

On the energy landscape of the mixed even p -spin model

Wei-Kuo Chen*

Madeline Handschy†

Gilad Lerman‡

March 27, 2017

Abstract

We investigate the energy landscape of the mixed even p -spin model with Ising spin configurations. We show that for any given energy level between zero and the maximal energy, with overwhelming probability there exist exponentially many distinct spin configurations such that their energies stay near this energy level. Furthermore, their magnetizations and overlaps are concentrated around some fixed constants. In particular, at the level of maximal energy, we prove that the Hamiltonian exhibits exponentially many orthogonal peaks. This improves the results of Chatterjee [20] and Ding-Eldan-Zhai [29], where the former established a logarithmic size of the number of the orthogonal peaks, while the latter proved a polynomial size. Our second main result obtains disorder chaos at zero temperature and at any external field. As a byproduct, this implies that the fluctuation of the maximal energy is superconcentrated when the external field vanishes and obeys a Gaussian limit law when the external field is present.

AMS 2000 subject classifications. 60K35, 60G15, 82B44

Keywords and phrases. Disorder chaos, energy landscape, multiple peaks, Parisi formula, Sherrington-Kirkpatrick model

Contents

1	Introduction	2
1.1	The mixed even p -spin model	3
1.2	Energy landscape	4
1.3	Chaos in disorder	4
1.4	Main ideas of proofs	6
1.5	Open questions	7
1.6	Organization of the paper	7
2	Parisi formula	8
2.1	Uniqueness	8
2.2	Optimality	13
3	Establishing disorder chaos	16
3.1	Guerra-Talagrand inequality	16
3.2	Determination of $q_{t,h}$	18
3.3	Control of the GT bound	19
3.4	Proof of Theorem 2	24

*School of Mathematics, University of Minnesota. Email: wkchen@umn.edu

†School of Mathematics. University of Minnesota. Email: hands014@umn.edu

‡School of Mathematics, University of Minnesota. Email: lerman@math.umn.edu

4	Establishing exponential number of multiple peaks	25
4.1	Proof of Proposition 1	25
4.2	Proof of Theorem 1	30
Appendix		32

1 Introduction

Initially invented by theoretical physicists, spin glass models are disordered spin systems created to explain the strange magnetic behavior of certain alloys. These models are typically formulated as families of highly correlated random variables, called Hamiltonians or energies in physics, indexed by configuration spaces with given metrics. This framework has also been extended to a variety of disordered models in other scientific disciplines including computer science [8, 28, 32, 38] and neural networks [1, 27, 37, 43].

Understanding the energy landscape of spin glass systems is a challenging endeavor. When the Hamiltonian of a spin glass model is defined as a Gaussian field indexed by a differentiable manifold, it is natural to study the distribution of the extrema and critical points of the Hamiltonian. Over the past two decades, physicists have intensively investigated such problems for the spherical pure p -spin model; see Charbonneau et al. [18], Fyodorov-William [31], Fyodorov [30], and Kurchan-Parisi-Virasoro [36]. Rigorous mathematical results for the spherical pure p -spin and mixed p -spin models have also appeared in recent years. The description of complexity was presented in Auffinger-Ben Arous [3], Auffinger-Ben Arous-Cerny [4], and Subag [46], the statistics were studied in Subag [47] and Subag-Zeitouni [48], and the formula for the maximal energy was obtained in Chen-Sen [26] and Jagannath-Tobasco [34].

Studies related to energy statistics of spin glass models with discrete configuration spaces appear in Bovier-Kurkova [11, 12, 13, 14, 15, 16], Bovier-Klimovsky [10], and Ben Arous-Gaynard-Kuptsov [2]. These works consider a wide class of examples including the random energy model, the generalized random energy model, the Sherrington-Kirkpatrick (SK) model, the pure p -spin model and the random partition problems. More recently, Chatterjee [20] revolutionized the study of energy landscapes of spin glass models by establishing the so-called *superconcentration*, *disorder chaos*, and *multiple peaks* via an interpolation method for general discrete Gaussian fields. In the context of the SK model, his results imply that near the maximal energy the system exhibits multiple peaks in the sense that there exist logarithmically many spin configurations around the maximal energy and they are nearly orthogonal to each other. This picture extends to any energy level. Another paper of Chatterjee [19] establishes relationships between superconcentration, disorder chaos, and multiple peaks for more general Gaussian fields than spin glass models. Chatterjee's general results [19, 20, 21] on multiple peaks were improved by Ding-Eldan-Zhai [29] and as a byproduct, they showed that the number of orthogonal peaks in the SK model is at least of polynomial order.

Theoretical physicists believe that the multiple peaks discussed above occupy a large portion of the configuration space (see, e.g., Mézard-Parisi-Virasoro [39]). The main goal of this paper is to establish a description of this belief for the SK model as well as the mixed p -spin model. We show that for any given energy level between zero and the maximal energy, one can construct exponentially many spin configurations with energies around the given level. Furthermore, they possess the same magnetization and are equidistant from each other. In particular, at the level of maximal energy, our result strengthens the aforementioned results in [21, 29] by showing that the number of orthogonal peaks is of exponential order.

1.1 The mixed even p -spin model

We first review the mixed even p -spin model with Ising spin configurations. For any integer $N \geq 1$, let $\Sigma_N := \{-1, +1\}^N$ be the set of spin configurations. The Hamiltonian of the mixed even p -spin model is defined as

$$H_N(\sigma) = \sum_{p \in 2\mathbb{N}} c_p H_{N,p}(\sigma)$$

for $\sigma \in \Sigma_N$, where the sum is over all even p and $H_{N,p}$ is the pure p -spin Hamiltonian defined by

$$H_{N,p}(\sigma) = \frac{1}{N^{(p-1)/2}} \sum_{1 \leq i_1, \dots, i_p \leq N} g_{i_1, \dots, i_p} \sigma_{i_1} \cdots \sigma_{i_p}.$$

Here, for all $1 \leq i_1, \dots, i_p \leq N$ and $p \in 2\mathbb{N}$, g_{i_1, \dots, i_p} 's are i.i.d. standard Gaussian random variables. We assume that the sequence $(c_p)_{p \in 2\mathbb{N}}$ satisfies $c_p \neq 0$ for at least one p and it decays fast enough, for instance, $\sum_{p \in 2\mathbb{N}} 2^p c_p^2 < \infty$, so that the infinite sum H_N converges a.s. Under these assumptions, one readily computes that

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

where

$$\xi(s) := \sum_{p \in 2\mathbb{N}} c_p^2 s^p \tag{1}$$

and

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

is the overlap between σ^1 and σ^2 . The classical SK model corresponds to $\xi(s) = s^2/2$. If one replaces the space Σ_N by the sphere $\{\sigma \in \mathbb{R}^N : \sum_{i=1}^N \sigma_i^2 = N\}$, then the above Hamiltonian H_N is called the spherical mixed even p -spin model.

The mixed even p -spin model with external field is defined by

$$H_N^h(\sigma) = H_N(\sigma) + h N m_N(\sigma),$$

where $h \geq 0$ denotes the strength of the external field and

$$m_N(\sigma) := \frac{1}{N} \sum_{i=1}^N \sigma_i$$

is called the magnetization of σ . Denote by L_N^h the maximal energy of H_N^h , that is,

$$L_N^h := \max_{\sigma \in \Sigma_N} \frac{H_N^h(\sigma)}{N}.$$

Here, unlike the tradition in physics, we consider the maximum rather than the minimum of the energy. Nonetheless, they differ essentially only by a negative sign. It is well-known (see, for instance, [41]) that for all $h \in \mathbb{R}$,

$$M(h) := \lim_{N \rightarrow \infty} L_N^h \tag{2}$$

exists and is equal to a nonrandom number a.s., which can be computed through the Parisi formula for the free energy at positive temperature. Recently, Auffinger-Chen [7] proved that $M(h)$ can also be expressed as a Parisi-type formula for the limiting maximal energy L_N^h (see (6) below).

1.2 Energy landscape

We study the energy landscape of the Hamiltonian H_N . In order to investigate the behavior of the spin configurations at different energy levels, we introduce an auxiliary function,

$$E(h) := M(h) - hM'(h), \quad \forall h \geq 0, \quad (3)$$

where the differentiability of M is guaranteed by Proposition 8 in Subsection 4.1. The following proposition states that by varying h between zero and infinity, the function $E(h)$ continuously scans all energy levels of H_N between zero and its maximal energy.

Proposition 1. *$E(h)$ is nonincreasing and continuous with $E(0) = M(0)$ and $\lim_{h \rightarrow \infty} E(h) = 0$.*

Our main result on the energy landscape of the Hamiltonian H_N is formulated in Theorem 1 below. It uses the notion of the overlap constant, q_h . We postpone its definition to Subsection 3.2, but remark that q_h depends only on h and ξ and satisfies $0 < q_h < 1$ if $h \neq 0$ and $q_h = 0$ if $h = 0$.

Theorem 1. *Let $h \geq 0$. For any $\varepsilon > 0$, there exists a constant $K > 0$ depending on h, ξ, ε such that for any $N \geq 1$, with probability at least $1 - Ke^{-N/K}$, there exists $S_N(h) \subset \Sigma_N$ such that the following hold:*

- (i) $|S_N(h)| \geq e^{N/K}$,
- (ii) For any $\sigma \in S_N(h)$, $\left| \frac{H_N(\sigma)}{N} - E(h) \right| < \varepsilon$.
- (iii) For any $\sigma \in S_N(h)$, $|m_N(\sigma) - M'(h)| < \varepsilon$.
- (iv) For any distinct $\sigma, \sigma' \in S_N(h)$, $|R(\sigma, \sigma') - q_h| < \varepsilon$.

Theorem 1 means that sufficiently near the energy level $E(h)$ of H_N/N , there exist exponentially many spin configurations that are approximately equidistant from each other and their magnetizations are near $M'(h)$. Both $q_h = 0$ and $M'(h) = 0$ when $h = 0$. Consequently, Theorem 1 assures the existence of exponentially many orthogonal peaks with zero magnetization.

As we mentioned before, in the SK model the energy landscape at the maximal energy was studied previously by Chatterjee [20]. He showed that with probability at least $1 - K(\log N)^{-1/12}$ for some constant K , there exists a set of spin configurations of size at least $(\log N)^{1/8}$ that satisfy (ii) and (iv) with $q_h = 0$. In Ding-Eldan-Zhai [29], the size of this set was improved to be of polynomial order. Both works specified the dependence of ε on N . Theorem 1 here improves these results (without quantitative dependence of ε on N) by extending the number of orthogonal peaks to be of exponential order and establishing analogous geometric structure at all energy levels.

1.3 Chaos in disorder

We formulate our second main result: disorder chaos at zero temperature for the Hamiltonian H_N^h . This is interesting on its own, but is also crucial to the construction of the set $S_N(h)$ of Theorem 1. Generally, chaos in spin glasses studies the instability of the system subject to a small perturbation of certain external parameters such as temperature, disorder, or external field. It is a very old subject that has received a lot of attention in the physics community; see Rizzo [44] for an up-to-date survey. In recent years, intensive mathematical progress on chaos in disorder has been made for the mixed even p -spin models: At positive temperature, Chatterjee [20] and Chen [22, 24] considered the Ising spin case, while Chen-Hsieh-Hwang-Sheu [25] studied the spherical spin case. More recently, disorder chaos at zero temperature was obtained in the spherical model by Chen-Sen [26].

Our investigation aims to establish chaos in disorder for the Hamiltonian H_N^h at zero temperature. More precisely, we are interested in the behavior of the maximizer of L_N^h when a perturbation is applied to the disorder. Let $t \in [0, 1]$ be a coupling parameter. Consider two Hamiltonians

$$H_{N,t}^{1,h}(\sigma^1) := \sqrt{t}H_N(\sigma^1) + \sqrt{1-t}H_N^1(\sigma^1) + h \sum_{i=1}^N \sigma_i^1,$$

$$H_{N,t}^{2,h}(\sigma^2) := \sqrt{t}H_N(\sigma^2) + \sqrt{1-t}H_N^2(\sigma^2) + h \sum_{i=1}^N \sigma_i^2,$$

where H_N^1 and H_N^2 are two i.i.d. copies of H_N . One immediately sees that $H_{N,t}^1 = H_{N,t}^2$ if $t = 1$, while for $0 < t < 1$, there is a certain amount of decoupling between the two systems controlled by the parameter t .

We measure chaos in disorder by the cross overlap between the maximizers of $H_{N,t}^{1,h}$ and $H_{N,t}^{2,h}$. By cross overlap, we mean the overlap between spin configurations from different Hamiltonians. Let $\sigma_{t,h}^1$ and $\sigma_{t,h}^2$ be any maximizers of $H_{N,t}^{1,h}$ and $H_{N,t}^{2,h}$, respectively. Note that the Hamiltonians H_N, H_N^1, H_N^2 are summations of even p -spin interactions. If $h \neq 0$, then these maximizers are almost surely unique. Nonetheless, if $h = 0$, then the two Hamiltonians $H_{N,t}^{1,0}$ and $H_{N,t}^{2,0}$ are symmetric, i.e., $H_{N,t}^{j,0}(\sigma) = H_{N,t}^{j,0}(-\sigma)$ for $j = 1, 2$ and yield

$$\text{Argmax}_{\sigma \in \Sigma_N} H_{N,t}^{1,0}(\sigma) = \{\sigma_{t,0}^1, -\sigma_{t,0}^1\},$$

$$\text{Argmax}_{\sigma \in \Sigma_N} H_{N,t}^{2,0}(\sigma) = \{\sigma_{t,0}^2, -\sigma_{t,0}^2\},$$

from which one sees that the absolute value of the cross overlap between any maximizers of $H_{N,t}^{1,h}$ and $H_{N,t}^{2,h}$ is always equal to

$$|R(\sigma_{t,0}^1, \sigma_{t,0}^2)|.$$

Based on these, we measure the instability of the system with respect to disorder perturbation by $R(\sigma_{t,h}^1, \sigma_{t,h}^2)$ when $h \neq 0$ and by $|R(\sigma_{t,0}^1, \sigma_{t,0}^2)|$ when $h = 0$. It is clear that these two quantities are equal to one if $t = 1$. However, if $0 < t < 1$, Theorem 2 below shows that they are concentrated around a constant strictly less than one no matter how close t is to 1.

Theorem 2. *Let $t \in (0, 1)$ and $\varepsilon > 0$. The following two statements hold for any ξ defined in (1):*

(i) *If $h = 0$, there exists a constant $K > 0$ depending on t, ξ, ε such that*

$$\mathbb{P}(|R(\sigma_{t,0}^1, \sigma_{t,0}^2)| \geq \varepsilon) \leq K \exp\left(-\frac{N}{K}\right), \quad \forall N \geq 1.$$

(ii) *If $h \neq 0$, there exists a constant $K > 0$ depending on t, h, ξ, ε such that*

$$\mathbb{P}(|R(\sigma_{t,h}^1, \sigma_{t,h}^2) - q_{t,h}| \geq \varepsilon) \leq K \exp\left(-\frac{N}{K}\right), \quad \forall N \geq 1,$$

where the constant $q_{t,h}$ relies only on ξ, h, t and satisfies $q_{t,h} \in (0, q_h)$. Here, q_h is the same constant as the one in Theorem 1.

Theorem 2(i) means that as long as there is no external field, the maximizers of $H_{N,t}^{1,0}$ are orthogonal to the maximizers of $H_{N,t}^{2,0}$. In the case that $h \neq 0$, Theorem 2(ii) states that $\sigma_{t,h}^1$ and $\sigma_{t,h}^2$ preserve a strictly positive distance $\sqrt{2(1 - q_{t,h})}$ as N tends to infinity. The key point is that even when t approaches 1, this distance remains at least $\sqrt{2(1 - q_h)} > 0$. This indicates instability of the maximizers due to the perturbation of the disorder. Theorem 2 was previously obtained in the context of the spherical mixed even p -spin model by Chen-Sen [26]. However, as explained in Subsection 1.4, the proof of Theorem 2 is fundamentally different than that of the spherical case.

Finally, we derive two consequences of Theorem 2 on the fluctuation properties of the maximal energy:

Theorem 3. *For any ξ defined in (1), the following two limits hold.*

(i) *If $h = 0$, then*

$$\lim_{N \rightarrow \infty} N \text{Var}(L_N^h) = 0.$$

(ii) *If $h > 0$, then*

$$\lim_{N \rightarrow \infty} d_{TV}(\sqrt{N}(L_N^h - \mathbb{E}L_N^h), g) = 0,$$

where d_{TV} is the total variation distance and g is a centered Gaussian random variable with variance $\int_0^1 \xi(q_{t,h}) dt$.

The Poincaré inequality implies that for all $h \geq 0$, there exists some $C > 0$ such that $\text{Var}(L_N^h) \leq C/N$ for any $N \geq 1$. Theorem 3(i) states that if the external field vanishes, L_N^h is superconcentrated in the sense that it has a faster self-averaging rate than the Poincaré bound. We emphasize that this result applies to the SK model and it thus solves Open Problem 3.2 in Chatterjee [21]. We refer the readers to Chatterjee [19] for some earlier results on superconcentration, where quantitative bounds were obtained for certain choices of ξ , which exclude the SK model. While L_N^h exhibits superconcentration in the absence of external field, Theorem 3(ii) shows that L_N^h has a Gaussian fluctuation if the external field is present. In other words, the Poincaré inequality is sharp whenever $h > 0$. An analogue of Theorem 3 was established by Chen-Sen [26] for the spherical mixed p -spin model. Its derivation relies on an analogue of Theorem 2 for the spherical case and an application of Stein's method for normal approximation. The proof of Theorem 3 is practically identical to that of the spherical case in Chen-Sen [26].

1.4 Main ideas of proofs

Our proof of Theorem 1 is motivated by Chatterjee [20]. To construct multiple peaks, he first established chaos in disorder for the SK model at positive temperature and zero external field by deriving upper bounds for the moments of the cross overlap. These bounds depend on the temperature and the coupling parameter t and they asymptotically vanish as N tends to infinity when $t < 1$. He then used these bounds with an adjusted temperature to select logarithmically many orthogonal peaks. The same bounds were later employed in Ding-Eldan-Zhai [29] to obtain a polynomially many peaks. The construction of peaks in these works requires a careful adjustment of both the coupling parameter and the inverse temperature. Such a simultaneous adjustment is restrictive and makes it very difficult to derive an exponential number of orthogonal peaks.

To overcome this difficulty, we verify disorder chaos at zero temperature with exponential tail control in Theorem 2, instead of considering the moments of the cross overlap at positive temperature.

We explain the basic idea of establishing this in the case where $h = 0$ (the case of $h \neq 0$ is similar). We show that for any $0 < t, \varepsilon < 1$, there exist constants $\eta, K > 0$ depending only on ξ, t, ε such that with probability at least $1 - Ke^{-N/K}$,

$$\begin{aligned} & \frac{1}{N} \max_{(\sigma^1, \sigma^2) \in \Sigma_N^2: |R(\sigma^1, \sigma^2)| \geq \varepsilon} (H_{N,t}^{1,0}(\sigma^1) + H_{N,t}^{2,0}(\sigma^2)) \\ & \leq \frac{1}{N} \max_{\sigma^1 \in \Sigma_N} H_{N,t}^{1,0}(\sigma^1) + \frac{1}{N} \max_{\sigma^2 \in \Sigma_N} H_{N,t}^{2,0}(\sigma^2) - \eta \end{aligned} \quad (4)$$

for all $N \geq 1$. If this is valid, the probability of $|R(\sigma_{t,0}^1, \sigma_{t,0}^2)| \geq \varepsilon$ must be bounded by $Ke^{-N/K}$ due to the positivity of η and the optimality of $\sigma_{t,0}^1, \sigma_{t,0}^2$, which yields Theorem 2(i). Generally, proving (4) is a very challenging task for arbitrary Gaussian fields. To accomplish this in the current setting we rely on the fact that the two sides of (4) can be controlled by the Parisi formula and the Guerra-Talagrand bound at zero temperature. The analysis of these formula and bound is based on an analogous study of the mixed p -spin model at positive temperature in the framework of Chen [24] and the recent development of the Parisi formula at zero temperature by Auffinger-Chen [6]. Once chaos in disorder at zero temperature is established, the construction of nearly orthogonal peaks in Theorem 1 is similar to Chatterjee [20]. Taking into account the external field allows us to select exponentially many equidistant spin configurations at all energy levels of H_N .

As mentioned before, a version of Theorem 2 for the spherical mixed p -spin model appeared in Chen-Sen [26]. The approaches in both the Ising and spherical cases require a series of sophisticated techniques and results about the properties of Parisi's formulas at positive and zero temperatures as well as Guerra-Talagrand bounds. In the spherical case, these formulas and bounds admit explicit and simple expressions. They yield great simplifications in controlling the coupled maximal energy of $H_{N,t}^{1,h}$ and $H_{N,t}^{2,h}$. However, in the Ising case, the Parisi formulas and Guerra-Talagrand bounds are formulated in a more delicate way through a semi-linear parabolic differential equation and its two-dimensional extension, called the Parisi PDEs. The analysis in the present paper is generally more subtle than that in the spherical case.

1.5 Open questions

In view of Theorem 1, there are two closely related open questions of great interest. The first is to establish a version of Theorem 1 with an error term ε depending on N .

The second problem is to further explore the structure of the energy landscape of H_N . From the methodology presented in the rest of the paper, it seems possible that one can construct another set $S'_N(h) \subset \Sigma_N$ at the same energy level as that of $S_N(h)$ such that the spin configurations within $S'_N(h)$ are again equidistant to each other and also preserve a constant distance to all elements in $S_N(h)$. Besides, it also seems possible that for any $0 \leq h < h'$, one can construct $S_N(h)$ and $S_N(h')$ at different energy levels such that Theorem 1 holds for both energy levels and the distance between $S_N(h)$ and $S_N(h')$ is a fixed constant. These can be verified if one can establish results for chaos in external field h and chaos in mixture parameters $(c_p)_{p \geq 2}$, similar to chaos in disorder in Theorem 2.

1.6 Organization of the paper

The paper is organized as follows. Section 2 focuses on the Parisi formula for the maximal energy L_N^h and establishes the uniqueness of the Parisi measure. Furthermore, Section 2 proves the non-triviality of the Parisi measure and also derives some consistency equations for this measure. Section 3 presents the proof for Theorem 2 based on the results in Section 2. Section 4 verifies Proposition 1 and Theorem 1. The proof of Theorem 3 is omitted as it is identical to those of Theorems 4 and 5 in

[26]. Some technical results regarding the regularity of the Parisi PDEs at zero temperature are left to the appendix. The paper frequently uses ideas of [24] and skips some proofs of results whenever they clearly follow an identical argument of [24].

Acknowledgements. The authors are indebted to N. Krylov and M. Safonov for illuminating discussions on the regularity properties of the Parisi PDE. They thank S. Chatterjee, D. Panchenko, and the anonymous referees for a number of suggestions regarding the presentation of the paper. The research of W.-K. C. is partially supported by NSF grant DMS-16-42207 and Hong Kong research grants council GRF-14-302515. The research of M. H. and G. L. is partially supported by NSF grant DMS-14-18386.

2 Parisi formula

The Parisi formula for the free energy in the mixed even p -spin model was first verified in the celebrated work of Talagrand [49]. Later Panchenko [42] validated it for more general mixtures including odd p -spin interactions. More recently, Auffinger-Chen [6] extended Parisi's formula to the maximal energy of H_N^h . Their formulation is described below. Let \mathcal{U} be the collection of all functions γ on $[0, 1)$ induced by some measure μ , i.e., $\gamma(s) = \mu([0, s])$ and satisfying

$$\int_0^1 \gamma(s) ds < \infty.$$

We equip \mathcal{U} with the L^1 distance with respect to the Lebesgue measure on $[0, 1)$. For each $\gamma \in \mathcal{U}$, let Φ_γ be the weak solution of

$$\partial_s \Phi_\gamma(s, x) = -\frac{\xi''(s)}{2} (\partial_{xx} \Phi_\gamma(s, x) + \gamma(s) (\partial_x \Phi_\gamma(s, x))^2) \quad (5)$$

for $(s, x) \in [0, 1) \times \mathbb{R}$ with boundary condition $\Phi_\gamma(1, x) = |x|$. Here the existence of Φ_γ is assured by Proposition 2 below. We call Φ_γ the Parisi PDE solution. Define

$$\mathcal{P}_h(\gamma) = \Phi_\gamma(0, h) - \frac{1}{2} \int_0^1 \xi''(s) s \gamma(s) ds, \quad \gamma \in \mathcal{U}.$$

The Parisi formula [6, Theorem 1] asserts that the limiting maximal energy of H_N^h can be written as a variational problem,

$$M(h) = \lim_{N \rightarrow \infty} L_N^h = \inf_{\gamma \in \mathcal{U}} \mathcal{P}_h(\gamma), \text{ a.s.} \quad (6)$$

Note that from [6], \mathcal{P}_h is a continuous functional and the minimizer of the variational problem exists. As one will see in Subsection 2.1, this minimizer is indeed unique. It will be called the Parisi measure and denoted by γ_h throughout the remainder of the paper.

2.1 Uniqueness

Uniqueness of the minimizer of the Parisi formula for the free energy of the mixed p -spin model was proved in Auffinger-Chen [5]. Our main result here is an extension of [5] at zero temperature.

Theorem 4. *The Parisi formula (6) has a unique minimizer.*

The proof of Theorem 4 will be based on [5]. The aim is to show that the functional $\gamma \mapsto \Phi_\gamma$ is strictly convex on the space \mathcal{U} via the stochastic optimal control representation for Φ_γ (see Theorem 5 below). While the formula considered in [5] restricts to those γ with $\gamma(1-) \leq 1$ and the boundary condition of the Parisi PDE has nice regularities, the added difficulties in the current situation are that it is possible that $\gamma(1-) = \infty$ and the boundary condition has a nondifferentiable point at 0. The following proposition shows that we still have good regularity properties on the spatial derivatives of Φ_γ as long as the time variable is away from 1. This is enough for us to prove Theorem 4. Denote by $\mathcal{U}_d \subset \mathcal{U}$ the set of all γ 's induced by atomic measures and by \mathcal{U}_c the set of all γ 's that are induced by measures without atoms. Note that $\gamma \in \mathcal{U}_c$ means that γ is continuous on $[0, 1)$.

Proposition 2. *Let $\gamma \in \mathcal{U}$. The following statements hold:*

(i) *The weak solution Φ_γ exists and is unique. Moreover, it satisfies*

$$|\Phi_\gamma(s, x) - \Phi_{\gamma'}(s, x)| \leq \frac{3}{2} \int_0^1 \xi''(s) |\gamma(s) - \gamma'(s)| ds \quad (7)$$

for any $(s, x) \in [0, 1] \times \mathbb{R}$ and $\gamma, \gamma' \in \mathcal{U}$.

(ii) *For any $k \geq 1$, $\partial_x^k \Phi_\gamma$ exists and is continuous on $[0, 1) \times \mathbb{R}$. Furthermore, for any $s_1 \in (0, 1)$,*

$$\sup_{(s, x) \in [0, s_1] \times \mathbb{R}} |\partial_x^k \Phi_\gamma(s, x)| < F_k(\gamma(s_1)),$$

where F_k is a continuous function defined on $[0, \infty)$ independent of γ .

(iii) *In either \mathcal{U}_d or \mathcal{U}_c , there exists $(\gamma_n)_{n \geq 1}$ with weak limit γ such that for any $k \geq 0$,*

$$\lim_{n \rightarrow \infty} \sup_{(s, x) \in [0, s_1] \times [-M, M]} |\partial_x^k \Phi_{\gamma_n}(s, x) - \partial_x^k \Phi_\gamma(s, x)| = 0$$

for any $s_1 \in (0, 1)$ and $M > 0$.

We defer the proof of Proposition 2 to the appendix. In what follows, we recall the stochastic optimal control for Φ_γ from [6]. Let $W = (W(w))_{0 \leq w \leq 1}$ be a standard Brownian motion. For $0 \leq r < s \leq 1$, denote by $D[r, s]$ the collection of all progressively measurable processes u on $[r, s]$ with respect to the filtration generated by W and satisfying $\sup_{0 \leq s \leq 1} |u(s)| \leq 1$. We equip the space $D[r, s]$ with the metric

$$d_0(u, u') := \left(\mathbb{E} \int_r^s \mathbb{E} |u(w) - u'(w)|^2 dw \right)^{1/2}. \quad (8)$$

Let $\gamma \in \mathcal{U}$. For any $x \in \mathbb{R}$ and $u \in D[r, s]$, define

$$F^{r,s}(u, x) = \mathbb{E} [C^{r,s}(u, x) - L^{r,s}(u)],$$

where

$$C^{r,s}(u, x) = \Phi_\gamma \left(t, x + \int_r^s \gamma(w) \xi''(w) u(w) dw + \int_r^s \xi''(w)^{1/2} dW(w) \right),$$

$$L^{r,s}(u) = \frac{1}{2} \int_r^s \gamma(w) \xi''(w) u(w)^2 dw.$$

Theorem 5. Let $0 \leq r \leq s \leq 1$. For any $\gamma \in \mathcal{U}$,

$$\Phi_\gamma(r, x) = \max \{F^{r,s}(u, x) | u \in D[r, s]\}, \quad (9)$$

where the maximum in (9) is attained by

$$u_\gamma(w) = \partial_x \Phi_\gamma(w, X_\gamma(w)). \quad (10)$$

Here $(X_\gamma(w))_{r \leq w \leq s}$ is the strong solution to

$$\begin{aligned} dX_\gamma(w) &= \gamma(w)\xi''(w)\partial_x \Phi_\gamma(w, X_\gamma(w))dw + \xi''(w)^{1/2}dW(w), \\ X_\gamma(r) &= x. \end{aligned}$$

Theorem 5 is essentially taken from [6], where it was shown to be valid for $\gamma \in \mathcal{U}_c$ by a direct application of Itô's formula. For arbitrary $\gamma \in \mathcal{U}$, it remains true by using Proposition 2 combined with an approximation argument similar to the proof of [5, Theorem 3]. We omit the details here. Next, we show that Φ_γ is convex in γ , which will play an essential role in proving the strict convexity of $\gamma \mapsto \Phi_\gamma$.

Lemma 1. For $\gamma_0, \gamma_1 \in \mathcal{U}$, $x_0, x_1 \in \mathbb{R}$ and $\theta \in [0, 1]$, denote

$$\begin{aligned} \gamma_\theta &= (1 - \theta)\gamma_0 + \theta\gamma_1, \\ x_\theta &= (1 - \theta)x_0 + \theta x_1. \end{aligned} \quad (11)$$

Then

$$\Phi_{\gamma_\theta}(r, x_\theta) \leq (1 - \theta)\Phi_{\gamma_0}(r, x_0) + \theta\Phi_{\gamma_1}(r, x_1) \quad (12)$$

for any $r \in [0, 1]$.

Proof. Let $\gamma_0, \gamma_1 \in \mathcal{U}$, $x_0, x_1 \in \mathbb{R}$, $\theta \in [0, 1]$, $0 \leq r \leq s \leq 1$ and $u \in D[r, s]$. Let $a = 0, \theta$, or 1 . Denote by

$$F_a^{r,s}, C_a^{r,s}, L_a^{r,s},$$

the functionals defined in the variational formulas corresponding respectively to γ_a :

$$\Phi_{\gamma_a}(s, x_a) = \max \{F_\theta^{r,s}(u, x_a) | D[r, s]\}. \quad (13)$$

Let $0 \leq r \leq 1$ and take $s = 1$. Suppose that $u \in D[r, 1]$. Observe that

$$L_\theta^{r,1}(u) = (1 - \theta)L_0^{r,1}(u) + \theta L_1^{r,1}(u)$$

and from the convexity of $|x|$,

$$C_\theta^{r,1}(u, x_\theta) \leq (1 - \theta)C_0^{r,1}(u, x_0) + \theta C_1^{r,1}(u, x_1).$$

These together imply

$$F_\theta^{r,1}(u, x_\theta) \leq (1 - \theta)F_0^{r,1}(u, x_0) + \theta F_1^{r,1}(u, x_1).$$

Since this is true for any $u \in D[r, 1]$, the representation formula (13) gives (12). \square

In order to prove strict convexity, we need two results regarding the uniqueness as well as some properties of the optimizer in (9). The first result gives the uniqueness of the maximizer u_γ if $\gamma(r) > 0$.

Lemma 2 (Uniqueness). *Let $\gamma \in \mathcal{U}$ and $0 \leq r < s \leq 1$ with $\gamma(r) > 0$. If u attains the maximal value of the variational representation (9), then $u = u_\gamma$.*

Lemma 2 is proved using Proposition 2(ii) in an argument identical to that of [24, Lemma 5]. We will not reproduce the details here. Lemma 3 below gives the usual results that $\partial_x \Phi_\gamma(w, X_\gamma(w))$ is a martingale and $\partial_{xx} \Phi_\gamma(w, X_\gamma(w))$ is a semi-martingale.

Lemma 3. *Let $\gamma \in \mathcal{U}$, $0 \leq r \leq s < 1$, and $x \in \mathbb{R}$. For any $r \leq a \leq b \leq s$, we have*

$$\partial_x \Phi_\gamma(b, X_\gamma(b)) - \partial_x \Phi_\gamma(a, X_\gamma(a)) = \int_a^b \xi''(w)^{1/2} \partial_{xx} \Phi_\gamma(w, X_\gamma(w)) dW(w) \quad (14)$$

and

$$\begin{aligned} & \partial_{xx} \Phi_\gamma(b, X_\gamma(b)) - \partial_{xx} \Phi_\gamma(a, X_\gamma(a)) \\ &= - \int_a^b \gamma(w) \xi''(w) (\partial_{xx} \Phi_\gamma(w, X_\gamma(w)))^2 dr + \int_a^b \xi''(w)^{1/2} \partial_x^3 \Phi_\gamma(w, X_\gamma(w)) dW(w). \end{aligned} \quad (15)$$

If $\gamma \in \mathcal{U}_c$, Lemma 3 is easily verified by a standard application of Itô's formula as $\partial_t \partial_{xx} \Phi_\gamma$ is continuous on $[0, 1) \times \mathbb{R}$ in this case. For the general case, one can argue by an approximation procedure via Proposition 2(iii), similar to that of [5, Proposition 3] with some minor modifications. Again we will omit these details. Next, we state a crucial property of $\partial_x \Phi_\gamma$.

Lemma 4. *For any $\gamma \in \mathcal{U}$ and $s \in [0, 1)$, $\partial_x \Phi_\gamma(s, \cdot)$ is odd and strictly increasing.*

Proof. Since the boundary condition $|x|$ is even, it is easy to see that $\Phi_\gamma(s, \cdot)$ is also even and thus $\partial_x \Phi_\gamma(s, \cdot)$ is odd. From Lemma 1, we know that $\Phi_\gamma(s, \cdot)$ is convex. This implies that $\partial_x \Phi_\gamma(s, \cdot)$ is nondecreasing. To see that $\partial_x \Phi_\gamma(s, \cdot)$ is strictly increasing, for any two distinct $x_0, x_1 \in \mathbb{R}$ and $\theta \in (0, 1)$, let u_γ^θ be the maximizer and X_γ^θ be the SDE solution in the representation (9) for $\Phi_\gamma(s, x_\theta)$. Note that from Girsanov's theorem [35, Theorem 5.1], the distribution of

$$\Delta := \int_s^1 \xi'' \gamma u_\gamma^\theta dw + \int_s^1 \xi''^{1/2} dW$$

is Gaussian under some change of measure. Therefore, Δ is supported on \mathbb{R} , so

$$\mathbb{P}((x_0 + \Delta)(x_1 + \Delta) < 0) > 0.$$

Thus with positive probability,

$$\begin{aligned} |X_\gamma^\theta| &= |(1 - \theta)(x_0 + \Delta) + \theta(x_1 + \Delta)| \\ &< (1 - \theta)|x_0 + \Delta| + \theta|x_1 + \Delta|. \end{aligned}$$

Using this inequality, (9) implies

$$\Phi_\gamma(s, x_\theta) < (1 - \theta)\Phi_\gamma(s, x_0) + \theta\Phi_\gamma(s, x_1).$$

This gives the strict convexity of $\Phi_\gamma(s, \cdot)$. If $\partial_x \Phi_\gamma(s, x_0) = \partial_x \Phi_\gamma(s, x_1)$ for two distinct x_0 and x_1 , then $\Phi_\gamma(s, x) = cx + d$ for any x between x_0 and x_1 , where c, d are constants. However, this contradicts the strict convexity of $\Phi_\gamma(s, \cdot)$. □

Lemma 5 establishes the strict convexity of the PDE solution Φ_γ in γ .

Lemma 5. *Let $\gamma_0, \gamma_1 \in \mathcal{U}$, $x_0, x_1 \in \mathbb{R}$, and $\theta \in (0, 1)$. Recall the notations x_θ and γ_θ from (11). If $\gamma_0 \neq \gamma_1$, we have*

$$\Phi_{\gamma_\theta}(0, x_\theta) < (1 - \theta)\Phi_{\gamma_0}(0, x_0) + \theta\Phi_{\gamma_1}(0, x_1).$$

Proof. We adapt a slightly simplified version of the argument of [5, Theorem 4]. Suppose that $x_0, x_1 \in \mathbb{R}$, $\gamma_0 \neq \gamma_1$ and $\theta \in (0, 1)$. Using the right-continuity of γ_0, γ_1 , without loss of generality, we may assume that there exist some $0 < r < s < 1$ such that

$$\gamma_0 > \gamma_1 \text{ on } [r, s]. \quad (16)$$

First we claim that

$$\Phi_{\gamma_\theta}(r, x_\theta) < (1 - \theta)\Phi_{\gamma_0}(r, x_0) + \theta\Phi_{\gamma_1}(r, x_1). \quad (17)$$

Suppose, on the contrary, that equality holds. Recall the notations used in the proof of Lemma 1. Let $u_{\gamma_0}, u_{\gamma_\theta}, u_{\gamma_1}$ be the corresponding maximizers of (13) generated by (10). From Lemma 1,

$$F_\theta^{r,s}(u_{\gamma_\theta}, x_\theta) \leq (1 - \theta)F_0^{r,s}(u_{\gamma_0}, x_0) + \theta F_1^{r,s}(u_{\gamma_1}, x_1). \quad (18)$$

Note that

$$F_0^{r,s}(u_{\gamma_\theta}, x_0) \leq \Phi_{\gamma_0}(r, x_0), \quad F_1^{r,s}(u_{\gamma_\theta}, x_1) \leq \Phi_{\gamma_1}(r, x_1), \quad F_\theta^{r,s}(u_{\gamma_\theta}, x_\theta) = \Phi_{\gamma_\theta}(r, x_\theta).$$

Consequently, from (18) and the assumption that

$$\Phi_{\gamma_\theta}(r, x_\theta) = (1 - \theta)\Phi_{\gamma_0}(r, x_0) + \theta\Phi_{\gamma_1}(r, x_1),$$

we obtain

$$\begin{aligned} F_0^{r,s}(u_{\gamma_\theta}, x_0) &= \Phi_{\gamma_0}(r, x_0), \\ F_1^{r,s}(u_{\gamma_\theta}, x_1) &= \Phi_{\gamma_1}(r, x_1). \end{aligned}$$

In other words, u_{γ_θ} realizes the maximum of the representations for $\Phi_{\gamma_0}(r, x_0)$ and $\Phi_{\gamma_1}(r, x_1)$. Now from (16) and Lemma 2, we conclude $u_{\gamma_0} = u_{\gamma_\theta}$ with respect to the metric d_0 defined in (8). Since u_{γ_0} and u_{γ_θ} are continuous on $[r, s]$, we have

$$\partial_x \Phi_{\gamma_0}(w, X_{\gamma_0}(w)) = u_{\gamma_0}(w) = u_{\gamma_\theta}(w) = \partial_x \Phi_{\gamma_\theta}(w, X_{\gamma_\theta}(w)) \quad (19)$$

for all $r \leq w \leq s$, where $(X_{\gamma_0}(w))_{r \leq w \leq s}$ and $(X_{\gamma_\theta}(w))_{r \leq w \leq s}$ satisfy respectively,

$$\begin{aligned} dX_{\gamma_0}(w) &= \gamma_0(w)\xi''(w)\partial_x \Phi_{\gamma_0}(w, X_{\gamma_0}(w))dr + \xi''(w)^{1/2}dB(w), \\ X_{\gamma_0}(w) &= x_0, \\ dX_{\gamma_\theta}(w) &= \gamma_\theta(w)\xi''(w)\partial_x \Phi_{\gamma_\theta}(w, X_{\gamma_\theta}(w))dr + \xi''(w)^{1/2}dB(w), \\ X_{\gamma_\theta}(w) &= x_\theta. \end{aligned}$$

From (14), (19) and Itô's isometry, we obtain

$$\partial_{xx} \Phi_{\gamma_0}(w, X_{\gamma_0}(w)) = \partial_{xx} \Phi_{\gamma_\theta}(w, X_{\gamma_\theta}(w)), \quad \forall w \in [r, s]. \quad (20)$$

Combined with (15), this gives

$$\int_r^s \gamma_0(w) \xi''(w) \mathbb{E}(\partial_{xx} \Phi_{\gamma_0}(w, X_0(w)))^2 dw = \int_r^s \gamma_\theta(w) \xi''(w) \mathbb{E}(\partial_{xx} \Phi_{\gamma_\theta}(w, X_\theta(w)))^2 dw. \quad (21)$$

For any $r < w < s$, neither $\mathbb{E}(\partial_{xx} \Phi_{\gamma_0}(w, X_{\gamma_0}(w)))^2$ nor $\mathbb{E}(\partial_{xx} \Phi_{\gamma_\theta}(w, X_{\gamma_\theta}(w)))^2$ is equal to zero. In fact, from Girsanov's theorem [35, Theorem 5.1], the distribution of $X_{\gamma_0}(w)$ is Gaussian with some change of measure and therefore is supported on \mathbb{R} . If, for instance,

$$\mathbb{E}(\partial_{xx} \Phi_{\gamma_0}(w, X_{\gamma_0}(w)))^2 = 0,$$

then $\partial_{xx} \Phi_{\gamma_0}(w, \cdot) = 0$, which contradicts Lemma 4. Now, from (16) and (20),

$$\int_r^s \gamma_0(w) \xi''(w) \mathbb{E}(\partial_{xx} \Phi_{\gamma_0}(w, X_{\gamma_0}(w)))^2 dr > \int_r^s \gamma_\theta(w) \xi''(w) \mathbb{E}(\partial_{xx} \Phi_{\gamma_\theta}(w, X_{\gamma_\theta}(w)))^2 dr.$$

This contradicts (21), so the inequality (17) holds. This completes the claim (17).

Next, from the above claim, for any $y_0, y_1 \in \mathbb{R}$,

$$\Phi_{\gamma_\theta}(w, y_\theta) < (1 - \theta) \Phi_{\gamma_0}(w, y_0) + \theta \Phi_{\gamma_1}(w, y_1).$$

Therefore,

$$C_\theta^{0,s}(u_{\gamma_\theta}, x_\theta) < (1 - \theta) C_0^{0,s}(u_{\gamma_\theta}, x_0) + \theta C_1^{0,s}(u_{\gamma_\theta}, x_1).$$

This together with

$$L_\theta^{0,s}(u_{\gamma_\theta}) = (1 - \theta) L_0^{0,s}(u_{\gamma_\theta}) + \theta L_1^{0,s}(u_{\gamma_\theta})$$

gives

$$F_\theta^{0,s}(u_{\gamma_\theta}, x_\theta) < (1 - \theta) F_0^{0,s}(u_{\gamma_\theta}, x_0) + \theta F_1^{0,s}(u_{\gamma_\theta}, x_1).$$

Since u_{γ_θ} is the maximizer for $\Phi_{\gamma_\theta}(0, x_\theta)$, clearly

$$\Phi_{\gamma_\theta}(0, x_\theta) = F_\theta^{0,t}(u_{\gamma_\theta}, x_\theta) < (1 - \theta) \Phi_{\gamma_0}(0, x_0) + \theta \Phi_{\gamma_1}(0, x_1).$$

This completes our proof. □

Proof of Theorem 4. Note that $\int_0^1 s \xi''(s) \gamma(s) ds$ is linear in γ . From the strict convexity of Φ_γ in Lemma 5, Theorem 2.1 follows. □

2.2 Optimality

In this subsection, we establish some consequences of the optimality of the Parisi measure. These will be of great use when we turn to the proof of disorder chaos. Our first main results are Theorem 6 and Proposition 3 below.

Theorem 6. γ_h is not identically equal to zero.

Let μ_h be the measure induced by γ_h . From Theorem 6, it is clear that $\Omega_h := \text{supp} \mu_h$ is not empty. This fact will play an essential role throughout the rest of the paper. Next, we formulate some consistency conditions for γ_h .

Proposition 3. For any $s \in \Omega_h$, we have that

$$\mathbb{E}(\partial_x \Phi_{\gamma_h}(s, X_{\gamma_h}(s)))^2 = s, \quad (22)$$

$$\xi''(s) \mathbb{E}(\partial_{xx} \Phi_{\gamma_h}(s, X_{\gamma_h}(s)))^2 \leq 1, \quad (23)$$

where $X_{\gamma_h} = (X_{\gamma_h}(w))_{0 \leq w < 1}$ satisfies the following stochastic differential equation with initial condition $X_{\gamma_h}(0) = h$.

$$dX_{\gamma_h}(w) = \xi''(w) \gamma_h(w) \partial_x \Phi_{\gamma_h}(w, X_{\gamma_h}(w)) dw + \xi''(w)^{1/2} dW(w).$$

Recall from Subsection 2.1 that \mathcal{P}_h defines a strictly convex functional on \mathcal{U} . Our approach to proving Theorem 6 and Proposition 3 relies on the directional derivative of the Parisi functional derived in Proposition 4.

Proposition 4. Let $\gamma_0, \gamma \in \mathcal{U}$. For any $\theta \in [0, 1]$, define $\gamma_\theta = (1 - \theta)\gamma_0 + \theta\gamma$. Then

$$\left. \frac{d\mathcal{P}_h(\gamma_\theta)}{d\theta} \right|_{\theta=0} = \frac{1}{2} \int_0^1 \xi''(s) (\gamma(s) - \gamma_0(s)) (\mathbb{E}u_{\gamma_0}(s)^2 - s) ds,$$

where $d\mathcal{P}_h(\gamma_\theta)/d\lambda|_{\theta=0}$ is the right-derivative at 0. In addition, the following statements are equivalent:

(i) γ_0 is the Parisi measure,

(ii) $\left. \frac{d\mathcal{P}_h(\gamma_\theta)}{d\theta} \right|_{\theta=0} \geq 0$ for any $\gamma \in \mathcal{U}$.

An analogue of Proposition 4 was established in [24, Theorem 2] for the Parisi functional in the formulation of the Parisi formula for the limiting free energy. One may find that exactly the same argument also applies to prove Proposition 4, so we once again do not reproduce the proof here.

Proof of Theorem 6. We argue by contradiction. Assume that $\gamma_h \equiv 0$. Then the PDE (5) reduces to a heat equation, in which case, one can solve

$$\Phi_{\gamma_h}(s, x) = \mathbb{E}|x + \sqrt{\xi'(1) - \xi'(s)}Z|,$$

where Z is a standard normal random variable. Thus,

$$\partial_x \Phi_{\gamma_h}(s, x) = \mathbb{E} \text{sign}(x + \sqrt{\xi'(1) - \xi'(s)}Z)$$

and

$$\mathbb{E}u_{\gamma_h}(s)^2 = \mathbb{E}_{Z'} (\mathbb{E}_Z \text{sign}(h + \sqrt{\xi'(s)}Z' + \sqrt{\xi'(1) - \xi'(s)}Z))^2,$$

where Z' is standard normal independent of Z and \mathbb{E}_Z and $\mathbb{E}_{Z'}$ are the expectations with respect to Z and Z' , respectively. We now compute the derivative of $\mathbb{E}u_{\gamma_h}(s)^2$. To this end, define

$$\begin{aligned} f(s) &= \mathbb{E}_{Z'} g(s, h + \sqrt{s}Z')^2, \\ g(s, x) &= \mathbb{E}_Z \text{sign}(x + \sqrt{1-s}Z). \end{aligned}$$

Observe that

$$\begin{aligned} g(s, x) &= 1 - 2\mathbb{P}(x + \sqrt{1-s}Z < 0) \\ &= 1 - 2 \int_{-\infty}^0 \frac{1}{\sqrt{2\pi(1-s)}} \exp\left(-\frac{(z-x)^2}{2(1-s)}\right) dz \\ &= 1 - 2 \int_0^{\infty} \frac{1}{\sqrt{2\pi(1-s)}} \exp\left(-\frac{(z+x)^2}{2(1-s)}\right) dz, \end{aligned}$$

from which one can directly check that

$$\partial_s g = -\frac{1}{2}\partial_{xx}g, \quad \forall (s, x) \in [0, 1) \times \mathbb{R}.$$

Set $Z_s = h + \sqrt{s}Z'$. Then

$$\begin{aligned} f'(s) &= 2\mathbb{E}\partial_s g(s, Z_s)g(s, Z_s) + \frac{1}{\sqrt{s}}\mathbb{E}Z'\partial_x g(s, Z_s)g(s, Z_s) \\ &= 2\mathbb{E}\partial_s g(s, Z_s)g(s, Z_s) + \mathbb{E}\partial_{xx}g(s, Z_s)g(s, Z_s) + \mathbb{E}(\partial_x g(s, Z_s))^2 \\ &= \mathbb{E}(\partial_x g(s, Z_s))^2. \end{aligned}$$

To compute the last term, note that

$$\begin{aligned} \partial_x g(s, x) &= 2 \int_0^\infty \frac{(z+x)}{\sqrt{2\pi}(1-s)^{3/2}} \exp\left(-\frac{(z+x)^2}{2(1-s)}\right) dz \\ &= -\frac{2}{\sqrt{2\pi}(1-s)} \exp\left(-\frac{(z+x)^2}{2(1-s)}\right) \Big|_0^\infty \\ &= \frac{2}{\sqrt{2\pi}(1-s)} \exp\left(-\frac{x^2}{2(1-s)}\right). \end{aligned}$$

This gives

$$\mathbb{E}(\partial_x g(s, Z_s))^2 = \frac{2}{\pi(1-s)} \frac{1}{\sqrt{2\pi s}} \int_{-\infty}^\infty \exp\left(-\frac{z^2}{(1-s)} - \frac{(z-h)^2}{2s}\right) dz.$$

Here,

$$\begin{aligned} 2sz^2 + (1-s)(z-h)^2 &= 2sz^2 + (1-s)z^2 - 2h(1-s)z + (1-s)h^2 \\ &= (1+s)z^2 - 2h(1-s)z + (1-s)h^2 \\ &= (1+s)\left(z - \frac{(1-s)h}{1+s}\right)^2 - \frac{(1-s)^2}{1+s}h^2 + (1-s)h^2 \\ &= (1+s)\left(z - \frac{(1-s)h}{1+s}\right)^2 + \frac{2s(1-s)}{1+s}h^2. \end{aligned}$$

Therefore,

$$\begin{aligned} f'(s) &= \mathbb{E}(\partial_x g(s, Z_s))^2 \\ &= \frac{2}{\pi(1-s)} \frac{1}{\sqrt{2\pi s}} \int_{-\infty}^\infty \exp\left(-\frac{(1+s)}{2s(1-s)}\left(z - \frac{(1-s)h}{1+s}\right)^2\right) dz \exp\left(-\frac{h^2}{1+s}\right) \\ &= \frac{2}{\pi\sqrt{1-s^2}} \exp\left(-\frac{h^2}{1+s}\right). \end{aligned} \tag{24}$$

To compute the derivative of $\mathbb{E}u_{\gamma_h}(s)^2$, we write

$$\mathbb{E}u_{\gamma_h}(s)^2 = \mathbb{E}_{Z'}(\mathbb{E}_Z \text{sign}(h/\sqrt{\xi'(1)} + \sqrt{\xi'(s)/\xi'(1)}Z' + \sqrt{1-\xi'(s)/\xi'(1)}Z))^2.$$

Using (24) and the chain rule, we obtain

$$(\mathbb{E}u_{\gamma_h}(s)^2)' = \frac{2\xi''(s)}{\pi\sqrt{\xi'(1)^2 - \xi'(s)^2}} \exp\left(-\frac{h^2}{\xi'(1) + \xi'(s)}\right).$$

Finally, from this formula,

$$\begin{aligned}
\mathbb{E}u_{\gamma_h}(s)^2 - s &= (1 - \mathbb{E}u_{\gamma_h}(1)^2) - (s - \mathbb{E}u_{\gamma_h}(s)^2) \\
&= \int_s^1 (w - \mathbb{E}u_{\gamma_h}(w)^2)' dw \\
&= \int_s^1 1 - (\mathbb{E}u_{\gamma_h}(w)^2)' dw \\
&< 0,
\end{aligned}$$

whenever s is sufficiently close to 1. However, for any $\gamma \in \mathcal{U}$ with $\gamma = 0$ on $[0, s_1]$ for s_1 being arbitrarily close to 1, we have

$$\frac{d}{d\lambda} \mathcal{P}_h((1-\lambda)\gamma_h + \lambda\gamma) \Big|_{\lambda=0} = \frac{1}{2} \int_{s_1}^1 \xi''(s) \gamma(s) (\mathbb{E}u_{\gamma_h}(s)^2 - s) ds < 0,$$

which contradicts Proposition 4(ii). □

Proof of Proposition 3. Using the minimality of γ_h , one can vary γ in Proposition 4 to deduce (22), while the validity of (23) is assured by (22) and the equation (14). For detail, we refer the readers to [24, Proposition 1] for an identical proof. □

3 Establishing disorder chaos

Chaos in disorder at positive temperature has been intensively studied in recent years, see [20, 22, 25, 24]. One of the major approaches is to control the coupled free energy with overlap constraint via a two-dimensional version of the Guerra-Talagrand (GT) inequality. Lately, this method was pushed forward to show chaos in disorder at zero temperature in [26] for the spherical mixed even p -spin models. In this section, we utilize the ideas from [24, 26] to establish Theorem 2.

3.1 Guerra-Talagrand inequality

We begin by recalling the GT inequality for the coupled free energy in terms of Parisi's PDE from [24]. Let

$$V_N := \{R(\sigma^1, \sigma^2) : \sigma^1, \sigma^2 \in \Sigma_N\}. \quad (25)$$

Let $q \in V_N$ be fixed. For $\beta > 0$, define the coupled free energy as

$$F_{N,\beta}(q) := \frac{1}{N\beta} \mathbb{E} \log \sum_{R(\sigma^1, \sigma^2)=q} \exp \beta (H_{N,t}^1(\sigma^1) + H_{N,t}^2(\sigma^2)).$$

Denote $\iota = 1$ if $q \geq 0$ and $\iota = -1$ if $q < 0$. Define

$$T(s) = \begin{bmatrix} \xi''(s) & \iota t \xi''(s) \\ \iota t \xi''(s) & \xi''(s) \end{bmatrix}$$

for $s \in [0, |q|)$ and

$$T(s) = \begin{bmatrix} \xi''(s) & 0 \\ 0 & \xi''(s) \end{bmatrix} \quad (26)$$

for $s \in [|q|, 1]$. Let \mathcal{M} be the collection of all functions α on $[0, 1]$ induced by a probability measure μ , i.e., $\alpha(s) = \mu([0, s])$. For $\alpha \in \mathcal{M}$, consider the following PDE,

$$\partial_s \Psi_{\alpha, \beta} = -\frac{1}{2} (\langle T, \nabla^2 \Psi_{\alpha, \beta} \rangle + \beta \alpha \langle T \nabla \Psi_{\alpha, \beta}, \nabla \Psi_{\alpha, \beta} \rangle) \quad (27)$$

for $(\lambda, s, \mathbf{x}) \in \mathbb{R} \times [0, 1] \times \mathbb{R}^2$ with boundary condition

$$\Psi_{\alpha, \beta}(\lambda, 1, \mathbf{x}) = \frac{1}{\beta} \log(\cosh(\beta x_1) \cosh(\beta x_2) \cosh(\beta \lambda) + \sinh(\beta x_1) \sinh(\beta x_2) \sinh(\beta \lambda)). \quad (28)$$

The GT inequality established in [24, Theorem 6] states that for any $\beta > 0$ and $\alpha \in \mathcal{M}$,

$$F_{N, \beta}(q) \leq \frac{2 \log 2}{\beta} + \Psi_{\alpha, \beta}(\lambda, 0, h, h) - \lambda q - \left(\int_0^1 \beta \alpha(s) s \xi''(s) ds + t \int_0^{|q|} \beta \alpha(s) s \xi''(s) ds \right). \quad (29)$$

Note that the form here is slightly different than the one appeared in [24, Equation (45)] by a change of variable in β .

To use this bound in our case, we need to convert it into the form for the coupled maximal energy by sending the inverse temperature β to infinity. For any $\gamma \in \mathcal{U}$ with $\gamma(1-) < \infty$, letting $\alpha = \gamma/\beta$, from (27) and (28), we obtain that in the limit $\beta \rightarrow \infty$,

$$F_N(q) := \frac{1}{N} \mathbb{E} \max_{R(\sigma^1, \sigma^2)=q} (H_{N, t}^1(\sigma^1) + H_{N, t}^2(\sigma^2)) \leq \Lambda(\lambda, \gamma, q), \quad (30)$$

where

$$\Lambda(\lambda, \gamma, q) := \Psi_\gamma(\lambda, 0, h, h) - \lambda q - \left(\int_0^1 \gamma(s) s \xi''(s) ds + t \int_0^{|q|} \gamma(s) s \xi''(s) ds \right). \quad (31)$$

Let

$$g(\lambda, \mathbf{x}) := \max(x_1 + x_2 + \lambda, -x_1 - x_2 + \lambda, x_1 - x_2 - \lambda, -x_1 + x_2 - \lambda). \quad (32)$$

Here, Ψ_γ is the weak solution to the following equation,

$$\partial_s \Psi_\gamma = -\frac{1}{2} (\langle T, \nabla^2 \Psi_\gamma \rangle + \gamma \langle T \nabla \Psi_\gamma, \nabla \Psi_\gamma \rangle)$$

for $(\lambda, s, \mathbf{x}) \in \mathbb{R} \times [0, 1] \times \mathbb{R}^2$ with boundary condition

$$\Psi_\gamma(\lambda, 1, \mathbf{x}) = g(\lambda, \mathbf{x}), \quad (33)$$

where $\nabla^2 \Psi_\gamma$ is the Hessian and $\nabla \Psi_\gamma$ is the gradient of Ψ_γ in \mathbf{x} in the classical sense. With an approximation procedure, one may relax the condition $\gamma(1-) < \infty$ to obtain the validity of (30) for arbitrary $\gamma \in \mathcal{U}$. We comment that these results can be justified by a similar treatment as [6] and Proposition 2 with some minor adjustments. The fact that Ψ_γ is two-dimensional does not influence the argument in any essential way.

3.2 Determination of $q_{t,h}$

We first define the overlap constant q_h as the minimum of Ω_h . Recall that μ_h is the measure induced by γ_h and Ω_h is the support of μ_h . From Theorem 6, $\Omega_h \neq \emptyset$ and $q_h < 1$. From Proposition 3, letting $s = q_h$ gives

$$\mathbb{E}(\partial_x \Phi_{\gamma_h}(q_h, h + \chi))^2 = q_h, \quad (34)$$

$$\xi''(q_h) \mathbb{E}(\partial_{xx} \Phi_{\gamma_h}(q_h, h + \chi))^2 \leq 1 \quad (35)$$

for χ a centered Gaussian random variable with variance $\xi'(q_h)$. From the first equation and Lemma 4, $q_h > 0$ if $h \neq 0$. Define

$$\psi_t(s) = \mathbb{E} \partial_x \Phi_{\gamma_h}(q_h, h + \chi_1(s)) \partial_x \Phi_{\gamma_h}(q_h, h + \chi_2(s))$$

for $s \in [-q_h, q_h]$, where $\chi_1(s)$ and $\chi_2(s)$ are jointly Gaussian with mean zero and covariance

$$\begin{aligned} \mathbb{E}(\chi_1(s))^2 &= \mathbb{E}(\chi_2(s))^2 = \xi'(q_h), \\ \mathbb{E}\chi_1(s)\chi_2(s) &= t\xi'(s). \end{aligned}$$

Assume that $h = 0$. If $q_h > 0$, then ψ_1 has two fixed points 0 and q_h , which contradicts [23, Theorem 5]. Thus, $q_h = 0$ if $h = 0$.

The overlap constant $q_{t,h}$ of Theorem 2 is determined by the fixed point of ψ_t .

Proposition 5. *For any $t \in [0, 1]$, ψ_t maps $[-q_h, q_h]$ into itself and has a unique fixed point $q_{t,h}$, where $q_{t,h} = 0$ if $h = 0$ and $q_{t,h} > 0$ if $h > 0$. In addition, $t \mapsto q_{t,h}$ is continuous with $q_{1,h} = q_h$.*

Proof. By the Cauchy-Schwarz inequality and (34), it is easy to see that ψ_t maps $[-q_h, q_h]$ into itself. If $h = 0$, then ψ_t is defined only at the origin 0 and obviously 0 is the unique fixed point of ψ_t . For $h \neq 0$, we use (34), (35), and [23, Theorem 5] to obtain the existence and uniqueness of the fixed point of ψ_t . Moreover, observe that

$$\psi_t(0) = (\mathbb{E} \partial_x \Phi_{\gamma_h}(q_h, h + \chi))^2.$$

From Lemma 4, we see that $\psi_t(0) > 0$ and thus, $q_{t,h} > 0$. Finally, from (35) and the Cauchy-Schwarz inequality, using Gaussian integration by parts gives

$$\begin{aligned} \psi'_t(s) &= t\xi''(s) \mathbb{E} \partial_{xx} \Phi_{\gamma_h}(q_h, h + \chi_1(s)) \partial_{xx} \Phi_{\gamma_h}(q_h, h + \chi_2(s)) \\ &< \xi''(q_h) \mathbb{E}(\partial_{xx} \Phi_{\gamma_h}(q_h, h + \chi))^2 \\ &\leq 1 \end{aligned}$$

for $(s, t) \in (-q_h, q_h) \times (0, 1)$. From this, $(\psi_t(s) - s)' < 0$. By the implicit function theorem, there exists a continuous function $q(t)$ on $(0, 1)$ such that $\psi_t(q(t)) = q(t)$. From the uniqueness of $q_{t,h}$, it follows that $q_{t,h} = q(t)$ is continuous on $(0, 1)$. To obtain the continuity of $t \mapsto q_{t,h}$ at 0 and 1. Denote $a = \limsup_{t \rightarrow 1^-} q_{t,h}$. From the continuity of $\psi_t(s)$ in both variables (s, t) ,

$$\psi_1(a) = \limsup_{t \rightarrow 1^-} \psi_t(q_{t,h}) = \limsup_{t \rightarrow 1^-} q_{t,h} = a,$$

which implies that $a = q_{1,h}$ by the uniqueness of $q_{1,h}$. Similarly, one may argue that $\liminf_{t \rightarrow 1^-} q_{t,h} = q_{1,h}$. Thus, $t \mapsto q_{t,h}$ is continuous at 1. Similarly, we can conclude that this function is also continuous at 0. The equation $q_{1,h} = q_h$ is valid directly by the definition of ψ_1 . \square

3.3 Control of the GT bound

Recall the GT bound from (30). Although $F_N(q)$ is defined only on V_N , the functional $\Lambda(\lambda, \gamma, q)$ is indeed well-defined on $\mathbb{R} \times \mathcal{U} \times [-1, 1]$. This allows us to define

$$\Lambda(q) = \inf_{\lambda \in \mathbb{R}, \gamma \in \mathcal{U}} \Lambda(\lambda, \gamma, q)$$

for $q \in [-1, 1]$. The main goal of this subsection is to get the following controls for $\Lambda(q)$.

Proposition 6. *Let $t \in (0, 1)$. Recall the overlap constant $q_{t,h}$ from Proposition 5.*

- (i) *For any $q \in [-q_h, q_h] \setminus \{q_{t,h}\}$, $\Lambda(q) < 2M(h)$,*
- (ii) *For any $q \notin [-q_h, q_h]$, we have $\Lambda(q) < 2M(h)$.*

The remainder of the subsection is devoted to establishing Proposition 6. First of all, we adapt the stochastic optimal control representation for the PDE Ψ_γ in a similar manner as Theorem 5. Denote by $\mathbf{W} = \{\mathbf{W}(w) = (W_1(w), W_2(w)), \mathcal{G}_w, 0 \leq w \leq 1\}$ a two-dimensional Brownian motion, where $(\mathcal{G}_w)_{0 \leq w \leq 1}$ satisfies the usual condition (see Definition 2.25 [35]). For $0 \leq r < s \leq 1$, denote by $\mathcal{D}[r, s]$ the space of all two-dimensional progressively measurable processes $v = (v_1, v_2)$ with respect to $(\mathcal{G}_w)_{r \leq w \leq s}$ satisfying $\sup_{r \leq w \leq s} |v_1(w)| \leq 1$ and $\sup_{r \leq w \leq s} |v_2(w)| \leq 1$. Endow the space $\mathcal{D}[s, t]$ with the norm

$$\|v\|_{r,s} = \left(\mathbb{E} \int_r^s (v_1(w)^2 + v_2(w)^2) dw \right)^{1/2}. \quad (36)$$

Define a functional

$$\mathcal{F}_\gamma^{r,s}(\lambda, v, \mathbf{x}) = \mathbb{E} [\mathcal{C}_\gamma^{r,s}(\lambda, v, \mathbf{x}) - \mathcal{L}_\gamma^{r,s}(v)]$$

for $(\lambda, v, \mathbf{x}) \in \mathbb{R} \times \mathcal{D}[r, s] \times \mathbb{R}^2$, where

$$\begin{aligned} \mathcal{C}_\gamma^{r,s}(\lambda, v, \mathbf{x}) &:= \Psi_\gamma \left(\lambda, s, \mathbf{x} + \int_r^s \alpha_\gamma(w) T(w) v(w) dw + \int_r^s T(w)^{1/2} d\mathbf{W}(w) \right), \\ \mathcal{L}_\gamma^{r,s}(v) &:= \frac{1}{2} \int_r^s \alpha_\gamma(w) \langle T(w) v(w), v(w) \rangle dw. \end{aligned}$$

The following is an analogue of Theorem 5.

Theorem 7. *For any $\gamma \in \mathcal{U}$,*

$$\Psi_\gamma(\lambda, s, \mathbf{x}) = \max \{ \mathcal{F}_\gamma^{r,s}(\lambda, v, \mathbf{x}) \mid v \in \mathcal{D}[r, s] \}. \quad (37)$$

The maximum of (37) is attained by $v_\gamma(r) = \nabla \Psi_\gamma(\lambda, r, \mathbf{X}_\gamma(r))$, where the two-dimensional stochastic process $(\mathbf{X}_\gamma(w))_{r \leq w \leq s}$ satisfies

$$\begin{aligned} d\mathbf{X}_\gamma(w) &= \gamma(w) T(w) \nabla \Psi_\gamma(\lambda, w, \mathbf{X}_\gamma(w)) dw + T(w)^{1/2} d\mathbf{W}(w), \\ \mathbf{X}_\gamma(r) &= \mathbf{x}. \end{aligned}$$

The proof of Theorem 7 is a standard computation of Itô's formula. We invite the readers to check either [6, Theorem 2] or [24, Theorem 5]. The following lemma listed some key properties of Ψ_γ when $\lambda = 0$. It will be used in the proof of Proposition 6(i).

Lemma 6. Assume $|q| < 1$. Let $\gamma \in \mathcal{U}$. For any $r \in [|q|, 1)$ and $\mathbf{x} \in \mathbb{R}^2$,

$$\Psi_\gamma(0, r, \mathbf{x}) = \Phi_\gamma(r, x_1) + \Phi_\gamma(r, x_2), \quad (38)$$

and $\Psi_\gamma(\lambda, r, \mathbf{x})$ is partially differentiable in λ in the classical sense with

$$\partial_\lambda \Psi_\gamma(0, r, \mathbf{x}) = \partial_x \Phi_\gamma(r, x_1) \partial_x \Phi_\gamma(r, x_2). \quad (39)$$

The proof of Lemma 6 utilizes the variational representations for $\Phi_\gamma(r, x_1)$, $\Phi_\gamma(r, x_2)$, and $\Psi_\gamma(0, r, \mathbf{x})$ introduced above. While the validity of (38) follows directly from these representations, the proof of the partial differentiability of $\Psi_\gamma(\lambda, r, \mathbf{x})$ in λ is relatively difficult and depends on the following technical lemma concerning about the right differentiability of some functions defined in terms of maximization problems.

Lemma 7 ([24, Lemma 2]). Let K be a metric space and I be an interval with right open edge. Let f be a real-valued function on $K \times I$ and $f_0(y) = \sup_{a \in K} f(a, y)$. Suppose that there exists a K -valued continuous function $a(y)$ on I such that $f_0(y) = f(a(y), y)$ and $\partial_y f$ is continuous on $K \times I$, then f_0 is right-differentiable with derivative $\partial_y f(a(y), y)$ for all $y \in I$.

Proof of Lemma 6. Assume that $|q| < 1$ and $\gamma \in \mathcal{U}$ is fixed. Let $r \in [|q|, 1]$ and $s = 1$ be fixed. Recall the function g from (32). From Theorem 7 associated to the pair (r, s) ,

$$\Psi_\gamma(\lambda, r, \mathbf{x}) = \max \{ \mathcal{F}_\gamma^{r,1}(\lambda, v, \mathbf{x}) \mid v \in \mathcal{D}[r, 1] \}. \quad (40)$$

From (26) and (33),

$$\Psi_\gamma(\lambda, r, \mathbf{x}) = \max_{v \in \mathcal{D}[r,1]} \mathbb{E} \left[g \left(\lambda, \mathbf{x} + \int_r^1 \xi'' \gamma v dw + \int_r^1 \sqrt{\xi''} d\mathbf{W} \right) - \frac{1}{2} \int_r^1 \xi'' \gamma \langle v, v \rangle dw \right].$$

Since $g(0, \mathbf{y}) = |y_1| + |y_2|$, the foregoing equation becomes

$$\begin{aligned} \Psi_\gamma(\lambda, r, \mathbf{x}) &= \max_{v=(v_1, v_2) \in \mathcal{D}[r,1]} \left\{ \mathbb{E} \left[\left| x_1 + \int_r^1 \xi'' \gamma v_1 dw + \int_r^1 \sqrt{\xi''} dW_1 \right| - \frac{1}{2} \int_r^1 \xi'' \gamma v_1^2 dw \right] \right. \\ &\quad \left. + \mathbb{E} \left[\left| x_2 + \int_r^1 \xi'' \gamma v_2 dw + \int_r^1 \sqrt{\xi''} dW_2 \right| - \frac{1}{2} \int_r^1 \xi'' \gamma v_2^2 dw \right] \right\} \\ &= \Phi_\gamma(r, x_1) + \Phi_\gamma(r, x_2) \end{aligned}$$

by applying Theorem 5 with the same pair (r, s) . This gives (38).

To establish the partial differentiability of Ψ_γ in λ , observe that g is Lipschitz in λ and it can be written as

$$g(\lambda, \mathbf{y}) = \max(|y_1 + y_2| + \lambda, |y_1 - y_2| - \lambda).$$

Note that

$$\partial_\lambda g(\lambda, \mathbf{y}) = \text{sign}(2\lambda + |y_1 + y_2| - |y_1 - y_2|) = 2I_{\{2\lambda + |y_1 + y_2| - |y_1 - y_2| > 0\}} - 1 \text{ a.e.},$$

where $\text{sign}(a) := 1$ if $a \geq 0$ and $:= -1$ if $a < 0$ and $I_{\{2\lambda + |y_1 + y_2| - |y_1 - y_2| > 0\}}$ is the indication function on $\{2\lambda + |y_1 + y_2| - |y_1 - y_2| > 0\}$. By the dominated convergence theorem, for any $v \in \mathcal{D}[r, 1]$,

$$\partial_\lambda \mathcal{F}_\gamma^{r,1}(\lambda, v, \mathbf{x}) = 2\mathbb{P} \left(2\lambda + |Y_{\gamma,1}^v + Y_{\gamma,2}^v| > |Y_{\gamma,1}^v - Y_{\gamma,2}^v| \right) - 1,$$

where $\mathbf{Y}_\gamma^v = (Y_{\gamma,1}^v, Y_{\gamma,2}^v)$ is defined by

$$\mathbf{Y}_\gamma^v = \mathbf{x} + \int_r^1 \gamma \xi'' v dw + \int_r^1 \sqrt{\xi''} d\mathbf{W}.$$

Here, observe that the equality

$$2\lambda + |Y_{\gamma,1}^v + Y_{\gamma,2}^v| = |Y_{\gamma,1}^v - Y_{\gamma,2}^v|$$

implies that $Y_{\gamma,1}^v$ or $Y_{\gamma,2}^v$ must be equal to λ or $-\lambda$, which happens with zero probability. Consequently,

$$\mathbb{P}\left(2\lambda + |Y_{\gamma,1}^v + Y_{\gamma,2}^v| = |Y_{\gamma,1}^v - Y_{\gamma,2}^v|\right) = 0.$$

This guarantees that $\partial_\lambda \mathcal{F}_\gamma^{r,1}(\lambda, v, \mathbf{x})$ is continuous in (λ, v) . Next, we recall that from Theorem 7, the optimizer of (40) is given by $v_\gamma^{r,\lambda,\mathbf{x}}(w) = \nabla \Psi_\gamma(\lambda, w, \mathbf{X}_\gamma^{r,\lambda,\mathbf{x}}(w))$, where $\mathbf{X}_\gamma^{r,\lambda,\mathbf{x}}$ satisfies

$$\begin{aligned} d\mathbf{X}_\gamma^{r,\lambda,\mathbf{x}}(w) &= \gamma(w)\xi''(w)\nabla \Psi_\gamma(\lambda, w, \mathbf{X}_\gamma^{r,\lambda,\mathbf{x}}(w))dw + \sqrt{\xi''(w)}d\mathbf{W}(w), \quad r \leq w \leq 1, \\ \mathbf{X}_\gamma^{r,\lambda,\mathbf{x}}(r) &= \mathbf{x}. \end{aligned}$$

Since g is uniformly Lipschitz, we can apply the same argument as Proposition 2 to obtain that $\nabla \Psi_\gamma(\lambda, s, \mathbf{x})$ is Lipschitz in $(\lambda, \mathbf{x}) \in \mathbb{R}^3$ uniformly over all $s \in [0, s_0]$ for any $0 < s_0 < 1$. As a result, an application of the Gronwall inequality yields (see, e.g., the argument of [5, Theorem 3]) that for any $s_0 \in (0, 1)$, there exists K_0 such that with probability 1,

$$\sup_{w \in [0, s_0]} |\mathbf{X}_\gamma^{r,\lambda,\mathbf{x}}(w) - \mathbf{X}_\gamma^{r,\lambda',\mathbf{x}}(w)| \leq K_0 |\lambda - \lambda'|, \quad \forall \lambda, \lambda' \in \mathbb{R}.$$

Therefore, the process $\lambda \mapsto v_\gamma^{r,\lambda,\mathbf{x}}$ is continuous, i.e., $\lim_{\lambda' \rightarrow \lambda} \|v_\gamma^{r,\lambda',\mathbf{x}} - v_\gamma^{r,\lambda,\mathbf{x}}\|_{r,1} = 0$, where $\|\cdot\|_{r,1}$ is defined through (36). This together with the continuity of $\partial_\lambda \mathcal{F}_\gamma^{r,1}(\lambda, v, \mathbf{x})$ in (λ, v) established above implies that $\Psi_\gamma(\lambda, r, \mathbf{x})$ is right partially differentiable at all $\lambda \in \mathbb{R}$ by Lemma 7. Furthermore, this right derivative is equal to $\partial_\lambda \mathcal{F}_\gamma^{r,1}(\lambda, v_\gamma^{r,\lambda,\mathbf{x}}, \mathbf{x})$ and is continuous in λ . It is a well-known result (see, e.g., [17]) that if a function has a continuous right derivative on an open interval U , then it is differentiable on U . From these, we obtain the partial differentiability of $\Psi_\gamma(\lambda, r, \mathbf{x})$ in λ .

To see how (39) is obtained, we note that when $\lambda = 0$, $g(0, \mathbf{y}) = |y_1| + |y_2|$, which implies $\nabla \Psi_\gamma(0, w, \mathbf{y}) = (\partial_x \Phi_\gamma(w, y_1), \partial_x \Phi_\gamma(w, y_2))$. Thus, the stochastic processes $X_{\gamma,1}^{r,0,\mathbf{x}}$ and $X_{\gamma,2}^{r,0,\mathbf{x}}$ are exactly the minimizers of the variational representations (9) for $\Phi_\gamma(r, x_1)$ and $\Phi_\gamma(r, x_2)$ with respect to independent standard Brownian motions W_1 and W_2 , respectively. Thus,

$$\begin{aligned} \partial_\lambda \Psi_\gamma(0, r, \mathbf{x}) &= \mathbb{E} \text{sign}(X_{\gamma,1}^{r,0,\mathbf{x}}) \cdot \mathbb{E} \text{sign}(X_{\gamma,2}^{r,0,\mathbf{x}}) \\ &= \partial_x \Phi_\gamma(r, x_1) \partial_x \Phi_\gamma(r, x_2), \end{aligned}$$

where the last equality is obtained by noting that $\partial_x \Phi_\gamma(r, x_j) = \mathbb{E} \text{sign}(X_{\gamma,j}^{r,0,\mathbf{x}})$, which can be derived by the above argument. This completes our proof. \square

Proof of Proposition 6(i). If $h = 0$, from Subsection 3.2 and Proposition 5, we see that $q_h = 0$, so $[-q_h, q_h] \setminus \{q_{t,h}\} = \emptyset$ and (i) is vacuously valid. Assume now $h \neq 0$ and $q \in [-q_h, q_h] \setminus \{q_{t,h}\}$. Recall

the Gaussian random variables $\chi, \chi_1(s), \chi_2(s)$ and the function ψ_t introduced in Subsection 3.2. Note that $\gamma_h = 0$ on $[0, q_h]$ and $q_h < 1$. From (38),

$$\begin{aligned}\Psi_{\gamma_h}(0, 0, h, h) &= \mathbb{E}\Psi_{\gamma_h}(0, q_h, h + \chi_1(q), h + \chi_2(q)) \\ &= \mathbb{E}\Phi_{\gamma_h}(q_h, h + \chi_1(q)) + \mathbb{E}\Phi_{\gamma_h}(q_h, h + \chi_2(q)) \\ &= 2\mathbb{E}\Phi_{\gamma_h}(q_h, h + \chi) \\ &= 2\Phi_{\gamma_h}(0, h).\end{aligned}$$

This implies $\Lambda(0, \gamma_h, q) = 2\mathcal{P}_h(\gamma_h) = 2M(h)$ because

$$\int_0^{|q|} \gamma_h(s) s \xi''(s) ds = 0$$

in Λ (see (31)) for $|q| \leq q_h$. On the other hand, from (39),

$$\begin{aligned}\partial_\lambda \Psi_{\gamma_h}(0, 0, h, h) &= \mathbb{E}\partial_\lambda \Psi_{\gamma_h}(0, q_h, h + \chi_1(q), h + \chi_2(q)) \\ &= \mathbb{E}\partial_x \Phi_{\gamma_h}(q_h, h + \chi_1(q)) \partial_x \Phi_{\gamma_h}(q_h, h + \chi_2(q)) \\ &= \psi_t(q).\end{aligned}$$

Therefore,

$$\partial_\lambda \Lambda(0, \gamma_h, q) = \psi_t(q) - q.$$

Since ψ_t has a unique fixed point $q_{t,h}$ by Proposition 5, depending on the sign of $\psi_t(q) - q$, we may vary λ slightly to obtain $\Lambda(q) \leq \Lambda(\lambda, \gamma_h, q) < \Lambda(0, \gamma_h, q) = 2M(h)$. \square

To show Proposition 6(ii), we need the following crucial lemma, which allows us to quantify the difference between the one-dimensional and two-dimensional Parisi PDEs in an elementary way:

Proposition 7. *Assume that $|q| > q_h$. Define $\gamma_t \in \mathcal{U}$ by*

$$\gamma_t = \begin{cases} \frac{\gamma_h}{1+t}, & \text{on } [0, |q|], \\ \gamma_h, & \text{on } [|q|, 1]. \end{cases}$$

The following two statements hold:

- (i) *If $v_{\gamma_t} = (v_1, v_2)$ is the maximizer to the variational problem (37) for $\Psi_{\gamma_t}(0, 0, \mathbf{x})$ using $(r, s) = (0, |q|)$, then*

$$\Psi_{\gamma_t}(0, 0, \mathbf{x}) \leq \Phi_{\gamma_h}(0, x_1) + \Phi_{\gamma_h}(0, x_2) - \frac{t(1-t)}{2(1+t)^2} \int_0^{|q|} \gamma_h \xi'' \mathbb{E}(v_1 - v_2)^2 dw. \quad (41)$$

- (ii) *Define*

$$(u_1(w), u_2(w)) = \frac{1}{1+t} T(w) v_{\gamma_t}(w), \quad (42)$$

$$(B_1(r), B_2(r)) = \frac{1}{\xi''(w)^{1/2}} T(w)^{1/2} \mathbf{W}(w) \quad (43)$$

for $0 \leq r \leq |q|$. If

$$\Psi_{\gamma_t}(0, 0, \mathbf{x}) = \Phi_{\gamma_h}(0, x_1) + \Phi_{\gamma_h}(0, x_2),$$

then u_1 and u_2 are the maximizers of the variational problems (9) for $\Phi_{\gamma_h}(0, x_1)$ and $\Phi_{\gamma_h}(0, x_2)$ using $(r, s) = (0, |q|)$ with respect to the standard Brownian motions B_1 and B_2 . Moreover, on the interval $[q_h, |q|]$,

$$\begin{aligned} u_1(w) &= \partial_x \Phi_{\gamma_h}(w, X_{1, \gamma_h}(w)), \\ u_2(w) &= \partial_x \Phi_{\gamma_h}(w, X_{2, \gamma_h}(w)), \end{aligned}$$

where $(X_{1, \gamma_h}(w))_{0 \leq w \leq |q|}$ and $(X_{2, \gamma_h}(w))_{0 \leq w \leq |q|}$ satisfy

$$\begin{aligned} dX_{1, \gamma_h}(w) &= \gamma_h(w) \xi''(w) \partial_x \Phi_{\gamma_h}(w, X_{1, \gamma_h}(w)) dw + \xi''(w)^{1/2} dB_1(w), \\ dX_{2, \gamma_h}(w) &= \gamma_h(w) \xi''(w) \partial_x \Phi_{\gamma_h}(w, X_{2, \gamma_h}(w)) dw + \xi''(w)^{1/2} dB_2(w) \end{aligned}$$

with initial condition $X_{\gamma_h}(0) = x_1$ and $X_{\gamma_h}(0) = x_2$.

Theorem 7 and Proposition 7 are analogous to the results of Theorem 5 and Proposition 5 in [24], which were used to control the PDE solution $\Psi_{\alpha, \beta}$ in the GT bound (29). Although the space \mathcal{U} is different from \mathcal{M} in [24], the same proofs there carry through in our case. We do not reproduce the proofs here.

Proof of Proposition 6(ii). Let $q \notin [-q_h, q_h]$. Observe that

$$\Lambda(q) \leq \Lambda(0, \gamma_t, q) = \Psi_{\gamma_t}(0, 0, h, h) - \int_0^1 w \xi''(w) \gamma_h(w) dw.$$

It suffices to show that $\Psi_{\gamma_t}(0, 0, h, h) < 2\Phi_{\gamma_h}(0, h)$. If, on the contrary, the two sides are the same, then Proposition 7(ii) readily implies

$$\begin{aligned} u_1(w) &= \partial_x \Phi_{\gamma_h}(w, X_{1, \gamma_h}(w)), \\ u_2(w) &= \partial_x \Phi_{\gamma_h}(w, X_{2, \gamma_h}(w)), \end{aligned}$$

for $w \in [q_h, |q|]$, where $X_{1, \gamma_h} = (X_{1, \gamma_h}(w))_{0 \leq w \leq |q|}$ and $X_{2, \gamma_h} = (X_{2, \gamma_h}(w))_{0 \leq w \leq |q|}$ satisfy

$$\begin{aligned} X_{1, \gamma_h}(w) &= h + \int_0^w \gamma_h(a) \xi''(a) \partial_x \Phi_{\gamma_h}(a, X_{1, \gamma_h}(a)) da + \int_0^w \xi''(a)^{1/2} dB_1(a), \\ X_{2, \gamma_h}(w) &= h + \int_0^w \gamma_P(a) \xi''(a) \partial_x \Phi_{\gamma_h}(a, X_{2, \gamma_h}(a)) da + \int_0^w \xi''(a)^{1/2} dB_2(a). \end{aligned} \tag{44}$$

On the other hand, from (41), $v_1 = \iota v_2$ on $[q_h, |q|]$, which together with (42) leads to $u_1 = \iota u_2$ on $[q_h, |q|]$. Now since $\partial_x \Phi_{\gamma_h}(w, \cdot)$ is a strictly increasing odd function by Lemma 4,

$$X_{1, \gamma_h}(w) = \iota X_{2, \gamma_h}(w), \quad \forall w \in [q_h, |q|].$$

Consequently, from (44),

$$0 = X_{1, \gamma_h}(w) - \iota X_{2, \gamma_h}(w) = (1 - \iota)h + \int_0^w \xi''(a)^{1/2} d(B_1 - \iota B_2)(a), \quad \forall w \in [q_h, |q|].$$

This forces $B_1 = \iota B_2$. However, from (43), $\iota w = \mathbb{E}B_1(w)B_2(w) = \iota t w$ for $w \in [q_h, |q|]$, a contradiction since $t \in (0, 1)$. This finishes our proof. \square

3.4 Proof of Theorem 2

Before we start the proof, we recall some concentration inequalities for the extreme values of Gaussian processes. They will be used several times in this subsection and Section 4. Recall the Hamiltonians, H_N^h , $H_{N,t}^{1,h}$, and $H_{N,t}^{2,h}$ from Section 1. Observe that if $h = 0$, for any nonempty $A \subset \Sigma_N$ and $\hat{A} \subset \Sigma_N \times \Sigma_N$,

$$\max_{\sigma \in A} \mathbb{E} \left(\frac{H_N^h(\sigma)}{N} \right)^2 = \frac{\xi(1)}{N}$$

and

$$\max_{(\sigma^1, \sigma^2) \in \hat{A}} \mathbb{E} \left(\frac{H_{N,t}^{1,h}(\sigma^1) + H_{N,t}^{2,h}(\sigma^2)}{N} \right)^2 = \frac{2}{N} \max_{(\sigma^1, \sigma^2) \in \hat{A}} (\xi(1) + t\xi(R(\sigma^1, \sigma^2))) \leq \frac{4\xi(1)}{N}.$$

From the Gaussian concentration of measure (see [9]), the random variables

$$\begin{aligned} L_N^h(A) &:= \max_{\sigma \in A} \frac{H_N^h(\sigma)}{N}, \\ \hat{L}_{N,t}^h(\hat{A}) &:= \max_{R(\sigma^1, \sigma^2) \in \hat{A}} \frac{H_{N,t}^{1,h}(\sigma^1) + H_{N,t}^{2,h}(\sigma^2)}{N} \end{aligned}$$

satisfy

$$\mathbb{P}(|L_N^h(A) - \mathbb{E}L_N^h(A)| \geq l) \leq 2 \exp\left(-\frac{Nl^2}{4\xi(1)}\right). \quad (45)$$

and

$$\mathbb{P}(|\hat{L}_{N,t}^h(\hat{A}) - \mathbb{E}\hat{L}_{N,t}^h(\hat{A})| \geq l) \leq 2 \exp\left(-\frac{Nl^2}{8\xi(1)}\right). \quad (46)$$

In addition, we note a well-known upper bound for the Gaussian extremal process (see, e.g., [9]) that, for $h = 0$, gives

$$\mathbb{P}|L_N^0(A)| \leq \sqrt{2\xi''(1) \log 2}. \quad (47)$$

A crucial feature of these inequalities is that the upper bounds are valid independent of the choice of A , \hat{A} and external field h .

Proof of Theorem 2. To prove Theorem 2, let $t \in (0, 1)$ be fixed. Recall V_N from (25). For any $\varepsilon > 0$, let $I := [-1, 1] \setminus (q_{t,h} - \varepsilon, q_{t,h} + \varepsilon)$ and $I_N := V_N \cap I$. Define

$$\hat{A}_N = \{(\sigma^1, \sigma^2) \in \Sigma_N^2 : R(\sigma^1, \sigma^2) \in I_N\}$$

and for $q \in V_N$,

$$\hat{A}_{N,q} = \{(\sigma^1, \sigma^2) \in \Sigma_N^2 : R(\sigma^1, \sigma^2) = q\}.$$

Write

$$\mathbb{E}\hat{L}_{N,t}^h(\hat{A}_N) = \mathbb{E} \max_{q \in I_N} (\hat{L}_{N,t}^h(\hat{A}_{N,q}) - \mathbb{E}\hat{L}_{N,t}^h(\hat{A}_{N,q}) + \mathbb{E}\hat{L}_{N,t}^h(\hat{A}_{N,q})).$$

Since I_N contains at most $2N + 1$ elements, this and the uniform concentration (46) together imply

$$\begin{aligned} \limsup_{N \rightarrow \infty} \mathbb{E} \hat{L}_{N,t}^h(\hat{A}_N) &\leq \limsup_{N \rightarrow \infty} \max_{q \in I_N} \mathbb{E} \hat{L}_{N,t}^h(\hat{A}_{N,q}) \\ &\leq \lim_{N \rightarrow \infty} \max_{q \in I_N} \Lambda(q) \\ &\leq \max_{q \in S} \Lambda(q). \end{aligned}$$

Note that one may use the variational representation (37) for Ψ_γ to obtain the continuity of $\Lambda(\lambda, \gamma, q)$ (see [24, Theorem 4]), which implies that $\Lambda(q)$ is upper semi-continuous on $[-1, 1]$. Since $\Lambda(q) < 2M(h)$ for all $q \in [-1, 1] \setminus \{q_{t,h}\}$ by Proposition 6, $\max_{q \in S} \Lambda(q) < 2M(h)$ and thus,

$$\limsup_{N \rightarrow \infty} \frac{1}{N} \mathbb{E} \max_{R(\sigma^1, \sigma^2) \in I_N} (H_{N,t}^{1,h}(\sigma^1) + H_{N,t}^{2,h}(\sigma^2)) = \limsup_{N \rightarrow \infty} \mathbb{E} \hat{L}_{N,t}^h(\hat{A}_N) < 2M(h).$$

Using (46) again, there exists a constant $K > 0$ such that the following event holds with probability at least $1 - K \exp(-N/K)$,

$$\frac{1}{N} \max_{R(\sigma^1, \sigma^2) \in I_N} (H_{N,t}^{1,h}(\sigma^1) + H_{N,t}^{2,h}(\sigma^2)) < \frac{1}{N} \left(\max_{\sigma^1 \in \Sigma_N} H_{N,t}^{1,h}(\sigma^1) + \max_{\sigma^2 \in \Sigma_N} H_{N,t}^{2,h}(\sigma^2) \right). \quad (48)$$

Note that

$$H_{N,t}^{1,h}(\sigma_{t,h}^1) = \max_{\sigma \in \Sigma_N} H_{N,t}^{1,h}(\sigma), \quad H_{N,t}^{2,h}(\sigma_{t,h}^2) = \max_{\sigma \in \Sigma_N} H_{N,t}^{2,h}(\sigma).$$

Assume that $h = 0$. Since $q_{t,0} = 0$, the set I_N is symmetric with respect to the origin. As a result, if $|R(\sigma_{t,h}^1, \sigma_{t,h}^2)| \in I_N$, then $R(\sigma_{t,h}^1, \sigma_{t,h}^2) \in I_N$, which violates (48). This means that the event $|R(\sigma_{t,h}^1, \sigma_{t,h}^2)| \in I_N$ has probability at most $K \exp(-N/K)$, which completes the proof of Theorem 2(i). Similarly, if $h \neq 0$ and $R(\sigma_{t,h}^1, \sigma_{t,h}^2) \in I_N$, then this again violates (48) and we conclude that the event $R(\sigma_{t,h}^1, \sigma_{t,h}^2) \in I_N$ has probability at most $K \exp(-N/K)$. The property that $q_{t,h} \in (0, q_h)$ follows directly from Proposition 5. This finishes the proof of Theorem 2(ii). \square

4 Establishing exponential number of multiple peaks

This section establishes the main result on the energy landscape of H_N . Subsection 4.1 proves Proposition 1 and Subsection 4.2 verifies Theorem 1 by using chaos in disorder.

4.1 Proof of Proposition 1

We will prove Proposition 1. We first establish some concentration inequalities for the Hamiltonian H_N and the magnetization m_N when they are evaluated at the maximizer of L_N^h . Recall $M(h)$ from (2).

Proposition 8. *M is continuously differentiable on \mathbb{R} with*

$$M'(h) = \partial_x \Phi_{\gamma_h}(0, h).$$

In particular, $M'(0) = 0$.

Proof. This proof is exactly the same as that of [40, Theorem 1]. Let $\delta > 0$. Using the Parisi formula (6),

$$\frac{M(h + \delta) - M(h)}{\delta} \leq \frac{\Phi_{\gamma_h}(0, h + \delta) - \Phi_{\gamma_h}(0, h)}{\delta}$$

and

$$\frac{M(h) - M(h - \delta)}{\delta} \geq \frac{\Phi_{\gamma_h}(0, h) - \Phi_{\gamma_h}(0, h - \delta)}{\delta}.$$

Applying Proposition 2(iii) yields

$$\limsup_{\delta \downarrow 0} \frac{M(h + \delta) - M(h)}{\delta} \leq \partial_x \Phi_{\gamma_h}(0, h)$$

and

$$\liminf_{\delta \downarrow 0} \frac{M(h) - M(h - \delta)}{\delta} \geq \partial_x \Phi_{\gamma_h}(0, h),$$

which completes the proof of the differentiability of M . Using Lemma 4 gives $M'(0) = 0$. Finally, since M is convex and differentiable on \mathbb{R} , it is a classical result in convex analysis (see, e.g., [45, Theorem 25.3]) that these two assumptions automatically imply the continuity of M' and thus, M is continuously differentiable. This finishes our proof. \square

Let $\sigma_h := \text{Argmax}_{\sigma \in \Sigma_N} H_N^h(\sigma)$. In what follows, we show that the magnetization of σ_h is concentrated around a constant.

Proposition 9. *For any $\varepsilon > 0$, there exists some $K > 0$ such that*

$$\mathbb{P}\left(\left|m_N(\sigma_h) - M'(h)\right| \geq \varepsilon\right) \leq K \exp\left(-\frac{N}{K}\right)$$

for any $N \geq 1$.

Proof. Let $\varepsilon > 0$. Define

$$B_N^+ = \max_{m_N(\sigma) > M'(h) + \varepsilon} \frac{H_N^h(\sigma)}{N},$$

$$B_N^- = \max_{m_N(\sigma) < M'(h) - \varepsilon} \frac{H_N^h(\sigma)}{N}.$$

Then for any $\lambda > 0$,

$$\begin{aligned} \mathbb{E}B_N^+ &\leq \mathbb{E} \max_{m_N(\sigma) > M'(h) + \varepsilon} \left(\frac{H_N(\sigma)}{N} + (h + \lambda)m_N(\sigma) - \lambda M'(h) \right) - \lambda \varepsilon \\ &\leq \mathbb{E}L_N^{h+\lambda} - \lambda(M'(h) + \varepsilon). \end{aligned}$$

and

$$\begin{aligned} \mathbb{E}B_N^- &\leq \mathbb{E} \max_{m_N(\sigma) < M'(h) - \varepsilon} \left(\frac{H_N(\sigma)}{N} + (h - \lambda)m_N(\sigma) + \lambda M'(h) \right) - \lambda \varepsilon \\ &\leq \mathbb{E}L_N^{h-\lambda} - \lambda(\varepsilon - M'(h)). \end{aligned}$$

Passing to the limit,

$$\limsup_{N \rightarrow \infty} \mathbb{E} B_N^\pm \leq M(h \pm \lambda) - \lambda(\varepsilon \pm M'(h)). \quad (49)$$

By Proposition 8, the λ -derivative at zero on the right-hand side of (49) is equal to $-\varepsilon$. Thus, we can choose λ small enough so that

$$\limsup_{N \rightarrow \infty} \mathbb{E} B_N^\pm \leq M(h) - \varepsilon'.$$

for some $\varepsilon' > 0$. From the Gaussian concentration of measure (45), there exists a constant $K > 0$ such that as long as N is big enough, the following holds with probability at least $1 - K \exp(-N/K)$,

$$B_N^\pm \leq L_N(h) - \frac{\varepsilon'}{2}.$$

Therefore, the probability that σ_h satisfies $|m_N(\sigma_h) - M'(h)| > \varepsilon$ is at most $K \exp(-N/K)$, which finishes the proof. \square

Recall the function $E(h)$ from (3). Next, we show that $H_N(\sigma_h)/N$ is essentially near the energy level $E(h)$.

Proposition 10. *Let $h \geq 0$. For any $\varepsilon > 0$, there exists some $K > 0$ such that*

$$\mathbb{P}\left(\left|\frac{H_N(\sigma_h)}{N} - E(h)\right| \geq \varepsilon\right) \leq K \exp\left(-\frac{N}{K}\right)$$

for all $N \geq 1$.

Proof. From the definition of σ_h , we write

$$\begin{aligned} \frac{H_N(\sigma_h)}{N} - E(h) &= \left(L_N^h - \mathbb{E}L_N^h\right) \\ &\quad + \left(\mathbb{E}L_N^h - M(h)\right) \\ &\quad + h(M'(h) - m_N(\sigma_h)). \end{aligned}$$

Now the first term on the right-hand side can be controlled by applying the Gaussian concentration of measure to L_N^h . The second term can be handled by noting that $M(h) = \lim_{N \rightarrow \infty} \mathbb{E}L_N^h$. Control of the last term follows from Proposition 9. \square

Next we develop an auxiliary lemma that will be used to show that $\lim_{h \rightarrow \infty} E(h) = 0$ in Proposition 1.

Lemma 8. *If $m_N(\sigma^1) = k_1/N$ and $m_N(\sigma^2) = k_2/N$, then*

$$R(\sigma^1, \sigma^2) \geq \frac{1}{4} \left(\frac{3k_1}{N} + \frac{3k_2}{N} - 2 \right).$$

Proof. Set

$$\begin{aligned} P_1 &= \{1 \leq i \leq N : \sigma_i^1 = 1\}, \\ P_2 &= \{1 \leq i \leq N : \sigma_i^2 = 1\}. \end{aligned}$$

Note that $|P_1^c| = (N - k_1)/2$ and $|P_2^c| = (N - k_2)/2$ and that

$$\begin{aligned} |P_1 \cap P_2| &= \frac{1}{2}(|P_1| + |P_2| - |P_1 \cap P_2^c| - |P_2 \cap P_1^c|) \\ &\geq \frac{1}{2}(|P_1| + |P_2| - |P_2^c| - |P_1^c|) \\ &= \frac{1}{2}(k_1 + k_2). \end{aligned}$$

Lemma 8 then follows by

$$\begin{aligned} NR(\sigma^1, \sigma^2) &= |P_1 \cap P_2| + |P_1^c \cap P_2^c| - |P_1 \cap P_2^c| - |P_1^c \cap P_2| \\ &\geq |P_1 \cap P_2| - |P_2^c| - |P_1^c| \\ &\geq \frac{1}{2}\left(k_1 + k_2 - \frac{(N - k_1)}{2} - \frac{(N - k_2)}{2}\right) \\ &= \frac{1}{4}(3k_1 + 3k_2 - 2N). \end{aligned}$$

□

Lemma 9. $\lim_{h \rightarrow \infty} E(h) = 0$.

Proof. The argument contains three major steps. For any $\varepsilon > 0$, define the set

$$A_N(h, \varepsilon) := \{\sigma \in \Sigma_N : |m_N(\sigma) - M'(h)| < \varepsilon\}.$$

First, we verify that $\lim_{h \rightarrow \infty} M'(h) = 1$. Observe that using (45) and (47) with $h = 0$ and $A = \Sigma_N$, we have that, with probability at least $1 - Ke^{-KN}$,

$$-C \leq \max_{\sigma \in \Sigma_N} \frac{H_N(\sigma)}{N} < C,$$

where $C, K > 0$ are constants independent of N . This inequality implies that, with probability at least $1 - Ke^{-KN}$,

$$-\frac{C}{h} + \max_{\sigma \in \Sigma_N} m_N(\sigma) \leq \max_{\sigma \in \Sigma_N} \left(\frac{H_N(\sigma)}{Nh} + m_N(\sigma) \right) \leq \frac{C}{h} + \max_{\sigma \in \Sigma_N} m_N(\sigma).$$

Therefore, from the definition of $M(h)$ and the Borel-Cantelli lemma,

$$\lim_{h \rightarrow \infty} \frac{M(h)}{h} = \lim_{N \rightarrow \infty} \max_{\sigma \in \Sigma_N} m_N(\sigma) = 1. \quad (50)$$

On the other hand, from Proposition 9, a similar reasoning also gives that for any $\varepsilon > 0$,

$$\begin{aligned} \lim_{h \rightarrow \infty} \frac{M(h)}{h} &= \lim_{h \rightarrow \infty} \lim_{N \rightarrow \infty} \max_{A_N(h, \varepsilon)} \left(\frac{H_N(\sigma)}{Nh} + m_N(\sigma) \right) \\ &= \lim_{h \rightarrow \infty} \lim_{N \rightarrow \infty} \max_{A_N(h, \varepsilon)} m_N(\sigma). \end{aligned} \quad (51)$$

Note that $M'(h)$ is a nondecreasing continuous function. If $\lim_{h \rightarrow \infty} M'(h) \neq 1$, then one can choose ε small enough such that the above limits (50) and (51) contradict each other. Therefore, we conclude that $\lim_{h \rightarrow \infty} M'(h) = 1$.

Next, consider an auxiliary free energy,

$$V_N(s, h) = \frac{1}{N\beta_h} \mathbb{E} \log \sum_{A_N(h, \varepsilon_h)} \exp(\beta_h \sqrt{s} H_N(\sigma)),$$

where ε_h is any positive function of h satisfying $\lim_{h \rightarrow \infty} \varepsilon_h = 0$ and

$$\beta_h := (1 - M'(h) + \varepsilon_h)^{-1/2}.$$

Using Gaussian integration by parts, we get

$$\begin{aligned} \partial_s V_N(s, h) &= \frac{\beta_h}{2} \mathbb{E} \langle (\xi(1) - \xi(R(\sigma^1, \sigma^2))) \rangle' \\ &\leq \frac{\beta_h}{2} \max \{ \xi(1) - \xi(R(\sigma^1, \sigma^2)) : \sigma^1, \sigma^2 \in A_N(h, \varepsilon_h) \} \\ &\leq \frac{\beta_h \xi''(1)}{2} \left(1 - \frac{1}{2} (3M'(h) - 3\varepsilon_h - 1) \right) \\ &= \frac{3\xi''(1)}{4} (1 - M'(h) + \varepsilon_h)^{1/2}. \end{aligned}$$

Here, σ^1 and σ^2 in the first equality are two i.i.d. samplings from the Gibbs expectation with respect to the partition function,

$$\sum_{A_N(h, \varepsilon_h)} \exp(\beta_h \sqrt{s} H_N(\sigma))$$

and $\langle \cdot \rangle'$ denotes the corresponding Gibbs average. The second inequality also used Lemma 8. In conclusion, we obtain that

$$\begin{aligned} |V_N(1, h) - V_N(0, h)| &\leq \max_{0 \leq s \leq 1} \partial_s V_N(s, h) \\ &\leq \frac{3\xi''(1)}{4} (1 - M'(h) + \varepsilon_h)^{1/2}. \end{aligned} \tag{52}$$

The two terms on the left-hand side have the limits given in equations (53) and (54). Note that $\lim_{h \rightarrow \infty} \beta_h = \infty$. Using the evident inequalities

$$\mathbb{E} \max_{A_N(h, \varepsilon_h)} \frac{H_N(\sigma)}{N} \leq V_N(1, h) \leq \frac{\log 2}{\beta_h} + \mathbb{E} \max_{A_N(h, \varepsilon_h)} \frac{H_N(\sigma)}{N},$$

we deduce that

$$\limsup_{h \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E} \max_{A_N(h, \varepsilon_h)} \frac{H_N(\sigma)}{N} = \limsup_{h \rightarrow \infty} \limsup_{N \rightarrow \infty} V_N(1, h). \tag{53}$$

In addition, since

$$0 < V_N(0, h) = \frac{1}{N\beta_h} \log |A_N(h, \varepsilon_h)| \leq \frac{\log 2}{\beta_h},$$

we also have

$$\limsup_{h \rightarrow \infty} \limsup_{N \rightarrow \infty} |V_N(0, h)| = 0. \tag{54}$$

As a result, plugging (53) and (54) into (52) yields

$$\limsup_{h \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E} \max_{A_N(h, \varepsilon_h)} \frac{H_N(\sigma)}{N} = 0. \quad (55)$$

Finally, observe that from Proposition 9, for any $\varepsilon, h > 0$

$$M(h) = \lim_{N \rightarrow \infty} \mathbb{E} \max_{A_N(h, \varepsilon)} \left(\frac{H_N(\sigma)}{N} + hm_N(\sigma) \right).$$

This implies that

$$|E(h)| \leq \varepsilon h + \lim_{N \rightarrow \infty} \mathbb{E} \max_{A_N(h, \varepsilon)} \frac{H_N(\sigma)}{N}.$$

Since (55) holds for any ε_h with $\lim_{h \rightarrow \infty} \varepsilon_h = 0$, letting $\varepsilon = \varepsilon_h = 1/h^2$ in the last inequality and sending $h \rightarrow \infty$, we finish the proof. \square

Proof of Proposition 1. Clearly $E(0) = M(0)$. From Proposition 10 and Lemma 9, E is continuous and satisfies $\lim_{h \rightarrow \infty} E(h) = 0$. To prove monotonicity, for any $h' > h \geq 0$, write

$$E(h') - E(h) = (M(h') - M(h) - (h' - h)M'(h')) + h(M'(h) - M'(h')).$$

Since M is convex, the first bracket on the right-hand side is not positive and neither is the other term. Therefore, E is nonincreasing. \square

4.2 Proof of Theorem 1

Recall that q_h is the smallest number in the support of the Parisi measure μ_h . Also, recall the constant $q_{t,h} \leq q_h$ from Proposition 5.

Proof of Theorem 1. Fix $h \geq 0$. For a given $\varepsilon > 0$, fix $t \in (0, 1)$ such that

$$((1 - \sqrt{t}) + \sqrt{1 - t})(\sqrt{2\xi''(1) \log 2} + 1) < \frac{\varepsilon}{2} \quad (56)$$

and

$$0 \leq q_{t,h} - q_h \leq \frac{\varepsilon}{2}. \quad (57)$$

Let $n \geq 2$ be an integer, which will be specified later. Consider n i.i.d. copies H_N^1, \dots, H_N^n of H_N . Set

$$H_{N,t}^{\ell,h}(\sigma^\ell) = \sqrt{t}H_N(\sigma^\ell) + \sqrt{1-t}H_N^\ell(\sigma^\ell) + h \sum_{i=1}^N \sigma_i^\ell.$$

Denote by $\sigma_{t,h}^1, \dots, \sigma_{t,h}^n$ the maximizers of $H_{N,t}^{1,h}, \dots, H_{N,t}^{n,h}$ over the configuration space Σ_N , respectively. We need the following control:

1. From Theorem 2, there exists some $K_1 > 0$ independent of n such that for any $1 \leq \ell < \ell' \leq n$, the following inequality holds,

$$\mathbb{P}\left(|R(\sigma_{t,h}^\ell, \sigma_{t,h}^{\ell'}) - q_{t,h}| \geq \frac{\varepsilon}{2}\right) \leq K_1 \exp\left(-\frac{N}{K_1}\right),$$

from which

$$\begin{aligned} & \mathbb{P}\left(\max_{1 \leq \ell < \ell' \leq n} |R(\sigma_{t,h}^\ell, \sigma_{t,h}^{\ell'}) - q_{t,h}| \geq \frac{\varepsilon}{2}\right) \\ & \leq \sum_{1 \leq \ell < \ell' \leq n} \mathbb{P}\left(|R(\sigma_{t,h}^\ell, \sigma_{t,h}^{\ell'}) - q_{t,h}| \geq \frac{\varepsilon}{2}\right) \\ & \leq n^2 K_1 \exp\left(-\frac{N}{K_1}\right). \end{aligned} \quad (58)$$

2. From Proposition 10, there exists $K_2 > 0$ independent of n such that

$$\mathbb{P}\left(\max_{1 \leq \ell \leq n} \left|\frac{H_{N,t}^{\ell,0}(\sigma_{t,h}^\ell)}{N} - E(h)\right| \geq \frac{\varepsilon}{2}\right) \leq n K_2 \exp\left(-\frac{N}{K_2}\right). \quad (59)$$

3. From Proposition 9, there exists $K_3 > 0$ independent of n such that

$$\mathbb{P}\left(\max_{1 \leq \ell \leq n} |m_N(\sigma_{t,h}^\ell) - M'(h)| \geq \varepsilon\right) \leq n K_3 \exp\left(-\frac{N}{K_3}\right). \quad (60)$$

4. Note that

$$\left|\frac{H_{N,t}^{\ell,0}(\sigma_{t,h}^\ell)}{N} - \frac{H_N(\sigma_{t,h}^\ell)}{N}\right| \leq (1 - \sqrt{t}) \frac{|L_N|}{N} + \sqrt{1-t} \frac{|L_N^\ell|}{N},$$

where $L_N := N^{-1} \max_{\sigma \in \Sigma_N} H_N(\sigma)$ and $L_N^\ell := N^{-1} \max_{\sigma \in \Sigma_N} H_N^\ell(\sigma)$. Here, from (47), for any $N \geq 1$,

$$\mathbb{E}|L_N| \leq \sqrt{2\xi''(1) \log 2}.$$

Using the Gaussian concentration of measure (45) for $N^{-1}L_N, N^{-1}L_N^1, \dots, N^{-1}L_N^n$ and the above inequality, there exists some $K_4 > 0$ independent of n such that

$$\mathbb{P}\left(\frac{1}{N} \max(|L_N|, |L_N^1|, \dots, |L_N^n|) \geq \sqrt{2\xi''(1) \log 2} + 1\right) \leq (n+1) K_4 \exp\left(-\frac{N}{K_4}\right). \quad (61)$$

From above, we see that (58), (59), (60), and (61) are valid with probability at least

$$1 - n^2 K_1 \exp\left(-\frac{N}{K_1}\right) - n K_2 \exp\left(-\frac{N}{K_2}\right) - n K_3 \exp\left(-\frac{N}{K_3}\right) - (n+1) K_4 \exp\left(-\frac{N}{K_4}\right).$$

Consequently, the following three statements hold with probability at least $1 - n^2 K_5 e^{-N/K_5}$ for $K_5 := K_1 + K_2 + K_3 + K_4$. First, from (59), (61), and then (56), for $1 \leq \ell \leq n$,

$$\left|\frac{H_N(\sigma_{t,h}^\ell)}{N} - E(h)\right| \leq ((1 - \sqrt{t}) + \sqrt{1-t}) (\sqrt{2\xi''(1) \log 2} + 1) + \frac{\varepsilon}{2} < \varepsilon.$$

Second, for $1 \leq \ell \leq n$,

$$|m_N(\sigma_{t,h}^\ell) - M'(h)| < \varepsilon.$$

Finally, from (57) and (58), for $1 \leq \ell < \ell' \leq n$,

$$|R(\sigma_{t,h}^\ell, \sigma_{t,h}^{\ell'}) - q_h| \leq \frac{\varepsilon}{2} + |q_{t,h} - q_h| < \varepsilon.$$

Note that the constants K_1, K_2, K_3, K_4 are independent of n . Now take n to be the largest integer less than $e^{N/3K_5}$ and set $S_N(h) = \{\sigma_{t,h}^1, \dots, \sigma_{t,h}^n\}$. Then the announced statement in Theorem 1 holds with $K = 3K_5$. □

Appendix

This appendix is devoted to establishing Proposition 2. We prove a priori estimate first:

Lemma 10. *Let $0 < r_0 < r_1 < r_2 < \infty$. Suppose that $\kappa_1, \kappa_2 \in L^\infty([r_0, r_2] \times \mathbb{R})$ and $g \in L^\infty(\mathbb{R})$ with $\|\kappa_i\|_\infty \leq C_i$ for $i = 1, 2$ and $\|g\|_\infty \leq C_0$. Assume that u is the classical solution to*

$$\partial_r u(r, x) = \partial_{xx} u(r, x) + \kappa_1(r, x) \partial_x u(r, x) + \kappa_2(r, x), \quad \forall (r, x) \in (r_0, r_2] \times \mathbb{R}$$

with initial condition $u(r_0, x) = g(x)$. If

$$r \mapsto \|\partial_x u(r, \cdot)\|_\infty$$

is continuous on $[r_0, r_2]$, then there exists a nonnegative continuous function F on $[0, \infty)^3$ depending only on r_0, r_2 such that

$$\sup_{(r,x) \in [r_1, r_2] \times \mathbb{R}} |\partial_x u(r, x)| \leq F(C_0, C_1, C_2).$$

Proof. From the Duhamel principle,

$$u(r, x) = P_{r-r_0} g(x) + \int_{r_0}^r P_{r-w} (\kappa_1(w, \cdot) \partial_x u(w, \cdot) + \kappa_2(w, \cdot))(x) dw,$$

where for any $f \in L^\infty(\mathbb{R})$,

$$P_a f(x) := \frac{1}{\sqrt{4\pi a}} \int_{\mathbb{R}} f(y) e^{-\frac{(x-y)^2}{4a}} dy.$$

A direct computation leads to

$$\partial_x u(r, x) = \partial_x (P_{r-r_0} g(x)) + \int_{r_0}^r \partial_x (P_{r-w} (\kappa_1(w, \cdot) \partial_x u(w, \cdot) + \kappa_2(w, \cdot))(x)) dw.$$

Note that

$$\partial_x (P_{r-r_0} g(x)) = \frac{-1}{2(r-r_0)\sqrt{4\pi(r-r_0)}} \int_{\mathbb{R}} (x-y)g(y) e^{-\frac{(x-y)^2}{4(r-r_0)}} dy,$$

from which

$$|\partial_x (P_{r-r_0} g(x))| \leq \frac{C_0}{2(r-r_0)\sqrt{4\pi(r-r_0)}} \int_{\mathbb{R}} |x-y| e^{-\frac{(x-y)^2}{4(r-r_0)}} dy = \frac{C_0}{\sqrt{\pi(r-r_0)}}.$$

A similar computation also yields that

$$\int_{r_0}^r \partial_x (P_{r-w}(\kappa_1(w, \cdot) \partial_x u(w, \cdot) + \kappa_2(w, \cdot))(x)) dw \leq \int_{r_0}^r \frac{C_1 \|\partial_x u(w, \cdot)\|_\infty + C_2}{\sqrt{\pi(r-w)}} dw.$$

As a result,

$$\|\partial_x u(r, \cdot)\|_\infty \leq \phi_0(r) + C_1 \int_{r_0}^r \frac{\|\partial_x u(w, \cdot)\|_\infty}{\sqrt{\pi(r-w)}} dw,$$

where

$$\phi_0(r) := \frac{C_0}{\sqrt{\pi(r-r_0)}} + \frac{2C_2 \sqrt{r-r_0}}{\sqrt{\pi}}.$$

From the Gronwall inequality,

$$\begin{aligned} \|\partial_x u(r, \cdot)\|_\infty &\leq \phi_0(r) + \int_{r_0}^r \frac{C_1 \phi_0(w)}{\sqrt{\pi(r-w)}} \exp\left(\int_w^r \frac{dl}{\sqrt{\pi(r-l)}}\right) dw \\ &= \phi_0(r) + \int_{r_0}^r \frac{C_1 \phi_0(w)}{\sqrt{\pi(r-w)}} \exp\left(2\sqrt{\frac{r-w}{\pi}}\right) dw \\ &\leq \phi(r), \end{aligned}$$

where

$$\phi(r) := \phi_0(r) + \frac{C_1}{\pi} \exp\left(2\sqrt{\frac{r_2-r_0}{\pi}}\right) (\pi C_0 + 4C_2 \sqrt{r_2-r_0}),$$

and the last inequality was obtained by using $\int_{r_0}^r 1/\sqrt{(r-w)(w-r_0)} dw = \pi$. This finishes our proof. \square

With the help of Lemma 10, we obtain some controls on the spacial derivatives of Φ_γ .

Lemma 11. *Let $s_1 \in (0, 1)$. For any $\gamma \in \mathcal{U}_d$,*

$$\sup_{(s,x) \in [0,1] \times \mathbb{R}} |\partial_x \Phi_\gamma(s, x)| \leq 1 \tag{62}$$

and for $k \geq 2$

$$\sup_{(s,x) \in [0,s_1] \times \mathbb{R}} |\partial_x^k \Phi_\gamma(s, x)| \leq F_k(\gamma(s_1)), \tag{63}$$

where F_k is nonnegative continuous on $[0, \infty)$ depending on s_1 only and independent of γ .

Proof. Assume that $\gamma \in \mathcal{U}_d$. One can explicitly solve Φ_γ in the classical sense by performing the Cole-Hopf transformation. In fact, if $\gamma = \sum_{i=0}^m a_i 1_{[q_i, q_{i+1})}$ for some sequences

$$\begin{aligned} 0 &= q_0 < q_1 < \dots < q_m < q_{m+1} = 1, \\ 0 &\leq a_0 \leq a_1 \leq \dots \leq a_{m-1} \leq a_m < \infty, \end{aligned}$$

then for any $1 \leq i \leq m$,

$$\Phi_\gamma(s, x) = \frac{1}{a_i} \log \mathbb{E} \exp a_i \Phi_\gamma(q_{i+1}, x + z \sqrt{\xi'(q_{i+1}) - \xi'(s)})$$

for all $(s, x) \in [q_i, q_{i+1}) \times \mathbb{R}$, where z is a standard normal random variable. Using the initial condition $\Phi_\gamma(1, x) = |x|$ and an iteration argument, it can be easily checked that for any $k \geq 0$, $\partial_x^k \Phi \in C([0, 1) \times \mathbb{R})$ and for any $k \geq 0$ and $s_0 \in (0, 1)$, $\partial_s \partial_x^k \Phi(s, \cdot) \in C(\mathbb{R})$. Furthermore, it can be verified that (62) holds, and for any $s_0 \in (0, 1)$, there exists a constant C such that

$$\sup_{x \in \mathbb{R}} |\partial_x^k \Phi_\gamma(s, x) - \partial_x^k \Phi_\gamma(s', x)| \leq C|s - s'|, \quad \forall s, s' \in [0, 1), x \in \mathbb{R}, k \geq 2.$$

This inequality implies that $s \mapsto \|\partial_x^k \Phi_\gamma(s, \cdot)\|_\infty$ is a continuous function for $s \in [0, 1)$. Now define $\zeta(s) = (\xi'(1) - \xi'(s))/2$. Fix $0 < s_1 < s_0 < 1$. Set $r_0 = \zeta(s_0)$, $r_1 = \zeta(s_1)$, and $r_2 = \zeta(0)$. Evidently, the function $u(r, x) := \partial_x \Phi_\gamma(\zeta^{-1}(r), x)$ satisfies

$$\partial_r u(r, x) = \partial_{xx} u(r, x) + \kappa_1(r, x) \partial_x u(r, x) + \kappa_2(r, x)$$

for $(r, x) \in [r_0, r_2] \times \mathbb{R}$ with initial condition $\partial_x u(r_0, x) = \partial_x \Psi_\gamma(\zeta^{-1}(r_0), x)$, where

$$\begin{aligned} \kappa_1(r, x) &:= \gamma(\zeta^{-1}(r)) \partial_x u(r, x), \\ \kappa_2(r, x) &:= 0. \end{aligned}$$

Since $\|\kappa_1\|_\infty \leq \gamma(\zeta^{-1}(r_0)) < \infty$, we can apply Lemma 10 to get that

$$\begin{aligned} \sup_{(s, x) \in [0, s_1] \times \mathbb{R}} |\partial_x \Phi_\gamma(s, x)| &= \sup_{(r, x) \in [r_1, r_2] \times \mathbb{R}} |\partial_x u(r, x)| \\ &\leq F(1, \gamma(\zeta^{-1}(r_1)), 0) \\ &= F(1, \gamma(s_1), 0), \end{aligned}$$

where F is a nonnegative continuous function on $[0, \infty)^3$ depending only on s_1 . Letting $F_1(y) = F(1, y, 0)$ gives (63) with $k = 1$. For $k \geq 2$, note that

$$\partial_t (\partial_x^k \Phi_\gamma(s, x)) = \frac{\xi''(s)}{2} \left(\partial_{xx} (\partial_x^k \Phi_\gamma(s, x)) + 2\gamma(s) \partial_x \Phi_\gamma(s, x) (\partial_x^{k+1} \Phi_\gamma(s, x)) + K(s, x) \right),$$

where $K(s, x)$ is the sum of products of spatial derivatives of Φ_γ of order at most k . One may use a similar argument as the case $k = 1$ together with an induction procedure to obtain the proof of (63) for all $k \geq 2$. □

Following a similar definition of the weak solution for the original Parisi PDE in Jagannath-Tobasco [33], we define the weak solution of the Parisi PDE (5) as follows.

Definition 1 (Weak solution). *Let $\gamma \in \mathcal{U}$. Let Φ be a continuous function on $[0, 1] \times \mathbb{R}$ with essentially bounded weak derivative $\partial_x \Phi$. We say that Φ is a weak solution to*

$$\partial_s \Phi(s, x) = -\frac{\xi''(s)}{2} (\partial_{xx} \Phi(s, x) + \gamma(s) (\partial_x \Phi(s, x))^2)$$

on $[0, 1) \times \mathbb{R}$ with $\Phi(1, x) = |x|$ if

$$\int_0^1 \int_{\mathbb{R}} \left(-\Phi \partial_s \phi + \frac{\xi''(s)}{2} (\Phi \partial_{xx} \phi + \gamma(s) (\partial_x \Phi)^2 \phi) \right) dx ds + \int_{\mathbb{R}} \phi(1, x) |x| dx = 0$$

for all smooth ϕ on $(0, 1) \times \mathbb{R}$ with compact support.

Proof of Proposition 2. Let us pick any $(\gamma_n)_n \subset \mathcal{U}_d$ with weak limit γ . From the estimates in Lemma 11, one readily sees that for any $k \geq 0$, the sequence $(\partial_x^k \Phi_{\gamma_n})_{n \geq 1}$ is equicontinuous and uniformly bounded on $[0, s_1] \times \mathbb{R}$ for all $s_1 \in (0, 1)$. Thus, from the Arzela-Ascoli theorem combined with a diagonal process, on any $[0, s_1] \times [-M, M]$, $(\partial_x^k \Phi_{\gamma_n})_{n \geq 1}$ converges uniformly on any compact subset of $[0, 1) \times \mathbb{R}$ for any $k \geq 0$. Here, without loss of generality, we use the same sequence $(\gamma_n)_{n \geq 1}$ instead of adapting a subsequence for notational convenience. The above discussion makes the following well-defined,

$$\Phi_\gamma(s, x) := \lim_{n \rightarrow \infty} \Phi_{\gamma_n}(s, x)$$

for $(s, x) \in [0, 1] \times \mathbb{R}$. Note that for $s \in [0, 1)$, $\Phi_\gamma(s, x)$ is differentiable in $x \in \mathbb{R}$ and satisfies $\|\partial_x \Phi_\gamma(s, \cdot)\|_\infty < \infty$. Since Φ_{γ_n} satisfies

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

for $(s, x) \in [0, 1) \times \mathbb{R}$ with boundary condition $\Phi_{\gamma_n}(1, x) = |x|$, passing to the limit gives

$$\int_0^1 \int_{\mathbb{R}} \left(-\Phi_\gamma \partial_s \phi + \frac{\xi''(s)}{2} (\Phi_\gamma \partial_{xx} \phi + \gamma(s) (\partial_x \Phi_\gamma)^2 \phi) \right) dx ds + \int_{\mathbb{R}} \phi(1, x) |x| dx = 0$$

for all smooth functions ϕ on $(0, 1] \times \mathbb{R}$ with compact support. Therefore, Φ_γ exists in the weak sense and the fact that it satisfies the Lipschitz property (7) follows by applying (9) and noting that $|x|$ has Lipschitz constant 1. Note that this Lipschitz property also implies that the definition of Φ_γ is independent of the choice of the sequence $(\gamma_n)_{n \geq 1}$. To see the uniqueness of Φ_γ , it can be obtained via a fixed point argument identical to [33, Lemma 13]. These together give (i). The above discussion and Lemma 11 imply (ii). Since one can also pick this sequence (γ_n) from \mathcal{U}_c and perform a similar procedure as above, we also obtain (iii). □

References

- [1] D. J. Amit, H. Gutfreund, and H. Sompolinsky. Spin-glass models of neural networks. *Phys. Rev. A*, 32:1007–1018, Aug 1985.
- [2] G. Ben Arous, V. Gayrard, and A. Kuptsov. A new REM conjecture. In *In and out of equilibrium. 2*, volume 60 of *Progr. Probab.*, pages 59–96. Birkhäuser, Basel, 2008.
- [3] A. Auffinger and G. Ben Arous. Complexity of random smooth functions on the high-dimensional sphere. *Ann. Probab.*, 41(6):4214–4247, 2013.
- [4] A. Auffinger, G. Ben Arous, and J. Černý. Random matrices and complexity of spin glasses. *Comm. Pure Appl. Math.*, 66(2):165–201, 2013.
- [5] A. Auffinger and W.-K. Chen. The Parisi formula has a unique minimizer. *Comm. Math. Phys.*, 335(3):1429–1444, 2015.
- [6] A. Auffinger and W.-K. Chen. Parisi formula for the ground state energy in the mixed p-spin model. *ArXiv e-prints*, June 2016.
- [7] A. Auffinger and W.-K. Chen. The Legendre structure of the Parisi formula. *Comm. Math. Phys.*, pages 1–20, 2016.

- [8] S. Bhamidi, P. S. Dey, and A. B. Nobel. Energy Landscape for large average submatrix detection problems in Gaussian random matrices. *ArXiv e-prints*, November 2012.
- [9] S. Boucheron, G. Lugosi, and P. Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013. A nonasymptotic theory of independence, With a foreword by Michel Ledoux.
- [10] A. Bovier and A. Klimovsky. Fluctuations of the partition function in the generalized random energy model with external field. *J. Math. Phys.*, 49(12):125202, 27, 2008.
- [11] A. Bovier and I. Kurkova. Derrida’s generalised random energy models. I. Models with finitely many hierarchies. *Ann. Inst. H. Poincaré Probab. Statist.*, 40(4):439–480, 2004.
- [12] A. Bovier and I. Kurkova. Derrida’s generalized random energy models. II. Models with continuous hierarchies. *Ann. Inst. H. Poincaré Probab. Statist.*, 40(4):481–495, 2004.
- [13] A. Bovier and I. Kurkova. Energy statistics in disordered systems: the local REM conjecture and beyond. *Acta Phys. Polon. B*, 36(9):2621–2634, 2005.
- [14] A. Bovier and I. Kurkova. Local energy statistics in disordered systems: a proof of the local REM conjecture. *Comm. Math. Phys.*, 263(2):513–533, 2006.
- [15] A. Bovier and I. Kurkova. A tomography of the GREM: beyond the REM conjecture. *Comm. Math. Phys.*, 263(2):535–552, 2006.
- [16] A. Bovier and I. Kurkova. Local energy statistics in spin glasses. *J. Stat. Phys.*, 126(4-5):933–949, 2007.
- [17] A. Bruckner. *Differentiation of real functions*, volume 5 of *CRM Monograph Series*. American Mathematical Society, Providence, RI, second edition, 1994.
- [18] P. Charbonneau, J. Kurchan, G. Parisi, P. Urbani, and F. Zamponi. Fractal free energy landscapes in structural glasses. *Nature communications*, 5, 2014.
- [19] S. Chatterjee. Chaos, concentration, and multiple valleys. *ArXiv e-prints*, October 2008.
- [20] S. Chatterjee. Disorder chaos and multiple valleys in spin glasses. *ArXiv e-prints*, July 2009.
- [21] S. Chatterjee. *Superconcentration and related topics*. Springer Monographs in Mathematics. Springer, Cham, 2014.
- [22] W.-K. Chen. Disorder chaos in the Sherrington-Kirkpatrick model with external field. *Ann. Probab.*, 41(5):3345–3391, 2013.
- [23] W.-K. Chen. Chaos in the mixed even-spin models. *Comm. Math. Phys.*, 328(3):867–901, 2014.
- [24] W.-K. Chen. Variational representations for the Parisi functional and the two-dimensional Guerra-Talagrand bound. *ArXiv e-prints*, January 2015.
- [25] W.-K. Chen, H.-W. Hsieh, C.-R. Hwang, and Y.-C. Sheu. Disorder chaos in the spherical mean-field model. *J. Stat. Phys.*, 160(2):417–429, 2015.
- [26] W.-K. Chen and A. Sen. Parisi formula, disorder chaos and fluctuation for the ground state energy in the spherical mixed p -spin models. *ArXiv e-prints*, December 2015.

- [27] A. Choromanska, M. Henaff, M. Mathieu, G. Ben Arous, and Y. LeCun. The loss surfaces of multilayer networks. In G. Lebanon and S. V. N. Vishwanathan, editors, *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics, AISTATS 2015, San Diego, California, USA, May 9-12, 2015*, volume 38 of *JMLR Workshop and Conference Proceedings*. JMLR.org, 2015.
- [28] A. Dembo, A. Montanari, and S. Sen. Extremal cuts of sparse random graphs. *ArXiv e-prints*, March 2015.
- [29] J. Ding, R. Eldan, and A. Zhai. On multiple peaks and moderate deviations for the supremum of a Gaussian field. *Ann. Probab.*, 43(6):3468–3493, 2015.
- [30] Y. V. Fyodorov. Complexity of random energy landscapes, glass transition, and absolute value of the spectral determinant of random matrices. *Phys. Rev. Lett.*, 92(24):240601, 4, 2004.
- [31] Y. V. Fyodorov and I. Williams. Replica symmetry breaking condition exposed by random matrix calculation of landscape complexity. *J. Stat. Phys.*, 129(5-6):1081–1116, 2007.
- [32] D. Gamarnik and Q. Li. Finding a Large Submatrix of a Gaussian Random Matrix. *ArXiv e-prints*, February 2016.
- [33] A. Jagannath and I. Tobasco. A dynamic programming approach to the Parisi functional. *Proc. Amer. Math. Soc.*, 144(7):3135–3150, 2016.
- [34] A. Jagannath and I. Tobasco. Low temperature asymptotics of spherical mean field spin glasses. *ArXiv e-prints*, February 2016.
- [35] I. Karatzas and S. E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.
- [36] J. Kurchan, G. Parisi, and M. A. Virasoro. Barriers and metastable states as saddle points in the replica approach. *Journal de Physique I*, 3(8):1819–1838, 1993.
- [37] Y. C. Lee, G. Doolen, H. H. Chen, G. Z. Sun, T. Maxwell, H. Y. Lee, and C. L. Giles. Machine learning using a higher order correlation network. *Physica D: Nonlinear Phenomena*, 22(1):276 – 306, 1986.
- [38] M. Mézard and A. Montanari. *Information, physics, and computation*. Oxford graduate texts. Oxford University Press, Oxford, 2009. Autre tirage : 2010, 2012.
- [39] M. Mézard, G. Parisi, and M. A. Virasoro. *Spin glass theory and beyond*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific Publishing Co., Inc., Teaneck, NJ, 1987.
- [40] D. Panchenko. On differentiability of the Parisi formula. *Electron. Commun. Probab.*, 13:241–247, 2008.
- [41] D. Panchenko. *The Sherrington-Kirkpatrick model*. Springer Monographs in Mathematics. Springer, New York, 2013.
- [42] D. Panchenko. The Parisi formula for mixed p -spin models. *Ann. Probab.*, 42(3):946–958, 2014.
- [43] P. Peretto and J. J. Niez. Long term memory storage capacity of multiconnected neural networks. *Biological Cybernetics*, 54(1):53–63, 1986.

- [44] T. Rizzo. Chaos in mean-field spin-glass models. In *Spin glasses: statics and dynamics*, volume 62 of *Progr. Probab.*, pages 143–157. Birkhäuser Verlag, Basel, 2009.
- [45] R. Tyrrell Rockafellar. *Convex analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.
- [46] E. Subag. The complexity of spherical p -spin models - a second moment approach. *ArXiv e-prints*, April 2015.
- [47] E. Subag. The geometry of the Gibbs measure of pure spherical spin glasses. *ArXiv e-prints*, April 2016.
- [48] E. Subag and O. Zeitouni. The extremal process of critical points of the pure p -spin spherical spin glass model. *ArXiv e-prints*, September 2015.
- [49] M. Talagrand. The Parisi formula. *Ann. of Math. (2)*, 163(1):221–263, 2006.