Phase Transitions in Random Tensors with Multiple Spikes

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Madeline Curtis Handschy

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Wei-Kuo Chen, Gilad Lerman

May, 2019

Acknowledgements

I am deeply grateful to the following people for their support during my time in graduate school.

Most importantly, I am thankful to my advisors, Professors Wei-Kuo Chen and Gilad Lerman for constantly challenging me to learn more and become a better mathematician.

To my mom for always having the right mix of commiseration, sarcasm, and dark humor as well as patience in excising all of the unnecessary commas from this thesis.

To my dad for his heartfelt support and advice. Let's go drink bourbon down there where the train goes slow

To Lizzy. I can't even put it into words, but she knows.

To Kyle for being there every day and helping me keep things in perspective.

Dedication

To Marcia. Thank you for your patience, optimism, friendship, and everything.

Abstract

This thesis is concerned with the problem of detecting and recovering a low-rank tensor in noise. A spiked random tensor is composed of a symmetric Gaussian p-tensor and a fixed number of spikes. Each spike is a rank one p-tensor formed by a vector whose entries are drawn i.i.d. from a probability measure on the real line with bounded support. Each spike is weighted by a signal-to-noise ratio (SNR). For a random tensor with a single spike, it is possible to detect the presence of the spike when the SNR exceeds a critical threshold, and impossible when the SNR is below this threshold. For a random tensor with multiple spikes, detection of the low-rank structure is possible when the SNR of at least one spike exceeds its critical threshold. Additionally, recovery of the spikes by the minimum mean square error estimator has the same phase transition. When at least one SNR is above its critical threshold, the minimum mean square error estimator performs better than a random guess.

It is shown that the spike detection problem is equivalent to distinguishing between the high- and low-temperature regimes of certain mean field spin glass models. The set of SNRs for which detection is impossible is equal to the high-temperature regime of a certain p-spin model. Thus the main tools to investigate the detection problem come from the study of spin glasses.

Contents

Chapter 1

Introduction and Main Results

1.1 Motivation and Outline

This thesis addresses the question of when it is possible to detect and recover underlying low-rank structure from a random tensor formed as the sum of a fixed number of rank-one tensors and a tensor of white Gaussian noise. This model is known as the spiked tensor model. These results were first presented by Chen, Handschy, and Lerman in [1].

The detection of low-rank structure in tensors is motivated by the question of when Principal Component Analysis (PCA) can uncover linear low-rank structure in noisy data. PCA is equivalent to finding the eigen-decomposition of the sample covariance matrix of observed data points. Suppose data points $x_1, \ldots, x_L \in \mathbb{R}^N$ are drawn independently from the multi-variate Gaussian distribution $\mathcal{N}(0, I + \beta uu^T)$ with $u \in \mathbb{R}^N$ a unit vector, and $\beta > 0$ a constant known as the signal-to-noise ratio (SNR). Under this model, each data point is composed of the sum of a signal component and a noise component, and the question is whether PCA can detect the presence of the signal component and recover the signal.

Suppose $N/L \to \gamma < 1$ as $L \to \infty$. When $\beta = 0$, the eigenvalues of the sample covariance matrix follow the Marchenko-Pastur distribution [2]; however, when β is large enough, as $N \to \infty$, the largest eigenvalue of the sample covariance matrix falls outside the support of the Marchenko-Pastur distribution. In particular, when $\beta \leq \sqrt{\gamma}$, the eigenvalues follow the Marchenko-Pastur distribution, but when $\beta > \sqrt{\gamma}$, the largest eigenvalue 'pops out' of the support of the Marchenko-Pastur distribution, a phenomenon known as the BBP transition [3, 4].

In [5], the phase transition of spike detection is extended to spike recovery by PCA. More precisely, when $\beta > \sqrt{\gamma}$, there is a non-trivial asymptotic correlation between the signal u and the top eigenvector of the sample covariance matrix. Thus one can approximately recover u using PCA. However, when $\beta \leq \sqrt{\gamma}$, this asymptotic correlation is zero. Extension of detection and recovery to the case where $\gamma \geq 1$ is also established in [5]

The spiked Wigner matrix is another common signal-plus-noise model that exhibits a similar phase transition. An $N \times N$ Gaussian Wigner matrix is a symmetric matrix with independent entries $W_{ij} \sim \mathcal{N}(0, 1/2)$ for $1 \leq i < j \leq N$ and $W_{ii} \sim \mathcal{N}(0, 1)$ for $1 \leq i \leq N$. The spiked Wigner matrix with SNR $\beta > 0$ is defined as $T = W + \frac{\beta}{\sqrt{N}} uu^T$ with spike $u \in \mathbb{R}^N$ a vector with entries sampled i.i.d. from a probability distribution on $\mathbb{R}$ with bounded support. When the SNR β is below a critical threshold, the eigenvalue distribution of T follows Wigner's semi-circle law and detection of the low-rank structure is impossible. Once the value of β exceeds the critical threshold, the largest eigenvalue jumps away from the support of the Wigner semi-circle law and the top eigenvector nontrivially correlates with the signal [6, 7]. In this case, one can detect and approximately recover the signal by PCA.

This thesis studies the extension of the detection and recovery problems for the spiked Wigner model to higher-order tensors in the case of a single spike and in the case of multiple spikes. In particular, we show that the detection and recovery problem both exhibit sharp phase transitions. We define detection in terms of the total variation distance between the spiked and unspiked tensors and recovery in terms of the minimum mean square error estimator. One formula for the total variation distance between the spiked and unspiked random tensors involves their likelihood ratio, which turns out to be the free energy of certain p-spin spin glass models which are introduced in Chapter 2. The results on recovery by the minimum mean square error estimator follow from the results on detection.

The remainder of the present chapter introduces the main results on detection and recovery. Section 1.2 formally defines the spiked random tensor and the concept of spike detection. Section 1.3 states the main results for detection in the case of a single spike, Section 1.4 states the main results for detection in the case of multiple spikes, and

Section 1.5 gives the result for recovery by the minimum mean square error estimator.

Chapter 2 defines the p-spin mean field models relevant to the spike detection problem and states critical results for these models. Chapter 3 gives the proofs of the main theorems on detection and recovery assuming that the results stated in Chapter 2 hold. Finally, Chapter 4 presents the proofs of the spin glass results of Chapter 2.

1.2 Symmetric Gaussian Tensors

We now define the symmetric Gaussian tensor and the total variation distance between two random tensors. Additionally, we explain what it means for two sequences of random tensors to be distinguishable, a concept that defines the problem of spike detection.

Fix $p \in \mathbb{N}$. For any integer $N \geq 1$, define Ω_N to be the set of all real-valued p-tensors $Y = (y_{i_1,\ldots,i_p})_{1\leq i_1,\ldots,i_p\leq N}$. The inner product of two tensors is defined as

$$\langle Y, Y' \rangle = \sum_{1\leq i_1,\ldots,i_p\leq N} y_{i_1,\ldots,i_p} y'_{i_1,\ldots,i_p}.$$

Given a vector $u \in \mathbb{R}^N$ create a rank one p-tensor by taking the outer product

$$(u^{\otimes p})_{i_1,\ldots,i_p} = u_{i_1} \cdots u_{i_p}.$$

For $Y \in \Omega_N$ and a permutation π of the set $\{1, 2, \ldots, p\}$, define the permuted tensor Y^π by

$$Y^\pi_{i_1,\ldots,i_p} = y_{\pi(i_1),\ldots,\pi(i_p)}.$$

A tensor is symmetric when $Y = Y^\pi$ for all permutations π of the set $\{1, 2, \ldots, p\}$. From now on, Y will denote a random p-tensor with i.i.d. standard Gaussian entries. Given such a tensor Y, the symmetric Gaussian p-tensor is

$$W = \frac{1}{p!} \sum_\pi Y^\pi. \tag{1.1}$$

Given any two symmetric random p-tensors $U, V \in \Omega_N$, the total variation distance between U, V is

$$d_{TV}(U, V) = \sup_A |P(U \in A) - P(V \in A)|,$$

where the supremum is taken over all sets A in the Borel σ-algebra generated by symmetric p-tensors.

Definition 1.2.1. Two sequences of symmetric random tensors $U_N, V_N \in \Omega_N$ are indistinguishable if

$$\lim_{N\to\infty} d_{TV}(U_N, V_N) = 0,$$

and distinguishable if

$$\lim_{N\to\infty} d_{TV}(U_N, V_N) = 1.$$

From the definition of total variation distance, when U_N and V_N are distinguishable, there exists a sequence of measurable sets A_N such that $\lim_{N\to\infty} P(U_N \in A_N) = 1$ and $\lim_{N\to\infty} P(V_N \in A_N) = 0$. For example, consider the $N \times N$ spiked Wigner matrix $T_N = W_N + \frac{\beta}{\sqrt{N}} uu^T$. As mentioned in the previous section, if the SNR β is below a critical threshold, in the limit, the eigenvalues of T follow the Wigner semi-circle law, but if the SNR is above this threshold, the largest eigenvalue falls outside the support of the Wigner semi-circle distribution. If β is greater than than the critical SNR, and A_N is the event that the eigenvalue distribution of the matrix follows the Wigner semi-circle law, then $\lim_{N\to\infty} P(W_N \in A_N) = 1$ and $\lim_{N\to\infty} P(T_N \in A_N) = 0$, so the spiked and unspiked Wigner matrices are distinguishable.

More generally, suppose U_N, V_N are sequences of random tensors. Let 1_A denote the indicator function of a set A. Given a statistical hypothesis test $S_N : \Omega_N \to \{0, 1\}$ such that $S_N(\omega) = 1_{A_N}(\omega)$, the sum of Type I errors, or false positives, and Type II errors, or false negatives, satisfy the relationship

$$\min_A \{\text{Type I errors + Type II errors}\} = \text{ 1 - total variation distance.}$$

Thus, if U_N, V_N are distinguishable, then there must exist statistical hypothesis tests S_N that distinguish the two tensors in the sense that

$$\lim_{N\to\infty} \left(P(S_N(U_N) = 1) + P(S_N(V_N) = 0)\right) = 0.$$

On the other hand, when U_N, V_N are indistinguishable, there is no statistical hypothesis test that can distinguish between the two tensors.

1.3 Results for Detection with a Single Spike

We now consider the case of a spiked random tensor with a single spike. As in the case of the spiked Wigner matrix, we wish to know for which SNRs one can detect the presence of the spike to distinguish between the spiked tensor and a tensor of pure noise. The main result on spike detection for a single spike, Theorem 1, states that the detection problem exhibits a sharp phase transition, and Theorem 2 gives a method to compute the location of this phase transition.

Let Λ be a bounded subset of $\mathbb{R}$ and μ a probability measure on the Borel σ-field of Λ satisfying $\int a\mu(da) = 0$. Assume that $u_1, \dots, u_N$ are i.i.d. samples from μ that are also independent of the symmetric Gaussian tensor W. Set $u = (u_1, \dots, u_N)$. Given $\beta > 0$, the spiked random tensor (with a single spike) with signal-to-noise ratio β is

$$T = W + \frac{\beta}{N^{(p-1)/2}} u^{\otimes p}.$$

Definition 1.3.1. Detection of the spike $u^{\otimes p}$ is possible if W, T are distinguishable according to Definition 1.2.1. Detection is not possible if W, T are indistinguishable.

For each probability space (Λ, μ) there exists a critical SNR β_c depending on p such that detection is possible only when the SNR satisfies $\beta_c < \beta$.

Theorem 1. *Assume that μ is centered. Then for any $p \geq 3$ there exists a constant $\beta_c > 0$ such that*

(i) If $0 < \beta < \beta_c$ then detection is impossible.

(ii) If $\beta > \beta_c$ then detection is possible.

In [8], Perry, et al. show that for the Rademacher prior, where entries of u take values ± 1 with probability $1/2$, and the sparse Rademacher prior, where entries of u take values $\pm 1/\sqrt{\rho}$ with probability $\rho/2$ and 0 with probability $1 - \rho$, there exist upper and lower bounds β_- and β_+ with $\beta_- \leq \beta_+$, such that for $\beta < \beta_-$ detection is impossible and for $\beta > \beta_+$, detection is possible. Chen [9] shows that in fact $\beta_- = \beta_+$ in the case of the Rademachor prior, and Theorem 1 closes the gap between β_- and β_+ for a more general class of priors which includes the Rademacher and sparse Rademacher priors.

The proof of Theorem 1 is given in Chapter 3. Lemma 2 in Section 2.4 relates the total variation distance between T and W to the scalar-valued p-spin model defined in Section 2.2. The bulk of the work to prove Theorem 1 lies in investigating the behavior of the scalar-valued p-spin model at low vs. high values of β.

While Theorem 1 gives the existence of a phase transition in the detection problem, Theorem 2 provides a way to determine the value of the critical SNR β_c. Set

$$\xi(s) = \frac{s^p}{2}, \tag{1.2}$$

and

$$v_* = \int a^2 \mu(da). \tag{1.3}$$

For $b, s \geq 0$, define

$$\gamma_b(s) = \mathbb{E}\left[\frac{\left(\int aZ(a, b\xi'(s))\mu(da)\right)^2}{\int Z(a, b\xi'(s))\mu(da)}\right], \tag{1.4}$$

where $Z(a,t)$ is the geometric Brownian motion

$$Z(a,t) = \exp\left(aB_t - \frac{a^2 t}{2}\right)$$

with B_t a standard Brownian motion. Define an auxiliary function $\Gamma_b(v) : [0, \infty) \to \mathbb{R}$ by

$$\Gamma_b(v) = \int_0^v \xi''(s)(\gamma_b(s) - s)ds. \tag{1.5}$$

Theorem 2. *If $p \geq 3$ and μ is centered, then β_c is the largest β such that*

$$\sup_{v \in [0, v_*]} \Gamma_\beta(v) = 0.$$

For many priors, the function $\gamma_b(s)$ is straightforward to compute, meaning it is possible to numerically integrate equation (1.5) to evaluate the auxiliary function $\Gamma_b(s)$ at various values of β which allows one approximate β_c. For example, consider the sparse Rademacher prior where entries of $u = (u_1, \dots, u_N)$ are independently sampled from the distribution

$$\frac{\rho}{2}\delta_{-1/\sqrt{\rho}} + (1-\rho)\delta_0 + \frac{\rho}{2}\delta_{1/\sqrt{\rho}},$$

with parameter $\rho \in (0,1]$ that controls the sparsity of the vector u. When $\rho = 1$, this is the Rademacher prior. Figure 1.1 shows the critical SNR for the sparse Rademacher prior for tensors of order $p = 3, 4, 5, 10$ and sparsity parameter $\rho = .1, .2, .3, \dots, 1$. The solid line in each plot is the function

$$H(\rho) = 2\sqrt{-\rho\log\rho - (1-\rho)\log(1-\rho) + \rho\log 2},$$

which is the upper bound for β_c derived in [8].

For each combination of p and ρ,

$$\gamma_\beta(s) = \mathbb{E}\left[\frac{\rho\exp\left(-\frac{\beta^2 ps^{p-1}}{\rho}\right)\sinh^2\left(g\sqrt{\frac{\beta^2 ps^{p-1}}{2\rho}}\right)}{\rho\exp\left(-\frac{\beta^2 ps^{p-1}}{4\rho}\right)\cosh\left(g\sqrt{\frac{\beta^2 ps^{p-1}}{2\rho}}\right) + (1-\rho)}\right],$$

with g a standard Gaussian. Note that $v_* = \int a^2\mu(da) = 1$ for all parameters $\rho \in (0,1]$. To determine the critical SNR, the NIntegrate function from Wolfram Mathematica was used to compute test values of $\Gamma_b(v)$ for values of v in the interval $(0,1]$ in increments of .001. The critical value of β_c is reported as the largest b such that $\Gamma_b(v) \le 0$ for all test values of v, where the function $\Gamma_b(v)$ was computed for values of b in increments of .001.

The critical value β_c exhibits a tension between the sparsity of the vector u, which is expected to have ρ non-zero entries, and the magnitude of the non-zero entries, $1/\sqrt{\rho}$. For small values of ρ, the vector u is very sparse, but the non-zero entries of u are very large in absolute value, and these very large values are easy to detect. As ρ grows, the magnitude of the non-zero entries decreases, and the detection threshold β_c accordingly must increase. Eventually, the sparsity parameter ρ is large enough that there are enough non-zero entries for detection, even though the magnitude of these entries is smaller, and β_c begins to decrease. As p, the order of the tensor grows, the proportion of non-zero entries for a given ρ decreases, so maximum value of β_c occurs at higher values of ρ for higher-order tensors.

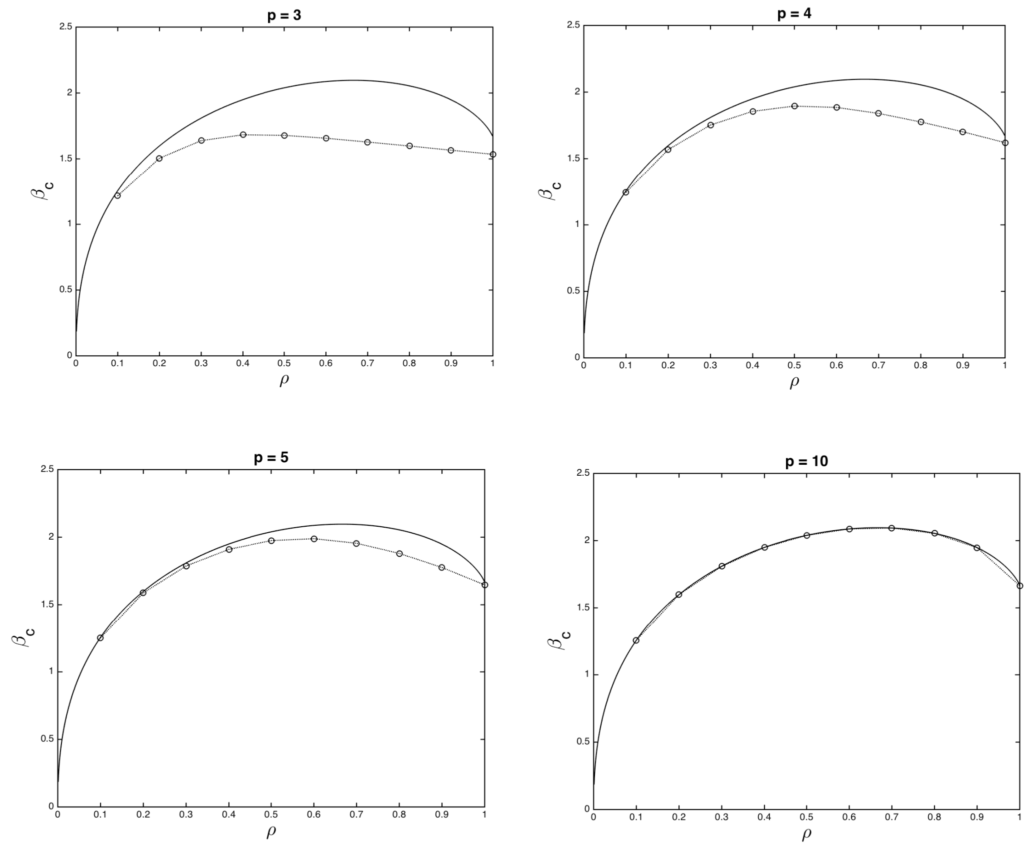

Figure 1.1: Numerical simulations for the critical value β_c with sparse Rademacher prior and various values of p. The top left plot is for $p = 3$, the top right for $p = 4$, the bottom left for $p = 5$ and the bottom right for $p = 10$. The open circles are the simulated critical values β_c. The dashed curve interpolates between these points and the solid curve describes the function $H(\rho)$.

1.4 Results for Detection with Multiple Spikes

The spiked random tensor with multiple spikes is formed as the sum of a Gaussian tensor and a linear combination of $k > 1$ rank one tensors weighted by possibly different SNRs $\beta_1, \ldots, \beta_k$. Theorem 3, the main result of this section, relates the set of SNRs for which detection is possible in the case of multiple spikes back to Theorem 1, the detection result for the case of a single spike. It is reasonable to guess that the presence of multiple rank one tensors would act as additional noise, increasing the SNR necessary for detection. Theorem 3 is remarkable because it shows that this does not happen: adding additional spikes does not add additional noise.

Fix $k > 1$. Consider bounded sets $\Lambda_1, \dots, \Lambda_k \subset \mathbb{R}$ and probability measures $\mu_1, \dots, \mu_k$ on the Borel σ-algebras of $\Lambda_1, \dots, \Lambda_k$ respectively. For each $1 \leq r \leq k$, assume that $u_1(r), \dots, u_N(r)$ are independent samples from μ_r that are also independent of the Gaussian tensor W. Define $u(r) = (u_1(r), \dots, u_N(r))$. Assume additionally that vectors $u(r), u(r')$ are independent of one another when $r \neq r'$. For $\beta_1, \dots, \beta_k > 0$, define $\bar{\beta} = (\beta_1, \dots, \beta_k)$. The spiked random tensor with k spikes and signal-to-noise ratio vector $\bar{\beta}$ is

$$T_k = W + \frac{1}{N^{(p-1)/2}} \sum_{r=1}^{k} \beta_r u(r)^{\otimes p}.$$

The definition of detection in the case of multiple spikes is identical to the single spike case.

Definition 1.4.1. Detection of the spikes $u(r)^{\otimes p}$ is possible if W, T_k are distinguishable according to Definition 1.2.1. Detection is not possible if W, T_k are indistinguishable.

For $1 \leq r \leq k$, let $\beta_{r,c}$ be the critical threshold for spike detection in the random tensor

$$W + \frac{\beta_r}{N^{(p-1)/2}} u(r)^{\otimes p}.$$

Theorem 3 states that in the case of multiple spikes, detection is possible when at least one signal-to-noise ratio exceeds the critical signal-to-noise ratio of its associated single-spike random tensor.

Theorem 3. *Assume that $\mu_1, \dots, \mu_k$ are centered. For $p \geq 3$,*

(i) If $\bar{\beta} \in (0, \beta_{1,c}) \times \cdots \times (0, \beta_{k,c})$ then detection is impossible.

(ii) If $\bar{\beta} \in (\beta_{1,c}, \infty) \times \cdots \times (\beta_{k,c}, \infty)$ then detection is possible.

The likelihood ratio of T_k and W is the free energy of the vector-valued p-spin model defined in Section 2.3. Unlike the scalar-valued case, the vector-valued p-spin model involves interactions between the different spikes. It turns out that, in the large-system limit, these interactions do not contribute to the behavior of the vector-valued model, and Theorem 3 follows.

1.5 Results for Recovery by MMSE

In addition to asking when one can detect the presence or absence of a spike, one may also ask if it is possible to recover any information about the spike itself. In this section we show that recovery of the spike by the minimum mean square error (MMSE) estimator has the same phase transition as the detection problem. Let $\hat{\theta} = (\hat{\theta}_{i_1,\ldots,i_p})_{1\leq i_1,\ldots,i_p\leq N}$ denote an $\mathbb{R}^{N^p}$-valued random variable generated by the σ-field $\sigma(T_k)$. The estimators $\hat{\theta}$ may depend on T_k, the vectors $u(r)$, or other randomness. The minimum mean square error is

$$\mathrm{MMSE}_N(\bar{\beta}) = \min_{\hat{\theta}} \frac{1}{N^p} \sum_{1\leq i_1,\ldots,i_p\leq N} \mathbb{E}\left(\sum_{r=1}^{k} u_{i_1}(r)\cdots u_{i,p}(r) - \hat{\theta}_{i_1,\ldots,i_p}\right)^2. \tag{1.6}$$

The minimum is achieved by the MMSE estimator

$$\hat{\theta}^{\mathrm{MMSE}}_{i_1,\ldots,i_p} = \sum_{r=1}^{k} \beta_r \mathbb{E}(u_{i_1}\cdots u_{i_p} \mid T_k).$$

A dummy estimator is any estimator that does not depend on the randomness of the vectors $u(r)$. The best dummy estimator is

$$\hat{\theta}_{i_1,\ldots,i_p} = \mathbb{E}\sum_{r=1}^{k} \beta_r u_{i_1}(r)\cdots u_{i_p}(r).$$

Therefore, replacing the minimum in equation (1.6) by the minimum over only dummy estimators gives the bound

$$\mathrm{MMSE}_N(\bar{\beta}) \leq \frac{1}{N^p} \sum_{1\leq i_1,\ldots,i_p\leq N} \left(\mathbb{E}\left(\sum_{r=1}^{k} \beta_r u_{i_1}(r)\cdots u_{i_p}(r)\right)^2 - \left(\mathbb{E}\sum_{r=1}^{k} \beta_r u_{i_1}(r)\cdots u_{i_p}(r)\right)^2\right).$$

Since the $u(r)$ are independent, when taking $N\to\infty$, the law of large numbers gives

$$\limsup_{N\to\infty} \mathrm{MMSE}_N(\bar{\beta}) \leq \mathrm{DMSE}(\bar{\beta}) := \sum_{r=1}^{N} \beta_r v_{*,r}^p$$

where $v_{*,r} = \int a^2 \mu_r(da)$.

Theorem 4 below states the precise result on recovery of the spikes by the MMSE estimator.

Theorem 4. *For $p \geq 3$,*

(i) If $\bar{\beta} \in (0, \beta_{1,c}) \times \cdots \times (0, \beta_{k,c})$ then $\limsup_{N\to\infty} \text{MMSE}_N(\bar{\beta}) = \text{DMSE}(\bar{\beta})$

(ii) If $\bar{\beta} \in (\beta_{1,c}, \infty) \times \cdots \times (\beta_{k,c}, \infty)$ then $\limsup_{N\to\infty} \text{MMSE}_N(\bar{\beta}) < \text{DMSE}(\bar{\beta})$

Lesieur et al. [10] prove the same result by computing the limiting mutual information between W and T_k. The proof of Theorem 4 included here uses a different method that relies on the detection results of the preceding sections.

1.5.1 Performance of Approximate Message Passing

Despite the fact that Theorem 4 gives us information about where the MMSE estimator performs better than random guessing, computation of the MMSE estimator is intractable. In the matrix case ($p = 2$), the performance of the approximate message passing (AMP) algorithm is well studied, and it is conjectured [10] that AMP achieves the best possible mean square error of any polynomial time algorithm. For the spiked matrix with Rademacher prior, [11] shows that, in the large-system limit, AMP is Bayes optimal and recovery by AMP exhibits the same phase transition at $\beta = 1$ as recovery by spectral methods such as PCA. For general priors, there is typically an 'easy' region of parameters β where AMP is Bayes optimal and MMSE $<$ DMSE, a 'hard' region where AMP is suboptimal, and an 'impossible' region where MMSE $=$ DMSE. See, for example, [12, 13, 14, 15] for a discussion of the performance of AMP in low-rank matrix estimation in different settings.

For centered priors, the tensor case is drastically different than the matrix case. In the tensor case with spikes from mean-zero priors, there is no 'easy' region AMP both achieves the minimum mean square error and performs better than a dummy estimator. Let MSE_{AMP} denote the mean square error achievable by AMP. For $\beta > \beta_c$, Lesieur, et al. [10] show that MMSE $< \text{MSE}_{AMP}$, and MMSE $= \text{MSE}_{AMP}$ only for $\beta < \beta_c$. However, from Theorem 4, $\text{MSE}_{AMP} = \text{MMSE} = \text{DMSE}$ when $\beta < \beta_c$. Thus, in the region where MMSE $<$ DMSE we unfortunately also have MMSE $< \text{MSE}_{AMP}$.

AMP is a variation of Belief Propagation (BP) algorithms used for inference on graphical models. Given a probability distribution that takes the form of a product of factors, one may form a bipartite graph consisting of one vertex for each factor of the probability distribution and one vertex for each variable. A factor vertex and a variable vertex are connected if the factor takes the variable as an argument. In belief propagation, messages are passed between connected factor and variable nodes. Messages are chosen so that, in the limit, they give information about the probability distribution in question. For example, in different versions of BP, the messages may converge to the marginal distributions of the variables or the max-marginals.

AMP takes advantage of symmetry in the structure of the graphical model to reduce the amount of information passed at each iteration of the algorithm and consequently reduce the complexity. Instead of keeping track of the entire probability distributions passed between nodes, AMP only keeps track of the mean and variance of the distributions. Additionally, updates for these parameters are calculated using mean field approximations that exploit the weak dependence of the incoming distribution on the index of the variable node. A key feature of AMP is an 'Onsager term' corresponding to a correction between the mean-field approximation and the original cavity field. This term arises naturally in the derivation of AMP as shown below.

We give a detailed derivation of AMP from BP in the case of recovery by MMSE for a tensor with a single spike. This closely follows the analysis in [16]. The case of k spikes is no harder and essentially only entails replacing scalar products by vector products throughout the derivation. However, for the sake of more concise notation, we only derive the single-spike case. Consider the spiked tensor

$$T = \frac{1}{N^{(p-1)/2}} u^{\otimes p} + \frac{1}{\beta} W.$$

This has the same signal-to-noise ratio as the original spiked tensor, but the parameter β is considered as part of the noise tensor instead of part of the spike. Due to symmetry, the final algorithm only depends on observing the tensor elements with $i_1 < \cdots < i_p$, but the derivation is easier with all elements included.

For simpler notation, we will use lower-case letters to denote a tensor index, for example $a = (i_1, \ldots, i_p)$. To develop AMP from BP, we need to distinguish between the modes of the tensor and create a variable for u_i appearing in the first mode, a variable for u_i appearing in the second node, ..., and variable for u_i appearing in the p-th mode. Let capital letters $1 \leq A, B, C, \cdots \leq p$ denote the place in the product $u_{i_1} \cdots u_{i_p}$. For example, the notation u_{iA} will mean that entry u_i stands in the A-th spot in the product and u_{jB} will mean entry u_j stands in the B-th place in the product. This will distinguish between different orderings of the same product. As we develop AMP, symmetry will remove this dependence on ordering, but it is helpful in developing the algorithm.

Let $U = u^{\otimes p}$. We are interested in the posterior distribution

$$P(U|T) = \frac{1}{Z(T,U)} \prod_{i=1}^{N} \mu(u_i) \prod_{a} P(T_a \mid U_a),$$

where $Z(T,U)$ is the appropriate normalizing constant. The factor graph of the posterior distribution consists of $\binom{N}{p} \cdot p!$ factor nodes, each corresponding to an index $(i_1, \ldots, i_p)$, and $N \cdot p$ variable nodes, each corresponding to a variable u_{iA} in position A of the product. Note that we do not include the factors $\mu(u_i)$ in the factor graph because, in BP, the messages at any factor node depending on a single variable never change. A factor node a and variable node iA are connected if i is the A-th element of the list $a = (i_1, \ldots, i_p)$.

Note that each likelihood is Gaussian with

$$P(T_a \mid U_a) \sim \mathcal{N}(U_a, 1/\beta^2).$$

Define a cost function by

$$g(T_a, U_a) = \log P(T_a \mid U_a) = -\frac{\beta^2}{2}(T_a - U_a)^2 - \frac{1}{2}\log\frac{2\pi}{\beta^2}.$$

Thus the posterior distribution can be written

$$P(U|T) = \frac{1}{Z(T,U)} \prod_{i=1}^{N} \mu(u_i) \prod_{a} e^{g(T_a,U_a)}. \tag{1.7}$$

The version of BP presented here is designed to compute the marginal posterior distributions of the variables $u_1, \ldots, u_N$. The simplification to AMP keeps track of only the first and second moments of these marginals. The means of the marginals are the estimators of the true signal.

For each factor node $a = (i_1, \ldots, i_p)$, denote the set of variable nodes connected to a by $\partial a = \{i_1 1, \ldots, i_p P\}$. Also, for each variable node i, denote the set of factor nodes connected to iA by $\partial iA = \{a \mid iA \in a\}$. Each iteration of BP consists of a set of messages passed from variable nodes to their neighboring factor nodes and a set of messages passed from the factor nodes back to the variable nodes. Let $\eta^t_{iA\to a}(x_i)$ denote the message passed from node iA to node a at step t of the algorithm, and similarly, $\tilde{\eta}^t_{a\to iA}(x_i)$ the messages passed from node a to node iA. The BP updates proceed by

$$\eta^t_{iA\to a}(u_{iA}) = \frac{\mu(u_{iA})}{Z^t_{iA\to a}} \prod_{b\in\partial iA\setminus a} \tilde{\eta}^{t-1}_{b\to iA}(u_{iA}) \tag{1.8}$$

$$\tilde{\eta}^t_{b\to iA}(u_{iA}) = \frac{1}{Z^t_{b\to iA}} \int e^{g(T_b, N^{(1-p)/2}U_b)} \prod_{jB\in\partial b\setminus iA} d\eta^t_{jB\to b}(u_{jB}). \tag{1.9}$$

The normalizing constants $Z^t_{iA\to a}$ and $Z^t_{b\to iA}$ are chosen to make $\eta^t_{iA\to a}$ and $\tilde{\eta}^t_{b\to iA}$ probability measures. At each step, $\eta^t_{iA\to a}$ is the product of incoming messages from the factor nodes. These messages are weighted by the cost function at that factor node, marginalized, and passed back to the variable nodes.

The simplification to AMP proceeds in two steps. First, the messages are expanded to quadratic order, after which it is only necessary to keep track of the coefficients of the linear and quadratic terms of the expansion rather than entire probability distribution. In the second step, mean-field approximations are used to remove the dependence on the factor nodes.

Step 1: Quadratic Approximation

Define

$$S_b = \frac{\partial}{\partial U_b} g(T_b, U_b)\Big|_{U_b=0}$$

$$R_b = \frac{\partial^2}{\partial U_b^2} g(T_b, U_b)\Big|_{U_b=0} + \left(\frac{\partial}{\partial U_b} g(T_b, U_b)\Big|_{U_b=0}\right)^2.$$

Let $\mathcal{E}$ denote any error term of order $O(N^{1-p})$ or smaller. These terms will vanish in the large-system limit. Expanding $g(T_b, N^{(1-p)/2}U_b)$ about $U_b = 0$ gives

$$g(T_b, N^{(1-p)/2}U_b) = g(T_b, 0) + \frac{1}{N^{(p-1)/2}} S_b U_b + \frac{1}{N^{p-1}} \frac{R_b - S_b^2}{2} U_b^2 + \mathcal{E}, \tag{1.10}$$

and expanding $\exp g(T_b, N^{(1-p)/2}U_b)$ about $U_b = 0$ gives

$$e^{g(T_b, N^{(1-p)/2}U_b)} = e^{g(T_b,0)} \left(1 + \frac{1}{N^{(p-1)/2}} S_b U_b + \frac{1}{N^{p-1}} R_b U_b^2\right) + \mathcal{E}. \tag{1.11}$$

Plugging the second expansion (1.11) into the factor-to-variable message (1.9) gives

$$\tilde{\eta}_{b\to iA}(u_{iA}) = \frac{e^{g(T_b,0)}}{Z_{b\to iA}} \left(1 + \frac{1}{N^{(p-1)/2}} S_b \int U_b \prod_{j\in\partial b\setminus i} d\eta_{jB\to b}(u_{jB}) + \frac{1}{N^{p-1}} R_b \int U_b^2 \prod_{j\in\partial b\setminus i} d\eta_{jB\to b}(u_{jB})\right) + \mathcal{E}. \tag{1.12}$$

Recall that $U_b = u_{iA} \prod_{jB\in\partial b\setminus iA} u_{jB}$. Define $\hat{u}_{jB\to b}$ and $\hat{\sigma}_{jB\to b}$ as the mean and variance, respectively, of the distribution $\eta_{jB\to b}(u_{jB})$. Expanding the two integrals above,

$$\int U_b \prod_{jB\in\partial b\setminus iA} d\eta_{jB\to b}(u_{jB}) = u_{iA} \prod_{jB\in\partial b\setminus i} \int u_{jB} d\eta_{jB\to b}(u_{jB}) = u_{iA} \prod_{jB\in\partial b\setminus iA} \hat{u}_{jB\to b},$$

and

$$\int U_b^2 \prod_{jB\in\partial b\setminus iA} d\eta_{jB\to b}(u_{jB}) = u_{iA}^2 \prod_{jB\in\partial b\setminus iA} \int u_{jB}^2 d\eta_{jB\to b}(u_{jB}) = u_{iA}^2 \prod_{jB\in\partial b\setminus iA} (\hat{\sigma}_{jB\to b} + \hat{u}_{jB\to b}^2).$$

We can then consider $\tilde{\eta}_{b\to iA}$ as a power series in u_{iA}:

$$\tilde{\eta}_{b\to iA}(u_{iA}) = \frac{e^{g(T_b,0)}}{Z_{b\to iA}}\left(1 + \frac{u_{iA}}{N^{(p-1)/2}} S_b \prod_{jB\in\partial b\setminus iA} \hat{u}_{jB\to b} + \frac{u_{iA}^2}{N^{p-1}} R_b \prod_{jB\in\partial b\setminus iA} (\hat{\sigma}_{jB\to b} + \hat{u}_{jB\to b}^2)\right) + \mathcal{E}.$$

Using the equivalence of the two expansions (1.10) and (1.11) gives

$$\begin{aligned}\tilde{\eta}_{b\to iA}(u_{iA}) = &\frac{e^{g(T_b,0)}}{Z_{b\to iA}} \exp\left(\frac{1}{N^{(p-1)/2}} S_b u_{iA} \prod_{jB\in b\setminus iA} \hat{u}_{jB\to b} + \frac{1}{N^{p-1}}\frac{R_b}{2} u_{iA}^2 \prod_{jB\in\partial b\setminus iA} (\hat{\sigma}_{jB\to b} + \hat{u}_{jB\to b}^2)\right.\\ &\left.- \frac{1}{N^{p-1}} S_b u_{iA}^2 \prod_{jB\in\partial b\setminus iA} \hat{u}_{jB\to b}^2\right) + \mathcal{E}. \qquad (1.13)\end{aligned}$$

Next, substituting expression (1.13) into the variable-to-factor messages (1.8) gives

$$\begin{aligned}\eta_{iA\to a}(u_{iA}) = \frac{\mu(u_{iA})}{Z_{iA\to a}} \exp&\left(\frac{u_{iA}}{N^{(p-1)/2}} \sum_{b\in\partial iA\setminus a} S_b \prod_{jB\in\partial b\setminus iA} \hat{u}_{jB\to b} + \frac{u_{iA}^2}{2N^{p-1}} \sum_{b\in\partial iA\setminus a} R_b \prod_{jB\in\partial b\setminus iA} (\hat{\sigma}_{jB\to b} + \hat{u}_{jB\to b}^2)\right.\\ &\left.- \frac{u_{iA}^2}{2N^{p-1}} S_b \prod_{jB\in\partial b\setminus iA} \hat{u}_{jB\to b}^2\right) + \mathcal{E}.\end{aligned}$$

All constant terms have been absorbed into the normalization constant $Z_{iA\to a}$. Recall that the original error terms on the expansions were of order $O(N^{3(1-p)/2})$. The product in the above expression is a product over $p-1$ factors. Since $p \geq 3$, the combined error is of order $O(N^{1-p})$, so we can still represent it by $\mathcal{E}$.

Define constants

$$B_{iA\to a} = \frac{1}{N^{(p-1)/2}} \sum_{b\in\partial iA\setminus a} S_b \prod_{jB\in\partial b\setminus iA} \hat{u}_{jB\to b}$$

and

$$A_{iA\to a} = \frac{1}{N^{p-1}}\left(\sum_{b\in\partial iA\setminus a} S_b^2 \prod_{jB\in\partial b\setminus iA} \hat{u}_{jB\to b}^2 - \sum_{b\in\partial iA\setminus a} R_b \prod_{jB\in\partial b\setminus iA} (\hat{\sigma}_{jB\to b} + \hat{u}_{jB\to b}^2)\right)$$

so that

$$\eta_{iA\to a}(u_{iA}) = \frac{\mu(u_{iA})}{Z_{iA\to a}} e^{B_{iA\to a}u_{iA} - \frac{A_{iA\to a}u_{iA}^2}{2}}.$$

The normalization constant is

$$Z_{iA\to a} = \int d\mu(u_{iA}) e^{B_{iA\to a}u_{iA} - \frac{A_{iA\to a}u_{iA}^2}{2}}.$$

Define a function

$$f(A, B) = \frac{d}{dB} \log \int d\mu(u_{iA}) e^{Bu_{iA} - \frac{Au_{iA}^2}{2}}.$$

Recall that $\hat{u}_{iA\to a}$ and $\hat{\sigma}_{iA\to a}$ are the mean and variance of $\mu_{iA\to a}(u_{iA})$. We can write

$$\hat{u}_{iA\to a} = f(A_{iA\to a}, B_{iA\to a})$$

and

$$\hat{\sigma}_{iA\to a} = \frac{\partial}{\partial B} f(A_{iA\to a}, B_{iA\to a}).$$

Reintroducing the parameter t to keep track of the iterations, we have a quadratic approximation to belief propagation that proceeds by updates

$$\begin{aligned}
B_{iA\to a}^t &= \frac{1}{N^{(p-1)/2}} \sum_{b\in\partial iA\setminus a} S_b \prod_{jB\in\partial b\setminus iA} \hat{u}_{jB\to b}^t \\
A_{iA\to a}^t &= \frac{1}{N^{p-1}} \left(\sum_{b\in\partial iA\setminus a} S_b^2 \prod_{jB\in\partial b\setminus i} (\hat{u}^t)_{jB\to b}^2 - \sum_{b\in\partial iA\setminus a} R_b \prod_{jB\in\partial b\setminus iA} (\hat{\sigma}_{jB\to b}^2 + (\hat{u}^t)_{jB\to b}^2) \right) \\
\hat{u}_{iA\to a}^{t+1} &= f(A_{iA\to a}^t, B_{iA\to a}^t) \\
\hat{\sigma}_{iA\to a}^{t+1} &= \frac{\partial}{\partial B} f(A_{iA\to a}^t, B_{iA\to a}^t).
\end{aligned}$$

This algorithm necessitates updating constants rather than probability distributions at each iteration.

Step 2: Mean Field Approximation

In the preceding quadratic approximation to BP, the constants $B_{iA\to a}$, $A_{iA\to a}$, $\hat{u}_{iA\to a}$, and $\hat{\sigma}_{iA\to a}$ depend only weakly on the target node a. The next simplification exploits this to remove the dependence on the target node a resulting in a set of constants

$B_{iA}, A_{iA}, \hat{u}_{iA}, \hat{\sigma}_{iA}$ that depend only on the variable node, reducing the number of constants updated at each step.

Define

$$A_{iA}^t = \frac{1}{N^{p-1}} \left(\sum_{b \in \partial iA} S_b^2 \prod_{jB \in \partial b \setminus iA} (\hat{u}^t)_{jB \to b}^2 - \sum_{b \in \partial iA \setminus a} R_b \prod_{jB \in \partial b} (\hat{\sigma}_{jB \to b}^2 + (\hat{u}^t)_{jB \to b}^2) \right) \tag{1.14}$$

Since $A_{iA \to a} - A_{iA} \sim O(N^{1-p})$, we may freely replace $A_{iA \to a}$ by A_{iA} and disregard the error terms. Next, define

$$B_{iA}^t = \frac{1}{N^{(p-1)/2}} \sum_{b \in \partial iA} S_b \prod_{jB \in \partial b \setminus iA} \hat{u}_{jB \to b}^t, \tag{1.15}$$

and define the correction term by

$$\delta B_{iA \to a}^t = B_{iA}^t - B_{iA \to a}^t = \frac{1}{N^{(p-1)/2}} S_a \prod_{jB \in \partial a \setminus iA} \hat{u}_{jB \to a}^t.$$

We also need to define statistics $\hat{u}_{iA}$ and $\hat{\sigma}_{iA}$ that do not depend on a target variable. Define $\hat{u}_{iA}^t = f(A_{iA}^{t-1}, B_{iA}^{t-1})$ and $\hat{\sigma}_{iA}^t = \frac{\partial}{\partial B} f(A_{iA}^{t-1}, B_{iA}^{t-1})$. The correction term for the mean is

$$\delta \hat{u}_{iA \to a}^t = \hat{u}_{iA}^t - \hat{u}_{iA \to a}^t = f(A_{iA}^{t-1}, B_{iA}^{t-1}) - f(A_{iA \to a}^{t-1}, B_{iA \to a}^{t-1}) = f(A_{iA}^{t-1}, B_{iA}^{t-1}) - f(A_{iA}^{t-1}, B_{iA \to a}^{t-1}) + \mathcal{E}.$$

The last equality uses the fact that the $A_{iA}^{t-1} - A_{iA \to a}^{t-1} \sim O(N^{1-p})$. Using the identity $B_{iA \to a}^{t-1} = \delta B_{iA}^{t-1} - \delta B_{iA \to a}^{t-1}$, we can consider the function $f(A_{iA}^{t-1}, B_{iA \to a}^{t-1})$ as a function of $\delta B_{iA \to a}^t$ and expand to linear order about $\delta B_{iA \to a}^t$ to get

$$\begin{aligned} \delta \hat{u}_{iA \to a}^t &= f(A_{iA}^{t-1}, B_{iA}^{t-1}) - \left(f(A_{iA}^{t-1}, B_{iA}^{t-1}) - \frac{\partial}{\partial B} f(A_{iA}^{t-1}, B_{iA}^{t-1})(B_{iA}^{t-1} - B_{iA \to a}^{t-1}) \right) \\ &= \frac{\partial}{\partial B} f(A_{iA}^{t-1}, B_{iA}^{t-1})(B_{iA}^{t-1} - B_{iA \to a}^{t-1}) \\ &= \hat{\sigma}_{iA}^t \frac{1}{N^{(p-1)/2}} S_a \prod_{jB \in \partial a \setminus iA} \hat{u}_{jB \to a}^{t-1}. \end{aligned} \tag{1.16}$$

We need only expand to linear order because all quadratic terms are of order $O(N^{1-p})$ and can be neglected. Next, write $\hat{u}^{t-1}_{jB\to a} = \hat{u}^{t-1}_{jB} - \delta\hat{u}^{t-1}_{jB\to a}$ to get

$$\frac{1}{N^{(p-1)/2}} \prod_{jB\in\partial a\setminus iA} \hat{u}^{t-1}_{jB\to a} = \frac{1}{N^{(p-1)/2}} \prod_{jB\in\partial a\setminus iA} (\hat{u}^{t-1}_{jB} - \delta\hat{u}^{t-1}_{jB\to a}) = \frac{1}{N^{(p-1)/2}} \prod_{jB\in\partial a\setminus iA} \hat{u}^{t-1}_{jB} + \mathcal{E}.$$

Substituting this into equation (1.16) gives

$$\delta\hat{u}^{t}_{iA\to a} = \hat{\sigma}^{t}_{iA} \frac{1}{N^{(p-1)/2}} S_a \prod_{jB\in\partial a\setminus iA} \hat{u}^{t-1}_{iA}. \tag{1.17}$$

To remove the dependence of B^t_{iA} on the factor node through the terms $\hat{u}^t_{jB\to b}$, recall that $\hat{u}^t_{jB\to b} = \hat{u}^t_{jB} - \delta\hat{u}^t_{jB\to b}$. Substituting this into (1.15) gives

$$\begin{aligned} B^t_{iA} &= \frac{1}{N^{(p-1)/2}} \sum_{b\in\partial iA} S_b \prod_{jB\in\partial b\setminus iA} (\hat{u}^t_j - \delta\hat{u}^t_{jB\to b}) \\ &= \frac{1}{N^{(p-1)/2}} \sum_{b\in\partial iA} S_b \left(\prod_{jB\in\partial b\setminus iA} \hat{u}^t_{jB} - \sum_{jB\in\partial b\setminus iA} \delta\hat{u}^t_{jB\to b} \prod_{kC\in\partial b\setminus\{iA,jB\}} \hat{u}^t_{kC} \right) + \mathcal{E}. \end{aligned}$$

Next, substituting the expression (1.17) for $\delta\hat{u}^t_{jB\to b}$ gives

$$B^t_{iA} = \frac{1}{N^{(p-1)/2}} \sum_{b\in\partial iA} S_b \prod_{jB\in\partial b\setminus iA} \hat{u}^t_{jB} - \frac{\hat{u}^{t-1}_{iA}}{N^{p-1}} \sum_{b\in\partial iA} S_b^2 \hat{\sum}_{jB\in\partial b\setminus iA} \hat{\sigma}^t_{jB} \prod_{kC\in\partial b\setminus\{iA,jB\}} \hat{u}^t_{kC} \hat{u}^{t-1}_{kC}.$$

The second term in the preceding expression is called the Onsager correction term and corrects for the difference between the mean-field approximation where one sums over all $b \in \partial iA$ and the original cavity field where one sums over $b \in \partial iA \setminus a$. This term is closely related to the TAP equation for spin glasses and decouples the iterations of the algorithm. The original BP algorithm is guaranteed to converge to the true marginal distributions if the underlying graph is a tree. In the case of a tree graph, the messages never backtrack. The graph for spiked tensors is quite loopy, and the Onsager term corrects for backtracking caused by these loops.

All that is left is to remove the dependence of A^t_{iA} on the target nodes. Recall that

$$S_b = \frac{\partial}{\partial U} g(T_b, U_b)\bigg|_{U_b=0} = \beta^2 T_b.$$

Therefore, taking expectation with respect to the posterior distribution, one sees that

$$\mathbb{E}_P S_b^2 = \beta^2,$$

where $\mathbb{E}_P$ denotes expectation with respect to the posterior. Then, since

$$R_b = \frac{\partial^2}{\partial U^2} g(T_b, U_b)\bigg|_{U_b=0} + S_b^2 = -\beta^2 + S_b^2,$$

we see that $\mathbb{E}_P R_b = 0$. The quantities S_b and R_b are self-averaging, so replacing S_b^2 and R_b in equation (1.14) by the averages gives

$$\begin{aligned} A^t_{iA} &= \frac{\beta^2}{N^{p-1}} \sum_{b\in\partial iA} \prod_{jB\in\partial b\setminus iA} (\hat{u}^t)^2_{jB\to b} \\ &= \frac{\beta^2}{N^{p-1}} \prod_{B\neq A} \sum_{j=1}^{N} (\hat{u}^t_{jB})^2. \end{aligned}$$

Notice that A^t_{iA} does not actually depend on the variable node iA. Because of this, we may define

$$A^t = \frac{\beta^2}{N^{p-1}} \prod_{B\neq A} \sum_{j=1}^{N} (\hat{u}^t_{jB})^2.$$

A similar computation for the Onsager correction term gives

$$\begin{aligned} &\frac{\hat{u}^{t-1}_{iA}}{N^{p-1}} \sum_{b\in\partial iA} S_b^2 \sum_{jB\in\partial b\setminus iA} \hat{\sigma}^t_{jB} \prod_{kC\in\partial b\setminus\{iA,jB\}} \hat{u}^t_{kC}\hat{u}^{t-1}_{kC} \\ &= \hat{u}^{t-1}_{iA}\frac{\beta^2}{N^{p-1}} \sum_{b\in\partial iA} \sum_{jB\in\partial b\setminus iA} \hat{\sigma}^t_{jB} \prod_{kC\in\partial b\setminus\{iA,jB\}} \hat{u}^t_{kC}\hat{u}^{t-1}_{kC} \\ &= \hat{u}^{t-1}_{iA}\frac{\beta^2}{N^{p-1}} \sum_{B\neq A} \sum_{j} \hat{\sigma}^t_{jB} \prod_{C\neq A,B} \left(\sum_{k=1}^{N} \hat{u}^t_{kC}\hat{u}^t_{kC} \right) \end{aligned}$$

Finally, symmetry between modes of the tensor gives the simplest version of AMP:

$$
\begin{aligned}
B_i^t &= \frac{\beta^2}{N^{(p-1)/2}} \sum_{i_2<\cdots<i_p} T_{i,i_2,\ldots,i_p} \hat{u}_{i_1}^t \cdots \hat{u}_{i_p}^t - \frac{(p-1)\beta^2}{N^{p-1}} \left(\sum_{j=1}^N \hat{\sigma}_j^t \right) \left(\sum_{j=1}^N \hat{u}_j^t \hat{u}_j^{t-1} \right)^{p-2} \hat{u}_i^{t-1} \\
A^t &= \frac{\beta^2}{N^{p-1}} \left(\sum_{j=1}^N \hat{u}_j^t \hat{u}_j^t \right)^{p-1} \\
\hat{u}_i^{t-1} &= \lambda \hat{u}_i^t + (1-\lambda) f(A, B_i) \\
\hat{\sigma}_i^{t-1} &= \lambda \hat{\sigma}_i^t + (1-\lambda) \frac{d}{dB} f(A, B_i).
\end{aligned}
$$

The parameter $\lambda \in [0,1)$ is a damping constant that controls the step-size at each iteration. Setting $\lambda = 0$ gives the straightforward AMP which can sometimes oscillate around a fixed value instead of converging. Setting $\lambda > 0$ helps avoid this behavior. Other damping schemes may also be considered, but this is the most straightforward.

The performance of AMP is generally analyzed by studying the overlap between the estimator and the ground truth, defined as

$$
M^t = \frac{1}{N} \sum_{i=1}^N \hat{u}_i^t u_i.
$$

The updates to this quantity are the state evolution equations

$$
\begin{aligned}
M^{t+1} &= \mathbb{E}_{z,x_0} \left[\frac{d}{db} \log Z(\hat{M}^t, \hat{M}^t u_0 + (\hat{M}^t)^{1/2} z) u_0 \right] \\
\hat{M}^t &= \beta^2 (M^t)^{p-1} \qquad (1.18)
\end{aligned}
$$

with u_0 the ground truth and $z \sim \mathcal{N}(0,1)$. The expectation is with respect to both the Gaussian random variable z as well as the distribution of the ground truth. The state evolution equations are derived from the AMP algorithm by computing updates for the quantity M^t. See [16] for a careful derivation of the state evolution equations.

AMP converges to a stationary point of the state evolution equations. When the expected value of the signal prior is zero, zero is a fixed point of the stationary equations meaning that AMP may converge to a solution $\hat{u}$ such that the overlap of $\hat{u}$ and the true signal u_0 is zero, so the estimate is completely uninformative. The value β_c of Theorem

1 is the border between the hard and impossible regions, and, assuming $\beta > \beta_c$, whether or not zero is a stable fixed point determines whether one is in the easy or hard region. Denote the transition between the easy and hard regions by β_{Alg}.

As an example, consider the sparse Rademacher prior (which is the Rademacher prior when $\rho = 1$). A fixed point of the state evolution equations (1.18) satisfies

$$M = \rho \mathbb{E}_z \tanh\left(\beta^2 M^{p-1} + \frac{\beta M^{p-1/2}}{\sqrt{\rho}} z\right) \frac{1}{\rho + (1-\rho)\frac{e^{\beta^2 M^{p-1}/2\rho}}{\cosh(\beta^2 M^{p-1} + \frac{\beta M^{p-1/2}}{\sqrt{\rho}} z)}}.$$

For the matrix case with Rademacher prior ($p = 2$, $\rho = 1$), this simplifies to

$$M = \mathbb{E}_z \tanh(\beta^2 M + \beta\sqrt{M} z).$$

Cleary $M = 0$ is a fixed point. Gaussian integration by parts yields

$$\frac{d}{dM}\mathbb{E}_z \tanh(\beta^2 M + \beta\sqrt{M} z)\Big|_{M=0} = \beta^2.$$

Thus the fixed point $M = 0$ is stable for $\beta < 1$, so $\beta_{Alg=1}$. This means that the easy region is $\beta > 1$. It is well known that in this setting, $\beta_c = 1$, so in fact for the matrix case with Rademacher prior, as mentioned above, no hard phase exists. The matrix case with sparse Rademacher prior exhibits a phase transition in ρ. For large enough ρ, zero is the only stable solution of the state evolution equation and also $\beta_{Alg} = \beta_c$, so the performance of AMP is similar to the Rademacher case where there is no hard phase. For smaller ρ, there are multiple stable fixed points of the state evolution equations and additionally $\beta_{Alg} > \beta_c$, so a hard phase exists. See [15] for a detailed discussion of the sparse Rademacher matrix case.

When $p \geq 3$ and the spike prior has mean zero, zero is a stable fixed point of the state evolution equations for all β. For example, in the Rademacher case, it is straightforward to compute that

$$\frac{d}{dM}\mathbb{E}_z \tanh(\beta^2 M^{p-1} + \beta M^{(p-1)/2} z)\Big|_{M=0} = 0.$$

This means there is no easy region for spike recovery by AMP in the tensor case.

In Figure 1.2, we compare the performance of AMP for the Rademacher prior in the case of $p = 2$ and $p = 3$ when AMP is initialized at an 'uninformative' vector that has no correlation with the ground truth [1]. For the Rademacher prior, there are two fixed points of AMP, one with low error that has mean square error less than 1, and one with high error with mean square error equal to 1. For $p = 2$ and $\beta < 1$, AMP always converges to the high-error fixed point. However, for $\beta > 1$, as β grows, AMP converges to the low-error solution more frequently. In stark contrast, for $p = 3$, AMP never converges to the high-error solution, regardless of the value of β.

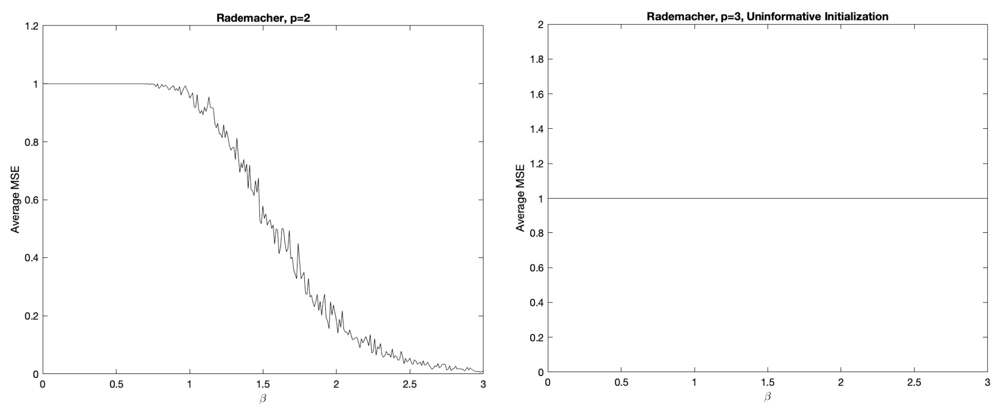

Figure 1.2: *Average mean-square error for spike recovery by AMP for* $N = 50$ *and 50 trials.* For $p = 2$ and $p = 3$ we run 50 trials of AMP and plot the averaged mean square error for each values of β between 0 and 3 in increments of .01. For each trial, we initialize $u_i^0 \sim N(0, 1)$. This is an 'uninformative' initialization that has no correlation with the ground truth. The plot on the left is for $p = 2$ and clearly shows the phase transition at $\beta = 1$. The plot on the right is for $p = 3$. For $p = 3$, AMP always converges to the high-error fixed point.

In Figure 1.3, we show some results for AMP for the spiked tensor with Rademacher prior and $p = 3$ initialized at a vector u^0 that is correlated with the ground truth vector. For each run of AMP, we initialize $u_i^0 = u_i + \mathcal{N}(0, .01)$, where $u = (u_1, \ldots, u_N)$ is the ground truth. As β surpasses the critical threshold $\beta_c = 1.535$, some runs of AMP converge to the low-error fixed point. For each value of $\beta \in [0, 3]$ at increments of .01, we run AMP until we achieve 20 convergent runs. As β increases, more of the runs

[1] Code is available at `https://github.com/mchandschy/UMN_Phd_Thesis_AMP`

oscillate and do not converge, but the percentage of runs that converge to the high-error fixed point goes to zero. Figure 1.3 shows the average mean square error across the 20 convergent runs, the total number of runs performed, and the percentage of runs converging to the low- and high- error fixed points. We perform all runs with $\lambda = 0$.

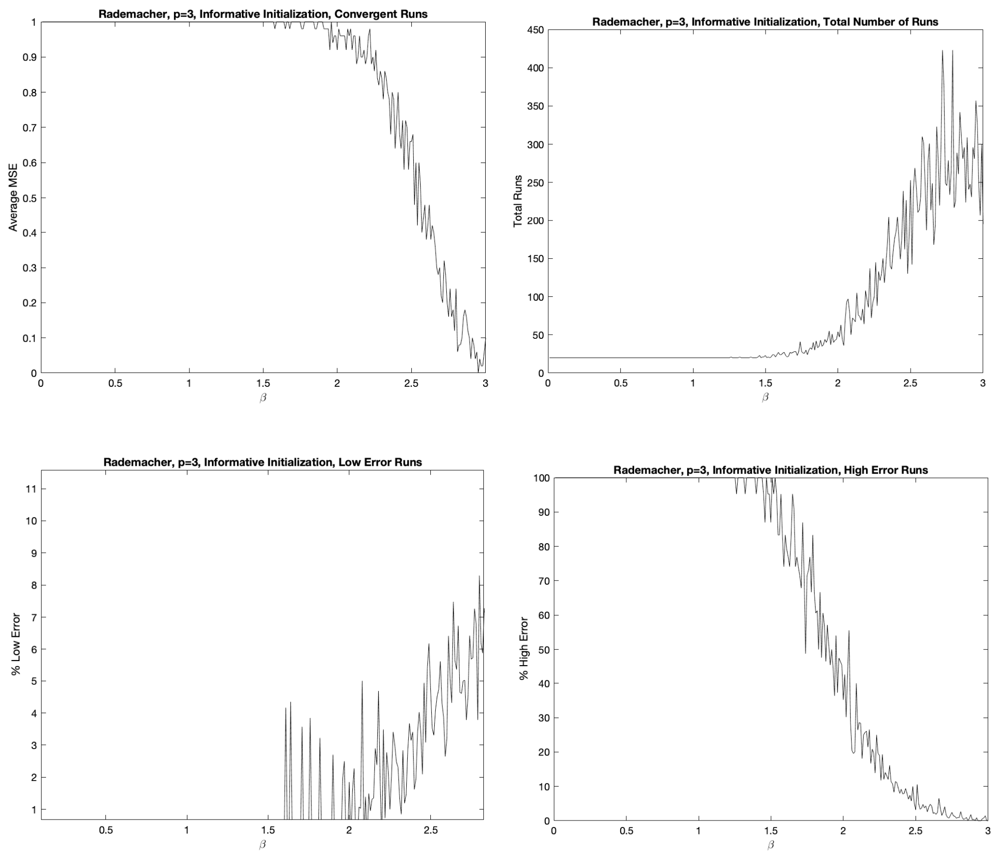

Figure 1.3: *AMP for* $N = 50$ *with informative initialization.* For each value of β, we run AMP until we achieve 20 convergent runs. The top left plot shows the average mean square error over the 20 convergent runs. The top right plot shows the total number of runs performed for each value of β. The bottom left and bottom right plots, respectively, show the percentage of runs that converge to the low-error solution and the percentage of runs that converge to the high-error solution.

Chapter 2

Pure p-Spin Models

This chapter introduces the mean field spin glass models related to the spike detection problem and explains how these models relate to the total variation distance between the spiked and unspiked random tensors. Section 2.2 introduces the scalar-valued p-spin model corresponding to the single-spike case, and Section 2.3 introduces the vector-valued p-spin model corresponding to the case of multiple spikes. In Section 2.4, Lemmas 1 and 2 relate the free energy of these spin glass models to the total variation distance between the spiked and un-spiked random tensors. Finally, Section 2.5 outlines how the results about the spin-glass models are used to prove Theorems 1, 2, 3, and 4.

2.1 A Brief History of Spin Glasses

In the late 1950s, physicists began to study the magnetic resonance of alloys composed of ions in non-magnetic metal in an effort to study magnetic interactions between the ions [17]. In studies of manganese ions in non-magnetic copper, Owen, Browne, Knight, and Kittel [18] noticed that alloys composed of .1-10% manganese exhibited anomalous behavior at low temperatures. Observations of the specific heat - the amount of heat required to raise the temperature of a unit mass of material by one degree Celcius - suggested a magnetic phase transition. Below a critical temperature, the magnetic spins froze in random directions rather than settling into a more orderly ferromagnetic or anti-ferromagnetic configuration [19, 20]. Inspired by this disordered magnetic structure, these alloys came to be known as 'classic spin glasses;' the name is an analogy to the disordered molecular structure of a glass compared to the regular molecular structure of a crystal.

The mathematical study of spin glasses began with the insight by Edwards and Anderson [21] to model the interactions between the ions by the random Heisenberg model. In this model, known as the Edwards-Anderson (EA) spin glass model, Ising spins $\sigma_1, \ldots, \sigma_N \in \{+1, -1\}$ are situated on a lattice where spin σ_i is an assignment of the value $+1$ or -1 to site i. Interactions between spins at neighboring sites on the lattice are modeled by independent standard Gaussians, where $g_{ij} \sim \mathcal{N}(0,1)$ models the interaction between the spins at sites i and j. A choice of $\sigma_i = +1$ or $\sigma_i = -1$ for each site on the lattice is called a spin configuration. Given a spin configuration, the energy of the system is given by the Hamiltonian $\sum_{\langle ij \rangle} g_{ij}\sigma_i\sigma_j$. The notation $\langle i, j \rangle$ indicates that the sum is taken only over neighboring sites on the lattice.

In the Edwards-Anderson model, the interaction between two spins depends on the distance between them: two spins interact only if they are neighbors. The EA model and other distance-dependent models are difficult to analyze, which led to the introduction of the so-called 'infinite range' models in which distance is ignored. The Sherrington-Kirkpatrick (SK) model for spin-glasses, originally formulated in [22], generalizes the EA model to an infinite range model by instead summing over *all* pairs of sites. The SK model is one of the most well-studied and well-understood spin glass models. For an in depth review of the SK model, see [23, 24, 25]. The models studied here are an extension of the SK model, where spins take on values in a bounded subset of $\mathbb{R}$ and the sum is over all interactions between sets of p spins.

2.2 The Scalar-Valued p-Spin Model

Recall the symmetric random tensor W defined in equation (1.1). Recall also the probability space (Λ, μ) from which the entries of the vector u are drawn. Given a spin configuration $\sigma = (\sigma_1, \ldots, \sigma_N) \in \Lambda^N$, the Hamiltonian of the scalar-valued pure p-spin model is

$$X_N(\sigma) = \frac{1}{N^{(p-1)/2}} \sum_{1 \le i_1, \ldots, i_p \le N} Y_{i_1, \ldots, i_p} \sigma_{i_1} \cdots \sigma i_p = \frac{1}{N^{(p-1)/2}} \langle Y, \sigma^{\otimes p} \rangle.$$

By the symmetry of the Hamiltonian, one may replace Y by W to get $X_N(\sigma) = \frac{1}{N^{(p-1)/2}}\langle W, \sigma^{\otimes p}\rangle$. Let σ^1, σ^2 denote two different spin configurations. The Hamiltonian $X_N(\sigma)$ is a Gaussian process indexed by Λ^N with covariance structure

$$\mathbb{E}X_N(\sigma^1)X_N(\sigma^2) = N(R(\sigma^1,\sigma^2))^p,$$

where $R(\sigma^1,\sigma^2)$ is the overlap between configurations σ^1, σ^2 defined as

$$R(\sigma^1,\sigma^2) = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^1\sigma_i^2. \tag{2.1}$$

Define the normalized Hamiltonian at inverse temperature β by

$$H_{N,\beta}(\sigma) = \beta X_N(\sigma) - \frac{\beta^2 N}{2}R(\sigma^1,\sigma^2)^p. \tag{2.2}$$

It is normalized in the sense that

$$\mathbb{E}e^{H_{N,\beta}(\sigma)} = 1.$$

The Hamiltonian 2.2 induces a natural probability measure on the space Λ^N known as the Gibbs measure and defined by

$$G_{N,\beta}(\sigma) = \frac{e^{H_{N,\beta}}\mu^{\otimes N}(d\sigma)}{Z_{N,\beta}}.$$

The normalizing constant $Z_{N,\beta}$, called the partition function, is

$$Z_{N,\beta} = \int_{\Lambda^N} e^{H_{N,\beta}(\sigma)}\mu^{\otimes N}(d\sigma).$$

Let $\langle\cdot\rangle_\beta$ denote expectation with respect to the Gibbs measure. At larger values of β (low temperature), the Gibbs measure concentrates on spin configurations σ that maximize $H_{N,\beta}(\sigma)$. In contrast, smaller values of β (high temperature) do not amplify the differences between values of $H_{N,\beta}(\sigma)$ for different configurations σ. Consequently, at high temperatures, sampling from the Gibbs measure is likely to produce common, as opposed to large, values of $H_N(\sigma)$.

The free energy associated to $H_{N,\beta}$ is

$$F_N(\beta) = \frac{1}{N} \log Z_{N,\beta}.$$

Denote the thermodynamic limit of the free energy by $F(\beta)$:

$$F(\beta) = \lim_{N\to\infty} F_N(\beta).$$

In Chapter 4 it is shown that for every $\beta > 0$ this limit exists and is non-random. For all $N \geq 1$ and $\beta > 0$, Jensen's inequality gives

$$\mathbb{E}F_N(\beta) \leq 0.$$

Therefore, taking the limit as $N \to \infty$ shows that $F(\beta) \leq 0$ for all $\beta > 0$. Define the high-temperature regime of the scalar-valued model by

$$\mathcal{R} = \{\beta \mid F(\beta) = 0\},$$

and define $\beta_c = \sup \mathcal{R}$. Proposition 1 states that the high-temperature regime is the interval $\mathcal{R} = (0, \beta_c)$, the same set of parameters given in Theorem 1 for which detection is impossible.

Proposition 1. If $p \geq 2$, then $\mathcal{R} = (0, \beta_c)$. Also, for $\beta > 0$, $\beta \in \mathcal{R}$ if and only if $\sup_{v\in(0,v_*]} \Gamma_\beta(v) \leq 0$.

The following two results describe the behavior of the scalar-valued model in the high-temperature regime. Theorem 5 states that in the high-temperature regime the overlap of two spin configurations σ^1, σ^2 sampled according to the Gibbs measure concentrates at 0, and Proposition 2 gives control of the fluctuation of the free energy $F_N(\beta)$ in the high temperature regime. The two results are used together to prove Theorem 1.

Theorem 5. *For $p \geq 2$, $m \in \mathbb{N}$ and $0 < \beta < \beta_c$, there exists a constant K depending only on p, m, β such that*

$$\mathbb{E}\left\langle |R(\sigma^1, \sigma^2)|^{2m} \right\rangle_s \leq \frac{K}{N^m}$$

for all $s \in [0, \beta]$ and all $N \geq 1$.

Proposition 2. For $p \geq 2$ and $0 < \beta < \beta_c$ there exists a constant K depending only on p, β such that

$$P\left(|F_N(\beta)| \geq \ell\right) \leq \frac{K}{\ell^2 N^{p/2+1}}$$

for all $\ell > 0, N \geq 1$.

2.3 The Vector-Valued p-Spin Model

Recall the probability spaces $(\Lambda_1, \mu_1), \ldots, (\Lambda_k, \mu_k)$ from which the entries of vectors $u(1), \ldots, u(k)$, respectively, are drawn. Define

$$(\bar{\mu}, \bar{\Lambda}) = \left(\prod_{r=1}^{k} \mu_r, \prod_{r=1}^{k} \Lambda_r \right).$$

Recall also SNR vector of the spiked tensor with multiple spikes, $\bar{\beta} = (\beta_1, \ldots, \beta_k)$. A spin configuration $\bar{\sigma} \in \bar{\Lambda}^N$ is a matrix whose rows are scalar-valued spins $\sigma(1) \in \Lambda_1^N, \ldots, \sigma(k) \in \Lambda_k^N$.

For each $1 \leq r \leq k$, the pure, scalar-valued p-spin model is

$$X_N(\sigma(r)) = \frac{1}{N^{(p-1)/2}} \langle Y, \sigma(r)^{\otimes p} \rangle.$$

The normalized vector-valued Hamiltonian is

$$H_{N,\bar{\beta}}(\bar{\sigma}) = \frac{1}{N^{(p-1)/2}} \sum_{r=1}^{k} \beta_r X_N(\sigma(r)) - \sum_{1 \leq r, r' \leq k} \frac{\beta_r \beta_{r'}}{2} N R(\sigma(r), \sigma(r'))^p.$$

The Gibbs measure corresponding to the vector-valued Hamiltonian is

$$G_{N,\bar{\beta}}(\sigma) = \frac{e^{H_{N,\bar{\beta}}(\bar{\sigma})} \bar{\mu}^{\otimes N}(d\bar{\sigma})}{Z_{N,\bar{\beta}}}$$

with partition function

$$Z_{N,\bar{\beta}} = \int e^{H_{N,\bar{\beta}}(\bar{\sigma})} \bar{\mu}^{\otimes N}(d\bar{\sigma}).$$

The associated free energy is

$$F_N(\bar{\beta}) = \frac{1}{N} \log Z_{N,\bar{\beta}}.$$

Define $F(\bar{\beta}) = \lim_{N\to\infty} F_N(\bar{\beta})$. When p is even, Panchenko [26] shows that this limit exists for the pure p-spin model with vector spins, which can be extended to the present case using an argument similar to the proof of Proposition 4. When $k = 1$, the limit exists for odd p, but it is an open question whether it exists for odd p and $k > 1$.

As in the scalar-valued case, an application of Jensen's inequality to $\mathbb{E}F_N(\bar{\beta})$ shows that $F(\bar{\beta}) \leq 0$. Define the high-temperature regime of the vector-valued model by

$$\bar{\mathcal{R}} = \{\bar{\beta} \mid F(\bar{\beta}) = 0\}.$$

For $1 \leq r \leq k$, let $\beta_{r,c}$ be the critical value separating the high-temperature and low-temperature regimes of the marginal scalar-valued model $H_{N,\beta(r)}(\sigma(r))$. Theorem 6 states that the high-temperature regime $\bar{\mathcal{R}}$ is the product of the high-temperature regimes of these marginal systems. In this region, the effect of the cross-overlap terms $R(\sigma(r), \sigma(r'))^p$ is negligible and the k marginal systems $H_{N,\beta_r}(\sigma(r))$ act like independent systems.

Theorem 6. *For $p \geq 3$, $\bar{\mathcal{R}} = (0, \beta_{1,c}) \times \cdots \times (0, \beta_{k,c})$.*

To prove Theorem 6, we must show that the cross-overlap terms $R(\sigma(r), \sigma(r'))^p$ with $r \neq r'$ concentrate at zero. This result is given in Section 4.6 which presents the proof of Theorem 6. An earlier version of the proof of Theorem 3 required versions of Theorem 5 and Proposition 2 to control the overlaps and free energy fluctuation, stated as Theorem 7 and Proposition 3 below. These results are unnecessary for the proof method presented here, but are still included as interesting results in the study of spin glasses.

Theorem 7. *Assume that $p \geq 2$ is even, $m \in \mathbb{N}$ and $\bar{\beta} = (\beta_1, \ldots, \beta_k)$ satisfies $0 < \beta_r < \beta_{r,c}$ for all $1 \leq r \leq k$. Then there exists a constant $K > 0$ depending only k, p, m, and $\bar{\beta}$ such that for any $1 \leq r \leq k$ and $s \in [0, 1]$*

$$\mathbb{E} \left\langle |R(\sigma^1(r), \sigma^2(r)|^{2m} \right\rangle_{s\bar{\beta}} \leq \frac{K}{N^m}$$

for all $N \geq 1$.

Proposition 3. For $p \geq 2$ and $\bar{\beta}$ as in Theorem 7, there exists a constant $K > 0$ depending only on $k, p, \bar{\beta}$ such that for any $\ell > 0$

$$P\left(|F_N(\bar{\beta})| > \ell\right) \leq \frac{K}{\ell^2 N^{p/2+1}}$$

for all $N \geq 1$.

2.4 Total Variation Distance

Chapter 1 gave an interpretation of total variation distance in terms of the sum of Type I and Type II errors of a statistical hypothesis test. The present section relates total variation distance to the likelihood ratio test and the free energy of the scalar- and vector-valued spin glasses in Lemmas 1 and 2, respectively.

Lemma 1 gives a formula in terms of the ratio of the densities of T and W or T_k and W. The result is [9, Lemma 2], but the proof is reproduced here for completeness. In Lemma 2, the connections to spin-glasses becomes clear. The ratio of the densities in Lemma 1 is $NF_N(\beta)$ in the case of a single spike and $NF_N(\bar{\beta})$ in the case of multiple spikes.

Lemma 1. Suppose U, V are two N-dimensional random vectors with densities f_U, f_V, respectively, with $f_U \neq 0$ and $f_V \neq 0$ almost everywhere. Then

$$d_{TV}(U, V) = \int_0^1 P\left(\frac{f_U(V)}{f_V(V)} < x\right) dx = \int_0^1 P\left(\frac{f_U(U)}{f_V(U)} > \frac{1}{x}\right) dx.$$

Proof. First,

$$d_{TV}(U, V) = \frac{1}{2}\int_{\mathbb{R}^n} |f_V(x) - f_U(x)| dx = \int_{f_U \leq f_V} (f_V(x) - f_U(x)) dx.$$

To see the second equality, simply note that

$$\frac{1}{2}\int_{\mathbb{R}^n} |f_V(x) - f_U(x)| dx = \frac{1}{2}\left(\int_{f_U \leq f_V} (f_V(x) - f_U(x)) dx + \int_{f_U > f_V} (f_U(x) - f_V(x)) dx\right),$$

and write

$$\int_{f_U > f_V} (f_U(x) - f_V(x))dx = \left(1 - \int_{f_U \le f_V} f_U(x)dx\right) - \left(1 - \int_{f_U \le f_V} f_V(x)dx\right)$$
$$= \int_{f_U \le f_V} (f_V(x) - f_U(x))dx.$$

To see the first equality, define $A = \{x \mid f_U(x) \le f_V(x)\}$. Then, for any measurable set B,

$$\begin{aligned}
|P(V \in B) - P(U \in B)| &= \left|\int_B (f_V(x) - f_U(x))dx\right| \\
&= \left|\int_{B \cap A} (f_V(x) - f_U(x))dx + \int_{B \cap A^c} (f_V(x) - f_U(x))dx\right| \\
&\le \max\left\{\left|\int_{B \cap A} (f_V(x) - f_U(x))dx\right|, \left|\int_{B \cap A^c} (f_V(x) - f_U(x))dx\right|\right\}.
\end{aligned}$$

This inequality holds because, by the definition of the set A, the first integral is positive and the second is negative. Continuing,

$$\begin{aligned}
|P(V \in B) - P(U \in B)| &\le \max\left\{\left|\int_A (f_V(x) - f_U(x))dx\right|, \left|\int_{A^c} (f_V(x) - f_U(x))dx\right|\right\} \\
&= \max\{|P(V \in A) - P(U \in A)|, |P(U \in A^c) - P(V \in A^c)|\} \\
&= |P(V \in A) - P(U \in A)|.
\end{aligned}$$

Thus $d_{TV}(U, V) = |P(V \in A) - P(U \in A)|$. Finally,

$$\begin{aligned}
|P(V \in A) - P(U \in A)| &= \int_A (f_V(x) - f_U(x))dx \\
&= \frac{1}{2}\left(\int_A (f_V(x) - f_U(x))dx + 1 - \int_A f_U(x)dx - 1 + \int_A f_V(x)dx\right) \\
&= \frac{1}{2}\left(\int_A (f_V(x) - f_U(x))dx + \int_{A^c} (f_U(x)dx - f_V(x))dx\right) \\
&= \frac{1}{2}\left(\int_{\mathbb{R}^N} |f_V(x) - f_U(x)|dx\right).
\end{aligned}$$

To finish the proof,

$$\begin{aligned}\int_0^1 P\left(\frac{f_U(V)}{f_V(V)} < x\right) dx &= \int_0^1 \int_{\mathbb{R}^N} 1_{\frac{f_U(r)}{f_V(r)} < x}(r) f_V(r) dr dx \\ &= \int_{\mathbb{R}^N} \int_0^1 1_{\frac{f_U(r)}{f_V(r)} < x}(r) dx f_V(r) dr \\ &= \int_{\mathbb{R}^N} 1_{\frac{f_U(r)}{f_V(r)} \leq 1}(r) \left(1 - \frac{f_U(r)}{f_V(r)}\right) f_V(r) dr \\ &= \int_{f_U \leq f_V} (f_V(r) - f_U(r)) dr.\end{aligned}$$

This shows the first equality of the lemma. For the second equality, simply switch the roles of U, V.

□

Lemma 2. For any $\beta \in (0, \infty)$ and $\bar{\beta} \in (0, \infty)^k$ the total variation distances can be written as

$$d_{TV}(W, T) = \int_0^1 P\left(F_N(\beta) < \frac{\log x}{N}\right) dx$$

and

$$d_{TV}(W, T_k) = \int_0^1 P\left(F_N(\bar{\beta}) < \frac{\log x}{N}\right) dx.$$

Proof. The Gaussian random tensor W has density $f_W(w) = \frac{1}{C}\exp(-\frac{1}{2}\langle w, w\rangle)$. Here C is a normalizing constant whose specific value is not important. For any Borel measurable set A,

$$\begin{aligned}P(T_k \in A) &= P\left(W + \frac{1}{N^{(p-1)/2}} \sum_{r=1}^{k} \beta_r u(r)^{\otimes p} \in A\right) \\ &= \mathbb{E}_u P_W\left(W \in A - \frac{1}{N^{(p-1)/2}} \sum_{r=1}^{k} \beta_r u(r)^{\otimes p}\right) \\ &= \mathbb{E}_u \int_{A - \frac{1}{N^{(p-1)/2}} \sum_{r=1}^{k} \beta_r u(r)^{\otimes p}} f_W(w) dw,\end{aligned}$$

where $\mathbb{E}_u$ denotes expectation with respect to the randomness in $u(1), \ldots, u(k)$ only, and P_W denotes probability with respect to $f_W(w)$ only. Performing the change of variables $w \mapsto w - \frac{1}{N^{(p-1)/2}} \sum_{r=1}^k \beta_r u(r)^{\otimes p}$ and then using Fubini's theorem to change the order of integration gives

$$\begin{aligned} P(T_k \in A) &= \mathbb{E}_u \int_A f_W \left(w - \frac{1}{N^{(p-1)/2}} \sum_{r=1}^k \beta_r u(r)^{\otimes p} \right) dw \\ &= \int_A \mathbb{E}_u f_W \left(w - \frac{1}{N^{(p-1)/2}} \sum_{r=1}^k \beta_r u(r)^{\otimes p} \right) dw. \end{aligned}$$

Therefore

$$\begin{aligned} f_{T_k}(w) &= \mathbb{E}_u f_W \left(w - \frac{1}{N^{(p-1)/2}} \sum_{r=1}^k \beta_r u(r)^{\otimes p} \right) \\ &= \frac{1}{C} \mathbb{E}_u \exp \left(-\frac{1}{2} \left\langle w - \frac{1}{N^{(p-1)/2}} \sum_{r=1}^k \beta_r u(r)^{\otimes p}, w - \frac{1}{N^{(p-1)/2}} \sum_{r=1}^k \beta_r u(r)^{\otimes p} \right\rangle \right) \\ &= f_W(w) \mathbb{E}_u \exp \left(\frac{1}{N^{(p-1)/2}} \sum_{r=1}^k \beta_r \langle w, u(r)^{\otimes p} \rangle - \frac{1}{2N^{p-1}} \sum_{1 \le r, r' \le k} \beta_r \beta_{r'} \langle u(r)^{\otimes p}, u(r')^{\otimes p} \rangle \right) \\ &= f_W(w) \int e^{H_{N,\bar{\beta}}(\sigma)} \mu^{\otimes N}(d\sigma). \end{aligned}$$

Since $f_W(w) \neq 0$ almost everywhere, dividing by $f_W(w)$ and taking the log gives $\log(f_{T_k}(w)/f_W(w)) = N F_N(\bar{\beta})$. Plugging this in to Lemma 1 gives

$$d_{TV}(W, T_k) = \int_0^1 P\left(F_N(\bar{\beta}) < \frac{\log x}{N} \right) dx.$$

The expression for total variation distance in the single-spike case comes from using $k = 1$ in the preceding argument. □

2.5 Structure of the Proofs of Theorems 1, 2, 3, and 4

The remainder of the thesis is devoted to proving Theorems 1, 2, 3, and 4. Chapter 3 presents the proofs of Theorems 1, 2, 3, and 4 assuming the results of the present chapter. It is clear from the representation of the total variation distance given in Lemma 2 that to solve the detection problem it is necessary to study the behavior of the free energies $F_N(\beta)$ and $F_N(\bar{\beta})$. By Proposition 1, the high-temperature region of the scalar-valued spin glass is the interval $(0, \beta_c)$. Proposition 2 bounds the fluctuations of the free energy $F_N(\beta)$ for β in this interval. This result, combined with Lemma 2, is used to prove Theorem 1. To prove Theorem 3, the free energy of the vector-valued spin glass is related to the free energies of the marginal scalar-valued systems, and then the proof proceeds in a manner identical to that of Theorem 1. Theorem 2 follows almost immediately from Proposition 1.

Chapter 4 builds the theory and results needed to prove the structure of the high-temperature regimes of the scalar- and vector-valued models as well as the concentration of overlaps used to prove Propositions 2 and 3. The key result of Chapter 4 is a cavity method to control the moments of the spin overlaps. For many models, including the SK model, the cavity method only works in a small subset of the high-temperature regime; however, the cavity method presented here holds in the entirety of the high-temperature regime for both the scalar- and vector-valued p-spin models.

Chapter 3

Proofs of Detection and Recovery Results

This chapter presents the proofs of Theorems 1, 2, 3, and 4: the results on the detection problem for a single spike, a method to compute the critical SNR, the result for the detection of multiple spikes, and recovery by the minimum mean square error, respectively. These proofs rely on the results of Sections 2.2 and 2.3, which control the fluctuations of the free energies of the scalar- and vector-valued p-spin models. The present chapter assumes that these results hold.

3.1 Detection with a Single Spike: Proofs of Theorems 1 and 2

Theorem 1 states that detection is not possible when $0 \leq \beta < \beta_c$, and detection is possible when $\beta > \beta_c$. The proof that detection is impossible for $\beta \in [0, \beta_c)$ relies on the fact that the high-temperature regime of the scalar-valued p-spin model is the interval $(0, \beta_c)$ and the behavior of the free energy $F_N(\beta)$ for β in this set. The proof that detection is possible when $\beta > \beta_c$ depends on the behavior of the free energy and a simple application of the dominated convergence theorem.

Proof. (Theorem 1) Recall that the high-temperature regime of the spin glass $H_N(\sigma)$ is defined as $\mathcal{R} = \{\beta > 0 \mid F(\beta) = 0\}$, and $\beta_c = \sup \mathcal{R}$. By Proposition 1, $\mathcal{R} = [0, \beta_c)$.

First suppose that $\beta \in \mathcal{R}$, so that $F(\beta) = 0$. From Lemma 2, the total variation distance between W and T is

$$d_{TV}(W,T) = \int_0^1 P\left(F_N(\beta) < \frac{\log x}{N}\right) dx.$$

Performing the change of variables $y = -\log x$ gives

$$d_{TV}(W,T) = \int_0^\infty P\left(F_N(\beta) < \frac{-y}{N}\right) e^{-y} dy \leq \int_0^\infty P\left(F_N(\beta) < \frac{-y}{N}\right) dy.$$

The inequality holds since the integrand is positive and $e^{-y} \leq 1$ for all $0 \leq y < \infty$.

For any $\varepsilon > 0$, splitting the right-most integral into two pieces at ε gives

$$\begin{aligned} d_{TV}(W,T) &\leq \int_0^\varepsilon P\left(F_N(\beta) < \frac{-y}{N}\right) dy + \int_\varepsilon^\infty P\left(F_N(\beta) < \frac{-y}{N}\right) dy \\ &\leq \int_0^\varepsilon dy + \int_\varepsilon^\infty P\left(|F_N(\beta)| > \frac{y}{N}\right) dy \\ &= \varepsilon + \int_\varepsilon^\infty P\left(|F_N(\beta)| > \frac{y}{N}\right) e^{-y} dy. \end{aligned}$$

Since $0 < \beta < \beta_c$, by Proposition 2, there exists a constant $K > 0$ depending only on p, β such that for all $\ell > 0$ and all $N \geq 1$,

$$P\left(|F_N(\beta)| > \ell\right) \leq \frac{K}{\ell^2 N^{\frac{p}{2}+1}}.$$

Therefore,

$$d_{TV}(W,T) \leq \varepsilon + \frac{K}{N^{\frac{p}{2}-1}} \int_\varepsilon^\infty \frac{1}{y^2} dy = \varepsilon + \frac{K}{\varepsilon N^{\frac{p}{2}-1}}.$$

Choose $\varepsilon = N^{-\frac{p-2}{4}}$ so that

$$d_{TV}(W,T) \leq \frac{1+K}{N^{\frac{p-2}{4}}}.$$

Taking $N \to \infty$ shows that $\lim_{N\to\infty} d_{TV}(W,T) = 0$, so W, T are indistinguishable and detection is impossible.

Next consider $\beta > \beta_c$. Since $\beta \notin \mathcal{R}$, $F_N(\beta) < 0$. Recall that $F_N(\beta) \to F(\beta)$ almost surely, therefore $F_N(\beta) - \frac{\log x}{N} \to F(\beta)$ almost surely for any $x > 0$. Almost sure convergence implies convergence in distribution, therefore

$$\lim_{N\to\infty} P\left(F_N(\beta) - \frac{\log x}{N} < 0\right) = P(F(\beta) < 0) = 1.$$

The dominated convergence theorem allows the interchange of the limit and the integral so

$$\begin{aligned}
\lim_{N\to\infty} d_{TV}(W,T) &= \lim_{N\to\infty} \int_0^1 P\left(F_N(\beta) - \frac{\log x}{N} < 0\right) dx \\
&= \int_0^1 \lim_{N\to\infty} P\left(F_N(\beta) - \frac{\log x}{N} < 0\right) dx \\
&= 1.
\end{aligned}$$

Thus W, T are distinguishable and detection is possible. □

Theorem 2 follows directly from Proposition 1, which gives an alternate characterization of the high-temperature regime in terms of the auxiliary function $\Gamma_b(v)$.

Proof. (Theorem 2) Suppose, for the sake of contradiction, that $\sup_{v\in[0,v_*]} \Gamma_{\beta_c}(v) < 0$. Since $\Gamma_\beta(v)$ is continuous in β, it is therefore possible to find $\beta > \beta_c$ such that $\sup_{v\in[0,v_*]} \Gamma_\beta(v) \leq 0$. By Proposition 1, this means $\beta \in \mathcal{R}$, which contradicts the maximality of β_c. □

3.2 Detection with Multiple Spikes: Proof of Theorem 3

Theorem 3 states that for a random tensor with multiple spikes, detection is possible when the SNR for at least one of the spikes exceeds the critical threshold for the corresponding single-spike tensor. The proof relies on Lemma 3, a version of the triangle inequality for total variation distance that bounds $d_{TV}(W, T_k)$ by the sum of the total variation distances between W and the single-spike tensors. Then it is possible to use Proposition 2 and the arguments in the proof of Theorem 1 to control each individual single-spike total variation distance.

Lemma 3. Assume that Y_1, Y_2 are independent random p-tensors which are also independent of W. Then $d_{TV}(W, W+Y_1+Y_2) \leq d_{TV}(W, W+Y_1) + d_{TV}(W, W+Y_2)$.

Proof. From the triangle inequality,

$$\begin{aligned}
d_{TV}(W, W+Y_1+Y_2) &= \sup_A |P(W\in A) - P(W+Y_1+Y_2\in A)| \\
&= \sup_A |P(W\in A) - P(W+Y_1\in A) + P(W+Y_1\in A) - P(W+Y_1+Y_2\in A)| \\
&\le \sup_A |P(W\in A) - P(W+Y_1\in A)| \\
&+ \sup_A |P(W+Y_1\in A) - P(W+Y_1+Y_2\in A)| \\
&= d_{TV}(W, W+Y_1) + d_{TV}(W+Y_1, W+Y_1+Y_2).
\end{aligned}$$

Since Y_1 and Y_2 are independent,

$$\begin{aligned}
d_{TV}(W+Y_1, W+Y_1+Y_2) &= \sup_A |\mathbb{E}_{Y_1}(P(W\in A-Y_1) - P(W+Y_2\in A-Y_1))| \\
&\le \mathbb{E}_{Y_1}\sup_A |P(W\in A-Y_1) - P(W+Y_2\in A-Y_1)| \\
&\le \mathbb{E}_{Y_1}\sup_A |P(W\in A) - P(W+Y_2\in A)| \\
&= \sup_A |P(W\in A) - P(W+Y_2\in A)| \\
&= d_{TV}(W, W+Y_2).
\end{aligned}$$

□

A previous version of the proof of Theorem 3 controlled the fluctuations of free energy $F_N(\bar{\beta})$ in the high-temperature regime using Proposition 3 in much the same way that Proposition 2 is used in the proof of Theorem 1; however, Lemma 3 renders this argument unnecessary. Theorem 7 and Proposition 3 are still interesting in the study of spin glasses, and are still included here even though they are not necessary to solve the detection problem.

Proof. (Theorem 3) First assume that $\bar{\beta} \in (0, \beta_{1,c}) \times \cdots \times (0, \beta_{k,c})$. For $1 \le r \le k$, define single spike models

$$T_{k,r} = W + \frac{\beta_r}{N^{(p-1)/2}} u(r)^{\otimes p}.$$

By Lemma 3,

$$d_{TV}(W,T_k) \leq \sum_{r=1}^{k} d_{TV}(W,T_{k,r}).$$

Combining the claim of Proposition 2 and the arguments in the proof of Theorem 1, for each $1 \leq r \leq k$ there exist constants $K_r > 0$ such that $d_{TV}(W,T_{k,r}) \leq K_r/N^{\frac{p-2}{4}}$. Therefore

$$\lim_{N\to\infty} d_{TV}(W,T_k) \leq \lim_{N\to\infty} \sum_{r=1}^{k} \frac{K_r}{N^{\frac{p-2}{4}}} = 0,$$

so $W, T_{k,r}$ are indistinguishable and detection is impossible.

Next assume that $\bar{\beta} \notin (0,\beta_{1,c}] \times \cdots \times (0,\beta_{k,c}]$. By Theorem 6, the high-temperature regime is $\mathcal{R} = (0,\beta_{1,c}) \times \cdots \times (0,\beta_{k,c})$, so $F(\bar{\beta}) < 0$. Using the representation of total variation distance in Lemma 2,

$$\begin{aligned}
\lim_{N\to\infty} d_{TV}(W,T_k) &\geq \liminf_{N\to\infty} \int_0^1 P\left(F_N(\bar{\beta}) < \frac{\log x}{N}\right) dx \\
&\geq \int_0^1 \liminf_{N\to\infty} P\left(F_N(\bar{\beta}) < \frac{\log x}{N}\right) dx \\
&= \int_0^1 P(F(\bar{\beta}) < 0) dx \\
&= 1.
\end{aligned}$$

The second line uses Fatou's lemma. Thus W, T_k are distinguishable and detection is possible. □

3.3 Spike Recovery by MMSE: Proof of Theorem 4

In this section we prove Theorem 4, which states that recovery of the spike by the minimum mean square error estimator has the same phase transition as the spike detection problem. The proof requires an auxiliary Hamiltonian and an auxiliary mean square error. Lemma 4 states several properties of the auxiliary mean square error, and these properties are used to prove Theorem 4

Fix an SNR vector $\bar\beta = (\beta_1, \dots, \beta_k)$. For $t \geq 0$ define a new random tensor $T_k(t)$ by

$$T_k(t) = W + \sqrt{\frac{t}{N^{p-1}}} \sum_{r=1}^{k} \beta_r u(r)^{\otimes p},$$

and define an auxiliary Hamiltonian

$$AH_{N,t}(\bar\sigma) = \frac{\sqrt{t}}{N^{(p-1)/2}} \sum_{r=1}^{k} \beta_r \langle T_k(t), \sigma(r)^{\otimes p}\rangle - \frac{t}{2} \sum_{1\leq r,r'\leq k} \beta_r\beta_{r'} N R(\sigma(r), \sigma(r'))^p.$$

When $t = 1$, the auxiliary Hamiltonian is the Hamiltonian $H_{N,\bar\beta}(\bar\sigma)$ with disorder Y replaced by the random tensor T_k. Expanding $T_k(t)$, the auxiliary Hamiltonian can also be written as

$$AH_{N,t}(\bar\sigma) = \frac{\sqrt{t}}{N^{(p-1)/2}} \sum_{r=1}^{k} \beta_r \langle Y, \sigma(r)^{\otimes p}\rangle - \frac{t}{2} \sum_{1\leq r,r'\leq k} \beta_r\beta_{r'} N R(\sigma(r), \sigma(r'))^p + t \sum_{1\leq r,r'\leq k} \beta_r\beta_{r'} N R(\sigma(r), u(r'))^p.$$

Let $AF_{N,t}(\bar\beta)$ denote the free energy of the auxiliary Hamiltonian and $AG_{N,t}$ the associated Gibbs measure. Let $\langle\cdot\rangle_t^A$ denote the corresponding Gibbs average. Note that $AF_{N,1}(\bar\beta) = F_N(\bar\beta)$.

A key component of the proof of Theorem 4 is the fact that the auxiliary Gibbs measure $AG_{N,t}$ is equal to the conditional distribution of $(u(1), \dots, u(k))$ given $T_k(t)$. Let $\mathbb{E}_u$ denote the expected value in the randomness of $u(1), \dots, u(k)$ only. Recall the density of the symmetric Gaussian tensor $f_W(w) = \frac{1}{C}\exp(-\frac{1}{2}\langle w, w\rangle)$. Since $u(1), \dots, u(k)$ are independent of W, the conditional probability is

$$\begin{aligned}
&P((u(1), \dots, u(k)) \in B \mid T_k(t) = w) \\
&= \frac{\mathbb{E}_u\left[f_W\left(w - \sqrt{\frac{t}{N^{p-1}}}\sum_{r=1}^k \beta_r u(r)^{\otimes p}\right)\cdot 1_{(u(1),\dots,u(k))\in B}\right]}{\mathbb{E}_u\left[f_W\left(w - \sqrt{\frac{t}{N^{p-1}}}\sum_{r=1}^k \beta_r u(r)^{\otimes p}\right)\right]} \\
&= \frac{\int_B \exp\left(\sqrt{\frac{t}{N^{p-1}}}\sum_{r=1}^k \beta_r\langle w, \sigma(r)^{\otimes p}\rangle - \frac{t}{2}\sum_{r,r'}\beta_r\beta_{r'} N R(\sigma(r),\sigma(r'))^p\right)\mu^{\otimes N}(d\bar\sigma)}{\int \exp\left(\sqrt{\frac{t}{N^{p-1}}}\sum_{r=1}^k \beta_r\langle w, \sigma(r)^{\otimes p}\rangle - \frac{t}{2}\sum_{r,r'}\beta_r\beta_{r'} N R(\sigma(r),\sigma(r'))^p\right)\mu^{\otimes N}(d\bar\sigma)}.
\end{aligned}$$

Taking $w = T_k(t)$ gives $AG_{N,t}(B)$.

Define an auxiliary minimum mean square error by

$$\mathrm{MMSE}_N^A(\bar\beta,t)=\min_{\hat\theta}\frac{1}{N^p}\sum_{1\le i_1,\dots,i_p\le N}\mathbb{E}\left(\sum_{r=1}^k\beta_r u_{i_1}(r)\cdots u_{i_p}(r)-\hat\theta_{i_1,\dots,i_p}\right)^2$$

where the minimum is taken over all random variables $\hat\theta$ generated by the sigma field $\sigma(T_k(t))$. When $t=1$, we have $\mathrm{MMSE}_N^A(\bar\beta,1)=\mathrm{MMSE}_N(\bar\beta)$. Lemma 4 below gives a representation of $\mathrm{MMSE}_N^A(\bar\beta,t)$ in terms of the auxiliary free energy. Setting $t=1$ and using this representation will help prove Theorem 4.

Lemma 4. The following hold.

(i) $\mathbb{E}AF_N(\bar\beta,t)$ is a nondecreasing, nonnegative, and convex function of t.

(ii) $\mathrm{MMSE}_N^A(\bar\beta,t)=\sum_{1\le r,r'\le k}\beta_r\beta_{r'}\mathbb{E}R(u(r),u(r'))^p-2\frac{d}{dt}\mathbb{E}AF_N(\bar\beta,t)$.

Proof. By Gaussian integration by parts, the derivative of $\mathbb{E}AF_N(\bar\beta,t)$ with respect to t is

$$\frac{d}{dt}\mathbb{E}AF_N(\bar\beta,t)=\sum_{1\le r,r'\le k}\beta_r\beta_{r'}\left(-\frac{1}{2}\mathbb{E}\langle R(\sigma^1(r),\sigma^2(r'))\rangle_t^A+\mathbb{E}\langle R(\sigma(r),u(r'))^p\rangle_t^A\right).$$

Recall that $AG_{N,t}(\cdot)=P((u(1),\dots,u(k))\in\cdot\mid T_k(t))$. Therefore

$$\begin{aligned}\mathbb{E}\langle R(\sigma(r),u(r'))^p\rangle_t^A&=\mathbb{E}\left[\mathbb{E}\left[\langle R(\sigma(r),u(r'))^p\rangle_t^A\mid T_k(t)\right]\right]\\&=\mathbb{E}\langle R(\sigma^1(r),\sigma^2(r'))^p\rangle_t^A.\end{aligned}$$

Plugging in to the derivative,

$$\frac{d}{dt}\mathbb{E}AF_N(\bar\beta,t)=\frac{1}{2}\sum_{1\le r,r'\le k}\beta_r\beta_{r'}\mathbb{E}\langle R(\sigma^1(r),\sigma^2(r'))^p\rangle_t^A.$$

Note that the minimizer of $\mathrm{MMSE}_N^A(\bar\beta,t)$ occurs at

$$\hat\theta^A_{i_1,\dots,i_p}:=\sum_{r=1}^k\beta_r\mathbb{E}[u_{i_1}(r)\cdots u_{i_p}(r)|T_k(t)]=\sum_{r=1}^k\beta_r\langle\sigma_{i_1}\cdots\sigma_{i_p}\rangle_t^A.$$

Therefore

$$\begin{aligned}
\mathrm{MMSE}_N^A(\bar\beta,t) &= \frac{1}{N^p}\sum_{1\le i_1,\dots,i_p\le N}\mathbb{E}\left(\sum_{r=1}^k \beta_r u_{i_1}(r)\cdots u_{i_p}(r) - \sum_{r=1}^k \beta_r\langle\sigma_{i_1}\cdots\sigma_{i_p}\rangle_t^A\right)^2\\
&= \sum_{r,r'=1}^k \beta_r\beta_{r'}\left(\mathbb{E}R(u(r),u(r'))^p - 2\mathbb{E}\langle R(u^1(r),\sigma^2(r'))^p\rangle_t^A + \mathbb{E}\langle R(\sigma^1(r),\sigma^2(r'))^p\rangle_t^A\right)\\
&= \sum_{r,r'=1}^k \beta_r\beta_{r'}\left(\mathbb{E}R(u(r),u(r'))^p - 2\mathbb{E}\langle R(\sigma^1(r),\sigma^2(r'))^p\rangle_t^A\right)\\
&= \sum_{r,r'=1}^k \beta_r\beta_{r'}\mathbb{E}R(u(r),u(r'))^p - 2\frac{d}{dt}\mathbb{E}AF_N(\bar\beta,t).
\end{aligned}$$

This completes the proof of part (ii).

To prove part (i), set $\hat\theta_{i_1,\dots,i_p}=0$ to get the upper bound for $\mathrm{MMSE}_N^A(\bar\beta,t)$

$$\mathrm{MMSE}_N^A(\bar\beta,t) \le \sum_{r,r'=1}^k \beta_r\beta_{r'}\mathbb{E}R(u(r),u(r'))^p.$$

Thus we must have

$$\sum_{r,r'=1}^k \beta_r\beta_{r'}\mathbb{E}R(u(r),u(r'))^p - 2\frac{d}{dt}\mathbb{E}AF_N(\bar\beta,t) \le \sum_{r,r'=1}^k \beta_r\beta_{r'}\mathbb{E}R(u(r),u(r'))^p,$$

so we can conclude that $\frac{d}{dt}\mathbb{E}AF_N(\bar\beta,t)\ge 0$. Therefore $\mathbb{E}AF_N(\bar\beta,t)$ is non-decreasing in t. It is straightforward to compute that $AF_N(\bar\beta,0)=0$; therefore, $\mathbb{E}AF_N(\bar\beta,t)$ is non negative for $t\ge 0$.

Finally, if $\mathrm{MMSE}_N^A(\bar\beta,t)$ is non-increasing in t, then $\frac{d}{dt}\mathbb{E}AF_N(\bar\beta,t)$ must be non-decreasing in t and therefore $\mathbb{E}AF_N(\bar\beta,t)$ must be convex in t. To show that $\mathrm{MMSE}_N^A(\bar\beta,t)$ is non-increasing in t, note that

$$\frac{1}{\sqrt t}T_k(t) = \frac{1}{N^{(p-1)/2}}\sum_{r=1}^k \beta_r u(r)^{\otimes p} + \frac{1}{\sqrt t}W.$$

Let W' be an independent copy of W. For $t' > t$,

$$\frac{1}{\sqrt{t'}}T_k(t') + \sqrt{\frac{1}{t} - \frac{1}{t'}}W' = \frac{1}{N^{(p-1)/2}}\sum_{r=1}^{k}\beta_r u(r)^{\otimes p} + \frac{1}{\sqrt{t'}}W + \sqrt{\frac{1}{t} - \frac{1}{t'}}W'.$$

This is equal in distribution to $\frac{1}{\sqrt{t}}T_k(t)$ for any $0 \le t < t'$. Since W' and $T_k(t')$ are independent,

$$\mathbb{E}[u_{i_1}\cdots u_{i_p} \mid T_k(t')] = \mathbb{E}[u_{i_1}\cdots u_{i_p} \mid T_k(t'), W'] = \mathbb{E}[u_{i_1}\cdots u_{i_p} \mid T_k(t'), T_k(t)].$$

Therefore

$$\begin{aligned}
\mathrm{MMSE}_N^A(\bar{\beta}, t') &= \frac{1}{N^p}\sum_{1\le i_1,\ldots,i_p\le N}\mathbb{E}\left(\sum_{r=1}^{k}\beta_r\left(u_{i_1}\cdots u_{i_p} - \mathbb{E}[u_{i_1}\cdots u_{i_p} \mid T_k(t'), T_k(t)\right)\right)^2 \\
&\le \frac{1}{N^p}\sum_{1\le i_1,\ldots,i_p\le N}\mathbb{E}\left(\sum_{r=1}^{k}\beta_r\left(u_{i_1}\cdots u_{i_p} - \mathbb{E}[u_{i_1}\cdots u_{i_p} \mid T_k(t)\right)\right)^2 \\
&= \mathrm{MMSE}_N^A(\bar{\beta}, t).
\end{aligned}$$

Conditioning only on $T_k(t)$ instead of on $T_k(t)$ and $T_k(t')$, we are given less information, so the estimate is not as good and the inequality holds. This concludes the proof of part (i). □

3.3.1 Proof of Theorem 4

We now turn to the proof of Theorem 4. The key tool is Lemma 4 (ii). We show that when $\bar{\beta} \in (0, \beta_{1,c}) \times \cdots \times (0, \beta_{k,c})$, in the limit $\frac{d}{dt}\mathbb{E}AF_N(\bar{\beta}, 1)$ is equal to zero. Then an application of the strong law of large numbers shows that

$$\limsup_{N\to\infty}\mathrm{MMSE}_N(\bar{\beta}) = \limsup_{N\to\infty}\mathrm{MMSE}_N^A(\bar{\beta}, 1) = \mathrm{DMSE}\,.$$

On the other hand, when $\bar{\beta} \notin (0, \beta_{1,c}) \times \cdots \times (0, \beta_{k,c})$, we show that $\frac{d}{dt}\mathbb{E}AF_N(\bar{\beta}, 1)$ is strictly positive in the limit and consequently

$$\limsup_{N\to\infty}\mathrm{MMSE}_N(\bar{\beta}) < \mathrm{DMSE}\,.$$

Recall that Lemma 1 gives two representations of the total variation distance between random tensors U, V with densities f_U, f_V:

$$d_{TV}(U,V) = \int_0^1 P\left(\frac{f_U(V)}{f_V(V)} < x\right) dx = \int_0^1 P\left(\frac{f_U(U)}{f_V(U)} > \frac{1}{x}\right) dx.$$

The result of Lemma 2 comes from setting $U = T_k$ and $V = W$ and using the first expression above. Using the second expression gives

$$d_{TV}(T_k, W) = \int_0^1 P\left(AF_N(\bar{\beta}, 1) > -\frac{\log x}{N}\right) dx.$$

Proof. (*Theorem 4*): We begin by proving statement (i) of Theorem 4. Suppose $\bar{\beta} \in (0, \beta_{1,c}) \times \cdots \times (0, \beta_{k,c})$. By Fatou's Lemma,

$$0 \le \int_0^1 \liminf_{N\to\infty} P\left(AF_N(\bar{\beta}, 1) > -\frac{\log x}{N}\right) dx \le \liminf_{N\to\infty} \int_0^1 P\left(AF_N(\bar{\beta}, 1) > -\frac{\log x}{N}\right) dx = \liminf_{N\to\infty} d_{TV}(T_k, W).$$

Theorem 3 guarantees that

$$\lim_{N\to\infty} d_{TV}(T_k, W) = 0,$$

so we must in fact have

$$\liminf_{N\to\infty} P\left(AF_N(\bar{\beta}, 1) > -\frac{\log x}{N}\right) = 0 \tag{3.1}$$

for all $x \in [0, 1]$.

For $\varepsilon > 0$, define $B_N(\varepsilon) = \{AF_N(\bar{\beta}, 1) \le \varepsilon\}$. By equation (3.1), for all $\varepsilon > 0$ we have

$$\limsup_{N\to\infty} P(B_N(\varepsilon)) = 1.$$

Write

$$\begin{aligned}\mathbb{E}AF_N(\bar{\beta}, 1) &= \mathbb{E}AF_N(\bar{\beta}, 1)1_{B_N(\varepsilon)} + \mathbb{E}AF_N(\bar{\beta}, 1)1_{B_N(\varepsilon)^c} \\ &\le \varepsilon + \mathbb{E}AF_N(\bar{\beta}, 1)1_{B_N(\varepsilon)^c}.\end{aligned}$$

Applying Hölder's inequality yields

$$\begin{aligned}\mathbb{E}AF_N(\bar{\beta},1) &\leq \varepsilon + \left(\mathbb{E}AF_N(\bar{\beta},1)^2\right)^{1/2}\left(\mathbb{E}I^2_{B_N(\varepsilon)^c}\right)^{1/2} \\ &= \varepsilon + \left(\mathbb{E}AF_N(\bar{\beta},1)^2\right)^{1/2} P(B_N(\varepsilon))^{1/2}.\end{aligned}$$

Since vector entries $u_i(r)$ all come from bounded sets, the quantity $\left(\mathbb{E}AF_N(\bar{\beta},1)^2\right)^{1/2}$ is bounded. We may therefore take the lim sup on both sides above and then take $\varepsilon \to 0$ to see that

$$\limsup_{N\to\infty} \mathbb{E}AF_N(\bar{\beta},1) \leq 0.$$

By Lemma 4 part (i), $\mathbb{E}AF_N(\bar{\beta},t)$ is non-negative, convex, and non-decreasing in t, so for all $t \in [0,1]$,

$$\lim_{N\to\infty} \mathbb{E}AF_N(\bar{\beta},t) = 0.$$

Therefore,

$$\lim_{N\to\infty} \frac{d}{dt}\mathbb{E}AF_N(\bar{\beta},t) = 0.$$

By Lemma 4 part (ii),

$$\mathrm{MMSE}^A_N(\bar{\beta},t) = \sum_{1\leq r,r'\leq k} \beta_r\beta_{r'}\mathbb{E}R(u(r),u(r'))^p - 2\frac{d}{dt}\mathbb{E}AF_N(\bar{\beta},t).$$

Since the derivative is eqal to zero for all $t \in [0,1]$,

$$\begin{aligned}\limsup_{N\to\infty} \mathrm{MMSE}_N(\bar{\beta}) &= \limsup_{N\to\infty} \mathrm{MMSE}^A_N(\bar{\beta},1) \\ &= \limsup_{N\to\infty} \frac{1}{N^p}\sum_{1\leq r,r'\leq k} \beta_r\beta_{r'}\mathbb{E}R(u(r),u(r'))^p \\ &= \limsup_{N\to\infty} \frac{1}{N^p}\sum_{1\leq r,r'\leq k} \beta_r\beta_{r'}\mathbb{E}\sum_{1\leq i_1,\ldots i_p\leq N} u_{i_1}(r)\cdots u_{i_p}(r)u_{i_1}(r')\cdots u_{i_p}(r') \\ &= \limsup_{N\to\infty} \frac{1}{N^p}\sum_{1\leq i_1,\ldots,i_p\leq N} \mathbb{E}\left(\sum_{r=1}^k \beta_r u_{i_1}(r)\cdots u_{i_p}(r)\right)^2 \\ &= \mathrm{DMSE}(\bar{\beta}).\end{aligned}$$

The last equality is by the strong law of large numbers. This proves statement (i).

The key to prove statement (ii) is the fact that

$$\liminf_{N\to\infty} \frac{d}{dt}\mathbb{E}AF_N(\bar{\beta},1) > 0.$$

This, combined with the strong law of large numbers, will show that in the limit $\limsup_{N\to\infty} \mathrm{MMSE}_N^A(\bar{\beta},1) < \mathrm{DMSE}$.

Suppose that $\bar{\beta} \in (\beta_{1,c},\infty)\times\cdots\times(\beta_{k,c},\infty)$. Define an interpolating Hamiltonian and free energy by

$$IH_{N,s}(\bar{\sigma}) = H_{N,\bar{\beta}}(\bar{\sigma}) + s\sum_{1\le r,r'\le k}\beta_r\beta_{r'}NR(\sigma(r),\sigma(r'))^p$$

and

$$IF_N(\bar{\beta},s) = \frac{1}{N}\log\int \exp\left(IH_{N,\bar{\beta}}(\bar{\sigma})\right)\mu^{\otimes N}(d\bar{\sigma}),$$

respectively. This Hamiltonian interpolates between the original Hamiltonian $H_{N,\bar{\beta}}(\bar{\sigma})$ when $t=0$ and the auxiliary Hamiltonian $AH_{N,1}(\bar{\sigma})$ when $t=1$. Therefore $IF_N(\bar{\beta},1) = AF_N(\bar{\beta},1)$ and $IF_N(\bar{\beta},0) = F_N(\bar{\beta})$. Note that $IF_N(\bar{\beta},s)$ is convex in s, the proof of which easily follows from an application of Hölder's inequality.

The derivative $\frac{d}{dt}\mathbb{E}AF_N(\bar{\beta},t)$ was computed in the proof of Lemma 4 using Gaussian integration by parts, and the derivative $\frac{d}{ds}IF_N(\bar{\beta},s)$ is straightforward to compute. Comparing these two derivitavies,

$$\frac{d}{dt}\mathbb{E}AF_N(\bar{\beta},1) = \frac{1}{2}\frac{d}{ds}IF_N(\bar{\beta},1) \ge \frac{1}{2}\frac{d}{ds}\mathbb{E}IF_N(\bar{\beta},s) \tag{3.2}$$

for all $s\in[0,1]$. The inequality is by convexity. Since the function $\frac{d}{ds}\mathbb{E}IF_N(\bar{\beta},s)$ is continuous for $s\in[0,1]$, it achieves its maximum on this interval. Therefore, from this and equation (3.2) the Lipschitz constants of the functions $\mathbb{E}IF_N(\bar{\beta},s)$ are uniformly bounded, and therefore the functions are equicontinuous. Thus, by the Arzela-Ascoli Theorem, there exists a uniformly convergent subsequence $\mathbb{E}IF_{N_n}(\bar{\beta},s)$. Denote the limit along this subsequence by $IF(\bar{\beta},s)$. Along the subsequence we also have

$$\lim_{n\to\infty}\frac{d}{dt}\mathbb{E}AF_{N_n}(\bar{\beta},1) = 2\lim_{n\to\infty}\frac{d}{ds}\mathbb{E}IF_{N_n}(\bar{\beta},1).$$

Since the functions $\mathbb{E}IF_N(\bar{\beta}, s)$ are convex in s, the function $IF(\bar{\beta}, s)$ is also convex. Since convex functions have at most countably many points of discontinuity, there exists a point $s_0 \in (0, 1)$ such that $IF(\bar{\beta}, s_0)$ is differentiable. Also, at this point,

$$\lim_{n\to\infty} \frac{d}{ds}\mathbb{E}IF_N(\bar{\beta}, s_0) = \frac{d}{ds}IF(\bar{\beta}, s_0). \tag{3.3}$$

Recall from Lemma 4 (i) that $\mathbb{E}AF_N(\bar{\beta}, 1) \geq 0$. By Theorem 6, $\bar{\beta}$ is not in the high-temperature regime so

$$F(\bar{\beta}) = \limsup_{N\to\infty} \mathbb{E}F_N(\bar{\beta}) < 0.$$

Since $IF_N(\bar{\beta}, 1) = AF_N(\bar{\beta}, 1)$ and $IF_N(\bar{\beta}, 0) = F_N(\bar{\beta})$, by the above observations we have that

$$IF(\bar{\beta}, 0) = \lim_{n\to\infty} \mathbb{E}IF_{N_n}(\bar{\beta}, 0) = \limsup_{N\to\infty} \mathbb{E}F_N(\bar{\beta}) < 0 \leq \liminf_{N\to\infty} \mathbb{E}AF_N(\bar{\beta}, 1) = IF(\bar{\beta}, 1).$$

Thus at s_0 we in fact have $\frac{d}{ds}IF(\bar{\beta}, s_0) > 0$. Combining this with equations (3.2) and (3.3) gives

$$\liminf_{N\to\infty} \frac{d}{dt}\mathbb{E}AF_N\bar{\beta}, 1 > 0. \tag{3.4}$$

Finally, Lemma 4 part (ii) states that

$$\limsup_{N\to\infty} \mathrm{MMSE}_N(\bar{\beta}) = \limsup_{N\to\infty} \mathrm{MMSE}_N^A(\bar{\beta}, 1).$$

Applying the law of large numbers to the right-hand side gives

$$\limsup_{N\to\infty} \mathrm{MMSE}_N(\bar{\beta}) = \mathrm{DMSE} - 2\liminf_{N\to\infty} \frac{d}{dt}\mathbb{E}AF_N(\bar{\beta}, 1) < \mathrm{DMSE}\,.$$

The inequality is by equation (3.4). This proves part (ii) of Theorem 4. □

Chapter 4

Overlap Concentration: Proofs of Spin-Glass Results

Theorem 1 relies on using Proposition 2 to bound the fluctuations of the free energy $F_N(\beta)$ in the high temperature regime $\mathcal{R} = (0, \beta_c)$. The bound of Proposition 2 requires delicate control of the expected value of the Gibbs average of overlaps $R(\sigma^1, \sigma^2)$, and the present chapter builds the machinery to achieve this control.

Section 4.1 introduces the Parisi formula for the scalar-valued p-spin model, which gives a formula for the limiting free energy $F(\beta)$. Computing the Parisi formula requires minimizing a functional over a set of probability measures. In Section 4.5, it is shown that in the high-temperature regime, the optimal probability measure is a Dirac measure. Also, it is shown that the auxiliary function $\Gamma_b(v)$ is increasing in b, and these two results are combined to prove Proposition 1.

Section 4.5 presents the proofs of Theorem 5 and Proposition 2. Theorem 5 gives a bound for the moments $\mathbb{E}\langle R(\sigma^1, \sigma^2)^{2m} \rangle_\beta$. Section 4.4 presents a cavity argument to prove this bound. The cavity method for the scalar- and vector-valued p-spin models holds for the entire high-temperature regime, a stark contrast to the cavity method for the Sherrington-Kirkpatrick model which only holds in a subset of the high-temperature regime. In addition to a bound on the even moments, the proof of Theorem 5 relies on a bound of $\mathbb{E}\langle I(|R(\sigma^1, \sigma^2)| > \varepsilon) \rangle_\beta$. The bound is given in section 4.3 using the Guerra-Talagrand 1-replica symmetry breaking bound.

Finally, Section 4.6 presents the proofs of Theorem 7 and Proposition 3. These mainly follow the same steps as the proofs of Theorem 5 and Proposition 2, respectively, with the added complication that they require control of the total overlap. In addition to showing that the moments $R(\sigma^1(r), \sigma^2(r))$ concentrate at zero, which is a direct consequence of Theorem 5, it must also be shown that cross-overlap terms $R(\sigma^1(r), \sigma^2(r'))$ concentrate at zero.

4.1 Parisi Formula

Define $\mathcal{V} = \{u^2 \mid u \in \Lambda\}$. For any $v \in V$, denote by $\mathcal{M}_v$ the set of all cumulative distribution functions of probability measures on the interval $[0, v]$. For $\alpha \in \mathcal{M}_v$ and $\lambda \in \mathbb{R}$, define the Parisi functional by

$$\mathcal{P}_{\beta,v}(\alpha, \lambda) = \Phi_{\beta,v,\alpha}(0, 0, \lambda) - \lambda v - \frac{\beta^2}{2} \int_0^v \alpha(s)\xi''(s)sds,$$

where $\Phi_{\beta,v,\alpha}(s, x, \lambda) : [0, v] \times \mathbb{R} \times \mathbb{R}$ is the weak solution to the PDE

$$\partial_s \Phi_{\beta,v,\alpha} = -\frac{\beta^2 \xi''}{2}(\partial_{xx}\Phi_{\beta,v,\alpha} + \alpha(s)(\partial_x \Phi_{\beta,v,\alpha})^2)$$

with boundary condition

$$\Phi_{\beta,v,\alpha}(v, x, \lambda) = \log \int e^{xa+\lambda a^2} \mu(da).$$

The Parisi formula was first established for mixed even p-spin models with Ising spins by Talagrand [27]. Panchenko extended this to the generalized SK models with spins coming from bounded sets [28] and to the mixed even p-spin model with vector spins [26]. In the present case, the Parisi formula states that

$$\lim_{N\to\infty} \frac{1}{N} \log \int e^{\beta X_N(\sigma)} \mu^{\otimes N}(d\sigma) = \sup_{v\in\mathcal{V}} \inf_{(\alpha,\lambda)\in\mathcal{M}_v\times\mathbb{R}} \mathcal{P}_{\beta,v}(\alpha, \lambda).$$

Define

$$\mathcal{Q}_{\beta,v}(\alpha, \lambda) = \mathcal{P}_{\beta,v}(\alpha, \lambda) - \frac{\beta^2 v^p}{2}.$$

The Parisi formula of Proposition 4 is given in terms of the functionals $\mathcal{Q}_{\beta,\alpha}$ which account for the normalizing term added to the Hamiltonian $H_N(\beta)$.

Proposition 4. (Parisi Formula) For any $\beta > 0$,

$$F(\beta) = \lim_{N\to\infty} F_N(\beta) = \sup_{v\in\mathcal{V}} \inf_{\alpha,\lambda} \mathcal{Q}_{\beta,v}(\alpha,\lambda).$$

Proof. For any measurable set $A \subset \mathcal{V}$ define the free energy restricted to A by

$$F_N(\beta, A) = \frac{1}{N}\log\int_{\sigma:R(\sigma,\sigma)\in A} e^{H_{N,\beta}(\sigma)}\mu^{\otimes N}(d\sigma).$$

For any $\eta > 0$ and $v \in \mathcal{V}$, define $A_\eta(v) = (v-\eta, v+\eta)$. For σ such that $R(\sigma,\sigma) \in A_\eta(v)$, note that

$$-\frac{\beta^2}{2}(v+\eta)^p + \frac{1}{N}\log\int e^{\beta X_N(\sigma)}\mu^{\otimes N}(d\sigma) \le F_N(\beta, A_\eta(v)) \le -\frac{\beta^2}{2}(v-\eta)^p + \frac{1}{N}\log\int e^{\beta X_N(\sigma)}\mu^{\otimes N}(d\sigma).$$

From [28, Theorem 1], it is known that

$$\lim_{\eta\downarrow 0}\lim_{N\to\infty}\frac{1}{N}\int_{\sigma:R(\sigma,\sigma)\in A_\eta(v)} e^{\beta X_N(\sigma)}\mu^{\otimes N}(d\sigma) = \inf_{\mathcal{M}_v,\mathbb{R}} \mathcal{P}_{\beta,v}(\alpha,\lambda),$$

thus

$$\lim_{\eta\downarrow 0}\lim_{N\to\infty} F_N(\beta, A_\eta(v)) = \inf_{\mathcal{M}_v\times\mathbb{R}} \mathcal{Q}_{\beta,v}(\alpha,\lambda).$$

Therefore, for any $\delta > 0$ there exists $\eta(v), N(v)$ such that for $N > N(v)$ and $0 \le \eta < \eta(v)$

$$|F_N(\beta, A_\eta(v)) - \inf_{\mathcal{M}_v\times\mathbb{R}} \mathcal{Q}_{\beta,v}(\alpha,\lambda)| < \delta. \tag{4.1}$$

Note that the set of sets $\{A_\eta(v) \mid v \in \mathcal{V}\}$, where for each v, the radius η is chosen so that $\eta < \eta(v)$, forms an open cover of $\mathcal{V}$. Since $\mathcal{V}$ is bounded and any accumulation point of $\mathcal{V}$ is contained in at least one set $A_\eta(v)$ by construction, we can pass to a finite sub-cover $\{A_\eta(v_j) \mid 1 \le j \le n\}$ for $\bar{\mathcal{V}}$, the closure of $\mathcal{V}$. This collection of sets is also a cover of $\mathcal{V}$.

For each $1 \le j \le n$,

$$F_N(\beta, A_\eta(v_j)) \le F_N(\beta). \tag{4.2}$$

Also,

$$F_N(\beta) \leq \frac{1}{N} \log \sum_{j=1}^{n} \exp N F_N(\beta, A_\eta(v_j)) \leq \frac{\log n}{N} + \max_{1 \leq j \leq n} F_N(\beta, A_\eta(v_j)). \tag{4.3}$$

For N large enough, $\frac{\log n}{N} < \delta$. Rearranging equation 4.1 gives

$$\inf_{\mathcal{M}_{v_j} \times \mathbb{R}} \mathcal{Q}(\alpha, \lambda) - \delta < F_N(\beta, A_\eta(v_j)) < \inf_{\mathcal{M}_{v_j} \times \mathbb{R}} \mathcal{Q}(\alpha, \lambda) + \delta.$$

Combining with equation (4.2) to get a lower bound and equation (4.3) to get an upper bound gives

$$\inf_{\mathcal{M}_{v_j} \times \mathbb{R}} \mathcal{Q}(\alpha, \lambda) - \delta < F_N(\beta) < 2\delta + \max_{1 \leq j \leq n} \inf_{\mathcal{M}_{v_j} \times \mathbb{R}} \mathcal{Q}(\alpha, \lambda).$$

Finally, since the left-hand inequality holds for all $1 \leq j \leq n$, we may take the maximum over j to get

$$\max_{1 \leq j \leq n} \inf_{\mathcal{M}_{v_j} \times \mathbb{R}} \mathcal{Q}(\alpha, \lambda) - \delta \leq \liminf_{N \to \infty} F_N(\beta) \leq \limsup_{N \to \infty} F_N(\beta) \leq 2\delta + \max_{1 \leq j \leq n} \inf_{\mathcal{M}_{v_j} \times \mathbb{R}} \mathcal{Q}(\alpha, \lambda).$$

Since $\inf_{\mathcal{M}_{v_j} \times \mathbb{R}} \mathcal{Q}(\alpha, \lambda)$ is continuous in v, taking $\delta \to 0$ gives the result.

□

4.2 Structure of $\mathcal{R}$: Proof of Proposition 1

This section presents the proof of Proposition 1, a key result that states that the high-temperature regime of the scalar-valued p-spin model takes the form of an interval, $\mathcal{R} = [0, \beta_c)$. Two technical lemmas, presented in Subsection 4.2.1, are needed. The first lemma uses Itô's formula to compute the differential of $\gamma_\beta(s)$ to show that this process is a sub-martingale. It will follow that $\Gamma_b(v)$ is increasing in b. The second lemma computes the optimizers of the Parisi formula when $\beta \in \mathcal{R}$. In Subsection 4.2.2 these results are combined to prove Proposition 1.

4.2.1 Two Technical Lemmas

Recall that

$$\gamma_\beta(s) = \mathbb{E}\left[\frac{\left(\int aZ(a, b\xi'(s))\mu(da)\right)^2}{\int Z(a, b\xi'(s))\mu(da)}\right],$$

where $Z(a,t) = e^{aB_t - a^2t/2}$ is a geometric Brownian motion and $\xi(s) = s^p/2$. Recall also that the auxiliary function for computing β_c is $\Gamma_\beta(v) = \int_0^v \xi''(s)(\gamma_\beta(s) - s)ds$. Lemma 5 shows that $\gamma_\beta(s)$ is strictly increasing in β and it follows directly that $\Gamma_\beta(v)$ is also strictly increasing in β.

Lemma 5. If $0 < \beta < \beta'$ then $\gamma_\beta(s) < \gamma_{\beta'}(s)$ for all $s > 0$.

Proof. Note that $Z(a,0) = 1$. Thus $\gamma_\beta(0) = \mathbb{E}\left(\int a\mu(da)\right)^2 = 0$ since we assume μ is centered. For $j = 0, 1, 2, 3$, set

$$g_j(t,x) = \int a^j e^{ax - \frac{a^2 t}{2}} \mu(da).$$

Set $X_t = g_1(t, B_t)^2$ and $Y_t = g_0(t, B_t)^{-1}$. Setting $t = \beta\xi'(s)$ gives $\gamma_\beta(s) = \mathbb{E}X_tY_t$. □

Using Itô's formula to compute dX_t gives

$$\begin{aligned}
dX_t &= \left(2g_1\partial_t g_1 + \frac{1}{2}\frac{\partial}{\partial B_t}(2g_1\partial_{B_t} g_1)\right)dt + 2g_1\partial_x g_1 dB_t \\
&= \left(2g_1\partial_t g_1 + g_1\partial_{xx} g_1 + (\partial_x g_1)^2\right)dt + 2g_1\partial_x g_1 dB_t \\
&= -g_1g_3dt + 2g_1g_2dB_t + g_1g_3dt + g_2^2dt \\
&= g_2^2dt + 2g_1g_2dB_t.
\end{aligned}$$

Also,

$$\begin{aligned}
dY_t &= \left(-\frac{\partial_t g_0}{g_0^2} - \frac{1}{2}\frac{d}{dx}\frac{\partial g_0}{g_0^2}\right)dt - \frac{\partial_x g_0}{g_0^2}dB_t \\
&= \left(-\frac{\partial_t g_0}{g_0^2} - \frac{1}{2}\left(\frac{\partial_{xx} g_0}{g_0^2} - 2\frac{(\partial_x g_0)^2}{g_0^3}\right)\right)dt - \frac{\partial_x g_0}{g_0^2}dB_t \\
&= \frac{1}{2}\frac{g_2}{g_0^2}dt - \frac{1}{2}\frac{g_2}{g_0^2} + \frac{g_1^2}{g_0^3}dt - \frac{g_1}{g_0^2}dB_t \\
&= \frac{g_1^2}{g_0^3}dt - \frac{g_1}{g_0^2}dB_t.
\end{aligned}$$

Using the product rule for Itô processes, $d(X_tY_t) = X_t dY_t + Y_t dX_t + d\langle X_t, Y_t\rangle$, gives

$$\begin{aligned}
d(X_tY_t) &= g_1^2\left(\frac{g_1^2}{g_0^3}dt - \frac{g_1}{g_0^2}dB_t\right) + g_0^{-1}\left(g_2^2 dt + 2g_1g_2 dB_t\right) - \frac{2g_1^2 g_2}{g_0^2}dt \\
&= \left(\frac{g_1^4}{g_0^3} + \frac{g_2^2}{g_0} - \frac{2g_1^2g_2}{g_0^2}\right)dt + \left(-\frac{g_1}{g_0^2} + \frac{2g_1g_2}{g_0}\right)dB_t \\
&= g_0\left(\frac{g_1^2}{g_0^2} - \frac{g_2}{g_0}\right)^2 dt + \left(-\frac{g_1}{g_0^2} + \frac{2g_1g_2}{g_0}\right)dB_t.
\end{aligned}$$

Since the drift term is positive, the process X_tY_t is a sub-martingale, meaning $\mathbb{E}X_tY_t \leq \mathbb{E}X_{t'}Y_{t'}$ for any $0 \leq t < t'$.

If equality holds for some $t < t'$, then

$$0 = \mathbb{E}X_{t'}Y_{t'} - \mathbb{E}X_tY_t = \int_t^{t'} \mathbb{E}\left(g_0\left(\frac{g_1^2}{g_0^2} - \frac{g_2}{g_0}\right)^2\right)ds.$$

It follows that we must have

$$0 = \frac{g_1^2}{g_0^2} - \frac{g_2}{g_0},$$

or equivalently

$$\left(\frac{\int a e^{aB_s - \frac{a^2 s}{2}}}{\int e^{aB_s - \frac{a^2 s}{2}}}\right)^2 = \frac{\int a^2 e^{aB_s - \frac{a^2 s}{2}}}{\int e^{aB_s - \frac{a^2 s}{2}}}$$

for all $t \leq s \leq t'$. Interpreting the above as a Gibbs average, we have $\langle a\rangle^2 = \langle a^2\rangle$. By Jensen's inequality, $\langle a\rangle^2 \leq \langle a^2\rangle$, with equality if and only if there is exactly one value $a \in \Lambda$. This gives a contradiction since we assume that there is more than one value in the set Λ. Therefore the equality is strict. Setting $t = \beta\xi'(s)$ shows that $\gamma_\beta(s)$ is strictly increasing in β.

The next lemma gives the optimizers of the Parisi formula when $F(\beta) = 0$. Recall that $v_* = \int a^2 \mu(da)$. Define $\alpha_v \in \mathcal{M}_v$ by $\alpha_v(s) = 1$ for $s \in [0, v]$. Define

$$\lambda_* = -\frac{\beta^2 p v_*^{p-1}}{2} = -\frac{\beta^2 \xi'(v_*)}{2}.$$

Lemma 6. The following two statements hold:

(i) If $v \neq v_*$ then $\inf_\lambda \mathcal{Q}_{\beta,v}(\alpha_v, \lambda) < 0$

(ii) If $v = v_*$ then $\inf_\lambda \mathcal{Q}_{\beta,v}(\alpha_v, \lambda) = 0$ where the minimizer is given by $\lambda = \lambda_*$.

If for some $\beta > 0$, the supremum over $v \in \mathcal{V}$ in the Parisi formula occurs at $v \neq v_*$, Lemma 6 (i) states that

$$F(\beta) = \inf_{(\alpha,\lambda)\in\mathcal{M}_v\times\mathbb{R}} \mathcal{Q}_{\beta,v}(\alpha, \lambda) \leq \inf_{(\alpha_v,\lambda)\in\mathcal{M}_v\times\mathbb{R}} \mathcal{Q}_{\beta,v}(\alpha, \lambda) < 0,$$

and therefore $\beta \notin \mathcal{R}$. Thus if $\beta \in \mathcal{R}$, then the supremum must occur at v_* and in this case Lemma 6 (ii) says that it is enough to minimizer over all CDFs α_v rather than the all $\alpha \in \mathcal{M}_v$. Since $\alpha_v(s) = 1$ for all $s \in [0, v]$, this means that the Parisi measure is a Dirac measure at zero. Since the Parisi measure gives the limiting overlap distribution, Lemma 6 implies that the scalar-valued p-spin model follows our intuition that in the high-temperature regime, the overlaps concentrate at a single value; in this case the overlaps $R(\sigma^1, \sigma^2)$ concentrate at zero. This is the first step toward proving the concentration results formally stated in Theorem 5 and 7.

Proof. Take $z \sim \mathcal{N}(0, 1)$. With α_v, the Parisi PDE $\Phi_{\beta,\alpha}$ can be solved using the Cole-Hopf transform (see Appendix A.3) giving the Parisi functional

$$\begin{aligned}\mathcal{P}_{\beta,v}(\alpha_v, \lambda) &= \log \mathbb{E} \int e^{\beta z\sqrt{\xi'(v)}a+\lambda a^2} \mu(da) - \lambda v - \frac{\beta^2}{2}\int_0^v \xi''(s) s\, ds \\ &= \log \int e^{\frac{\beta^2 p v^{p-1}}{2}a^2+\lambda a^2} \mu(da) - \lambda v - \frac{\beta^2(p-1)}{2} v^p.\end{aligned}$$

Therefore,

$$\begin{aligned}\inf_{\lambda} \mathcal{Q}_{\beta,v}(\alpha_v,\lambda) &= \inf\left(\log\int e^{\frac{\beta^2 p v^{p-1}}{2}a^2+\lambda a^2}\mu(da) - \lambda v - \frac{\beta^2(p-1)}{2}v^p - \frac{\beta^2}{2}v^p\right)\\ &= \inf\left(\log\int e^{\left(\frac{\beta^2 p v^{p-1}}{2}+\lambda\right)a^2}\mu(da) - v\left(\frac{\beta^2 p v^{p-1}}{2}+\lambda\right)\right).\end{aligned}$$

Set $\lambda' = \frac{\beta^2 p v^{p-1}}{2} + \lambda$ so that the above becomes

$$\inf_{\lambda} \mathcal{Q}_{\beta,v}(\alpha_v,\lambda) = \inf_{\lambda'}\left(\log\int e^{\lambda' a^2}\mu(da) - \lambda' v\right).$$

Define $F(v,\lambda) = \log\int e^{\lambda a^2}\mu(da) - \lambda v$. Using Hölder's inequality gives

$$\begin{aligned}F(v, t\lambda_1 + (1-t)\lambda_2) &= \log\int e^{t\lambda_1 a^2}e^{(1-t)\lambda_2 a^2}\mu(da) - t\lambda_1 v - (1-t)\lambda_2 v\\ &\le \log\left(\left(\int e^{\lambda_1 a^2}\mu(da)\right)^{1/t}\left(\int e^{\lambda_2 a^2}\mu(da)\right)^{1/1-t}\right) - t\lambda_1 v - (1-t)\lambda_2 v\\ &= t\left(\log\int e^{\lambda_1 a^2}\mu(da) - \lambda_1 v\right) + (1-t)\left(\log\int e^{\lambda_2 a^2}\mu(da) - \lambda_2 v\right)\\ &= tF(v,\lambda_1) + (1-t)F(v,\lambda_2).\end{aligned}$$

Thus $F(v,\lambda)$ is convex in the argument λ. The derivative of F with respect to λ is

$$\partial_\lambda F(v,\lambda) = \frac{\int a^2 e^{\lambda a^2}\mu(da)}{\int e^{\lambda a^2}\mu(da)} - v.$$

Plugging in $v = v^*$,

$$\partial_\lambda F(v_*,\lambda) = \frac{\int a^2 e^{\lambda a^2}\mu(da)}{\int e^{\lambda a^2}\mu(da)} - \int a^2\mu(da).$$

To find the minimizer, note that

$$\partial_\lambda F(v_*,0) = \int a^2\mu(da) - \int a^2\mu(da) = 0,$$

so $F(v_*,\lambda')$ is minimized at $\lambda' = 0$ which is equivalent to $\lambda = -\frac{\beta^2 p v_*^{p-1}}{2}$. This completes the proof of part (ii).

To prove part (i), notice that $F(v,0) = \log \int 1\mu(da) = \log 1 = 0$. Also, if $v \neq v_*$, then $\partial_\lambda F(v,0) \neq 0$, and $\lambda = 0$ is not a minimizer. Therefore we must have

$$\inf_\lambda F(v,\lambda) < 0$$

for $v \neq v_*$. □

4.2.2 Proof of Proposition 1

Recall that Proposition 1 states that $\mathcal{R} = [0, \beta_c)$ and that $\beta \in \mathcal{R}$ if and only if $\sup_{v\in[0,v_*]} \Gamma_\beta(v) \leq 0$.

Proof. First suppose $\beta \in \mathcal{R}$. Then, by definition, $F(\beta) = 0$. By Proposition 4,

$$0 = F(\beta) = \sup_v \inf_{\alpha,\lambda} \mathcal{Q}_{\beta,v}(\alpha,\lambda).$$

By Lemma 6, if $v \neq v_*$ then

$$\inf_{\alpha,\lambda} \mathcal{Q}_{\beta,v}(\alpha,\lambda) \leq \inf_\lambda \mathcal{Q}_{\beta,v}(\alpha_v,\lambda) < 0,$$

so it must be the case that

$$F(\beta) = \inf_{\alpha,\lambda} \mathcal{Q}_{\beta,v_*}(\alpha,\lambda).$$

Again from Lemma 6, $(\alpha_{v_*}, \lambda_*)$ is a minimizer of the right-hand side above above, so

$$F(\beta) = \mathcal{Q}_{\beta,v_*}(\alpha_{v_*},\lambda_*).$$

Recall that

$$\mathcal{Q}_{\beta,v_*}(\alpha,\lambda_*) = \Phi_{\beta,v_*,\alpha}(0,0,\lambda_*) - \frac{\beta^2}{2}\int_0^{v_*} \alpha(s)\xi''(s)ds + \frac{\beta^2(p-1)v_*^p}{2}.$$

Since the boundary condition $\Phi_{\beta,v_*,\alpha}(v,x,\lambda)$ is convex, the map $(\alpha,\lambda) \in \mathcal{M}_{v_*} \times \mathbb{R} \mapsto \mathcal{Q}_{\beta,v_*}(\alpha,\lambda)$ is convex. Set $\alpha_\theta = (1-\theta)\alpha_{v_*} + \theta\alpha$ and $\lambda_\theta = (1-\theta)\lambda_* + \theta\lambda$ for $\theta \in [0,1]$. By convexity, since $(\alpha_{v_*}, \lambda_*)$ is a minimizer of $\mathcal{Q}_{\beta,v_*}$, the derivative of $\mathcal{Q}_{\beta,v_*}(\alpha_\theta,\lambda_\theta)$ with respect to θ as $\theta \to 0$ from the right must be non-negative. Computing this derivative yields

$$\frac{d}{d\theta}\mathcal{Q}_{\beta,v_*}(\alpha_\theta,\lambda_\theta)\Big|_{\theta=0} = \frac{\beta^2}{2}\int_0^{v_*}\xi''(s)(\alpha(s)-\alpha_{v_*}(s))(\gamma_\beta(s)-s)ds + \left(\int a^2\mu(da) - v_*\right)(\lambda-\lambda_*)$$
$$= \frac{\beta^2}{2}\int_0^{v_*}\xi''(s)(\alpha(s)-\alpha_{v_*}(s))(\gamma_\beta(s)-s)ds.$$

Using the fact that $\alpha_{v_*} \equiv 1$ on $[0,\alpha_{v_*}]$, write

$$\begin{aligned}\int_0^{v_*}\xi''(s)(\alpha(s)-\alpha_{v_*}(s))(\gamma_\beta(s)-s)ds &= \int_0^{v_*}\int_0^{s}\xi''(s)(\gamma_\beta(s)-s)\alpha(dv)ds - \int_0^{v_*}\xi''(s)(\gamma_\beta(s)-s)ds.\\ &= \int_0^{v_*}\int_v^{v_*}\xi''(s)(\gamma_\beta(s)-s)ds\alpha(dv) - \int_0^{v_*}\xi''(s)(\gamma_\beta(s)-s)ds\\ &= \int_v^{v_*}\xi''(s)(\gamma_\beta(s)-s)ds - \int_0^{v_*}\xi''(s)(\gamma_\beta(s)-s)ds\\ &= -\int_0^{v}\xi''(s)(\gamma_\beta(s)-s)ds\\ &= -\Gamma_\beta(v).\end{aligned}$$

Thus the optimality condition $\frac{d}{d\theta}\mathcal{Q}_{\beta,v_*}(\alpha_\theta,\lambda_\theta)\big|_{\theta=0} \geq 0$ translates to $\Gamma_\beta(v) \leq 0$ for all $v \in [0,v_*)$, or equivalently, $\sup_{v\in(0,v_*]}\Gamma_\beta(v) \leq 0$. The reverse direction is identical.

To show that $\mathcal{R}$ is an interval, recall from Lemma 5 that $\Gamma_\beta(v)$ is increasing in β. By this and the preceding argument, if $\beta \in \mathcal{R}$ then $\beta' \in \mathcal{R}$ for all $0 < \beta' \leq \beta$ since $\Gamma_{\beta'}(v) < \Gamma_\beta(v)$ for all $v \in [0,v_*]$ which implies that

$$\sup_{v\in[0,v_*]}\Gamma_{\beta'}(v) < \sup_{v\in[0,v_*]}\Gamma_{\beta'}(v) \leq 0.$$

Since $\beta_c = \sup\mathcal{R}$, we must have $\sup_{v\in[0,v_*]}\Gamma_\beta(v) > 0$ for all $\beta > \beta_c$. Suppose that $\sup_{v\in[0,v_*]}\Gamma_{\beta_c}(v) < 0$. Since $\Gamma_\beta(v)$ is continuous in β, it is therefore possible to find $\beta' > \beta_c$ such that $\sup_{v\in[0,v_*]}\Gamma_{\beta'}(v) \leq 0$. Thus $\beta' \in \mathcal{R}$, which contradicts the maximality of β_c. Therefore, $\mathcal{R} = (0,\beta_c]$ and β_c is the largest β such that $\sup_{v\in[0,v_*]}\Gamma_\beta(v) = 0$. $\square$

4.3 Guerra Talagrand 1-RSB Bound and Overlap Concentration

Theorem 5 states that for the vector-valued spin glass, the overlaps $R(\sigma^1(r), \sigma^2(r))$ concentrate at zero for spins σ^1, σ^2 sampled from the Gibbs measure associated to the vector-valued Hamiltonian. The first step towards this result is the Guerra-Talagrand 1-replica symmetry breaking bound which controls the free energy of two coupled copies of the Hamiltonian $H_{N,\beta}(\sigma)$.

Set $M_2(\mathbb{R})$ as the set of all real-valued 2×2 matrices with the metric

$$\|V - V'\|_{\max} = \max_{1 \le r, r' \le 2} |V_{r,r'} - V'_{r,r'}|.$$

The inner product on this space is

$$\langle V, V' \rangle = \sum_{i,j} V_{ij} V'_{ij}.$$

Given two spin configurations σ^1, σ^2 define their overlap matrix by

$$\mathbf{R}(\sigma^1, \sigma^2) = \begin{bmatrix} R(\sigma^1, \sigma^1) & R(\sigma^1, \sigma^2) \\ R(\sigma^2, \sigma^1) & R(\sigma^2, \sigma^2) \end{bmatrix}.$$

Given any set $A \subseteq M_2(\mathbb{R})$ such that all elements of A are positive semi-definite, write '$\int_A$' to mean the integral over the set $\{(\sigma^1, \sigma^2) : \mathbf{R}(\sigma^1, \sigma^2) \in A\}$. For any such A, define the coupled free energy restricted to the set A by

$$CF_N(\beta, A) = \frac{1}{N} \log \int_A e^{H_{N,\beta}(\sigma^1) + H_{N,\beta}(\sigma^2)} \mu^{\otimes N}(d\sigma^1) \mu^{\otimes N}(d\sigma^2). \tag{4.4}$$

Recall that $\mathcal{V} = \{v^2 \mid v \in \Lambda\}$ and $\mathcal{M}_v$ is the set of all cumulative distribution functions on the interval $[0, v]$. Fix $v \in \mathcal{V}$ and fix $v_0 \in \mathbb{R}$ such that $v_0 > 0$ and

$$V := \begin{bmatrix} v & v_0 \\ v_0 & v \end{bmatrix}$$

is positive semi-definite. This is satisfied when $0 < v_0 \le v$. Define a function $T : [0, v] \to$

$M_2(\mathbb{R})$ by

$$T(s) = \begin{cases} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} & s \in [0, v_0) \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} & s \in [v_0, v] \end{cases}.$$

For any $\alpha \in \mathcal{M}_v$ let $\Psi_{\beta,V,\alpha}(s,x,\lambda)$ be the weak solution to the following PDE for $(s,x,\lambda) \in [0,v) \times \mathbb{R}^2 \times M_2(\mathbb{R})$:

$$\partial_s \Psi_{\beta,V,\alpha} = -\frac{\beta^2 \xi''}{2} \left(\langle \nabla^2 \Psi_{\beta,V,\alpha}, T \rangle + \alpha \langle T \nabla \Psi, \nabla \Psi \rangle \right)$$

with boundary condition

$$\Psi_{\beta,V,\alpha}(v,x,\lambda) = \log \int e^{\langle a,x \rangle + \langle \lambda a, a \rangle} \mu \times \mu(da).$$

Here, the gradient refers to the derivatives with respect to $x = (x_1, x_2)$ only. The existence of the solution $\Psi_{\beta,V,\alpha}$ is shown in [29]. Define a Parisi functional

$$\mathcal{P}_{\beta,V}(\alpha,\lambda) = \Psi_{\beta,V,\alpha}(0,0,\lambda) - \langle \lambda, V \rangle - \beta^2 \left(\int_0^v \xi''(s) s\alpha(s) ds + \int_0^{v_0} \xi''(s) s\alpha(s) ds \right).$$

The function $T(s)$ together with the CDF α play the role of the functional order parameter represting possible distributions for each entry of the overlap matrix $\mathbf{R}(\sigma^1, \sigma^2)$.

For $\eta > 0$ define

$$A_\eta(V) = \{ V' \in M_2(\mathbb{R}) \mid \|V' - V\|_{\max} < \eta \}.$$

The Guerra-Talagrand bound of [30] states that if p is even, then for any $(\alpha, \lambda) \in \mathcal{M}_v \times M_2(\mathbb{R})$,

$$\lim_{\eta \downarrow 0} \limsup_{N \to \infty} \frac{1}{N} \log \int_{A_\eta(V)} e^{\beta X_N(\sigma^1) + \beta X_N(\sigma^2)} \mu^{\otimes N}(d\sigma^1) \mu^{\otimes N}(d\sigma^2) \leq \mathcal{P}_{\beta,V}(\alpha,\lambda).$$

Set $\mathcal{Q}_{\beta,V}(\alpha,\lambda) = \mathcal{P}_{\beta,V}(\alpha,\lambda) - \beta^2 v^p$. Thus when p is even, the coupled free energy of equation (4.4) restricted to $A_\eta(V)$ is bounded in a similar way:

$$\begin{aligned} CF_N(\beta, A_\eta(V)) &= \frac{1}{N}\log\int_{A_\eta(V)} e^{H_{N,\beta}(\sigma^1)+H_{N,\beta}(\sigma^2)}\mu^{\otimes N}(d\sigma^1)\mu^{\otimes N}(d\sigma^2) \\ &= \frac{1}{N}\log\int_{A_\eta(V)} e^{\beta X_N(\sigma^1)+\beta X_N(\sigma^2)-\frac{\beta^2 N}{2}(R(\sigma^1,\sigma^1)^p+R(\sigma^2,\sigma^2)^p)}\mu^{\otimes N}(d\sigma^1)\mu^{\otimes N}(d\sigma^2) \\ &\le \frac{1}{N}\log\int_{A_\eta(V)} e^{\beta X_N(\sigma^1)+\beta X_N(\sigma^2)-\beta^2 N(v-\eta)^p}\mu^{\otimes N}(d\sigma^1)\mu^{\otimes N}(d\sigma^2) \\ &= \frac{1}{N}\log\int_{A_\eta(V)} e^{\beta X_N(\sigma^1)+\beta X_N(\sigma^2)}\mu^{\otimes N}(d\sigma^1)\mu^{\otimes N}(d\sigma^2) - \beta^2 N(v-\eta)^p. \end{aligned}$$

Thus

$$\lim_{\eta\downarrow 0}\limsup_{N\to\infty}\mathbb{E}CF_N(\beta, A_\eta(V)) \le \mathcal{Q}_{\beta,V}(\alpha,\lambda). \tag{4.5}$$

Whether the same bound holds for odd p for general choices of α, λ is still an open question. Chen [9] showed that specific choices of α, λ give the Guerra-Talagrand 1-Replica Symmetry Breaking (1-RSB) Bound for even and odd p. Specifically, for $0 < v_0 < v$, set $\mathcal{M}_{v,v_0}$ to be the set of all $\alpha \in \mathcal{M}_v$ such that $\alpha \equiv c$ on $[0, v_0)$ for some constant $c \le 1$ and $\alpha \equiv 1$ on $[v_0, v]$. This type of measure is known as 1-RSB. A replica symmetric measure is a Dirac measure. If the Parisi measure is replica symmetric, all overlaps concentrate at a single value. If the Parisi measure is 1-RSB, the overlaps concentrate at two values - there is one level of replica symmetry breaking. Proposition 5 states that equation 4.5 holds for 1-RSB measures. The proof of Proposition 5 in the present setting is nearly identical to the proof of [9, Proposition 2], and is not reproduced here.

Proposition 5. For $p \ge 2, V \in \mathcal{V}, v_0 \in [0, v], \lambda \in M_2(\mathbb{R})$ and $\alpha \in \mathcal{M}_{v,v_0}$,

$$\lim_{\eta\downarrow 0}\limsup_{N\to\infty}\mathbb{E}CF_N(\beta, A_\eta(v)) \le \mathcal{Q}_{\beta,V}(\alpha,\lambda).$$

Proposition 5 is important because if it can be shown that $\mathcal{Q}_{\beta,V}(\alpha,\lambda) < 0$, then the coupled free energy $CF_N(\beta, A_\eta(V))$ exhibits a free energy cost: it is strictly negative. Then, applying a covering argument similar to that used in the proof of Proposition 4 shows that the overlaps $R(\sigma^1, \sigma^2)$ are concentrated at zero, as formally stated in

Proposition 6 below.

Proposition 6. Assume that $0 < \beta < \beta_c$ and that $s_0 \in (0,1)$. For any $\varepsilon > 0$, there exists a constant $K > 0$, depending only on β, s_0, and ε such that if σ^1, σ^2 are i.i.d. samples from $G_{N,s\beta}$ then

$$\mathbb{E}\langle I(|R(\sigma^1,\sigma^2)| \geq \varepsilon)\rangle_{s\beta} \leq Ke^{-N/K} \tag{4.6}$$

for all $N \geq 1$ and all $s \in [s_0, 1]$.

Proof. Let $0 < \varepsilon < v_*$ be fixed. Assume that $v_0 \in [\varepsilon, v_*]$. Fix a diagonal matrix $\lambda \in M_2(\mathbb{R})$ with diagonal entries $\lambda_{1,1} = \lambda_{2,2} = -\beta^2\xi'(v_*)/2$. Let $\alpha \in \mathcal{M}_{v_*}$ satisfy $\alpha \equiv 0$ on $[0, v_0)$ and $\alpha \equiv 1$ on $[v_0, v_*]$. Define

$$\alpha_\theta(s) = \begin{cases} \frac{1-\theta}{2} & s \in [0, v_0) \\ 1 & s \in [v_0, v_*] \end{cases}$$

for $\theta \in [0,1]$.

Using the Cole-Hopf transformation, one may compute that

$$\begin{aligned} \mathcal{Q}_{\beta,V}(\alpha_\theta,\lambda) = 2\left(\frac{1}{1-\theta}\log \mathbb{E}g_0(\beta^2\xi'(v_0), B_{\beta\xi'(v_0)^{1/2}})^{1-\theta} + \frac{\beta^2}{2}\xi'(v_*)v_*\right) \\ - \beta^2\left((1-\theta)\int_0^{v_0}\xi''(s)sds + \int_{v_0}^{v_*}\xi''(s)sds\right) - \beta^2 v_*^p, \end{aligned}$$

where $g_0(t,x) = \int e^{ax-a^2t/2}\mu(da)$ and B_t is a standard Brownian motion. Also,

$$\partial_\theta \mathcal{Q}_{\beta,V}(\alpha_\theta,\lambda)|_{\theta=0} = \beta^2\Gamma_\beta(v_0).$$

By the monotonicity of $\gamma_\beta(v_0)$ in Lemma 5,

$$\partial_\theta \mathcal{Q}_{\beta,V}(\alpha_\theta,\lambda)|_{\theta=0} < \beta_c^2\Gamma_{\beta_c}(v_0) \leq 0.$$

Thus $\mathcal{Q}_{\beta,V}(\alpha_\theta,\lambda)$ is decreasing at $\theta = 0$. Since $\mathcal{Q}_{\beta,v}(\alpha_\theta,\lambda)$ is continuous in (v_0,θ), there exists $\delta > 0$ such that

$$\sup_{v_0\in[\varepsilon,v_*]}\inf_{\theta\in[0,1]}\mathcal{Q}_{\beta,V}(\alpha_\theta,\lambda) \leq -\delta.$$

Combining this bound with Proposition 5, we see that

$$\lim_{\eta\downarrow 0}\limsup_{N\to\infty}\mathbb{E}CF_N(\beta, A_\eta(V)) < -\delta. \tag{4.7}$$

Since we assume that $\beta \in \mathcal{R}$, we have that $F(\beta) = 0$, so the above is

$$\lim_{\eta\downarrow 0}\limsup_{N\to\infty}\mathbb{E}CF_N(\beta, A_\eta(V)) < F(\beta) - \delta.$$

For each V, there exists $N(v)$ and $\eta(V)$ such that for all $N \geq N(V)$ and $\eta < \eta(V)$

$$\mathbb{E}CF_N(\beta, A_\eta(V)) < F(\beta) - \delta.$$

Then By Gaussian concentration of measure (see Appendix A.2), there exists $K > 0$ such that

$$P\left(|CF_N(\beta, A_\eta(V)) - \mathbb{E}CF_N(\beta, A_\eta(V))| \geq \frac{\delta}{2}\right) \leq Ke^{-N/K}.$$

Thus with probability at least $1 - Ke^{-N/K}$,

$$CF_N(\beta, A_\eta(V)) \leq F(\beta) - \frac{\delta}{2}. \tag{4.8}$$

For each $v_0 \in [\varepsilon, v_*]$, choose $0 < \eta < \eta_V$. Let

$$A_\varepsilon = \left\{ V = \begin{bmatrix} v_1 & v_0 \\ v_0 & v_1 \end{bmatrix} \mid \varepsilon \leq v_0 \leq v_*, |v_1 - v_*| \leq \varepsilon, V \succeq 0 \right\}.$$

For small enough ε the set of sets $A_\eta(V)$ cover A_ε. Thus, as in the proof of Proposition 5, we may choose a finite subcover $A_\eta(V_j)$ for $1 \leq j \leq n$. Following the arguments of Proposition 5, for each $1 \leq j \leq n$

$$CF_N(\beta, A_\varepsilon) \leq \frac{1}{N}\log\sum_{j=1}^{n}\exp NF_N(\beta, A_\eta(V_j)) \leq \frac{\log n}{N} + \max_{1\leq j\leq n} CF_N(\beta, A_\eta(V_j)) < \frac{\log n}{N} + F(\beta) - \frac{\delta}{2}.$$

For N large enough, $\log n/N < \delta/4$, and

$$CF_N(\beta, A_\varepsilon) < F(\beta) - \frac{\delta}{4}.$$

Rearranging this expression gives

$$\langle I(\mathbf{R} \in A_\varepsilon)\rangle_\beta \leq e^{-N\delta/4}.$$

with probability at least $1 - Ke^{-N/K}$. Thus there exists $K' > 0$ such that

$$\mathbb{E}\,\langle I(\mathbf{R} \in A_\varepsilon)\rangle_\beta \leq K'e^{-N/K'}.$$

When p is even, $H_{N,\beta}(\sigma) = H_{N,\beta}(-\sigma)$, so for even p,

$$\mathbb{E}\langle I(|R(\sigma^1,\sigma^2)| \leq \varepsilon, |R(\sigma^1,\sigma^1)| \in A_\varepsilon(v_*), |R(\sigma^2,\sigma^2)| \in A_\varepsilon(v_*))\rangle_\beta \leq 2K'e^{-N/K'}. \quad (4.9)$$

When p is odd, use Jensen's inequality on the expected value to see that

$$\mathbb{E}CF_N(\beta, A_\eta(V)) \leq \frac{1}{N}\log \int_{\sigma|R(\sigma)\in A_{\eta(V)}} e^{\beta^2 R(\sigma^1,\sigma^2)^p}\mu^{\otimes N}(d\sigma^1, d\sigma^2) < \beta^2(v_0+\eta)^p < 0$$

for any $v_0 \leq -\varepsilon$ and $0 < \eta < \varepsilon/2$. Again using a covering argument and Gaussian concentration of measure gives the existence of a constant $K'' > 0$ such that for odd p

$$\mathbb{E}\langle I(R(\sigma^1,\sigma^2) \leq -\varepsilon, R(\sigma^1,\sigma^1) \in A_\varepsilon(v_*), R(\sigma^2,\sigma^2) \in A_\varepsilon(v_*))\rangle_\beta \leq K''e^{-N/K''}. \quad (4.10)$$

Combining equations 4.9, and 4.10, there exists a constant $L > 0$ such that

$$\mathbb{E}\langle I(|R(\sigma^1,\sigma^2)| \geq \varepsilon, |R(\sigma^1,\sigma^1)| \in A_\varepsilon(v_*), |R(\sigma^2,\sigma^2)| \in A_\varepsilon(v_*))\rangle_\beta \leq Le^{-N/L}. \quad (4.11)$$

Also, from Proposition 7 stated in Section 4.6, which controls the total overlap for the vector-valued Hamiltonian, there exists a constant $L' > 0$ such that

$$\mathbb{E}\langle I(R(\sigma^1,\sigma^1) \notin A_\varepsilon(v))\rangle_\beta \leq L'e^{-N/L'}. \quad (4.12)$$

Combining equations 4.11 and 4.12,

$$\mathbb{E}\langle I(|R(\sigma^1,\sigma^2)| \geq \varepsilon)\rangle_\beta \leq Le^{-N/L} + 2L'e^{-N/L'}.$$

Finally, we show that (4.6) holds for all $s \in [s_0, 1]$. Denote the coupled free energy of two pure p-spin models by

$$CF_N^X(\beta, A_\eta(V)) = \frac{1}{N}\log \int_{A_\eta(V)} e^{\beta X_N(\sigma^1)+\beta X_N(\sigma^2)} \mu^{\otimes N}(d\sigma^1)\mu^{\otimes N}(d\sigma^2).$$

Since $\mathbb{E}CF_N^X(\beta, A_\eta(V))$ is convex in the temperature parameter β, the convergence of

$$\limsup_{N\to\infty} \mathbb{E}CF_N^X(\beta, A_\eta(V))$$

is uniform.

For $\sigma^1, \sigma^2 \in A_\eta(V)$, we know that both $|R(\sigma^1,\sigma^1)-v_*| < \eta$ and $|R(\sigma^2,\sigma^2)-v_*| < \eta$. This restriction on the overlaps gives the upper and lower bounds

$$-\beta^2(v_*+\eta)+\limsup_{N\to\infty} \mathbb{E}CF_N^X(\beta, A_\eta(V)) \leq \limsup_{N\to\infty} \mathbb{E}CF_N(\beta, A_\eta(V)) \leq \limsup_{N\to\infty} \mathbb{E}CF_N^X(\beta, A_\eta(V))-\beta^2(v_*-\eta).$$

Thus the convergence of $\mathbb{E}CF_N(\beta, A_\eta(V))$ is uniform. For every $v_0 \in [\varepsilon, v]$, there exists $\eta(v_0)$ and $N(v_0)$ such that for all $\eta < \eta(v_0)$ and $N \geq N(v_0)$,

$$\mathbb{E}CF_N(\beta, A_\eta(V)) < -\frac{\delta}{2}.$$

We may pass to a finite sub-cover of $[\varepsilon, v]$ and consider only the sets $A_\eta(V_j)$ for $1 \leq j \leq n$, where the matrix V_j has off-diagonal entries $v_j \in [\varepsilon, v]$ and diagonal entries v.

Similar to the methods in the proof of Proposition 4, for any $1 \leq j \leq n$

$$CF_N(\beta, A_\eta(V_j)) \leq CF_N\left(\beta, \bigcup_{j=1}^n A_\eta(V_j)\right) \leq \frac{1}{N}\sum_{j=1}^n \exp NCF_N(\beta, A_\eta(V_j)) \leq \frac{\log n}{N} + \max_{1\leq j\leq n} CF_N(\beta, A_\eta(V_j)).$$

We also notice that the error probability for the Gaussian concentration of measure inequality can be uniformly controlled in temperature. Furthermore, the auxiliary function Γ_β is continuous. Combining these facts, we conclude that all coupled free energies for temperatures $s\beta$ exhibit a uniform energy cost. That is, there exists a $\delta > 0$ such that a bound of the form (4.7) holds for all $s\beta$. This in turn implies that (4.6) holds and completes our proof. □

4.4 Cavity Argument

The cavity method is used to control the even overlap moments. At its heart, the cavity method is induction on N, the number of spins in the system. The method creates an interpolating path between a system with $N-1$ spins and a system of N spins and controls how much the overlap moments change along the path. The goal is to show that this change is not 'too large' and thus control the overlap moments of the N spin system based on bounds known for the $N-1$ spin system.

The cavity method here depends on a good interpolating Hamiltonian with $H_{N,\bar{\beta}}(\bar{\sigma})$ at one end of the interpolating path and a well-behaved Hamiltonian at the other. While the previous sections focused exclusively on the scalar-valued model, most proofs in this section are presented for the vector-valued model only, but the argument is extended to the scalar-valued model by setting $k=1$.

To define the interpolating Hamiltonian, for any $A \subseteq \{1, 2, \dots, p\}$, define $I_A = \{(i_1, \dots, i_p) \in \{1, \dots, N\}^p\}$ where $i_s = N$ exactly when $s \in A$. Define

$$X_N^A(\sigma(r)) = \frac{1}{N^{(p-1)/2}} \sum_{(i_1,\dots,i_p)\in I_A} Y_{i_1,\dots,i_p} \sigma_{i_1}(r) \cdots \sigma_{i_p}(r).$$

For any sets A, A' with $A \neq A'$,

$$\mathbb{E} X_N^A(\sigma(r)) X_N^A(\sigma(r')) = \frac{1}{N^{j-1}} (R_{1,2}^{-,r,r'})^{p-|A|} (\varepsilon_1^r \varepsilon_2^r)^{|A|}$$

and

$$\mathbb{E} X_N^A(\sigma(r)) X_N^{A'}(\sigma(r')) = 0.$$

Here, for more compact notation define,

$$R_{1,2}^{r,r} = R(\sigma^1(r), \sigma^2(r)) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^1(r) \sigma_i^2(r')$$

and

$$R_{1,2}^{-,r,r} = \frac{1}{N}\sum_{i=1}^{N-1}\sigma_i^1(r)\sigma_i^2(r').$$

The purpose of the interpolating Hamiltonian is to decouple the N-th spin $\bar{\sigma}_N = (\sigma_N(1), \ldots, \sigma_N(k))$. Notice that $A = \varnothing$ is the only set that does not depend on $\bar{\sigma}_N$. The processes $X_N^{\varnothing}(\sigma(r))$ make up the 'well-behaved' end of the path. For $t \in [0,1]$ define the interpolating Hamiltonian by

$$\begin{aligned} H_{N,\bar{\beta},t}(\bar{\sigma}) = &\sum_{r=1}^{k}\beta_r\left(X_N^{\varnothing}(\sigma(r)) + \sqrt{t}\sum_{j=1}^{p}\sum_{A:|A|=j}X_N^A(\sigma(r))\right) \\ &- \sum_{1\le r,r'\le k}\beta_r\beta_{r'}\left((R^{-,r,r'})^p + \frac{t}{N^{j-1}}\sum_{j=1}^{p}\sum_{A:|A|=j}(R^{-,r,r'})^{p-j}(\varepsilon^r\varepsilon^{r'})^j\right). \end{aligned}$$

One may check that when $t = 1$, the interpolating Hamiltonian is equal to the original Hamiltonian, $H_{N,\bar{\beta},1}(\bar{\sigma}) = H_{N,\bar{\beta}}$. Also, when $t = 0$, the interpolating Hamiltonian is equal to H_{N-1} at the temperature vector with entries

$$\beta_r\frac{(N-1)^{(p-1)/2}}{N^{(p-1)/2}}.$$

Let $\nu_{\bar{\beta},t}(f) = \mathbb{E}\langle f\rangle_{\bar{\beta},t}$ denote the expected Gibbs average for the Gibbs measure associated to $H_{N,\bar{\beta},t}$. When $t = 1$, we write $\nu_{\bar{\beta},1}(f) = \nu_{\bar{\beta}}(f)$.

4.4.1 Technical Lemmas

This section presents several technical lemmas necessary to the cavity method. Lemmas 7 and 8 together bound $\nu_{\bar{\beta},t}(\cdot)$ by $\nu_{\bar{\beta}}(t)$. Lemma 9 shows that $R_{1,2}^{r,r'}$ is close to $R_{1,2}^{-,r,r'}$: overlaps at either end of the interpolating path are close. Finally, Lemma 10 gives a bound for the moments of $R_{1,2}^{-,r,r'}$ based on bounds for the moments of $R_{1,2}^{r,r}$.

Define $\varepsilon_\ell^r = \sigma_N^\ell(r)$.

Lemma 7. If f is a real-valued, bounded function of $\bar{\sigma}^1, \ldots, \bar{\sigma}^n$, then

$$\nu'_{\bar{\beta},t}(f) = \sum_{r,r'=1}^{k}\beta_r\beta_{r'}\sum_{j=1}^{p}\binom{p}{j}\frac{1}{N^j}\left(\sum_{1\le\ell<\ell'\le n}\nu_{t,\bar{\beta}}(f(R_{\ell,\ell'}^{-,r,r'})^{p-j}(\varepsilon_\ell^r\varepsilon_{\ell'}^{r'})^j\right).$$

Proof. Computing the derivative is straightforward but tedious. First, taking the derivative in t yields

$$\nu'_{\bar\beta,t}(f) = \sum_{r=1}^{k}\sum_{j=1}^{p}\sum_{A:|A|=j}\left(\frac{\beta_r}{\sqrt{t}}\mathbb{E}\left\langle f\sum_{\ell=1}^{n} X_N^A(\sigma^\ell(r))\right\rangle - \frac{\beta^2}{2N^{j-1}}\mathbb{E}\left\langle f\sum_{\ell=1}^{n}(R_{\ell,\ell}^{-,r,r})^{p-j}(\varepsilon_\ell^r\varepsilon_\ell^r)^j\right\rangle\right.$$
$$\left. - n\,\frac{\beta_r}{2\sqrt{t}}\mathbb{E}\left\langle fX_N^A(\sigma^{n+1}(r))\right\rangle - \frac{\beta^2}{2N^{j-1}}\mathbb{E}\left\langle f(R_{n+1,n+1}^{-'r,r})^{p-j}(\varepsilon_{n+1}\varepsilon_{n+1})^j\right\rangle\right)$$
$$= \sum_{r=1}^{k}\sum_{j=1}^{p}\sum_{A:|A|=j}\left(\frac{\beta_r}{\sqrt{t}}\mathbb{E}\left\langle f\sum_{\ell=1}^{n} X_N^A(\sigma^\ell(r))\right\rangle - n\frac{\beta_r}{2\sqrt{t}}\mathbb{E}\left\langle fX_N^A(\sigma^{n+1}(r))\right\rangle\right).$$

The second and fourth terms cancel by symmetry between sites.

By Gaussian integration by parts, the above becomes

$$\sum_{1\le r,r'\le k}\frac{\beta_r\beta_{r'}}{2}\sum_{j=1}^{p}\sum_{A:|A|=j}\left(\sum_{1\le \ell,\ell'\le n}\mathbb{E}\langle f\mathbb{E}X_N^A(\sigma(r))X_N^A(\sigma(R'))\rangle - n\sum_{\ell\le n}\mathbb{E}\langle f\mathbb{E}X_N^A(\sigma^\ell(r))X_N^A(\sigma^{n+1}(r'))\rangle\right.$$
$$\left. -n\sum_{\ell\le n+1}\mathbb{E}\langle f\mathbb{E}X_N^A(\sigma^{n+1}(\sigma(r))X_N^A(\sigma^\ell(r'))\rangle + n(n+1)\mathbb{E}\langle f\mathbb{E}X_N^A(\sigma^{n+1}(r))X_N^A(\sigma^{n+2}(r))\rangle\right).$$

Taking the expectation inside the Gibbs average and combining like terms using symmetry between sites gives the result. □

Lemma 8. For f a non-negative and bounded function of $\bar\sigma^1,\dots,\bar\sigma^r$

$$\nu_{\bar\beta,t}(f) \le \exp\left(n^2 2^{p+1}M^{2p}\sum_{1\le r,r'\le k}\beta_r\beta_{r'}\right)\nu_{\bar\beta}(f).$$

Proof. For any pair $1\le \ell,\ell'\le n+2$ and any pair $1\le r,r'\le k$, the overlap terms in the expression for $\nu'_{\bar{\beta},t}(f)$ can be bounded by a constant as follows:

$$
\begin{aligned}
|(R^{-,r,r'}_{\ell,\ell'})^{p-j}(\varepsilon^r_\ell \varepsilon^{r'}_{\ell'})^j| &\leq \frac{1}{N^{p-j}} \sum_{1\leq i_1,\ldots,i_{p-j}\leq N-1} |\sigma^\ell_{i_1}(r)\sigma^{\ell'}_{i_1}(r')\cdots\sigma^\ell_{i_{p-j}}(r)\sigma^{\ell'}_{i_{p-j}}(r')||\varepsilon^r_\ell \varepsilon^{r'}_{\ell'}|^j \\
&\leq \frac{1}{N^{p-j}} \sum_{1\leq i_1,\ldots,i_{p-j}\leq N-1} M^{2p} \\
&= \frac{(N-1)^{p-j}M^{2p}}{N^{p-j}} \\
&\leq M^{2p}. \qquad (4.13)
\end{aligned}
$$

Using this to bound $|\nu'_{\bar{\beta},t}(f)|$ gives

$$
\begin{aligned}
|\nu'_{\bar{\beta},t}(f)| &\leq M^{2p} \sum_{1\leq r,r'\leq k} \beta_r\beta_{r'} \sum_{j=1}^{p} \binom{p}{j} 2n^2 \nu_{\bar{\beta},t}(f) \\
&\leq M^{2p} \sum_{1\leq r,r'\leq k} \beta_r\beta_{r'} 2^{p+1} n^2 \nu_{\bar{\beta},t}(f).
\end{aligned}
$$

This first inequality uses the bound in equation (4.13), the fact that $1/N^{j-1} \leq 1$ and the fact that $\nu'_{\bar{\beta},t}(f)$ is composed of $2n^2$ terms, and the second inequality uses the fact that $\sum_{j=1}^{p}\binom{p}{j} = 2^p - 1 \leq 2^p$.

In particular

$$
-\nu'_{\bar{\beta},t}(f) \leq M^{2p} 2^{p+1} n^2 \nu_{\bar{\beta},t}(f) \sum_{1\leq r,r'\leq k} \beta_r\beta_{r'}.
$$

Since f is non-negative, $\nu_{\bar{\beta},t}(f)$ is non-negative for every $t \in [0,1]$, thus dividing the above by $\nu_{\bar{\beta},t}(f)$ does not change the direction of the inequality. Integrating the result gives

$$
\begin{aligned}
-\int_0^t \frac{\nu'_{\bar{\beta},t}(f)}{\nu_{\bar{\beta},f}(f)} dt &\leq \int_t^1 n^2 2^{p+1} M^{2p} \sum_{1\leq r,r'\leq k} \beta_r\beta_{r'} dt \\
\log\left(\frac{\nu_{\bar{\beta},t}(f)}{\nu_{\bar{\beta},1}(f)}\right) &\leq n^2 2^{p+1} M^{2p} \sum_{1\leq r,r'\leq k} \beta_r\beta_{r'}.
\end{aligned}
$$

Exponentiating both sides and multiplying by $\nu_{\bar{\beta}}(f) \geq 0$ gives the result. □

Lemma 9. For any $m \geq 1$ and any pair $1 \leq r, r' \leq k$,

$$|(R_{1,2}^{r,r'})^{m+1} - (R_{1,2}^{-,r,r'})^{m+1}| \leq \frac{M^2 m}{N} \left(|R_{1,2}^{r,r'}|^m + |R_{1,2}^{-,r,r'}|^m \right).$$

Proof. For any $x, y \in \mathbb{R}$,

$$|x^{m+1} - y^{m+1}| = \left| (x-y) \sum_{\ell=0}^{m} x^\ell y^{m-\ell} \right| \leq |x-y| \sum_{\ell=0}^{m} |x|^\ell |y|^{m-\ell}.$$

Without loss of generality, assume that $|x| \leq |y|$. Then

$$|x-y| \sum_{\ell=0}^{m} |x|^\ell |y|^{m-\ell} \leq |x-y| \sum_{\ell=0}^{m} |y|^m \leq |x-y| \sum_{\ell=0}^{m} (|x|^m + |y|^m) = m|x-y|(|x|^m + |y|^m).$$

Using this inequality gives

$$\begin{aligned}
|(R_{1,2}^{r,r'})^{m+1} - (R_{1,2}^{-,r,r'})^{m+1}| &\leq m |R_{1,2}^{r,r'} - R_{1,2}^{-,r,r'}| \left(|R_{1,2}^{r,r'}|^m + |R_{1,2}^{-,r,r'}|^m \right) \\
&= \frac{m}{N} \left| \sum_{i=1}^{N} \sigma_i^1(r) \sigma_i^2(r') - \sum_{i=1}^{N-1} \sigma_i^1(r) \sigma_i^2(r') \right| \left(|R_{1,2}^{r,r'}|^m + |R_{1,2}^{-,r,r'}|^m \right) \\
&= \frac{m}{N} |\sigma_N^1(r) \sigma_N^2(r')| \left(|R_{1,2}^{r,r'}|^m + |R_{1,2}^{-,r,r'}|^m \right) \\
&\leq \frac{mM^2}{N} \left(|R_{1,2}^{r,r'}|^m + |R_{1,2}^{-,r,r'}|^m \right).
\end{aligned}$$

□

Lemma 10. Assume that there exists some $K \geq 1$ such that

$$\nu_{\bar{\beta}}((R_{1,2}^{r,r'})^{2j}) \leq \frac{K}{N^j}$$

for any $0 \leq j \leq m$. Then

$$\nu_{\bar{\beta}}((R_{1,2}^{-,r,r'})^{2m}) \leq \frac{2^{2m} M^{4m} K}{N^m}.$$

Proof. Write

$$R_{1,2}^{-,r,r'} = R_{1,2}^{r,r'} - \frac{1}{N} \varepsilon_1^r \varepsilon_2^{r'}.$$

Using the binomial theorem to expand $(R^{-,r,r'}_{1,2})^m$ gives

$$\nu_{\bar\beta}\left((R^{-,r,r'}_{1,2})^{2m}\right) \le \sum_{j=0}^{2m}\binom{2m}{j}\frac{1}{N^{2m-j}}\nu_{\bar\beta}\left(|R^{r,r'}_{1,2}|^j|\varepsilon_1^r\varepsilon_2^r|^{2m-j}\right)$$
$$\le \sum_{j=0}^{2m}\binom{2m}{j}\left(\frac{M^2}{N}\right)^{2m-j}\nu_{\bar\beta}(|R^{r,r'}_{1,2}|^j). \tag{4.14}$$

For any $0 \le j \le 2m$, choose j_1, j_2 such that $0 \le j_1, j_2 \le m$ and $j_1 + j_2 = j$. Then, by the Cauchy-Schwarz inequality,

$$\nu_{\bar\beta}(|R^{r,r}_{1,2}|^j) = \nu_{\bar\beta}(|R^{r,r}_{1,2}|^{j_1}|R^{r,r}_{1,2}|^{j_2}) \le \nu_{\bar\beta}(|R^{r,r}_{1,2}|^{2j_1}|)^{1/2}\nu_{\bar\beta}(|R^{r,r}_{1,2}|^{2j_2}|)^{1/2} \le \frac{K}{N^{j_1/2+j_2/2}} = \frac{K}{N^{j/2}}.$$

Substituting this in 4.14 gives the bound

$$\nu_{\bar\beta}\left((R^{-,r,r'}_{1,2})^{2m}\right) \le \sum_{j=0}^{2m}\binom{2m}{j}\left(\frac{M^2}{N}\right)^{2m-j}\frac{K}{N^{j/2}}$$
$$= K\left(\frac{M^2}{N}+\frac{1}{\sqrt{N}}\right)^{2m}$$
$$\le K\left(\frac{M^2+1}{\sqrt{N}}\right)^{2m}$$
$$\le \frac{K2^{2m}M^{4m}}{N^m}.$$

The last inequality used the fact that $M \ge 1$. □

4.4.2 Cavity Argument

We now turn to the cavity argument to bound the even moments of the overlaps $R^{r,r}_{1,2}$.

Lemma 11. Let m be a non-negative integer and $\bar\beta \in (0,\infty)^k$. Assume that there exists a constant $K_0 \ge 1$ such that

$$\max_{1\le r\le k}\nu_{\bar\beta}((R^{r,r}_{1,2})^{2j}) \le \frac{K_0}{N^j}$$

for all $0 \leq j \leq m$ and all $N \geq 1$. Then, for all $N \geq 1$,

$$\max_{1\leq r\leq k} \nu_{\bar{\beta}}((R_{1,2}^{r,r})^{2(m+1)}) \leq K_1(\bar{\beta}) \max_{1\leq r\leq k} \nu_{\bar{\beta}}(|R_{1,2}^{r,r}|^{2m+3}) + \frac{K_2(\bar{\beta})}{N^{m+1}}$$

where $K_1(\bar{\beta}), K_2(\bar{\beta})$ are two non-negative continuous functions of $\bar{\beta}$ that are independent of N. In addition, $K_1(\bar{\beta}) \leq K_1(\bar{\beta}')$ whenever $\beta_r \leq \beta_r'$ for all $1 \leq r \leq k$. Furthermore, $K_1(\bar{\beta}) = 0$ if and only if $\bar{\beta} = 0$.

The proof proceeds in four main steps. First $\nu_{\bar{\beta}}((R_{1,2}^{r,r})^{2m+2})$ is bounded by $\nu_{\bar{\beta}}((R_{1,2}^{-,r,r})^{2m+2})$ and a remainder of the appropriate order $O(N^{-m})$. In the second and third steps, $\nu_{\bar{\beta}}((R_{1,2}^{-,r,r})^{2m+2})$ is bounded by the derivative $\nu_{\bar{\beta},t}'((R_{1,2}^{r,r})^{2m+2})$ which is controlled by Lemmas 7 and 8. Finally step 4 follows the same pattern as step 1.

Proof. **Step 1:** By symmetry between sites, write

$$\nu_{\bar{\beta}}((R_{1,2}^{r,r})^{2m+2}) = \nu_{\bar{\beta}}(\varepsilon_1^r \varepsilon_2^r (R_{1,2}^{r,r})^{2m+1}).$$

Next, set

$$\mathcal{E} = \nu_{\bar{\beta}}(\varepsilon_1^r \varepsilon_2^r ((R_{1,2}^{r,r})^{2m+1} - (R_{1,2}^{-,r,r})^{2m+1}))$$

so that

$$\nu_{\bar{\beta}}((R_{1,2}^{r,r})^{2m+2}) = \nu_{\bar{\beta}}(\varepsilon_1^r \varepsilon_2^r (R_{1,2}^{-,r,r})^{2m+1}) + \mathcal{E}.$$

Applying the triangle inequality and then Lemma 9 controls $|\mathcal{E}|$ by

$$|\mathcal{E}| \leq M^2 \nu_{\bar{\beta}}(|(R_{1,2}^{r,r})^{2m+1} - (R_{1,2}^{-,r,r})^{2m+1}|) \leq \frac{2mM^4}{N}(\nu_{\bar{\beta}}(|R_{1,2}^{r,r}|^{2m}) + \nu_{\bar{\beta}}(|R_{1,2}^{-,r,r}|^{2m})). \tag{4.15}$$

By the assumptions,

$$\nu_{\bar{\beta}}(|R_{1,2}^{r,r}|^{2j}) \leq \frac{K_0}{N^j} \tag{4.16}$$

for all $1 \leq j \leq m$, so by Lemma 10,

$$\nu_{\bar{\beta}}(|R_{1,2}^{-,r,r}|^{2m}) \leq \frac{2^{2m} M^{4m} K_0}{N^m}. \tag{4.17}$$

Combining the bounds of equations (4.16) and (4.17) and plugging in to equation 4.15 gives

$$|\mathcal{E}| \leq \frac{C_1}{N^{m+1}},$$

where $C_1 = 2mM^4K_0(1 + 2^{2m}M^{4m})$. So far we have

$$\nu_{\bar{\beta}}((R_{1,2}^{r,r})^{2m+2}) \leq \nu_{\bar{\beta}}(\varepsilon_1^r\varepsilon_2^r(R_{1,2}^{-,r,r})^{2m+1}) + \frac{C_1}{N^{m+1}} \tag{4.18}$$

Step 2: For each $1 \leq r \leq k$, define $f_r = \varepsilon_1^s\varepsilon_2^r(R_{1,2}^{-,r,r})^{2m+1}$ so that equation (4.18) reads

$$\nu_{\bar{\beta}}((R_{1,2}^{r,r})^{2m+2}) \leq \nu_{\bar{\beta}}(f_r) + \frac{C_1}{N^{m+1}}.$$

Since $H_{N,\bar{\beta},0}(\sigma)$ does not depend on σ_N,

$$\nu_{\bar{\beta},0}(f_r) = \nu_{\bar{\beta},0}(\varepsilon_1^r\varepsilon_2^{r'})\nu_{\bar{\beta},0}(R_{1,2}^{-,r,r}) = 0.$$

It is easy to see that $\nu_{\bar{\beta},0}(\varepsilon_1^r\varepsilon_2^{r'}) = 0$ since spins $\sigma_N(r)$ and $\sigma_N(r')$ are independent under the Gibbs measure corresponding to $H_{N,\bar{\beta},0}$.

By the mean value theorem, there exists $c \in (0,1)$ such that

$$\nu'_{\bar{\beta},c}(f_r) = \nu_{\bar{\beta}}(f_r) - \nu_{\bar{\beta},0}(f_r) = \nu_{\bar{\beta}}(f_r).$$

Thus

$$\nu_{\bar{\beta}}((R_{1,2}^{r,r})^{2m+2}) \leq \sup_{0\leq t\leq 1} |\nu'_{\bar{\beta},t}(f_r)| + \frac{C_1}{N^{m+1}}. \tag{4.19}$$

Since f_r is a function of two replicas $\bar{\sigma}^1, \bar{\sigma}^2$, applying Lemma 7 with $f = f_r$ and $n = 2$

bounds the derivative $\nu'_{\bar\beta,t}(f_r)$ by

$$|\nu'_{\bar\beta,t}(f_r)| \leq \sum_{1\leq s,s'\leq k} \beta_s\beta_{s'} \sum_{j=1}^{p} \binom{p}{j} \frac{(M^2)^{j+1}}{N^{j-1}} \Big(\nu_{\bar\beta,t}(|R^{-,r,r}_{1,2}|^{2m+1}|R^{-,s,s'}_{1,2}|^{p-j}) + 2\nu_{\bar\beta,t}(|R^{-,r,r}_{1,2}|^{2m+1}|R^{-,s,s'}_{1,3}|^{p-j}) + 2\nu_{\bar\beta,t}(|R^{-,r,r}_{1,2}|^{2m+1}|R^{-,s,s'}_{2,3}|^{p-j}) + 3\nu_{\bar\beta,t}(|R^{-,r,r}_{1,2}|^{2m+1}|R^{-,s,s'}_{3,4}|^{p-j})\Big). \quad (4.20)$$

For $1 \leq j \leq p$ set

$$\tau_j^1 = \frac{2m+1+p-j}{2m+1} \quad \text{and} \quad \tau_j^2 = \frac{\tau_j^1}{\tau_j^1 - 1}.$$

Note that $1/\tau_j^1 + 1/\tau_j^2 = 1$. Thus, applying Hölder's inequality to equation (4.20) yields

$$\begin{aligned} |\nu'_{\bar\beta,t}(f_r)| \leq \sum_{1\leq s,s'\leq k} \beta_s\beta_{s'} \sum_{j=1}^{p} \binom{p}{j} \frac{(M^2)^{j+1}}{N^{j-1}} &\Big(\nu_{\bar\beta,t}(|R^{-,r.r}_{1,2}|^{2m+1+p-j}|)^{1/\tau_j^1}\nu_{\bar\beta,t}(R^{-,s.s'}_{1,2}|^{2m+1+p-j})^{1/\tau_j^2} \\ &+ 2\nu_{\bar\beta,t}(|R^{-,r,r}_{1,2}|^{2m+1+p-j}|)^{1/\tau_j^1}\nu_{\bar\beta,t}(R^{-,s,s'}_{1,3}|^{2m+1+p-j})^{1/\tau_j^2} \\ &+ 2\nu_{\bar\beta,t}(|R^{-,r,r}_{1,2}|^{2m+1+p-j})^{1/\tau_j^1}\nu_{\bar\beta,t}(|R^{-,s,s'}_{2,3}|^{2m+1p-j})^{1/\tau_j^2} \\ &+ 3\nu_{\bar\beta,t}(|R^{-,r,r}_{1,2}|^{2m+1+p-j})^{1/\tau_j^1}\nu_{\bar\beta,t}(|R^{-,s,s'}_{3,4}|^{2m+1+p-j})^{1/\tau_j^2}\Big). \\ = 8\sum_{1\leq s,s'\leq k} \beta_s\beta_{s'} \sum_{j=1}^{p} \binom{p}{j} \frac{(M^2)^{j+1}}{N^{j-1}} &\nu_{\bar\beta,t}(|R^{-,r,r}_{1,2}|^{2m+1+p-j})^{1/\tau_j^1}\nu_{\bar\beta,t}(|R^{-,s,s'}_{1,2}|^{2m+1+p-j})^{1/\tau_j^2}. \end{aligned}$$

Applying the Cauchy-Schwarz inequality to the terms involving cross-overlaps $R^{-,s,s'}_{\ell,\ell'}$ with $s \neq s'$ gives

$$\nu_{\bar\beta,t}(|R^{s,s'}_{\ell,\ell'}|^{2m+1+p-j})^{1/\tau_j^2} \leq \nu_{\bar\beta,t}(|R^{-,s,s}_{1,2}|^{2m+1+p-j})^{1/2\tau_j^2}\nu_{\bar\beta,t}(|R^{-,s',s'}_{1,2}|^{2m+1+p-j})^{1/\tau_2 j^2}.$$

Plugging this in to the previous expression and using Lemma 8 to bound $\nu_{\bar\beta,t}(\cdot)$ by $\nu_{\bar\beta}(\cdot)$

gives

$$|\nu'_{\bar\beta,t}(f_r)| \le C(\bar\beta) \sum_{1\le s,s'\le k} \beta_s\beta_{s'} \sum_{j=1}^{p} \binom{p}{j} \frac{(M^2)^{j+1}}{N^{j-1}} \Big(\nu_{\bar\beta}(|R^{-,r,r}_{1,2}|^{2m+1+p-j})^{1/\tau^1_j}$$
$$\times \nu_{\bar\beta}(|R^{-,s,s}_{1,2}|^{2m+1+p-j})^{1/2\tau^2_j}$$
$$\times \nu_{\bar\beta}(|R^{-,s',s'}_{1,2}|^{2m+1+p-j})^{1/2\tau^2_j}\Big), \quad (4.21)$$

where

$$C(\bar\beta) = 8\exp\left(2^{p+3}M^{2p}\sum_{1\le r,r'\le k}\beta_r\beta_{r'}\right).$$

Step 3: In this step, the goal is to reduce the power $2m+1+p-j$ of the overlaps to the power $2m+p$, which is accomplished by considering two different cases: the case when $j=1$ and the case when $j>1$. When $j=1$, the terms of the sum are

$$\nu_{\bar\beta}(|R^{-,r,r}_{1,2}|^{2m+p})^{1/\tau^1_j}\nu_{\bar\beta}(|R^{-,s,s}_{1,2}|^{2m+p})^{1/2\tau^2_j}\nu_{\bar\beta}(|R^{-,s',s'}_{1,2}|^{2m+p})^{1/2\tau^2_j},$$

so the overlaps already have the desired power and no work is required.

When $j>1$, pulling out $1+p-j$ powers of the overlap and using the fact that $R^{-,r,r}_{1,2} \le M^2$ shows that the terms of the sum are bounded above by

$$(M^2)^{1+p-j}\nu_{\bar\beta}(|R^{-,r,r}_{1,2}|^{2m})^{1/\tau^1_j}\nu_{\bar\beta}(|R^{-,s,s}_{1,2}|^{2m})^{1/2\tau^2_j}\nu_{\bar\beta}(|R^{-,s',s'}_{1,2}|^{2m})^{1/2\tau^2_j}.$$

The assumptions of Lemma 10 are satisfied, so there exists a constant K_0 such that

$$(M^2)^{1+p-j}\nu_{\bar\beta}(|R^{-,r,r}_{1,2}|^{2m})^{1/\tau^1_j}\nu_{\bar\beta}(|R^{-,s,s}_{1,2}|^{2m})^{1/2\tau^2_j}\nu_{\bar\beta}(|R^{-,s',s'}_{1,2}|^{2m})^{1/2\tau^2_j} \le (M^2)^{1+p-j}\frac{2^{2m}M^{4m}K_0}{N^m}.$$

The sum over all terms $j>1$ in equation (4.21) is upper bounded by

$$\sum_{j=2}^{p}\binom{p}{j}\frac{(M^2)^{j+1}}{N^{j-1}}\frac{2^{2m}M^{4m}K_0}{N^m} \le \frac{2^{2m+p}(M^2)^{2m+p+1}K_0}{N^{m+1}}. \quad (4.22)$$

Plugging the bound in equation (4.22) into (4.21) and then plugging this back into

4.19 finally gives the bound

$$\nu_{\bar\beta}((R_{1,2}^{r,r})^{2m+2}) \le C(\bar\beta)C_2 \sum_{1\le s,s'\le k} \beta_s\beta_{s'}\nu_{\bar\beta}(|R_{1,2}^{-,r,r}|^{2m+p})^{1/\tau_j^1}\nu_{\bar\beta}(|R_{1,2}^{-,s,s}|^{2m+p})^{1/2\tau_j^2}\nu_{\bar\beta}(|R_{1,2}^{-,s',s'}|^{2m+p})^{1/2\tau_j^2} + \frac{C(\bar\beta)C_3}{N^{m+1}}, \tag{4.23}$$

where

$$C_2 := pM^4 \quad \text{and} \quad C_3 := 2^{2m+p}(M^2)^{2m+p+1}K_0 \sum_{1\le r,r'\le k}\beta_r\beta_{r'} + C_1.$$

Step 4: Similar to Step 1, Lemmas 9 and 10 relate the overlaps $R_{1,2}^{-,r,r}$ back to overlaps $R_{1,2}^{r,r}$. First

$$\nu_{\bar\beta}(|R_{1,2}^{-,r,r}|^{2m+p})^{1/\tau_j^1}\nu_{\bar\beta}(|R_{1,2}^{-,s,s}|^{2m+p})^{1/2\tau_j^2}\nu_{\bar\beta}(|R_{1,2}^{-,s',s'}|^{2m+p})^{1/2\tau_j^2} \le (M^2)^{p-3}\max_{1\le r\le k}\nu_{\bar\beta}(|R_{1,2}^{-,r,r}|^{2m+3}).$$

As in Step 1, add and subtract $\nu_{\bar\beta}(|R_{1,2}^{r,r}|^{2m+3})$ so that, by Lemma 9,

$$\begin{aligned}\max_{1\le r\le k}\nu_{\bar\beta}(|R_{1,2}^{-,r,r}|^{2m+3}) &= \max_{1\le r\le k}\nu_{\bar\beta}(|(R_{1,2}^{-,r,r})^{2m+3} + (R_{1,2}^{r,r})^{2m+3} - (R_{1,2}^{r,r})^{2m+3}|) \\ &\le \max_{1\le r\le k}\left(\nu_{\bar\beta}(|R_{1,2}^{r,r}|^{2m+3}) + \frac{2m+3}{N}M^2\left(\nu_{\bar\beta}(|R_{1,2}^{-,r,r}|^{2m+2}) + \nu_{\bar\beta}(|R_{1,2}^{r,r}|^{2m+2})\right)\right) \\ &\le \max_{1\le r\le k}\nu_{\bar\beta}(|R_{1,2}^{r,r}|^{2m+3}) + \frac{2m+3}{N}(M^2)^3\left(\max_{1\le r\le k}\nu_{\bar\beta}(|R_{1,2}^{-,r,r}|^{2m}) + \max_{1\le r\le k}(|R_{1,2}^{r,r}|^{2m})\right).\end{aligned} \tag{4.24}$$

By the assumptions,

$$\max_{1\le r\le k}\nu\bar\beta(|R_{1,2}^{r,r}|^{2m}) \le \frac{K_0}{N^m}. \tag{4.25}$$

By Lemma 10,

$$\max_{1\le r\le k}\nu\bar\beta(|R_{1,2}^{r,r}|^{2m}) \le 2^{2m}M^{4m}K_0/N^m. \tag{4.26}$$

Plugging equations (4.25) and (4.26) into equation (4.24) gives

$$\max_{1\le r\le k} \nu_{\bar\beta}(|R_{1,2}^{-,r,r}|^{2m+3}) \le \max_{1\le r\le k} \nu_{\bar\beta}(|R_{1,2}^{r,r}|^{2m+3})\frac{C_4}{N^{m+1}}.$$

where

$$C_4 = (2m+3)M^6K_0(1+2^{2m}M^{4m}).$$

Set

$$K_1(\bar\beta) = C(\bar\beta)C_2 \sum_{1\le s,s'\le k} \beta_s\beta_{s'} \quad \text{and} \quad K_2(\bar\beta) = C(\bar\beta)C_3 + C_4.$$

Finally, plugging back into equation 4.23 gives

$$\nu_{\bar\beta}((R_{1,2}^{r,r})^{2m+3} \le K_1(\bar\beta) \max_{1\le r\le k} \nu_{\bar\beta}(|R_{1,2}^{r,r}|^{2m+3}) + \frac{K_2(\bar\beta)}{N^{m+1}}.$$

□

4.5 Proof of Theorem 5 and Proposition 2

With the bound of the cavity method in hand, it is now possible to prove Theorem 5 and Proposition 2. Recall that Theorem 5 states that if $0 < \beta < \beta_c$, then there exists a constant $K > 0$ independent of N such that

$$\mathbb{E}\langle |R(\sigma^1,\sigma^2)|^{2m}\rangle_s \le \frac{K}{N^m}$$

for all $s \in [0,\beta]$ and all $N \ge 1$. Proposition 2 uses Theorem 5 to bound the fluctuations of the free energy $F_N(\beta)$. For more compact notation, define $R_{1,2} = R(\sigma^1,\sigma^2)$.

Proof. (Theorem 5) Fix $0 < \beta < \beta_c$. The proof is by induction on m. When $m = 0$, for all $s \in [0,1]$,

$$\nu_{s\beta}(|R_{1,2}|^0) = 1 = \frac{1}{N^0}.$$

Assume that for some $m \ge 0$ there exists a constant $K > 0$ such that

$$\nu_{s\beta}(|R_{1,2}|^{2m}) \le \frac{K}{N^m}$$

for all $s \in [0,\beta]$ and all $N \geq 1$. From Lemma 11 there exist non-negative continuous functions $K_1(s\beta), K_2(s\beta)$ such that

$$\nu_{s\beta}(|R_{1,2}|^{2m+2}) \leq K_1(s\beta)\nu_{s\beta}(|R_{1,2}|^{2m+3}) + \frac{K_2(s\beta)}{N^{m+1}}.$$

The function $K_1(s\beta)$ is non-decreasing in s and $K_1(0) = 0$. Define

$$s_0 = \sup\left\{s \in [0,1] \mid K_1(s\beta)M^2 \leq \frac{1}{2}\right\}.$$

Note that the set of s for which $K_1(s\beta)M^2 \leq 1/2$ is an interval as $K_1(s\beta)$ is increasing in s. The proof is split into two cases.

Case 1: Assume $s \in [0, s_0]$. Then, observing that $|R_{1,2}| \leq M^2$,

$$\begin{aligned}\nu_{s\beta}(|R_{1,2}|^{2m+2}) &\leq K_1(s\beta)\nu_{s\beta}(|R_{1,2}|^{2m+3}) + \frac{K_2(s\beta)}{N^{m+1}} \\ &\leq K_1(s_0\beta)M^2\nu_{s\beta}(|R_{1,2}|^{2m+2}) + \frac{K_2(s\beta)}{N^{m+1}} \\ &\leq \frac{1}{2}\nu_{s\beta}(|R_{1,2}|^{2m+2}) + \frac{K_2(s\beta)}{N^{m+1}}.\end{aligned}$$

Rearranging,

$$\nu_{s\beta}(|R_{1,2}|^{2m+2}) \leq \frac{2K_2(s\beta)}{N^{m+1}}.$$

Case 2: Assume $s \in (s_0, 1]$. Choose $\varepsilon > 0$ so that $\varepsilon \max_{s\in[s_0,1]} K_1(s\beta) < \frac{1}{2}$. This is valid since $K_1(s\beta)$ is continuous on the compact set $[s_0, 1]$. By Proposition 6, for any $\varepsilon > 0$ there exists $K' > 0$ such that

$$\nu_{s\beta}\left(I(|R_{1,2}| \geq \varepsilon)\right) \leq K'e^{-N/K'}$$

for all $s \in [s_0, 1]$ and all $N \geq 1$. Note that K' is independent of s. Next,

$$\begin{aligned}\nu_{s\beta}(|R_{1,2}|^{2m+3}) &= \nu_{s\beta}(|R_{1,2}|^{2m+3}I(|R_{1,2}| > \varepsilon)) + \nu_{s\beta}(|R_{1,2}|^{2m+3}I(|R_{1,2}| \leq \varepsilon)) \\ &\leq M^{2(2m+3)}\nu_{s\beta}(I(|R_{1,2}| > \varepsilon)) + \varepsilon\nu_{s\beta}(|R_{1,2}|^{2m+2}I(|R_{1,2}| \leq \varepsilon)) \\ &\leq M^{2(2m+3)}K'e^{-N/K'} + \varepsilon\nu_{s\beta}(|R_{1,2}|^{2m+2}).\end{aligned}$$

Therefore

$$\begin{aligned}\nu_{s\beta}(|R_{1,2}|^{2m+2}) &\leq K_1(s\beta)\nu_{s\beta}(|R_{1,2}^{2m+3}) + \frac{K_2(s\beta)}{N^{m+1}}\\ &\leq K_1(s\beta)\left(M^{2(2m+3)}K'e^{-N/K'} + \varepsilon\nu_{s\beta}(|R_{1,2}|^{2m+2})\right) + \frac{K_2(s\beta)}{N^{m+1}}\\ &\leq \frac{1}{2}\nu_{s\beta}(|R_{1,2}|^{2m+2}) + K_1(s\beta)M^{2(2m+3)}K'e^{-N/K'} + \frac{2K_2(s\beta)}{N^{m+1}}.\end{aligned}$$

Rearranging,

$$\begin{aligned}\nu_{s\beta}(|R_{1,2}^{2m+2}) &\leq 2K_1(s\beta)M^{2(2m+3)}K'e^{-N/K'} + \frac{K_2(s\beta)}{N^{m+1}}\\ &\leq \frac{2K_1(s\beta)M^{2(2m+3)} + 2K_2(s\beta)}{N^{m+1}}.\end{aligned}$$

The second inequality holds for large enough N. Setting

$$K = \sup_{s\in[0,s_0]} 2K_2(s\beta) + \sup_{s\in[s_0,1]} (2K_1(s\beta)M^{2(2m+3)} + 2K_2(s\beta))$$

gives

$$\nu_{s\beta}(|R_{1,2}|^{2m+2}) \leq \frac{K}{N^{m+1}}$$

for all $N \geq 1$ and all $s \in [0,1]$ as desired. □

Chebyshev's inequality and the Gaussian Poincarè inequality relate the quantity $P(|F_N(\beta)| \geq \ell)$ to the quantity $\nu_\beta(R_{1,2}^p)$. This, together with the bound of Theorem 5, controls the free energy fluctuations as desired for Proposition 2.

Proof. (Proposition 2) For any $\ell > 0$, Chebyshev's inequality gives

$$P\left(|F_N(\beta)| \geq \ell\right) \leq \frac{\mathbb{E}F_N(\beta)^2}{\ell^2} = \frac{1}{\ell^2}\left(\mathrm{Var}(F_N(\beta)) + (\mathbb{E}F_N(\beta))^2\right).$$

Since

$$\partial_{g_{i_1,\ldots,i_p}} F_N(\beta) = \frac{\beta}{N^{(p+1)/2}}\langle\sigma_{i_1}\cdots\sigma_{i_p}\rangle_\beta,$$

the Gaussian Poincaré inequality gives

$$\mathrm{Var}(F_N(\beta)) \leq \frac{\beta^2}{N^{p+1}}\mathbb{E}\sum_{1\leq i_1,\ldots,i_p}\langle\sigma_{i_1}\cdots\sigma_{i_p}\rangle_\beta^2 = \frac{\beta^2}{N}\mathbb{E}\langle R(\sigma^1,\sigma^2)^p\rangle_\beta.$$

Also, Gaussian integration by parts gives

$$\frac{\partial}{\partial\beta}\mathbb{E}F_N(\beta) = -\beta\mathbb{E}\langle R(\sigma^1,\sigma^2)^p\rangle_\beta,$$

so

$$\begin{aligned}\mathbb{E}F_N(\beta) &= \mathbb{E}F_N(0) + \int_0^\beta \mathbb{E}F_N(s)ds\\ &= -\int_0^\beta s\mathbb{E}\langle R(\sigma^1,\sigma^2)^p\rangle_s ds\end{aligned}$$

Thus

$$(\mathbb{E}F_N(\beta))^2 = \left(\int_0^\beta s\mathbb{E}\langle R(\sigma^1,\sigma^2)^p\rangle_s ds\right)^2.$$

Combining, we so far have

$$P(|F_N(\beta)|\geq\ell) \leq \frac{\beta^2}{\ell^2 N}\mathbb{E}\langle R(\sigma^1,\sigma^2)^p\rangle_\beta + \frac{1}{\ell^2}\left(\int_0^\beta s\mathbb{E}\langle R(\sigma^1,\sigma^2)^p\rangle_s ds\right)^2.$$

Finally, from Theorem 5 there exists a constant $K > 0$ such that

$$\begin{aligned}P(|F_N(\beta)|\geq\ell) &\leq \frac{\beta^2 K}{\ell^2 N^{p/2}N} + \frac{1}{\ell^2}\left(\int_0^\beta s\frac{K}{N^{p/2}}ds\right)^2\\ &= \frac{\beta^2 K}{\ell^2 N^{p/2}N} + \frac{\beta^4K^2}{4N^p\ell^2}.\\ &\leq \frac{\beta^2}{\ell^2 N^{(p+2)/2}}(1+\beta^2 K).\end{aligned}$$

□

4.6 Proof of Theorem 6

The most interesting characteristic of the detection problem is that additional spikes do not add additional noise in the spiked tensor model. This results directly from the structure of high-temperature regime of the vector-valued model $H_{N,\bar{\beta}}(\bar{\sigma})$. The high-temperature regime of the vector-valued model is simply the product of the high-temperature regimes of the marginal scalar-valued systems $H_{N,\beta}(\sigma(r))$, as stated in Theorem 6. To prove that the high-temperature regime has this structure, we need concentration of the total overlap, as stated below in Proposition 7.

Recall that $v_{*,r} = \int a^2 \mu_r(da)$, and define a diagonal matrix $V_* \in \mathbb{R}^{k\times k}$ with entries $(V_*)_{r,r} = v_{*,r}$. For any $\varepsilon > 0$ and $V \in \mathbb{R}^{k\times k}$

$$A_\varepsilon(V) = \{V' \in \mathbb{R}^{k\times k} \mid \|V - V'\|_{\max} < \varepsilon\}.$$

For any $\bar{\sigma} \in \bar{\Lambda}$, define a matrix $R(\bar{\sigma}) \in \mathbb{R}^{k\times k}$ with entries $(R(\bar{\sigma}))_{r,r'} = R(\sigma(r), \sigma(r'))$. Concentration of the total overlap means that $R(\sigma)$ is close to V_* with high probability, as stated formally below.

Proposition 7. Assume that $\bar{\beta} \in \bar{\mathcal{R}}$. Let $\bar{\sigma}$ be sampled from $G_{N,\bar{\beta}}$. Then for any $\varepsilon > 0$ there exist positive constants K, δ such that for any $N \geq 1$,

$$\mathbb{E}\langle I(R(\bar{\sigma}) \notin A_\varepsilon(V_*))\rangle_{\bar{\beta}} \leq Ke^{-N/K}. \tag{4.27}$$

Also, with probability at least $1 - Ke^{-N/K}$,

$$F_N(\bar{\beta}, A_\varepsilon(V_*)^c) \leq F_N(\bar{\beta}) - \delta. \tag{4.28}$$

Proof. Suppose that equation 4.28 hold holds with probability $1 - Ke^{-N/K}$ for some $K > 0$. That is,

$$\frac{1}{N}\log\int_{A_\varepsilon(V_*)^c} e^{H_{N,\bar{\beta}}(\bar{\sigma})}\mu(d\bar{\sigma}) \leq \frac{1}{N}\log\int e^{H_{N,\bar{\beta}}(\bar{\sigma})}\mu(d\bar{\sigma}) - \delta.$$

Multiplying by N, combining the logarithms and exponentiating both sides gives

$$\langle I(R(\bar{\sigma}) \notin A_\varepsilon(V_*))\rangle_{\bar{\beta}} = \frac{\int_{A_\varepsilon(V_*)^c} e^{H_{N,\bar{\beta}}(\bar{\sigma})}\mu(d\bar{\sigma})}{\int e^{H_{N,\bar{\beta}}(\bar{\sigma})}\mu(d\bar{\sigma})} < e^{-N\delta}.$$

Thus equation (4.27) holds. We therefore need only prove equation 4.28.

For $1 \le r, r' \le k$, define

$$A_\varepsilon(r, r') = \{\bar{\sigma} \mid |R(\sigma(r), \sigma(r')) - (V_*)_{r,r'}| > \varepsilon.$$

Notice that

$$A_\varepsilon(V_*)^c \subset \bigcup_{r,r'} A_\varepsilon(r, r').$$

Thus

$$\begin{aligned}\limsup_{N\to\infty} \mathbb{E}F_N(\bar{\beta}, A_\varepsilon(V_*)^c) &\le \limsup_{N\to\infty} \mathbb{E}\frac{1}{N}\log\sum_{r,r'}\int_{A_\varepsilon(r,r')} e^{H_{N,\bar{\beta}}(\bar{\sigma})}\bar{\mu}^{\otimes N}(\bar{\sigma}) \\ &\le \limsup_{N\to\infty} \frac{1}{N}\int_{A_\varepsilon(V_*)^c} \bar{\mu}^{\otimes N}(d\bar{\sigma}). \qquad (4.29)\end{aligned}$$

The second inequality results from applying Jenson's inequality to $E\log(x)$.

Since the coordinates of $\sigma(r)$ are i.i.d. with distribution μ_r, and $\sigma(r)$ is independent of $\sigma(r')$ for any $r \ne r'$, each set $A_\varepsilon(r, r')$ is bounded away from $\int ab\mu_r(da)\mu_{r'}(db)$. Also, for every r, the set Λ_r is bounded so the sets $A_\varepsilon(r, r')$ are bounded as well. Thus, by a stronger version of Cramér's Theorem for large deviations, see for example the proof of [31, Theorem 2.2.3], there exists a positive constant δ such that

$$\sum_{1\le r,r'\le k}\int_{\bar{\sigma}\in A_\varepsilon(r,r')} \mu^{\otimes N}(d\bar{\sigma}) \le e^{-N\delta}.$$

Therefore there exists $\delta > 0$ such that

$$\limsup_{N\to\infty} \mathbb{E}F_N(\bar{\beta}, A_\varepsilon(V_*)^c) \le F(\bar{\beta}) - \delta.$$

From this, the result follows by applying the Gaussian concentration of measure inequality to both free energies of the above inequality.

□

We now use the overlap concentration established above to prove Theorem 6, the structure of the high temperature regime for the vector-valued model.

Proof. (Theorem 6): First we show that $\bar{\mathcal{R}} \subseteq (0, \beta_{1,c}] \times \cdots \times (0, \beta_{k,c}]$. Suppose that $\bar{\beta} = (\beta_1, \ldots, \beta_k) \in \bar{\mathcal{R}}$. Then $F(\bar{\beta}) = 0$ by definition of $\bar{\mathcal{R}}$. If $R(\bar{\sigma}) \in A_\varepsilon(V_*)$, then for $r \neq r'$,

$$|R(\sigma(r), \sigma(r'))^p - (V_*^{r,r'})^p| = |R(\sigma(r), \sigma(r'))^p| \leq \varepsilon^p.$$

From Proposition 7, for any $\varepsilon > 0$ there exists a constant $K > 0$ such that for all $N \geq 1$,

$$\mathbb{E}\langle I(R(\bar{\sigma}) \in A_\varepsilon(V^*)\rangle \geq 1 - Ke^{-N/K} \tag{4.30}$$

Using Gaussian concentration of measure and an argument similar to the beginning of the proof of Proposition 7, it is possible to show that the equation (4.31) implies that

$$\limsup_{N\to\infty} F_N(\beta, A_\varepsilon(V_*)) \geq \limsup_{N\to\infty} F_N(\bar{\beta}) = 0.$$

By Jensen's inequality,

$$\limsup_{N\to\infty} F_N(\beta, A_\varepsilon(V_*)) \leq 0,$$

therefore

$$\limsup_{N\to\infty} F_N(\bar{\beta}, A_\varepsilon(V_*)) = \limsup_{N\to\infty} F_N(\bar{\beta}).$$

As a result, for $\bar{\sigma} \in A_\varepsilon(V_*)$ we can substitute ε^p for the overlap terms $R(\sigma(r), \sigma(r'))$ with $r \neq r'$ in the Hamiltonian $H_{N,\bar{\beta}}(\bar{\sigma})$ to get

$$\begin{aligned}\limsup_{N\to\infty} F_N(\bar{\beta}) &\leq \limsup_{N\to\infty} \frac{1}{N} \log \int_{A_\varepsilon(V_*)} \exp\left(\sum_{r=1}^k H_{N,\beta_r}(\sigma(r))\right) \bar{\mu}^{\otimes N} - \frac{\varepsilon^p}{2} \sum_{r\neq r'} \beta_r \beta_{r'} \\ &\leq \limsup_{N\to\infty} \sum_{r=1}^k F_{N,r}(\beta_r) - \frac{\varepsilon^p}{2} \sum_{r\neq r'} \beta_r \beta_{r'}.\end{aligned}$$

Taking $\varepsilon \to 0$ gives

$$0 = F(\bar{\beta}) \leq \sum_{r=1}^{k} F_r(\beta_r).$$

Since $F_r(\beta_r) \leq 0$ for all $1 \leq r \leq k$, it must be that $F_r(\beta_r) = 0$ for all $1 \leq r \leq k$. Therefore $\beta_r \in (0, \beta_{r,c}]$, so $\bar{\mathcal{R}} \in (0, \beta_{1,c}] \times \cdots \times (0, \beta_{k,c}]$.

We next prove that $(0, \beta_{1,c}] \times \cdots \times (0, \beta_{k,c}] \subset \bar{\mathcal{R}}$ by contradiction, and distinguish between two cases.

Case 1: Suppose $\bar{\beta} \in (0, \beta_{1,c}) \times \cdots \times (0, \beta_{k,c})$, but $\bar{\beta} \notin \bar{\mathcal{R}}$. This means that $F(\beta_r) = 0$ for $1 \leq r \leq k$, but $F(\bar{\beta}) < 0$. Thus there exists some $\eta > 0$ such that $F(\bar{\beta}) < -\eta$, so for large enough N, $\mathbb{E}F_N(\bar{\beta}) < -\eta$. By Gaussian concentration of measure, there exists a constant $K > 0$ such that

$$F_N(\bar{\beta}, A_\varepsilon(V_*)) \leq F_N(\bar{\beta}) < -\frac{\eta}{2} \tag{4.31}$$

with probability at least $1 - Ke^{-N/K}$.

For any $\varepsilon > 0$,

$$F_N(\bar{\beta}, A_\varepsilon(V_*)) \leq F_N(\bar{\beta}) < -\frac{\eta}{2}$$

for all $N \geq 1$, where the first inequality is by the definitions of the two free energies. By definition of the set $A_\varepsilon(V_*)$,

$$F_N(\bar{\beta}, A_\varepsilon(V_*)) \geq \frac{1}{N} \log \int_{A_\varepsilon(V_*)} \exp\left(\sum_{r=1}^{k} H_{N,\beta_r}(\sigma(r))\right) \bar{\mu}^{\otimes N}(d\bar{\sigma}) - \frac{\varepsilon^p}{2} \sum_{r \neq r'} \beta_r \beta_{r'}. \tag{4.32}$$

Combining equations (4.31) and (4.32), we have

$$\frac{1}{N} \log \int_{A_\varepsilon(V_*)} \exp\left(\sum_{r=1}^{k} H_{N,\beta_r}(\sigma(r))\right) \bar{\mu}^{\otimes N}(d\bar{\sigma}) < \frac{\varepsilon^p}{2} \sum_{r \neq r'} \beta_r \beta_{r'} - \frac{\eta}{2}.$$

Choose ε small enough so that $\frac{\varepsilon^p}{2} \sum_{r \neq r'} \beta_r \beta_{r'} < \frac{\eta}{4}$. Then

$$\frac{1}{N} \log \int_{A_\varepsilon(V_*)} \exp\left(\sum_{r=1}^{k} H_{N,\beta_r}(\sigma(r))\right) \bar{\mu}^{\otimes N}(d\bar{\sigma}) < -\frac{\eta}{4}.$$

Since it is assumed that $F(\beta_r) = 0$ for all $1 \leq r \leq k$, the above is equivalent to

$$\frac{1}{N} \log \int_{A_\varepsilon(V_*)} \exp\left(\sum_{r=1}^{k} H_{N,\beta_r}(\sigma(r))\right) \bar{\mu}^{\otimes N}(d\bar{\sigma}) < \sum_{r=1}^{k} F_N(\beta_r) - \frac{\eta}{4}.$$

Let $\langle\cdot\rangle'$ denoe the Gibbs average with respect to the product measure $\prod_{r=1}^{k} G_{N,\beta_r}(d\sigma(r))$. An argument identical to the proof of Proposition 7 gives the existence of a constant $K > 0$ such that

$$E\langle I(R(\bar{\sigma}) \in A_\varepsilon(V_*))\rangle' \leq Ke^{-N/K}. \tag{4.33}$$

For a sample $\bar{\sigma}$ from $\prod_{r=1}^{k} G_{N,\beta_r}(\sigma(r))$, the spins $\sigma(1), \ldots, \sigma(k)$ are independent of one another, so when $k = 1$, Proposition 7 gives

$$\lim_{N\to\infty} \mathbb{E}\langle I(|R(\sigma(r), \sigma(r)) - v_{r,*}| \leq \varepsilon)\rangle' = 1. \tag{4.34}$$

Also, by Proposition 6, for any $\varepsilon > 0$,

$$\lim_{N\to\infty} \mathbb{E}\langle I(R(\sigma^1(r), \sigma^2(r))| \leq \varepsilon)\rangle' = 1. \tag{4.35}$$

For $r \neq r'$, using the independence of $\sigma(1), \ldots, \sigma(k)$ gives

$$\begin{aligned}
\mathbb{E}\langle R(\sigma(r), \sigma(r'))^2\rangle' &= \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbb{E}\langle \sigma_i(r)\sigma_j(r)\sigma_i(r')\sigma_j(r')\rangle' \\
&= \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbb{E}\langle \sigma_i(r)\sigma_j(r)\rangle\langle \sigma_i(r')\sigma_j(r')\rangle'.
\end{aligned}$$

Applying Cauchy-Schwarz then gives

$$\begin{aligned}
\mathbb{E}\langle R(\sigma(r), \sigma(r'))^2\rangle' &\leq \left(\frac{1}{N^2} \sum_{i,j} \mathbb{E}(\langle \sigma_i(r)\sigma_j(r)\rangle')^2\right)^{1/2} \left(\frac{1}{N^2} \sum_{i,j} \mathbb{E}(\langle \sigma_i(r')\sigma_j(r')\rangle')^2\right)^{1/2} \\
&= \left(\mathbb{E}(\langle R(\sigma^1(r), \sigma^2(r))\rangle')^2\right)^{1/2} \left(\mathbb{E}(\langle R(\sigma^1(r'), \sigma^2(r'))\rangle')^2\right)^{1/2}.
\end{aligned}$$

Therefore, by equation 4.35,

$$\lim_{N\to\infty} \mathbb{E}\langle R(\sigma(r),\sigma(r'))^2\rangle' = 0.$$

This and 4.34 give

$$\lim_{N\to\infty} \mathbb{E}\langle I(R(\bar\sigma) \in A_\varepsilon(V_*))\rangle' = 1,$$

In other words, the total overlap $R(\bar\sigma)$ is close to the optimal matrix V_* which contradicts equation 4.33.

Case 2: Assume $\bar\beta =\in (0,\beta_{1,c}] \times \cdots \times (0,\beta_{k,c}]$ and $\bar\beta \notin (0,\beta_{1,c}) \times \cdots \times (0,\beta_{k,c})$. In other words, $\beta_r = \beta_{r,c}$ for at least one value $1 \le r \le k$. Each of the free energies $F_1(\beta_1),\ldots,F_k(\beta_k)$ are continuous functions of $\beta_1,\ldots,\beta_k$ respectively. Thus $F(\bar\beta)$ can be approximated by $F(\bar\beta')$ for $\bar\beta' \in (0,\beta_{1,c}) \times \cdots \times (0,\beta_{k,c})$. Then from Case I the conclusion holds. □

4.7 Proof of Theorem 7 and Proposition 3

The proof of Theorem 7 is nearly identical to the proof of Theorem 5, so most of the details are omitted. Originally, the result of Theorem 7 was used in the proof of detection in the case of multiple spikes, but the proof technique has changed from the original version rendering Theorem 7 unnecessary; however, it is still included in this thesis as an interesting result in the study of the vector-valued p-spin model.

Proof. (Theorem 7): By Lemma 11, if there exists a constant $K_0 \ge 1$ such that

$$\max_{1\le r\le k} \nu_{\bar\beta}((R^{r,r}_{1,2})^{2j}) \le \frac{K_0}{N^j}$$

for all $0 \le j \le m$ and all $N \ge 1$ then

$$\max_{1\le r\le k} \nu_{\bar\beta}((R^{r,r}_{1,2})^{2m+2}) \le K_1(\bar\beta) \max_{1\le r\le k} \nu_{\bar\beta}(|R^{r,r}_{1,2}|^{2m+3}) + \frac{K_2(\bar\beta)}{N^{m+1}}$$

for all $N \ge 1$. The proof of Theorem 7 follows exactly the same steps as the proof of Theorem 5. □

Proof. (Proposition 3) Using the Chebyshev inequality, Gaussian Poincaré inequality and Gaussian integration by parts as in the proof of Proposition 2 yields

$$P(|F_N(\bar{\beta})| \geq \ell) \leq \frac{2}{\ell^2 N} \sum_{1 \leq r, r' \leq k} \beta_r \beta_{r'} \mathbb{E}\langle |R_{1,2}^{r,r'}|^p \rangle_{\bar{\beta}} + \frac{2}{\ell^2} \left(\sum_{1 \leq r, r' \leq k} \beta_r \beta_{r'} \int_0^1 s \mathbb{E}\langle |R_{1,2}^{r,r,}|^p \rangle \right)^2 .$$

Applying the result of Theorem 7 gives the desired result. □

References

[1] Wei-Kuo Chen, Madeline Handschy, and Gilad Lerman. Phase transition in random tensors with multiple spikes. *arXiv preprint arXiv:1809.06790*, 2018.

[2] V. A Marchenko and L. A. Pastur. Distribution of eigenvalues for some sets of random matrices. *Mathematics of the USSR-Sbornik*, 1(4):457, 1967.

[3] Jinho Baik, Gerard Ben Arous, Sandrine Peche, et al. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(1643-1697), 2005.

[4] J. Baik and J. Silverstein. Eigenvalues of large sample covariance matrices of spiked population models. *Journal of Multivariate Analysis*, 97(6):1382–1408, 2006.

[5] D. Paul. Asymptotics of the leading sample eigenvalues for a spiked covariance model. *Statistica Sinica*, 17(4):1617–1642, 2007.

[6] F. Benaych-Georges and R. R. Nadakuditi. The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Advances in Mathematics*, 227(1):494–521, 2011.

[7] F. Benaych-Georges and R. R. Nadakuditi. The singular values and vectors of low rank perturbations of large rectangular random matrices. *Journal of Multivariate Analysis*, 111:120–135, 2012.

[8] Amelia Perry, Alexander S Wein, and Afonso Bandeira. Statistical limits of spiked tensor models. *arXiv preprint arXiv:1612.07728*, 2016.

[9] Wei-Kuo Chen. Phase transition in the spiked random tensor with rademacher prior. *arXiv preprint arXiv:1712.01777*, 2017.

[10] Thibault Lesieur, Leo Miolane, Marc Lelarge, Florent Krzakala, and Lenka Zdeborova. Statistical and computational phase transitions in spiked tensor estimation. In *Information Theory (ISIT), 2017 IEEE International Symposium on*, pages 511–515. IEEE, 2017.

[11] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. Asymptotic mutual information for the binary stochastic block model. In *2016 IEEE International Symposium on Information Theory (ISIT)*, pages 185–189. IEEE, 2016.

[12] Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, pages 1–71, 2017.

[13] Yash Deshpande and Andrea Montanari. Information-theoretically optimal sparse pca. In *2014 IEEE International Symposium on Information Theory*, pages 2197–2201. IEEE, 2014.

[14] Andrea Montanari and Ramji Venkataramanan. Estimation of low-rank matrices via approximate message passing. *arXiv preprint arXiv:1711.01682*, 2017.

[15] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(7):073403, 2017.

[16] Jonathan Kadmon and Surya Ganguli. Statistical mechanics of low-rank tensor decomposition. In *Advances in Neural Information Processing Systems*, pages 8201–8212, 2018.

[17] Philip W Anderson. Spin glass i: A scaling law rescued. *Physics Today*, 41(1):9–11, 1988.

[18] J Owen, M Browne, WD Knight, and Ch Kittel. Electron and nuclear spin resonance and magnetic susceptibility experiments on dilute alloys of mn in cu. *Physical Review*, 102(6):1501, 1956.

[19] JRL De Almeida and David J Thouless. Stability of the sherrington-kirkpatrick solution of a spin glass model. *Journal of Physics A: Mathematical and General*, 11(5):983, 1978.

[20] JA Mydosh. Spin glasses and mictomagnets. In *AIP Conference Proceedings*, volume 24, pages 131–137. AIP, 1975.

[21] Samuel Frederick Edwards and Phil W. Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965, 1975.

[22] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(26):1792, 1975.

[23] Michel Talagrand. *Mean field models for spin glasses: Volume I: Basic examples*, volume 54. Springer Science and Business Media, 2010.

[24] Dmitry Panchenko. *The Sherrington-Kirkpatrick Model*. Springer-Verlag New York, 2013.

[25] Dmitry Panchenko. The sherrington-kirkpatrick model: an overview. *Journal of Statistical Physics*, 149(2):362–383, 2012.

[26] Dmitry Panchenko. Free energy in the mixed p-spin models with vector spins. 12 2015, 1512.04441.

[27] Michel Talagrand. The parisi formula. *Annals of Mathematics*, 221-263, 2006.

[28] Dmitry Panchenko. Free energy in the generalized sherrington-kirkpatrick mean field model. *Reviews in Mathematical Physics*, 17(07):793–857, 2005.

[29] Aukosh Jagannath and Ian Tobasco. A dynamic programming approach to the parisi functional. *Proceedings of the American Mathematical Society*, 144(7):3135–3150, 2016.

[30] Michel Talagrand. *Mean Field Models for Spin Glasses: Advanced replica-symmetry and low temperature*. Springer, 2011.

[31] Amir Dembo and Ofer Zeitouni. Large deviations techniques and applications. 1998. *Applications of Mathematics*, 38, 2011.

[32] Wei-Kuo Chen. Variational representations for the parisi functional and the two-dimensional guerra-talagrand bound. 01 2015, 1501.06635.

Appendix A

Appendix

A.1 Gaussian Integration by Parts

Gaussian integration by parts is used repeatedly in this thesis and the study of spin glasses in general. Take $g \sim \mathcal{N}(0, \sigma^2)$ and $F(x)$ any differentiable function that satisfies

$$\lim_{|x| \to \infty} F(x) e^{-x^2/2\sigma^2} = 0.$$

The most basic version of Gaussian integration by parts, see, for example [23, Appendix A.4] states that

$$\mathbb{E} g F(g) = \mathbb{E} g^2 \mathbb{E} F(g).$$

If $g, z_1, \ldots, z_n$ are Gaussians and $F(x) : \mathbb{R}^n \to \mathbb{R}$ is differentiable and satisfies

$$\lim_{\|x\| \to \infty} |F(x)| e^{-a\|x\|^2} = 0$$

for all $a > 0$, there is a multivariate Gaussian integration by parts [23] that states

$$\mathbb{E} g F(z_1, \ldots, z_n) = \sum_{\ell \leq n} \mathbb{E}(g z_\ell) \mathbb{E} \frac{\partial F}{\partial x_\ell}(z_1, \ldots, z_n).$$

The version used in this thesis extends the multi-variate version to families of Gaussians that are not necessarily finite. If $\mathbf{g} = (g(\rho))_{\rho \in U}$ is a Gaussian process indexed by $U \in \mathbb{R}^N$

and $F(\mathbf{g})$ is a differentiable function of $\mathbb{R}^U$ then given $\sigma \in U$, by [28, Lemma 4],

$$\mathbb{E}g(\sigma)F(\mathbf{g}) = \mathbb{E}\frac{\partial F}{\partial \mathbf{g}}[\mathbb{E}g(\sigma)g(\rho)],$$

where the right-hand side denotes the expectation of the variational derivative of F in the direction of $\mathbb{E}g(\sigma)g(\rho)$.

A.2 Gaussian Concentration of Measure

Gaussian concentration of measure is used repeatedly to show that a free energy and the expectation of that free energy are close with high probability. The key result is [23, Proposition 1.3.5] which states that if F is a Lipschitz function on $\mathbb{R}^M$ with Lipschitz constant A, and $g = (g_1, \ldots, g_M)$ with $g_1, \ldots, g_M$ i.i.d. standard Gaussians, then

$$P(|F(g) - \mathbb{E}F(g)| \geq t) \leq 2e^{-t^2/4A^2}. \tag{A.1}$$

For any $A \subseteq \bar{\Lambda}$, taking the gradient of $F_N(\bar{\beta}, A)$ in the distinct random variables $g_{i_1,\ldots,i_p}$ with $1 \leq i_1 < \cdots < i_p \leq N$, there exists a constant $K > 0$ such that

$$\|\nabla F_N(\bar{\beta}, A)\|_2^2 \leq \frac{K}{N}.$$

Plugging this in to A.1 gives

$$P(|F_N(\bar{\beta}, A) - \mathbb{E}F_N(\bar{\beta}, A)| \geq t) \leq 2e^{-t^2 N/4K}.$$

Thus there exist $\eta > 0$ and $K_0 > 0$ such that

$$P(|F_N(\bar{\beta}, A) - \mathbb{E}F_N(\bar{\beta}, A)| \geq \eta) \leq K_0 e^{-N/K_0},$$

which is the result used throughout this thesis.

A.3 Cole-Hopf Transformation

The Cole-Hopf transform is a transform used to solve Burgers equation

$$u_t + uu_x = \kappa u_{xx}. \tag{A.2}$$

Taking the derivative in x of equation (A.2) and setting $U_x = u$ implies U satisfies the Hamilton-Jacobi equation

$$U_t + \frac{(U_t)^2}{2} = \kappa U_{xx}, \tag{A.3}$$

which is similar to the Parisi PDE. If ϕ satisfies the heat equation $\phi_t = \kappa\phi_{xx}$, the Cole-Hopf transform

$$B(x,t) = -2\kappa \log[\phi(x,t)]$$

solves equation (A.3).

Recall that $\Phi_{\beta,v,\alpha}(s,x,\lambda) : [0,v] \times \mathbb{R} \times \mathbb{R}$ is the weak solution to the PDE

$$\partial_s \Phi_{\beta,v,\alpha} = -\frac{\beta^2 \xi''}{2}(\partial_{xx}\Phi_{\beta,v,\alpha} + \alpha(s)(\partial_x \Phi_{\beta,v,\alpha})^2)$$

with boundary condition

$$\Phi_{\beta,v,\alpha}(v,x,\lambda) = \log \int e^{xa+\lambda a^2} \mu(da).$$

When $\alpha(s)$ is the CDF of a finitely supported probability measure, the Cole-Hopf transformation can be used to solve the Parisi PDE as shown in the following adaptation of [32, Lemma 3].

Suppose $0 \le a < b \le v$. Let A be a smooth function on $\mathbb{R}$ with $\limsup_{x\to\infty} |A(x)|/|x| < \infty$. Suppose $\alpha(s) = m$ on the interval $[a,b]$. Let $z(s)$ be a Gaussian random variable with covariance

$$\mathbb{E}z(s)^2 = \xi''(b) - \xi''(s).$$

Then

$$\Phi_{\beta,v,\alpha}(s,x,\lambda) = \frac{1}{m}\log \mathbb{E} \exp mA(x + \beta z(s))$$

satisfies the Parisi PDE on the interval $[a, b]$. Note that the boundary condition of the Parisi PDE satisfies the growth requirement, so the preceding result can be used to solve the Parisi PDE backward from v.