MIT Open Access Articles

The overlap gap property in principal submatrix recovery

Massachusetts Institute of Technology

The overlap gap property in principal submatrix recovery

THE OVERLAP GAP PROPERTY IN PRINCIPAL SUBMATRIX RECOVERY

DAVID GAMARNIK, AUKOSH JAGANNATH, AND SUBHABRATA SEN

ABSTRACT. We study support recovery for a $k \times k$ principal submatrix with elevated mean λ/N, hidden in an $N \times N$ symmetric mean zero Gaussian matrix. Here $\lambda > 0$ is a universal constant, and we assume $k = N\rho$ for some constant $\rho \in (0,1)$. We establish that there exists a constant $C > 0$ such that the MLE recovers a constant proportion of the hidden submatrix if $\lambda \geq C\sqrt{\frac{1}{\rho} \log \frac{1}{\rho}}$, while such recovery is information theoretically impossible if $\lambda = o(\sqrt{\frac{1}{\rho} \log \frac{1}{\rho}})$. The MLE is computationally intractable in general, and in fact, for $\rho > 0$ sufficiently small, this problem is conjectured to exhibit a *statistical-computational gap*. To provide rigorous evidence for this, we study the likelihood landscape for this problem, and establish that for some $\varepsilon > 0$ and $\sqrt{\frac{1}{\rho} \log \frac{1}{\rho}} \ll \lambda \ll \frac{1}{\rho^{1/2+\varepsilon}}$, the problem exhibits a variant of the *Overlap-Gap-Property (OGP)*. As a direct consequence, we establish that a family of local MCMC based algorithms do not achieve optimal recovery. Finally, we establish that for $\lambda > 1/\rho$, a simple spectral method recovers a constant proportion of the hidden submatrix.

1. INTRODUCTION

In this paper, we study support recovery for a planted principal submatrix in a large symmetric Gaussian matrix. Formally, we observe a symmetric matrix $A = (A_{ij}) \in \mathbb{R}^{N \times N}$,

$$A_{ij} = \theta_{ij} + W_{ij}. \tag{1.1}$$

Throughout, we assume that W is a GOE random matrix; in other words, $\{W_{ij} : i \leq j\}$ are independent Gaussian random variables, with $\{W_{ij} : i < j\}$ i.i.d. $\mathcal{N}(0, 1/N)$, and $\{W_{ii} : 1 \leq i \leq N\}$ i.i.d. $\mathcal{N}(0, 2/N)$. Regarding the mean matrix $\mathbf{\Theta} = (\theta_{ij})$, we assume that there exists $U \subset [N] := \{1, 2, \cdots, N\}$, $|U| = k$, such that

$$\theta_{ij} = \begin{cases} \frac{\lambda}{N} & \text{if } i, j \in U \\ 0 & \text{o.w.,} \end{cases}$$

where $\lambda > 0$ is a constant independent of N. Equivalently, the observed matrix A may be re-written as

$$A = \frac{\lambda}{N} vv^T + W, \tag{1.2}$$

where $v = (v_i) \in \{0,1\}^N$, with $\sum_i v_i = k$. Throughout the subsequent discussion, we will denote the set of such boolean vectors as $\Sigma_N(\frac{k}{N})$.

In the setting introduced above, the following statistical questions are natural.

(1) (Detection) Can we detect the presence of the planted submatrix, i.e., can we consistently test

$$H_0 : \lambda = 0 \quad \text{vs.} \quad H_1 : \lambda > 0.$$

(2) (Recovery) How large should λ be, such that the support of v can be recovered *approximately*?

(3) (Efficient Recovery) When can support recovery be accomplished using a computationally feasible procedure?

Here, we study support recovery in the special case $k = N\rho$, for some $\rho \in (0,1)$. To ensure that this problem is well-defined for all N, we work along a sequence $\rho_N \to \rho$ such that $N\rho_N \in \mathbb{N}$ and $N\rho \in (N\rho_N - 1/2, N\rho_N + 1/2]$. Note that in this case, the corresponding submatrix detection question [27] is trivial, and a test which rejects for large values of the sum of the matrix entries consistently detects the underlying submatrix for any $\lambda > 0$. Motivated by the sparse PCA problem, we will study support recovery in this setup in the double limit $\rho \to 0$, following $N \to \infty$. Deshpande-Montanari [36] initiated a study of the problem (1.2), and established that Bayes optimal recovery of the matrix $\boldsymbol{\Theta}$ can be accomplished using an Approximate Message Passing based algorithm, whenever $\rho > 0$ is sufficiently large (specifically, $\rho > 0.041$). In [58, 59], the authors analyze optimal Bayesian estimation in the $\rho \to 0$ regime, and based on the behavior of the fixed points of a state-evolution system, conjecture the existence of an algorithmically hard phase in this problem; specifically, they conjecture that the minimum signal λ required for accurate support recovery using feasible algorithms should be significantly higher than the information theoretic minimum. This conjecture has been repeatedly quoted in various follow up works [15, 56, 57], but to the best of our knowledge, it has not been rigorously established in the prior literature. In this paper, we study the likelihood landscape of this problem, and provide rigorous evidence to the veracity of this conjecture.

From a purely conceptual viewpoint, the existence of a computationally hard phase in problem (1.2) is particularly striking. In the context of rank one matrix estimation contaminated with additive Gaussian noise (1.2), it is known that if the spike v is sampled uniformly at random from the unit sphere, PCA recovers the underlying signal, whenever its detection is possible [66]. In contrast, for rank one tensor estimation under additive Gaussian noise [75], there exists an extensive gap between the threshold for detection [50], and the threshold where tractable algorithms are successful [18, 47, 83]. Thus at first glance, the matrix and tensor problems appear to behave very differently. However, as the present paper establishes, once the planted spike is sufficiently sparse, a hard phase re-appears in the matrix problem.

This problem has natural connections to the planted clique problem [5], sparse PCA [6], bi-clustering [23, 40, 78], and community detection [1, 64, 67]. All these problems are expected to exhibit a statistical-computational gap—there are regimes where optimal statistical performance might be impossible to achieve using computationally feasible statistical procedures. The potential existence of such fundamental computational barriers has attracted significant attention recently in Statistics, Machine Learning, and Theoretical Computer Science. A number of widely distinct approaches have been used to understand this phenomenon better—prominent ones include average case reductions [22, 25, 26, 29, 45, 61], convex relaxations [14, 30, 37, 46, 60, 62], query lower bounds [39, 76], and direct analysis of specific algorithms [12, 32, 55]. The submatrix recovery problem itself has been investigated from a number of distinct perspectives. We defer an in-depth discussion of these approaches to the end of the Introduction.

In comparison to the approaches originating from Computer Science and Optimization, a completely different perspective to this problem comes from statistical physics, particularly the study of spin glasses. This approach seeks to understand algorithmic hardness in random optimization problems as a consequence of the underlying geometry of the problem— specifically, the structure of the near optimizers. The Overlap Gap property (OGP) has emerged as a recurrent theme in this context (for an illustration, see Fig 1.3). At a high level, the Overlap Gap Property (OGP) looks at approximate optimizers in a problem, and establishes that any two such approximate optimizers must either be close to each other, or far away. In other words, the one-dimensional set of distances

2

between the near optimal states is disconnected. This property has been established for various problems arising from theoretical computer science and combinatorial optimization, for instance random constraint satisfaction [2, 41, 63], Max Independent Set [40, 74], and a maxcut problem on hypergraphs [31]. Further, OGP has been shown to act as a barrier to the success of a family of "local algorithms" on sparse random graphs [33, 31, 40, 41]. This perspective has been introduced to the study of inference problems arising in Statistics and Machine Learning [18, 40, 42, 43, 44] in recent works by the first two authors which has yielded new insights into algorithmic barriers in these problems. As an aside, we note that exciting new developments in the study of mean field spin glasses [3, 65, 81], establish that in certain problems without OGP, it is actually possible to obtain approximate optimizers using appropriately designed polynomial time algorithms. This lends further credence to the belief that OGP captures a fundamental algorithmic barrier in random optimization problems— understanding this phenomenon in greater depth remains an exciting direction for future research.

1.1. **Results.** We initiate the study of *approximate support recovery* in the setting of (1.2). To introduce this notion, let us begin with the following definitions and observations. For $v \in \{0,1\}^N$, define the *support* of v to be the subset $S(v) \subset [N]$, such that

$$S(v) = \{k \in [N] : v_k = 1\}.$$

To estimate the support, it is evidently equivalent to produce an estimator $\hat{v}$ that takes values in the Boolean hypercube $\{0,1\}^N$. Observe that if $\hat{v} \in \Sigma_N(\rho)$ is drawn uniformly at random, then the intersection of the support of $\hat{v}$ and v satisfies

$$\frac{1}{N}|S(v) \cap S(\hat{v})| = \frac{1}{N}(\hat{v}, v) \to \rho^2,$$

where $(\cdot, \cdot)$ denotes the usual Euclidean inner product. For an estimator $\hat{v}$ of v, in the following, we call the normalized inner product $(\hat{v}, v)/N$ the *overlap* of the estimator with v, or simply the overlap. We are interested in the statistical and algorithmic feasibility of recovering a non-trivial fraction of the support. To this end, we introduce the notion of approximate recovery, which is defined as follows.

Definition 1.1 (Approximate Recovery). A sequence of estimators $\hat{v}_N = \hat{v}_N(A) \in \Sigma_N(\rho_N)$ is said to recover the support of v approximately if there exists $c > 0$, independent of N, ρ, such that $\langle v, \hat{v} \rangle > c\rho N$ with high probability as $N \to \infty$.

Observe that since $\hat{v}, v \in \Sigma_N(\rho_N)$, $\hat{v}$ recovers the support approximately if and only if it achieves a non-trivial Hamming loss, i.e., $d_H(v, \hat{v}) \leq 2N\rho(1-c)$ for some $c > 0$ with high probability. (Here d_H is the Hamming distance).

We study here the question of approximate support recovery in this context, and exhibit a regime of parameters (λ, ρ) where this problem exhibits OGP. This provides rigorous evidence to the existence of a computationally hard phase in this problem, as suggested in [58, 59]. To substantiate this claim, we establish that OGP acts as a barrier to the success of certain Markov-chain based algorithms. Finally, we show that for very large signal strengths λ, approximate support recovery is easy, and can be achieved by rounding the largest eigenvector of A.

To state our results in detail, we first introduce the Maximum Likelihood Estimator (MLE) in this setting. Let

$$H_N(x) = (x, Wx), \tag{1.3}$$

and consider the MLE for v,

$$\hat{v}_{ML} = \operatorname{argmax}_{x \in \Sigma_N(\rho_N)}(x, Ax) = \operatorname{argmax}_{x \in \Sigma_N(\rho_N)}\left\{H_N(x) + \frac{\lambda}{N}(x, v)^2\right\}. \tag{1.4}$$
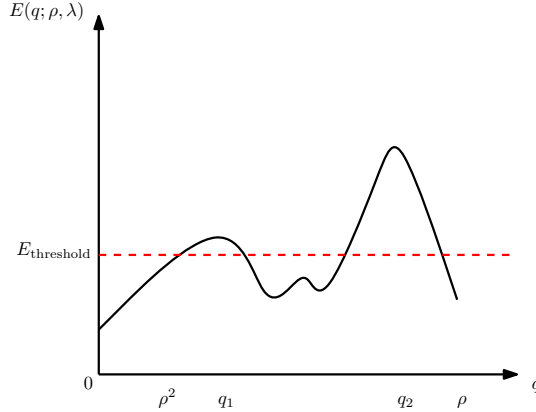
3

FIGURE 1. An illustration of OGP. Observe that under the conditions of Definition 1.3, there exists a certain threshold energy $E_{\text{threshold}}$ such that the set of possible overlaps $\{\frac{1}{N}(x,v) : x \in \Sigma_N(\rho), (x,Ax) > NE_{\text{threshold}}\}$ has a gap.

Our first result is information theoretic, and derives the minimum signal strength λ required for approximate support recovery

Theorem 1.2. *If* $\lambda = o\left(\sqrt{\frac{1}{\rho}\log\frac{1}{\rho}}\right)$, *approximate recovery is information theoretically impossible. On the other hand, for any* $\varepsilon > 0$, *if* $\lambda > (2+\varepsilon)\sqrt{\frac{1}{\rho}\log\frac{1}{\rho}}$, *the MLE recovers the support approximately.*

Thus for any $\varepsilon > 0$ and $\lambda > (2+\varepsilon)\sqrt{\frac{1}{\rho}\log\frac{1}{\rho}}$, there exists at least one estimator, namely, the MLE, which performs approximate recovery. However, the MLE is computationally intractable in general. Our next result analyzes the likelihood landscape for this problem, and establishes that the problem exhibits OGP for certain (λ, ρ) parameters in this phase.

To this end, we introduce a version of *the overlap gap property* in this context. Consider the constrained maximum likelihood,

$$E_N(q; \rho, \lambda) = \frac{1}{N} \max_{\substack{x \in \Sigma_N(\rho_N) \\ (x,v) = Nq}} H_N(x) + \lambda q^2, \tag{1.5}$$

which denotes the maximum likelihood subject to the additional constraint of achieving overlap q with v. For any $q \in [0, \rho]$, fix a sequence $q_N \to q$ such that $Nq_N \in \mathbb{N}$ and $Nq \in [Nq_N - 1/2, Nq_N + 1/2)$. We establish in Theorem 2.1 below that for all $q \in [0, \rho]$, we have that

$$E_N(q_N; \rho_N, \lambda) \to E(q; \rho, \lambda)$$

as $N \to \infty$ and that $E(q; \rho, \lambda)$ can be computed by a deterministic Parisi-type [69] variational problem. In the subsequent, we refer to $E(q; \rho, \lambda)$ as the *constrained ground state energy*, or simply *the constrained energy*. We are now in the position to define the overlap gap property.

Definition 1.3. For some $\varepsilon > 0$, we say that the model (1.2) with sparsity ρ and signal-to-noise ratio $\lambda > 0$ exhibits the *ε-overlap gap property (ε-OGP)* if there exist three points $w < x < y < \rho$ such that:

(1) $w < \rho^2 < x$, $y < \rho^{1+\varepsilon}$,
(2) $\max\{E(w; \rho, \lambda), E(y; \rho, \lambda)\} < E(x; \rho, \lambda)$,
(3) $\sup_{(w,\varepsilon\rho]} E(q; \rho, \lambda) < \sup_{[\varepsilon\rho,\rho]} E(q; \rho, \lambda)$.

4

4

Our main result establishes that the planted submatrix problem (1.2) admits the overlap gap property in the limit of high sparsity and moderate signal-to-noise ratios.

Theorem 1.4. *For any $\alpha < 2 - \sqrt{2} \approx 0.586$ and $C_1 > 2$, there exist $\varepsilon > 0$, ~~and C_2~~ such that for ρ sufficiently small, for all $C_1 \sqrt{\frac{1}{\rho} \log \frac{1}{\rho}} < \lambda < \left(\frac{1}{\rho}\right)^{\alpha}$, the planted sub-matrix problem has the ε-overlap gap property.*

Note that by Theorem 1.2, approximate recovery is possible in the entire regime covered by Theorem 1.4; however, the likelihood-landscape exhibits OGP in this part of the parameter space. Put simply, the overlap gap property states that the MLE achieves an overlap q_2 that is substantially better than ρ^2, but in the interval $[0, q_2]$, the constrained energy has another local maximum. Heuristically, this suggests that a local optimization procedure, if initialized uniformly at random (and thus starting at overlap ρ^2), will get trapped at a local maximum q_1 which is sub-optimal as compared to the true global optimum in terms of both the likelihood and the overlap. We illustrate this notion visually in Figure 1.1.

Observe that the hard phase becomes more prominent as $\rho \to 0$.

Finally, we establish that the OGP established above acts as a barrier to a family of local MCMC type algorithms. To this end, consider a Gibbs distribution on the configuration space $\Sigma_N(\rho_N)$ with

$$\pi_\beta(\{x\}) \propto \exp(\beta(x, Ax)),$$

for some $\beta > 0$ and A defined as in (1.2). Note that for any fixed N, as $\beta \to \infty$, the sample from π_β approximates the MLE 1.4. Thus a simple proxy for the MLE seeks to sample from the distribution π_β for β sufficiently large. It is natural to use local Markov chains to sample from this distribution. Specifically, construct a graph with vertices being the elements of $\Sigma_N(\rho_N)$, and add an edge between two states $x, x' \in \Sigma_N(\rho_N)$ if they are at Hamming distance 2. Finally, let Q_x denote a nearest-neighbor Markov chain on this graph started from $X_0 = x$ that is reversible with respect to the stationary distribution π_β. The following theorem establishes hitting time bounds for any such Markov Chain.

Theorem 1.5. *For (λ, ρ) as in Theorem 1.4, there exist points $a < \rho^2 < b$ with $b/\rho \to 0$ as $\rho \to 0$ and an $h > 0$ such that for some $c > 0$, with probability $1 - O(e^{-cN})$, if $\mathcal{I} = (a, b)$, then the exit time of $\mathcal{I}$, $\tau_{\mathcal{I}^c}$, satisfies*

$$\int Q_x(\tau_{\mathcal{I}^c} \leq T) d\pi_\beta(x|\mathcal{I}) \leq T e^{-chN}.$$

The proof of this result immediately follows by combining Theorem 1.4 with Corollary 3.4. Thus OGP indeed acts as a barrier to these algorithms, and furnishes rigorous evidence of a hard phase in this problem.

As an aside, we also note that Theorem 2.1 implies

$$\lim_{N \to \infty} \frac{1}{N} \max_{x \in \Sigma_N(\rho_N)} H_N(x) = \max_{q \in [0,\rho]} E(q; \rho, 0) \quad \text{a.s.}$$

Our proof of Theorem 1.4 establishes that $\{E(q; \rho, 0) : q \in [0, \rho]\}$ is maximized at $q = \rho^2$. Thus, observing that $W \stackrel{d}{=} (G + G^T)/\sqrt{2N}$, where $G = (G_{ij}) \in \mathbb{R}^{N \times N}$ is a matrix of i.i.d. $\mathcal{N}(0, 1)$ random variables, we have,

$$\lim_{N \to \infty} \frac{1}{N^{3/2}} \max_{x \in \Sigma_N(\rho_N)} (x, Gx) = \frac{1}{\sqrt{2}} E(\rho^2; \rho, 0) \quad \text{a.s.} \tag{1.6}$$

To each $N\rho_N \times N\rho_N$ principal submatrix of the matrix G, assign a score, which corresponds to the sum of its entries. The LHS of (1.6) represents the largest score, as we scan over all $N\rho_N \times N\rho_N$ principal submatrices of G. In [23], the authors derive upper and lower bounds on the maximum average value of $k \times k$ submatrices in an iid Gaussian matrix, for $k \leq \exp(o(\log N))$. Our corollary

extends the results of [23] to the case of principal submatrices with size $N\rho_N \times N\rho_N$, and provides a tight first order characterization of the maximum. We note that in spin-glass terminology, the score of the largest $N\rho_N \times N\rho_N$ submatrix (not necessarily principal) corresponds to the ground state of a bipartite spin-glass model[17], which is out of reach of current techniques.

Returning to the planted model (1.2), we complete our discussion by establishing that when the signal λ is appropriately large, approximate support recovery is easily achieved using a spectral algorithm. To this end, we introduce the following two-step estimation algorithm, which rounds the leading eigenvector of A.

(1) Let $\hat{v} = (\hat{v}_i)$ denote the eigenvector corresponding to the largest eigenvalue of the matrix A, with $\|\hat{v}\|^2 = N\rho$. Denote $\tilde{S} = \{i \in [N] : \hat{v}_i \geq \frac{\delta}{2}\}$.

(2) If $|\tilde{S}| < \rho N$, sample $(\rho N - |\tilde{S}|)$ elements at random from $[N]\backslash\tilde{S}$ and augment to the set $\tilde{S}$ in order to construct $\hat{S}$.

(3) Otherwise, sample ρN elements uniformly at random from $\tilde{S}$, and denote this set as $\hat{S}$.

(4) Finally, construct $v_{\hat{S}} = \mathbf{1}_{\hat{S}}$, i.e., a vector with ones corresponding to the entries in $\hat{S}$, and zeros otherwise.

Observe that the set $\hat{S}$ depends on δ. We keep this dependence implicit for notational convenience. Then we have the following lemma.

Lemma 1.6. *For any $\varepsilon > 0$ and $\lambda > (1+\varepsilon)\frac{1}{\rho}$, there exists $\delta := \delta(\varepsilon)$ such that with high probability as $N \to \infty$, $v_{\hat{S}}$ recovers the support approximately.*

1.2. **Comparison with existing results.** In this section, we compare our approach with existing results, and summarize our main technical contributions. The information theoretic limits for exact recovery of planted submatrices under Gaussian additive noise was derived in [28]. [29] studies the submatrix localization problem under additive sub-Gaussian noise, and characterizes the information theoretic threshold for exact recovery. Further, they derive a computational lower bound via an average case reduction to a variant of the planted clique problem. Their results establish a crucial distinction between submatrix detection and localization problems—in contrast to the detection problem, there always exists a gap between the information theoretic threshold and the computational one in the localization problem. [12, 55] study the performance of specific thresholding approaches for the submatrix localization problem. We emphasize that this line of work focuses on *exact recovery*, and thus these results are not directly comparable with our results.

Closer in spirit to our work are the recent results in [38]—they study exact recovery for the planted Wigner model (1.2). Their results correspond to planted $k \times k$ principal submatrices with $k = o(n)$, and thus do not compare directly with our results. Based on the low degree likelihood ratio method, they suggest the existence of a hard regime if

$$\sqrt{\frac{k}{N}} \leq \lambda\rho \leq \min\left\{1, \frac{k}{\sqrt{N}}\right\}.$$

Note that while their conjecture is restricted to the regime $k = o(N)$, their thresholds formally agree with our conclusions upon setting $k = N\rho$. In particular, it suggests that the threshold for the spectral algorithm derived in Lemma 1.6 is the computational threshold for this problem. Improving our OGP result to the entire regime $\sqrt{\frac{1}{\rho}\log\frac{1}{\rho}} \ll \lambda \ll \frac{1}{\rho}$ is an intriguing open problem. We leave this for future study. Motivated by these conjectures, [20] have studied the regime $k = o(N)$ using the lens of OGP, and have provided rigorous structural evidence to the existence of a hard phase for exact recovery.

In [13], the authors study detection and approximate recovery in for the planted Wigner model (1.2). They assume that $N\rho$ many non-zero coordinates are chosen uniformly at random, and each non-zero entry assumes $\{\pm\rho^{-1/2}\}$ with probability $1/2$. Using a sophisticated conditional second

moment approach, they identified $\lambda = 2\sqrt{\rho \log \frac{1}{\rho}}(1 + o_\rho(1))$ as the sharp threshold for detection and weak recovery, where $o_\rho(\cdot)$ represents a term that converges to zero as $\rho \to 0$. Further, the authors conjecture a hard phase in the regime $2\sqrt{\rho \log \frac{1}{\rho}} \ll \lambda \ll \frac{1}{\rho}$. We note that their information theoretic threshold is sharper than our Theorem 1.2, while the conjectured hard regime matches our predictions, upto the leading constant in the lower bound. It should be possible to extend our methods to establish OGP in this problem. We leave this for future investigations.

The results of [57] derive the sharp (up to constant) threshold for approximate recovery in a related Bayesian problem. Assuming a product prior on the spike v (1.2), they characterize the limiting (normalized) mutual information. We note that in this Bayesian setting, the analysis is substantially simplified by the "Nishimori Identity"—in statistical physics parlance, the problem is "replica symmetric", and thus the limiting mutual information is expressed in terms of a simple scalar optimization problem. In contrast, the MLE corresponds to the ground state of the model, and in general exhibits "full replica symmetry breaking". To analyze its performance, we need to leverage recent advances in the study of mean field spin glasses, specifically related to the analysis of Parisi type formulas. We believe that our approach can also be useful in analyzing the landscape of the posterior distribution in high dimensional inference problems—we leave this for future study. Finally, the limiting (normalized) mutual information has been derived in a version of this problem, corresponding to the $k = o(N)$ setting [16]. These results have been derived by an appropriate extension of the *adaptive interpolation method*.

Since the completion of the first version of this article, [77] has provided further evidence to the existence of a computationally hard phase in the spiked Wigner problem using the low degree framework.

1.3. **Technical Contributions.** Let us now discuss the main technical contributions of our paper. The proof of Theorem 1.4 is in two parts: we develop a variational formula for the large N limit of the maximum likelihood constrained to a given overlap, and analyze this variational problem.

We begin by viewing $\Sigma_N(\rho_N)$ as the configuration space of a spin system with energy H_N. We compute the free energy of this spin system on the subspace of fixed overlaps. On this subspace, H_N can be viewed as a two species generalized Sherrington-Kirkpatrick model, and we compute the free energy via Guerra interpolation and the Aizenman–Sims–Star scheme [4]. Due to the special symmetries of the problem considered here, we can develop a Parisi-type formula for the free energy as the sum of two Parisi type functionals and a Lagrange multiplier term in the spirit of [70], with modifications to account for the change in alphabet, the additional constraints, and the "multiple species". The key observation is that, due to these symmetries, we can decouple the two species through a Lagrange multiplier argument. With the positive temperature free energy in hand, it remains to compute the zero temperature limit. We do so using the results of [51], which computed the zero-temperature Γ-limit of the non-linear term in general Parisi-type functionals.

Let us pause to comment on the relation between this approach and prior work. Free energies of spin glasses with multiple species and non-symmetric alphabets have been studied in other works [71, 73, 72, 49], and typically require substantially more sophisticated tools than what we use here. In particular, prior results crucially use Panchenko's synchronization mechanism. We emphasize that the free energy we consider here cannot be directly obtained from these works. In principle, one might compute it by extending the work of [49] to more general alphabets — we expect this would require significantly more advanced machinery from the theory of spin glasses and variational calculus. Furthermore, the variational problems obtained by this approach are typically intractable. Our approach bypasses these issue, and identifies a simple formula in this special case.

Let us now turn to the analysis of the "ground state energy" functional. We establish the overlap gap property by analyzing the sign changes of the first derivative of the restricted ground state

energy functional. This analysis requires a precise understanding of the scaling properties of the derivative as $\rho \to 0$. To this end, we utilize the natural connection of Parisi functionals with a family of SDEs, which, in turn, can be explicitly analyzed in the regime $\rho \to 0$.

Outline. The remainder of this paper is structured as follows. Theorem 1.2 is established in Section 7. The first part follows by a data processing argument, and the second follows by a Slepian-type bound. Theorem 1.4 is the main contribution of this paper, and is established in Section 2. The proof depends crucially on Theorem 2.1, which derives a Parisi type variational problem for the restricted energy $E(\cdot; \rho, \lambda)$. In turn, to derive Theorem 2.1, we first establish a finite temperature Parisi formula (Proposition 4.1) in Section 6, and subsequently compute a limit of this formula as the temperature converges to zero (see Sections 4 and 5). Similar zero temperature formulae have been instrumental in establishing properties of mean field spin glasses in the low-temperature regime (see e.g. [8, 9, 53, 51]). We emphasize that even with Theorem 2.1, the proof of Theorem 1.4 is subtle, and critically depends on understanding the scaling of the variational formula as $\rho \to 0$. Lemma 1.6 is established in Section 8. Finally, Section 3 studies the effect of OGP on this problem, and establishes that certain local Markov Chain based recovery algorithms are stymied by this structural barrier.

2. Proof of the overlap gap property

To establish Theorem 1.4, we first require the following variational formula for the limiting constrained energy (1.5). Let $\mathscr{M}([0, q])$ denote the space of non-negative Radon measures on $[0, \rho]$ equipped with the weak-* topology. Let $\mathcal{A}$ denote the set

$$\mathcal{A} = \{\nu \in \mathscr{M}([0, q]) : \nu = m(s)ds + c\delta_\rho, \ m(s) \geq 0 \text{ non-decreasing}\}$$

For $\nu \in \mathcal{A}$ and $\mathbf{\Lambda} = (\Lambda_1, \Lambda_2) \in \mathbb{R}^2$, let $u_\nu^i : [0, \rho] \times \mathbb{R} \to \mathbb{R}$ solve the Cauchy problem

$$\begin{cases} \partial_t u_\nu^i + 2(\partial_x^2 u_\nu^i + m(t)(\partial_x u_\nu^i)^2) = 0 & (t, x) \in [0, \rho) \times \mathbb{R} \\ u_\nu^i(\rho, x) = (x + \Lambda_i + 2c)_+, & x \in \mathbb{R} \end{cases} \quad (2.1)$$

where $(\cdot)_+$ denotes the positive part, and ∂_x^2 is the Laplace operator. Note that any $\nu \in \mathcal{A}$ is locally absolutely continuous with respect to the Lebesgue measure on $(0, 1)$, so that m is almost surely well-defined. As explained in [51, Appendix A], there is a unique weak solution to this PDE. Consider then the functional

$$\mathcal{P}(\nu, \mathbf{\Lambda}) = \rho u_\nu^1(0, 0) + (1 - \rho) u_\nu^2(0, 0) - \Lambda_1 q - \Lambda_2(\rho - q) - 2\int s d\nu(s).$$

We then have the following,

Theorem 2.1. *For q_N, ρ_N as above, we have that*

$$E(q; \rho, \lambda) := \lim_{N \to \infty} E_N(q_N; \rho_N, \lambda) = \lambda q^2 + \min_{\nu \in \mathcal{A}, \Lambda \in \mathbb{R}^2} \mathcal{P}(\nu, \mathbf{\Lambda}),$$

almost surely.

8

We defer the proof of this result to Section 4. Let us now complete the proof of the overlap gap property, Theorem 1.4, assuming Theorem 2.1.

Proof of Theorem 1.4. By Theorem 2.1, we have that

$$E(q; \rho, \lambda) = \lambda q^2 + \min_{\nu \in \mathcal{A}, \mathbf{\Lambda}} \mathcal{P}(\nu, \mathbf{\Lambda}).$$

Observe that setting $c = \nu(\{1\})$, one may make the linear transformation $\Lambda \mapsto \Lambda + 2c$ without changing the value of the functional. Thus it suffices to consider the problem

$$\min_{\nu \in \mathcal{A}_0, \Lambda \in \mathbb{R}^2} \mathcal{P}(\nu, \Lambda) \tag{2.2}$$

where $\mathcal{A}_0$ are those $\nu \in \mathcal{A}$ with $\nu(\{\rho\}) = c = 0$. By Lemma B.2, we have that $\mathcal{P}$ is jointly strictly convex in (m, Λ) when restricted to this subspace. Thus the minimum (ν_*, Λ_*) is unique. As such we can restrict this problem further to the compact set $S = \{\nu : ||\nu||_{TV} \leq ||\nu_*||\} \cup B_\delta(\Lambda_*)$ for some $\delta > 0$. As $\lambda q^2 + \mathcal{P}(\nu, \Lambda)$ is convex in q, we may use Danskin's envelope theorem, to show that

$$E(q; \rho, \lambda) = \lambda q^2 + \min_S \mathcal{P}(\nu, \Lambda)$$

is differentiable in q with derivative

$$\frac{\partial E}{\partial q} = 2\lambda q + (\Lambda_2^* - \Lambda_1^*), \tag{2.3}$$

where Λ_i^* denote the maximizers of (2.2) corresponding to q. On the other hand one can show that u_ν^i is classically differentiable in its final time data, and thus in Λ_i for $t < \rho$ by a standard differentiable dependence argument (see, e.g., [19, Lemma A.5] for a similar agument),. Thus we may differentiate in Λ to obtain the following fixed point equation for the optimal Λ^* :

$$\frac{\partial}{\partial \Lambda_1} u_\nu^1(0, 0) = \frac{q}{\rho}$$
$$\frac{\partial}{\partial \Lambda_2} u_\nu^2(0, 0) = \frac{\rho - q}{1 - \rho}. \tag{2.4}$$

Recall that u_ν^i solves the PDE (2.1), and thus $\partial_{\Lambda_i} u_\nu^i$ is given by the solution $w^i = \partial_{\Lambda_i} u_\nu^i$ to

$$\begin{cases} \partial_t w^i + L_i w^i = 0 \\ w^i(\rho, x) = \mathbb{1}_{x + \Lambda_i \geq 0}, \end{cases} \tag{2.5}$$

where $L_i = 2\partial_x^2 + 4m(t)\partial_x u_\nu^i \partial_x$ is an elliptic operator. Observe that L_i is the infinitesimal generator of the diffusion X_t^i, given by the solution to the stochastic differential equation

$$dX_t^i = 2dB_t + 4m(t)\partial_x u_\nu^i(t, X_t^i)dt. \tag{2.6}$$

By Ito's lemma we have, for $t \in (0, \rho)$, $x \in \mathbb{R}$, the stochastic representation formula for $\partial_{\Lambda_i} u_\nu^i$,

$$\partial_{\Lambda_i} u_\nu^i(t, x) = \mathbb{P}(X_\rho^i + \Lambda_i \geq 0 | X_t^i = x).$$

In particular, for $t = x = 0$, we have

$$\partial_{\Lambda_i} u_\nu^i(0, 0) = \mathbb{P}(X_\rho^i + \Lambda_i \geq 0 | X_0^i = 0). \tag{2.7}$$

Thus (2.4) and (2.3) identifies Λ_i^* with quantiles of X_ρ^i.

Next, we derive bounds on Λ_1^*, Λ_2^* by analyzing the diffusion itself. Since $\partial_x u_\nu^i$ is weakly differentiable, we see that it weakly solves (2.5) as well. Thus

$$\partial_x u_\nu^i(t, x) = \mathbb{P}(X_\rho^i + \Lambda_i \geq 0 | X_t^i = x) \tag{2.8}$$

as well. In particular we obtain the maximum principle $0 \leq \partial_x u_\nu^i \leq 1$. Finally we require the following estimate on $\int_0^\rho m(t)dt$ for the unique minimizer $m(s)$.

9

Lemma 2.2. *We have that* $\int_0^\rho m(t)dt \le \frac{1}{2}\sqrt{\rho \log(1/\rho)}$.

We defer the proof of this estimate to Section 5 and complete the proof assuming this estimate.

Using $0 \le \partial_x u_\nu^i \le 1$ to control the drift term in (2.6), we bound the above probability as

$$\mathbb{P}\left(2B_\rho \ge -\Lambda_i | B_0 = 0\right) \le \mathbb{P}\left(X_\rho^i \ge -\Lambda_i | X_0^i = 0\right) \le \mathbb{P}\left(2B_\rho \ge -\Lambda_i - 4\int_0^\rho m(t)dt\right).$$

Combining this with the stationary point conditions for Λ_j^* (2.4), and Lemma 2.2, we have, setting Φ as the CDF of a standard Gaussian random variable,

$$2\sqrt{\rho}\,\Phi^{-1}\left(\frac{q}{\rho}\right) - 2\sqrt{\rho \log \frac{1}{\rho}} \le \Lambda_1^* \le 2\sqrt{\rho}\,\Phi^{-1}\left(\frac{q}{\rho}\right)$$

$$2\sqrt{\rho}\,\Phi^{-1}\left(\frac{\rho-q}{1-\rho}\right) - 2\sqrt{\rho \log \frac{1}{\rho}} \le \Lambda_2^* \le 2\sqrt{\rho}\,\Phi^{-1}\left(\frac{\rho-q}{1-\rho}\right). \tag{2.9}$$

Armed with these bounds on the optimizers Λ_1^*, Λ_2^*, we can complete the proof in a relatively straightforward manner. First, observe that at $q = \rho^2$, (2.4) ensures that $\Lambda_1^* = \Lambda_2^*$, and consequently, (2.3) implies

$$\frac{\partial}{\partial q}E(\rho^2; \rho, \lambda) = 2\lambda\rho^2 > 0.$$

Finally, we evaluate $\partial_q E(q, \rho, \lambda)$ at $q = \rho^{2-\delta}$, In particular, since $\alpha < 2 - \sqrt{2}$ and $\lambda \le \rho^{-\alpha}$, choose δ such that

$$\alpha < \frac{3 - 2\delta}{2} < 2 - \sqrt{2}.$$

In this case, we again have, using (2.3),

$$\partial_q E(\rho^{2-\delta}; \rho, \lambda) = 2\rho^{2-\delta}\lambda + (\Lambda_2^* - \Lambda_1^*). \tag{2.10}$$

Again, using the bounds derived in (2.9), we have,

$$\partial_q E(\rho^{2-\delta}; \rho, \lambda) \le 2\rho^{2-\delta}\lambda + 2\sqrt{\rho}\,\Phi^{-1}\left(\frac{\rho - \rho^{2-\delta}}{1-\rho}\right) - 2\sqrt{\rho}\,\Phi^{-1}(\rho^{1-\delta}) + 2\sqrt{\rho \log \frac{1}{\rho}},$$

$$\le 2\lambda\rho^{2-\delta} - 2\sqrt{2\rho \log \frac{1}{\rho}}(1 + o_\rho(1)) + (2\sqrt{(1-\delta)} + \sqrt{2})\sqrt{2\rho \log \frac{1}{\rho}})$$

$$\le \sqrt{2\rho \log \frac{1}{\rho}}\left(\lambda\sqrt{\frac{\rho^{3-2\delta}}{\log(\frac{1}{\rho})}} - 2 + 2\sqrt{(1-\delta)} + \sqrt{2} + o(1)\right).$$

If we take ρ sufficiently small, we obtain

$$\partial_q E(\rho^{2-\delta}; \rho, \lambda) < 0$$

by our choice of δ.

The above calculation establishes that there exists and $\varepsilon' > 0$ such that for any $C_1 > 0$, $\rho > 0$ sufficiently small, and $\lambda \in \left(C_1\sqrt{\frac{\log \frac{1}{\rho}}{\rho}}, \rho^{-\alpha}\right)$, there exist $\rho^2 < \tilde{q}_0 < \rho^{1+\varepsilon'}$ such that

$$\partial_q E(\rho^2; \rho, \lambda) > 0, \quad \partial_q E(\tilde{q}_0; \rho, \lambda) < 0.$$

Thus there exist $w < \rho^2 < x < y = \tilde{q}_0$ such that

$$\max\{E(w; \rho, \lambda), E(y; \rho, \lambda)\} < E(x; \rho, \lambda).$$

10

Next, we claim that if we choose $C_1 > 2$, then $\max_{(w, \theta \rho)} E(q; \rho, \lambda) < \max_{(\theta \rho, \rho)} E(q; \rho, \lambda)$ for $0 < \theta < \theta_0$ for some $\theta_0 = \theta_0(C_1) > 0$. Assuming this claim, if we take $\varepsilon = \varepsilon' \wedge \theta_0$, we see that the points $w < x < y$ satisfy (i)-(iii) in definition 1.3. In particular, ε-OGP holds as desired.

As the proof of this claim is subsumed by our proof of Part 2 of Theorem 1.2 we end by proving said theorem. (Specifically, see (2.13) below and take $\theta_0(C_1) = c(C_1 - 2)$ there.) $\qquad\square$

Proof of Part 2 of Theorem 1.2. Fix $\varepsilon > 0$, and let $0 < c = c(\varepsilon) < 1$, to be specified later. Recall again that by Slepian's comparison inequality [24], comparing H_N to an IID process with the same variance, yields

$$\mathbb{E} \max_{x \in A} H_N(x) \leq \sqrt{2 \max_x \mathrm{Var}(H_N(x)) \log|A|}, \tag{2.11}$$

for any $A \subset \Sigma_N$. Applying (2.11) in the case $A = \Sigma_N(\rho_N)$, yields

$$\lim_{N \to \infty} \frac{1}{N} \mathbb{E} \Big[\max_{x \in \Sigma_N(\rho_N)} (x, Wx) \Big] \leq 2\sqrt{\rho^3 \log \frac{1}{\rho}} (1 + o_\rho(1)). \tag{2.12}$$

By (2.11) and the concentration of Gaussian maxima [24, Theorem 5.8], with high probability as $N \to \infty$,

$$\frac{1}{N} \max_{x \in \Sigma_N(\rho_N):(x,v) \leq cN\rho} \Big[\frac{\lambda}{N}(x, v)^2 + H_N(x) \Big] \leq \lambda c^2 \rho^2 + 2\sqrt{\rho^3 \log \frac{1}{\rho}} (1 + o_\rho(1)) + o_N(1).$$

On the other hand, plugging-in $x = v$,

$$\frac{1}{N} \max_{x \in \Sigma_N(\rho_N):(x,v) > cN\rho} \Big[\frac{\lambda}{N}(x, v)^2 + H_N(x) \Big] \geq \lambda \rho^2 + o_N(1).$$

Thus selecting c such that $(2 + \varepsilon) \cdot (1 - c^2) > 2$, we have

$$\frac{1}{N} \max_{x \in \Sigma_N(\rho_N):(x,v) \leq cN\rho} \Big[\frac{\lambda}{N}(x, v)^2 + H_N(x) \Big] < \frac{1}{N} \max_{x \in \Sigma_N(\rho_N):(x,v) > cN\rho} \Big[\frac{\lambda}{N}(x, v)^2 + H_N(x) \Big] \tag{2.13}$$

for all small enough $\rho > 0$, implying $(\hat{x}_{\mathrm{MLE}}, v) \geq cN\rho$ with high probability as $N \to \infty$. We conclude that the MLE recovers the support approximately in this regime. $\qquad\square$

3. From Overlap Gap Property to Free Energy Wells

In this section, we will show that the overlap gap property implies a hardness type result for Monte Carlo Markov chains. To this end, let us first define the relevant dynamics. Fix $\beta > 0$ and consider the Gibbs distribution

$$\pi_\beta(dx) \propto e^{\beta(x, Ax)} dx,$$

where dx is the counting measure on $\Sigma_N(\rho)$. Construct a graph G_N with vertices $\Sigma_N(\rho)$, and add an edge between $x, x' \in \Sigma_N(\rho)$ if and only if their Hamming distance is exactly 2. Let $(X_t)_{t \geq 0}$ denote any nearest neighbor Markov chain on G_N, reversible with respect to the stationary distribution π_β. By this we mean that the transition matrix Q for X_t satisfies detailed balance with respect to π_β and $Q(x, y) = 0$ if x and y are not connected by an edge. We show here that when OGP holds, if we run the Markov Chain with initial data, $\pi_\beta(dx|\mathcal{I})$, where $\mathcal{I}$ is as in Theorem 1.5 and β is sufficiently large, it takes at least exponential time for the chain to hit the region of order ρ overlap.

11

3.1. **Free energy wells.** Let $G = (V, E)$ be a finite graph and ν denote a probability measure on V. For $a \in \mathbb{R}$ and an $\epsilon > 0$, let $B_\epsilon(a) = [a - \epsilon, a + \epsilon]$. For any function $f : V \to \mathbb{R}$, consider the following "rate function",

$$I_f(a; \epsilon) = -\log \nu(\{x : f(x) \in B_\epsilon(a)\}).$$

For any two vertices $x, y \in V$ we say that $x \sim y$ if x and y are connected by an edge. We say that a function $f : V \to \mathbb{R}$ is K-Lipschitz if there is some K such that

$$\max_{x \sim y} |f(x) - f(y)| \leq K.$$

Definition 3.1. We say that f has an ϵ-*free energy well* of depth h in $[a, b]$ if there exists a $c \in [a, b]$ such that $B_\epsilon(a), B_\epsilon(b)$ and $B_\epsilon(c)$ are disjoint and

$$\min \{I_f(a, \epsilon), I_f(b, \epsilon)\} - I_f(c, \epsilon) \geq h.$$

We then have the following whose proof is an adaptation of [18, Theorem 7.4] to this setting. See also [44].

Theorem 3.2. *Let X_t denote a nearest neighbor Markov chain on G which is reversible with respect to ν. If f is ϵ-Lipschitz and has an ϵ-free energy well of depth h in $[a, b]$, then for $\mathscr{A} = f^{-1}([a, b])$, the exit time of $\mathscr{A}$, denoted $\tau_{\mathscr{A}^c}$, satisfies*

$$\int Q_x(\tau_{\mathscr{A}^c} \leq T) d\nu(x | \mathscr{A}) \leq Te^{-h}$$

for any T, where Q_x is the law of X_t.

Proof. In the following, let $\mathscr{A}_{-\epsilon} = f^{-1}([a, b] \backslash (B_\epsilon(a) \cup B_\epsilon(b)))$, $\mathscr{A} = f^{-1}([a, b])$, and $B_\epsilon = \mathscr{A} \backslash \mathscr{A}_{-\epsilon}$. Let us define the boundary of $\mathscr{A}$ to be the set

$$\partial \mathscr{A} = \{x \in \mathscr{A} : \exists y \in \mathscr{A}^c : x \sim y\}.$$

Observe that since f is ϵ-Lipschitz, $\partial \mathscr{A} \subset B_\epsilon$.

Let $\bar{X}_t$ be the Markov chain defined on $\mathscr{A}$ which is X_t reflected at the boundary of $\mathscr{A}$. That is, $\bar{X}_t$ has transition matrix $(\bar{Q}(x, y))$ which is identical to $\mathscr{A}$ if $x \in \mathscr{A} \backslash \partial \mathscr{A}$ and $y \in \mathscr{A}$ and for $x \in \partial \mathscr{A}$,

$$\bar{Q}(x, y) \propto \begin{cases} Q(x, y) & y \in \mathscr{A} \\ 0 & \text{else.} \end{cases}$$

Note that by detailed balance $\bar{X}_t$ is reversible with respect to $\tilde{\nu} = \nu(\cdot | \mathscr{A})$, the invariant measure of X_t conditioned on $\mathscr{A}$. Let $\tau_{\partial \mathscr{A}}$ denote the first time either X_t or $\bar{X}_t$ hits ∂A. Note that for $t \leq \tau_{\partial \mathscr{A}}$, the Markov Chains X_t and $\bar{X}_t$, started from a common state in $\mathscr{A}$ follow the same trajectory. As a result,

$$\int_{\mathscr{A}} Q_x(\tau_{\partial \mathscr{A}} < T) d\tilde{\nu}(x) = \int_{\mathscr{A}} \bar{Q}_x(\tau_{\partial \mathscr{A}} < T) d\tilde{\nu}(x).$$

We now estimate the right hand side. Since $\partial \mathscr{A} \subset B_\epsilon$,

$$\int_{\mathscr{A}} \bar{Q}_x(\tau_{\partial \mathscr{A}} < T) d\tilde{\nu} \leq \int_{\mathscr{A}} \bar{Q}_x(\exists i \leq T : X_i \in B_\epsilon) d\tilde{\nu} \leq \sum_i \int_{\mathscr{A}} \bar{Q}_x(X_i \in B_\epsilon) d\tilde{\nu} = T\tilde{\nu}(B_\epsilon) \leq Te^{-h},$$

where the last equality follows by stationarity and the last inequality follows by assumption that f has an ϵ-free energy well of height h. $\qquad \square$

3.2. **From the Overlap Gap Property to Free energy wells at low temperature.** We now establish that if the overlap gap property holds, then the overlap $m(x) = \frac{1}{N}(x, v)$ has a free energy well for $\beta > 0$ sufficiently large.

In the following, let

$$\Sigma_N(\rho, q) = \left\{ x \in \Sigma_N(\rho) : \sum_{i=1}^{N\rho} x_i = Nq, \ \sum_{i=N\rho+1}^{N} x_i = N(\rho - q) \right\}. \tag{3.1}$$

In defining the set $\Sigma_N(\rho, q)$, we implicitly use the distributional invariance of A under row/column permutations to assume, without loss of generality, that v is of the form

$$v = (1, \ldots, 1, 0, \ldots, 0), \tag{3.2}$$

where the first $N\rho_N$ entries are 1 and the remaining are 0. For $\lambda, \beta > 0$, let

$$F_N(\lambda, \beta, \rho, q) = \frac{1}{N} \mathbb{E} \log \left(\sum_{x \in \Sigma_N(\rho, q)} \exp \left(\beta \big(\lambda N q^2 + H_N(x) \big) \right) \right), \tag{3.3}$$

where H_N is defined by (1.3) and $\Sigma_N(\rho, q)$ is defined in (3.1). Let $F_N(\lambda, \beta, \rho_N)$ be defined similarly except with the sum running over the set $\Sigma_N(\rho_N)$. Using [51, Lemma 2.6], we have,

$$\left| \frac{1}{\beta} F_N(\lambda, \beta, \rho_N, q_N) - \mathbb{E}\left[\frac{1}{N} \max_{x \in \Sigma_N(\rho_N, q_N)} (x, Ax) \right] \right| \leq \frac{\log|\Sigma_N(\rho_N, q_N)|}{N\beta} \leq \frac{C(\rho, q)}{\beta}, \tag{3.4}$$

for some $C(\rho, q) > 0$ independent of N. Observe that by combining this bound with (1.5) implies that if the limit $F(\lambda, \beta, \rho, q) = \lim F_N(\lambda, \beta, \rho_N, q_N)$ exists, then

$$E(q; \rho, \lambda) = \lim_{\beta \to \infty} \frac{1}{\beta} F(\lambda, \beta, \rho, q).$$

Theorem 3.3. *For any $\varepsilon, \lambda, \rho > 0$, if the ε-overlap gap property holds, then there are points $w < \rho^2 < x < y < \rho$ with $y = o_\rho(\rho)$ such that for N sufficiently large and any $\delta > 0$, the overlap $m(x)$ has an δ-free energy well of depth Nh in $[w, y]$ with probability $1 - O(e^{-cN})$.*

Proof. By definition of the ε-overlap gap property, we may take $w < \rho^2 < x < y = o_\rho(\rho)$, such that

$$\max\{E(w; \rho, \lambda), E(y; \rho, \lambda)\} < E(x; \rho, \lambda).$$

Furthermore, by continuity of the map $q \mapsto E(q; \rho, \lambda)$ (see (2.3)), we may assume that there is an $\delta > 0$ and $h > 0$, such that

$$\max_{q \in B_\delta(w) \cup B_\delta(y)} E(q; \rho, \lambda) - \min_{q' \in B_\delta(x)} E(q'; \rho, \lambda) \leq -h$$

and such that $B_\delta(w), B_\delta(y)$, and $B_\delta(x)$ are pairwise disjoint. By (3.4), we then have that

$$\max_{q \in B_\delta(w) \cup B_\delta(y)} F_N(\lambda, \beta, q) - \min_{q' \in B_\delta(x)} F_N(\lambda, \beta, q') \leq -\frac{h}{2}$$

for β sufficiently large.

For a set $E \subset \Sigma_N$, let $Z_N(E) = \sum_{x \in E} e^{\beta(x, Ax)}$, and let $\mathcal{A}_N = m(\Sigma_N(\rho))$ denote the image of the overlap function. Note that $|\mathcal{A}_N \cap B_\delta(a)| \leq C \cdot N \cdot \delta$ for some constant $C > 0$. By a union bound, we have

$$\left| \frac{1}{N} \log Z_N(B_\delta(a)) - \max_{q \in B_\delta(a) \cap \mathcal{A}_N} \frac{1}{N} \log Z_N(\Sigma_N(\rho, q)) \right| \leq C \cdot \delta.$$

13

Recall that by Gaussian concentration of measure [69, Theorem 1.2], there is a $C > 0$ such that for $N \geq 1$ and $\eta > 0$,

$$\max_{q \in \mathcal{A}_N} \frac{1}{N} \log \mathbb{P}\Big(|F_N(\lambda, \beta, q) - \frac{1}{N} \log\left(Z_N(\Sigma_N(\rho, q))\right)| \geq \eta\Big) \leq -C\eta^2$$

$$\frac{1}{N} \log \mathbb{P}\Big(|F_N(\lambda, \beta) - \frac{1}{N} \log Z_N(\Sigma_N(\rho))| \geq \eta\Big) \leq -C\eta^2.$$

If we take δ and η sufficiently small, we see that

$$\max\Big\{\frac{1}{N} \log Z_N(B_\delta(w)), \frac{1}{N} \log Z_N(B_\delta(y))\Big\} - \frac{1}{N} \log Z_N(B_\delta(x)) \leq -\frac{h}{4},$$

with probability $1 - O(e^{-cN})$, where we have combined the above concentration bounds with a union bound, using the fact again that $|\mathcal{A}_N| \leq C \cdot N$. Setting I_m as the rate-function corresponding to the overlap m with respect to the measure $\pi_\beta \propto \exp(\beta(x, Ax))$ on $\Sigma_N(\rho)$, and subtracting the above from $\frac{1}{N} \log Z_N(\Sigma_N(\rho))$, we have,

$$\min\{I_m(a; \delta), I_m(b; \delta)\} - I_m(q_1; \delta) \geq N\frac{h}{4}$$

with probability $1 - O(e^{-cN})$. This yields the desired result. $\qquad\square$

Observe that $m(x)$ is $1/N$-Lipschitz. Combining Theorem 3.2 with Theorem 3.3 then immediately yields the following corollary.

Corollary 3.4. *For any $\varepsilon, \lambda, \rho > 0$, if the ε-overlap gap property holds, then there are points $w < \rho^2 < x < y < \rho$ with $y = o_\rho(\rho)$ and an $h > 0$ such that with probability $1 - O(e^{-c'N})$, if $\mathcal{I} = (a, b)$, then the exit time of $\mathcal{I}$, $\tau_{\mathcal{I}^c}$, satisfies*

$$\int Q_x(\tau_{\mathcal{I}^c} \leq T) d\pi(x|\mathcal{I}) \leq T e^{-chN}$$

for some $c > 0$.

4. VARIATIONAL FORMULA FOR CONSTRAINED ENERGY

We establish Theorem 2.1 in this section. To begin we introduce a relaxation of the optimization problem (1.4), called the "positive temperature free energy" of the problem. Recall that we may assume, without loss of generality, that v is of the form

$$v = (1, \ldots, 1, 0, \ldots 0), \tag{4.1}$$

where the first $N\rho_N$ entries are 1 and the remaining are 0. Recall that ρ_N is a sequence such that $N\rho_N \in \mathbb{N}$ and $\rho_N \to \rho$. For any $q \in [0, \rho]$, fix a sequence $q_N \to q$ such that $Nq_N \in \mathbb{N}$ and $Nq \in [Nq_N - 1/2, Nq_N + 1/2)$. In the subsequent, we will refer to a sequence $(\rho_N, q_N) \to (\rho, q)$ that satisfies these conditions as an *admissible sequence*. We begin by deriving the following formula for the limiting free energy, $F(\beta, \rho, q) = \lim F_N(0, \beta, \rho_N, q_N)$, where F_N is as in (3.3) and $\beta > 0$ is fixed.

To this end, let $\Lambda_1, \Lambda_2 \in \mathbb{R}$, and $\mu \in \mathcal{M}_1([0, q])$, where $\mathcal{M}_1([0, q])$ is the space of probability measures on $[0, q]$ equipped with the weak-* topology. Let $u^i_{\mu,\beta}$ be the unique weak solutions to the Cauchy problem

$$\begin{cases} \partial_t u^i_{\mu,\beta} + 2(\partial_x^2 u^i_{\mu,\beta} + \beta\mu([0, t])(\partial_x u^i_{\mu,\beta})^2) = 0 & (t, x) \in [0, \rho) \times \mathbb{R} \\ u^i_{\mu,\beta}(\rho, x) = \frac{1}{\beta} \log\left(1 + \exp(\beta(x + \Lambda_i))\right). \end{cases} \tag{4.2}$$

14

For a definition of weak solution as well as well-posedness of the Cauchy problem see [52, Sec. 2] (alternatively, see [19, Appendix A]). Consider the functional

$$\mathcal{P}_\beta(\mu, \Lambda_1, \Lambda_2; q) = \rho u^1_{\mu,\beta}(0,0) + (1-\rho)u^2_{\mu,\beta}(0,0) + \Lambda_1 q - \Lambda_2(\rho - q) - 2\int_0^\rho s\beta\mu([0,s])ds \quad (4.3)$$

We then have the following.

Proposition 4.1. *For $\beta > 0$ and any admissible sequence $(\rho_N, q_N) \to (\rho, q)$ we have that $F(\beta, \rho, q) = \lim_{N\to\infty} F_N(0, \beta, \rho_N, q_N)$ exists and satisfies*

$$\frac{1}{\beta}F(\beta, \rho, q) = \min_{\substack{\mu \in \mathcal{M}_1([0,q]), \\ \Lambda_1, \Lambda_2 \in \mathbb{R}}} \mathcal{P}_\beta(\mu, \Lambda_1, \Lambda_2; q) + \lambda q^2. \quad (4.4)$$

In particular, this minimum is achieved.

We defer the proof of this result to Section 6.

4.1. **Proof of variational formula.** We now compute the zero-temperature limit of the positive temperature problem.

Theorem 4.2. *We have that*

$$\lim_{\beta\to\infty} \frac{1}{\beta}F(\beta, \rho, q) = \min_{\nu \in \mathcal{A}, \mathbf{\Lambda} \in \mathbb{R}^2} \mathcal{P}(\nu, \mathbf{\Lambda}; q) + \lambda q^2.$$

Recall that by (3.4), this immediately implies Theorem 2.1.

To this end, we study the convergence of the above variational problem as $\beta \to \infty$. Let us first recall the notion of sequential Γ-convergence.

Definition 4.3. Let X be a topological space. We say that a sequence of functionals $F_n : X \to [-\infty, \infty]$ *sequentially Γ-converges* to $F : X \to [-\infty, \infty]$ if

(1) The $\Gamma - \underline{\lim}$ inequality holds: For every x and sequence $x_n \to x$,

$$\varliminf_{n\to\infty} F_n(x_n) \geq F(x).$$

(2) The $\Gamma - \overline{\lim}$ inequality holds: For every x, there exists a sequence $x_n \to x$ such that

$$\varlimsup_{n\to\infty} F_n(x_n) \leq F(x).$$

For a sequence of functionals F_β indexed by a real parameter β, we say that F_β sequentially Γ-converges to F if for any sequence $\beta_n \to \infty$, the sequence F_{β_n} sequentially Γ converges to F.

Recall that in [51, Theorem 3.2] it was shown that the functionals

$$\mathcal{F}_\beta(\nu, \Lambda_i) = \begin{cases} u^i_{\mu,\beta}(0,0) & \nu = \beta\mu([0,s])ds, \quad \mu \in \mathcal{M}_1([0,q]) \\ +\infty & \text{o.w.} \end{cases}$$

sequentially Γ-converges to the solution of (2.1),

$$\mathcal{F}_\beta(\nu, \Lambda_i) \xrightarrow{\Gamma} u^i_\nu(0,0). \quad (4.5)$$

Let

$$\mathcal{G}_\beta(\nu, \Lambda_1, \Lambda_2) = \rho\mathcal{F}_\beta(\nu, \Lambda_1) + (1-\rho)\mathcal{F}_\beta(\nu, \Lambda_2)$$
$$\mathcal{G}(\nu, \mathbf{\Lambda}) = \rho u^1_\nu(0,0) + (1-\rho)u^2_\nu(0,0).$$

15

Furthermore, let

$$\mathcal{E}_\beta(\nu, \boldsymbol{\Lambda}) = \mathcal{G}_\beta(\nu, \Lambda_1, \Lambda_2) - \Lambda_1 q - \Lambda_2(\rho - q) - \int s d\nu(s),$$

$$\mathcal{E}(\nu, \boldsymbol{\Lambda}) = \mathcal{G}(\nu, \boldsymbol{\Lambda}) - \Lambda_1 q - \Lambda_2(\rho - q) - \int s d\nu(s).$$

The preceding results (with minor modification) will yield the following result.

Lemma 4.4. *We have that* $\mathcal{E}_\beta \xrightarrow{\Gamma} \mathcal{E}$.

Proof. By (4.5),we have the $\Gamma - \underline{\lim}$ inequality: for any sequence $(\nu_\beta, \boldsymbol{\Lambda}^\beta)$,

$$\underline{\lim}\, \mathcal{G}_\beta(\nu_\beta, \Lambda_1^\beta, \Lambda_2^\beta) \geq \rho \underline{\lim}\, \mathcal{F}_\beta(\nu_\beta, \Lambda_i^\beta) + (1 - \rho) \underline{\lim}\, \mathcal{F}_\beta(\nu_\beta, \Lambda_i^\beta) = \mathcal{G}(\nu, \boldsymbol{\Lambda}).$$

It remains to prove the $\Gamma - \overline{\lim}$ upper bound. This will follow since the recovery sequence in the $\Gamma - \overline{\lim}$ upperbound of (4.5) does not depend on the functional itself.

More precisely, fix $(\nu, \boldsymbol{\Lambda})$. Consider the sequence $(\nu_\beta, \boldsymbol{\Lambda}^\beta)$ with $\boldsymbol{\Lambda}^\beta = \boldsymbol{\Lambda}$, and ν_β constructed as in [53, Lemma 2.1.2] (see alternatively [51, Lemma 3.4]). Consequently, we have that

$$\lim \mathcal{G}_\beta(\nu_\beta, \boldsymbol{\Lambda}) = \lim \mathcal{F}_\beta(\nu_\beta, \Lambda_1) + \mathcal{F}_\beta(\nu_\beta, \Lambda_2) = \mathcal{G}(\nu, \boldsymbol{\Lambda}),$$

along this sequence as each of the summands converge by applying [51, Lemma 3.3] termwise. □

Lemma 4.5. *Any sequence of minimizers of* $\mathcal{E}_\beta$ *is pre-compact. Furthermore, any limit point of such a sequence is a minimizer of* $\mathcal{E}$ *and*

$$\lim \min \mathcal{E}_\beta = \min \mathcal{E}.$$

In fact, this sequence is unique, however we will not require this. The compactness of ν_β is established in the following lemma, whose proof is deferred to Section 5.

Lemma 4.6. *For every* $\beta > 0$,

$$\|\nu_\beta\|_{TV} \leq \frac{1}{2}\sqrt{\rho \log(1/\rho)}.$$

Proof of Lemma 4.5. The compactness of ν_β follows from Lemma 4.6. On the other hand, by Lemma 5.1 below, Λ_i^β lie in a uniformly bounded set. Thus the sequence is pre-compact. The second half of the result follows by the fundamental theorem of Γ-convergence. □

Proof of Theorem 4.2. This follows by combining Lemma 4.4-4.5, and Proposition 4.1. □

Proof of Theorem 2.1. This follows immediately upon combining (3.4) with Theorem 4.2. □

5. Bounds on optimal measures

In this section, we prove Lemmas 2.2 and 4.6. To this end, we need the following useful notation. Let $\boldsymbol{\Lambda} = (\Lambda_1, \Lambda_2)$, and let us abuse notation to denote by $d\rho$, the two atomic measure on $\{1, 2\}$,

$$d\rho = \rho\delta_1 + (1 - \rho)\delta_2.$$

Let $v_i(t, x)$ be given by the change of variables

$$v_i(t, x) = \beta u_{\mu,\beta}^i(t, x/\beta; \Lambda_i/\beta). \tag{5.1}$$

Note that v_i solves

$$\begin{cases} \partial_t v + 2\beta^2(\partial_x^2 v + \mu([0, s])(\partial_x v)^2) = 0 & (t, x) \in [0, \rho) \times \mathbb{R}^2 \\ v(\rho, x) = \log(1 + \exp(x + \Lambda)), \end{cases} \tag{5.2}$$

16

with $\Lambda = \Lambda_i$. Throughout the following when studying the problem at finite β, it is helpful to rewrite the functional $\mathcal{P}_\beta$ to simplify notation. To this end, consider the following functional,

$$P_\beta(\mu, \Lambda_1, \Lambda_2; q) = \int v_i(0,0) d\rho(i) - q\Lambda_1 - (\rho - q)\Lambda_2 - 2\beta^2 \int s\mu(s)ds.$$

and notice that

$$P_\beta(\mu, \Lambda_1, \Lambda_2; q) = \beta \mathcal{P}_\beta(\mu, \mathbf{\Lambda}/\beta, q). \tag{5.3}$$

In the following, it will be useful to note the first order optimality conditions for this functional. To this end, let $\hat{X}_s$ solve the stochastic differential equation

$$d\hat{X}_s^i = 4\beta^2 \mu([0,t])\partial_x v_i(s, \hat{X}_s^i) + 2\beta\, dB_t \tag{5.4}$$

with initial data $\hat{X}_0^i = 0$, and let

$$G_{\mu,\mathbf{\Lambda}}(t) = \int_t^\rho 4\beta^2 \Big(\int \mathbb{E}(\partial_x v_i)^2(s, \hat{X}_s^i)d\rho(i) - s \Big)ds.$$

Lemma 5.1. *We have the following.*

(1) *For every $\beta > 0$ there is a unique minimizing triple $(\mu, \mathbf{\Lambda})$ of P_β. Furthermore, the set of optimizing $\{\mathbf{\Lambda}_\beta\}_\beta$ lie in a compact subset of $\mathbb{R}^2$.*

(2) *This triple satisfies the optimality conditions*

$$\mu(G_{\mu,\mathbf{\Lambda}}(s) = \min_s G_{\mu,\mathbf{\Lambda}}(s)) = 1 \tag{5.5}$$

$$\int \partial_{\Lambda_i} v_i(0,0)d\rho(i) = \rho. \tag{5.6}$$

(3) *In particular, for any $q \in \mathrm{supp}(\mu)$,*

$$q = \int \mathbb{E}(\partial_x v_i)^2(q, \hat{X}_q^i)d\rho(i). \tag{5.7}$$

Proof. We begin with item (1), and first prove the uniqueness of the minimizing pair. To see this, we note that by [51, Lemma 4.2], the map

$$(\mu, \Lambda) \mapsto w(0,0)$$

where w weakly solves (5.2) is strictly convex. Thus P_β is strictly convex.

To show existence of a minimizing pair, note that since $\mathcal{M}_1([0,q])$ is weak-* compact, it suffices to show that $\mathbf{\Lambda}$ lives in a compact subset of $\mathbb{R}^2$. By the parabolic comparison principle (see [51, Lemma 4.6] in this case), it follows that, for any μ, $v_i(0,0) \geq \log(1 + \exp(\Lambda_i))$. Thus

$$P_\beta(\mu, \Lambda_1, \Lambda_2; q) \geq \log(1 + \exp(\Lambda_1)) - q\Lambda_1 + \log(1 + \exp(\Lambda_2)) - (\rho - q)\Lambda_2 - \int sds,$$

which diverges to infinity as $\Lambda_1, \Lambda_2 \to \pm\infty$, from which the compactness result follows. In fact, this shows that the set is β independent.

It remains to derive the optimality conditions. For item (3), note that the fixed point equation (5.7) for the support follows upon differentiating G and applying (5.5). To obtain (5.6), first note that v_i are differentiable in Λ_i — this follows by a classical differentiable dependence argument, see, e.g., [19, Lemma A.5]. Explicitly differentiating the functional in Λ_i, (5.6) then follows upon observing the relation

$$\rho\Big(\frac{q}{\rho}\Big) + (1 - \rho)\Big(\frac{\rho - q}{1 - \rho}\Big) = \rho.$$

The first-order stationary condition for G (5.5) then follows by first fixing Λ_i and computing the first variation of the maps

$$\mu \mapsto v_i(0,0; \Lambda_i).$$

This has been done in [51, Lemma 4.3] following [54, Lemma 3.2.1]. In particular, this yields the following first variation formula for P_β: if μ_t is a weak-* right differentiable path in $\mathcal{M}_1([0, q])$ ending at μ, in the sense that

$$\dot{\mu} = \lim_{t \to 0^+} \frac{\mu_t - \mu}{t}$$

exists weak-*, then

$$\frac{d}{dt}|_{t=0} P_\beta(\mu_t, \mathbf{\Lambda}; q) = \int G_{\mu, \mathbf{\Lambda}}(s) d\dot{\mu}.$$

By the first order optimality condition for convex functions, it follows that the righthand side is non-negative for all such paths if and only if we choose μ_0 to be the optimizer of $P_\beta(\cdot, \mathbf{\Lambda}; q)$. If we then take the path $\mu_t = t\delta_{s_0} + (1 - t)\mu$, we see that

$$G_{\mu, \mathbf{\Lambda}}(s_0) \geq \int G_{\mu, \mathbf{\Lambda}} d\mu,$$

from which (5.5) follows. $\qquad\square$

Recall the function F introduced in Proposition 4.1. Armed with Lemma 5.1, we next establish a formula for the β-derivative of F.

Lemma 5.2. *We have that*

$$\partial_\beta F(\beta, \rho, q) = 2\beta \int (\rho^2 - q^2) d\mu(q).$$

Proof. We start with (4.3) and (4.4), and observe that we can equivalently express

$$F(\beta, \rho, q) = \min_{\Lambda_i, \mu} P_\beta(\mu, \Lambda_1, \Lambda_2; q).$$

By the same argument as [51, Theorem 4.1], we see that

$$\partial_\beta v_i(0, 0) = \beta \left\{ 4\rho \mathbb{E} \left(\partial_x^2 v_i(\rho, \hat{X}_\rho^i) + (\partial_x v_i)^2(\rho, \hat{X}_\rho^i) \right) - 4 \int_0^\rho s \mathbb{E}(\partial_x v_i)^2(s, \hat{X}_s^i) d\mu \right\}, \qquad (5.8)$$

where $\hat{X}^i$ solves the stochastic differential equation (5.4). [19, Lemma A.5] implies that $w_i = \partial_{\Lambda_i} v(t, x)$ weakly solves

$$\begin{cases} \partial_t w_i + L_i w = 0 & (t, x) \in [0, \rho) \times \mathbb{R} \\ w_i(T, x) = \partial_{\Lambda_i} \log(1 + \exp(x + \Lambda_i)) & x \in \mathbb{R}, \end{cases}$$

where L_i is the infinitesimal generator for $\hat{X}^i$. Therefore,

$$\mathbb{E} \left[\partial_{\Lambda_i} v_i(\rho, \hat{X}_\rho^i) | \hat{X}_0^i = 0 \right] = \partial_{\Lambda_i} v_i(0, 0).$$

By direct computation and (5.2),

$$\partial_x^2 v_i(\rho, x) + (\partial_x v_i)^2(\rho, x) = \partial_{\Lambda_i} v_i(\rho, x).$$

Combining these observations with (5.8), and using the optimality conditions (5.6) and (5.7) yields

$$\partial_\beta F(\beta, \rho, q) = \rho \partial_\beta v_1(0, 0) + (1 - \rho) \partial_\beta v_2(0, 0) - 4\beta \int_0^\rho s \mu([0, s]) ds$$

$$= \beta \left[4\rho \cdot \rho - 4 \int_0^\rho s^2 d\mu(s) \right] - 2\beta \int_0^\rho (\rho^2 - s^2) d\mu(s)$$

$$= 2\beta \int (\rho^2 - s^2) d\mu(s)$$

as desired. $\qquad\square$

18

We now turn to the proof of Lemmas 2.2 and 4.6.

Proofs of Lemmas 2.2 and 4.6. Let us begin by observing that Lemma 2.2 follows from Lemma 4.6. To see this, recall the functional $\mathcal{P}_\beta$ from (4.4) and let μ_β denote the corresponding minimizer. Then, by Lemma 4.5, there is a limit point of the sequence $\nu_\beta = \beta\mu_\beta([0,s])dt$, call it $\nu = \tilde{m}(t)dt + c\delta_\rho$, and any such limit point is a minimizer of $\mathcal{P}$. Observe that, by the reduction from (2.2), and the strict convexity of the corresponding problem, $\tilde{m} = m$. We now observe that along any subsequence converging to ν,

$$\int_0^\rho m(t)dt \leq \lim_{\beta\to\infty} \beta \int_0^\rho \mu_\beta([0,t])dt. \tag{5.9}$$

The desired bound then follows from Lemma 4.6.

Let us now turn to the proof of Lemma 4.6. By Fubini's theorem, we have that

$$\beta \int_0^\rho \mu_\beta([0,t])dt = \beta \int_0^\rho (\rho - q)d\mu_\beta(q) \leq \frac{1}{2\rho}\beta \int_0^\rho (\rho^2 - q^2)d\mu_\beta(q) = \frac{1}{4\rho}\partial_\beta F(\beta,\rho,q),$$

where the last inequality follows using Lemma 5.2. Next, recall $F_N(\beta,\rho_N,q_N)$ from (3.3). By differentiation, observe that $F_N(\beta,\rho_N,q_N)$ is convex in β, and thus Proposition 4.1 implies that $F(\beta,\rho,q)$ is convex as well. Thus $\partial_\beta F_N(\beta,\rho_N,q_N) \to \partial_\beta F(\beta,\rho,q)$, by Griffith's lemma for convex functions. Consequently, we have,

$$\beta \int_0^\rho \mu_\beta([0,t])dt \leq \frac{1}{4\rho}\lim_{N\to\infty}\partial_\beta F_N(\beta,\rho_N,q_N).$$

Finally, note that

$$\partial_\beta F_N(\beta,\rho_N,q_N) = \frac{1}{N}\mathbb{E}[\langle H_N\rangle] \leq \frac{1}{N}\mathbb{E}\max_{\Sigma_N(\rho_N,q_N)} H_N \leq 2\sqrt{\rho^3\log\frac{1}{\rho}}(1 + o_N(1)),$$

where $\langle\cdot\rangle$ denotes integration with respect to the Gibbs measure $\pi(\{\sigma\}) \propto \exp(-H_N(\sigma))\mathbb{1}_{\Sigma_N(\rho_N,q_N)}$, the first equality follows by Gaussian integration-by-parts [69, Lemma 1.1], and the last bound follows by bounding $\langle H_N\rangle$ by the maximum and applying Slepian's comparison inequality (2.11) with $A = \Sigma_N(\rho_N,q_N)$. Thus we obtain,

$$\int_0^\rho \beta\mu_\beta([0,t])dt \leq \frac{1}{2}\sqrt{\rho\log 1/\rho},$$

as desired. $\qquad\square$

6. Proof of Free energy formula

In this section, we aim to prove that for $(\rho_N, q_N) \to (\rho, q)$ admissible,

$$\lim_{N\to\infty} F_N(\beta,\rho_N,q_N) = \inf_{\substack{\mu\in\mathscr{M}_1([0,\rho])\\ \mathbf{\Lambda}\in\mathbb{R}^2}} P_\beta(\mu,\mathbf{\Lambda};q), \tag{6.1}$$

where we recall the functional P_β from (5.3). Note that this agrees with the statement of Proposition 4.1 after dividing through by β.

To this end, first write v as in (4.1). We may then view $x \in \{0,1\}^N$ as $x \in \{0,1\}^{N_1} \times \{0,1\}^{N_2}$ where $N_1 = N\rho_N$. We call x the configuration, and x_i the spin of the i particle. We call $I_1 = [N_1]$ the *first species* of particles and $I_2 = [N]\backslash I_1$ the *second species* of particles. We may then view the log-likelihood, H_N, as the *Hamiltonian* for a *two-species spin glass* model,

$$H_N(x) = \frac{\sqrt{2}}{\sqrt{N}}\sum_{i,j} g_{ij}x_ix_j,$$

where here g_{ij} are i.i.d. $\mathcal{N}(0,1)$ (and in particular, (g_{ij}) is not symmetric). Thus our goal is to compute the free energy of this two-species spin glass model constrained to certain classes of configurations.

In the following, it will be useful to define the *overlap* between two points $\sigma^1, \sigma^2 \in \Sigma_N(\rho)$ by

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

When the notation is unambiguous we will also denote $R(\sigma^i, \sigma^j) = R_{ij}$. (In particular, we let $R_{11} = R(\sigma^1, \sigma^1)$). It will also be useful to define the *intraspecies overlaps*,

$$R^{(1)}(\sigma^1, \sigma^2) = \frac{1}{N_1} \sum_{i=1}^{N_1} \sigma_i^1 \sigma_i^2, \quad R^{(2)}(\sigma^2, \sigma^2) = \frac{1}{N_2} \sum_{i=N_1+1}^{N} \sigma_i^1 \sigma_i^2.$$

Note that for $i, j = 1, 2$, we have $R_{ij} = \rho_N R_{ij}^{(1)} + (1 - \rho_N) R_{ij}^{(2)}$. The overlap distribution plays a central role in the analysis of spin glass systems as it is at the heart of the (non-rigorous) replica symmetry breaking method [?] as well as the Aizenman–Sims–Starr approach [4].

6.1. **The Upper Bound.** Let $\mathcal{A}_r$ denote the rooted tree of depth r where each non-leaf vertex has countably many children, i.e., the first r levels of the Ulam-Harris tree. Note that we may view the leaves of this tree as the set $\mathbb{N}^r$, where $\alpha = (\alpha_1, \ldots, \alpha_r)$ denotes the root leaf path $\emptyset \to \alpha$. We denote this path by $p(\alpha)$. For more on this notation see Appendix A.

Theorem 6.1. *Suppose that (ρ, q) is such that $\Sigma_N(\rho, q)$ is non-empty. Then for $N \geq 1$, we have that*

$$\mathbb{E}F_N(\beta, \rho, q) \leq \inf_{\mu, \Lambda} P_\beta(\mu, \Lambda; q). \tag{6.2}$$

Proof. Let us begin with the case where μ has finite support. Fix $r \geq 1$ and a pair of sequences

$$0 = Q_0 < \ldots < Q_r = \rho$$
$$0 = \mu_{-1} < \mu_0 < \ldots < \mu_r = 1,$$

and let μ be such that $\mu(\{Q_k\}) = \mu_k - \mu_{k-1}$. Let $(v_\alpha)_{\alpha \in \mathbb{N}^r}$ denote a Ruelle probability cascade (RPC) corresponding to parameters (μ_ℓ) (For a definition of Ruelle probability cascades see Appendix A.). Let $(Z(\alpha))$ denote the centered Gaussian process on $\mathbb{N}^r$ with covariance

$$\mathbb{E}Z(\alpha)Z(\gamma) = 4\beta^2 Q_{|\alpha \wedge \gamma|}, \tag{6.3}$$

where $|\alpha \wedge \gamma|$ denotes the depth of the least common ancestor of α and γ in $\mathcal{A}_r$, and let $(Z_i(\alpha))$ denote i.i.d. copies of this process. Similarly, let $(Y(\alpha))$ be the centered Gaussian process on $\mathbb{N}^r$ with covariance

$$\mathbb{E}Y(\alpha)Y(\gamma) = 2\beta^2 Q_{|\alpha \wedge \gamma|}^2. \tag{6.4}$$

Take the processes Y and Z to be independent of each other and H_N. For $t \in [0, 1]$, define the interpolating Hamiltonian $H_t(\sigma, \alpha) : \Sigma_N(\rho, q) \times \mathbb{N}^r \to \mathbb{R}$ given by

$$H_t(\sigma, \alpha) = \sqrt{t}(\beta H_N(\sigma) + \sqrt{N}Y(\alpha)) + \sqrt{1-t} \left(\sum_{i=1}^{N} Z_i(\alpha)\sigma_i \right).$$

Finally let

$$\phi(t) = \frac{1}{N} \mathbb{E} \log \sum_{\alpha, \sigma} v_\alpha \exp(H_t(\sigma, \alpha)).$$

To control this, let us recall Gaussian integration-by-parts for Gibbs measures (see, e.g., [69, Lemma 1.1]):

20

Lemma. *(Integration by parts for Gibbs measures) Let $\mathcal{X}$ be at most countable and let $(u(x)),(v(x))$ be centered Gaussian processes on $\mathcal{X}$ with mutual covariance $C(x,y) = \mathbb{E}u(x)v(y)$. Let G be the probability measure with $G(\{x\}) \propto \exp(v(x))$. We have the identity*

$$\mathbb{E}\int_{\mathcal{X}} u(x)dG(x) = \mathbb{E}\int\int_{\mathcal{X}^2} C(x^1,x^1) - C(x^1,x^2)dG^{\otimes 2}. \tag{6.5}$$

If we let $(\sigma^\ell,\alpha^\ell)$ denote independent draws from the Gibbs measure

$$\pi_t(\{\sigma,\alpha\}) \propto v_\alpha \exp(H_t(\sigma,\alpha)),$$

then integrating by parts, we have that

$$\phi'(t) = \frac{1}{N}\mathbb{E}\langle \partial_t H_t(\sigma,\alpha)\rangle = \frac{1}{N}\mathbb{E}\langle C_{11} - C_{12}\rangle$$

where

$$C_{ij} = \mathbb{E}\partial_t H_t(\sigma^i,\alpha^i)\cdot H_t(\sigma^j,\alpha^j) = \frac{N}{2}\beta^2(R_{ij} - Q_{|\alpha^i \wedge \alpha^j|})^2$$

for $i,j = 1,2$. In the case that $i = j$ observe that, by definition of $\Sigma_N(\rho,q)$, $R_{11} = \rho = Q_r$ so that $C_{11} = 0$. Thus $\phi' \leq 0$. The result then follows by comparing the boundary conditions. In particular, re-arranging the inequality $\phi(1) \leq \phi(0)$ yields

$$\mathbb{E}F_N \leq \frac{1}{N}\mathbb{E}\log\sum_\alpha v_\alpha \sum_{\sigma\in\Sigma_N(\rho,q)} \exp(\sum_i Z_i(\alpha)\sigma_i) - \frac{1}{N}\mathbb{E}\log\sum_\alpha v_\alpha \exp(\sqrt{N}Y(\alpha)), \tag{6.6}$$

By Corollary A.3, the second term is equal to

$$\frac{1}{N}\mathbb{E}\log\sum_\alpha v_\alpha e^{\sqrt{N}Y(\alpha)} = \frac{1}{2}\int 4\beta^2 s\mu([0,s])ds.$$

It remains to upper-bound the first term.

To this end, observe that the set $\Sigma_N(\rho,q)$ is defined via a constraint on the intra-species overlaps. If we add Lagrange multipliers, Λ_1 and Λ_2, for the constraints that $R_{11}^{(1)} = \rho$ and $R_{11}^{(2)} = (\rho-q)$, the first term in (6.6) is equal to

$$\frac{1}{N}\mathbb{E}\log\sum_\alpha v_\alpha \sum_{\sigma\in\Sigma_N(\rho,q)} \exp\Big(\sum_i Z_i(\alpha)\sigma_i + \Lambda_1 N_1 R_{11}^{(1)} + \Lambda_2 N_2 R_{11}^{(2)}\Big) - \Lambda_1 q - \Lambda_2(\rho-q)$$

$$\leq \frac{1}{N}\mathbb{E}\log\sum_\alpha v_\alpha \sum_{\sigma\in\Sigma_N} \exp\Big(\sum_i Z_i(\alpha)\sigma_i + \Lambda_1 N_1 R_{11}^{(1)} + \Lambda_2 N_2 R_{11}^{(2)}\Big) - \Lambda_1 q - \Lambda_2(\rho-q), \tag{6.7}$$

where the inequality follows by the set containment $\Sigma_N(\rho,q) \subseteq \Sigma_N = \{0,1\}^N$. Observe that the summation in σ is now over a product space, and that $R_{11}^{(\ell)} = \frac{1}{N_1}\sum_{i\leq N_1}\sigma_i^\ell$, since $\sigma_i \in \{0,1\}$. Thus the first term in this display is of the form

$$\frac{1}{N}\mathbb{E}\log\sum_\alpha v_\alpha \prod_{i=1}^{N_1}(1+\exp(Z_i(\alpha)+\Lambda_1)) \prod_{i=N_1+1}^{N}(1+\exp(Z_i(\alpha)+\Lambda_2)).$$

Applying (both parts of) Theorem A.2 we see that this is given by

$$\rho\mathbb{E}\log\sum_\alpha v_\alpha(1+\exp(Z(\alpha)+\Lambda_1))+(1-\rho)\mathbb{E}\log\sum v_\alpha(1+\exp(Z(\alpha)+\Lambda_2)) = \rho\varphi_\mu^1(0,0)+(1-\rho)\varphi_\mu^2(0,0),$$
$$\tag{6.8}$$

where here we used that $N = N_1 + N_2 = \rho N + (1-\rho)N$, and $\varphi_\mu^i = v_i$ for v_i as in (5.1). This gives the desired upper bound,

$$\mathbb{E}F_N(\beta,\rho,q) \leq P_\beta(\mu,\Lambda,q) \tag{6.9}$$

21

for μ of finite support.

We obtain the result for general μ by continuity. Specifically, recall that u_μ^i is weak-* continuous in μ (see [52, Sec 2.4]) and the last term in the definition of P_β is integration of μ against a continuous function. (Alternatively, we may use Lemma 6.5 below for both terms.) Thus P_β weak-* continuous in μ. Since the set of measures of finite support is weak-* dense in $\mathscr{M}_1([0,\rho])$, we obtain (6.9) for all $\mu \in \mathscr{M}_1([0,\rho])$ and Λ. Minimizing yields the desired bound. □

6.2. **Lower bound via Aizenman-Sims-Starr scheme.** It now remains to prove the matching lower bound. In the following let $Z_N(E) = \sum_{\sigma \in E} \exp(H(\sigma))$. We begin with the following inequality: for any $M \geq 1$,

$$\varliminf_{N\to\infty} \frac{1}{N} \mathbb{E} \log Z_N(\Sigma_N(\rho_N, q_N)) \geq \varliminf_{N\to\infty} \frac{1}{M} \left[\mathbb{E} \log Z_{N+M}(\Sigma_{N+M}(\rho_N, q_N)) - \mathbb{E} \log Z_N(\Sigma_N(\rho_N, q_N)) \right].$$

We call these new M coordinates *cavity coordinates*. Let us take $M = M_1 + M_2$ where we add M_1 cavity coordinates to the first species and M_2 to the second. Throughout the following we will take the admissible sequence to be such that

$$|\rho_N - \rho|, |q_N - q| \leq \frac{C}{N}. \tag{6.10}$$

(The bound for q_N follows by definition of admissibility.)

Let us now decompose the Hamiltonian into the part induced by the cavity and the remaining. We set

$$H'_N(\sigma) = \frac{\sqrt{2}\beta}{\sqrt{N+M}} \sum_{i,j=1}^N g_{ij}\sigma_i\sigma_j,$$

$$z_{N,i}(\sigma) = \frac{\sqrt{2}\beta}{\sqrt{N+M}} \sum_{j=1}^N (g_{ij} + g_{ji})\sigma_j,$$

$$y_N(\sigma) = \frac{\sqrt{2}\beta}{\sqrt{N(N+M)}} \sum_{i,j=1}^N g'_{ij}\sigma_i\sigma_j,$$

where the g'_{ij} are independent of the g_{ij}. It then follows that

$$\beta H_{N+M}(\sigma) = H'_N(\sigma) + \sum_{i=N+1}^{N+M} z_{N,i}(\sigma)\sigma_i + r_1(\sigma)$$

$$\beta H_N(\sigma) = H'_N(\sigma) + \sqrt{M} y_N(\sigma) + r_2(\sigma),$$

where r_i are centered Gaussian processes with $\mathrm{Var}(r_i) = O(\frac{1}{N})$. Then for $J = \{N+1, \ldots, N+M\}$,

$$\mathbb{E} \log Z_{N+M}(\Sigma_{N+M}(\rho_{N+M}, q_{N+M})) = \mathbb{E} \log \sum_{\sigma \in \Sigma_{N+M}(\rho_{N+M}, q_{N+M})} \exp(H'_N(\sigma)) \cdot \prod_{i \in J} \exp(z_{N,i}(\sigma)\sigma_i) + o(1)$$
$$\tag{6.11}$$

$$\mathbb{E} \log Z_N(\Sigma_N(\rho_N, q_N)) = \mathbb{E} \log \sum_{\sigma \in \Sigma_N(\rho_N, q_N)} \exp(H'_N(\sigma) + \sqrt{M} y_N(\sigma)) + o(1), \tag{6.12}$$

where we have eliminated the r_i dependence on the right hand side by the following lemma.

Lemma 6.2. *Let $\mathcal{X}$ be a finite set. Let $r : \mathcal{X} \to \mathbb{R}$ be a centered Gaussian process on $\mathcal{X}$ and let $H(x)$ be a Gaussian process independent of r. Then*

$$|\mathbb{E} \log \sum_{x \in \mathcal{X}} \exp(H(x) + r(x)) - \mathbb{E} \log \sum_{x \in \mathcal{X}} \exp(H(x))| \leq \max_{x \in \mathcal{X}} \mathrm{Var}(r(x)).$$

Proof. Let $Y_t(x) = H(x) + \sqrt{t}r(x)$ and

$$\psi(t) = \mathbb{E}\log\sum\exp Y_t.$$

Then if we let $G_t \in Pr(\mathcal{X})$ be $G_t(\{x\}) \propto e^{Y_t}$ then we have

$$\partial_t\psi(t) = \mathbb{E}\int\partial_t Y_t dG_t = \mathbb{E}\int C(x^1, x^1) - C(x^1, x^2)dG_t^{\otimes 2}$$

where

$$C(x, y) = \mathbb{E}\partial_t Y_t(x)Y_t(y) = \frac{1}{2}\mathbb{E}r_t(x)r_t(y) \leq \frac{1}{2}\max\mathrm{Var}(r).$$

Consequently $|\partial_t\psi(t)| \leq \mathrm{Var}(r)$. From which it follows that $|\psi(1) - \psi(0)| \leq \max\mathrm{Var}(r)$. Evaluating ψ at 0 and 1 yields the desired bound. $\qquad\square$

6.2.1. *Reduction to continuous functionals.* Our goal now is to compute the limit of the difference of (6.11) and (6.12). We begin by showing that their difference can be related to the difference of two continuous functionals on an appropriate space of probability measures. To this end, we first claim that, upon passing to a subsequence in N, we may choose a sequence $(r_M, u_M) \to (\rho, q)$ such that

$$\Sigma_{N+M}(\rho_{N+M}, q_{N+M}) \supset \Sigma_N(\rho_N, q_N) \times \Sigma_M(r_M, u_M)$$

where here we have abused notation and view the cavity coordinates $\Sigma_M = \Sigma_{M_1} \times \Sigma_{M_2}$ in this product as being distributed between the species in such a way that $N_1 + M_1$ is the size of the first species.

To see that such a sequence exists, observe that if we define $r_M(N)$ and $u_M(N)$ by

$$Mr_M(N) = (N + M)\rho_{N+M} - N\rho_N$$
$$Mu_M(N) = (N + M)q_{N+M} - Nq_N,$$

then by the choice of the sequence (ρ_N, q_N) (see (6.10)), these are both integers bounded by M. As such, we may pass to a subsequence along which the pair converges. In particular, eventually along this subsequence in N, (Mr_M, Mu_M) will be constant. The desired properties of (r_M, u_M) follows immediately by definition of (ρ_N, q_N).

Using this sequence, we may lower bound (6.11) by

$$\mathbb{E}\log\sum_{\sigma\in\Sigma_{N+M}(\rho_{N+M}, q_{N+M})}\exp(\beta H_{N+M}(\sigma))$$

$$\geq \mathbb{E}\log\sum_{\sigma\in\Sigma_N(\rho_N, q_N)}\exp(H'(\sigma)) \cdot \sum_{\epsilon\in\Sigma_M(r_M, u_M)}\prod_{i\in J}\exp(z_{N,i}(\sigma)\epsilon_i) + o(1). \tag{6.13}$$

Furthermore, since

$$\mathbb{E}y_N(\sigma^1)y_N(\sigma^2) = \frac{2\beta^2}{N(N + M)}\left(\sigma^1 \cdot \sigma^2\right)^2 = 2\beta^2\frac{N}{N + M}R_{12}^2 = 2\beta^2 R_{12}^2 + o(1),$$

if we let $(y'(\sigma))$ denote the Gaussian process with covariance

$$\mathbb{E}y'(\sigma^1)y'(\sigma^2) = 2\beta^2 R_{12}^2,$$

then we may apply Lemma 6.2 to express (6.12) as

$$\mathbb{E}\log Z_N\left(\Sigma_N(\rho_N, q_N)\right) = \mathbb{E}\log\sum_{\sigma\in\Sigma_N(\rho_N, q_N)}\exp(H_N'(\sigma) + \sqrt{M}y'(\sigma)) + o(1). \tag{6.14}$$

Define G' to be the Gibbs distribution corresponding to H' on the subset $\Sigma_N(\rho_N, q_N)$,

$$G'(\{\sigma\}) \propto \exp(H'(\sigma))\mathbb{1}_{\sigma\in\Sigma_N(\rho_N, q_N)},$$

23

23

where we keep the dependence on N implicit. Let $\langle \cdot \rangle_{G'}$ denote expectation with respect to this measure. Subtracting (6.14) from (6.13), we may lower bound the difference of (6.11) and (6.12) by

$$\mathbb{E}\log\left\langle \sum_{\epsilon\in\Sigma_M(r_M,u_M)} \exp(\sum_i z_{N,i}(\sigma)\epsilon_i) \right\rangle_{G'} - \mathbb{E}\log\left\langle \exp(\sqrt{M}y_N(\sigma)) \right\rangle_{G'},$$

up to a $o(1)$ correction. We now provide alternative representations for these two terms.

To this end, let $\mathscr{R}$ be the space of infinite Gram arrays with entries bounded by 1. Let $\mathscr{M}_{exch}$ be the subspace of probability measures on $\mathscr{R}$ that are *weakly exchangeable*, i.e., if R is drawn from some $\mathcal{R}\in\mathscr{M}_{exch}$, then

$$(R_{\ell\ell'}) \overset{(d)}{=} (R_{\pi(\ell)\pi(\ell')}),$$

for any permutation $\pi:\mathbb{N}\to\mathbb{N}$ which permutes finitely many numbers. We call such an R a *Gram-de Finetti array*.

By standard arguments, one can show that there are weak-* continuous functionals $\mathcal{F}_{i,M}:\mathscr{M}_{exch}\to\mathbb{R}$ such that if we let $\mathcal{R}_N$ denote the law of the Gram array formed by the overlaps of $(\sigma^\ell)\sim(G')^{\otimes\infty}$—called the *overlap distribution* corresponding to G'—then

$$\mathcal{F}_{1,M}(\mathcal{R}_N) = \frac{1}{M}\mathbb{E}\log\left\langle \sum_{\epsilon\in\Sigma_M(r_M,u_M)} \exp(\sum_i z_{N,i}(\sigma)\epsilon_i) \right\rangle_{G'} \tag{6.15}$$

$$\mathcal{F}_{2,M}(\mathcal{R}_N) = \frac{1}{M}\mathbb{E}\log\left\langle \exp(\sqrt{M}y_N(\sigma)) \right\rangle_{G'}. \tag{6.16}$$

More precisely, this continuity result is an immediate consequence of the following general continuity theorem for functionals of this form.

Let $(w_\alpha)_{\alpha\in\mathscr{A}}$ be the weights of a (possibly random) probability measure on a countable set $\mathscr{A}$. Denote the measure by $\Gamma(\{\alpha\})=w_\alpha$. Let $K_\mathscr{A}=(K_{\alpha\gamma})_{\alpha,\gamma\in\mathscr{A}}$ be a (possibly infinite) Gram matrix. We think of the matrix $K_\mathscr{A}$ as the collection of allowed values for an overlap and call it an *abstract overlap structure* (defined by w_α).

Fix $S\subseteq\Sigma_M$ and consider the expressions

$$f_{1,M} = \frac{1}{M}\mathbb{E}\log\left[\sum_{\alpha\in\mathscr{A}} w_\alpha \sum_{\sigma\in S}\exp\left(\sum_{i=1}^M Z_i(\alpha)\sigma_i\right)\right] \tag{6.17}$$

$$f_{2,M} = \frac{1}{M}\mathbb{E}\log\left[\sum_{\alpha\in\mathscr{A}} w_\alpha \exp(\sqrt{M}Y(\alpha))\right]. \tag{6.18}$$

Here $Z_i(\alpha)$ are iid copies of a centered Gaussian process $Z(\alpha)$ with covariance

$$\mathrm{Cov}(Z(\alpha),Z(\gamma)) = C_Z(K_{\alpha\gamma}),$$

for some continuous function C_Z. Similarly $Y(\alpha)$ is a centered Gaussian process with covariance

$$\mathrm{Cov}(Y(\alpha),Y(\gamma)) = C_Y(K_{\alpha\gamma}),$$

for some continuous function C_Y.

Let $(\alpha^\ell)_{\ell\geq 1}$ be iid draws from Γ, and consider the random matrix

$$A = \left(K_{\alpha^\ell\alpha^{\ell'}}\right)_{\ell,\ell'}. \tag{6.19}$$

Let us denote its law by $\mathcal{R}_A$. We then have the following result from [73, Lemma 8].

Theorem. *The quantities $f_{i,M}$ depend on $\mathscr{A}$, (w_α), and $R_\mathscr{A}$ only through the law of A. In particular, there are weak-* continuous functionals $\tilde{\mathcal{F}}_{i,M}:\mathscr{M}_{exch}\to\mathbb{R}$ such that $f_{i,M}=\tilde{\mathcal{F}}_{i,M}(\mathcal{R}_A)$.*

24

The representations in (6.15)-(6.16) follow from observing that the righthand sides of these displays are of the form (6.17)-(6.18) with $C_Z(t) = 4\beta^2 t$ and $C_Y(t) = 2\beta^2 t^2$ (see (6.3)-(6.4)) which are continuous in t so that the above theorem applies. Let us now show that we can slightly modify $\mathcal{R}_N$ in such a way that we can compute these limits explicitly.

6.2.2. *Perturbation of overlap distribution.* Before computing this limit, we begin by observing that we may assume, up to a perturbation, that these measures satisfy what is called the *Ghirlanda-Guerra identities* which are defined as follows. Let $R = (R_{\ell\ell'})_{\ell,\ell'\geq 1}$ satisfy $R \sim \mathcal{R}$ for some $\mathcal{R} \in \mathcal{M}_{exch}$.

Definition 6.3. Let $R = (R_{\ell,\ell'})$ be a Gram-de Finnetti array with $|R_{ij}| \leq 1$. We say that the law of $(R_{\ell,\ell'})$ satisfies the *Ghirlanda-Guerra identitites* if for every $n \geq 1$, $f \in L^\infty(\mathbb{R}^{n^2})$ and $p \geq 1$,

$$\mathbb{E}[f((R_{ij})_{i,j\in[n]}) \cdot R^p_{1,n+1}] = \frac{1}{n}\Big[\mathbb{E}f((R_{ij})_{i,j\in[n]}) \cdot \mathbb{E}R^p_{12} + \sum_{\ell=2}^N \mathbb{E}[f((R_{ij})_{i,j\in[n]}) \cdot R^p_{1\ell}]\Big].$$

We begin by observing that we may perturb the Hamiltonian H so that it satisfies the Ghirlanda-Guerra identities.

Lemma 6.4. *Let*

$$h_N(\sigma) = \sum \frac{x_p}{2^p} g_p(\sigma)$$

where $(x_p) \in [1,2]^{\mathbb{N}}$*, and*

$$g_p(\sigma) = \frac{1}{N^{\frac{p}{2}}} \sum g_{i_1\ldots i_p} \sigma_{i_1} \cdots \sigma_{i_p}$$

where $(g_{i_1\ldots,i_p})$ *are i.i.d. and independent of W. Let* $s_N \to \infty$ *with* $N^{1/4} \ll s_N \ll N^{1/2}$*. There is a sequence of choices of parameters* (x_p^N) *such that*

$$\lim |\frac{1}{N}\mathbb{E}\log \sum \exp(\beta H_N(\sigma) + s_N\beta h_N(\sigma)) - \frac{1}{N}\mathbb{E}\log \sum \exp(\beta H(\sigma))| = 0 \qquad (6.20)$$

and such that if R denotes the limiting law of the Gram array formed by overlaps of i.i.d. samples from the Gibbs measure $\pi_{pert}(\{\sigma\}) \propto \exp(-\beta H_N(\sigma) - \beta h(\sigma))$*, then R satisfies the Ghirlanda-Guerra identities.*

Proof Sketch . Results of this type are standard in the spin glass literature. For a textbook presentation see [69, Chap. 3]. We only sketch the key points here and how they differ from [69].

For any sequence of choices of parameters, by the condition on s_N, (6.20) holds by an application of Jensen's inequality. One can then show that if we choose the parameters (x_p) to be drawn i.i.d. from the uniform measure on $[1,2]$, then

$$\mathbb{E}_x \mathbb{E} \langle |(g_p - \mathbb{E}\langle g_p\rangle)|\rangle \to 0.$$

Consequently for any choice of $n \geq 1$ and $f \in L^\infty(\mathbb{R}^n)$, we obtain, conditionally on x,

$$\mathbb{E} \langle f(R^n)g_p\rangle = \mathbb{E} \langle f(R^n)\rangle \cdot \mathbb{E} \langle g_p\rangle + o(1).$$

Integrating by parts (see [69, Exer. 1.1]), we obtain

$$\mathbb{E} \langle f(R^n)(R^p_{11} + R^p_{12} + \cdots + R^p_{1n} - nR^p_{1n+1}\rangle = \mathbb{E} \langle f\rangle \mathbb{E} \langle R^p_{11} - R^p_{12}\rangle,$$

which yields, upon re-arrangement,

$$\frac{1}{n}(\mathbb{E} \langle fR^p_{11}\rangle - \mathbb{E} \langle f\rangle \mathbb{E} \langle R^p_{11}\rangle) + \frac{1}{n}\left(\mathbb{E} \langle f\rangle \mathbb{E} \langle R^p_{12}\rangle + \sum_{i=2}^n \mathbb{E} \langle f \cdot R^p_{1i}\rangle\right) = \mathbb{E} \langle fR^p_{1n+1}\rangle + o(1).$$

The main issue in settings where the Gibbs measure is not on the discrete hypercube $\{-1,1\}^N$ are the terms with the self-overlap, R_{11}. This, however, is not an issue in our setting as R_{11} is deterministic, so that the first term in the above vanishes identically for all N.

The remaining argument is then unchanged from [69, Sec 3.2]. In particular, by this vanishing, we obtain that the error between the left and right hand sides of the Ghirlanda-Guerra identity for any p by this argument. The existence of a suitable sequence then follows by the probabilistic method as in [69, Lemma 3.3]. $\qquad\square$

As a consequence, it suffices to evaluate the limits of $\mathcal{F}_{i,M}$ along the sequence of overlap distributions, $\mathcal{R}'_N$, corresponding to the perturbed Gibbs distribution $G'_{pert,N}(\{\sigma\}) \propto \exp(\beta H_N(\sigma) + \beta h_N(\sigma))$, where h_N is obtained from the preceding lemma (and we still restrict to $\Sigma_N(\rho_N, q_N)$). As $\mathscr{R}$ is compact, this sequence has weak-* limit points. we see that any weak-* limit point of this sequence satisfies the Ghirlanda-Guerra identities. Let us now briefly recall the structure of the subspace of $\mathscr{M}_{exch}$ which satisfy these identities.

To this end, we first recall the construction of overlap distributions corresponding to Ruelle probability cascades. Fix $r \geq 1$ and a pair of sequences

$$0 = Q_0 < \ldots < Q_r = \rho \qquad 0 = \mu_{-1} < \mu_0 < \ldots < \mu_r = 1.$$

Let $(v_\alpha)_{\alpha \in \mathbb{N}^r}$ denote the weights of a Ruelle probability cascade corresponding to the parameters $(\mu_\ell)_{\ell=-1}^r$. Corresponding to this sequence, we may define the (random) probability measure on the ball of ℓ_2 of radius ρ given by

$$\pi = \sum v_\alpha \delta_{s_\alpha} \qquad s_\alpha = \sum_{\gamma \in p(\alpha)} \sqrt{Q_{|\gamma|} - Q_{|\gamma|-1}} e_\gamma, \tag{6.21}$$

where $(e_\alpha)_{\alpha \in \mathcal{A}_r \setminus \{\emptyset\}}$ is an orthonormal basis of ℓ_2. Let $(\sigma^\ell) \sim \pi^{\otimes \infty}$ and let $\mathcal{R}$ denote the law of their corresponding Gram matrix. If we let $\mu \in \mathscr{M}_1([0, \rho])$ be defined by $\mu(\{Q_\ell\}) = \mu_\ell - \mu_{\ell-1}$, then by Lemma A.1, μ is the law of the first off-diagonal entry of the gram array. In particular, μ is a sufficient statistic for the family of such $\mathcal{R}$ and we will denote this law by $\mathcal{R}(\mu)$.

Let $\mathscr{M}_f \subseteq \mathscr{M}_1([0, \rho])$ be those measures of finite support. As a consequence of Panchenko's Ultrametricity theorem [68] and the Baffioni-Rossati theorem [10] and [82, Theorem 15.3.6], the space of Gram-de Finnetti arrays satisfying the Ghirlanda–Guerra identities is given by the closure of the set $\mathscr{G} = \{\mathcal{R}(\mu) : \mu \in \mathscr{M}_f\}$ in the weak-* topology. We denote a law of this type by $\mathcal{R}(\mu)$. In particular, if $(\mathcal{R}_n) \subset \mathscr{G}$ is a sequence of laws in this closure, then the law of the off-diagonals of $\mathcal{R}_n$ converge to the law of the off diagonals of $\mathcal{R}$ if and only if $\mu_n(\cdot) = \mathcal{R}_n(R_{12} \in \cdot)$ have $\mu_n \to \mu$. For a precise statement of these two results, see [69, Theorem 2.13, Theorem 2.17]. (As we will always be using this result in the case that the diagonal terms of the array are a deterministic constant, c_n, with $c_n \to \rho$ constant, this convergence is effectively for the entire array for our purposes.)

Thus it suffices to compute these limiting functionals on Ruelle cascades and then, take their limits (provided they exist). To carry out this program, it will be important to note that, restricted to the space of Ruelle probability cascades, these functionals are in fact Lipschitz.

6.2.3. *Lipschitz Continuity.* Let us restrict our attention to functionals of the form $\mathcal{F}_{i,M}$ on the space of Ruelle probability cascades. If we let $\mathcal{R}$ denote the law of the overlap array corresponding to a Ruelle cascade as above, then by construction

$$\mathcal{F}_{1,M}(\mathcal{R}) = \frac{1}{M} \mathbb{E} \log \sum v_\alpha \sum_{\sigma \in \Sigma_M(\rho, q)} \exp(\sum_i Z_i(\alpha)\sigma_j)$$

$$\mathcal{F}_{2,M}(\mathcal{R}) = \frac{1}{M} \mathbb{E} \log \sum v_\alpha \exp(\sqrt{M} Y(\alpha)).$$

Let us now understand these functionals in more detail.

26

To this end, let
$$R_{\mathbb{N}^r} = (s_\alpha \cdot s_{\alpha'})_{\alpha, \alpha' \in \mathbb{N}^r}$$
denote the collection of overlaps defined by s_α from (6.21), we may define the functional $f_{1,M}(\mu; S)$ to be the functional as in (6.17) with weights given by the v_α and abstract overlap structure $R_{\mathbb{N}^r}$ for some choice $S \subset \Sigma_M(\rho)$. Define $f_{2,M}(\mu; S)$ similarly with (6.18). In this case, if we let $\mathcal{R}$ denote such an overlap distribution and let μ denote the law of the first off diagonal, we have that
$$f_{i,M}(\mu; S) = \mathcal{F}_{i,M}(\mathcal{R}) \quad i = 1, 2.$$
In particular, this functional depends on $\mathcal{R}$ only through μ. Let us study the regularity of the map $\mu \mapsto f_{1,M}(\mu; S)$.

To this end, equip $\mathscr{M}_1([0, \rho])$ with the the Kantorovich metric,
$$d(\mu, \nu) = \int_0^1 |\mu^{-1}(s) - \nu^{-1}(s)| ds = \int_0^\rho |\mu([0, s]) - \nu([0, s])|,$$
where for a probability measure μ, we let μ^{-1} denote its quantile function. Recall that d metrizes the weak-* topology on $\mathscr{M}_1([0, \rho])$. We then have the following.

Lemma 6.5. *The following holds for any $M \geq 1$.*

(1) *For any $S \subset \Sigma_M(\rho)$, the map $\mu \mapsto f_{1,M}(\mu; S)$ is Lipschitz (uniformly in M, ρ).*

(2) *The map $\mu \mapsto f_{2,M}(\mu)$ is Lipschitz (uniformly in M, ρ).*

In particular, these functionals are well-defined and Lipschitz (uniformly in M, ρ) on all of $\mathscr{M}_1([0, \rho])$.

Proof. We begin with the first claim. To this end, fix two measures $\mu, \tilde{\mu}$ of finite support. Consider their quantile functions μ^{-1} and $\tilde{\mu}^{-1}$. Observe that we may view these as monotone paths from 0 to ρ indexed by $[0, 1]$. In particular, we may parameterize these paths (and thus the measures) as follows. Since these paths have finitely many jumps, there is some $r \geq 1$ and some sequence
$$0 = x_{-1} < x_0 \leqslant ... < x_r = 1$$
such the jumps in either path are given by an increasing subset of the times (x_ℓ). Furthermore, we may parameterize the supports of μ and $\tilde{\mu}$ by two finite sequences
$$0 = Q_0 \leq ... \leq Q_r = \rho$$
$$0 = \tilde{Q}_0 \leq \cdots \leq \tilde{Q}_r = \rho,$$
(here we allow repetition) that satisfy $Q_k = \mu^{-1}(x_k)$ and $\tilde{Q}_k = \tilde{\mu}^{-1}(x_k)$.

Let (v_α) denote the Ruelle probability cascade corresponding to (x_r),
$$H_t(\epsilon, \alpha) = \sqrt{t} \sum Z_i(\alpha) \epsilon_i + \sqrt{1-t} \sum \tilde{Z}_i(\alpha) \epsilon_i,$$
and
$$\Phi(t) = \frac{1}{M} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^r} v_\alpha \sum_{\sigma \in S} e^{H_t(\sigma, \alpha)}.$$
Since the measure induced by x and the sequence Q is μ and likewise for $x, \tilde{Q}$, and $\tilde{\mu}$, we have that
$$\Phi(0) = f_{1,M}(\tilde{\mu}; S) \quad \Phi(1) = f_{1,M}(\mu; S).$$
Differentiating in time we find that
$$\Phi'(t) = \frac{1}{M} \mathbb{E} \langle \partial_t H_t(\sigma, \alpha) \rangle_t,$$
where $\langle \cdot \rangle_t$ denotes integration with respect to the measure $G_t(\epsilon, \alpha) \propto v_\alpha \exp(H_t(\epsilon, \alpha))$. Observe that
$$\frac{1}{M} \mathbb{E} \partial_t H_t(\epsilon, \alpha) H_t(\epsilon', \alpha') = 2\beta^2 \left(Q_{|\alpha \wedge \alpha'|} - \tilde{Q}_{|\alpha \wedge \alpha'|} \right) R(\epsilon, \epsilon').$$

27

Since $|R(\epsilon, \epsilon')| \leq \rho$, and $Q_r = \tilde{Q}_r = 1$, we see by integration-by-parts (6.5) that

$$|\Phi'(t)| \leq 2\beta^2 \rho \mathbb{E} \left\langle |Q_{\alpha^1 \wedge \alpha^2} - \tilde{Q}_{\alpha^1 \wedge \alpha^2}| \right\rangle_t,$$

where α^1, α^2 are (the second coordinates of) two independent draws from G_t.

Now since the law of $H_t(\epsilon, \alpha)$ is invariant for $0 \leq t \leq 1$ we may apply the Bolthausen-Sznitman invariance, specifically Theorem A.2, to find that the marginal of G_t on $\mathbb{N}^r$ is equal to v_α for all t. Consequently

$$\mathbb{E} \left\langle |Q_{\alpha^1 \wedge \alpha^2} - \tilde{Q}_{\alpha^1 \wedge \alpha^2}| \right\rangle = \mathbb{E} \sum_{\alpha, \alpha'} v_\alpha v_{\alpha'} |Q_{|\alpha \wedge \alpha'|} - \tilde{Q}_{|\alpha \wedge \alpha'|}| = \sum_{1 \leq k \leq r} |Q_k - \tilde{Q}_k| \mathbb{E} \sum_{|\alpha \wedge \alpha'| = k} v_\alpha v_{\alpha'}$$

$$= \sum_{1 \leq k \leq r} |Q_k - \tilde{Q}_k| (x_k - x_{k-1}) = \int_0^1 |\mu^{-1}(x) - \tilde{\mu}^{-1}(x)| dx$$

where the second to last inequality follows by Lemma A.1. Combining this with the preceding yields

$$|f_{1,M}(\mu; S) - f_{1,M}(\tilde{\mu}, S)| \leq 2\beta^2 \rho d(\mu, \tilde{\mu}),$$

which yields the desired since $\rho \leq 1$.

We now turn to the second claim. This follows by direct calculation. Applying Theorem A.2 and Corollary A.3, yields

$$f_{2,M}(\mu) = 2\beta^2 \int_0^\rho t\mu([0,t]) dt,$$

from which it follows that

$$|f_{2,M}(\mu) - f_{2,M}(\tilde{\mu})| \leq 2\beta^2 \rho \int_0^\rho |\mu([0,t]) - \tilde{\mu}([0,t])| dt = 2\beta^2 \rho d(\mu, \tilde{\mu})$$

which yields the desired as $\rho \leq 1$. $\qquad\qquad\square$

6.2.4. *Computing limits.* Let $\mathcal{R}$ be a limiting overlap distribution corresponding to G'_{pert}, and μ denote the corresponding law of R_{12}. Let μ_r denote a sequence of discretizations of μ with finitely many atoms with $\mu_r \to \mu$ weak-* and $\mu_r(\{\rho\}) > 0$. Since ρ is always charged for this sequence, the overlap distirbution $\mathcal{R}(\mu_r)$ has diagonal equal to ρ. Thus, as explained at the end of Section 6.2.2, the laws $\mathcal{R}(\mu_r) \to \mathcal{R}$ weak-* since μ determines the off-diagonal of $\mathcal{R}$ and the diagonals are constant an equal to ρ.

We first observe that $\mathcal{F}_{2,M}$ may be handled as previously: we have that

$$\lim_{N \to \infty} \mathcal{F}_{2,M}(\mathcal{R}'_N) = \mathcal{F}_{2,M}(\mathcal{R}) = \lim_{r \to \infty} \frac{1}{M} \mathbb{E} \log \sum v_\alpha e^{\sqrt{M} Y(\alpha)}$$

$$= \lim_{r \to \infty} 2\beta^2 \int_0^\rho s\mu_r([0,s]) ds = 2\beta^2 \int_0^\rho s\mu([0,s]) ds,$$

where the first two equalities follows by continuity of $\mathcal{F}_{2,M}$, the third by Corollary A.3, and the final follows by weak-* continuity. It remains to consider the first term, $\mathcal{F}_{1,M}$.

Let

$$A_M = \lim_{N \to \infty} \frac{1}{M} \mathbb{E} \log \left\langle \sum_{\epsilon \in \Sigma_M(r_M, u_M)} \exp(\sum_i z_{N,i}(\sigma)\epsilon_i) \right\rangle_{G'_{pert}}.$$

Then by Lemma 6.5, it follows that

$$A_M \geq \frac{1}{M} \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^r} v_\alpha \sum_{\substack{\epsilon \in \Sigma_M(r_M, u_M)}} \exp\left(\sum_i Z_i(\alpha)\epsilon_i\right) - Kd(\mu_r, \mu),$$

where (v_α) are the weights of the Ruelle probability cascade with parameters (μ_r), and K does not depend on M. Recall from the proof of Theorem 6.1 that we were able to produce an upper bound for this term where Λ_i play the role of Lagrange multipliers for the constraints defining $\Sigma_M(r_M, u_M)$. One can also produce the following matching lower bound by minimizing over the choice of Lagrange multipliers whose proof is deferred to the next section.

Theorem 6.6. *Let $\mu \in \mathscr{M}_1([0, \rho])$ have finite support. Let $(\rho_M, q_M) \to (\rho, q)$ admissible. We have that*

$$\lim_{M \to \infty} f_{1,M}(\mu, \Sigma_M(\rho_M, q_M)) = \inf_{\Lambda_1, \Lambda_2} \{\rho \varphi_\mu^1(0,0) + (1-\rho)\varphi_\mu^2(0,0) - \Lambda_1 q - \Lambda_2(\rho - q)\}.$$

In particular, we have that, for each $r \geq 1$,

$$\lim_{M \to \infty} \mathcal{F}_{1,M}(\mu_r) = \lim_{M \to \infty} f_1(\mu_r, \Sigma_M(\rho_M, q_M)) = \inf_{\Lambda_1, \Lambda_2} \{\rho \varphi_{\mu_r}^1(0,0) + (1-\rho)\varphi_{\mu_r}^2(0,0) - \Lambda_1 q - \Lambda_2(\rho - q)\}.$$

As the functionals $\mu \mapsto f_1(\mu, \Sigma_M(\rho_M, q_M))$ are uniformly Lipschitz by Lemma 6.5, so is their limit. In particular we see that

$$\lim_{M \to \infty} A_M \geq \inf_{\Lambda_1, \Lambda_2} \{\rho \varphi_\mu^1(0,0) + (1-\rho)\varphi_\mu^2(0,0) - \Lambda_1 q - \Lambda_2(\rho - q)\} - 2Kd(\mu_r, \mu),$$

for some K. If we then send $r \to \infty$, we see that we have

$$\varliminf_{N \to \infty} \tilde{F}_N \geq \inf_{\Lambda_1, \Lambda_2} P_\beta(\mu, \Lambda_1, \Lambda_2). \tag{6.22}$$

This yields the desired lower bound.

Proof of Proposition 4.1. Upon combining (6.2) with (6.22) we obtain (6.1). Recalling the equality $P_\beta = \beta \mathcal{P}_\beta$ and dividing by β yields the result, except with an infimum. We then recall Lemma 5.1 to conclude that this infimum is actually achieved. $\square$

6.3. **Removing constraints via Lagrange multipliers.** We now turn to the proof of Theorem 6.6. The lemmas mentioned in this proof will be proved in the following section. Let $\mathcal{S} = \{(x, y) \in [0,1]^2 : y \leq x\}$.

Proof of Theorem 6.6. Observe that as in (6.7) we were able to obtain the corresponding upper bound. The question is then to show that the infimum is achieved. We follow a strategy similar to [73, Sec. 3]. In the following, it will be useful to recall that the quantities we wish to compute are invariant under permutations of the entries of the vector v.

For a point $\mathbf{x} = (\rho, q) \in \mathcal{S}$ and $\epsilon > 0$, consider the set

$$\Sigma_M^\epsilon(\mathbf{x}) = \left\{\sigma \in \Sigma_M : \frac{1}{M} \sum_{i=1}^M \sigma_i \in [\rho - \epsilon, \rho + \epsilon], \frac{1}{M} \sum_{i=1}^{\lceil M\rho \rceil} \sigma_i \in [q - \epsilon, q + \epsilon]\right\}.$$

We begin by observing that the error caused by this ϵ dilation is negligible. Let $\mathcal{S}_M \subseteq \mathcal{S}$ be those $\mathbf{x} \in \mathcal{S}$ such that $\Sigma_M(\mathbf{x}) = \Sigma_M(x_1, x_2)$ is non-empty.

Lemma 6.7. *There is an $C > 0$ such that for $\epsilon > \delta > 0$ sufficiently small and any $M \geq 1$ we have*

$$\sup_{x \in \mathcal{S}_M} |f_{1,M}(\mu, \Sigma_M^\epsilon(\mathbf{x})) - f_1(\mu, \Sigma_M(\mathbf{x}))| \leq C\sqrt{\epsilon}$$

$$\sup_{x \in \mathcal{S}} |f_{1,M}(\mu, \Sigma_M^\epsilon(\mathbf{x})) - f_1(\mu, \Sigma_M^\delta(\mathbf{x}))| \leq C\sqrt{\epsilon - \delta}$$

Next, we begin by observing that on such sets, the limit of these functionals exist and is concave.

29

Lemma 6.8. *For each $\epsilon > 0$, the limit*

$$f_1(\mu, \epsilon, \mathbf{z}) = \lim_{M \to \infty} f_{1,M}(\mu, \Sigma_M^\epsilon(\mathbf{z}))$$

exists and is concave.

Combining these claims we find that

$$f_1(\mu; \rho, q) = \lim_{M \to \infty} f_{1,M}(\mu, \Sigma_M(\rho_M, q_M)) = \lim_{\epsilon \to 0} f_1(\mu; \rho, q, \epsilon). \tag{6.23}$$

exists and is concave in (ρ, q). Furthermore, applying the first claim again, we find that is 1/2-Hölder. In particular, it is uniformly continuous on $\mathcal{S}$.

It remains for us to relate $f_1(\mu; \rho, q)$ to Parisi-type functionals. To this end, we observe the following. For any $\boldsymbol{\Lambda} = (\Lambda_1, \Lambda_2) \in \mathbb{R}^2$, let

$$F(\boldsymbol{\Lambda}) = \rho \varphi_\mu^1(0, 0; \Lambda_1) + (1 - \rho) \varphi_\mu^2(0, 0; \Lambda_2) - \Lambda_1 q - \Lambda_2(\rho - q)$$

(here we have made the dependence of φ^i on Λ_i explicit for clarity). We then have the following.

Lemma 6.9. *For any (Λ_1, Λ_2), we have that*

$$F(\boldsymbol{\Lambda}) = \max_{\mathbf{x} \in \mathcal{S}} \{f_1(\mu, \mathbf{x}) + \Lambda_1 q + \Lambda_2(\rho - q)\} \tag{6.24}$$

To complete the result, we recall that $f_1(\mu, \mathbf{x})$ is concave and continuous in $\mathbf{x}$. Thus, we may take its Legendre transform and apply (6.24) to find that

$$f_1(\mu, \mathbf{x}) = \inf_{\boldsymbol{\Lambda}} \{F(\boldsymbol{\Lambda}) - \Lambda_1 q - \Lambda_2(\rho - q)\}$$

as desired. $\qquad\qquad\square$

6.4. **Proof of lemmas used Theorem 6.6.**

Proof of Lemma 6.7 . We prove the first bound. The second follows by the same argument upon noting that $\Sigma_M^\delta(\rho, q) \subseteq \Sigma_M^\epsilon(\rho, q)$.

Let $\pi : \Sigma_M^\epsilon(\mathbf{x}) \to \Sigma_M(\mathbf{x})$ be the map from a point in $\Sigma_N^\epsilon(\mathbf{x})$ to the closest point in $\Sigma_M(\mathbf{x})$ with respect to the Hamming distance. Note that in the ℓ_2 distance we have $d(\mathbf{x}, \pi(\mathbf{x})) \leq C\epsilon N$. (Throughout this proof C will be a constant that may change from line to line.) Let $\mathbf{Z}(\alpha) = (Z_i(\alpha))$, and let $\tilde{\mathbf{Z}}(\alpha)$ be an independent copy of $\mathbf{Z}$. Let $\mathbf{Z}_t(\alpha, \sigma) = \sqrt{t}\mathbf{Z}(\alpha) \cdot \epsilon + \sqrt{1 - t}\tilde{\mathbf{Z}}(\alpha) \cdot \pi(\sigma)$ and consider

$$\phi(t) = \frac{1}{M} \mathbb{E} \log \sum v_\alpha \sum_{\sigma \in \Sigma_M^\epsilon(\mathbf{x})} e^{\mathbf{Z_t}(\alpha, \sigma)},$$

Since

$$\frac{1}{M} \mathbb{E} \partial_t \mathbf{Z}_t(\alpha, \sigma) \mathbf{Z_t}(\alpha', \sigma') = 2\beta^2 \{Q_{|\alpha \wedge \alpha'|}[R(\sigma, \sigma') - R(\pi(\sigma), \pi(\sigma'))]\}$$

so that $|\phi'| \leq C\epsilon$.

$$\phi(0) = \frac{1}{M} \mathbb{E} \log \sum v_\alpha \sum_{\sigma \in \Sigma_M(\mathbf{x})} |\pi^{-1}(\{\sigma\})| e^{\mathbf{Z}(\alpha) \cdot \sigma}.$$

Now by a standard counting argument,

$$\frac{1}{M} \log \max |\pi^{-1}(\{\sigma\})| \leq \log 2 - J(1 - C\epsilon) \leq C\sqrt{\epsilon}$$

where $J(x) = \frac{1}{2}(1+x)\log(1+x) + \frac{1}{2}(1-x)\log(1-x)$, where the last inequality holds for ϵ sufficiently small. Thus

$$f_{1,M}(\mu, \Sigma_M(\mathbf{x})) \leq \phi(0) \leq f_{1.M}(\mu, \Sigma_M(\mathbf{x})) + C\sqrt{\epsilon}.$$

On the other hand, by the preceding, $|\phi(1) - \phi(0)| \leq C\epsilon$. Combining these inequalities and recalling the definition of $\phi(1)$ then yields the claim. $\qquad\square$

Proof of Lemma 6.8. To see this, note that for any $M_1, M_2 \geq 1$, let $M = M_1 + M_2$ and $\theta = \frac{M_1}{M}$. For two points $\mathbf{x}, \mathbf{y} \in \mathcal{S}$ let $\mathbf{z} = \theta\mathbf{x} + (1 - \theta)\mathbf{y}$. Let $i : \Sigma_{M_1}^\epsilon(\mathbf{x}) \times \Sigma_{M_2}^\epsilon(\mathbf{y}) \hookrightarrow \Sigma_M$ be the map that takes a point (u, v) to the point $\pi(u, v)$ whose first $\lceil M_1 x_1 \rceil$ coordinates are those of u and next $\lceil M_2 y_1 \rceil$ points are those of v, and whose remaining coordinates are given by those of u and v in that order. Evidently i is an injection. In fact its image is contained in $\Sigma_M^{\epsilon + 1/M}(\mathbf{z})$. Therefore, applying the second inequality from Lemma 6.7, we see that

$$f_{1,M}(\mu, \Sigma_M^\epsilon(\mathbf{z})) \geq \lambda f_{1,M_1}(\mu, \Sigma_M^\epsilon(\mathbf{x})) + (1 - \lambda)f_{1,M_2}(\mu, \Sigma_M^\epsilon(\mathbf{y})) - \sqrt{1/M}.$$

Taking $\mathbf{x} = \mathbf{y}$, we see that the sequence $a_M = M f_{1,M}(\mu, \Sigma_M^\epsilon(\mathbf{z}))$ satisfies

$$a_{m_1 + m_2} + \phi(m_1 + m_2) \geq a_{m_1} + a_{m_2},$$

for $\phi(t) = \sqrt{t}$. As ϕ is increasing and satisfies $\int_1^\infty \phi(t)t^{-2}dt < \infty$, we may apply the de Brujin-Erdos superadditivity lemma (see below) to find that the desired limit exists. Furthermore, by taking the limits in the case $\mathbf{x} \neq \mathbf{y}$, we see that the limit is concave. $\qquad\square$

We have used here the following classical result of de Brujin-Erdös [35, Theorem 23] (or see [79, Theorem 1.9.2]).

Theorem. *Let (a_n) be a real sequence and ϕ a non-decreasing function with $\int_1^\infty \phi(t)t^{-2}dt < \infty$. If (a_n) satisfies the superadditivity criterion*

$$a_{n+m} + \phi(n + m) \geq a_n + a_m \qquad \frac{1}{2}n \leq m \leq 2n,$$

then $\lim_n a_n/n$ converges to $\sup_n a_n/n$.

Proof of Lemma 6.9. This will follow by calculations involving the Ruelle Probability Cascades along with a standard covering argument. Consider the functional

$$F(\mathbf{\Lambda}, S) = \frac{1}{M}\mathbb{E}\log\sum_\alpha v_\alpha \sum_{\sigma \in S}\exp\left(\sum_i Z_i(\alpha)\sigma_i + \Lambda_1 M_1 R_{11}^{(1)} + \Lambda_2 M_2 R_{11}^{(2)}\right).$$

Then

$$F(\mathbf{\Lambda}, \Sigma_N) = F(\mathbf{\Lambda}) \qquad \text{and} \qquad F(\Lambda, \Sigma_M(\rho, q)) = f_{1,M}(\mu, \Sigma_N(\rho, q)) + \Lambda_1 q + \Lambda_2(\rho - q),$$

where the first equality again follows by (6.8).

Since f_1 from (6.23) is uniformly continuous on $\mathcal{S}$, thus suffices to show that

$$F(\mathbf{\Lambda}) - \max_{\mathbf{x} \in \mathcal{S}_M} F(\mathbf{\Lambda}, \Sigma_M(\mathbf{x})) \to 0.$$

To this end, we begin by present a formula for $F(\mathbf{\Lambda}; S)$. Its proof follows from an abstract version of Theorem A.2. Let $(z_p)_{p \leq r}$ denote a collection of independent centered Gaussian random variables with

$$\mathbb{E}z_p^2 = 4\beta^2(Q_p - Q_{p-1}),$$

and let $(z_{p,i})_{p \leq r, i \in [N]}$ be i.i.d. copies of this process. Consider

$$X_r(\mathbf{\Lambda}, S) = \log\sum_{\sigma \in S}\exp\left(\sum z_{p,i}\sigma_i + \Lambda N_1 R_{11}^{(1)} + \Lambda N_2 R_{11}^{(2)}\right).$$

Define now the sequence of random variables $(X_\ell)_{\ell=0}^r$ by

$$X_\ell = \frac{1}{\mu_\ell}\mathbb{E}_p\exp(\mu_\ell X_{\ell+1}),$$

31

where $\mathbb{E}_p$ is expectation in z_p and $\mu_\ell = \mu([0, Q_\ell])$. As a consequence of [69, Theorem 2.9], we have that

$$F(\mathbf{\Lambda}, S) = \frac{1}{M} X_0.$$

Let us now prove the desired limit. By construction,

$$\exp(X_r(\Lambda, \Sigma_N)) = \sum_{\mathbf{x} \in \mathcal{S}_N} \exp(X_r(\Lambda, \Sigma_M(\mathbf{x})).$$

Since $\mu_\ell \leq 1$ and $\frac{\mu_\ell}{\mu_{\ell+1}} \leq 1$, and $(\sum a_i)^s \leq \sum a_i^s$ for $s \leq 1$ and $a_i \geq 0$, we have by an inductive argument that

$$\exp(\mu_p X_p) = \mathbb{E}_p \exp(\mu_p X_{p+1}) \leq \sum_{\mathbf{x} \in \mathcal{S}_M} \exp(\mu_p X_p).$$

Applying this in the case $p = 0$, we see that

$$F(\mathbf{\Lambda}) = \frac{1}{M} X_0(\mathbf{\Lambda}, \Sigma_M) \leq \frac{1}{M\mu_0} \log \sum_{\mathbf{x} \in \mathcal{S}_M} \exp \mu_0 X_0(\mathbf{\Lambda}, \Sigma_M(\mathbf{x}))$$

$$\leq \frac{1}{M\mu_0} \log |\mathcal{S}_M| + \max_{\mathbf{x} \in \mathcal{S}_M} F(\mathbf{\Lambda}, \Sigma_M(\mathbf{x}))$$

On the other hand by (6.2) we have that

$$F(\mathbf{\Lambda}) \geq \max_{\mathbf{x} \in \mathcal{S}_M} F(\mathbf{\Lambda}, \Sigma_M(\mathbf{x})).$$

Since $\mathrm{card}(\mathcal{S}_M)$ is of at most polynomial growth in M, the result follows by the squeeze theorem. $\quad\square$

7. Statistical feasibility of Estimation: proof of Theorem 1.2

We prove Theorem 1.2 in two parts. Recall that the second part was already proved in Section 2. It remains to show the first part of Theorem 1.2, regarding the impossibility of approximate support recovery. We follow a rate distortion argument from [84, Theorem 2]. In the following, for $a, b \in (0,1)$, let

$$h(a) = -a \log a - (1-a) \log(1-a)$$

Proof of part 1 of Theorem 1.2. We will establish the following: for any $0 < c < 1$, if $\lambda < \frac{c^{3/2}}{2} \sqrt{\frac{1}{\rho} \log(\frac{1}{\rho})}$ then

$$\overline{\lim_N} \frac{1}{N} \sup_{\hat{v} \in \Sigma_N(\rho)} \inf_{v \in \Sigma_N(\rho)} \mathbb{E}(\hat{v}, v) \leq c.$$

Observe that $\inf_{v \in \Sigma_N(\rho)} \mathbb{E}(\hat{v}, v) \leq \mathbb{E}_{v \sim U(\Sigma_N(\rho))}(\hat{v}, v)$, where $\mathbb{E}_{v \sim U(\Sigma_N(\rho))}[\cdot]$ denotes the expectation in A and v, where now $v \sim U(\Sigma_N(\rho))$ instead of being fixed. It thus suffices to upper bound

$$\overline{\lim_N} \frac{1}{N} \sup_{\hat{v} \in \Sigma_N(\rho)} \mathbb{E}_{v \sim U(\Sigma_N(\rho))}(\hat{v}, v) \leq c. \tag{7.1}$$

We do so by an information theoretic argument. For two random variables, X, Y, let $H(X)$ denote the Shannon entropy, $H(X|Y)$ denote the conditional entropy, and let

$$I(X;Y) = H(X) - H(X|Y),$$

denote their mutual information.

To begin, observe that v and $\hat{v}$ are conditionally independent given A, and thus by the data processing inequality [34], $I(v; \hat{v}) \leq I(v; A)$. Similarly, note that A and v are conditionally independent given $\frac{\lambda}{N} v v^{\mathrm{T}}$, and in fact v is uniquely determined given $\frac{\lambda}{N} v v^{\mathrm{T}}$, and thus $I(v; A) \leq I(\frac{\lambda}{N} v v^{\mathrm{T}}; A)$. Combining, we have that $I(v; \hat{v}) \leq I(\frac{\lambda}{N} v v^{\mathrm{T}}; A)$.

Now, we note that given $\{\frac{\lambda}{N}v_iv_j : i < j\}$, $\{A_{ij} : i < j\}$ is a product distribution, and thus

$$I\Big(\frac{\lambda}{N}vv^{\mathrm{T}}; A\Big) \leq \sum_{i<j} I\Big(\frac{\lambda}{N}v_iv_j; A_{ij}\Big) \leq \frac{N^2}{4}\log\Big(1 + \frac{\lambda^2\rho^2}{N}\Big) \leq \frac{N\lambda^2\rho^2}{4}. \qquad (7.2)$$

The second inequality above follows using the capacity of a Gaussian additive channel [34].

On the other hand, we have that

$$I(v;\hat{v}) = H(v) - H(v|\hat{v}) = \log\binom{N}{N\rho} - H(v|\hat{v}).$$

Suppose now that $\mathbb{E}_{v\sim U(\Sigma_N(\rho))}[(v,\hat{v})] \geq cN\rho$ for some $c > 0$. Set $f(\hat{v}) = \mathbb{E}_{v\sim U(\Sigma_N(\rho))}[(v,\hat{v})|\hat{v}]$. By the Paley-Zygmund inequality,

$$\mathbb{P}_{v\sim U(\Sigma_N(\rho))}\Big[f(\hat{v}) \geq \frac{cN\rho}{2}\Big] \geq \frac{1}{4}\frac{(\mathbb{E}_{v\sim\Sigma_N(\rho)}[f(\hat{v})])^2}{\mathbb{E}_{v\sim\Sigma_N(\rho)}[(f(\hat{v}))^2]} \geq \frac{c^2}{4}$$

which is bounded away from zero, uniformly in N and ρ. The inequality above follows from the observations that $0 \leq f(\hat{v}) \leq N\rho$, and $\mathbb{E}_{v\sim U(\Sigma_N(\rho))}[f(\hat{v})] \geq cN\rho$. We set $\mathscr{E} = \{f(\hat{v}) \geq \frac{cN\rho}{2}\}$ to obtain

$$H(v|\hat{v}) = \mathbb{E}_{v\sim U(\Sigma_N(\rho))}\Big[\mathbf{1}_{\mathscr{E}}g(\hat{v})\Big] + \mathbb{E}_{v\sim U(\Sigma_N(\rho))}\Big[\mathbf{1}_{\mathscr{E}^c}g(\hat{v})\Big],$$

where we use $g(x)$ to denote the entropy of the conditional distribution of v given $\hat{v} = x$. On the event $\mathscr{E}^c$, we use the trivial bound $g(\hat{v}) \leq \log\binom{N}{N\rho}$. On the event $\mathscr{E}$, we upper bound the conditional entropy as follows.

Fix $x \in \mathscr{E}$. Using the sub-additivity of conditional entropy, we have,

$$g(x) \leq \sum_{i:x_i=1} H(v_i|\hat{v}=x) + \sum_{i:x_i=0} H(v_i|\hat{v}=x) \leq N\rho\log 2 + N(1-\rho)\cdot\frac{1}{N(1-\rho)}\sum_{i:x_i=0} H(v_i|\hat{v}=x)$$

$$\leq N\rho\log 2 + N(1-\rho)h\Big(\frac{1}{N(1-\rho)}\sum_{i:x_i=0}\mathbb{E}_{v\sim U(\Sigma_N(\rho))}[v_i|\hat{v}=x]\Big),$$

where the last inequality follows by the concavity of the Shannon entropy h. Observe that on the event $\mathscr{E}$, $\mathbb{E}_{v\sim U(\Sigma_N(\rho))}[\sum_{i:x_i=0} v_i|\hat{v}=x] \leq N\rho(1-c/2)$. For ρ sufficiently small $\rho(1-c/2)/(1-\rho) < 1/2$, and thus maximizing the binary entropy, we obtain the improved upper bound

$$g(x) \leq N\rho\log 2 + N(1-\rho)h\Big(\frac{\rho(1-c/2)}{1-\rho}\Big).$$

Combining, we have,

$$H(v|\hat{v}) \leq \mathbb{P}_{v\sim U(\Sigma_N(\rho))}(\mathscr{E})\Big[N\rho\log 2 + N(1-\rho)h\Big(\frac{\rho(1-c/2)}{1-\rho}\Big)\Big] + \mathbb{P}_{v\sim U(\Sigma_N(\rho))}(\mathscr{E}^c)\log\binom{N}{N\rho}.$$

Therefore,

$$I(v;\hat{v}) \geq \log\binom{N}{N\rho} - \Big[\mathbb{P}_{v\sim U(\Sigma_N(\rho))}(\mathscr{E})\Big(N\rho\log 2 + N(1-\rho)h\Big(\frac{\rho(1-c/2)}{1-\rho}\Big)\Big)$$

$$+ \mathbb{P}_{v\sim U(\Sigma_N(\rho))}(\mathscr{E}^c)\log\binom{N}{N\rho}\Big]$$

$$\geq \mathbb{P}_{v\sim U(\Sigma_N(\rho))}(\mathscr{E})\log\binom{N}{N\rho} - N\mathbb{P}_{v\sim U(\Sigma_N(\rho))}(\mathscr{E})\Big[\rho\Big(1-\frac{c}{2}\Big)\log\frac{1}{\rho} + o\Big(\rho\log\frac{1}{\rho}\Big)\Big] \geq \frac{c^3}{16}N\rho\log\frac{1}{\rho}$$

33

for N sufficiently large, and ρ sufficiently small. Combining this with (7.2), we obtain

$$\frac{c^3}{16}N\rho\log\frac{1}{\rho} \le \frac{N\lambda^2\rho^2}{4} \implies \lambda^2 \ge \frac{c^3}{4}\cdot\frac{1}{\rho}\log\frac{1}{\rho}.$$

This contradicts our choice of λ. Thus the upper bound (7.1) holds whenever $\lambda \le c'\sqrt{\frac{1}{\rho}\log\frac{1}{\rho}}$ for some $c' < c^{3/2}/2$. This completes the proof. $\square$

8. A SIMPLE ROUNDING SCHEME WHEN $\lambda > \frac{1}{\rho}$

In this section, we establish that if the SNR is significantly large, it is algorithmically easy to approximately recover the support of the hidden principal sub-matrix. Specifically, we establish that for $\lambda > 1/\rho$, there exists a simple spectral algorithm which approximately recovers the hidden support. To this end, we start with a simple lemma, which will be critical in the subsequent analysis.

Lemma 8.1. *Let (X, Y) be jointly distributed random variables, with $X \in \{0, 1\}$. Further, assume $\mathbb{P}[X = 1] = \rho$, $\mathbb{E}[Y^2] = \rho$, and $\mathbb{E}[XY] \ge \delta\rho$ for some universal constant $\delta > 0$. Then*

$$\frac{\delta^2}{4}\rho \le \mathbb{P}\Big[Y > \frac{\delta}{2}\Big] \le \Big(1 + \frac{4}{\delta^2}\Big)\rho. \qquad and \qquad \mathbb{P}\Big[X = 1 | Y > \frac{\delta}{2}\Big] > \frac{\delta^4}{16}.$$

Proof. First, note that $\mathbb{E}[XY] \ge \delta\rho$ implies $\mathbb{E}[Y|X = 1] \ge \delta$. Further, $\mathbb{E}[Y^2] = \rho$ implies $\mathbb{E}[Y^2|X = 1] \le 1$. By the Paley-Zygmund inequality,

$$\mathbb{P}\Big[Y > \frac{\delta}{2}|X = 1\Big] \ge \mathbb{P}\Big[Y > \frac{1}{2}\mathbb{E}[Y|X = 1]|X = 1\Big] \ge \frac{1}{4}\frac{(\mathbb{E}[Y|X = 1])^2}{\mathbb{E}[Y^2|X = 1]} \ge \frac{\delta^2}{4}.$$

This implies the lower bound

$$\mathbb{P}\Big[Y > \frac{\delta}{2}\Big] \ge \rho\mathbb{P}\Big[Y > \frac{\delta}{2}|X = 1\Big] \ge \rho\frac{\delta^2}{4}.$$

On the other hand, Chebychev inequality implies

$$\mathbb{P}\Big[Y > \frac{\delta}{2}|X = 0\Big] \le \frac{4}{\delta^2}\mathbb{E}[Y^2|X = 0] \le \frac{4\rho}{(1-\rho)\delta^2}.$$

Thus,

$$\mathbb{P}\Big[Y > \frac{\delta}{2}\Big] = (1-\rho)\mathbb{P}\Big[Y > \frac{\delta}{2}|X = 0\Big] + \rho\mathbb{P}\Big[Y > \frac{\delta}{2}|X = 1\Big] \le \rho\Big(1 + \frac{4}{\delta^2}\Big).$$

Finally, using Chebychev's inequality, $\mathbb{P}\Big[Y > \frac{\delta}{2}\Big] \le \frac{4\rho}{\delta^2}$. Bayes Theorem implies

$$\mathbb{P}\Big[X = 1 | Y > \frac{\delta}{2}\Big] = \frac{\rho\mathbb{P}\Big[Y > \frac{\delta}{2}|X = 1\Big]}{\mathbb{P}\Big[Y > \frac{\delta}{2}\Big]} \ge \frac{\delta^4}{16}.$$

This completes the proof. $\square$

Armed with Lemma 8.1, we turn to the proof of Lemma 1.6.

Proof of Lemma 1.6. Recall that $A = \lambda\rho\frac{v}{\sqrt{N\rho}}(\frac{v}{\sqrt{N\rho}})^T + W$, and thus for $\lambda\rho > 1 + \varepsilon$, the celebrated BBP phase transition [11, 21] implies that there exists a universal constant $\delta := \delta(\varepsilon) > 0$ such that w.h.p as $N \to \infty$

$$\frac{1}{N}(v, \hat{v}) \ge \delta\rho.$$

34

In the subsequent discussion, we condition on this good event. Consider the two dimensional measure $\mu_N = \frac{1}{N}\sum_{i=1}^N \delta_{(v_i, \hat{v}_i)}$. Set $(X, Y) \sim \mu_N$ and note that (X, Y) satisfy the theses of Lemma 8.1. Lemma 8.1 implies that

$$\frac{\delta^2}{4}\rho N \leq |\tilde{S}| \leq \left(1 + \frac{4}{\delta^2}\right)\rho N.$$

On the event $|\tilde{S}| < \rho N$, we have,

$$\frac{1}{N}(v, v_{\hat{S}}) \geq \frac{\delta^2}{4}\rho \cdot \frac{\delta^4}{16} = \frac{\delta^6}{64}\rho.$$

On the other hand, if $|\tilde{S}| > N\rho$,

$$(v, v_{\hat{S}}) \sim \text{Hyp}(|\tilde{S}|, |\tilde{S} \cap \{i : v_i = 1\}|, N\rho).$$

Thus we have,

$$\mathbb{E}[(v, v_{\hat{S}})] = N\rho \frac{|\tilde{S} \cap \{i : v_i = 1\}|}{|\tilde{S}|} = N\rho \mathbb{P}\Big[X = 1 \big| Y > \frac{\delta}{2}\Big] \geq N\rho \frac{\delta^4}{16}.$$

Further, direct computation reveals

$$\text{Var}[(v, v_{\hat{S}})] = N\rho \cdot pq \cdot \frac{|\tilde{S}| - N\rho}{|\tilde{S}| - 1},$$

where we denote $p := \frac{|\tilde{S} \cap \{i : v_i = 1\}|}{|\tilde{S}|} = 1 - q$. Upon observing that $pq \leq 1/4$, we have, $\text{Var}[(v, \hat{v})] \leq N\rho/4$. Thus by Chebychev inequality,

$$\mathbb{P}\Big[(v, v_{\hat{S}}) < N\rho \frac{\delta^4}{32}\Big] \leq \frac{N\rho}{\left(N\rho \frac{\delta^4}{32}\right)^2} = O\Big(\frac{1}{N}\Big).$$

This implies that $(v, v_{\hat{S}}) \geq N\rho \frac{\delta^4}{32}$ whp over the sampling process, and establishes that the constructed estimator recovers the support approximately. This completes the proof.

$\square$

Data availability statement: Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

Appendix A. Ruelle Probability Cascades

For the convenience of the reader, we briefly review here basic properties of Ruelle probability cascades (RPCs) (sometimes called, Derrida Ruelle Probability Cascades) used throughout this paper.

A.1. **Construction and basic properties.** Let us begin by recalling the construction of RPCs and some basic properties. See, e.g., [69] or [48, Sec. 3.3].

Fix $r \geq 1$ and let $\mathcal{A}_r$ be as in Section 6.1. We label the vertices of this tree as $\mathcal{A}_r = \mathbb{N}^0 \cup \mathbb{N}^1 \cup ... \cup \mathbb{N}^r$, where a vertex at depth k has label $\alpha = (\alpha^1, ..., \alpha^k)$ which corresponds to the root-vertex path, $\emptyset \to \alpha^1 \to (\alpha^1, \alpha^2) \to \cdots \to (\alpha^1, \ldots, \alpha^k)$. As above, we denote this path by $p(\alpha)$. Denote the depth of a vertex by $|\alpha|$ and let $\partial \mathcal{A}_r$ denote the leaves of $\mathcal{A}_r$.

For $r \geq 1$ and a fixed sequence $0 = \mu_{-1} < \mu_0 < \ldots < \mu_r = 1$, we construct the corresponding RPC as follows. Let $m_\theta(dx) = \theta x^{-\theta-1} dx$. For each non-leaf vertex $\alpha \in \mathcal{A}_r \backslash \partial \mathcal{A}_r$, we assign an independent copy of the Poisson point process $PPP(m_{\mu_{|\alpha|}}(dx))$ arranged in decreasing order, where we assign each child of α the term in the point process of corresponding rank. This yields a

35

collection $(u_\alpha)_{\alpha \in \mathcal{A}_r}$ of random variables. Let $w_\alpha = \prod_{\gamma \in p(\alpha)} u_\gamma$ and finally consider the normalized collection $(v_\alpha)_{\alpha \in \mathcal{A}_r}$ given by

$$v_\alpha = \frac{w_\alpha}{\sum_{|\beta| = |\alpha|} w_\beta}.$$

The *Ruelle probability cascade with parameters* $(\mu_k)_{k=-1}^r$ is the stochastic process $(v_\alpha)_{\alpha \in \partial \mathcal{A}_r}$.

It will also be helpful to note the following. Let $\mu \in \mathcal{M}_1([0,\rho])$ have finite support and consider the overlap distribution $\mathcal{R}(\mu)$ defined as in (6.21). We note here the following elementary consequence of the definition. For a proof see, e.g., [69, Eq. 2.82].

Lemma A.1. *Let $\mu \in \mathcal{M}_1([0,\rho])$ be of finite support and consider π as defined in* (6.21). *Then* $\mathbb{E}\pi(R_{12} = q) = \mu(\{q\})$.

A.2. **Calculating expectations and Parisi PDEs.** Let us now recall the following well-known result connecting Ruelle probability cascades to Parisi-type PDEs. (Recall again that we may view $\mathbb{N}^r = \partial \mathcal{A}_r$.) Results of this type appear in different notations throughout the spin glass literature and are sometimes referred to as consequences of the Bolthausen-Sznitman invariance of RPCs. The following results are taken from [7].

Theorem A.2 (Theorem 6 from [7]). *Fix $r \geq 1, T > 0$, and sequences*

$$0 = q_0 < q_1 < \ldots < q_r = T$$
$$0 = \mu_{-1} < \mu_0 < \ldots < \mu_r = 1.$$

Let $\psi \in C^1([0,T])$ be non-negative increasing and let $(g_\psi(\alpha))_{\alpha \in \mathbb{N}^r}$ denote the centered Gaussian process with covariance

$$\mathbb{E}g_\psi(\alpha)g_\psi(\beta) = \psi(q_{|\alpha \wedge \beta|}).$$

Finally, let $(v_\alpha)_{\alpha \in \mathbb{N}^r}$ be a Ruelle probability cascade with parameters (μ_k). Then we have the following.

(1) *For any smooth f of at most linear growth we have that*

$$\mathbb{E}\log \sum_{\alpha \in \mathbb{N}^r} v_\alpha \exp[f(g_\psi(\alpha))] = \phi_\mu(0,0),$$

where ϕ is the unique solution to

$$\partial_t \phi + \frac{\psi'}{2}\left(\Delta \phi + \mu[0,t](\partial_x \phi)^2\right) = 0$$
$$\phi(T,x) = f(x),$$

and $\mu \in \mathcal{M}_1([0,T])$ is given by $\mu(\{q_k\}) = \mu_k - \mu_{k-1}$.

(2) *If $(g_\psi^i)_{i=1}^M$ are iid copies of g_ψ and (f_i) are of at most linear growth, then*

$$\mathbb{E}\log \sum_\alpha v_\alpha \exp\left(\sum_i f_i(g_\psi^i(\alpha))\right) = \sum_i \mathbb{E}\log \sum v_\alpha \exp f_i(g_\psi(\alpha)).$$

We note here the following corollary which has appeared more or less verbatim in many papers and follows by applying item 1 in the above with $f(x) = x$ and the Cole-Hopf iteration.

Corollary A.3 (Proposition 7 from [7]). *We have that*

$$\mathbb{E}\log \sum_{\alpha \in \mathbb{N}^r} v_\alpha \exp[g_\psi(\alpha)] = \frac{1}{2}\int_0^T \psi'(s)\mu([0,s])ds.$$

The simplicity of the formula in this case follows from noting that the heat equation with initial data e^x is exactly solvable.

36

Appendix B. Strict convexity

To prove strict convexity of $\mathcal{P}$, let us introduce the following notation. For the sake of clarity, we make the dependence of u^i_ν on Λ_i explicit by writing $u_{\nu,\Lambda_i}(t,x) = u^i_\nu(t,x)$. Furthermore, as (2.1) is invariant under a spatial translation, we see that

$$u_{\nu,\Lambda}(t,x) = u_{\nu_0,0}(t, x + \Lambda + \nu(\{\rho\})).$$

where $\nu_0 = \nu - \nu(\{\rho\})\delta_\rho$. It will also be helpful to recall the dynamic programming principle for $u_{\nu,\Lambda_i}(t,x)$ from [51, Lemma 3.5].

Lemma. *For any $(\nu, \lambda) \in \mathcal{A} \times \mathbb{R}$ of the form $d\nu = mdt + c\delta_\rho$, we have that for any $t < t' \le \rho$,*

$$u_{\nu,\lambda}(t,x) = \sup_{\alpha \in \mathcal{B}_{t'}} \mathbb{E}\Big[u_{\nu,\lambda}(t', X^\alpha_{t'}) - \int_t^{t'} m(s)\alpha_s^2 ds\Big],$$

where X^α solves

$$\begin{cases} dX_s = m(s)\alpha_s^2 ds + \sqrt{2}dW_s \\ X_t = x, \end{cases}$$

W_s is a standard brownian motion and $\mathcal{B}_{t'}$ is the space of bounded stochastic processes that are progressively measurable with respect to the filtration of $(W_s)_{s \le t'}$. Furthermore, any optimal control α_s^* satisfies

$$m(s)\alpha_s^* = m(s)\partial_x u_{\nu,\lambda}(s, X_s) \quad a.s.$$

The proof will begin with the following observation.

Lemma B.1. *For any $\nu \in \mathcal{A}_0$, and any $t \in [0, \rho)$, $u_{\nu,0}(t,x)$ is strictly convex in x.*

Proof. Fix $x_0, x_1 \in \mathbb{R}$ distinct, $\theta \in (0,1)$, and let $x_\theta = \theta x_0 + (1-\theta)x_0$. Let (X^θ_s) denote the optimal trajectory corresponding to $u_{\nu,0}(t, x_\theta)$, and similarly let $\alpha_s^\theta = \partial_x u_{\nu,0}(s, X^\theta_s)$ denote a corresponding optimal control.

Observe that if we let $G_s = \int_0^s \sqrt{2}dW_{s_1} + \int_0^s 2m\left(\alpha^\theta_{s_1}\right)^2 ds_1$, then $X^\theta_s = G_s + x_\theta$. We first claim that the law of G_ρ charges any interval $(a, b) \subseteq \mathbb{R}$. As it is possible that $\int_0^\rho m^2(s)ds = \infty$, Novikov's condition does not apply, so we cannot apply Girsanov's theorem directly to G_ρ. We circumvent this by a localization argument as follows.

Fix $0 < s < \rho$. Since $0 \le \sup|\alpha_s^\theta| \le 1$ by (2.8), we have that $b_t = 2m(t)(\alpha_t^\theta)^2$ has $\sup_{0 \le t \le s}|b_t| < C(s)$ for some non-random $C(s) > 0$. By Girsanov's theorem [80, Lemma 6.4.1] there is a tilt of the law of G_s such that the law under this tilted measure is Gaussian. Thus $\mathbb{P}[G_s \in \mathcal{I}] > 0$ for any interval $\mathcal{I}$. Now, fix an interval $\mathcal{I} = (a, b)$.

Note that

$$G_\rho = G_{s'} + \int_{s'}^\rho \sqrt{2}dW_{s_1} + \int_{s'}^\rho 2m(\alpha^\theta_{s_1})^2 ds_1,$$

and thus

$$|G_\rho - G_{s'}| \le 2\int_{s'}^\rho mds_1 + \Big|\int_{s'}^\rho \sqrt{2}dW_{s_1}\Big|.$$

Further, $\int_{s'}^\rho \sqrt{2}dW_{s_1} \sim \mathcal{N}(0, 2(\rho - s'))$. Fix $C' > 0$, and let $(c, d) \subset \mathcal{I}$ such that

$$\min\{c - a, d - b\} > 2\int_{s'}^\rho mds_1 + C'\sqrt{2(\rho - s')}.$$

Such an interval always exists once $\rho - s'$ is sufficiently small. This implies

$$\mathbb{P}[G_\rho \in \mathcal{I}] \ge \mathbb{P}\Big[G_{s'} \in (c, d), |G_\rho - G_{s'}| \le 2\int_{s'}^\rho mds_1 + C'\sqrt{2(\rho - s')}\Big] > 0.$$

37

This, in turn, establishes that for any interval $\mathcal{I}$, $\mathbb{P}[G_\rho \in \mathcal{I}] > 0$.

Let $Y = G_\rho + x_1$ and $Z = G_\rho + x_0$, then we have that

$$u_{\nu,0}(t,x) = \mathbb{E}\left[\left(X_\rho^\theta\right)_+ - \int_t^\rho 2m\left(\alpha_s^\theta\right)^2 ds\right]$$

$$< \theta\mathbb{E}\left[Y_+ - \int_t^\rho 2m\left(\alpha_s^\theta\right)^2 ds\right] + (1-\theta)\mathbb{E}\left[Z_+ - \int_t^\rho 2m\left(\alpha_s^\theta\right)^2 ds\right]$$

$$\leq \theta u_{\nu,0}(t,x_0) + (1-\theta)u_{\nu,0}(t,x_1),$$

where in the second line we use that if $a < 0 < b$ then $(\theta a + (1-\theta)b)_+ < \theta a_+ + (1-\theta)b_+$ and that

$$\mathbb{P}(YZ < 0) = \mathbb{P}((x_0 + G_\rho)(x_1 + G_\rho) < 0) > 0. \qquad \square$$

Lemma B.2. *The functional $\mathcal{P}$ is strictly convex on $\mathcal{A}_0 \times \mathbb{R}$.*

Remark B.3. It will be easy to see from the proof that it is also convex on $\mathcal{A} \times \mathbb{R}$. Strict convexity, however, fails on this larger domain due to the invariance of the functional under the map $(mdt + c\delta_\rho, \Lambda_1, \Lambda_2) \mapsto (mdt, \Lambda_1 + 2c, \Lambda_2 + 2c)$.

Proof. This proof follows the approach of [52, Theorem 20]. Fix $(\nu_1, \lambda_1), (\nu_2, \lambda_2) \in \mathcal{A}_0 \times \mathbb{R}$, and $\theta \in [0,1]$, let

$$(\nu_\theta, \lambda_\theta) = \theta(\nu_1, \lambda_1) + (1-\theta)(\nu_2, \lambda_2),$$

and write $\nu_1 = m_1 dt$, $\nu_2 = m_2 dt$, $\nu_\theta = m_\theta dt$ with $m_\theta = \theta m_1 + (1-\theta)m_2$.

Let X_s^θ denote the optimal trajectory for the stochastic control problem for $u_{\nu_\theta, \lambda_\theta}$ and let $\alpha^\theta = \partial_x u_{\nu_\theta, \lambda_\theta}(s, X_s^\theta)$ the corresponding control. Finally let Y^θ, Z^θ solve the SDEs

$$dY^\theta = 2m_1\left(\alpha_s^\theta\right)^2 ds + \sqrt{2}dW_s$$

$$dZ^\theta = 2m_2(\alpha_s^\theta)^2 ds + \sqrt{2}dW_s$$

with $Y_0 = Z_0 = 0$.

Now, fix some $0 < t < \rho$. then by the dynamic programming principle,

$$u_{\nu_\theta, \lambda_\theta}(0,0) = \mathbb{E}\left[u_{\nu_\theta, \lambda_\theta}(t, X_t^\theta) - \int_0^t m_\theta(s)\left(\alpha_s^\theta\right)^2 ds\right].$$

Since the equation (2.1) is invariant under translations of space, we see that $u_{\nu,\lambda}(t,x) = u_{\nu,0}(t, x+\lambda)$ for any (ν, λ). Thus we may rewrite the above as

$$u_{\nu_\theta, \lambda_\theta}(0,0) = \mathbb{E}u_{\nu_\theta, 0}(t, X_t^\theta + \lambda_\theta) - \int_0^t m_\theta(s)\left(\alpha_s^\theta\right)^2 ds$$

$$\leq \theta\mathbb{E}\left(u_{\nu_\theta, 0}(t, Y_t + \lambda_1) - \int_0^t m_1(s)(\alpha_s^\theta)^2 ds\right) + (1-\theta)\mathbb{E}\left(u_{\nu_\theta, 0}(t, Z_t + \lambda_2) - \int_0^t m_2(\alpha_s^\theta)^2 ds\right)$$

$$\leq \theta u_{\nu_1, \lambda_1}(0,0) + (1-\theta)u_{\nu_2, \lambda_2}(0,0),$$

in the first inequality we have used the convexity of $u_{\nu,0}(t,0)$ in space. Note that in fact the first inequality is strict, provided

$$\mathbb{P}((Y_t + \lambda_1) \neq (Z_t + \lambda_2)) > 0.$$

In particular, it suffices to show that $Var(Y_t - Z_t) + |\lambda_1 - \lambda_2| > 0$. Thus if $\lambda_1 \neq \lambda_2$ we are done. If they are equal then we know that $m_1 \neq m_2$. In this case, by right continuity and monotonicity, there must be some $s < \tau < \rho$ such that $m_1(t') \neq m_2(t')$ on $[s, \tau]$ (that we can take $t < \rho$ follows from the fact that if $\nu_1 \neq \nu_2$ then m_1 and m_2 must differ on a set of positive lebesgue measure). In particular, we choose $t = \tau$ from now on.

38

Note that by Ito's lemma, our choice of α_s^θ is a martingale, with

$$\alpha_s^\theta - \alpha_0^\theta = \int_0^s \sqrt{2}\partial_{xx}u_{\nu,0}(s_1, X_{s_1}^\theta)dW_{s_1}.$$

Thus by Ito's isometry, if we let $\Delta_s = 2(m_1 - m_2)$

$$Var(Y_t - Z_t) = \mathbb{E}\Big(\int_0^t \Delta_s(\alpha_s^\theta - \alpha_0^\theta)ds\Big)^2 = \int\int_{[0,t]^2} \Delta_{s_1}\Delta_{s_2}K(s_1, s_2)ds_1 ds_2,$$

where

$$K(s,t) = \mathbb{E}(\alpha_s - \alpha_0)(\alpha_t - \alpha_0) = \int_0^{t\wedge s} 2\mathbb{E}\left(\partial_{xx}u_{\nu,0}(s_1, X_{s_1}^\theta)\right)^2 ds_1 = p(t \wedge s) = p(t) \wedge p(s),$$

where $p(t) = \int_0^t 2\mathbb{E}\left(\partial_{xx}u_{\nu,0}(s, X_s^\theta)\right)^2 ds$. Notice that since $t < \rho$, we have that $\Delta_t \in L^2([0,t])$. Thus to show positivity of the variance, it suffices to show that K is positive definite.

Since u is $C^2([0, t + \epsilon] \times \mathbb{R})$ for some $\epsilon > 0$ small enough, and u is strictly convex, we have that $\partial_{xx}u(t, x) > 0$ lebesgue a.e. x. Thus $p(t)$ is strictly increasing. Thus this kernel corresponds to a monotone time change of a Brownian motion, so that it is positive definite. □

References

[1] Emmanuel Abbe. Community detection and stochastic block models: recent developments. *The Journal of Machine Learning Research*, 18(1):6446–6531, 2017.

[2] Dimitris Achlioptas, Amin Coja-Oghlan, and Federico Ricci-Tersenghi. On the solution-space geometry of random constraint satisfaction problems. *Random Structures & Algorithms*, 38(3):251–268, 2011.

[3] Louigi Addario-Berry and Pascal Maillard. The algorithmic hardness threshold for continuous random energy models. *arXiv preprint arXiv:1810.05129*, 2018.

[4] Michael Aizenman, Robert Sims, and Shannon L Starr. Extended variational principle for the sherrington-kirkpatrick spin-glass model. *Physical Review B*, 68(21):214403, 2003.

[5] Noga Alon, Michael Krivelevich, and Benny Sudakov. Finding a large hidden clique in a random graph. *Random Structures & Algorithms*, 13(3-4):457–466, 1998.

[6] Arash A Amini and Martin J Wainwright. High-dimensional analysis of semidefinite relaxations for sparse principal components. In *2008 IEEE International Symposium on Information Theory*, pages 2454–2458. IEEE, 2008.

[7] Louis-Pierre Arguin. Spin glass computations and Ruelle's probability cascades. *J. Stat. Phys.*, 126(4-5):951–976, 2007.

[8] Antonio Auffinger and Wei-Kuo Chen. Parisi formula for the ground state energy in the mixed *p*-spin model. *The Annals of Probability*, 45(6B):4617–4631, Nov 2017.

[9] Antonio Auffinger, Wei-Kuo Chen, and Qiang Zeng. The sk model is full-step replica symmetry breaking at zero temperature. *arXiv preprint arXiv:1703.06872*, 2017.

[10] Francesco Baffioni and Francesco Rosati. Some exact results on the ultrametric overlap distribution in mean field spin glass models (i). *The European Physical Journal B-Condensed Matter and Complex Systems*, 17(3):439–447, 2000.

[11] Jinho Baik, Gérard Ben Arous, and Sandrine Péché. Phase transition of the largest eigenvalue for nonnull complex sample covariance matrices. *The Annals of Probability*, 33(5):1643–1697, 2005.

[12] Sivaraman Balakrishnan, Mladen Kolar, Alessandro Rinaldo, Aarti Singh, and Larry Wasserman. Statistical and computational tradeoffs in biclustering. In *NeurIPS 2011 workshop on computational trade-offs in statistical learning*, volume 4, 2011.

[13] Jess Banks, Cristopher Moore, Roman Vershynin, Nicolas Verzelen, and Jiaming Xu. Information-theoretic bounds and phase transitions in clustering, sparse pca, and submatrix localization. *IEEE Transactions on Information Theory*, 64(7):4872–4894, 2018.

[14] Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.

[15] Jean Barbier, Mohamad Dia, Nicolas Macris, Florent Krzakala, Thibault Lesieur, and Lenka Zdeborová. Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula. In *Proceedings of the 30th International Conference on Neural Information Processing Systems*, NIPS'16, page 424–432, 2016.

[16] Jean Barbier, Nicolas Macris, and Cynthia Rush. All-or-nothing statistical and computational phase transitions in sparse spiked matrix estimation. *arXiv preprint arXiv:2006.07971*, 2020.

[17] Adriano Barra, Giuseppe Genovese, and Francesco Guerra. Equilibrium statistical mechanics of bipartite spin systems. *Journal of Physics A: Mathematical and Theoretical*, 44(24):245002, 2011.

[18] Gérard Ben Arous, Reza Gheissari, and Aukosh Jagannath. Algorithmic thresholds for tensor pca. *arXiv preprint arXiv:1808.00921*, 2018.

[19] Gérard Ben Arous and Aukosh Jagannath. Spectral gap estimates in mean field spin glasses. *Communications in Mathematical Physics*, 361(1):1–52, 2018.

[20] Gérard Ben Arous, Alexander S Wein, and Ilias Zadik. Free energy wells and overlap gap property in sparse pca. In *Conference on Learning Theory*, pages 479–482. PMLR, 2020.

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

[22] Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *Conference on Learning Theory*, pages 1046–1066, 2013.

[23] Shankar Bhamidi, Partha S Dey, and Andrew B Nobel. Energy landscape for large average submatrix detection problems in gaussian random matrices. *Probability Theory and Related Fields*, 168(3-4):919–983, 2017.

[24] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities: A nonasymptotic theory of independence*. Oxford university press, 2013.

[25] Matthew Brennan, Guy Bresler, and Wasim Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. *arXiv preprint arXiv:1806.07508*, 2018.

[26] Matthew Brennan, Guy Bresler, and Wasim Huleihel. Universality of computational lower bounds for submatrix detection. *arXiv preprint arXiv:1902.06916*, 2019.

[27] Cristina Butucea and Yuri I Ingster. Detection of a sparse submatrix of a high-dimensional noisy matrix. *Bernoulli*, 19(5B):2652–2688, 2013.

[28] Cristina Butucea, Yuri I Ingster, and Irina A Suslina. Sharp variable selection of a sparse submatrix in a high-dimensional noisy matrix. *ESAIM: Probability and Statistics*, 19:115–134, 2015.

[29] T Tony Cai, Tengyuan Liang, and Alexander Rakhlin. Computational and statistical boundaries for submatrix localization in a large noisy matrix. *The Annals of Statistics*, 45(4):1403–1430, 2017.

[30] Venkat Chandrasekaran and Michael I Jordan. Computational and statistical tradeoffs via convex relaxation. *Proceedings of the National Academy of Sciences*, 110(13):E1181–E1190, 2013.

[31] Wei-Kuo Chen, David Gamarnik, Dmitry Panchenko, and Mustazee Rahman. Suboptimality of local algorithms for a class of max-cut problems. *The Annals of Probability*, 47(3):1587–1618, 2019.

[32] Yudong Chen and Jiaming Xu. Statistical-computational tradeoffs in planted problems and submatrix localization with a growing number of clusters and submatrices. *The Journal of Machine Learning Research*, 17(1):882–938, 2016.

[33] Amin Coja-Oghlan, Amir Haqshenas, and Samuel Hetterich. Walksat stalls well below satisfiability. *SIAM Journal on Discrete Mathematics*, 31(2):1160–1173, 2017.

[34] Thomas M Cover and Joy A Thomas. Elements of information theory john wiley & sons. *New York*, 68:69–73, 1991.

[35] N. G. de Bruijn and P. Erdös. Some linear and some quadratic recursion formulas. II. *Nederl. Akad. Wetensch. Proc. Ser. A.* **55** = *Indagationes Math.*, 14:152–163, 1952.

[36] Yash Deshpande and Andrea Montanari. Information-theoretically optimal sparse pca. In *2014 IEEE International Symposium on Information Theory*, pages 2197–2201. IEEE, 2014.

[37] Yash Deshpande and Andrea Montanari. Improved sum-of-squares lower bounds for hidden clique and hidden submatrix problems. In *Conference on Learning Theory*, pages 523–562, 2015.

[38] Yunzi Ding, Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Subexponential-time algorithms for sparse pca. *arXiv preprint arXiv:1907.11635*, 2019.

[39] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM (JACM)*, 64(2):8, 2017.

[40] David Gamarnik and Quan Li. Finding a large submatrix of a gaussian random matrix. *The Annals of Statistics*, 46(6A):2511–2561, 2018.

[41] David Gamarnik and Madhu Sudan. Performance of sequential local algorithms for the random nae-k-sat problem. *SIAM Journal on Computing*, 46(2):590–619, 2017.

[42] David Gamarnik and Ilias Zadik. High dimensional regression with binary coefficients. estimating squared error and a phase transtition. In *Conference on Learning Theory*, pages 948–953, 2017.

[43] David Gamarnik and Ilias Zadik. Sparse high-dimensional linear regression. algorithmic barriers and a local search algorithm. *arXiv preprint arXiv:1711.04952*, 2017.

[44] David Gamarnik and Ilias Zadik. The landscape of the planted clique problem: Dense subgraphs and the overlap gap property. *arXiv preprint arXiv:1904.07174*, 2019.

[45] Chao Gao, Zongming Ma, and Harrison H Zhou. Sparse cca: Adaptive estimation and computational barriers. *The Annals of Statistics*, 45(5):2074–2101, 2017.

[46] Samuel B Hopkins, Pravesh Kothari, Aaron Henry Potechin, Prasad Raghavendra, and Tselil Schramm. On the integrality gap of degree-4 sum of squares for planted clique. *ACM Transactions on Algorithms (TALG)*, 14(3):28, 2018.

[47] Samuel B Hopkins, Tselil Schramm, Jonathan Shi, and David Steurer. Fast spectral algorithms from sum-of-squares proofs: tensor decomposition and planted sparse vectors. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 178–191. ACM, 2016.

[48] Aukosh Jagannath. Approximate ultrametricity for random measures and applications to spin glasses. *Comm. Pure Appl. Math.*, 70(4):611–664, 2017.

[49] Aukosh Jagannath, Justin Ko, and Subhabrata Sen. Max κ-cut and the inhomogeneous potts spin glass. *The Annals of Applied Probability*, 28(3):1536–1572, 2018.

[50] Aukosh Jagannath, Patrick Lopatto, and Léo Miolane. Statistical thresholds for tensor PCA. *Ann. Appl. Probab.*, 30(4):1910–1933, 2020.

[51] Aukosh Jagannath and Subhabrata Sen. On the unbalanced cut problem and the generalized sherrington-kirkpatrick model. *Ann. Inst. Henri Poincar. D*, to appear.

[52] Aukosh Jagannath and Ian Tobasco. A dynamic programming approach to the parisi functional. *Proceedings of the American Mathematical Society*, 144(7):3135–3150, 2016.

[53] Aukosh Jagannath and Ian Tobasco. Low temperature asymptotics of spherical mean field spin glasses. *Communications in Mathematical Physics*, 352(3):979–1017, 2017.

[54] Aukosh Jagannath and Ian Tobasco. Some properties of the phase diagram for mixed p-spin glasses. *Probability Theory and Related Fields*, 167(3-4):615–672, 2017.

[55] Mladen Kolar, Sivaraman Balakrishnan, Alessandro Rinaldo, and Aarti Singh. Minimax localization of structural information in large noisy matrices. In *Advances in Neural Information Processing Systems*, pages 909–917, 2011.

[56] Florent Krzakala, Jiaming Xu, and Lenka Zdeborová. Mutual information in rank-one matrix estimation. In *2016 IEEE Information Theory Workshop (ITW)*, pages 71–75. IEEE, 2016.

[57] Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, 173(3-4):859–929, 2019.

[58] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Phase transitions in sparse pca. In *2015 IEEE International Symposium on Information Theory (ISIT)*, pages 1635–1639. IEEE, 2015.

[59] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. Constrained low-rank matrix estimation: Phase transitions, approximate message passing and applications. *Journal of Statistical Mechanics: Theory and Experiment*, 2017(7):073403, 2017.

[60] Tengyu Ma and Avi Wigderson. Sum-of-squares lower bounds for sparse pca. In *Advances in Neural Information Processing Systems*, pages 1612–1620, 2015.

[61] Zongming Ma and Yihong Wu. Computational barriers in minimax submatrix detection. *The Annals of Statistics*, 43(3):1089–1116, 2015.

[62] Raghu Meka, Aaron Potechin, and Avi Wigderson. Sum-of-squares lower bounds for planted clique. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 87–96. ACM, 2015.

[63] Marc Mézard, Thierry Mora, and Riccardo Zecchina. Clustering of solutions in the random satisfiability problem. *Physical Review Letters*, 94(19):197205, 2005.

[64] Andrea Montanari. Finding one community in a sparse graph. *Journal of Statistical Physics*, 161(2):273–299, 2015.

[65] Andrea Montanari. Optimization of the sherrington-kirkpatrick hamiltonian. *arXiv preprint arXiv:1812.10897*, 2018.

[66] Andrea Montanari, Daniel Reichman, and Ofer Zeitouni. On the limitation of spectral methods: From the gaussian hidden clique problem to rank-one perturbations of gaussian tensors. In *Advances in Neural Information Processing Systems*, pages 217–225, 2015.

[67] Cristopher Moore. The computer science and physics of community detection: Landscapes, phase transitions, and hardness. *arXiv preprint arXiv:1702.00467*, 2017.

[68] Dmitry Panchenko. The Parisi ultrametricity conjecture. *Ann. of Math. (2)*, 177(1):383–393, 2013.

[69] Dmitry Panchenko. *The Sherrington-Kirkpatrick model*. Springer, 2013.

[70] Dmitry Panchenko. The Parisi formula for mixed p-spin models. *Ann. Probab.*, 42(3):946–958, 2014.

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

[72] Dmitry Panchenko. Free energy in the mixed p-spin models with vector spins. *The Annals of Probability*, 46(2):865–896, 2018.

[73] Dmitry Panchenko. Free energy in the potts spin glass. *The Annals of Probability*, 46(2):829–864, 2018.

[74] Mustazee Rahman and Balint Virag. Local algorithms for independent sets are half-optimal. *The Annals of Probability*, 45(3):1543–1577, 2017.

[75] Emile Richard and Andrea Montanari. A statistical model for tensor pca. In *Advances in Neural Information Processing Systems*, pages 2897–2905, 2014.

[76] Benjamin Rossman. *Average-case complexity of detecting cliques*. PhD thesis, Massachusetts Institute of Technology, 2010.

[77] Tselil Schramm and Alexander S Wein. Computational barriers to estimation from low-degree polynomials. *arXiv preprint arXiv:2008.02269*, 2020.

[78] Andrey A Shabalin, Victor J Weigman, Charles M Perou, and Andrew B Nobel. Finding large average submatrices in high dimensional data. *The Annals of Applied Statistics*, 3(3):985–1012, 2009.

[79] J. Michael Steele. *Probability theory and combinatorial optimization*, volume 69 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

[80] Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional diffusion processes*. Classics in Mathematics. Springer-Verlag, Berlin, 2006. Reprint of the 1997 edition.

[81] Eliran Subag. Following the ground-states of full-rsb spherical spin glasses. *arXiv preprint arXiv:1812.04588*, 2018.

[82] Michel Talagrand. *Mean field models for spin glasses. Volume II*, volume 55 of *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]*. Springer, Heidelberg, 2011. Advanced replica-symmetry and low temperature.

[83] Alexander S Wein, Ahmed El Alaoui, and Cristopher Moore. The kikuchi hierarchy and tensor pca. *arXiv preprint arXiv:1904.03858*, 2019.

[84] Yihong Wu and Jiaming Xu. Statistical problems with planted structures: Information-theoretical and computational limits. *Information-Theoretic Methods in Data Science*, page 383, 2021.

SLOAN SCHOOL OF MANAGEMENT, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, gamarnik@mit.edu

DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCES AND DEPARTMENT OF APPLIED MATHEMATICS, UNIVERSITY OF WATERLOO, a.jagannath@uwaterloo.ca

DEPARTMENT OF STATISTICS, HARVARD UNIVERSITY subhabratasen@fas.harvard.edu