

PHASE TRANSITION IN RANDOM TENSORS WITH MULTIPLE INDEPENDENT SPIKES

BY WEI-KUO CHEN^{*}, MADELINE HANDSCHY[†] AND GILAD LERMAN[‡]

School of Mathematics, University of Minnesota, ^{}wkchen@umn.edu; [†]hands014@umn.edu; [‡]lerman@umn.edu*

Consider a spiked random tensor obtained as a mixture of two components: noise in the form of a symmetric Gaussian p -tensor for $p \geq 3$ and signal in the form of a symmetric low-rank random tensor. The latter is defined as a linear combination of k independent symmetric rank-one random tensors, referred to as spikes, with weights referred to as signal-to-noise ratios (SNRs). The entries of the vectors that determine the spikes are i.i.d. sampled from general probability distributions supported on bounded subsets of $\mathbb{R}$. This work focuses on the problem of detecting the presence of these spikes, and establishes the phase transition of this detection problem for any fixed $k \geq 1$. In particular, it shows that for a set of relatively low SNRs it is impossible to distinguish between the spiked and nonspiked Gaussian tensors. Furthermore, in the interior of the complement of this set, where at least one of the k SNRs is relatively high, these two tensors are distinguishable by the likelihood ratio test. In addition, when the total number of low-rank components, k , of the p -tensor of size N grows in the order $o(N^{(p-2)/4})$ as N tends to infinity, the problem exhibits an analogous phase transition. This theory for spike detection is also shown to imply that recovery of the spikes by the minimum mean square error exhibits the same phase transition. The main methods used in this work arise from the study of mean field spin glass models, where the phase transition thresholds are identified as the critical inverse temperatures distinguishing the high and low-temperature regimes of the free energies. In particular, our result formulates the first full characterization of the high temperature regime for vector-valued spin glass models with independent coordinates.

1. Introduction. This work studies the detection and recovery of a low-rank component in a particular random tensor and characterizes their corresponding phase transitions. In order to motivate this problem, we first discuss a simpler and widely-studied question: When can principal component analysis (PCA) detect and recover low-rank linear structures in noisy data? While detection only requires determining the presence or absence of low-rank structure, the task of recovery aims to reveal the concealed low-rank structure.

One common setting for addressing this question assumes data points $y_1, \dots, y_L \in \mathbb{R}^N$ drawn independently from the multivariate normal distribution $\mathcal{N}(0, I + \beta uu^T)$, where I is the N -dimensional identity matrix, which generates spherically symmetric Gaussian noise, u is a unit column vector in $\mathbb{R}^N$, which generates a rank-one signal, and $\beta > 0$ is the signal-to-noise ratio (SNR). Under this model, the observations y_i , $i = 1, \dots, L$, take the form $y_i = x_i + \varepsilon_i$, where x_i is proportional to the signal u with signal-to-noise ratio β , and ε_i is the Gaussian noise. The question is then whether or not it is possible to apply PCA to detect the presence of the signal u when given the data points $y_1, \dots, y_L$ with different choices of the SNR parameter β . The earlier result for this problem traces back to the fundamental work of Johnstone [35].

Received July 2019; revised July 2020.

MSC2020 subject classifications. 62F05, 62F10, 82D30, 82B26.

Key words and phrases. BBP transition, Parisi formula, principal component analysis, spin glass, spiked matrix, spiked tensor.

Assume that $N/L \rightarrow \gamma < 1$ as $L \rightarrow \infty$. When $\beta = 0$, the Marchenko–Pastur distribution [39] describes the limiting distribution of the eigenvalues of the sample covariance matrix. The well-known Baik–Ben Arous–Péché phase transition [3, 4] states that when $\beta \leq \sqrt{\gamma}$, the eigenvalues of this matrix still follow the Marchenko–Pastur distribution and thus detection of the low-rank sample is impossible by PCA. In contrast, when $\beta > \sqrt{\gamma}$, the largest eigenvalue of this matrix stays away from the typical location of the Marchenko–Pastur distribution and PCA can detect the presence of the signal. This phase transition of spike detection is extended in Paul [53] to spike recovery by PCA. More precisely, [53] shows that when $\beta > \sqrt{\gamma}$, there is a nontrivial asymptotic correlation between the top eigenvector of the sample covariance and u and thus one can approximately recover u by PCA. Moreover, when $\beta \leq \sqrt{\gamma}$ this asymptotic correlation is zero and PCA cannot recover u . Extension of detection and recovery to the case where $\gamma \geq 1$ is also established in [53].

Another common setting for studying the detection problem using PCA assumes a random matrix of the form $T = W + \beta N^{-1/2} uu^T$, where W is an $N \times N$ Gaussian Wigner matrix¹ and u is an N -dimensional random vector with i.i.d. entries sampled from a bounded distribution on $\mathbb{R}$. The parameter β is the SNR. We refer to the rank one component, uu^T as a spike and to T as a spiked random matrix. The problem is to detect the presence of the spike in T , or equivalently, to distinguish between T and W . This detection problem exhibits a phase transition similar to that of the previous setting, see Féral–Péché [28], Péché [54], and Benaych–Georges–Nadakuditi [15, 16]. When the SNR is below a certain critical threshold, the eigenvalue distribution of T follows Wigner’s semi-circle law and it is thus impossible to distinguish between T and W . Once the value of β exceeds this critical threshold, the largest eigenvalue jumps away from the typical location of the Wigner semi-circle law and the top eigenvector nontrivially correlates with the signal. Consequently, in this case, one can detect and approximately recover the signal by PCA. Recent studies of phase transitions in detection and recovery of low-rank signals in random matrices include Lelarge–Miolane [37], Miolane [42, 43], Montanari–Reichman–Zeitouni [44], Montanari–Richard [44, 46], Onatski–Moreira–Hallim [48], and Perry–Wein–Bandeira–Moitra [55].

The latter setting of low-rank detection in spiked random matrices has a natural higher-order generalization to spiked random tensors. This generalization considers the spiked symmetric random p -tensor

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

The first component, W , is the symmetric Gaussian p -tensor of size N^p , formally defined in Section 2.1. The second component is the signal, which is a linear combination of the spikes $u(1)^{\otimes p}, \dots, u(k)^{\otimes p}$. Here, $u(1), \dots, u(k)$ are N -dimensional vectors whose entries are i.i.d. sampled from probability measures $\mu_1, \dots, \mu_k$ supported on bounded subsets of the real line. We refer to $\hat{\beta} = (\beta_1, \dots, \beta_k)$ as the vector of SNRs. The detection problem under this setting asks whether identification of the low-rank signal $\sum_{r=1}^k \beta_r u(r)^{\otimes p}$ in the tensor T_k is possible for a given vector $\bar{\beta}$. The recovery problem seeks to recover, if possible, the low-rank signal for given values of $\bar{\beta}$. Answering these question of whether the spike is detectable or recoverable requires characterization of the phase transitions in $\bar{\beta}$ of the detection and recovery problems.

We remark that the generalized tensor setting is significantly more challenging than the above setting of detecting and recovering rank-one structure in matrices. The former setting involved the best rank-one approximation by PCA. However, for tensors, basic relevant

¹ W is a symmetric matrix with independent $W_{ij} \sim \mathcal{N}(0, 1/2)$ for $1 \leq i < j \leq N$ and $W_{ii} \sim \mathcal{N}(0, 1)$ for $1 \leq i \leq N$.

notion, such as rank and best low-rank approximation, are not obvious, see Kolda–Bader [36]. Furthermore, many common algorithms for computing these and related notion are NP-hard, see Hillar–Lim [30]. In this work, we study low-rank tensor detection and recovery by common theoretical tests and estimators, which are hard to compute. We leave the analysis of tractable procedures to future work. Following Chen [20], El Alaoui–Krzakala–Jordan [27], Montanari–Reichman–Zeitouni [44], Montanari–Richard [45], and Perry–Wein–Bandeira [56], we say that spike detection is impossible if the total variation distance between W and T_k vanishes when N tends to infinity. In other words, any statistical test fails to distinguish W and T_k (see Section 2.1). On the other hand, we say that detection is possible if this distance is one in the limit. This means that asymptotically one can find a statistical test, in particular, the likelihood ratio test, that distinguishes between W and T_k (see Section 2.1). For recovery, we follow [38] and use the minimum mean square error (MMSE) and its corresponding estimator.

Many recent works, which are reviewed in Section 2.5, have studied detection and recovery under the spiked random tensor model. Nevertheless, the optimal phase transition for low-rank detection in spiked random tensors has not yet been established. This paper aims to close this gap. Our main result states that there exist critical thresholds $\beta_{1,c}, \dots, \beta_{k,c}$ and a set of the form $\bar{\mathcal{R}} = (0, \beta_{1,c}] \times \dots \times (0, \beta_{k,c}]$ such that detection is impossible if $\bar{\beta} = (\beta_1, \dots, \beta_k)$ lies strictly in the interior of the set $\bar{\mathcal{R}}$. Furthermore, it is possible to detect the spike via the likelihood ratio test when $\bar{\beta} \notin \bar{\mathcal{R}}$. In other words, detection is possible only when at least one of $\beta_1, \dots, \beta_k$ exceeds its critical threshold; whereas, if $\beta_1, \dots, \beta_k$ are all smaller than their critical thresholds, one cannot detect the spike. Our result also allows the total number of spikes to grow with N . In particular, if $\mu_1 = \dots = \mu_k$ and $k = o(N^{(p-2)/4})$, then similar statements hold. A byproduct of these developments is a new proof for a recent result on the recovery problem by Lesieur–Miolane–Lelarge–Krzakala–Zdeborová [38] when assuming the same setting of the present paper. In essence, their result states that $\beta_{1,c}, \dots, \beta_{k,c}$ are the critical thresholds for the MMSE recovery problem.

Our approach is based on methodologies from the study of mean-field spin glass models. Roughly speaking, spin glasses are spin systems that exhibit both quenched disorder and frustration. That is, the interactions between sites are disordered and spin constraints cannot be simultaneously satisfied. These two features are commonly shared by many problems that involve randomized combinatorial optimization, see Mézard–Montanari [40] and Montanari–Sen [47]. The book, Mézard–Parisi–Virasoro [41], reviews the area of spin glasses from the point of view of physicists, whereas mathematical treatments of the subject appear in Talagrand [60, 61] and Panchenko [50].

Mean-field spin glasses are related to the detection problem by the following key observation: The total variation distance between W and T_k can be represented as an integral of the distribution function of the so-called free energy of the re-centered pure p -spin model with vector-valued spin configurations (see (3.2) and Lemma 4.2 below). From this observation, to study the detection problem, we need to understand the full high-temperature regime of this model and investigate a sharp upper bound on the fluctuation of the free energy for all values of the SNR vector $\bar{\beta}$. Our results reveal that $\bar{\mathcal{R}}$ is indeed the high-temperature regime of the free energy and its fluctuation is up to the order $N^{-(p/2+1)}$ when $\bar{\beta}$ lies in the interior of $\bar{\mathcal{R}}$ and is of order 1 when $\bar{\beta}$ lies in the complement of $\bar{\mathcal{R}}$. These allow us to completely characterize the phase transition of the detection problem. We mention that while our study of the high-temperature behavior of the free energy is mainly used to derive results for spike detection, it is also of independent interest in the field of spin glasses. Indeed, our result (see Theorem 3.2) gives the first full characterization of the high temperature regime for vector-valued spin glass models with independent coordinates.

2. Main results. This section states the main results of this paper and provides the necessary mathematical background. Additionally, it reviews prior results and describes the structure of the rest of the paper, particularly the structure of the proofs of Theorems 2.1—2.4. Section 2.1 defines the necessary terminology, especially, the distinguishability of two random tensors. Section 2.2 describes our main results for the detection problem in the case of a single spike. In particular, it introduces an auxiliary function that characterizes the high-temperature regime and allows one to simulate the critical SNR. Using this function, we demonstrate numerical simulations of the critical SNR for the sparse Rademacher prior. Section 2.3 states our main results for the detection problem in the case of multiple spikes. Section 2.4 mentions a result for recovery by MMSE that is later obtained from our results for spike detection. Section 2.5 surveys recent related results. Finally, Section 2.7 describes the organization of the proofs of the main results.

2.1. Settings and definitions. Let $p \geq 2$ be an integer. For any integer $N \geq 1$, denote by Ω_N the set of all real-valued p -tensors $Y = (Y_{i_1, \dots, i_p})_{1 \leq i_1, \dots, i_p \leq N}$ equipped with the Borel σ -field. The inner product of two p -tensors is

$$\langle Y, Y' \rangle = \sum_{1 \leq i_1, \dots, i_p \leq N} Y_{i_1, \dots, i_p} Y'_{i_1, \dots, i_p}.$$

Given a vector $u = (u_1, \dots, u_N) \in \mathbb{R}^N$, we form a rank-one p -tensor using the outer product by

$$(u^{\otimes p})_{i_1, \dots, i_p} = u_{i_1} \cdots u_{i_p} \quad \forall 1 \leq i_1, \dots, i_p \leq N.$$

Given $Y \in \Omega_N$ and a permutation π of the set $\{1, 2, \dots, p\}$, define Y^π by

$$Y_{i_1, \dots, i_p}^\pi = Y_{\pi(i_1), \dots, \pi(i_p)}.$$

A p -tensor is said to be symmetric if $Y_{i_1, \dots, i_p}^\pi = Y_{i_1, \dots, i_p}$ for all corresponding indices and permutations. Throughout the rest of the paper, we assume that Y is a random p -tensor and all entries in Y are i.i.d. standard Gaussian. The symmetric Gaussian p -tensor of size N^p is obtained by the averaging over all permutations in the symmetric group of N letters:

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

In the case $p = 2$, W is the Gaussian–Wigner matrix.

Next, we define the notion of distinguishability and indistinguishability between two random p -tensors in terms of the total variation distance. For any two random p -tensors U, V , denote by $d_{\text{TV}}(U, V)$ the total variation distance between U and V , that is,

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

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

DEFINITION 2.1. Let U_N, V_N be two sequences of random p -tensors. We say that they are distinguishable if

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

and are indistinguishable if

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

Distinguishability of U_N and V_N means that there exists a sequence of measurable subsets A_N of Ω_N such that $\lim_{N \rightarrow \infty} \mathbb{P}(U_N \in A_N) = 1$ and $\lim_{N \rightarrow \infty} \mathbb{P}(V_N \in A_N) = 0$. From this, if we consider a statistical test $S_N : \Omega_N \rightarrow \{0, 1\}$ defined by $S_N(w) = 0$ for $w \in A_N$ and $S_N(w) = 1$ for $w \notin A_N$, then as N approaches infinity, the sum of type I and type II errors approaches zero:

$$(2.1) \quad \begin{aligned} & \lim_{N \rightarrow \infty} (\mathbb{P}(S_N(U_N) = 1) + \mathbb{P}(S_N(V_N) = 0)) \\ &= \lim_{N \rightarrow \infty} (\mathbb{P}(U_N \notin A_N) + \mathbb{P}(V_N \in A_N)) = 0. \end{aligned}$$

This means that one can statistically distinguish U_N and V_N by the test S_N . Furthermore, if U_N and V_N have nonvanishing densities f_{U_N} and f_{V_N} , the well-known formula

$$d_{\text{TV}}(U_N, V_N) = \int_{f_{U_N} \geq f_{V_N}} (f_{U_N} - f_{V_N}) dw$$

implies that

$$d_{\text{TV}}(U_N, V_N) = \mathbb{P}(U_N \in A_N) - \mathbb{P}(V_N \in A_N)$$

for

$$A_N := \left\{ w \in \Omega_N \mid \frac{f_{U_N}(w)}{f_{V_N}(w)} \geq 1 \right\}.$$

Therefore, one can naturally use the likelihood ratio test to distinguish U_N and V_N . In contrast, when U_N and V_N are indistinguishable, any statistical test is powerless as in this case the total error approaches one as N tends to infinity.

REMARK 2.1. The setting of spiked matrices (see, e.g., [27, 62]) considers a weaker notion of distinguishability that requires the limiting total error, which appears on the left hand side of (2.1), to be less than 1. In our spiked tensor model with $p > 2$, the limiting total error converges to either zero or one and this weaker notion of distinguishability coincides with ours.

2.2. Main results for detection of a single spike. Let Λ be a bounded subset of $\mathbb{R}$ and μ be a probability measure on the Borel σ -field of Λ . Assume that $u_1, \dots, u_N$ are i.i.d. samplings from μ and are independent of W . Denote $u = (u_1, \dots, u_N)$. We refer to the random variable u as the prior. Consider the spiked random p -tensor T defined by

$$(2.2) \quad T = W + \frac{\beta}{N^{(p-1)/2}} u^{\otimes p}.$$

We say that detection of the spike $u^{\otimes p}$ in T is possible if W and T are distinguishable and detection is impossible if they are indistinguishable in the sense of Definition 2.1. Note that if $\int a \mu(da) \neq 0$, one can immediately detect the spike by noting that $Y_{i_1, \dots, i_p}$ are i.i.d. standard Gaussian and using the strong law of large number. Indeed,

$$\frac{1}{N^{(p+1)/2}} \sum_{i_1, \dots, i_p=1}^N W_{i_1, \dots, i_p} = \frac{1}{N^{(p+1)/2}} \sum_{i_1, \dots, i_p=1}^N Y_{i_1, \dots, i_p} \rightarrow 0,$$

while

$$\begin{aligned} \frac{1}{N^{(p+1)/2}} \sum_{i_1, \dots, i_p=1}^N T_{i_1, \dots, i_p} &= \frac{1}{N^{(p+1)/2}} \sum_{i_1, \dots, i_p=1}^N Y_{i_1, \dots, i_p} + \beta \left(\frac{\sum_{i=1}^N u_i}{N} \right)^p \\ &\rightarrow \beta \left(\int a \mu(da) \right)^p. \end{aligned}$$

We can thus restrict our discussion to the case when μ is centered, that is, when $\int_{\mathbb{R}} a \mu(da) = 0$. Our first result on spike detection is formulated as follows.

THEOREM 2.1. *Assume that μ is centered. 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 other words, β_c is the critical threshold that describes the phase transition of the detection problem. As we explained in Section 2.1, when detection is possible, one can use the likelihood ratio test, which uses the ratio of densities $f_T(w)/f_W(w)$, to distinguish between W and T . In Lemma 4.2 below, we relate this ratio to the free energy of the pure p -spin mean field spin glass model.

The precise value of β_c can be determined as follows. Let

$$\xi(s) = s^p$$

and

$$(2.3) \quad v_* = \int a^2 \mu(da).$$

For $a \in \mathbb{R}$ and $t > 0$, consider the geometric Brownian motion

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

where B_t is a standard Brownian motion. For $b \geq 0$, define an auxiliary function $\Gamma_b(v)$ on $[0, \infty)$ by

$$(2.4) \quad \Gamma_b(v) = \int_0^v \xi''(s)(\gamma_b(s) - s) ds,$$

where for $s \geq 0$,

$$(2.5) \quad \gamma_b(s) := \mathbb{E}\left[\frac{(\int a Z(a, b^2 \xi'(s)) \mu(da))^2}{\int Z(a, b^2 \xi'(s)) \mu(da)}\right].$$

The critical value β_c in Theorem 2.1 can be calculated as follows:

THEOREM 2.2. *If $p \geq 3$ and μ is centered, then β_c is the largest b such that $\sup_{v \in (0, v_*]} \Gamma_b(v) = 0$.*

As an example of Theorem 2.2, we demonstrate numerical simulations for estimating the critical threshold β_c for the sparse Rademacher prior, in which the entries $u_1, \dots, u_N$ in u are i.i.d. sampled from the probability distribution

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

with parameter $\rho \in (0, 1]$ that controls the sparsity of the prior. The case $\rho = 1$ corresponds to the usual Rademacher prior. If $\rho < 1$, the sparse Rademacher prior can be regarded as first uniformly sampling approximately ρN of the coordinates and then for these coordinates, sampling Bernoulli $\pm 1/\sqrt{\rho}$ random variables with equal probability. The remaining approximately $(1 - \rho)N$ coordinates are set to zero. From this construction, the second moment of $\|u\|/\sqrt{N}$ is of order 1. To simulate β_c according to the value established in Theorem 2.2, we numerically evaluate $\Gamma_b(v)$ for test values of v with increments .001 in the interval between 0

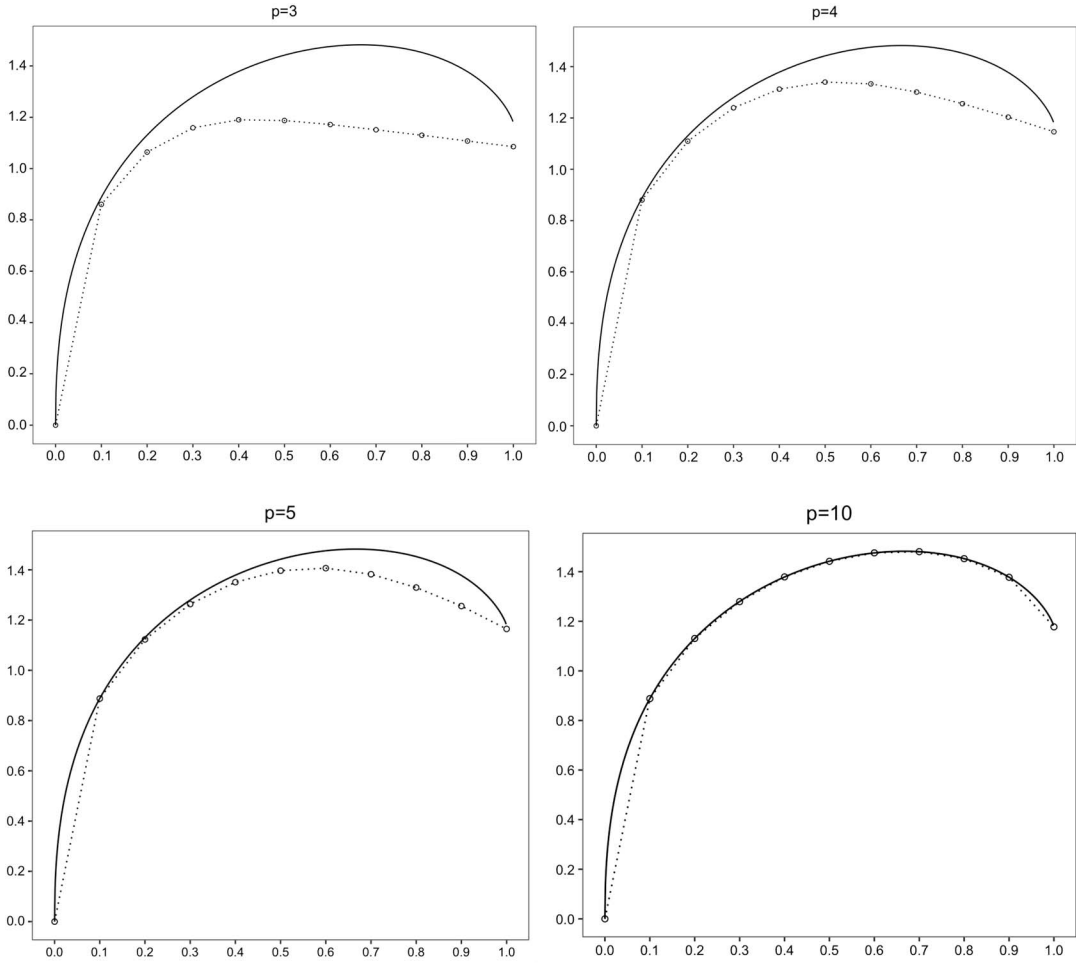


FIG. 1. Numerical simulations for the critical value β_c with sparse Rademacher prior and various values of p . The top left plot is for $p = 3$, the top right for $p = 4$, the bottom left for $p = 5$, and the bottom right for $p = 10$. The open circles are the simulated critical values β_c . The dashed curve interpolates between these points and the solid curve describes the function $H(\rho)$.

and $v_* = 1$. For this purpose, we have used the numerical integrator of Mathematica. The critical value β_c is the largest value b such that $\Gamma_b(v) \leq 0$ for all test values of v , where discrete positive values of b with increments 0.001 were tested. Figure 1 summarizes the numerical results for $p = 3, 4, 5, 10$ and $\rho = 0.1, 0.2, \dots, 1$.

The behavior of β_c is influenced by the proportion of zeros and the magnitude of the nonzero jumps. As can be seen, in each of the four figures there exists a threshold ρ_* (depending on p) such that $\rho \mapsto \beta_c$ is increasing on $[0, \rho_*]$ and decreasing on $[\rho_*, 1]$. Heuristically, in the interval $[0, \rho_*]$, the large fraction of the zeros dominates the small proportion of far jumps, whose magnitude $1/\sqrt{\rho}$ is large. On the other hand, in the interval $(\rho_*, 1]$, the far jumps overpower the small fraction of zeros and their magnitude has relatively low variation with ρ . In each subfigure of Figure 1, we indicate by a solid curve the following upper bound for β_c , which was pointed out in [56],

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

We note that as p increases the estimated values of β_c are closer to the ones of the upper bound $H(\rho)$. For $p = 3, 4, 5$, we see that if ρ is sufficiently small, then $H(\rho)$ is still a good approximation for β_c .

2.3. Main results for detection of multiple spikes. In this subsection, we study the case of more than one spike. Denote the number of spikes by k . Let $\Lambda_1, \dots, \Lambda_k$ be bounded subsets of $\mathbb{R}$ and $\mu_1, \dots, \mu_k$ be centered probability measures on the Borel σ -fields of $\Lambda_1, \dots, \Lambda_k$, respectively. For any $1 \leq r \leq k$, let $u_1(r), \dots, u_N(r)$ be i.i.d. samplings from μ_r and set $u(r) = (u_1(r), \dots, u_N(r))$. We assume that $u(1), \dots, u(k)$ are independent of each other and of W . For $\bar{\beta} = (\beta_1, \dots, \beta_k)$ with $\beta_1, \dots, \beta_k > 0$, the spiked tensor T_k is defined by

$$(2.6) \quad T_k = W + \frac{1}{N^{(p-1)/2}} \sum_{r=1}^k \beta_r u(r)^{\otimes p}.$$

In a manner similar to the previous subsection, we say that detection is possible if W and T_k are distinguishable and is impossible if they are indistinguishable. For $1 \leq r \leq k$, denote by $\beta_{r,c}$ the critical threshold obtained by plugging μ_r into Theorem 2.2. We extend Theorem 2.1 to the case of multiple spikes as follows.

THEOREM 2.3. *Assume that $\mu_1, \dots, \mu_k$ are centered. For $p \geq 3$, the following statements hold.*

- (i) *If $\bar{\beta} \in (0, \beta_{1,c}) \times \dots \times (0, \beta_{k,c})$, then detection is impossible;*
- (ii) *If $\bar{\beta} \notin (0, \beta_{1,c}] \times \dots \times (0, \beta_{k,c}]$, then detection is possible.*

Theorem 2.3 implies that in order to detect the spikes, at least one of the β_r 's has to exceed its own marginal critical threshold $\beta_{r,c}$. In particular, if all probability measures are the same, that is, $\mu_1 = \dots = \mu_k$, then the above result implies that W and T_k are indistinguishable if $\max_{1 \leq r \leq k} \beta_r < \beta_c$ and are distinguishable if $\max_{1 \leq r \leq k} \beta_r > \beta_c$, where β_c is the common threshold for all components.

REMARK 2.2. The first statement of Theorem 2.3 directly follows from Theorem 2.1 and a triangle inequality for the total variation distance, which is formulated in Lemma 4.3. The second statement of Theorem 2.3 is nontrivial and requires a thorough study of the high temperature regime of a spin glass model (see Section 3.2).

It is natural to ask whether this critical threshold β_c would change if one allows k to grow with N . We show that this is not the case if the growth of $k = k(N)$ is of certain polynomial order, which is sufficiently slow in comparison to the size of the p -tensor, N^p . To state our result, let μ be the probability measure considered in Section 2.2 and let β_c be the corresponding critical value provided by Theorem 2.2. Assume that $\mu_r = \mu$ for all $r \geq 1$ and that $(\beta_r)_{r \geq 1}$ is a sequence of SNRs satisfying $\sup_{r \geq 1} \beta_r < \infty$. Let T_k be the random tensor in (2.6) with μ_r and β_r for $1 \leq r \leq k$.

THEOREM 2.4. *Assume that $p \geq 3$ and $k = k(N)$ satisfies*

$$\lim_{N \rightarrow \infty} k(N) = \infty \quad \text{and} \quad \lim_{N \rightarrow \infty} \frac{k(N)}{N^{(p-2)/4}} = 0.$$

We have that:

- (i) *Detection is impossible if $\sup_{r \geq 1} \beta_r < \beta_c$.*
- (ii) *Detection is possible if either of the following two assumptions is satisfied:*
 - (a.1) *There exists a fixed $k_0 \in \mathbb{N}$ such that $\beta_r > \beta_c$ for at least one $r \leq k_0$ and $\sup_{r > k_0} \beta_r < \beta_c$.*
 - (a.2) *p is even and there exists a fixed $k_0 \in \mathbb{N}$ such that $\inf_{r \geq k_0} \beta_r > \beta_c$.*

REMARK 2.3. We expect that detection is possible if at least one of the β_r 's exceeds β_c . However, our proof requires more restrictive conditions for technical reasons. Note that condition (a.1) excludes the case of infinitely many β_r 's above β_c . On the other hand, condition (a.2) does not exclude this case, but requires p to be even.

As the number of independent spikes grows in N , it seems reasonable to believe that the critical threshold β_c should become smaller since now we have more spikes and it should be relatively easier to detect them in comparison to the case of a fixed finite number of spikes. However, Theorem 2.4 presents a counterintuitive result that if the total number of spikes is of smaller order than $N^{(p-2)/4}$, then the critical threshold remains unchanged. In particular, if we let $\beta_r = \tilde{\beta}/k$ for all $1 \leq r \leq k$ and use the average of the spikes, $U_k = k^{-1} \sum_{r=1}^k u(r)^{\otimes p}$, then we can write

$$T_{k(N)} = W + \frac{\tilde{\beta}}{N^{(p-1)/2}} U_{k(N)}.$$

If $k = k(N)$ satisfies the growth conditions above, Theorem 2.4 says that detection is impossible if $\tilde{\beta} < k\beta_c$ and detection is possible when $\tilde{\beta} > k\beta_c$ and p is even. Interestingly, this growth rate of $\tilde{\beta}$ required for detection, matches some recent results about algorithmic thresholds for spike recovery. In [45], Montanari and Richard established recovery via the tensor unfolding up to the threshold $N^{(\lceil p/2 \rceil - 1)/2}$ and predicted that the optimal threshold should be $N^{(p-2)/4}$. In addition, they obtained recovery via the tensor power iteration up to the threshold $N^{(p-1)/2}$ and conjectured that the true threshold for the power iteration and the approximate message passing algorithm is $N^{(p-2)/2}$ with no comment on the initialization. In [13], Ben Arous–Gheissari–Jagannath studied results for the Langevin dynamics and gradient descent and they gave recovery guarantees when $\tilde{\beta} > N^{(p-2)/2}$ for spherical and spin glass initial data. It was also known that the degree 4 sum-of-squares algorithm [32] and a related spectral algorithm [31] (for $p = 3$) have sharp recovery threshold $N^{(p-2)/4}$. More recently, a hierarchy of spectral methods following Kikuchi free energy has been proposed in the spiked tensor model [62], where it was shown that as long as the order parameter ℓ satisfies $\ell = o(N)$, then strong recovery and detection hold whenever $\tilde{\beta} \gg \ell^{-(p-2)/4} N^{(p-2)/4} \sqrt{\log N}$.

2.4. *Byproduct: Result for recovery by MMSE.* Recall the settings of Section 2.3. Let $\hat{\theta} = (\hat{\theta}_{i_1, \dots, i_p})$ be a $\mathbb{R}^{N^p}$ -valued bounded random variable generated by the σ -field $\sigma(T_k)$. We also allow the random variable $\hat{\theta}$ to be dependent on other randomness that are independent of the $u_i(r)$'s and T_k . The minimum mean square error (MMSE) is defined by

$$(2.7) \quad \text{MMSE}_N(\bar{\beta}) := \min_{\hat{\theta}} \frac{1}{N^p} \sum_{1 \leq i_1, \dots, i_p \leq N} \mathbb{E} \left(\sum_{r=1}^k \beta_r u_{i_1}(r) \cdots u_{i_p}(r) - \hat{\theta}_{i_1, \dots, i_p} \right)^2,$$

where the minimum is taken over all such $\hat{\theta}$. The minimizer to this problem is attained by the minimum mean square estimator,

$$\hat{\theta}^{\text{MMSE}} = \left(\sum_{r=1}^k \beta_r \mathbb{E}[u_{i_1}(r) \cdots u_{i_p}(r) | T_k] \right)_{1 \leq i_1, \dots, i_p \leq N}.$$

By restricting the minimum in the definition of $\text{MMSE}_N(\bar{\beta})$ to the so-called dummy estimators [37], that is, estimators where $\hat{\theta}$ is independent of $T_1, \dots, T_k$, one obtains a trivial upper

bound:

$$\begin{aligned} \text{MMSE}_N(\bar{\beta}) &\leq \frac{1}{N^p} \sum_{1 \leq i_1, \dots, i_p \leq N} \left(\mathbb{E} \left(\sum_{r=1}^k \beta_r u_{i_1}(r) \cdots u_{i_p}(r) \right)^2 \right. \\ &\quad \left. - \left(\mathbb{E} \left[\sum_{r=1}^k \beta_r u_{i_1}(r) \cdots u_{i_p}(r) \right] \right)^2 \right). \end{aligned}$$

Denote $v_{r,*} := \int a^2 \mu_r(da)$ for $1 \leq r \leq k$. Applying the strong law of large numbers yields

$$\limsup_{N \rightarrow \infty} \text{MMSE}_N(\bar{\beta}) \leq \text{DMSE}(\bar{\beta}) := \sum_{r=1}^k \beta_r^2 v_{r,*}^p.$$

Note that from the Gaussianity of Y , it can be directly checked that conditionally on T_k , the distribution of $(u(1), \dots, u(k))$ can be described by a Gibbs measure G_N^A on $\Lambda_1^N \times \cdots \times \Lambda_k^N$, see (5.2) below. Denote by $(\sigma(1), \dots, \sigma(k))$ a sampling from G_N^A and by $\langle \cdot \rangle^A$ the expectation associated to this measure. We show that

THEOREM 2.5. *For $p \geq 3$, the following statements hold.*

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

$$\begin{aligned} \lim_{N \rightarrow \infty} \text{MMSE}_N(\bar{\beta}) &= \text{DMSE}(\bar{\beta}), \\ \lim_{N \rightarrow \infty} \sum_{r,r'=1}^k \beta_r \beta_{r'} \mathbb{E} \left\langle \left(\frac{1}{N} \sum_{i=1}^N u_i(r) \sigma_i(r') \right)^p \right\rangle^A &= 0. \end{aligned}$$

(ii) *If $\bar{\beta} \notin (0, \beta_{1,c}] \times \cdots \times (0, \beta_{k,c}]$, then*

$$\begin{aligned} \limsup_{N \rightarrow \infty} \text{MMSE}_N(\bar{\beta}) &< \text{DMSE}(\bar{\beta}), \\ \liminf_{N \rightarrow \infty} \sum_{r,r'=1}^k \beta_r \beta_{r'} \mathbb{E} \left\langle \left(\frac{1}{N} \sum_{i=1}^N u_i(r) \sigma_i(r') \right)^p \right\rangle^A &> 0. \end{aligned}$$

This theorem asserts that if the SNRs of all marginal spikes are less than their critical thresholds, then when estimating for the tensor $\sum_{r=1}^k \beta_r u(r)^{\otimes p}$, the minimum mean square estimator is no better than a random guess. In contrast, if at least one of the SNRs of the marginal spikes is larger than its critical threshold, the minimum mean square estimator performs better than all dummy estimators. In the case that p is even, Theorem 2.5(i) further implies that the sampling $(\sigma(1), \dots, \sigma(k))$ does not provide useful information in recovering $(u(1), \dots, u(k))$ since their inner products are essentially zero, whereas Theorem 2.5(ii) shows that a weak form of recovery is possible as $(u(1), \dots, u(k))$ and $(\sigma(1), \dots, \sigma(k))$ are asymptotically correlated.

As mentioned before, the spike recovery in the random tensor for general priors via the MMSE was studied earlier by Lesieur–Miolane–Lelarge–Krzakala–Zdeborová [38]. They computed the limiting mutual information between W and T_k and used it to establish a result equivalent to Theorem 2.5. The proof of Theorem 2.5 relies heavily on our main results for the detection problem and presents a different approach than the one taken in [38].

2.5. Previous results. Understanding phase transitions of spike detection and recovery problems in spiked random matrices and tensors has received a lot of attention in the past several years. We summarize some recent works here.

Matrix case: $p = 2$. The PCA approach was studied by Baik–Ben Arous–Péché [3], Baik–Silverstein [4], Féral–Péché [28], Johnstone [35], Paul [53], and Péché [54]. Barbier–Dia–Macris–Krzakala–Lesieur–Zdeborová [6] studied the MMSE recovery problem in the spiked random matrix in (2.2) (see the setting in Section 2.4 with $p = 2$ and $k = 1$) by deriving a Parisi-type formula for the mutual information between W and T . Analogous study for the case of multiple spikes (2.6) was handled by Lelarge–Miolane [37], where $u(1), \dots, u(k)$ are assumed to have finite second moments and are allowed to be correlated. Similar result for the nonsymmetric case was pursued by Miolane [42].

As for the detection problem, under the same setting as (2.2), Alaoui–Krzakala–Jordan [27] obtained the same critical value β_c specified in equation (3.3) and Proposition 3.1 below. It was deduced that above β_c , detection is possible and below β_c , a weak form of detection remains possible in the sense that the limiting total error (the sum of type one and type two errors) of the likelihood ratio test between W and T is strictly less than one. Incidentally, we mention that when the results of [6, 37] apply to the case (2.2), β_c is also the critical threshold for recovery. In [26], El Alaoui and Jordan extended the results of [27] to the case of spiked rectangular matrices, where the spike is of the form uv^T and it was assumed that the entries of $u \in \mathbb{R}^M$, $v \in \mathbb{R}^N$ are chosen independently at random from possibly different priors and $M/N \rightarrow \alpha$. It was shown that for a set of parameters (α, β) the results of [27] hold. This set of parameters is sub-optimal for most priors as the spin-glass methods used fail near the boundaries of the optimal parameter space for the model of [26].

Tensor case: $p \geq 3$. Earlier results trace back to the works of Montanari–Richard [45] and Montanari–Reichman–Zeitouni [44], where the authors considered (2.6) with $k = 1$ and a spherical prior, that is, u in (2.2) is uniformly and independently sampled from the sphere, $\{x \in \mathbb{R}^N : \sum_{i=1}^N x_i^2 = N\}$. By adaptation of the second moment method, they showed that there exist β_- and β_+ such that detection is impossible for β below β_- and is possible for β above β_+ .

Lesieur–Miolane–Lelarge–Krzakala–Zdeborová [38] considered (2.6) with a general setting in which the vectors $(u_i(1), \dots, u_i(k))$ for $1 \leq i \leq N$ are i.i.d. sampled from a joint distribution with finite second moments. For centered priors, they proved that there exists a vector of critical thresholds $(\beta'_{1,c}, \dots, \beta'_{k,c})$ such that for any $\bar{\beta} = (\beta_1, \dots, \beta_k)$ satisfying $\beta_r > \beta'_{r,c}$ for $1 \leq r \leq k$, the MMSE estimator obtains a better error than any dummy estimator. Consequently, one can also detect the spike in that case. In addition, when $\bar{\beta}$ satisfies $\beta_r < \beta'_{r,c}$ for all $1 \leq r \leq k$, the MMSE estimator is statistically irrelevant to recover the spike. They did not provide results for the detection problem in this case. Notably, if $u(1), \dots, u(k)$ are chosen as in Section 2.3, our critical thresholds $\beta_{r,c}$ agree with $\beta'_{r,c}$ and as a consequence, their result in this case is the same as Theorem 2.5. Barbier–Macris [8] provided a different proof for the results of [38] by using stochastic interpolation. Analogous results to [38] were developed in nonsymmetric settings by Barbier–Macris–Miolane [10].

Perry–Wein–Bandeira [56] focused on $k = 1$ and three priors: the spherical prior, the Rademacher prior, and the sparse Rademacher prior. In these three settings, it was proved that there exist lower and upper bounds β'_- and β'_+ such that detection is not possible when $0 < \beta \leq \beta'_-$ and is possible when $\beta \geq \beta'_+$. In particular, their result in the spherical case improved the existing bounds in [44, 45] mentioned above. For the Rademacher prior, Chen [20] closed the gap between β'_- and β'_+ by showing that β_c in Theorem 2.2 is indeed the

critical threshold for detection. The present work extends the results of [20, 56] to a broader class of priors and also to $k > 1$.

Other related works. Since the likelihood ratio test and the MMSE estimator are often intractable to compute, it is natural to ask about the performance of tractable algorithms for detection and recovery for low-rank signals. The works [5, 7, 9, 22, 23, 37, 42, 46] studied the performance of the approximate message passing (AMP) algorithm in recovering the spike. See [12, 24, 34, 57] for the performance of AMP in compressed sensing. See [13] for the performance of the Langevin dynamics and the gradient decent in the spiked tensor model. The complexity of energy landscapes in spiked tensor models was studied in [14, 58].

2.6. Our approaches. As mentioned above, the work [20] considered the Gaussian p -tensor model for all $p \geq 3$ with a single spike (2.2) sampled from the Rademacher prior and it obtained the same result as Theorem 2.1. In the present paper, we extend [20] to general priors and to multiple spikes. In view of [20], the approach was based on a connection between the total variation distance of the pair (W, T) and the free energy (see (3.2)) of the pure p -spin spin glass model through an integral representation (see Lemma 4.2). From this, proving the impossibility of detection relies on knowing the decaying rate of the tail probability of the free energy in the high temperature regime. The core ingredient of obtaining this tail probability relied on a delicate study of the fluctuation of the free energy via the Parisi formula, the coupled free energy with overlap constraints, and the two-dimensional Guerra–Talagrand inequality.

Theorem 2.1 follows essentially from the same treatments as [20] by studying the fluctuation of a one-dimensional spin glass free energy (3.2). However, while the arguments in [20] were greatly simplified due to the simple structure of the Rademacher prior, there are a number of analytic obstacles in handling our generalization. For example, the main results in [20] critically relied on the strict monotonicity of γ_β in β (recalling (2.5)). The proof of this property used the symmetry of the Rademacher prior and it does not carry through in our setting. To prove our main results, we establish an analogous, though more general proposition, in Lemma 6.1 below, which requires a completely new argument.

Our approach to the high-dimensional generalization, Theorem 2.3, relies on the high temperature behavior of the free energy associated to the vector-valued pure p -spin spin glass model, see Section 3.2. In spin glasses, vector-valued models are usually harder as the spin components interact with each other in a highly complicated way. As a result, the analysis of the Parisi formula for the corresponding free energy and its coupled version becomes more involved. Nevertheless, to study the high temperature regime, we can directly handle the free energy by reducing the high-dimensional Hamiltonian to one-dimensional ones by exploiting the overlap constraints, see Section 9. This helps us avoid controlling the Parisi formula of the vector-valued model and greatly simplifies our argument. Ultimately, this leads to a full characterization of the high temperature regime of the vector-valued pure p -spin model (Theorem 3.2 below) and concludes Theorem 2.3. We expect that this approach is also applicable in characterizing the high temperature regimes for more general spin glass models, for instance, the vector-valued *mixed* p -spin model without external field.

REMARK 2.4. The assumption on the boundedness of the support of μ is used for technical purposes. For example, one can note that the estimates in Lemmas 8.2–8.4 use the bound M on the size of the support of μ . We believe that our results remain valid when μ is unbounded, but has exponential tail probability.

REMARK 2.5. We derive our main results in Section 2 by studying the pure p -spin mean-field spin glass model, whose one-dimensional Hamiltonian is defined in (3.1). As an alternative approach, one can study instead the planted model, whose Hamiltonian is formulated

in (5.1) with $t = 1$. It is natural for Bayesian estimation and can save some technical issues. Most notably, the analog of Proposition 7.1 for the planted model has a simpler proof. For details, we refer the readers to check [27], Proposition 13, and [25], Proposition 16, which are analogs of Proposition 7.1, but use the planted model for the one-dimensional spike detection problem in the matrix setting (i.e., $p = 2$). The planted model also yields an alternative scheme for determining the critical threshold β_c by using the replica symmetric Parisi formula [38], Theorem 1. Nevertheless, the model that we consider in this paper is essential to the field of spin glasses (see, e.g., Bolthausen [17], Section 6) and our study of its high-temperature behavior is of independent interest.

2.7. Structure of the rest of the paper. The key ingredient of this paper relies on an observation that the total variation distance between W and T_k can be expressed as an integral related to the free energy of the pure p -spin models with scalar- and vector-valued spin configurations (Lemma 4.2). Section 3 defines these models, characterizes their high-temperature regimes, and presents results on the fluctuation of the free energy and concentration of the overlap of the models. Section 4 establishes Theorems 2.1–2.4, while Section 5 presents the proof of Theorem 2.5. The rest of the sections are devoted to establishing the main results in Section 3. In Sections 6 and 9, we prove the asserted structures of the high-temperature regimes. These proofs are the most crucial components in this paper. Sections 7 and 8 establish the high-temperature behavior of the overlap and the free energy when $k = 1$.

3. Pure p -spin models. In this section, we introduce the pure p -spin mean field spin glass models with scalar-valued and vector-valued spin configurations and formulate some crucial results regarding their high-temperature behavior. Their proofs are deferred to later sections.

3.1. Scalar-valued model. Recall the random tensor Y from Section 2.1 and the probability space (Λ, μ) from Section 2.2. For any $\sigma \in \Lambda^N$, the Hamiltonian of the pure p -spin model is defined as

$$(3.1) \quad X_N(\sigma) = \frac{1}{N^{(p-1)/2}} \langle Y, \sigma^{\otimes p} \rangle = \frac{1}{N^{(p-1)/2}} \sum_{1 \leq i_1, \dots, i_p \leq N} Y_{i_1, \dots, i_p} \sigma_{i_1} \cdots \sigma_{i_p},$$

where the $Y_{i_1, \dots, i_p}$'s are i.i.d. standard Gaussian random variables. Note that by the symmetry of W , we also have the identity $X_N(s) = N^{-(p-1)/2} \langle W, \sigma^{\otimes p} \rangle$. For any two spin configurations σ^1 and σ^2 , the covariance of X_N can be computed as

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

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

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

Define the re-centered Hamiltonian $H_{N,\beta}(\sigma)$ by

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

Note that $\mathbb{E}e^{H_{N,\beta}(\sigma)} = 1$. Define the free energy and Gibbs measure respectively by

$$(3.2) \quad F_N(\beta) = \frac{1}{N} \log \int e^{H_{N,\beta}(\sigma)} \mu^{\otimes N}(d\sigma)$$

and

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

where $Z_{N,\beta}$ is the normalizing constant so that $G_{N,\beta}$ is a probability measure on Λ^N . Denote by $\langle \cdot \rangle_\beta$ the Gibbs expectation with respect to the i.i.d. samplings $\sigma, \sigma^1, \sigma^2, \dots$ from the Gibbs measure $G_{N,\beta}$.

A few properties of F_N are in position. First of all, an application of the Gaussian concentration of measures implies that $F_N(\beta)$ is concentrated around $\mathbb{E}F_N(\beta)$. Here, from the Jensen's inequality,

$$\mathbb{E}F_N(\beta) \leq \frac{1}{N} \log \int \mathbb{E} e^{H_{N,\beta}(\sigma)} \mu^{\otimes N}(d\sigma) = \frac{1}{N} \log 1 = 0.$$

In addition, $\mathbb{E}F_N(\beta)$ is a nonincreasing function since by using Gaussian integration by parts,

$$\begin{aligned} \frac{d}{d\beta} \mathbb{E}F_N(\beta) &= \frac{1}{N} \mathbb{E} \langle X_N(\sigma) \rangle_\beta - \beta \mathbb{E} \langle R(\sigma, \sigma)^p \rangle_\beta \\ &= -\beta \mathbb{E} \langle R(\sigma^1, \sigma^2)^p \rangle_\beta = -\beta \sum_{1 \leq i_1, \dots, i_p \leq N} \mathbb{E} \langle (\sigma_{i_1} \cdots \sigma_{i_p})^2 \rangle_\beta \leq 0. \end{aligned}$$

Second, it can be shown (see Proposition 6.1 below) that for all β , $\lim_{N \rightarrow \infty} \mathbb{E}F_N(\beta)$ exists. Denote this limit by $F(\beta)$. From above, $F(\beta)$ is nonpositive and nonincreasing. Define the high-temperature regime as

$$\mathcal{R} = \{\beta > 0 : F(\beta) = 0\}.$$

The low-temperature regime is define as $\mathcal{R}^c$. Set the critical threshold β_c by

$$(3.3) \quad \beta_c = \sup \mathcal{R}.$$

In spin glasses, the parameter β is understood as the (inverse) temperature parameter, while in the detection problem of (2.6), it is interpreted as the signal strength or SNR. These equivalent meanings of β are justified below in Lemma 4.2 via an integral representation for the total variation distance between W and T .

The following proposition shows that the high-temperature regime $\mathcal{R}$ is an interval and its right-end boundary is β_c . It also gives a characterization of $\mathcal{R}$ in terms of the constant v_* and the auxiliary function $\Gamma_b(v)$ defined in (2.3) and (2.4), respectively.

PROPOSITION 3.1. *For $p \geq 2$, $\mathcal{R} = (0, \beta_c]$. For $\beta > 0$, $\beta \in \mathcal{R}$ if and only if*

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

Next, we show that in the interior of the high-temperature regime, the overlap between two i.i.d. samples σ^1 and σ^2 is concentrated around zero.

THEOREM 3.1. *For $p \geq 2$, $m \in \mathbb{N}$, and $0 < \beta < \beta_c$, there exists a constant $K > 0$, depending only on p, m , and β , such that*

$$(3.4) \quad \mathbb{E} \langle |R(\sigma^1, \sigma^2)|^{2m} \rangle_{s\beta} \leq \frac{K}{N^m} \quad \forall s \in [0, 1], N \geq 1.$$

Furthermore, we control the fluctuation of the free energy as follows.

PROPOSITION 3.2. *For $p \geq 2$ and $0 < \beta < \beta_c$, there exists a constant K , depending only on p and β , such that*

$$\mathbb{P}(|F_N(\beta)| \geq l) \leq \frac{K}{l^{2N^{p/2+1}}} \quad \forall l > 0, N \geq 1.$$

In the case that μ is a uniform probability measure on $\{-1, 1\}$, the behavior of the overlap and the fluctuation of the free energy at high-temperature is well-understood. The case $p = 2$ corresponds to the famous Sherrington–Kirkpatrick (SK) model. In this case, Aizenman–Lebowitz–Ruelle [1] proved that $NF_N(\beta)$ converges to a Gaussian random variable when $\beta < \beta_c = 1$ and Talagrand [61], Chapters 11 and 13, obtained the moment control of Theorem 3.1. For $p \geq 3$, Bardina–Márquez–Carreras–Rovira–Tindel [11] established (3.4) for $\beta \ll \beta_c$. For even $p \geq 4$, Bovier–Kurkova–Löwe [18] showed that $N^{p/4+1/2}F_N(\beta)$ has a Gaussian fluctuation up to some temperature strictly less than β_c . More recently, Chen [20] obtained the same statements as Theorem 3.1 and Proposition 3.2 for this choice of (Λ, μ) . Our main contribution here is to establish concentration of the overlap and the fluctuation of the free energy up to the critical temperature for any spin configurations sampled from a probability measure on a bounded subset of the real line.

3.2. *High temperature regime of the vector-valued model.* Next we consider the pure p -spin model with k -dimensional vector-valued spin configurations, where $k \geq 2$. Recall the probability spaces $(\Lambda_1, \mu_1), \dots, (\Lambda_k, \mu_k)$ from Section 2.3. Set the product space and measure by

$$\begin{aligned} \bar{\Lambda} &= \Lambda_1 \times \dots \times \Lambda_k, \\ \bar{\mu} &= \mu_1 \otimes \dots \otimes \mu_k. \end{aligned}$$

For $\sigma(r) \in \Lambda_r^N$, $1 \leq r \leq k$, denote

$$\begin{aligned} \bar{\sigma}_i &= (\sigma_i(1), \dots, \sigma_i(k))^T \in \bar{\Lambda}, \quad 1 \leq i \leq N, \\ \bar{\sigma} &= (\bar{\sigma}_1, \dots, \bar{\sigma}_N) \in \bar{\Lambda}^N. \end{aligned}$$

In other words, the spin configuration $\bar{\sigma}$ is a $k \times N$ matrix: the rows are $\sigma(1) \in \Lambda_1^N, \dots, \sigma(k) \in \Lambda_k^N$ and the columns are $\bar{\sigma}_1, \dots, \bar{\sigma}_N \in \bar{\Lambda}$. Given $\bar{\beta} = (\beta_1, \dots, \beta_k)$ with $\beta_1, \dots, \beta_k > 0$, the re-centered pure p -spin Hamiltonian with vector-valued spin configurations is defined as

$$H_{N,\bar{\beta}}(\bar{\sigma}) = \sum_{r=1}^k \beta_r X_N(\sigma(r)) - \sum_{r,r'=1}^k \frac{\beta_r \beta_{r'}}{2} NR(\sigma(r), \sigma(r'))^p, \quad \bar{\sigma} \in \bar{\Lambda}^N.$$

Similar to the scalar-valued model, the free energy and the Gibbs measure are defined as

$$(3.5) \quad F_N(\bar{\beta}) = \frac{1}{N} \log \int e^{H_{N,\bar{\beta}}(\bar{\sigma})} \bar{\mu}^{\otimes N}(d\bar{\sigma})$$

and

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

where $Z_{N,\bar{\beta}}$ is the normalizing constant. Define

$$F(\bar{\beta}) = \limsup_{N \rightarrow \infty} F_N(\bar{\beta}).$$

There is a technical subtlety here that is not present in the previous subsection. In the case of even p , Panchenko [52] proved that if one drops the overlap term in $H_{N,\bar{\beta}}$, then the limiting free energy with overlap constraint exists. Consequently, one can show that $F(\bar{\beta}) = \lim_{N \rightarrow \infty} F_N(\bar{\beta})$ (see the proof of Proposition 6.1 below). When p is odd, this limit is preserved if $k = 1$, as explained in the previous subsection, but whether it is still true for $k \geq 2$ remains an open question.

An application of Jensen's inequality ensures that $F(\bar{\beta}) \leq 0$. The high-temperature regime is defined as

$$\bar{\mathcal{R}} = \{\bar{\beta} = (\beta_1, \dots, \beta_k) \mid \beta_r > 0 \text{ for all } 1 \leq r \leq k \text{ and } F(\bar{\beta}) = 0\}.$$

Again, while $\bar{\beta}$ is understood as the vector of SNRs in the detection problem, we read the entries of this vector as the temperature parameters in the setting of spin glass models. Let $\beta_{r,c}$ be the critical temperature obtained from Section 3.1 by taking $(\Lambda, \mu) = (\Lambda_r, \mu_r)$. The following theorem states that the high-temperature regime of the vector-valued p -spin model is equal to the product of the high-temperature regimes of the marginal systems.

THEOREM 3.2. *For $p \geq 3$, $\bar{\mathcal{R}} = (0, \beta_{1,c}] \times \dots \times (0, \beta_{k,c}]$.*

Theorem 3.2 highlights an interesting phenomenon: Although the Hamiltonian $H_{N,\bar{\beta}}$ involves interactions coming from the overlaps $R(\sigma(r), \sigma(r'))$ for all $r \neq r'$, in the high-temperature regime the marginal spin configurations $\sigma(1), \dots, \sigma(k)$ under $H_{N,\bar{\beta}}$ essentially interact with each other independently. Consequently, they behave like k independent one-dimensional systems associated to $H_{N,\beta_1}, \dots, H_{N,\beta_k}$. As a result, the high-temperature regime of $H_{N,\bar{\beta}}$ is simply the product of the high-temperature regimes of the marginal systems.

4. Establishing spike detection. This section proves the main theorems of this paper. Section 4.1 first expresses the total variation distance that appears in the detection problem in terms of the free energy of the pure p -spin model. Using this expression and results described in Section 3, Sections 4.2–4.3 conclude the proofs of Theorems 2.1–2.4.

4.1. Total variation distance. It is well known that one can relate the total variation distance between two continuous random variables to the ratio of their probability densities. See for instance [20], Lemma 1.

LEMMA 4.1. *If U and V are two N -dimensional random vectors with densities f_U and f_V , respectively, and $f_U(x), f_V(x) \neq 0$ a.e., then*

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

Recall T and T_k from (2.2) and (2.6). Note that W is a symmetric Gaussian p -tensor and the spikes are independent of W . From these, one can compute the density functions for W , T , and T_k explicitly and then apply Lemma 4.1 to the pairs (W, T) and (W, T_k) to get

LEMMA 4.2. *For any $\beta \in (0, \infty)$ and $\bar{\beta} \in (0, \infty)^k$,*

$$(4.1) \quad d_{\text{TV}}(W, T) = \int_0^1 \mathbb{P}(F_N(\beta) < N^{-1} \log x) dx,$$

$$(4.2) \quad d_{\text{TV}}(W, T_k) = \int_0^1 \mathbb{P}(F_N(\bar{\beta}) < N^{-1} \log x) dx.$$

For a detailed derivation, we refer the reader to [20], Lemma 2.

4.2. *Proof of Theorems 2.1 and 2.2.* PROOF OF THEOREM 2.1. Let β_c be the critical temperature defined in (3.3). Assume that $0 < \beta < \beta_c$. From (4.1), using change of variable $y = -\log x$ and writing $\int_0^\infty = \int_0^\varepsilon + \int_\varepsilon^\infty$ imply that

$$\begin{aligned} d_{\text{TV}}(W, T) &= \left(\int_0^\varepsilon + \int_\varepsilon^\infty \right) \mathbb{P}(F_N(\beta) < -N^{-1}y) e^{-y} dy \\ &\leq \int_0^\varepsilon e^{-y} dy + \int_\varepsilon^\infty \mathbb{P}(|F_N(\beta)| \geq N^{-1}y) e^{-y} dy \\ &\leq \varepsilon + \int_\varepsilon^\infty \frac{K}{y^2 N^{\frac{p}{2}-1}} e^{-y} dy \\ &\leq \varepsilon + \frac{K}{\varepsilon N^{\frac{p}{2}-1}} \quad \forall \varepsilon > 0, \end{aligned}$$

where the second inequality used Proposition 3.2. Letting $\varepsilon = N^{-(p-2)/4}$ yields

$$(4.3) \qquad d_{\text{TV}}(W, T) \leq \frac{1+K}{N^{\frac{p-2}{4}}}.$$

This implies that W and T are indistinguishable, so detection is impossible. Next, assume that $\beta > \beta_c$. Recall that $F_N(\beta)$ converges to $F(\beta)$ almost surely and note that $F(\beta) < 0$. It follows that

$$\lim_{N \rightarrow \infty} \mathbb{P}(F_N(\beta) - N^{-1} \log x < 0) = \mathbb{P}(F(\beta) < 0) = 1$$

and the dominated convergence theorem yields

$$\lim_{N \rightarrow \infty} d_{\text{TV}}(W, T) = \int_0^1 \mathbb{P}(F(\beta) < 0) dx = 1. \qquad \square$$

PROOF OF THEOREM 2.2. We have seen from the proof of Theorem 2.1 that the critical temperature β_c defined in (3.3) is the critical threshold for detection. In addition, from Proposition 3.1, we see that β_c satisfies $\sup_{v \in (0, v_*]} \Gamma_{\beta_c}(v) \leq 0$ and that any $\beta > 0$ satisfying $\sup_{v \in (0, v_*]} \Gamma_\beta(v) \leq 0$ must also satisfy $\beta \leq \beta_c$. From this, to complete the proof, it suffices to show that $\sup_{v \in (0, v_*]} \Gamma_{\beta_c}(v) = 0$. If on the contrary $\sup_{v \in (0, v_*]} \Gamma_{\beta_c}(v) < 0$, then there exists some $\beta > \beta_c$ such that $\sup_{v \in (0, v_*]} \Gamma_\beta(v) < 0$ since $\Gamma_\beta(v)$ is a continuous function in β and v . This contradicts the fact that $\beta \leq \beta_c$. $\square$

4.3. *Proof of Theorems 2.3 and 2.4.* The proof of Theorem 2.3 relies on the following simple lemma:

LEMMA 4.3. Assume that Y_1, Y_2, Y_3 are random vectors of the same size and Y_2 is independent of Y_1 and Y_3 . Then

$$d_{\text{TV}}(Y_1, Y_1 + Y_2 + Y_3) \leq d_{\text{TV}}(Y_1, Y_1 + Y_2) + d_{\text{TV}}(Y_1, Y_1 + Y_3).$$

PROOF. The assertion follows immediately by using the triangle inequality,

$$d_{\text{TV}}(Y_1, Y_1 + Y_2 + Y_3) \leq d_{\text{TV}}(Y_1, Y_1 + Y_2) + d_{\text{TV}}(Y_1 + Y_2, Y_1 + Y_2 + Y_3)$$

and noting that the independence between Y_2 and Y_1, Y_3 yields

$$\begin{aligned} &d_{\text{TV}}(Y_1 + Y_2, Y_1 + Y_2 + Y_3) \\ &= \sup_A |\mathbb{E}_{Y_2} [\mathbb{P}(Y_1 \in A - Y_2 | Y_2) - \mathbb{P}(Y_1 + Y_3 \in A - Y_2 | Y_2)]| \end{aligned}$$

$$\begin{aligned}
 &\leq \mathbb{E}_{Y_2} \left[\sup_A |\mathbb{P}(Y_1 \in A - Y_2 | Y_2) - \mathbb{P}(Y_1 + Y_3 \in A - Y_2 | Y_2)| \right] \\
 &= \mathbb{E}_{Y_2} \left[\sup_A |\mathbb{P}(Y_1 \in A) - \mathbb{P}(Y_1 + Y_3 \in A)| \right] \\
 &= d_{\text{TV}}(Y_1, Y_1 + Y_3),
 \end{aligned}$$

where $\mathbb{E}_{Y_2}$ is the expectation with respect to Y_2 only. $\square$

PROOF OF THEOREM 2.3. Let $\bar{\beta} = (\beta_1, \dots, \beta_k) \in (0, \beta_{1,c}) \times \dots \times (0, \beta_{r,c})$. For $1 \leq r \leq k$, set

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

From Lemma 4.3 and an induction argument,

$$(4.4) \quad d_{\text{TV}}(W, T_k) \leq \sum_{r=1}^k d_{\text{TV}}(W, T_{k,r}).$$

Since $\beta_r \in (0, \beta_{r,c})$, (4.3) implies that there exists a constant $K_r > 0$ such that for any $N \geq 1$,

$$(4.5) \quad d_{\text{TV}}(W, T_{k,r}) \leq \frac{K_r}{N^{(p-2)/4}}.$$

This together with (4.4) implies that detection is impossible. Next, assume that $\bar{\beta} \notin (0, \beta_{1,c}] \times \dots \times (0, \beta_{k,c}]$. Since

$$\limsup_{N \rightarrow \infty} F_N(\bar{\beta}) = F(\bar{\beta}) < 0 \quad \text{a.s.,}$$

the Fatou lemma yields that for any $x > 0$,

$$\begin{aligned}
 \liminf_{N \rightarrow \infty} \mathbb{P}(F_N(\bar{\beta}) < N^{-1} \log x) &= \liminf_{N \rightarrow \infty} \mathbb{E}[I(F_N(\bar{\beta}) < N^{-1} \log x)] \\
 &\geq \mathbb{E} \left[\liminf_{N \rightarrow \infty} I(F_N(\bar{\beta}) < N^{-1} \log x) \right] = 1,
 \end{aligned}$$

where $I(\cdot)$ is an indicator function. Using this, (4.2), and the Fatou lemma again, we arrive at

$$\begin{aligned}
 \liminf_{N \rightarrow \infty} d_{\text{TV}}(W, T_k) &= \liminf_{N \rightarrow \infty} \int_0^1 \mathbb{P}(F_N(\bar{\beta}) < N^{-1} \log x) dx \\
 &\geq \int_0^1 \liminf_{N \rightarrow \infty} \mathbb{P}(F_N(\bar{\beta}) < N^{-1} \log x) dx = \int_0^1 1 dx = 1.
 \end{aligned}$$

Thus, detection is possible. $\square$

PROOF OF THEOREM 2.4. Assume that $\sup_{r \geq 1} \beta_r < \beta_c$. From (4.4) and (4.5),

$$d_{\text{TV}}(W, T_k) \leq \frac{Kk}{N^{(p-2)/4}} \quad \forall N \geq 1,$$

where K is a universal constant independent of N . From the assumption on k , the right-hand side vanishes as N tends to infinity and this establishes the assertion (i).

Next, we establish (ii) assuming (a.1). Suppose that N satisfies $k = k(N) > k_0$. Denote

$$\Delta_k = \frac{1}{N^{(p-1)/2}} \sum_{k_0 < r \leq k} \beta_r u(r)^{\otimes p}.$$

Since $(u(r) : 1 \leq r \leq k_0)$ is independent of $(u(r) : r > k_0)$,

$$d_{\text{TV}}(T_k, T_{k_0}) \leq d_{\text{TV}}(W + \Delta_k, W) \leq \sum_{k_0 < r \leq k} d_{\text{TV}}(W, T_{k,r}),$$

where the second inequality used Lemma 4.3 and $T_{k,r}$ is defined in the proof of Theorem 2.3. Hence, from the triangular inequality and (4.5), there exists a positive constant K such that

$$d_{\text{TV}}(W, T_{k_0}) \leq d_{\text{TV}}(W, T_k) + d_{\text{TV}}(T_k, T_{k_0}) \leq d_{\text{TV}}(W, T_k) + \frac{Kk}{N^{(p-1)/4}} \quad \forall N \geq 1.$$

Here, since $\beta_r > \beta_c$ for at least one $r \leq k_0$, it means that $(\beta_1, \dots, \beta_{k_0}) \notin (0, \beta_c] \times \dots \times (0, \beta_c]$ and from Theorem 2.3, $d_{\text{TV}}(W, T_{k_0}) \rightarrow 1$. This implies that $d_{\text{TV}}(W, T_k) \rightarrow 1$ and the assertion (ii) follows under (a.1).

To establish Theorem 2.4 assuming (a.2), note that since p is even, dropping the overlap terms in $H_{N,\bar{\beta}}(\bar{\sigma})$ yields $F_N(\bar{\beta}) \leq \sum_{r=1}^k F_{N,r}(\beta_r)$. In addition, note that $\mathbb{E}F_{N,r}(\cdot)$ is the same function for every r and it can be checked, by using Gaussian integration by parts, that its derivative is uniformly bounded as long as β stays in a bounded interval. Hence, for every $r \geq 1$, $\mathbb{E}F_{N,r}(\cdot)$ is a sequence of equicontinuous functions. As a result, the assumption that $\inf_{r \geq k_0} \beta_r > \beta_c$ and $\sup_{r \geq k_0} \beta_r < \infty$ implies that there exist some $\delta > 0$ and $N_0 \geq 1$ such that $\mathbb{E}F_{N,r}(\beta_r) \leq -\delta$ for all $r \geq k_0$ and $N \geq N_0$. On the other hand, the Gaussian concentration inequality implies that there exists a universal constant $K > 0$ such that

$$\mathbb{P}(\Omega_{N,r,t}) \leq K e^{-t^2 N/K} \quad \forall N \geq 1, r \geq 1, t > 0,$$

where $\Omega_{N,r,t} := \{|F_{N,r}(\beta_r) - \mathbb{E}F_{N,r}(\beta_r)| \geq t\}$. Fix $0 < t \leq \delta/2$. From these, the probability of the event $\bigcap_{1 \leq r \leq k} \Omega_{N,r,t}^c$ is at least $1 - kK e^{-t^2 N/K}$. Furthermore, as long as $N \geq N_0$ satisfies $k = k(N) > k_0$, on this event,

$$\begin{aligned} \sum_{r=1}^k F_{N,r}(\beta_r) &= \sum_{r=1}^k (F_{N,r}(\beta_r) - \mathbb{E}F_{N,r}(\beta_r)) + \sum_{r=1}^k \mathbb{E}F_{N,r}(\beta_r) \\ &\leq \frac{\delta k}{2} - \sum_{k_0 < r \leq k} \delta = \frac{k_0 \delta}{2} - \frac{1}{2} \sum_{k_0 < r \leq k} \delta, \end{aligned}$$

where the first inequality used $\mathbb{E}F_{N,r}(\beta_r) \leq 0$. Hence, from the assumption on k and the Borel–Cantelli lemma, $\limsup_{N \rightarrow \infty} F_N(\bar{\beta}) = \infty$ a.s. and this implies that for all $x \in (0, 1)$,

$$\lim_{N \rightarrow \infty} \mathbb{P}(F_N(\bar{\beta}) \leq N^{-1} \log x) = 1.$$

From (4.2), the assertion (ii) follows. $\square$

5. Establishing spike recovery. We present the proof of Theorem 2.5 in this section. Recall that we handled the detection problem by means of the free energies of the spin glass models defined in Section 3. Our treatment for Theorem 2.5 will also rely on an auxiliary spin glass model, which arises naturally from the conditional distribution of $u(1), \dots, u(r)$ given T_k . This allows us to establish the so-called Nishimori identity and connect the MMSE to the free energy associated to this auxiliary spin system.

5.1. Nishimori identity. Recall the probability spaces (Λ_r, μ_r) , the product probability space $(\bar{\Lambda}, \bar{\mu})$, and the Hamiltonians $H_{N,\bar{\beta}}(\bar{\sigma})$ from Section 3.2. Fix a SNR vector $\bar{\beta}$. For any $t \geq 0$, define the random tensor $T_k(t)$ by

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

For $\bar{\sigma} \in \bar{\Lambda}^N$, define the auxiliary Hamiltonian

$$(5.1) \quad H_{N,t,\bar{\beta}}^A(\bar{\sigma}) = \frac{\sqrt{t}}{N^{(p-1)/2}} \sum_{r=1}^k \beta_r \langle T_k(t), \sigma(r)^{\otimes p} \rangle - \frac{t}{2} \sum_{r,r'=1}^k \beta_r \beta_{r'} N R(\sigma(r), \sigma(r'))^p.$$

For $t \geq 0$, define the auxiliary free energy and Gibbs measure by

$$F_N^A(t) = \frac{1}{N} \log \int e^{H_{N,t}^A(\bar{\sigma})} \bar{\mu}^{\otimes N}(d\bar{\sigma}) \quad \text{and} \quad G_{N,t}^A(d\bar{\sigma}) = \frac{e^{H_{N,t}^A(\bar{\sigma})} \bar{\mu}^{\otimes N}(d\bar{\sigma})}{\int e^{H_{N,t}^A(\bar{\sigma}')} \bar{\mu}^{\otimes N}(d\bar{\sigma}')}.$$

Denote by $\bar{\sigma}^1, \bar{\sigma}^2, \dots$ the i.i.d. samplings from $G_{N,t}^A$ and by $\langle \cdot \rangle_t^A$ the Gibbs expectation with respect to $G_{N,t}^A$.

A key observation here is that the distribution of $(u(1), \dots, u(k))$ conditionally on $T_k(t)$ is described by the Gibbs measure,

$$(5.2) \quad \mathbb{P}((u(1), \dots, u(k)) \in \cdot | T_k(t)) = G_{N,t}^A(\cdot).$$

To see this equation, one uses the fact that W is Gaussian and is independent of $u(r)$'s and then express the joint density of $u(1), \dots, u(k), T_k(t)$ in terms of the Gaussian density, see, for example, [20]. As a consequence, (5.2) implies the so-called Nishimori identity, namely,

$$(5.3) \quad \langle f(\bar{\sigma}^1, \dots, \bar{\sigma}^n, u(1), \dots, u(k)) \rangle_t^A = \langle f(\bar{\sigma}^1, \dots, \bar{\sigma}^n, \bar{\sigma}^{n+1}) \rangle_t^A$$

for any bounded measurable function f . One may find more general settings, for instance, in [38].

Consider the following auxiliary minimum mean square error

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

where the minimum is taken over all $\mathbb{R}^{N^p}$ -valued bounded random variables $\hat{\theta} = (\hat{\theta}_{i_1, \dots, i_p})$ that are generated by the σ -field $\sigma(T_k(t))$ and are allowed to depend on other randomness independent of both $u_i(r)$'s and T_k . The following lemma summarizes some key properties of $\mathbb{E} F_N^A(t)$ and relates the auxiliary minimum mean square error to the derivative of the free energy. These were originally discovered in [29, 63]. For completeness, we present their proofs here.

LEMMA 5.1. *The following statements hold:*

- (i) $\mathbb{E} F_N^A(t)$ is a nondecreasing, nonnegative, and convex function of t .
- (ii) $\frac{d}{dt} \mathbb{E} F_N^A(t) = \frac{1}{2} \sum_{r,r'=1}^k \beta_r \beta_{r'} \mathbb{E} \langle R(\sigma(r), u(r'))^p \rangle_t^A$.
- (iii) $\text{MMSE}_N^A(\bar{\beta}, t) = \sum_{r,r'=1}^k \beta_r \beta_{r'} \mathbb{E} R(u(r), u(r'))^p - 2 \frac{d}{dt} \mathbb{E} F_N^A(t)$.

PROOF. Using Gaussian integration by parts implies

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

From (5.3), (ii) follows. To establish (iii), note that the minimizer of MMSE_N^A is attained by the estimator

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

where the second equality used (5.3). Plugging this estimator into $\text{MMSE}_N^A(\bar{\beta}, t)$ and applying (5.3) yield (iii).

Finally, we prove (i). Note that setting $\hat{\theta}_{i_1, \dots, i_p} \equiv 0$ gives the upper bound

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

Combining this with (iii) shows that $\frac{d}{dt} \mathbb{E} F_N^A(t)$ is nonnegative, so $\mathbb{E} F_N^A(t)$ is nondecreasing in t . In addition, since $F_N^A(0) = 0$, we conclude that $\mathbb{E} F_N^A(t)$ is nonnegative. To establish the convexity of $\mathbb{E} F_N^A$ in t , from (iii) it suffices to show that $\text{MMSE}_N^A(\bar{\beta}, t)$ is nonincreasing in t . For any $0 \leq t < t'$, write

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

where W' is an independent copy of W and is also independent of $u(1), \dots, u(k)$. Write

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

It follows that

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

This establishes (i) and completes our proof. $\square$

5.2. Proof of Theorem 2.5. We prove Theorem 2.5(i) first. Assume that $\bar{\beta} \in (0, \beta_{1,c}) \times \cdots \times (0, \beta_{k,c})$. From Theorem 2.3, $d_{\text{TV}}(W, T_k) \rightarrow 0$. Note that

$$d_{\text{TV}}(W, T_k) = \int_0^1 \mathbb{P}(F_N(\bar{\beta}) < N^{-1} \log x) dx = \int_0^1 \mathbb{P}(F_N^A(1) > -N^{-1} \log x) dx.$$

Here the first equality is from Lemma 4.2, while the second equality follows from a similar argument as that for Lemma 4.2 by using the second equality in Lemma 4.1. By Fatou's lemma and the above display,

$$\liminf_{N \rightarrow \infty} \mathbb{P}(F_N^A(1) > -N^{-1} \log x) \rightarrow 0, \quad x \in (0, 1)$$

and consequently, $\limsup_{N \rightarrow \infty} \mathbb{P}(B_N(\varepsilon)) = 1$ for all $\varepsilon \in (0, 1)$, where $B_N(\varepsilon) := \{F_N^A(1) \leq \varepsilon\}$. Therefore, from Hölder's inequality,

$$\begin{aligned} \mathbb{E} F_N^A(1) &= \mathbb{E}[F_N^A(1); B_N(\varepsilon)] + \mathbb{E}[F_N^A(1); B_N(\varepsilon)^c] \\ &\leq \varepsilon + (\mathbb{E} F_N^A(1)^2)^{1/2} (\mathbb{P}(B_N(\varepsilon)^c))^{1/2}. \end{aligned}$$

Note that since $\mu_1, \dots, \mu_k$ are defined on bounded sets, one can verify that the second moment of the random variable $F_N^A(1)$ is bounded in N . As a result,

$$\limsup_{N \rightarrow \infty} \mathbb{E} F_N^A(1) \leq 0.$$

From Lemma 5.1(i), we then conclude that

$$\lim_{N \rightarrow \infty} \mathbb{E} F_N^A(t) = 0 \quad \forall t \in [0, 1].$$

Now the convexity of $\mathbb{E} F_N^A$ implies that $\lim_{N \rightarrow \infty} \frac{d}{dt} \mathbb{E} F_N^A(t) = 0$ for $t \in [0, 1]$. From Lemma 5.1(ii) and (iii) and the strong law of large numbers, Theorem 2.5(i) follows.

Next, we assume that $\bar{\beta} \notin (0, \beta_{1,c}] \times \cdots \times (0, \beta_{k,c}]$. For $s \in [0, 1]$, define an interpolating free energy by

$$F_N^I(s) = \frac{1}{N} \log \int \exp \left(H_{N, \bar{\beta}}(\bar{\sigma}) + s \sum_{r, r'=1}^k \beta_r \beta_{r'} N R(\sigma(r), u(r'))^p \right) \bar{\mu}^{\otimes N}(d\bar{\sigma}).$$

Note that when $t = 1$,

$$H_{N,t}^A(\bar{\sigma}) = H_{N, \bar{\beta}}(\bar{\sigma}) + \sum_{r, r'=1}^k \beta_r \beta_{r'} N R(\sigma(r), u(r'))^p.$$

This implies that $F_N^I(1) = F_N^A(1)$ and $F_N^I(0) = F_N(\bar{\beta})$. In addition, from Lemma 5.1(ii), and the convexity of $F_N^I(s)$,

$$(5.4) \quad \frac{d}{dt} \mathbb{E} F_N^A(1) = \frac{1}{2} \frac{d}{ds} \mathbb{E} F_N^I(1) \geq \frac{1}{2} \frac{d}{ds} \mathbb{E} F_N^I(s) \quad \forall s \in [0, 1].$$

Note that since $\mathbb{E} F_N^I$ is a family of equicontinuous and convex functions, one can pass to a subsequence $(N_n)_{n \geq 1}$ via a diagonalization procedure to show that $\mathbb{E} F_{N_n}^I$ is pointwise convergent along this subsequence. Furthermore, we can ensure that along this subsequence,

$$\lim_{n \rightarrow \infty} \frac{d}{dt} \mathbb{E} F_{N_n}^I(1) = \liminf_{N \rightarrow \infty} \frac{d}{dt} \mathbb{E} F_N^A(1).$$

Denote $F^I = \lim_{n \rightarrow \infty} \mathbb{E} F_{N_n}^I$. Note that on the one hand, $\mathbb{E} F_N^A(1) \geq 0$ by Lemma 5.1(i) and on the other hand, $F(\bar{\beta}) = \limsup_{N \rightarrow \infty} \mathbb{E} F_N(\bar{\beta}) < 0$ by Theorem 3.2. Using the identities $F_N^I(1) = F_N^A(1)$ and $F_N^I(0) = F_N(\bar{\beta})$ yields that $F^I(0) < 0 \leq F^I(1)$. Consequently, there exists some $s_0 \in (0, 1)$ such that F^I is differentiable at this point and

$$\lim_{n \rightarrow \infty} \frac{d}{ds} \mathbb{E} F_{N_n}^I(s_0) = \frac{d}{ds} F^I(s_0) > 0.$$

This and (5.4) together yield

$$\lim_{n \rightarrow \infty} \frac{d}{dt} \mathbb{E} F_{N_n}^A(1) \geq \frac{1}{2} \frac{d}{ds} F^I(s_0) > 0.$$

Finally, from this inequality, Lemma 5.1(ii) and (iii), and the strong law of large numbers, the assertion of Theorem 2.5(ii) follows.

6. Structure of the regime $\mathcal{R}$. In this section, we establish the proof of Proposition 3.1. It is based on a subtle control of the Parisi formula for the free energy. While a similar argument has appeared in [20] for the case that there is only one spike and it is sampled from the Rademacher prior, our argument here works for more general priors.

6.1. *The Parisi formula.* Recall the probability space (Λ, μ) from Section 3.1. Denote

$$(6.1) \quad \mathcal{V} = \{u^2 : u \in \Lambda\}.$$

Fix $v \in \mathcal{V}$ and let $\mathcal{M}_v$ be the space of probability measures on $[0, v]$. Recall that $\xi(s) = s^p$. For $\alpha \in \mathcal{M}_v$ and $\lambda \in \mathbb{R}$, define the Parisi functional by

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

where $\Phi_{\beta, v, \alpha}(0, 0, \lambda)$ is defined as the weak solution of the following PDE on $[0, v] \times \mathbb{R} \times \mathbb{R}$ (see [33]):

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

with the boundary condition

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

The Parisi formula states that

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

This formula was initially established by Talagrand [59] for the mixture of even p -spin Hamiltonians and $\Lambda = \{-1, 1\}$. Later it was generalized to arbitrary mixtures of pure p -spin Hamiltonians including odd p and any probability space (Λ, μ) with bounded $\Lambda \subset \mathbb{R}$ by Panchenko [49, 52]. The following proposition shows that the limiting free energy $F(\beta)$ can also be expressed as a Parisi-type formula.

PROPOSITION 6.1 (Parisi formula). *For any $\beta > 0$,*

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

where for $(\alpha, \lambda) \in \mathcal{M}_v \times \mathbb{R}$,

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

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

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

For any $\eta > 0$ and $v \in \mathcal{V}$, set $A_\eta(v) = (v - \eta, v + \eta)$. Note that it is already known from [52] that for any $v \in \mathcal{V}$,

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

From this, for any $\delta > 0$, there exist $\eta(v)$ and $N(v)$ such that for any $N \geq N(v)$

$$(6.2) \quad \left| F_N(\beta, A_\eta(v)) - \inf_{\mathcal{M}_v \times \mathbb{R}} \mathcal{Q}_{\beta, v}(\alpha, \lambda) \right| \leq \delta.$$

Note that $\mathcal{V}$ is bounded and that for any $\eta > 0$, $(A_\eta(v) : v \in \mathcal{V})$ forms an open covering for the closure of $\mathcal{V}$. From these, we can pass to a finite covering, $A_\eta(v_j)$ for $1 \leq j \leq n$, such

that (6.2) is valid. From this,

$$\begin{aligned} F_N(\beta, A_\eta(v_j)) &\leq F_N(\beta) \\ &\leq \frac{1}{N} \log \sum_{j=1}^n \exp N F_N(\beta, A_\eta(v_j)) \\ &\leq \frac{\log n}{N} + \max_{1 \leq j \leq n} F_N(\beta, A_\eta(v_j)) \end{aligned}$$

and hence, as long as N is large enough,

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

Thus,

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

This completes our proof by letting $\delta \downarrow 0$ and noting that $\inf_{\mathcal{M}_v \times \mathbb{R}} \mathcal{Q}_{\beta, v}(\alpha, \lambda)$ is continuous in v . $\square$

6.2. Two technical lemmas. Recall Γ_b from (2.4) and γ_b from (2.5). The following technical inequality establishes the strict monotonicity of γ_b in the temperature parameter b . This will be of great importance for the rest of this section as well as in Section 7.

LEMMA 6.1. *If $0 < \beta < \beta'$, then $\gamma_\beta(s) < \gamma_{\beta'}(s)$ for all $s > 0$.*

PROOF. Note that $\gamma_\beta(0) = 0$. Let B_t be a standard Brownian motion. Define

$$g_j(t, x) = \int a^j e^{ax - \frac{a^2 t}{2}} \mu(da) \quad \forall j = 0, 1, 2, 3.$$

Set $X_t = g_1(t, B_t)^2$ and $Y_t = g_0(t, B_t)^{-1}$. Note that $\gamma_\beta(s) = \mathbb{E} X_t Y_t$ if we let $t = \beta^2 \xi'(s)$. From Itô's formula,

$$\begin{aligned} dX_t &= 2g_1 \partial_t g_1 dt + 2g_1 \partial_x g_1 dB_t + (g_1 \partial_{xx} g_1 + (\partial_x g_1)^2) dt \\ &= -g_1 g_3 dt + 2g_1 g_2 dB_t + (g_1 g_3 + (g_2)^2) dt = (g_2)^2 dt + 2g_1 g_2 dB_t \end{aligned}$$

and

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

Now from the product rule,

$$\begin{aligned} d(X_t Y_t) &= X_t dY_t + Y_t dX_t + d\langle X_t, Y_t \rangle \\ &= g_1^2 \left(\frac{g_1^2}{g_0^3} dt - \frac{g_1}{g_0^2} dB_t \right) + g_0^{-1} ((g_2)^2 dt + 2g_1 g_2 dB_t) - \frac{2g_1^2 g_2}{g_0^2} dt \end{aligned}$$

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

where $\langle \cdot, \cdot \rangle$ is the quadratic variation and the third equality follows from the identity

$$\frac{g_1^4}{g_0^3} + \frac{g_2^2}{g_0} - \frac{2g_1^2 g_2}{g_0^2} = g_0 \left(\frac{g_1^4}{g_0^4} + \frac{g_2^2}{g_0^2} - 2 \frac{g_1^2 g_2}{g_0^2 g_0} \right) = g_0 \left(\frac{g_1^2}{g_0^2} - \frac{g_2}{g_0} \right)^2.$$

From this, we conclude that $X_t Y_t$ is a submartingale and thus $\mathbb{E} X_t Y_t \leq \mathbb{E} X_{t'} Y_{t'}$ for any $0 \leq t < t'$.

If equality holds for some $0 \leq t < t'$, then

$$\int_t^{t'} \mathbb{E} \left[g_0(s, B_s) \left(\frac{g_1(s, B_s)^2}{g_0(s, B_s)^2} - \frac{g_2(s, B_s)}{g_0(s, B_s)} \right)^2 \right] ds = \mathbb{E} X_{t'} Y_{t'} - \mathbb{E} X_t Y_t = 0.$$

This implies that

$$\left(\frac{\int_{\Lambda} a e^{a B_s - \frac{a^2 s}{2}} \mu(da)}{\int_{\Lambda} e^{a B_s - \frac{a^2 s}{2}} \mu(da)} \right)^2 = \frac{g_1(s, B_s)^2}{g_0(s, B_s)^2} = \frac{g_2(s, B_s)}{g_0(s, B_s)} = \frac{\int_{\Lambda} a^2 e^{a B_s - \frac{a^2 s}{2}} \mu(da)}{\int_{\Lambda} e^{a B_s - \frac{a^2 s}{2}} \mu(da)}$$

for all $t \leq s \leq t'$. From this, the necessary condition for obtaining equality in Jensen's inequality implies that

$$a' = \frac{\int_{\Lambda} a e^{a B_s - \frac{a^2 s}{2}} \mu(da)}{\int_{\Lambda} e^{a B_s - \frac{a^2 s}{2}} \mu(da)} \quad \forall a' \in \Lambda.$$

The above equation implies that Λ consists of a single element, which contradicts the assumption that μ is centered and Λ contains more than one element. Therefore, $\mathbb{E} X_t Y_t < \mathbb{E} X_{t'} Y_{t'}$ for any $0 \leq t < t'$. Finally, for $s > 0$ and $0 \leq \beta < \beta'$, plugging $t = \beta^2 \xi'(s)$ and $t' = \beta'^2 \xi'(s)$ into this inequality yields $\gamma_{\beta}(s) < \gamma_{\beta'}(s)$. $\square$

Recall the constant v_* from (2.3). Set the parameter

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

Recall the Parisi formula from Proposition 6.1. For any $v \in [0, v_*]$, define $\alpha_v \in \mathcal{M}_v$ by $\alpha_v(s) = 1$ for $s \in [0, v]$. The next lemma studies some variational properties of the functional $\mathcal{Q}_{\beta, v}$ defined in Proposition 6.1.

LEMMA 6.2. *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 $\inf_{\lambda} \mathcal{Q}_{\beta, v}(\alpha_v, \lambda) = 0$ and λ_* is a minimizer.*

PROOF. Note that

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

where z is standard Gaussian. Consequently,

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

where the right-hand side is obtained through a change of variable $\lambda \mapsto \lambda - \beta^2 \xi'(v)/2$. Define

$$F(v, \lambda) = \log \int \exp(\lambda a^2) \mu(da) - \lambda v.$$

Note that Hölder's inequality implies that $F(v, \cdot)$ is convex. If $v = v_*$, then $\partial_{\lambda} F(v, 0) = 0$ and thus $\lambda = 0$ is a minimizer of $F(v, \cdot)$. Recalling the substitution $\lambda \mapsto \lambda - \beta^2 \xi'(v)/2$, this means that property (ii) holds. To show (i), note that $F(v, 0) = 0$. If $v \neq v_*$, then $\partial_{\lambda} F(v, 0) = v_* - v \neq 0$. This means that 0 is not a minimizer of $F(v, \cdot)$ and therefore $\inf_{\lambda} F(v, \lambda) < 0$. This implies property (i) and completes our proof. $\square$

6.3. Proof of Proposition 3.1. First, we prove that for $\beta > 0$, $\beta \in \mathcal{R}$ if and only if $\sup_{v \in (0, v_*]} \Gamma_{\beta}(v) \leq 0$. Let $\beta \in \mathcal{R}$. From Proposition 6.1,

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

From Lemma 6.2(i), we see that for any $v \neq v_*$,

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

which implies that

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

From this and Lemma 6.2(ii), we conclude that $(\alpha_{v_*}, \lambda_*)$ is an optimizer of $\mathcal{Q}_{\beta, v_*}$. Now we use this conclusion to show that β must satisfy $\sup_{v \in (0, v_*]} \Gamma_{\beta}(v) \leq 0$ as follows. Note 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 in (x, λ) , an argument identical to that in [2] yields that $(\alpha, \lambda) \in \mathcal{M}_{v_*} \times \mathbb{R} \mapsto \mathcal{Q}_{\beta, v_*}(\alpha, \lambda)$ is a convex functional. For any $(\alpha, \lambda) \in \mathcal{M}_{v_*} \times \mathbb{R}$ and $\theta \in [0, 1]$, set

$$\alpha_{\theta} := (1 - \theta)\alpha_{v_*} + \theta\alpha \quad \text{and} \quad \lambda_{\theta} := (1 - \theta)\lambda_* + \theta\lambda.$$

The directional derivative of $\mathcal{Q}_{\beta, v_*}$ at $(\alpha_{v_*}, \lambda_*)$ can be computed as (see, e.g., [19], Theorem 2, and the derivation of (7.2) below),

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

where the derivative is from the right-hand side of 0. As a result, the optimality of $(\alpha_{v_*}, \lambda_*)$ implies that the last line of the above display is nonnegative. Write

$$\begin{aligned} &\int_0^{v_*} \xi''(s) (\alpha(s) - \alpha_{v_*}(s)) (\gamma_{\beta}(s) - s) ds \\ &= \int_0^{v_*} \int_0^s \xi''(s) (\gamma_{\beta}(s) - s) \alpha(da) ds - \int_0^{v_*} \xi''(s) (\gamma_{\beta}(s) - s) ds \\ &= \int_0^{v_*} \left(\int_a^{v_*} \xi''(s) (\gamma_{\beta}(s) - s) ds \right) \alpha(da) - \int_0^{v_*} \xi''(s) (\gamma_{\beta}(s) - s) ds. \end{aligned}$$

From this, the optimality of $(\alpha_{v_*}, \lambda_*)$ is equivalent to

$$\int_v^{v_*} \xi''(s)(\gamma_\beta(s) - s) ds \geq \int_0^{v_*} \xi''(s)(\gamma_\beta(s) - s) ds \quad \forall v \in [0, v_*],$$

and hence, this is also equivalent to $\Gamma_\beta(v) \leq 0$ for all $v \in [0, v_*]$. Conversely, if $\Gamma_\beta(v) \leq 0$ for all $v \in [0, v_*]$, then this inequality implies that the above directional derivative of $\mathcal{Q}_{\beta, v_*}$ is nonnegative. This means that $(\alpha_{v_*}, \lambda_*)$ is an optimizer of the variational problem $\inf_{\alpha, \lambda} \mathcal{Q}_{\beta, v_*}(\alpha, \lambda)$ and $\inf_{\alpha, \lambda} \mathcal{Q}_{\beta, v_*}(\alpha, \lambda) = \mathcal{Q}_{\beta, v_*}(\alpha_{v_*}, \lambda_*) = 0$. From these and Proposition 6.1, we arrive at $0 = \inf_{\alpha, \lambda} \mathcal{Q}_{\beta, v_*}(\alpha, \lambda) \leq F(\beta) \leq 0$ and hence, $F(\beta) = 0$. This establishes the statement that for $\beta > 0$, $\beta \in \mathcal{R}$ if and only if $\sup_{v \in (0, v_*]} \Gamma_\beta(v) \leq 0$.

Finally, the assertion $\mathcal{R} = (0, \beta_c]$ can be established similarly. Clearly, $\mathcal{R} \subseteq (0, \beta_c]$. If $0 < \beta \leq \beta_c$, then Lemma 6.1 and the above proof imply that $\Gamma_\beta(v) \leq \Gamma_{\beta_c}(v) \leq 0$ for all $v \in [0, v_*]$ and thus, $\beta \in \mathcal{R}$. This completes our proof.

7. Overlap concentration with exponential tail. Recall the probability space (Λ, μ) , the Gibbs measure $G_{N, \beta}$, and the Gibbs expectation $\langle \cdot \rangle_\beta$ from Section 3.1. The following proposition states that in the high-temperature regime, the overlap of two i.i.d. sampled spin configurations from $G_{N, \beta}$ is concentrated around the origin with overwhelming probability. This result will be essential when we later bound the overlap moments. Let $I(A)$ denote the indicator function of a set A .

PROPOSITION 7.1. *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 the following property holds for any $N \geq 1$ and $s \in [s_0, 1]$: For i.i.d. samplings σ^1 and σ^2 from $G_{N, s\beta}$,*

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

The rest of this section is devoted to proving Proposition 7.1.

7.1. The Guerra–Talagrand bound. Our main tool is the Guerra–Talagrand bound for the coupled free energy that we formulate as follows. Denote by $M_2(\mathbb{R})$ the space of all real-valued 2×2 matrices equipped with the metric

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

For $C, D \in M_2(\mathbb{R})$, denote by $\langle C, D \rangle$ the inner product of C and D , that is, $\langle C, D \rangle = \sum_{i, j=1}^2 C_{ij} D_{ij}$; when $x, y \in \mathbb{R}^2$, denote by $\langle x, y \rangle$ the usual scalar product between x and y . For any $\sigma^1, \sigma^2 \in \Lambda^N$, define the overlap matrix by

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

For any subset $A \subset M_2(\mathbb{R})$, define the coupled free energy restricted to A by

$$CF_N(\beta, A) = \frac{1}{N} \log \int_{\mathbf{R}(\sigma^1, \sigma^2) \in A} e^{H_{N, \beta}(\sigma^1) + H_{N, \beta}(\sigma^2)} \mu^{\otimes N}(d\sigma^1) \mu^{\otimes N}(d\sigma^2).$$

Recall the space $\mathcal{V}$ from (6.1) and recall that $\mathcal{M}_v$ is the set of all probability measures on $[0, v]$. Let $v \in \mathcal{V}$ and $v_0 \in [0, v]$ be fixed. Set

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

Let T be a $M_2(\mathbb{R})$ -valued function on $[0, v]$ defined by

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

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

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

with boundary condition $\Psi_{\beta, V, \alpha}(v, x, \lambda) = \log \int e^{\langle a, x \rangle + \langle \lambda a, a \rangle} (\mu \otimes \mu)(da)$. For the existence of $\Psi_{\beta, V, \alpha}$, we refer the readers to [33]. For $\alpha \in \mathcal{M}_v$ and $\lambda \in M_2(\mathbb{R})$, define

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

Denote $A_\eta(V) = \{V' \in M_2(\mathbb{R}) : \|V - V'\|_{\max} < \eta\}$. The Guerra-Talagrand inequality (see [61]) states that if p is even, then for any $(\alpha, \lambda) \in \mathcal{M}_v \times M_2(\mathbb{R})$,

$$(7.1) \quad \lim_{\eta \downarrow 0} \limsup_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \log \int_{\mathbf{R}(\sigma^1, \sigma^2) \in A_\eta(V)} e^{\beta X_N(\sigma^1) + \beta X_N(\sigma^2)} \mu^{\otimes N}(d\sigma^1) \mu^{\otimes N}(d\sigma^2) \leq \mathcal{P}_{\beta, V}(\alpha, \lambda).$$

When p is odd, the validity of this inequality is an open question. Nevertheless, in the case of the Rademacher prior, that is, $\mu = \delta_1/2 + \delta_{-1}/2$, the work [20] proved that for odd p , (7.1) remains valid if $\alpha \in \mathcal{M}_{v, v_0}$ for $0 \leq v_0 \leq v$, where

$$\mathcal{M}_{v, v_0} := \{\alpha \in \mathcal{M}_v : \alpha \text{ is a fixed constant on } [0, v_0) \text{ and } \alpha(s) = 1 \text{ on } [v_0, v]\}.$$

In view of the proof in [20], the argument does not rely on the measure μ in an essential way and it is applicable to the current general setting so that (7.1) remains valid for odd p and $(\alpha, \lambda) \in \mathcal{M}_{v, v_0} \times M_2(\mathbb{R})$. In the Appendix, we present a sketch of the proof for this inequality. Now substituting the overlap term in $H_{N, \beta}$ via the restriction $A_\eta(V)$ in $CF_N(\beta, A_\eta(V))$ yields

PROPOSITION 7.2 (Guerra-Talagrand Bound). *For any $p \geq 2$, $v \in \mathcal{V}$, $v_0 \in [0, v]$, $\lambda \in M_2(\mathbb{R})$ and $\alpha \in \mathcal{M}_{v, v_0}$,*

$$\lim_{\eta \downarrow 0} \limsup_{N \rightarrow \infty} \mathbb{E} CF_N(\beta, A_\eta(V)) \leq \mathcal{Q}_{\beta, V}(\alpha, \lambda) := \mathcal{P}_{\beta, V}(\alpha, \lambda) - \beta^2 v^p.$$

7.2. Proof of Proposition 7.1. Before proving Proposition 7.1, we need a lemma, which will also be used later in Section 9.

LEMMA 7.1. *Assume that $\beta \in \mathcal{R}$. For any $s_0 \in (0, 1]$ and $\varepsilon > 0$, there exists a constant $K > 0$ independent of N such that*

$$\mathbb{E} \langle I(|R(\sigma, \sigma) - v_*| \geq \varepsilon) \rangle_{s\beta} \leq K e^{-N/K} \quad \forall s \in [s_0, 1], N \geq 1,$$

where σ is a sampling from $G_{N, s\beta}$.

PROOF. From Jensen's inequality and $\mathbb{E} e^{H_{N, s\beta}}(\sigma) = 1$,

$$\begin{aligned} & \limsup_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \log \int_{R(\sigma, \sigma) \notin (v_* - \varepsilon, v_* + \varepsilon)} e^{H_{N, s\beta}(\sigma)} \mu^{\otimes N}(d\sigma) \\ & \leq \limsup_{N \rightarrow \infty} \frac{1}{N} \log \int_{R(\sigma, \sigma) \notin (v_* - \varepsilon, v_* + \varepsilon)} \mu^{\otimes N}(d\sigma). \end{aligned}$$

Observe that in the second integral $\sigma = (\sigma_1, \dots, \sigma_N)$ are i.i.d. random variables with respect to the measure μ , which has a bounded support and variance v_* . Using Cramér's theorem (see, e.g., [21], Theorem 2.2.3), there exists a $\delta > 0$ such that

$$\int_{R(\sigma, \sigma) \notin (v_* - \varepsilon, v_* + \varepsilon)} \mu^{\otimes N}(d\sigma) \leq e^{-N\delta} \quad \forall N \geq 1.$$

Hence, from the above inequalities and Proposition 3.1,

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \log \int_{R(\sigma, \sigma) \notin (v_* - \varepsilon, v_* + \varepsilon)} e^{H_{N, s\beta}(\sigma)} \mu^{\otimes N}(d\sigma) \leq -\delta = F(s\beta) - \delta \quad \forall s \in [s_0, 1].$$

Next, by using Gaussian integration by parts, one can compute the derivatives in the s variable for

$$\frac{1}{N} \mathbb{E} \log \int_{R(\sigma, \sigma) \notin (v_* - \varepsilon, v_* + \varepsilon)} e^{H_{N, s\beta}(\sigma)} \mu^{\otimes N}(d\sigma) \quad \text{and} \quad \mathbb{E} F_N(s\beta)$$

to show that the resulting derivatives are uniformly bounded over $s \in [s_0, 1]$. As a consequence, these two sequences of functions are equicontinuous. From this and the above inequality, there exists some $N_0 \geq 1$ such that

$$\frac{1}{N} \mathbb{E} \log \int_{R(\sigma, \sigma) \notin (v_* - \varepsilon, v_* + \varepsilon)} e^{H_{N, s\beta}(\sigma)} \mu^{\otimes N}(d\sigma) \leq \mathbb{E} F_N(s\beta) - \frac{\delta}{2} \quad \forall s \in [s_0, 1], N \geq N_0.$$

Finally, from this inequality and the Gaussian concentration inequality for $F_N(s\beta)$ and its restricted free energy, there exists a universal constant $K > 0$ such that for any $s \in [s_0, 1]$ and $N \geq N_0$, with probability at least $1 - Ke^{-N/K}$,

$$\frac{1}{N} \log \int_{R(\sigma, \sigma) \notin (v_* - \varepsilon, v_* + \varepsilon)} e^{H_{N, s\beta}(\sigma)} \mu^{\otimes N}(d\sigma) \leq F_N(s\beta) - \frac{\delta}{4},$$

which implies that

$$\mathbb{E} \langle I(|R(\sigma, \sigma) - v_*| \geq \varepsilon) \rangle_{s\beta} \leq e^{-N\delta/4} + Ke^{-N/K}.$$

This completes our proof. $\square$

Now we establish the proof of Proposition 7.1. Let $0 < \beta < \beta_c$, $0 < \varepsilon < v_*$, and $s_0 \in (0, 1)$ be fixed. Let $v = v_*$. Suppose that $v_0 \in [\varepsilon, v]$. For $s \in [0, 1]$, denote $\beta_s = s\beta$. Let $\lambda \in M_2(\mathbb{R})$ with $\lambda_{1,1} = \lambda_{2,2} = -\beta_s^2 \xi'(v_*)/2$ and $\lambda_{1,2} = \lambda_{2,1} = 0$. Let $\alpha \in \mathcal{M}_{v, v_0}$ satisfy $\alpha \equiv 0$ on $[0, v_0]$, and $\alpha \equiv 1$ on $[v_0, v]$. For $\theta \in [0, 1]$, set

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

Using the Cole–Hopf transformation, one can compute that

$$\begin{aligned} \mathcal{Q}_{\beta_s, v}(\alpha_\theta, \lambda) &= \frac{2}{1-\theta} \log \mathbb{E}[g_0(\beta_s^2 \xi'(v_0), \beta_s \xi'(v_0)^{1/2} z)^{1-\theta}] + \beta_s^2 \xi'(v) v \\ &\quad - \beta_s^2 \left((1-\theta) \int_0^{v_0} \xi''(r) r \, dr + \int_{v_0}^v \xi''(r) r \, dr \right) - \beta_s^2 v^p, \end{aligned}$$

where

$$g_0(t, x) := \int e^{ax - a^2 t/2} \mu(da)$$

and z is a standard normal random variable. A direct differentiation in θ yields

$$\partial_\theta \mathcal{Q}_{\beta_s, V}(\alpha_\theta, \lambda)|_{\theta=0} = -2\mathbb{E}g(t, \sqrt{t}z) \log g(t, \sqrt{t}z) + \beta_s^2 \int_0^{v_0} \xi''(s) ds, \quad t = \beta_s^2 \xi'(v_0).$$

To handle this equation, we use Gaussian integration by parts to get

$$\begin{aligned} & \frac{d}{dt} \mathbb{E}g(t, \sqrt{t}z) \log g(t, \sqrt{t}z) \\ &= \mathbb{E} \partial_t g(t, \sqrt{t}z) (\log g(t, \sqrt{t}z) + 1) \\ & \quad + \frac{1}{2} \left(\mathbb{E} \partial_{xx} g(t, \sqrt{t}z) (\log g(t, \sqrt{t}z) + 1) + \mathbb{E} \frac{\partial_x g(t, \sqrt{t}z)^2}{g(t, \sqrt{t}z)} \right) \\ &= -\frac{1}{2} \mathbb{E} \partial_{xx} g(t, \sqrt{t}z) (\log g(t, \sqrt{t}z) + 1) \\ & \quad + \frac{1}{2} \left(\mathbb{E} \partial_{xx} g(t, \sqrt{t}z) (\log g(t, \sqrt{t}z) + 1) + \mathbb{E} \frac{\partial_x g(t, \sqrt{t}z)^2}{g(t, \sqrt{t}z)} \right) \\ &= \frac{1}{2} \mathbb{E} \frac{\partial_x g(t, \sqrt{t}z)^2}{g(t, \sqrt{t}z)}, \end{aligned}$$

where the last equality used the observation $\partial_t g = -\partial_{xx} g/2$. Consequently,

$$\begin{aligned} & \partial_\theta \mathcal{Q}_{\beta_s, V}(\alpha_\theta, \lambda)|_{\theta=0} \\ &= 2\beta_s^2 \int_0^{v_0} \frac{1}{2} \xi''(r) \mathbb{E} \frac{\partial_x g(\beta_s^2 \xi'(r), \beta_s \sqrt{\xi'(r)}z)^2}{g(\beta_s^2 \xi'(r), \beta_s \sqrt{\xi'(r)}z)} dr - \beta_s^2 \int_0^{v_0} \xi''(s) ds \\ (7.2) \quad &= \beta_s^2 \int_0^{v_0} \xi''(r) (\gamma_{\beta_s}(r) - 1) dr \\ &< \beta_s^2 \int_0^{v_0} \xi''(r) (\gamma_{\beta_c}(r) - 1) dr = \beta_s^2 \Gamma_{\beta_c}(v_0) \leq 0 \end{aligned}$$

for all $0 < s \leq 1$, where the strict inequality used the monotonicity of the function $\gamma_b(r)$ in b by Lemma 6.1 and the last inequality used Proposition 3.1. Now since $\mathcal{Q}_{\beta_s, V}(\alpha_\theta, \lambda)$ is a continuous function in (s, v_0, θ) and $\mathcal{Q}_{\beta_s, V}(\alpha_\theta, \lambda)|_{\theta=0} = 0$, there exist $\delta, \delta' > 0$ such that

$$\sup_{s \in [s_0, 1]} \sup_{v_0 \in [\varepsilon, v + \delta']} \inf_{\theta \in [0, 1]} \mathcal{Q}_{\beta_s, V}(\alpha_\theta, \lambda) \leq -\delta.$$

Consequently, from Proposition 7.2, we see that the coupled free energy exhibits a free energy cost, that is, for any $v_0 \in [\varepsilon, v + \delta']$ and $s \in [s_0, 1]$,

$$(7.3) \quad \lim_{\eta \downarrow 0} \limsup_{N \rightarrow \infty} \mathbb{E} CF_N(\beta_s, A_\eta(V)) \leq -\delta.$$

Set

$$\begin{aligned} A^+ &= \{V \in M_2(\mathbb{R}) : V \text{ positive semi-definite with} \\ & \quad V_{12} = V_{21} \geq \varepsilon, V_{11}, V_{22} \in [v - \delta', v + \delta']\}, \\ A^- &= \{V \in M_2(\mathbb{R}) : V \text{ positive semi-definite with} \\ & \quad V_{12} = V_{21} \leq -\varepsilon, V_{11}, V_{22} \in [v - \delta', v + \delta']\}. \end{aligned}$$

Using the inequality

$$\log(x_1 + \cdots + x_k) \leq \log k + \max_{1 \leq \ell \leq k} \log x_\ell \quad \forall x_1, \dots, x_k > 0,$$

we see from the compactness of A^+ and (7.3) that for any $s \in [s_0, 1]$,

$$(7.4) \quad \limsup_{N \rightarrow \infty} \mathbb{E}CF_N(\beta_s, A^+) \leq -\frac{\delta}{2}.$$

To obtain the same inequality for A^- , we note that when p is even, $H_{N,\beta}(\sigma) = H_{N,\beta}(-\sigma)$, which implies that $\mathbb{E}CF_N(\beta_s, A^+) = \mathbb{E}CF_N(\beta_s, A^-)$ and thus,

$$\limsup_{N \rightarrow \infty} \mathbb{E}CF_N(\beta_s, A^-) \leq -\frac{\delta}{2}.$$

When p is an odd number, Jensen's inequality yields that

$$\begin{aligned} \limsup_{N \rightarrow \infty} \mathbb{E}CF_N(\beta_s, A^-) &\leq \limsup_{N \rightarrow \infty} \frac{1}{N} \log \int_{\mathbf{R}(\sigma^1, \sigma^2) \in A^-} e^{\beta^2 N R(\sigma^1, \sigma^2)^p} \mu^{\otimes N}(\sigma^1) \mu^{\otimes N}(\sigma^2) \\ &< -\beta^2 \varepsilon^p < 0. \end{aligned}$$

Combining these and Proposition 3.1 together implies that there exists a constant $\delta'' > 0$ such that for all $s \in [s_0, 1]$,

$$(7.5) \quad \limsup_{N \rightarrow \infty} \mathbb{E}CF_N(\beta_s, A^+ \cup A^-) < -\delta'' = 2F(\beta_s) - \delta''.$$

The rest of the proof follows essentially in the same way as that for Lemma 7.1. By computing the derivatives of $\mathbb{E}CF_N(\beta_s, A^+ \cup A^-)$ and $\mathbb{E}F_N(\beta_s)$ in s and using Gaussian integration by parts, it can be checked that these derivatives are uniformly bounded over $s \in [s_0, 1]$. Hence, $s \mapsto \mathbb{E}CF_N(\beta_s, A^+ \cup A^-)$ and $s \mapsto \mathbb{E}F_N(\beta_s)$ are two families of equicontinuous functions. From this and (7.5), there exists some $N_0 \geq 1$ such that as long as $N \geq N_0$ and $s \in [s_0, 1]$,

$$\mathbb{E}CF_N(\beta_s, A^+ \cup A^-) \leq 2\mathbb{E}F_N(\beta_s) - \frac{\delta''}{2}.$$

From the Gaussian concentration inequality for $CF_N(\beta_s, A^+ \cup A^-)$ and $F_N(\beta_s)$, there exists a constant $K > 0$ such that for any $N \geq N_0$ and $s \in [s_0, 1]$, with probability at least $1 - Ke^{-N/K}$,

$$CF_N(\beta_s, A^+ \cup A^-) \leq 2F_N(\beta_s) - \frac{\delta''}{4},$$

which leads to

$$\langle I(\mathbf{R}(\sigma^1, \sigma^2) \in A^+ \cup A^-) \rangle_{\beta_s} = e^{N(CF_N(\beta_s, A^+ \cup A^-) - 2F_N(\beta_s))} \leq e^{-\frac{\delta''N}{4}}.$$

By taking the expectation, there exists a constant $K' > 0$ such that for any $N \geq N_0$ and $s \in [s_0, 1]$,

$$(7.6) \quad \mathbb{E}\langle I(\mathbf{R}(\sigma^1, \sigma^2) \in A^+ \cup A^-) \rangle_{\beta_s} \leq K'e^{-N/K'}.$$

Finally, from Lemma 7.1, there exists some $K'' > 0$ such that for any $N \geq 1$, $s \in [s_0, 1]$, and $\ell = 1, 2$,

$$\mathbb{E}\langle I(R(\sigma^\ell, \sigma^\ell) \notin (v - \delta', v + \delta')) \rangle_{\beta_s} \leq K''e^{-N/K''}.$$

This together with (7.6) implies that for any $N \geq N_0$ and $s \in [s_0, 1]$,

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

and this completes the proof of Proposition 7.1.

8. Overlap concentration with moment control. As we have seen in Proposition 7.1, overlaps between i.i.d. samples of $G_{N,\beta}$ are concentrated around the origin with exponential tail control. The aim of this section is to establish the proof of Theorem 3.1, namely, the moment control of the overlap. The proof is based on the so-called cavity method in mean field spin glasses. As an immediate consequence of Theorem 3.1, we also present the proof of Proposition 3.2.

Briefly speaking, the cavity method is an induction argument that compares the systems of sizes N and $N - 1$ by parameterizing an interpolating path between the two systems and controlling the derivative in the parameter along this path. This technique is a very well-known tool in the physics literature, see [41]. Mathematically, it was implemented in the study of the high-temperature behavior for a number of mean field spin glass models by Talagrand [60]. For technical reasons, most of the existing results in [60] are valid only for a sub-region of the high-temperature regime and not up to the critical temperature. In the present paper, by adapting the argument in [61], Chapter 13, and [20], it turns out that from our understanding of the structure of the high-temperature regime $\mathcal{R}$ as well as the Parisi variational formula for the marginal free energy, we can show that the cavity method can indeed be applied throughout the entire high-temperature regime and ultimately it leads to the asserted moment control of the overlaps.

Before turning to the proof, we set some notation. For any $\ell, \ell' \geq 1$ and $1 \leq r, r' \leq k$, denote by $\sigma, \sigma^1, \sigma^2, \dots$ the spin configurations from Λ^N . Set overlaps

$$R_{\ell, \ell'} = R(\sigma^\ell, \sigma^{\ell'}), \quad R^- = \frac{1}{N} \sum_{i=1}^{N-1} \sigma_i \sigma_i, \quad R_{\ell, \ell'}^- = \frac{1}{N} \sum_{i=1}^{N-1} \sigma_i^\ell \sigma_i^{\ell'}.$$

In Section 8.1, we devise an interpolating system that connects the model of sizes $N - 1$ and N . Section 8.2 computes and bounds the derivative of the expectations of functions of the replicas sampled from the interpolating Gibbs measure along our interpolation. Additionally, this subsection presents some lemmas that are simple yet necessary to bound various powers of overlaps. These results are used in Section 8.3, where we present the cavity argument to establish an iterative inequality for the moments of the overlaps. Finally, Sections 8.4 and 8.5 prove Theorem 3.1 and Proposition 3.2, respectively.

8.1. Constructing an interpolation path. For each $S \subseteq \{1, \dots, p\}$, define I_S as the set of indices $(i_1, \dots, i_p) \in \{1, \dots, N\}^p$ such that $i_s = N$ for all $s \in S$ and $i_s < N$ otherwise. For example, if $p = 4$ and $S = \{1, 3\}$, then $I_S = \{(N, i, N, j) : 1 \leq i, j < N\}$. Define a Gaussian process indexed by S :

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

It is easy to check that

$$\mathbb{E} X_N^S(\sigma^1) X_N^S(\sigma^2) = N (R_{1,2}^-)^{p-|S|} (\sigma_N^1 \sigma_N^2)^{|S|}.$$

Notice that $S = \emptyset$ is the only set such that $X_N^S(\sigma)$ does not involve the last spin σ_N . For $t \in [0, 1]$, define the interpolating Hamiltonian by

$$\begin{aligned} H_{N,\beta,t}(\sigma) = & \beta \left(X_N^\emptyset(\sigma) + \sqrt{t} \sum_{j=1}^p \sum_{S \subset \{1, \dots, p\}: |S|=j} X_N^S(\sigma) \right) \\ & - \frac{\beta^2}{2} \left((R^-)^p + \frac{t}{N^{j-1}} \sum_{j=1}^p \sum_{S \subset \{1, \dots, p\}: |S|=j} (R^-)^{p-j} (\sigma_N \sigma_N)^j \right). \end{aligned}$$

From the binomial formula, one readily checks that when $t = 0$, $H_{N,\beta,0}(\sigma)$ is equal to H_{N-1} at a different temperature:

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

When $t = 1$, $H_{N,\beta,1}(\sigma)$ is simply the original Hamiltonian $H_{N,\beta}(\sigma)$. We define the Gibbs measure associated to $H_{N,\beta,t}$ in the same manner as $G_{N,\beta}$, that is,

$$G_{N,\beta,t}(d\sigma) = \frac{\exp H_{N,\beta,t}(\sigma) \mu^{\otimes N}(d\sigma)}{\int \exp H_{N,\beta,t}(\sigma') \mu^{\otimes N}(d\sigma')}.$$

As before, denote by $(\sigma^\ell)_{\ell \geq 1}$ a sequence of i.i.d. samples from $G_{N,\beta,t}$ and by $\langle \cdot \rangle_{\beta,t}$ the Gibbs average with respect to this sequence. For any bounded measurable function f of the sequence $(\sigma^\ell)_{\ell \geq 1}$, set $v_{\beta,t}(f) = \mathbb{E} \langle f \rangle_{\beta,t}$. When $t = 1$, we simply write $v_\beta(f) = v_{\beta,1}(f)$. We also denote the t -derivative of $v_{\beta,t}(f)$ by $v'_{\beta,t}(f)$.

8.2. Some auxiliary lemmas. We gather some lemmas that will be used in the proof of Theorem 3.1 below. As their proofs are fairly standard, we refer the readers to [60], Chapter 1, or [20]. First, we compute the derivative of $v_{\beta,t}(f)$.

LEMMA 8.1. *For any bounded function f of $\sigma^1, \dots, \sigma^n$, we have*

$$\begin{aligned} v'_{\beta,t}(f) &= \beta^2 \sum_{j=1}^p \binom{p}{j} \frac{1}{N^{j-1}} \left(\sum_{1 \leq \ell < \ell' \leq n} v_{\beta,t}(f(R_{\ell,\ell'}^-)^{p-j} (\sigma_N^\ell \sigma_N^{\ell'})^j) \right. \\ &\quad \left. - n \sum_{\ell \leq n} v_{\beta,t}(f(R_{\ell,n+1}^-)^{p-j} (\sigma_N^\ell \sigma_N^{n+1})^j) \right. \\ &\quad \left. + \frac{n(n+1)}{2} v_{\beta,t}(f(R_{n+1,n+2}^-)^{p-j} (\sigma_N^{n+1} \sigma_N^{n+2})^j) \right). \end{aligned}$$

Note that since Λ is bounded, there exists a constant $M > 1$ such that $\Lambda \subseteq [-M, M]$. The next lemma controls $v_{\beta,t}(f)$ by the terminal value $v_\beta(f)$.

LEMMA 8.2. *For any nonnegative and bounded function f of $\sigma^1, \dots, \sigma^n$, we have*

$$v_{\beta,t}(f) \leq \exp(n^2 2^{p+1} M^{2p} \beta^2) v_\beta(f).$$

In the proof of Theorem 3.1, it will sometimes be desirable to work with the overlaps $R_{1,2}^-$ instead of the overlaps $R_{1,2}$ and vice versa. Lemma 8.3 will allow us to replace $(R_{1,2})^m$ by $(R_{1,2}^-)^m$ (or vice versa). On the other hand, Lemma 8.4 states that we can also control the moments of $R_{1,2}^-$ by the bounds on $R_{1,2}$.

LEMMA 8.3. *For any $m \geq 1$,*

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

LEMMA 8.4. *Let $m \in \mathbb{N}$. Assume that there exists some $K \geq 1$ such that $v_\beta((R_{1,2})^{2j}) \leq K/N^j$ for any $0 \leq j \leq m$. Then*

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

8.3. *Cavity argument.* The following lemma is the key ingredient of our argument. It is obtained via a purely algebraic cavity computation and does not require any fact about the high-temperature behavior of the overlaps.

LEMMA 8.5. *Let m be a nonnegative integer and $\beta > 0$. Assume that there exists a constant $K_0 \geq 1$ such that for all $0 \leq j \leq m$ and $N \geq 1$*

$$v_\beta((R_{1,2})^{2j}) \leq \frac{K_0}{N^j}.$$

Then

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

for all $N \geq 1$, where K_1 and K_2 are two nonnegative continuous functions of β and they are independent of N . In addition, K_1 is nondecreasing with $K_1(0) = 0$ if and only if $\beta = 0$.

PROOF. We divide our proof into four steps.

Step 1: By symmetry between sites, write

$$v_\beta((R_{1,2})^{2m+2}) = v_\beta(\sigma_N^1 \sigma_N^2 (R_{1,2})^{2m+1}) = v_\beta(\sigma_N^1 \sigma_N^2 (R_{1,2}^-)^{2m+1}) + \mathcal{E}.$$

Here,

$$\mathcal{E} := v_\beta(\sigma_N^1 \sigma_N^2 ((R_{1,2})^{2m+1} - (R_{1,2}^-)^{2m+1}))$$

can be controlled by Lemmas 8.3 and 8.4 as follows:

$$\begin{aligned} |\mathcal{E}| &\leq M^2 v(|(R_{1,2})^{2m+1} - (R_{1,2}^-)^{2m+1}|) \\ &\leq \frac{2mM^4}{N} (v_\beta(|R_{1,2}|^{2m}) + v_\beta(|R_{1,2}^-|^{2m})) \leq \frac{C_1}{N^{m+1}}, \end{aligned}$$

where $C_1 := K_0(2mM^4 + m2^{2m+1}M^{4m+4})$. Thus, we arrive at

$$(8.1) \quad v_\beta((R_{1,2})^{2m+2}) \leq v_\beta(\sigma_N^1 \sigma_N^2 (R_{1,2}^-)^{2m+1}) + \frac{C_1}{N^{m+1}}.$$

Next, in order to control the right-hand side, we define $f = \sigma_N^1 \sigma_N^2 (R_{1,2}^-)^{2m+1}$. Recall that since μ is centered, $v_0(f) = 0$. This together with an application of the mean value theorem and (8.1) results in the inequality

$$(8.2) \quad v_\beta((R_{1,2})^{2m+2}) \leq v_\beta(f) + \frac{C_1}{N^{m+1}} \leq \sup_{0 \leq t \leq 1} |v'_{\beta,t}(f)| + \frac{C_1}{N^{m+1}}.$$

Step 2: We control $|v'_{\beta,t}(f)|$. Applying Lemma 8.1 with $n = 2$ and noting that $|\sigma_N^\ell \sigma_N^{\ell'}| \leq M^2$ for any $1 \leq \ell, \ell' \leq n$ yield the following bound on $|v'_{\beta,t}(f)|$:

$$\begin{aligned} &\beta^2 \sum_{j=1}^p \binom{p}{j} \frac{(M^2)^{j+1}}{N^{j-1}} (v_{\beta,t}(|R_{1,2}^-|^{2m+1} |R_{1,2}|^{p-j}) + 2v_{\beta,t}(|R_{1,2}^-|^{2m+1} |R_{1,3}^-|^{p-j}) \\ &\quad + 2v_{\beta,t}(|R_{1,2}^-|^{2m+1} |R_{2,3}^-|^{p-j}) + 3v_{\beta,t}(|R_{1,2}^-|^{2m+1} |R_{3,4}^-|^{p-j})). \end{aligned}$$

For each $1 \leq j \leq p$, set the Hölder conjugate exponents

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

By Hölder's inequality, for each $1 \leq j \leq p$ and each pair of replica indices $1 \leq \ell, \ell' \leq n$,

$$\begin{aligned} v_{\beta,t}(|R_{1,2}^-|^{2m+1}|R_{\ell,\ell'}^-|^{p-j}) &\leq v_{\beta,t}(|R_{1,2}^-|^{2m+1+p-j})^{1/\tau_j^1} v_{\beta,t}(|R_{\ell,\ell'}^-|^{2m+1+p-j})^{1/\tau_j^2} \\ &= v_{\beta,t}(|R_{1,2}^-|^{2m+1+p-j}). \end{aligned}$$

This inequality leads to

$$|v'_{\beta,t}(f)| \leq 8\beta^2 \sum_{j=1}^p \binom{p}{j} \frac{(M^2)^{j+1}}{N^{j-1}} v_{\beta,t}(|R_{1,2}^-|^{2m+1+p-j}).$$

Consequently, Lemma 8.2 allows us to replace $v_{\beta,t}$ by v_β on the right-hand side above to obtain that

$$(8.3) \quad |v'_{\beta,t}(f)| \leq 8e^{2p+3} M^{2p} \beta^2 \sum_{j=1}^p \binom{p}{j} \frac{(M^2)^{j+1}}{N^{j-1}} v_\beta(|R_{1,2}^-|^{2m+1+p-j}).$$

Step 3: We break the above sum into two pieces: When $j = 1$, $v_\beta(|R_{1,2}^-|^{2m+1+p-j}) = v_\beta(|R_{1,2}^-|^{2m+p})$; when $2 \leq j \leq p$, since $M > 1$,

$$v_\beta(|R_{1,2}^-|^{2m+1+p-j}) \leq M^{2(1+p-j)} v_\beta(|R_{1,2}^-|^{2m}) \leq \frac{2^{2m} M^{2(1+p-j+2m)} K_0}{N^m}.$$

The last inequality used the given assumption and Lemma 8.4. Let

$$C(\beta) := 8e^{2p+3} M^{2p} \beta^2, \quad C_2 := pM^4, \quad C_3 := 2^{2m+p} M^{2(2m+p+2)} K_0.$$

From (8.3) and the last two inequalities,

$$|v'_{\beta,t}(f)| \leq C(\beta) C_2 v_\beta(|R_{1,2}^-|^{2m+p}) + \frac{C(\beta) C_3}{N^{m+1}},$$

and subsequently, plugging this into (8.2) gives

$$(8.4) \quad v_\beta((R_{1,2})^{2m+2}) \leq C(\beta) C_2 v_\beta(|R_{1,2}^-|^{2m+p}) + \frac{C(\beta) C_3 + C_1}{N^{m+1}}.$$

Step 4: We may now perform a procedure similar to Step 1 to bring $R_{1,2}^-$ back to $R_{1,2}$. By Lemma 8.3 and $M > 1$,

$$\begin{aligned} v_\beta(|R_{1,2}^-|^{2m+p}) &\leq v_\beta(|R_{1,2}|^{2m+p}) + \frac{2m+p}{N} M^2 (v_\beta(|R_{1,2}^-|^{2m+p-1}) + v_\beta(|R_{1,2}|^{2m+p-1})) \\ &\leq v_\beta(|R_{1,2}|^{2m+p}) + \frac{2m+p}{N} M^{2p} (v_\beta(|R_{1,2}^-|^{2m}) + v_\beta(|R_{1,2}|^{2m})). \end{aligned}$$

Consequently, from Lemma 8.4, it follows that

$$v_\beta(|R_{1,2}^-|^{2m+p}) \leq v_\beta(|R_{1,2}|^{2m+p}) + \frac{C_4}{N^{m+1}}$$

for $C_4 := (2m+p)M^{2p}K_0(2^{2m}M^{4m}+1)$. Plugging this into (8.4) and noting that $p \geq 3$ imply

$$\begin{aligned} v_\beta((R_{1,2})^{2m+2}) &\leq C(\beta) C_2 v_\beta(|R_{1,2}|^{2m+p}) + \frac{C(\beta)(C_2 C_4 + C_3) + C_1}{N^{m+1}} \\ &\leq C(\beta) C_2 M^{2(p-3)} v_\beta(|R_{1,2}|^{2m+3}) + \frac{C(\beta)(C_2 C_4 + C_3) + C_1}{N^{m+1}}. \end{aligned}$$

Setting $K_1(\beta) = C(\beta) C_2 M^{2(p-3)}$ and $K_2(\beta) = C(\beta)(C_2 C_4 + C_3) + C_1$ completes the proof. $\square$

8.4. *Proof of Theorem 3.1.* We prove Theorem 3.1 by induction on $m \geq 0$. Clearly the case $m = 0$ is valid. Assume that for some $m \geq 0$, there exists a constant $K \geq 1$ such that (3.4) holds for all $N \geq 1$ and $s \in [0, 1]$. Our goal is to show that there exists some $K' \geq 1$ such that

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

for all $N \geq 1$ and $s \in [0, 1]$. Let K_1 and K_2 be the two nonnegative continuous functions from the statement of Lemma 8.5 so that for all $N \geq 1$ and $s \in [0, 1]$

$$(8.5) \quad v_{s\beta}(|R_{1,2}|^{2m+2}) \leq K_1(s\beta)v_{s\beta}(|R_{1,2}|^{2m+3}) + \frac{K_2(s\beta)}{N^{m+1}}.$$

Note that $K_1(s\beta)$ is a nondecreasing function in s and $K_1(0) = 0$. Set

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

Now we divide our proof into two cases:

Case 1: $s \in [0, s_0]$. Combining (8.5) and the observation that $|R_{1,2}| \leq M^2$ results in

$$\begin{aligned} v_{s\beta}((R_{1,2})^{2m+2}) &\leq K_1(s_0\beta)M^2 v_{s\beta}(|R_{1,2}|^{2m+2}) + \frac{K_2(s\beta)}{N^{m+1}} \\ &\leq \frac{1}{2} v_{s\beta}((R_{1,2})^{2m+2}) + \frac{K_2(s\beta)}{N^{m+1}} \quad \forall s \in [0, s_0]. \end{aligned}$$

This gives that

$$v_{s\beta}((R_{1,2})^{2m+2}) \leq \frac{2K_2(s\beta)}{N^{m+1}} \quad \forall s \in [0, s_0].$$

Case 2: $s \in (s_0, 1]$. Choose $\varepsilon > 0$ such that

$$(8.6) \quad \varepsilon \max_{s \in [s_0, 1]} K_1(s\beta) < \frac{1}{2}.$$

From Proposition 7.1, there exists a constant K'' independent of N and $s \in [s_0, 1]$ such that

$$v_{s\beta}(I(|R_{1,2}| > \varepsilon)) < K'' e^{-N/K''} \quad \forall N \geq 1.$$

Note that

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

Plugging this into (8.5) leads to

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

Thus, from (8.6), we conclude that for all $N \geq 1$ and $s \in [s_0, 1]$

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

Finally, from the above two cases, our proof is completed by taking K' as the supremum of this bound for $s \in [0, 1]$.

8.5. *Proof of Proposition 3.2.* The proof of Proposition 3.2 relies on the following bound: for any $p \geq 2$, $\beta > 0$, and $l > 0$,

$$\mathbb{P}(|F_N(\beta)| \geq l) \leq \frac{2\beta^2}{l^2 N} \mathbb{E}(|R_{1,2}|^p)_\beta + \frac{2}{l^2} \left(\int_0^\beta t \mathbb{E}(|R_{1,2}|^p)_t dt \right)^2.$$

This result is essentially taken from [20], Lemma 10. Although there the spin configurations are sampled from the uniform probability measure on the hypercube $\{-1, +1\}^N$, the same argument applies to the current general setting. Now, from this and the moment control in Theorem 3.1, there exists a constant $K > 0$ such that

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

This completes our proof.

9. Structure of the regime $\bar{\mathcal{R}}$. This section presents the proof of Theorem 3.2. Recall the probability spaces (Λ_r, μ_r) , the temperature vector $\bar{\beta} = (\beta_1, \dots, \beta_r)$, the Hamiltonian $H_{N, \bar{\beta}}$, the free energy $F_N(\bar{\beta})$, the Gibbs measure $G_{N, \bar{\beta}}$, and the critical temperatures $\beta_{r,c}$ from Section 3.2. For each $1 \leq r \leq k$, let H_{N, β_r} , $F_{N,r}(\beta_r)$, and $F_r(\beta_r)$ be the Hamiltonian, free energy, and limiting free energy, respectively, corresponding to the scalar-valued spin glass in Section 3.1 with temperature β_r and probability space (Λ_r, μ_r) . Denote $v_{r,*} = \int d^2 \mu_r(da)$.

9.1. *Concentration of total overlap.* Let $M_k(\mathbb{R})$ be the space of real-valued $k \times k$ matrices equipped with the metric

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

For any $\varepsilon > 0$ and $V \in M_k(\mathbb{R})$, let $A_\varepsilon(V)$ be the collection of all $V' \in M_k(\mathbb{R})$ with $\|V - V'\|_{\max} < \varepsilon$. Denote the total overlap matrix by

$$\mathbf{R}(\bar{\sigma}) = \mathbf{R}(\sigma(1), \dots, \sigma(k)) := (R(\sigma(r), \sigma(r'))_{1 \leq r, r' \leq k}.$$

Set $V_* = (V_{*,r,r'})_{1 \leq r, r' \leq k} \in M_k(\mathbb{R})$ where $V_{*,r,r} = v_{*,r}$ and $V_{*,r,r'} = 0$ for $r \neq r'$. For any measurable subset $A \subseteq M_k(\mathbb{R})$, define the restricted free energy $F_N(\bar{\beta}, A)$ as

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

For i.i.d. samplings $\bar{\sigma}, \bar{\sigma}^1, \bar{\sigma}^2, \dots$ from $G_{N, \bar{\beta}}$, denote by $\langle \cdot \rangle_{\bar{\beta}}$ the Gibbs expectation with respect to these random variables.

The following proposition states that the self-overlap of $\bar{\sigma}$ sampled from the Gibbs measure $G_{N, \bar{\beta}}$ is concentrated around V_* in the high-temperature regime $\bar{\mathcal{R}}$.

PROPOSITION 9.1. *Assume that $\bar{\beta} \in \bar{\mathcal{R}}$. Let $\bar{\sigma}$ be sampled from $G_{N, \bar{\beta}}$. For any $\varepsilon > 0$, there exist positive constants K and δ such that for any $N \geq 1$, with probability at least $1 - Ke^{-N/K}$,*

$$(9.1) \quad F_N(\bar{\beta}, A_\varepsilon(V_*)^c) \leq F_N(\bar{\beta}) - \delta.$$

PROOF. We adapt a similar argument as the one for Lemma 7.1. Let $\bar{\beta} \in \bar{\mathcal{R}}$. For $1 \leq r, r' \leq k$, define

$$A_\varepsilon(r, r') = \{V \in M_k(\mathbb{R}) \mid |V_{r,r'} - (V_*)_{r,r'}| < \varepsilon\}.$$

Note that $\bigcap_{1 \leq r, r' \leq k} A_\varepsilon(r, r') = A_\varepsilon(V_*)$. Using Jensen's inequality and $\mathbb{E}e^{H_{N,\bar{\beta}}(\bar{\sigma})} = 1$ yields

$$\limsup_{N \rightarrow \infty} \mathbb{E}F_N(\bar{\beta}, A_\varepsilon(V_*)^c) \leq \limsup_{N \rightarrow \infty} \frac{1}{N} \log \sum_{1 \leq r, r' \leq k} \int_{\mathbf{R}(\bar{\sigma}) \in A_\varepsilon(r, r')^c} \bar{\mu}^{\otimes N}(d\bar{\sigma}).$$

Note that the coordinates $\sigma_1(r), \dots, \sigma_N(r)$ of $\sigma(r)$ are i.i.d. with distribution μ_r , which has bounded support. Also, note that $\sigma(r)$ is independent of $\sigma(r')$ for any $r \neq r'$. In addition, the mean of $\sigma_1(r)\sigma_1(r')$ under $\bar{\mu}$ is equal to $(V_*)_{r,r'}$ for any $1 \leq r, r' \leq k$. By Cramér's theorem, there exists a positive constant δ such that

$$\sum_{1 \leq r, r' \leq k} \int_{\mathbf{R}(\bar{\sigma}) \in A_\varepsilon(r, r')^c} \mu^{\otimes N}(d\bar{\sigma}) \leq k^2 e^{-N\delta}.$$

Note that $F(\bar{\beta}) = 0$. It follows that

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

Finally, (9.1) follows by using the Gaussian concentration inequality for $F_N(\bar{\beta}, A_\varepsilon(V_*)^c)$ and $F_N(\bar{\beta})$. We omit the details here as they are the same as those in the proof of Lemma 7.1. $\square$

9.2. *Proof of Theorem 3.2:* $\bar{\mathcal{R}} \subseteq (0, \beta_{1,c}] \times \dots \times (0, \beta_{k,c}]$. Suppose that $\bar{\beta} = (\beta_1, \dots, \beta_k) \in \bar{\mathcal{R}}$. By the definition of $\bar{\mathcal{R}}$, $F(\bar{\beta}) = 0$. Let $\varepsilon > 0$. On the one hand, recall that from Proposition 9.1, there exist two positive constants $K, \delta > 0$ such that for any $N \geq 1$, with probability at least $1 - Ke^{-N/K}$,

$$(9.2) \quad F_N(\bar{\beta}, A_\varepsilon(V_*)^c) \leq F_N(\bar{\beta}) - \delta.$$

On the other hand, note that $V_{r,r'} = 0$ for $r \neq r'$ and that $\mathbf{R}(\bar{\sigma}) \in A_\varepsilon(V_*)$ implies

$$|R(\sigma(r), \sigma(r'))^p| = |R(\sigma(r), \sigma(r'))^p - V_{r,r',*}^p| \leq \varepsilon^p \quad \forall 1 \leq r \neq r' \leq k.$$

From this, we can bound the overlap terms and then release the constraint $A_\varepsilon(V_*)$ to get

$$(9.3) \quad F_N(\bar{\beta}, A_\varepsilon(V_*)^c) \leq \sum_{r=1}^k F_{N,r}(\beta_r) + \frac{\varepsilon^p}{2} \sum_{r \neq r'} \beta_r \beta_{r'}.$$

Note that

$$F_N(\bar{\beta}) = N^{-1} \log(e^{NF_N(\bar{\beta}, A_\varepsilon(V_*)^c)} + e^{NF_N(\bar{\beta}, A_\varepsilon(V_*)^c)})$$

and that

$$\log(x + y) \leq \log 2 + \max(\log x, \log y) \quad \forall x, y > 0.$$

From (9.2), (9.3), and these two displays, after taking $N \rightarrow \infty$,

$$\begin{aligned} 0 = F(\bar{\beta}) &\leq \max\left(F(\bar{\beta}) - \delta, \sum_{r=1}^k F_r(\beta_r) + \frac{\varepsilon^p}{2} \sum_{r \neq r'} \beta_r \beta_{r'}\right) \\ &= \max\left(-\delta, \sum_{r=1}^k F_r(\beta_r) + \frac{\varepsilon^p}{2} \sum_{r \neq r'} \beta_r \beta_{r'}\right). \end{aligned}$$

Consequently, we obtain

$$0 \leq \sum_{r=1}^k F_r(\beta_r) + \frac{\varepsilon^p}{2} \sum_{r \neq r'} \beta_r \beta_{r'}$$

and letting $\varepsilon \downarrow 0$ yields $\sum_{r=1}^k F_r(\beta_r) \geq 0$. Since $F_r(\beta_r) \leq 0$ for all $1 \leq r \leq k$, we must have $F_r(\beta_r) = 0$ for all r . Hence, $\beta_r \in (0, \beta_{r,c}]$ for all $1 \leq r \leq k$ by Proposition 3.1. This establishes that $\bar{\mathcal{R}} \subseteq (0, \beta_{1,c}] \times \cdots \times (0, \beta_{k,c}]$.

9.3. *Proof of Theorem 3.2:* $\bar{\mathcal{R}} \supseteq (0, \beta_{1,c}] \times \cdots \times (0, \beta_{k,c}]$. We divide our discussion into two cases.

Case 1: $\bar{\beta} \in (0, \beta_{1,c}) \times \cdots \times (0, \beta_{k,c})$, but $\bar{\beta} \notin \bar{\mathcal{R}}$. In this case, $F_r(\beta_r) = 0$ for $1 \leq r \leq k$ and $F(\bar{\beta}) < 0$. Then there exists a positive constant η such that $\mathbb{E}F_N(\bar{\beta}) < -\eta$ for large enough N . Note that for any $\varepsilon > 0$ and $N \geq 1$,

$$F_N(\bar{\beta}, A_\varepsilon(V_*)) \leq F_N(\bar{\beta}).$$

Thus, from the Gaussian concentration inequality for $F_N(\bar{\beta})$, there exists a constant $K > 0$ such that for any large enough N , with probability at least $1 - Ke^{-N/K}$,

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

Note that the off-diagonal entries of V_* are all zero. The restriction $A_\varepsilon(V_*)$ allows us to pull the off-diagonal entries of the total overlap outside of the free energy to get

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

Now, if we take $\varepsilon > 0$ with $\varepsilon^p \sum_{r \neq r'} \beta_r \beta_{r'} < \eta/2$, then the above inequality reduces to

$$(9.4) \quad \frac{1}{N} \log \int_{\mathbf{R}(\bar{\sigma}) \in A_\varepsilon(V_*)} \exp\left(\sum_{r=1}^k H_{N,\beta_r}(\sigma(r))\right) \bar{\mu}^{\otimes N}(d\bar{\sigma}) < -\frac{\eta}{4}.$$

Denote by $\langle \cdot \rangle'$ the Gibbs average with respect to the independent samplings

$$\bar{\sigma} = (\sigma(1), \dots, \sigma(k)), \quad \bar{\sigma}^1 = (\sigma^1(1), \dots, \sigma^1(k)), \quad \bar{\sigma}^2 = (\sigma^2(1), \dots, \sigma^2(k))$$

from the product measure $\prod_{r=1}^k G_{N,\beta_r}(d\sigma(r))$. The combination of (9.4), the fact that $F_r(\beta_r) = 0$, and the Gaussian concentration inequality for $F_r(\beta_r)$ implies that the self-overlap matrix $\mathbf{R}(\bar{\sigma})$ is concentrated around V_* in the sense that there exists a constant $K' > 0$ such that for sufficiently large N ,

$$(9.5) \quad \mathbb{E}\langle I(\mathbf{R}(\bar{\sigma}) \in A_\varepsilon(V_*)) \rangle' \leq K' e^{-N/K'}.$$

Next, in order to deduce a contradiction, we recall that Lemma 7.1 states

$$(9.6) \quad \lim_{N \rightarrow \infty} \mathbb{E}\langle I(|R(\sigma(r), \sigma(r)) - v_{r,*}| \leq \delta) \rangle' = 1.$$

Here we used the fact that for a sampling $\bar{\sigma}$ from $\langle \cdot \rangle'$, the components $\sigma(1), \dots, \sigma(k)$ are independent of each other. For the same reason, it also follows from Proposition 7.1 and the assumption $\bar{\beta} \in (0, \beta_{1,c}) \times \cdots \times (0, \beta_{k,c})$ that for any $\delta > 0$ and $1 \leq r \leq k$,

$$(9.7) \quad \lim_{N \rightarrow \infty} \mathbb{E}\langle I(|R(\sigma^1(r), \sigma^2(r))| \leq \delta) \rangle' = 1.$$

For $r \neq r'$, observe that

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

where the second equality holds since $\sigma_i(r)$, $\sigma_j(r)$ are independent of $\sigma_i(r')$, $\sigma_j(r')$ under $\langle \cdot \rangle'$. Consequently, an application of the Cauchy–Schwarz inequality implies that

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

where the last equality uses the identity

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

which also holds when r' replaces r . From the above inequality and (9.7),

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

which means that $R(\sigma(r), \sigma(r'))$ is essentially concentrated at 0 under $\langle \cdot \rangle'$. Combining the latter observation, the inequality (9.6), and the fact that the off-diagonal entries of V^* are all zero yields

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

However, this contradicts (9.5). Thus, we must have $\bar{\beta} \in \mathcal{R}$.

Case 2: $\bar{\beta} = (\beta_1, \dots, \beta_k) \in (0, \beta_{1,c}] \times \dots \times (0, \beta_{k,c}]$, but $\bar{\beta} = (\beta_1, \dots, \beta_k) \notin (0, \beta_{1,c}) \times \dots \times (0, \beta_{k,c})$. Note that the free energies $F, F_1, \dots, F_k$ are continuous functions of the temperature parameters. We can approximate $F(\bar{\beta})$ by $F(\bar{\beta}')$ for $\bar{\beta}' \in (0, \beta_{1,c}) \times \dots \times (0, \beta_{k,c})$. From this and Case 1, we see that $F(\bar{\beta}) = 0$ and so $\bar{\beta} \in \bar{\mathcal{R}}$.

APPENDIX: PROOF OF THE GUERRA–TALAGRAND BOUND

Recall X_N from Section 3.1 and recall $\mathcal{V}$, $\mathcal{M}_{v,v_0}$, and $\mathcal{P}_{\beta,V}$ from Section 7.1. The goal of this Appendix is to give a sketch of the proof for the Guerra–Talagrand inequality stated in (7.1). We follow the same argument in [20], Proposition 2.

Step 1: Fix v_0 and assume that $\alpha \in \mathcal{M}_{v,v_0}$ is of the form $\alpha(s) = m 1_{[0,v_0)}(s) + 1_{[v_0,v]}(s)$ for some $m \in [0, 1]$. Let $m_0 = 0 < m_1 = m < m_2 < m_3 = 1$. Let $(c_\tau)_{\tau \in \mathbb{N}^2}$ be the Ruelle probability cascades associated to $0 < m_1 < m_2 < 1$, see [61], Section 14.1. Set

$$\begin{aligned} \rho_0^{11} &= 0, & \rho_1^{11} &= v_0, & \rho_2^{11} &= v, \\ \rho_0^{12} &= 0, & \rho_1^{12} &= v_0, & \rho_2^{12} &= v_0, \\ \rho_0^{21} &= 0, & \rho_1^{21} &= v_0, & \rho_2^{21} &= v_0, \\ \rho_0^{22} &= 0, & \rho_1^{11} &= v_0, & \rho_2^{22} &= v. \end{aligned}$$

Assume that (z_1^1, z_1^2) and (z_2^1, z_2^2) are two independent Gaussian random vectors with mean zero and covariance,

$$\mathbb{E} z_a^\ell z_a^{\ell'} = \xi'(\rho_a^{\ell\ell'}) - \xi'(\rho_{a-1}^{\ell\ell'})$$

for $a = 1, 2$ and $\ell, \ell' = 1, 2$. Let $(z_{i,1,j_1}^1, z_{i,1,j_1}^2)_{j_1 \in \mathbb{N}}$ for $1 \leq i \leq N$ be i.i.d. copies of (z_1^1, z_1^2) and $(z_{i,2,j_1,j_2}^1, z_{i,2,j_1,j_2}^2)_{(j_1,j_2) \in \mathbb{N}^2}$ for $1 \leq i \leq N$ be i.i.d. copies of (z_2^1, z_2^2) . These are also independent of each other. For $0 \leq t \leq 1$, consider the interpolating Hamiltonian

$$X_{N,t}(\sigma^1, \sigma^2, \tau) := \sqrt{t}(X_N(\sigma^1) + X_N(\sigma^2)) + \sqrt{1-t} \sum_{1 \leq i \leq N} \sum_{\ell=1,2} \sigma_i^\ell (z_{i,1,j} + z_{i,2,\tau_1,\tau_2})$$

for $\sigma^1, \sigma^2 \in \Lambda^N$ and $\tau = (\tau_1, \tau_2) \in \mathbb{N}^2$. Define the interpolating free energy

$$\phi_{N,\eta}(t) = \frac{1}{N} \mathbb{E} \log \sum_{\tau \in \mathbb{N}^2} c_\tau \int_{\mathbf{R}(\sigma^1, \sigma^2) \in A_\eta(V)} e^{\beta X_{N,t}(\sigma^1, \sigma^2, \tau)} \mu^{\otimes N}(d\sigma^1) \mu^{\otimes N}(d\sigma^2).$$

Note that

$$\phi_{N,\eta}(1) = \frac{1}{N} \mathbb{E} \log \int_{\mathbf{R}(\sigma^1, \sigma^2) \in A_\eta(V)} e^{\beta X_N(\sigma^1) + \beta X_N(\sigma^2)} \mu^{\otimes N}(d\sigma^1) \mu^{\otimes N}(d\sigma^2),$$

and $\phi_{N,\eta}(0)$ involves only linear spin interactions.

Step 2: We proceed to compute the derivative of $\phi_{N,\eta}(t)$. Consider the Gibbs measure

$$G_{N,\eta,t}(d\sigma^1, d\sigma^2, \tau) := \frac{1}{Z_{N,\eta,t}} c_\tau e^{\beta X_{N,t}(\sigma^1, \sigma^2, \tau)} \mu^{\otimes N}(d\sigma^1) \mu^{\otimes N}(d\sigma^2)$$

for $(\sigma^1, \sigma^2, \tau) \in \Lambda^N \times \Lambda^N \times \mathbb{N}^2$ satisfying that $\mathbf{R}(\sigma^1, \sigma^2) \in A_\eta(V)$, where $Z_{N,\eta,t}$ is the normalizing constant. Let $\langle \cdot \rangle_{N,\eta,t}$ be the Gibbs expectation associated to this free energy. Denote by $(\sigma^1, \sigma^2, \tau)$ and $(\hat{\sigma}^1, \hat{\sigma}^2, \hat{\tau})$ two independent samplings from $G_{N,\eta,t}$. We set the overlaps between these two pairs by $Q^{\ell\ell'} = R(\sigma^\ell, \hat{\sigma}^{\ell'})$ for $1 \leq \ell, \ell' \leq 2$ and

$$\tau \wedge \hat{\tau} = \begin{cases} 0 & \text{if } \tau_1 \neq \tau_2, \\ 1 & \text{if } \tau_1 = \hat{\tau}_1, \tau_2 \neq \hat{\tau}_2, \\ 2 & \text{if } \tau_1 = \hat{\tau}_1, \tau_2 = \hat{\tau}_2. \end{cases}$$

From the same computation in [61], Chapter 15, that uses Gaussian integration by parts,

$$\phi'_{N,\eta}(t) = -\Pi_{N,\eta}(t) - E_{N,\eta}(t) + O(\eta)$$

for

$$\begin{aligned} \Pi_{N,\eta}(t) &:= \frac{\beta^2}{2} \sum_{\ell, \ell'=1,2} (\theta(\rho_2^{\ell\ell'}) - \mathbb{E} \langle \theta(\rho_{\tau \wedge \hat{\tau}}^{\ell\ell'}) \rangle_{N,\eta,t}), \\ E_{N,\eta}(t) &:= \frac{\beta^2}{2} \sum_{\ell, \ell'=1,2} \mathbb{E} \langle \Gamma(\rho_{\tau \wedge \hat{\tau}}^{\ell\ell'}, Q^{\ell\ell'}) \rangle_{N,\eta,t}, \end{aligned}$$

where $O(\eta)$ means that it uniformly vanishes as $\eta \downarrow 0$, $\theta(x) := x\xi'(x) - \xi(x)$, and $\Gamma(x, y) := \xi(y) - y\xi'(x) + \theta(x)$.

Step 3: We handle $\Pi_{N,\eta}(t)$ and $E_{N,\eta}(t)$ as follows. First, using the fact (see [61], Section 14.1) that

$$(A.8) \quad \mathbb{E} \langle I(\tau \wedge \hat{\tau} = a) \rangle = m_{a+1} - m_a, \quad 0 \leq a \leq 2$$

leads to

$$\Pi_{N,\eta}(t) = \beta^2(2m(\theta(v_0) - \theta(0)) + m_2(\theta(v) - \theta(v_0))).$$

The treatment for $E_{N,\eta}(t)$ is the harder part. If p is an even number, we obviously have

$$(A.9) \quad \Gamma(x, y) \geq 0 \quad \forall x, y \in \mathbb{R},$$

which implies $E_{N,\eta}(t) \geq 0$. If p is odd, (A.9) is no longer valid and we do not have an obvious sign for $E_{N,\eta}(t)$. The idea to overcome this difficulty is to add an asymptotically vanishing perturbation to the interpolating Hamiltonian such that the entries in the overlap matrix $(Q^{\ell\ell'})_{1 \leq \ell, \ell' \leq 2}$ are synchronized in the limit, see [51]. A crucial fact here is that this property will force the overlap matrix to be asymptotically positive semi-definite under $\mathbb{E}(\cdot)_{N,\eta,t}$ for all $0 \leq t \leq 1$ as $N \rightarrow \infty$ and $\eta \downarrow 0$, see [52]. For the precise choice of this perturbation, we refer the reader to the proof of [20], Proposition 2.

For clarity, we adapt the same notation for the free energy and the Gibbs expectation. Note that the procedure of adding an asymptotically vanishing perturbation will also not affect the value of the coupled free energy when $N \rightarrow \infty$ and $\eta \downarrow 0$ and the derivative of the corresponding $\phi_{N,\eta}(t)$ still has the same form. Now write

$$(A.10) \quad \begin{aligned} E_{N,\eta}(t) = & \frac{\beta^2}{2} \mathbb{E} \left\langle I(\tau \wedge \hat{\tau} = 0) \sum_{\ell, \ell' = 1, 2} \Gamma(Q^{\ell\ell'}, 0) \right\rangle_{N,\eta,t} \\ & + \frac{\beta^2}{2} \mathbb{E} \left\langle I(\tau \wedge \hat{\tau} = 1) \sum_{\ell, \ell' = 1, 2} \Gamma(Q^{\ell\ell'}, v_0) \right\rangle_{N,\eta,t} \\ & + \frac{\beta^2}{2} \mathbb{E} \left\langle I(\tau \wedge \hat{\tau} = 2) \sum_{\ell, \ell' = 1, 2} \Gamma(Q^{\ell\ell'}, \rho_2^{\ell\ell'}) \right\rangle_{N,\eta,t}. \end{aligned}$$

To hand this, we recall a lemma from [20], Lemma 11,² that for any $s \geq 0$ and $x, y, z \in \mathbb{R}$ so that

$$\begin{bmatrix} x & z \\ z & y \end{bmatrix}$$

is positive semi-definite, we have

$$\Gamma(s, x) + \Gamma(s, y) + 2\Gamma(s, z) \geq 0.$$

By using the positive semi-definiteness of the overlap matrix, this lemma implies that the first and second terms on the right-hand side of (A.10) are asymptotically nonnegative. Also, from (A.8), the third term of (A.10) is upper bounded by $C(1 - m_2)$. Hence, when p is odd, we have

$$\liminf_{\eta \downarrow 0} \liminf_{N \rightarrow \infty} E_{N,\eta}(t) \geq -C(1 - m_2).$$

To summarize, from the above discussion, no matter if p is even or odd,

$$(A.11) \quad \begin{aligned} & \lim_{\eta \downarrow 0} \limsup_{N \rightarrow \infty} \phi_{N,\eta}(1) \\ & \leq \lim_{\eta \downarrow 0} \limsup_{N \rightarrow \infty} \phi_{N,\eta}(0) + \lim_{\eta \downarrow 0} \limsup_{N \rightarrow \infty} \int_0^1 \phi'_{N,\eta}(t) dt \\ & \leq \lim_{\eta \downarrow 0} \limsup_{N \rightarrow \infty} \phi_{N,\eta}(0) \\ & \quad - \beta^2(2m(\theta(v_0) - \theta(0)) + m_2(\theta(v) - \theta(v_0))) + C(1 - m_2). \end{aligned}$$

²Although the statement there also requires $s \leq 1$ and $x, y, z \in [-1, 1]$, they are not actually needed in the proof. Hence the result is still valid under the present assumption.

Step 4: Our last step is to rearrange the right-hand side of the last inequality. Note that we can release the constraint $\mathbf{R}(\sigma^1, \sigma^2) \in A_\eta(V)$ in $\phi_{N,\eta}(0)$ by introducing a Lagrange variable $\lambda \in M_2(\mathbb{R})$ so that

$$\begin{aligned} \phi_{N,\eta}(0) &\leq \frac{1}{N} \mathbb{E} \log \sum_{\tau \in \mathbb{N}^2} c_\tau \int e^{\beta X_{N,0}(\sigma^1, \sigma^2, \tau) + N \langle \lambda, \mathbf{R}(\sigma^1, \sigma^2) \rangle} \mu^{\otimes N}(d\sigma^1) \mu^{\otimes N}(d\sigma^2) \\ &\quad - \langle \lambda, V \rangle + \eta. \end{aligned}$$

If we denote

$$\begin{aligned} w_{i,\tau}^1 &= z_{i,1,\tau_1}^1 + z_{i,2,\tau_1,\tau_2}^1, & w_{i,\tau}^2 &= z_{i,1,\tau_1}^2 + z_{i,2,\tau_1,\tau_2}^2, \\ w^1 &= z_1^1 + z_2^1, & w^2 &= z_1^2 + z_2^2, \end{aligned}$$

then the first term on the right-hand side of the last inequality can be written as

$$\begin{aligned} &\frac{1}{N} \mathbb{E} \log \sum_{\tau \in \mathbb{N}^2} c_\tau \prod_{i=1}^N \int \exp \beta \left(w_{i,\tau}^1 \sigma_i^1 + w_{i,\tau}^2 \sigma_i^2 + \sum_{\ell, \ell'=1,2} \lambda_{\ell \ell'} \sigma_i^\ell \sigma_i^{\ell'} \right) \mu(d\sigma_i^1) \mu(d\sigma_i^2) \\ &= \frac{1}{m_1} \log \mathbb{E}_1 \exp \frac{m_1}{m_2} \log \mathbb{E}_2 \left[\exp \left(m_2 \log \int e^{\beta w^1 a^1 + \beta w^2 a^2 + \langle \lambda a, a \rangle} \mu(da^1) \mu(da^2) \right) \right], \end{aligned}$$

where $\mathbb{E}_1$ is the expectation with respect to (z_1^1, z_1^2) only and $\mathbb{E}_2$ is the expectation with respect to (z_2^1, z_2^2) only. Here, this equality is valid by using [61], Theorem 14.2.1. From this, after sending $m_2 \rightarrow 1$, it can be checked directly by the Cole–Hopf transformation that the last equation is indeed equal to $\Psi_{\beta,V,\alpha}(0, 0, \lambda)$. On the other hand, the second and third terms in the last line of (A.11) together equal

$$\beta^2 \left(\int_0^v \xi''(s) s \alpha(s) ds + \int_0^{v_0} \xi''(s) s \alpha(ds) \right).$$

These and (A.11) complete our proof.

Acknowledgments. The authors thank the anonymous referees for providing many useful suggestions regarding the presentation of the paper and bringing an alternative approach of Section 2 to our attention.

W. K. Chen thanks the National Center for Theoretical Sciences and Academia Sinica in Taipei for their hospitality during his visit in June and July 2018, where part of the results and writings were completed. In addition, he is grateful to Lenka Zdeborová for many illuminating discussions.

Funding. The research of W.-K. Chen is partly supported by NSF Grants DMS-16-42207 and DMS-17-52184, and Hong Kong Research Grants Council GRF-14302515. The research of G. Lerman is partially supported by NSF Grants DMS-14-18386 and DMS-18-21266.

REFERENCES

- [1] AIZENMAN, M., LEBOWITZ, J. L. and RUELLE, D. (1987). Some rigorous results on the Sherrington–Kirkpatrick spin glass model. *Comm. Math. Phys.* **112** 3–20. [MR0904135](#)
- [2] AUFFINGER, A. and CHEN, W.-K. (2015). The Parisi formula has a unique minimizer. *Comm. Math. Phys.* **335** 1429–1444. [MR3320318](#) <https://doi.org/10.1007/s00220-014-2254-z>
- [3] BAIK, J., BEN AROUS, G. and PÉCHÉ, S. (2005). Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *Ann. Probab.* **33** 1643–1697. [MR2165575](#) <https://doi.org/10.1214/009117905000000233>

- [4] BAIK, J. and SILVERSTEIN, J. W. (2006). Eigenvalues of large sample covariance matrices of spiked population models. *J. Multivariate Anal.* **97** 1382–1408. [MR2279680](#) <https://doi.org/10.1016/j.jmva.2005.08.003>
- [5] BARBIER, J., DIA, M., MACRIS, N. and KRZAKALA, F. (2016). The mutual information in random linear estimation. In *54th Annual Allerton Conference on Communication, Control, and Computing* 625–632.
- [6] BARBIER, J., DIA, M., MACRIS, N., KRZAKALA, F., LESIEUR, T. and ZDEBOROVÁ, L. (2016). Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. *Adv. Neural Inf. Process. Syst.* **29** 424–432.
- [7] BARBIER, J., KRZAKALA, F., MACRIS, N., MIOLANE, L. and ZDEBOROVÁ, L. (2017). Phase transitions, optimal errors and optimality of message-passing in generalized linear models. arXiv e-prints.
- [8] BARBIER, J. and MACRIS, N. (2019). The adaptive interpolation method: A simple scheme to prove replica formulas in Bayesian inference. *Probab. Theory Related Fields* **174** 1133–1185. [MR3980313](#) <https://doi.org/10.1007/s00440-018-0879-0>
- [9] BARBIER, J., MACRIS, N., DIA, M. and KRZAKALA, F. (2020). Mutual information and optimality of approximate message-passing in random linear estimation. *IEEE Trans. Inf. Theory* **66** 4270–4303. [MR4130617](#) <https://doi.org/10.1109/TIT.2020.2990880>
- [10] BARBIER, J., MACRIS, N. and MIOLANE, L. (2017). The layered structure of tensor estimation and its mutual information. In *55th Annual Allerton Conference on Communication, Control, and Computing*.
- [11] BARDINA, X., MÁRQUEZ-CARRERAS, D., ROVIRA, C. and TINDEL, S. (2004). The p -spin interaction model with external field. *Potential Anal.* **21** 311–362. [MR2081143](#) <https://doi.org/10.1023/B:POTA.0000034325.04634.f5>
- [12] BAYATI, M. and MONTANARI, A. (2011). The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inf. Theory* **57** 764–785. [MR2810285](#) <https://doi.org/10.1109/TIT.2010.2094817>
- [13] BEN AROUS, G., GHEISSARI, R. and JAGANNATH, A. (2020). Algorithmic thresholds for tensor PCA. *Ann. Probab.* **48** 2052–2087. [MR4124533](#) <https://doi.org/10.1214/19-AOP1415>
- [14] BEN AROUS, G., MEI, S., MONTANARI, A. and NICA, M. (2019). The landscape of the spiked tensor model. *Comm. Pure Appl. Math.* **72** 2282–2330. [MR4011861](#) <https://doi.org/10.1002/cpa.21861>
- [15] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2011). The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices. *Adv. Math.* **227** 494–521. [MR2782201](#) <https://doi.org/10.1016/j.aim.2011.02.007>
- [16] BENAYCH-GEORGES, F. and NADAKUDITI, R. R. (2012). The singular values and vectors of low rank perturbations of large rectangular random matrices. *J. Multivariate Anal.* **111** 120–135. [MR2944410](#) <https://doi.org/10.1016/j.jmva.2012.04.019>
- [17] BOLTHAUSEN, E. (2018). Search results web results a Morita type proof of the replica-symmetric formula for SK. arXiv e-prints.
- [18] BOVIER, A., KURKOVA, I. and LÖWE, M. (2002). Fluctuations of the free energy in the REM and the p -spin SK models. *Ann. Probab.* **30** 605–651. [MR1905853](#) <https://doi.org/10.1214/aop/1023481004>
- [19] CHEN, W.-K. (2017). Variational representations for the Parisi functional and the two-dimensional Guerra–Talagrand bound. *Ann. Probab.* **45** 3929–3966. [MR3729619](#) <https://doi.org/10.1214/16-AOP1154>
- [20] CHEN, W.-K. (2019). Phase transition in the spiked random tensor with Rademacher prior. *Ann. Statist.* **47** 2734–2756. [MR3988771](#) <https://doi.org/10.1214/18-AOS1763>
- [21] DEMBO, A. and ZEITOUNI, O. (2010). *Large Deviations Techniques and Applications*. *Stochastic Modelling and Applied Probability* **38**. Springer, Berlin. Corrected reprint of the second (1998) edition. [MR2571413](#) <https://doi.org/10.1007/978-3-642-03311-7>
- [22] DESHPANDE, Y., ABBE, E. and MONTANARI, A. (2017). Asymptotic mutual information for the balanced binary stochastic block model. *Inf. Inference* **6** 125–170. [MR3671474](#) <https://doi.org/10.1093/imaiai/iaw017>
- [23] DESHPANDE, Y. and MONTANARI, A. (2014). Information-theoretically optimal sparse PCA. In *IEEE International Symposium on Information Theory* 2197–2201.
- [24] DONOHO, D. L., MALEKI, A. and MONTANARI, A. (2009). Message-passing algorithms for compressed sensing. *Proc. Natl. Acad. Sci. USA* **106** 18914–18919.
- [25] EL ALAOUI, A., KRZAKALA, F. and JORDAN, M. (2020). Fundamental limits of detection in the spiked Wigner model. *Ann. Statist.* **48** 863–885. [MR4102679](#) <https://doi.org/10.1214/19-AOS1826>
- [26] EL ALAOUI, A. and JORDAN, M. I. (2018). Detection limits in the high-dimensional spiked rectangular model. In *Conference on Learning Theory* 410–438.
- [27] EL ALAOUI, A., KRZAKALA, F. and JORDAN, M. (2017). Finite-size corrections and likelihood ratio fluctuations in the spiked Wigner model. arXiv e-prints.
- [28] FÉRAL, D. and PÉCHÉ, S. (2007). The largest eigenvalue of rank one deformation of large Wigner matrices. *Comm. Math. Phys.* **272** 185–228. [MR2291807](#) <https://doi.org/10.1007/s00220-007-0209-3>

- [29] GUO, D., WU, Y., SHAMAI, S. and VERDÚ, S. (2011). Estimation in Gaussian noise: Properties of the minimum mean-square error. *IEEE Trans. Inf. Theory* **57** 2371–2385. [MR2809096](#) <https://doi.org/10.1109/TIT.2011.2111010>
- [30] HILLAR, C. J. and LIM, L.-H. (2013). Most tensor problems are NP-hard. *J. ACM* **60** Art. 45, 39. [MR3144915](#) <https://doi.org/10.1145/2512329>
- [31] HOPKINS, S. B., SHI, J., SCHRAMM, T. and STEURER, D. (2016). Fast spectral algorithms from sum-of-squares proofs: Tensor decomposition and planted sparse vectors. In *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing* 178–191. ACM, New York. [MR3536564](#) <https://doi.org/10.1145/2897518.2897529>
- [32] HOPKINS, S. B., SHI, J. and STEURER, D. (2015). Tensor principal component analysis via sum-of-square proofs. In *Conference on Learning Theory* 956–1006.
- [33] JAGANNATH, A. and TOBASCO, I. (2016). A dynamic programming approach to the Parisi functional. *Proc. Amer. Math. Soc.* **144** 3135–3150. [MR3487243](#) <https://doi.org/10.1090/proc/12968>
- [34] JAVANMARD, A. and MONTANARI, A. (2013). State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference* **2** 115–144. [MR3311445](#) <https://doi.org/10.1093/imaiai/iat004>
- [35] JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.* **29** 295–327. [MR1863961](#) <https://doi.org/10.1214/aos/1009210544>
- [36] KOLDA, T. G. and BADER, B. W. (2009). Tensor decompositions and applications. *SIAM Rev.* **51** 455–500. [MR2535056](#) <https://doi.org/10.1137/07070111X>
- [37] LELARGE, M. and MIOLANE, L. (2017). Fundamental limits of symmetric low-rank matrix estimation. In *Proceedings of the 2017 Conference on Learning Theory* **PMLR 65** 1297–1301.
- [38] LESIEUR, T., MIOLANE, L., LELARGE, M., KRZAKALA, F. and ZDEBOROVÁ, L. (2017). Statistical and computational phase transitions in spiked tensor estimation. In *IEEE 2017 International Symposium* 511–515.
- [39] MARCHENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Math. USSR, Sb.* **1** 457.
- [40] MÉZARD, M. and MONTANARI, A. (2009). *Information, Physics, and Computation*. Oxford Graduate Texts. Oxford Univ. Press, Oxford. [MR2518205](#) <https://doi.org/10.1093/acprof:oso/9780198570837.001.0001>
- [41] MÉZARD, M., PARISI, G. and VIRASORO, M. A. (1987). *Spin Glass Theory and Beyond*. World Scientific Lecture Notes in Physics **9**. World Scientific Co., Inc., Teaneck, NJ. [MR1026102](#)
- [42] MIOLANE, L. (2017). Fundamental limits of symmetric low-rank matrix estimation: The non-symmetric case. arXiv e-prints.
- [43] MIOLANE, L. (2018). Phase transitions in spiked matrix estimation: Information-theoretic analysis. Preprint. Available at [arXiv:1806.04343](#).
- [44] MONTANARI, A., REICHMAN, D. and ZEITOUNI, O. (2017). On the limitation of spectral methods: From the Gaussian hidden clique problem to rank one perturbations of Gaussian tensors. *IEEE Trans. Inf. Theory* **63** 1572–1579. [MR3625981](#) <https://doi.org/10.1109/TIT.2016.2637959>
- [45] MONTANARI, A. and RICHARD, E. (2014). A statistical model for tensor PCA. *Neural Inf. Process. Syst.* **2** 2897–2905.
- [46] MONTANARI, A. and RICHARD, E. (2016). Non-negative principal component analysis: Message passing algorithms and sharp asymptotics. *IEEE Trans. Inf. Theory* **62** 1458–1484. [MR3472260](#) <https://doi.org/10.1109/TIT.2015.2457942>
- [47] MONTANARI, A. and SEN, S. (2016). Semidefinite programs on sparse random graphs and their application to community detection. In *STOC'16—Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing* 814–827. ACM, New York. [MR3536616](#) <https://doi.org/10.1145/2897518.2897548>
- [48] ONATSKI, A., MOREIRA, M. J. and HALLIN, M. (2013). Asymptotic power of sphericity tests for high-dimensional data. *Ann. Statist.* **41** 1204–1231. [MR3113808](#) <https://doi.org/10.1214/13-AOS1100>
- [49] PANCHENKO, D. (2005). Free energy in the generalized Sherrington–Kirkpatrick mean field model. *Rev. Math. Phys.* **17** 793–857. [MR2159369](#) <https://doi.org/10.1142/S0129055X05002455>
- [50] PANCHENKO, D. (2013). *The Sherrington–Kirkpatrick Model*. Springer Monographs in Mathematics. Springer, New York. [MR3052333](#) <https://doi.org/10.1007/978-1-4614-6289-7>
- [51] PANCHENKO, D. (2015). The free energy in a multi-species Sherrington–Kirkpatrick model. *Ann. Probab.* **43** 3494–3513. [MR3433586](#) <https://doi.org/10.1214/14-AOP967>
- [52] PANCHENKO, D. (2018). Free energy in the mixed p -spin models with vector spins. *Ann. Probab.* **46** 865–896. [MR3773376](#) <https://doi.org/10.1214/17-AOP1194>
- [53] PAUL, D. (2007). Asymptotics of sample eigenstructure for a large dimensional spiked covariance model. *Statist. Sinica* **17** 1617–1642. [MR2399865](#)

- [54] PÉCHÉ, S. (2006). The largest eigenvalue of small rank perturbations of Hermitian random matrices. *Probab. Theory Related Fields* **134** 127–173. [MR2221787](#) <https://doi.org/10.1007/s00440-005-0466-z>
- [55] PERRY, A., WEIN, A. S., BANDEIRA, A. and MOITRA, A. (2016). Optimality and sub-optimality of PCA for spiked random matrices and synchronization. arXiv e-prints.
- [56] PERRY, A., WEIN, A. S. and BANDEIRA, A. S. (2020). Statistical limits of spiked tensor models. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 230–264. [MR4058987](#) <https://doi.org/10.1214/19-AIHP960>
- [57] REEVES, G. and PFISTER, H. D. (2016). The replica-symmetric prediction for compressed sensing with Gaussian matrices is exact. arXiv e-prints.
- [58] ROS, V., BEN AROUS, G., BIROLI, G. and CAMMAROTA, C. (2018). Complex energy landscapes in spiked-tensor and simple glassy models: Ruggedness, arrangements of local minima and phase transitions. arXiv e-prints.
- [59] TALAGRAND, M. (2006). The Parisi formula. *Ann. of Math.* (2) **163** 221–263. [MR2195134](#) <https://doi.org/10.4007/annals.2006.163.221>
- [60] TALAGRAND, M. (2010). *Mean Field Models for Spin Glasses. Volume I: Basic Examples. Ergebnisse der Mathematik und Ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]* **54**. Springer, Berlin. [MR2731561](#) <https://doi.org/10.1007/978-3-642-15202-3>
- [61] TALAGRAND, M. (2011). *Mean Field Models for Spin Glasses. Volume II: Advanced Replica-Symmetry and Low Temperature. Ergebnisse der Mathematik und Ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]* **55**. Springer, Heidelberg. [MR3024566](#)
- [62] WEIN, A. S., EL ALAOUI, A. and MOORE, C. (2019). The Kikuchi hierarchy and tensor PCA. arXiv e-prints.
- [63] WU, Y. and VERDÚ, S. (2012). Functional properties of minimum mean-square error and mutual information. *IEEE Trans. Inf. Theory* **58** 1289–1301. [MR2932809](#) <https://doi.org/10.1109/TIT.2011.2174959>