^n\frac{\rho_s\rho_t}{\Delta_{s,t}}\mathbb{E}\left\langle x^\intercal M_sx^0-\frac{x^\intercal M_sx}{2}\right\rangle_{\tilde{Q}_t}\right)\\&\quad+\sum_{t=1}^n\rho_t\mathbb{E}\left\langle z^\intercal(\partial_\varepsilon\sqrt{\tilde{Q}_t+\varepsilon\tilde{M}_t}|_{\varepsilon=0})x\right\rangle\\&=\frac{\rho_s}{2}\left(-\sum_{t=1}^n\frac{\rho_t}{\Delta_{s,t}}\operatorname{Tr}(Q_tM_s)+2\sum_{t=1}^n\frac{\rho_t}{\Delta_{s,t}}\mathbb{E}\left\langle x^\intercal M_sx^0-\frac{x^\intercal M_sx^2}{2}\right\rangle_{\tilde{Q}_t}\right)\\&=\frac{\rho_s}{2}\left(-\sum_{t=1}^n\frac{\rho_t}{\Delta_{s,t}}\operatorname{Tr}(Q_tM_s)+\sum_{t=1}^n\frac{\rho_t}{\Delta_{s,t}}\mathbb{E}\langle x^\intercal M_sx^0\rangle_{\tilde{Q}_t}\right).\end{aligned}\tag{6.9}$$

(6.10)

We dealt with the square root that appeared in the first equality using the identity

$$\operatorname{Tr}\left(A\sqrt{\tilde{Q}_t}\partial_\varepsilon\sqrt{\tilde{Q}_t+\varepsilon\tilde{M}_t}|_{\varepsilon=0}\right)=\operatorname{Tr}\left(A\sqrt{\tilde{Q}_t}\partial_\varepsilon\sqrt{\tilde{Q}_t+\frac{\varepsilon\rho_s}{\Delta_{st}}M_s}|_{\varepsilon=0}\right)=\frac{\rho_s}{2\Delta_{st}}\operatorname{Tr}(AM_s),\tag{6.11}$$

for any symmetric matrix A, which implies that for standard Gaussian vectors z,

$$\mathbb{E}\big(x^\intercal\sqrt{\tilde{Q}_t}z\big)\big(z^\intercal(\partial_\varepsilon\sqrt{\tilde{Q}_t+\frac{\varepsilon\rho_s}{\Delta_{st}}M_s}|_{\varepsilon=0})x^2\big)=\operatorname{Tr}\left(x^2x^\intercal\sqrt{\tilde{Q}_t}\partial_\varepsilon\sqrt{\tilde{Q}_t+\frac{\varepsilon\rho_s}{\Delta_{st}}M_s}|_{\varepsilon=0}\right)=\frac{\rho_s}{2\Delta_{st}}\operatorname{Tr}\left(x^2x^\intercal M_s\right).$$

Therefore, Gaussian integration by parts implies that

$$\begin{aligned}\mathbb{E}\Big\langle z^{\mathsf{T}}(\partial_\varepsilon\sqrt{\tilde{Q}_t+\varepsilon\tilde{M}_s}|_{\varepsilon=0})x\Big\rangle_{\tilde{Q}_t} &= \mathbb{E}\Big\langle \mathbb{E}_z(x^{\mathsf{T}}\sqrt{\tilde{Q}_t}z)(z^{\mathsf{T}}(\partial_\varepsilon\sqrt{\tilde{Q}_t+\frac{\varepsilon\rho_s}{\Delta_{st}}M_s}|_{\varepsilon=0})x)\Big\rangle_{\tilde{Q}_t}\\ &\quad-\mathbb{E}\Big\langle \mathbb{E}_z((x^2)^{\mathsf{T}}\sqrt{\tilde{Q}_t}z)(z^{\mathsf{T}}(\partial_\varepsilon\sqrt{\tilde{Q}_t+\frac{\varepsilon\rho_s}{\Delta_{st}}M_s}|_{\varepsilon=0})x)\Big\rangle_{\tilde{Q}_t}\\ &= \frac{\rho_s}{\Delta_{st}}\mathbb{E}\Big\langle\frac{x^{\mathsf{T}}M_sx}{2}-\frac{x^{\mathsf{T}}M_sx^2}{2}\Big\rangle_{\tilde{Q}_t}\end{aligned}$$

where we used the same convention for the partial derivatives given in (6.4). At the critical point, the first derivative must vanish for all M_s, so the critical point equation simplifies to

$$\tilde{Q}_s=\sum_{t=1}^n\frac{\rho_t}{\Delta_{s,t}}\mathbb{E}\langle x^0x^{\mathsf{T}}\rangle_{\tilde{Q}_t}=\sum_{t=1}^n\frac{\rho_t}{\Delta_{s,t}}\mathbb{E}\langle x\rangle_{\tilde{Q}_t}\langle x\rangle_{\tilde{Q}_t}^{\mathsf{T}}. \tag{6.12}$$

by the Nishimori property. We first prove that the average $f(Q)=\mathbb{E}\langle x^0x^{\mathsf{T}}\rangle_Q$ is Lipschitz.

Lemma 6.3 (*Lipschitz Continuity*)

The functional $f:\mathbb{S}_\kappa^+\to\mathbb{S}_\kappa^+$ is Lipschitz in the space $\mathbb{S}_\kappa^+$ of non-negative $\kappa\times\kappa$ symmetric matrices: for any $Q,Q'\in\mathbb{S}_\kappa^+$,

$$\|f(Q)-f(Q')\|_2\le 3\kappa^2C^3\|Q-Q'\|_2$$

where C is the bound on the support of P_X and κ is the dimension.

Proof. The function $f:\mathbb{S}_\kappa\to\mathbb{R}^{\kappa\times\kappa}$ is also given by

$$f(Q)=\mathbb{E}\frac{\int x^0x^{\mathsf{T}}\exp\left(\left(Qx^0+\sqrt{Q}z\right)^{\mathsf{T}}x-\frac{x^{\mathsf{T}}Qx}{2}\right)dP_X(x)}{\int\exp\left(\left(Qx^0+\sqrt{Q}z\right)^{\mathsf{T}}x-\frac{x^{\mathsf{T}}Qx}{2}\right)dP_X(x)}=\mathbb{E}\langle x^0x^{\mathsf{T}}\rangle_Q.$$

Given arbitrary $Q_1,Q_2\ge 0$ we define $Q_\alpha=\alpha Q_1+(1-\alpha)Q_2=Q_2+\alpha(Q_1-Q_2)$ to be the interpolation between the matrices. We will show that the function

$$g_{k,k'}(\alpha):=f_{k,k'}(\alpha Q_1+(1-\alpha)Q_2)=f_{k,k'}(Q_2+\alpha(Q_1-Q_2))=\mathbb{E}\langle x^0(k)x(k')\rangle_{Q_\alpha}$$

has uniformly bounded derivative for $t\in[0,1]$. Let $M=Q_1-Q_2$. A similar computation as the derivation of the first variation in (6.10) via the Gaussian integration by parts computation on (**??**) implies that

$$\begin{aligned}|g'_{k,k'}(\alpha)| &= \Big|\mathbb{E}\Big\langle x^0(k)x(k')\Big(x^{\mathsf{T}}Mx^0+z^{\mathsf{T}}(\partial_\alpha\sqrt{Q+\alpha M})|_{\alpha=0}x-\frac{x^{\mathsf{T}}Mx}{2}\Big)\Big\rangle_Q\\ &\quad-\mathbb{E}\Big\langle x^0(k)x(k')\Big((x^2)^{\mathsf{T}}Mx^0+z^{\mathsf{T}}(\partial_\alpha\sqrt{Q+\alpha M})|_{\alpha=0}x^2-\frac{(x^2)^{\mathsf{T}}M(x^2)}{2}\Big)\Big\rangle_Q\Big|\\ &\le \Big|\mathbb{E}\Big\langle x^0(k)x(k')\Big(x^{\mathsf{T}}Mx^0-\frac{x^{\mathsf{T}}Mx^2}{2}\Big)\Big\rangle_Q\Big|+\Big|\mathbb{E}\Big\langle x^0(k)x(k')\Big((x^2)^{\mathsf{T}}Mx^0-\frac{(x^2)^{\mathsf{T}}Mx^3}{2}\Big)\Big\rangle_Q\Big|\\ &\le 2C^2\mathbb{E}\Big\langle\Big(\|M\|_2\|x^0x^{\mathsf{T}}\|_2+\frac{1}{2}\|M\|_2\|x^2x^{\mathsf{T}}\|_2\Big)\Big\rangle_Q\\ &\le 3\kappa C^3\|M\|_2.\end{aligned}$$

In the last second line, we used the Cauchy–Schwarz inequality on the Frobenius inner product and in the last line we used the fact that $\mathbb{E}\langle\cdot\rangle_Q$ is the average with respect to a probability measure.

By the mean value theorem, we can conclude that

$$|f_{k,k'}(Q_1)-f_{k,k'}(Q_2)|=|g_{k,k'}(1)-g_{k,k'}(0)|\le 3\kappa C^3\|Q_1-Q_2\|_2.$$

Lastly, we get our required estimate

$$\|f(Q_1)-f(Q_2)\|_2=\Big(\sum_{k,k'=1}^\kappa(f_{k,k'}(Q_1)-f_{k,k'}(Q_2))^2\Big)^{1/2}\le 3\kappa^2C^3\|Q_1-Q_2\|_2.$$

Taking $Q_1=\tilde{Q}_t$ and $Q_2=\tilde{Q}'_t$ gives our estimate. □

Using this Lipschitz continuity, we can do a fixed point argument to show that there exists a unique maximizer at 0 if the covariances $\frac{1}{\Delta_{s,t}^2}$ are sufficiently small.

Lemma 6.4 (*Uniqueness of a Maximizer at* 0)

Consider the model parameters defined in (6.3). The functional $\varphi(Q)$ has a unique maximizer at 0 if

$$\left\|\sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho}\right\|_{op} < \frac{1}{9\kappa^4 C^6}.$$

Proof. Following the computations leading to (6.7) applied to the general critical point equation (6.12), we see that

$$\|\sqrt{\rho}\tilde{Q}\|_2^2 = \mathrm{Tr}\left((\sqrt{\rho}\tilde{Q})^\intercal(\sqrt{\rho}\tilde{Q})\right) = \mathrm{Tr}\left((A(\sqrt{\rho}f(\tilde{Q}))^\intercal(A(\sqrt{\rho}f(\tilde{Q}))\right) \le \|A\|_{op}\|\sqrt{\rho}f(\tilde{Q})\|_2^2 \tag{6.13}$$

where

$$f(\tilde{Q}) = (\mathbb{E}\langle x^0 x^\intercal\rangle_{\tilde{Q}_1}, \dots, \mathbb{E}\langle x^0 x^\intercal\rangle_{\tilde{Q}_n}) \qquad A = \sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho}$$

We proved in Lemma 6.3 that $f(\tilde{Q}_t) = \mathbb{E}\langle x^0 x^\intercal\rangle_{\tilde{Q}_t}$ is Lipschitz in $\tilde{Q}_t$ with a Lipschitz constant that does not depend on t

$$\|f(\tilde{Q}_t) - f(\tilde{Q}_t')\|_2 \le 3\kappa^2 C^3 \|\tilde{Q}_t - \tilde{Q}_t'\|_2.$$

Recall that the prior is centered $f(0) = 0$, so

$$\|\sqrt{\rho}f(\tilde{Q})\|_2^2 = \sum_t \rho_t \|f(\tilde{Q}_t) - f(0)\|_2^2 \le 9\kappa^4 C^6 \sum_t \rho_t \|\tilde{Q}_t\|_2^2. \le 9\kappa^4 C^6 \|\sqrt{\rho}\tilde{Q}\|_2^2.$$

Applying this bound to (6.13) yields the result. □

This bound is clearly not tight, because it does not depend on the measure P_0 except through the support contained in $[-C, C]^\kappa$. We will show below that if the operator norm is sufficiently large, then a maximizer away from 0 will appear.

Lemma 6.5 (*Phase Transition with Centered Prior*)

Let $x \sim P_X$. The functional φ has a maximizer away from $Q = 0$ if

$$\left\|\sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho}\right\|_{op} > \frac{1}{\|\mathbb{E}xx^\intercal\|_{op}^2} = \frac{1}{\|\mathrm{Cov}(x)\|_{op}^2}.$$

Remark 6.6. If we have a Gaussian prior, then $x \sim N(0, I)$, so $\|\mathbb{E}xx^\intercal\|_{op} = \|I\|_{op} = 1$. Furthermore, it also agrees with condition (194) in [32] in the case when $\kappa = 1$.

Proof. We adapt the proof of the Gaussian prior to the general scenario. Consider an arbitrary perturbation $M = (M_1, \dots, M_n)$ of Q such that $Q + \varepsilon M \ge 0$ for all ε sufficiently small. The first and second variations are denoted by

$$\nabla_M \varphi(Q) = \partial_\varepsilon \varphi(Q + \varepsilon M)\Big|_{\varepsilon=0} \quad \text{and} \quad \nabla_M^2 \varphi(Q) = \partial_\varepsilon^2 \varphi(Q + \varepsilon M)\Big|_{\varepsilon=0}.$$

This can be computed explicitly using integration by parts and the Nishimori property,

$$\nabla_M \varphi(Q) = \sum_{s,u} \frac{\rho_s \rho_u}{2\Delta_{s,u}} \mathbb{E}\big\langle \mathrm{Tr}(x^\intercal M_s x_0)\big\rangle_{\tilde{Q}_u} - \sum_{s,u} \frac{\rho_s \rho_u}{2\Delta_{s,u}} \mathrm{Tr}(Q_u M_s)$$

and

$$\nabla_M^2 \varphi(Q) = \sum_{s,t,u} \frac{\rho_s \rho_t \rho_u}{4\Delta_{s,u}\Delta_{u,t}} \Big(\mathbb{E}\big\langle (x^\intercal M_s x_0)(x^\intercal M_t x_0)\big\rangle_{\tilde{Q}_u} - \mathbb{E}\big\langle (x_1^\intercal M_s x_0)(x_2^\intercal M_t x_0)\big\rangle_{\tilde{Q}_u}\Big)$$
$$- \sum_{s,t} \frac{\rho_s \rho_t}{2\Delta_{s,t}} \mathrm{Tr}(M_s M_t).$$

When $Q = 0$ the Hessian simplifies because $\langle\cdot\rangle_0$ is simply the average with respect to P_X which is centered,

$$\nabla_M^2 \varphi(0) = \sum_{s,t,u} \frac{\rho_s \rho_t \rho_u}{4\Delta_{s,u}^2 \Delta_{u,t}^2} \mathbb{E}\big\langle \mathrm{Tr}(M_s x_0 x^\intercal)^2\big\rangle_0 - \sum_{s,t} \frac{\rho_s \rho_t}{2\Delta_{s,t}^2} \mathrm{Tr}(M_s M_t). \tag{6.14}$$

For $x \sim P_X$, let $\boldsymbol{C} = \mathbb{E}xx^\mathsf{T} = \mathrm{Cov}(x)$. Let b denote the unit eigenvector corresponding to the largest eigenvalue of $\boldsymbol{C}$. In particular, we have $b^\mathsf{T}\boldsymbol{C}b = \|\boldsymbol{C}\|_{op}$. Suppose that

$$\|A\|_{op} = \left\|\sqrt{\boldsymbol{\rho}}\frac{1}{\Delta}\sqrt{\boldsymbol{\rho}}\right\|_{op} > \frac{1}{\mathbb{E}\|x_1x_2^\mathsf{T}\|_2^2} = \frac{1}{\|\boldsymbol{C}\|_{op}^2}. \tag{6.15}$$

The matrix $A = \sqrt{\boldsymbol{\rho}}\frac{1}{\Delta^2}\sqrt{\boldsymbol{\rho}}$ has non-negative entries, the eigenvector v associated with the largest eigenvalue has non-negative entries by the Perron–Frobenius Theorem. Therefore, we can take

$$M = (u_1\boldsymbol{B}, \dots, u_n\boldsymbol{B}) \qquad \text{where } \boldsymbol{B} = bb^\mathsf{T} \text{ and } u_i = \frac{1}{\sqrt{\rho_i}}v_i$$

Notice that

$$\mathbb{E}\big\langle \mathrm{Tr}(\boldsymbol{B}x_0x^\mathsf{T})^2\big\rangle_0 = \mathbb{E}\left\langle \sum_{i,j,k,l} b_ib_jb_kb_\ell x_ix_j^0x_kx_\ell^0\right\rangle_0 = (b^\mathsf{T}\boldsymbol{C}b)^2 = \|\,\mathrm{Cov}(x)\|_{op}^2.$$

since x and x^0 have the same distribution in the Bayes optimal case. The Hessian at 0 defined in (6.14) simplifies greatly for this choice of M,

$$\nabla_M^2\varphi(0) = \left(\|\boldsymbol{C}\|_{op}^2u^\mathsf{T}\rho\frac{1}{\Delta_{s,t}}\rho\frac{1}{\Delta_{s,t}}\rho u - u^\mathsf{T}\rho\frac{1}{\Delta_{s,t}}\rho u\right) = \|\boldsymbol{C}\|_{op}^2v^\mathsf{T}A^2v - v^\mathsf{T}Av \geq \|\boldsymbol{C}\|_{op}^2\|A\|_{op}^2 - \|A\|_{op}.$$

By our assumption (6.15), $\nabla_M^2\varphi(0) > 0$. Therefore, the function $g(\varepsilon) = \varphi(\varepsilon M)$ satisfies $g''(0) > 0$, so it is convex on $[0, \delta]$ for some $\delta > 0$. Furthermore, $g'(0) = 0$, so we can conclude that

$$\varphi(\delta M) = g(\delta) \geq g(0) = \varphi(0)$$

by convexity. Lastly, notice that $M \geq 0$ since the entries of $\vec{u}$ are non-negative, so M is in the domain of the optimization problem. □

Remark 6.7. Our choice of M in the proof is the choice of directional derivative that maximizes the Hessian at 0. This gives us the sharp condition when the Hessian of φ is no longer negative semidefinite.

6.3. **Comparison with the Naive BBP Transition.** In this section, we show that the threshold Lemma 6.5 is stronger than the transition computed by examining the BBP transition of a spiked matrix with homogeneous noise and inhomogenous spike. We will see that the transitions are equal if and only if the models are homogeneous.

We first consider the model with homogeneous noise, but inhomogeneous signal. We want to find the BBP transition of the matrix

$$\frac{1}{\sqrt{N}}Y^\Delta \odot \frac{1}{\Delta^{\odot\frac{1}{2}}} = \frac{x^0(x^0)^\mathsf{T}}{N} \odot \frac{1}{\Delta^{\odot\frac{1}{2}}} + \frac{1}{\sqrt{N}}G.$$

where G is a Gaussian Wigner matrix.

Lemma 6.8 (*BBP Transition*)

The matrix $\frac{1}{\sqrt{N}}Y^\Delta \odot \frac{1}{\Delta^{\odot\frac{1}{2}}}$ has an outlier iff

$$\left\|\sqrt{\rho}\frac{1}{\Delta^{\odot\frac{1}{2}}}\sqrt{\rho}\right\|_{op} > \frac{1}{\|\mathbb{E}xx^\mathsf{T}\|_{op}}.$$

Proof. It is well known that since G follows the GOE the largest eigenvalue of $\frac{1}{\sqrt{N}}Y^\Delta \odot \frac{1}{\Delta^{\odot\frac{1}{2}}}$ is given by $\lambda^* = \max_{i:\gamma_i>1}\{\gamma_i + \gamma_i^{-1}\}$ if $(\gamma_i)_{1\leq i\leq n}$ are the eigenvalues of

$$R = \frac{x^0(x^0)^\mathsf{T}}{N} \odot \frac{1}{\Delta^{\odot\frac{1}{2}}}$$

But exactly as in the proof of Theorem 3.14 and Remark 3.15, we can see that the eigenvalues of R are the same as those of

$$M_{st}^N = \frac{1}{N}\sum_{j=1}^{\kappa}\|v_s^j\|\|v_t^j\|\frac{1}{\Delta_{st}^{\odot\frac{1}{2}}}$$

a finite matrix whose top eigenvalue converges entrywise when N goes to infinity, by the law of large numbers, towards the top eigenvalue of the matrix

$$M_{st}^\infty = \|\mathbb{E}xx^\mathsf{T}\|_{op}\sqrt{\rho_s\rho_t}\frac{1}{\Delta_{st}^{\odot\frac{1}{2}}}$$

by the definition v^1 in Remark 3.15. We conclude that $\frac{1}{\sqrt{N}} Y^{\Delta} \odot \frac{1}{\Delta^{\odot \frac{1}{2}}}$ has an outlier iff

$$\left\| \sqrt{\rho} \frac{1}{\Delta^{\odot \frac{1}{2}}} \sqrt{\rho} \right\|_{op} > \frac{1}{\|\mathbb{E} x x^{\mathsf{T}}\|_{op}}$$

□

We now prove that the BBP transition is strictly weaker than the spin glass transition except in the homogeneous models. That is, if Δ is such that the BBP transition happens

$$\left\| \sqrt{\rho} \frac{1}{\Delta^{\odot \frac{1}{2}}} \sqrt{\rho} \right\|_{op} \geq \frac{1}{\|\mathbb{E} x x^{\mathsf{T}}\|_{op}}$$

then we are also in the information theoretically feasible detectable region

$$\left\| \sqrt{\rho} \frac{1}{\Delta} \sqrt{\rho} \right\|_{op} \geq \frac{1}{\|\mathbb{E} x x^{\mathsf{T}}\|_{op}^2}.$$

In particular, if it is possible to detect the signal using the spectral method, then it is also possible to detect is using any other method. The converse is false, so there exists some algorithms that beat the naive spectral ones.

This statement is true, and in fact a stronger statement holds, which clearly implies Proposition 2.18.

Lemma 6.9 (*Gap in Thresholds*)

We have

$$\left\| \sqrt{\rho} \frac{1}{\Delta^{\odot \frac{1}{2}}} \sqrt{\rho} \right\|_{op}^2 \leq \left\| \sqrt{\rho} \frac{1}{\Delta} \sqrt{\rho} \right\|_{op},$$

with equality holding if and only if there exists a constant c such that for all s, t, $\Delta_{s,t} = c$. In particular, if Δ satisfies

$$\left\| \sqrt{\rho} \frac{1}{\Delta^{\odot \frac{1}{2}}} \sqrt{\rho} \right\|_{op} \geq \frac{1}{\|\mathbb{E} x x^{\mathsf{T}}\|_{op}}$$

then

$$\left\| \sqrt{\rho} \frac{1}{\Delta} \sqrt{\rho} \right\|_{op} \geq \left\| \sqrt{\rho} \frac{1}{\Delta^{\odot \frac{1}{2}}} \sqrt{\rho} \right\|_{op}^2 \geq \frac{1}{\|\mathbb{E} x x^{\mathsf{T}}\|_{op}^2}.$$

Proof. First of all, observe that $\sqrt{\rho} \frac{1}{\Delta^{\odot \frac{1}{2}}} \sqrt{\rho}$ and $\sqrt{\rho} \frac{1}{\Delta} \sqrt{\rho}$ have non-negative entries, so the Perron–Frobenius theorem implies that the largest eigenvector has non-negative entries. Therefore, we can restrict ourselves to eigenvectors with non-negative entries to compute the maximum eigenvalues. But for any vector u with non-negative entries, by Cauchy-Schwartz's inequality

$$\begin{aligned} \langle u, \sqrt{\rho} \frac{1}{\Delta^{\odot \frac{1}{2}}} \sqrt{\rho} u \rangle &= \sum_{s,t} \frac{\rho_s^{\frac{1}{4}} u_s^{\frac{1}{2}} \rho_t^{\frac{1}{4}} u_t^{\frac{1}{2}}}{\Delta_{s,t}^{1/2}} \rho_s^{\frac{1}{4}} u_s^{\frac{1}{2}} \rho_t^{\frac{1}{4}} u_t^{\frac{1}{2}} \\ &\leq \left(\langle u, \sqrt{\rho} \frac{1}{\Delta} \sqrt{\rho} u \rangle \right)^{\frac{1}{2}} \left(\sum \rho_s^{\frac{1}{2}} u_s \right)^{\frac{1}{2}} \\ &\leq \left(\langle u, \sqrt{\rho} \frac{1}{\Delta} \sqrt{\rho} u \rangle \right)^{\frac{1}{2}} \|u\|_2^{\frac{1}{2}} \end{aligned}$$

where we used again Cauchy–Schwartz's inequality and the fact that $\sum \rho_s = 1$ to see that $\sum \rho_s^{\frac{1}{2}} u_s \leq \|u\|_2$. We deduce by taking the supremum over u so that $\|u\|_2 = 1$ that

$$\left\| \sqrt{\rho} \frac{1}{\Delta^{\odot \frac{1}{2}}} \sqrt{\rho} \right\|_{op} \leq \left\| \sqrt{\rho} \frac{1}{\Delta} \sqrt{\rho} \right\|_{op}^{1/2}.$$

Furthermore, equality holds only if there is equality in the above Cauchy-Schwartz inequality which happens when we take u to be the largest eigenvector of the matrix $\sqrt{\rho} \frac{1}{\Delta^{\odot \frac{1}{2}}} \sqrt{\rho}$. For this to happen, the largest eigenvector

must be $u = \sqrt{\rho}$ so there exists c such that

$$\frac{\rho_s^{\frac{1}{2}}\rho_t^{\frac{1}{2}}}{\Delta_{s,t}^{\frac{1}{2}}} = c\rho_s^{\frac{1}{2}}\rho_t^{\frac{1}{2}}, \quad \forall\, s,t$$

implying that $\Delta_{s,t}$ must be constant.

□

6.4. **Proof of Corollary 2.14 and Lemma 2.15.** We now explain how one can use the phase transition to recover the recovery transitions. Recall that the minimal matrix mean squared error is given by

$$\mathrm{MMSE}(N) = \frac{2}{N(N-1)}\sum_{i<j}\mathbb{E}(x_i^0\cdot x_j^0 - \mathbb{E}[x_i^0\cdot x_j^0 \mid Y])^2$$

The limit of the MMSE will follow from the following property of the maximizers of the replica symmetric functional φ defined in (2.13).

Lemma 6.10 (*Limit of the Overlaps*)

If $\frac{1}{\Delta}$ is positive definite, the maximizers of (2.13) satisfy

$$\lim_{N\to\infty}\sum_{s,t=1}^n \rho_s\rho_t\langle \mathrm{Tr}(R_{10}^s, R_{10}^t)\rangle = \sum_{s,t=1}^n \rho_s\rho_t\,\mathrm{Tr}(Q_sQ_t).$$

Furthermore, for any maximizer, the values of $\mathrm{Tr}(Q_s,Q_t)\geq 0$ are unique for fixed s,t.

Proof. Using the change of variables $\eta_{s,t} = \frac{1}{\Delta_{s,t}}$ we consider our Gaussian free energy

$$F_N(\eta) = \frac{1}{N}\mathbb{E}_Y \log\int e^{\sum_{s,t}\sum_{i\in I_s, j\in I_t}\frac{\sqrt{\eta_{s,t}}}{\sqrt{N}}g_{ij}(x_i\cdot x_j)+\frac{\eta_{st}}{N}(x_i^0\cdot x_j^0)(x_i\cdot x_j)-\frac{\eta_{st}}{2N}(x_i\cdot x_j)^2}\cdot dP_0^{\otimes N}(x)$$

where $\eta_{i,j} = \eta_{s,t}$ for $i\in I_s$ and $j\in I_t$, which is equivalent to (2.12). Notice that differentiating the free energy with respect to $\eta_{u,v} := \frac{1}{\Delta_{u,v}}$ and an application of the Nishimori property recovers the Gibbs averages,

$$\begin{aligned}
\partial_{\eta_{u,v}}F_N(\eta) &= \mathbb{E}\Big\langle \frac{1}{2N}\sum_{i\in I_u, j\in I_v}\frac{g_{ij}(x_i\cdot x_j)}{\sqrt{N\eta_{s,t}}} + \frac{\rho_u\rho_v}{2}\mathrm{Tr}(R_{1,0}^uR_{1,0}^v) - \frac{\rho_u\rho_v}{4}\mathrm{Tr}(R_{1,1}^uR_{1,1}^v)\Big\rangle \\
&= \mathbb{E}\Big\langle \frac{\rho_u\rho_v}{4}\mathrm{Tr}(R_{1,1}^uR_{1,1}^v) - \frac{\rho_u\rho_v}{4}\mathrm{Tr}(R_{1,2}^uR_{1,2}^v) + \frac{\rho_u\rho_v}{2}\mathrm{Tr}(R_{1,0}^uR_{1,0}^v) - \frac{\rho_u\rho_v}{4}\mathrm{Tr}(R_{1,1}^uR_{1,1}^v)\Big\rangle + o(1) \\
&= \mathbb{E}\Big\langle \frac{\rho_u\rho_v}{4}\mathrm{Tr}(R_{1,0}^uR_{1,0}^v)\Big\rangle + o(1)
\end{aligned}$$

where $\langle\cdot\rangle$ is the Gibbs average with respect to the Hamiltonian (2.11). On the other hand, for fixed $1\leq u,v\leq n$, consider the functional

$$\psi(\Delta) = \sup_Q\Big[-\sum_{s,t=1}^n\frac{\rho_s\rho_t}{4\Delta_{st}}\mathrm{Tr}(Q_sQ_t) + \sum_{s=1}^n\rho_s\mathbb{E}_{z,x^0}\ln\Big[\int e^{(\tilde{Q}_sx^0+\sqrt{\tilde{Q}_s}z)^\intercal x - \frac{x^\intercal\tilde{Q}_sx}{2}}dP_0(x)\Big].\Big]$$

When $\frac{1}{\Delta}$ is invertible, $\tilde{Q} = \frac{1}{\Delta}\rho Q$ implies that $Q = \rho^{-1}(\frac{1}{\Delta})^{-1}\tilde{Q}$. In this notation, it is understood that for $Q,\tilde{Q}\in(\mathbb{R}^{\kappa\times\kappa})^n$ and the entries fo the vector

$$\tilde{Q}_s = \sum_t\frac{1}{\Delta_{st}}\rho_tQ_t\in\mathbb{R}^{\kappa\times\kappa}$$

are non-negative because $Q_1,\dots,Q_n$ are. Under this change of variables, we have

$$\psi(\eta) = \sup_{\tilde{Q}}\Big[-\frac{1}{4}\mathrm{Tr}(\tilde{Q}^\intercal\Big(\frac{1}{\Delta}\Big)^{-1}\tilde{Q}) + \sum_{s=1}^n\rho_s\mathbb{E}_{z,x^0}\ln\Big[\int e^{(\tilde{Q}_sx^0+\sqrt{\tilde{Q}_s}z)^\intercal x-\frac{x^\intercal\tilde{Q}_sx}{2}}dP_0(x)\Big].\Big] \tag{6.16}$$

where the supremum is taken over non-negative matrices. Observe that the function we are optimizing is continuous. Moreover, the second term goes to infinity at most like $\left(\sum_s\mathrm{Tr}(\tilde{Q}_s^2)\right)^{1/2}$, whereas the first term goes to $-\infty$ like $\sum_s\mathrm{Tr}(\tilde{Q}_s^2)$ since we assumed that the smallest eigenvalue of $\frac{1}{\Delta}$ is positive. Hence, we can restrict the supremum to non-negative matrices such that $\sum_s\mathrm{Tr}(\tilde{Q}_s^2)$ is bounded by some finite M, which is a

compact space. Using the change of variables $\eta_{s,t} = \frac{1}{\Delta_{s,t}}$, the envelope theorem [33, Corollary 4] implies that at any point where ψ is differentiable,

$$\partial_{\eta_{u,v}}\psi(\eta) := \frac{d}{d\varepsilon}\psi(\eta_{u,v}+\varepsilon)\Big|_{\varepsilon=0} = \frac{\rho_u\rho_v}{4}\operatorname{Tr}(Q_u^*Q_v^*)$$

where Q^* is any maximizer of the right hand side of (6.16). Furthermore, if $Q^\dagger$ is another maximizer, then

$$\frac{\rho_u\rho_v}{4}\operatorname{Tr}(Q_u^*Q_v^*) = \frac{\rho_u\rho_v}{4}\operatorname{Tr}(Q_u^\dagger Q_v^\dagger)$$

so the value of $\operatorname{Tr}(Q_u^*Q_v^*)$ is unique even though the maximizers Q^* and $Q^\dagger$ may not be.

But ψ is convex in η so that $(\partial_{\eta_{u,v}}\psi(\eta))_{u,v}$ exists for almost every η, so does the limit of the derivative of the free energy Theorem 2.10 implies

$$\lim_{N\to\infty} F_N(\eta) = \psi(\eta) \qquad \text{and} \qquad \lim_{N\to\infty}\partial_{\eta_{u,v}}F_N(\eta) = \partial_{\eta_{u,v}}\psi(\eta)$$

for almost every η so

$$\lim_{N\to\infty}\sum_{u,v=1}^n \mathbb{E}\Big\langle \frac{\rho_u\rho_v}{4}\operatorname{Tr}(R_{1,0}^uR_{1,0}^v)\Big\rangle = \sum_{u,v=1}^n \rho_u\rho_v\operatorname{Tr}(Q_u^*Q_v^*),$$

finishing the proof. □

We now prove the statements about the MMSE.

Proof of Corollary 2.14. Recall that $\langle\cdot\rangle$ is the average with respect to $P(x\mid Y)$. The mean squared error can be simplified to

$$\begin{aligned}
&\frac{2}{N(N-1)}\sum_{i<j}\mathbb{E}(x_i^0\cdot x_j^0 - \mathbb{E}[x_i^0\cdot x_j^0\mid Y])^2\\
&= \frac{2}{N(N-1)}\sum_{i<j}\mathbb{E}\Big((x_i^0\cdot x_j^0)^2 - 2(x_i^0\cdot x_j^0)\mathbb{E}[x_i^0\cdot x_j^0\mid Y] + \mathbb{E}[x_i^0\cdot x_j^0\mid Y]^2\Big)\\
&= \frac{2}{N(N-1)}\sum_{i<j}\mathbb{E}\Big\langle (x_i^0\cdot x_j^0)^2 - 2(x_i^0\cdot x_j^0)(x_i^1\cdot x_j^1) + (x_i^1\cdot x_j^1)(x_i^2\cdot x_j^2)\Big\rangle\\
&= \mathbb{E}\Big\langle \operatorname{Tr}(R_{00},R_{00}) - 2\operatorname{Tr}(R_{10},R_{10}) + \operatorname{Tr}(R_{12},R_{12})\Big\rangle\\
&= \mathbb{E}\operatorname{Tr}(R_{00},R_{00}) - \mathbb{E}\langle\operatorname{Tr}(R_{10},R_{10})\rangle\\
&= \mathbb{E}\|xx^\intercal\|_2^2 - \sum_{s,t=1}^n \rho_s\rho_t\mathbb{E}\langle\operatorname{Tr}(R_{10}^s,R_{10}^t)\rangle.
\end{aligned}$$

By Lemma 6.10 it follows that

$$\operatorname{MMSE}(\Delta) = \mathbb{E}\|xx^\intercal\|_2^2 - \sum_{s,t=1}^n \rho_s\rho_t\operatorname{Tr}(Q_s,Q_t).$$

□

Lemma 2.15 is now immediate.

Proof of Lemma 2.15. By Lemma 6.4 and Corollary 2.14, if $\|\sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho}\|_{op} < \frac{1}{9\kappa^4C^6}$ then $Q_s = 0$ for all s, so $\lim_{N\to\infty}\operatorname{MMSE}(N) = \mathbb{E}_{P_0}\|xx^\intercal\|_2^2$. On the other hand, if $\|\sqrt{\rho}\frac{1}{\Delta}\sqrt{\rho}\|_{op} > \frac{1}{\|\mathbb{E}xx^\intercal\|_{op}^2} = \frac{1}{\|\operatorname{Cov}(x)\|_{op}^2}$ then there exists a s such that $Q_s > 0$, so $\lim_{N\to\infty}\operatorname{MMSE}(N) < \mathbb{E}_{P_0}\|xx^\intercal\|_2^2$. □

References

[1] Arka Adhikari and Christian Brennecke, *Free energy of the quantum sherrington-kirkpatrick spin-glass model with transverse field*, 2019.

[2] Oskari Ajanki, Laszlo Erdos, and Torben Krüger, *Universality for general wigner-type matrices*, Probability Theory and Related Fields **169** (2017).

[3] Diego Alberici, Francesco Camilli, Pierluigi Contucci, and Emanuele Mingione, *A statistical physics approach to a multi-channel wigner spiked model*, Europhysics Letters **136** (2021), no. 4, 48001.

[4] Greg W. Anderson, Alice Guionnet, and Ofer Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010. MR 2760897

[5] J. Baik, G. Ben Arous, and S. Péché, *Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices*, Ann. Probab. **33** (2005), no. 5, 1643–1697.

[6] Jean Barbier and Nicolas Macris, *The adaptive interpolation method: a simple scheme to prove replica formulas in bayesian inference*, Probability theory and related fields **174** (2019), no. 3, 1133–1185.

[7] Jean Barbier and Dmitry Panchenko, *Strong replica symmetry in high-dimensional optimal bayesian inference*, 2020.

[8] Jean Barbier and Galen Reeves, *Information-theoretic limits of a multiview low-rank symmetric spiked matrix model*, 2020 IEEE International Symposium on Information Theory (ISIT), IEEE, 2020, pp. 2771–2776.

[9] Erik Bates and Youngtak Sohn, *Crisanti–sommers formula and simultaneous symmetry breaking in multi-species spherical spin glasses*, Communications in Mathematical Physics (2022).

[10] ———, *Free energy in multi-species mixed p-spin spherical models*, Electronic Journal of Probability **27** (2022), no. none.

[11] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration inequalities*, Oxford University Press, Oxford, 2013, A nonasymptotic theory of independence, With a foreword by Michel Ledoux. MR 3185193

[12] Francesco Camilli, Pierluigi Contucci, and Emanuele Mingione, *An inference problem in a mismatched setting: a spin-glass model with mattis interaction*, SciPost Physics (2022).

[13] Philippe Carmona and Yueyun Hu, *Universality in sherrington–kirkpatrick's spin glass model*, Annales de l'Institut Henri Poincare (B) Probability and Statistics **42** (2006), no. 2, 215–222.

[14] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari, *Asymptotic mutual information for the balanced binary stochastic block model*, Information and Inference: A Journal of the IMA **6** (2017), no. 2, 125–170.

[15] Yash Deshpande and Andrea Montanari, *Information-theoretically optimal sparse pca*, 2014 IEEE International Symposium on Information Theory, IEEE, 2014, pp. 2197–2201.

[16] ———, *Sparse pca via covariance thresholding*, Advances in Neural Information Processing Systems **27** (2014).

[17] ———, *Finding hidden cliques of size in nearly linear time*, Foundations of Computational Mathematics **15** (2015), no. 4, 1069–1128.

[18] Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, Lenka Zdeborová, et al., *Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula*, Advances in Neural Information Processing Systems **29** (2016).

[19] David L Donoho, Matan Gavish, and Iain M Johnstone, *Optimal shrinkage of eigenvalues in the spiked covariance model*, arXiv preprint arXiv:1311.0851 (2013).

[20] ———, *Optimal shrinkage of eigenvalues in the spiked covariance model*, Annals of statistics **46** (2018), no. 4, 1742.

[21] Ahmed El Alaoui and Florent Krzakala, *Estimation in the spiked wigner model: a short proof of the replica formula*, 2018 IEEE International Symposium on Information Theory (ISIT), IEEE, 2018, pp. 1874–1878.

[22] Ahmed El Alaoui, Florent Krzakala, and Michael Jordan, *Fundamental limits of detection in the spiked wigner model*, The Annals of Statistics **48** (2020), no. 2, 863–885.

[23] Paul Erdõs, András Hajnal, and János Pach, *A ramsey-type theorem for bipartite graphs.*, Geombinatorics **10** (2000).

[24] Alice Guionnet and Jonathan Husson, *Large deviations for the largest eigenvalue of Rademacher matrices*, Ann. Probab. **48** (2020), no. 3, 1436–1465. MR 4112720

[25] Adel Javanmard and Andrea Montanari, *State evolution for general approximate message passing algorithms, with applications to spatial coupling*, Information and Inference: A Journal of the IMA **2** (2013), no. 2, 115–144.

[26] Brian Karrer and M. E. J. Newman, *Stochastic blockmodels and community structure in networks*, Phys. Rev. E **83** (2011), 016107.

[27] Justin Ko, *Free energy of multiple systems of spherical spin glasses with constrained overlaps*, Electron. J. Probab. **25** (2020), 34 pp.

[28] Florent Krzakala, Jiaming Xu, and Lenka Zdeborová, *Mutual information in rank-one matrix estimation*, 2016 IEEE Information Theory Workshop (ITW), IEEE, 2016, pp. 71–75.

[29] Marc Lelarge and Léo Miolane, *Fundamental limits of symmetric low-rank matrix estimation*, Probab. Theory Related Fields **173** (2019), no. 3-4, 859–929. MR 3936148

[30] Thibault Lesieur, Caterina De Bacco, Jess Banks, Florent Krzakala, Cris Moore, and Lenka Zdeborová, *Phase transitions and optimal algorithms in high-dimensional gaussian mixture clustering*, 2016 54th Annual Allerton Conference on Communication, Control, and Computing (Allerton), IEEE, 2016, pp. 601–608.

[31] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová, *Mmse of probabilistic low-rank matrix estimation: Universality with respect to the output channel*, 2015 53rd Annual Allerton Conference on Communication, Control, and Computing (Allerton), IEEE, 2015, pp. 680–687.

[32] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová, *Constrained low-rank matrix estimation: phase transitions, approximate message passing and applications*, J. Stat. Mech. Theory Exp. (2017), no. 7, 073403, 86. MR 3683819

[33] Paul Milgrom and Ilya Segal, *Envelope theorems for arbitrary choice sets*, Econometrica **70** (2002), no. 2, 583–601.

[34] Dmitry Panchenko, *The Sherrington-Kirkpatrick model*, Springer Monographs in Mathematics, Springer, New York, 2013. MR 3052333

[35] ———, *The free energy in a multi-species Sherrington-Kirkpatrick model*, Ann. Probab. **43** (2015), no. 6, 3494–3513. MR 3433586

[36] ———, *The free energy in a multi-species Sherrington-Kirkpatrick model*, Ann. Probab. **43** (2015), no. 6, 3494–3513. MR 3433586

[37] ———, *Free energy in the mixed p-spin models with vector spins*, Ann. Probab. **46** (2018), no. 2, 865–896. MR 3773376

[38] ———, *Free energy in the Potts spin glass*, Ann. Probab. **46** (2018), no. 2, 829–864. MR 3773375

[39] Amelia Perry, Alexander S Wein, Afonso S Bandeira, and Ankur Moitra, *Optimality and sub-optimality of pca for spiked random matrices and synchronization*, arXiv preprint arXiv:1609.05573 (2016).

[40] Sundeep Rangan and Alyson K Fletcher, *Iterative estimation of constrained rank-one matrices in noise*, 2012 IEEE International Symposium on Information Theory Proceedings, IEEE, 2012, pp. 1246–1250.

[41] Eliran Subag, *Tap approach for multi-species spherical spin glasses i: general theory*, 2021.

[42] ———, *Tap approach for multi-species spherical spin glasses ii: the free energy of the pure models*, 2021.