The SK Model Is Infinite Step
Replica Symmetry Breaking at Zero Temperature

ANTONIO AUFFINGER
Northwestern University

WEI-KUO CHEN
University of Minnesota

AND

QIANG ZENG
Northwestern University

1 Introduction and Main Results

1.1 The Model and Parisi's Solution

The study of mean field spin glass models is a very rich and important part of theoretical physics [20, 21, 32]. For mathematicians, it is a challenging program [26, 37, 41]. Roughly speaking, the main goal is to study the global maxima or, more generally, the "largest individuals" of a stochastic process with "high-dimensional" correlation structure.

The classic example of such a process is the mixed p-spin model. Its Hamiltonian (or energy) H_N is defined on the spin configuration space $\Sigma_N = \{-1, 1\}^N$ by

$$H_N(\sigma) = X_N(\sigma) + h \sum_{i=1}^{N} \sigma_i$$

for

$$X_N(\sigma) := \sum_{p \geq 2} c_p X_{N,p}(\sigma).$$

Here, the processes $X_{N,p}$ are pure p-spin Hamiltonians defined as

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

where $g_{i_1, \ldots, i_p}$ are i.i.d. standard normal variables for all $1 \leq i_1, \ldots, i_p \leq N$ and $p \geq 2$. The real sequence $(c_p)_{p \geq 2}$ satisfies $\sum_{p \geq 2} 2^p c_p^2 < \infty$. The parameter $h \in \mathbb{R}$ denotes the strength of the external field. By definition, the covariance of X_N can be computed as

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

where

$$\xi(s) := \sum_{p \geq 2} c_p^2 s^p$$

and

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

is the normalized inner product between σ^1 and σ^2, known as the overlap. The covariance structure of X_N is as rich as the structure of the metric space (Σ_N, d), where d is the normalized Hamming distance on Σ_N,

$$d(\sigma^1, \sigma^2) = \frac{1 - R(\sigma^1, \sigma^2)}{2}.$$

The problem of computing the maximum energy (or the ground state energy) of H_N as N diverges is a rather nontrivial task. Standard statistical mechanics deals with this problem by considering the Gibbs measure

$$G_{N,\beta}(\sigma) = \frac{1}{Z_{N,\beta}} e^{\beta H_N(\sigma)}$$

and the free energy

$$F_{N,\beta} = \frac{1}{\beta N} \log Z_{N,\beta},$$

where $Z_{N,\beta}$ is the partition function of H_N defined as

$$Z_{N,\beta} = \sum_{\sigma \in \Sigma_N} e^{\beta H_N(\sigma)}.$$

The parameter $\beta = 1/(kT) > 0$ is called the inverse temperature, where k is the Boltzmann constant and T is the absolute temperature. The main goal in this approach is to try to describe the large N limit of the sequences of the free energies $F_{N,\beta}$ and the Gibbs measures $G_{N,\beta}$. When the temperature T decreases, large values of H_N become more important (to both the partition function $Z_{N,\beta}$ and to the Gibbs measure $G_{N,\beta}$), and they prevail over the more numerous smaller values. Since H_N is a high-dimensional correlated field with a large number of points near its global maximum, this question becomes very challenging, especially for small values of T.

When $\xi(s) = s^2/2$ and $h = 0$, the model above is the famous Sherrington-Kirkpatrick (SK) model introduced in [34] as a mean field modification of the Edwards-Anderson model [14]. Using a nonrigorous replica trick and a replica symmetric hypothesis, Sherrington and Kirkpatrick [34] proposed a solution to the limiting free energy of the SK model. Their solution, however, was incomplete; an alternative solution was proposed in 1979 in a series of groundbreaking articles by Giorgio Parisi [28–31], where it was foreseen that:

(i) The limiting free energy is given by a variational principle, known as the Parisi formula.

(ii) The Gibbs measures are asymptotically ultrametric.

(iii) At low enough temperature, the symmetry of replicas is broken infinitely many times.

The first two predictions were confirmed in the past decade. Following the beautiful discovery of Guerra's broken replica symmetry scheme [16], the Parisi formula was proved in the seminal work of Talagrand [40] in 2006 under the convexity assumption of ξ. Later, in 2012, the ultrametricity conjecture was established by Panchenko [25] assuming the validity of the extended Ghirlanda-Guerra identities [15]. These identities are known to be valid for the SK model with an asymptotically vanishing perturbation and for generic models without any perturbation. Here, the model is said to be generic if the span of $\{1\} \cup \{s^p : c_p \neq 0\}$ is dense in the space of continuous functions $C[-1, 1]$ under the maximum norm. As a consequence of ultrametricity, the Parisi formula was further extended to all mixed p-spin models by Panchenko [27] utilizing the Aizenman-Sims-Starr scheme [2].

More precisely, the Parisi formula for general mixtures is stated as follows. Denote by $\mathcal{M}$ the collection of all cumulative distribution functions α on $[0, 1]$ and by $\alpha(ds)$ the probability induced by α. For $\alpha \in \mathcal{M}$, define

$$(1.1) \qquad \mathscr{P}_\beta(\alpha) = \frac{\log 2}{\beta} + \Psi_{\alpha,\beta}(0, h) - \frac{1}{2} \int_0^1 \beta \alpha(s) s \xi''(s) ds,$$

where $\Psi_{\alpha,\beta}(t, x)$ is the weak solution to the nonlinear parabolic PDE

$$\partial_t \Psi_{\alpha,\beta}(t, x) = -\frac{\xi''(t)}{2} \big(\partial_{xx} \Psi_{\alpha,\beta}(t, x) + \beta \alpha(t) (\partial_x \Psi_{\alpha,\beta}(t, x))^2 \big)$$

for $(t, x) \in [0, 1) \times \mathbb{R}$ with boundary condition

$$\Psi_{\alpha,\beta}(1, x) = \frac{\log \cosh \beta x}{\beta}.$$

For the existence and regularity of $\Psi_{\alpha,\beta}$, we refer the readers to [6, 17]. The Parisi formula [40] states that

$$(1.2) \qquad F_\beta := \lim_{N \to \infty} F_{N,\beta} = \inf_{\alpha \in \mathcal{M}} \mathscr{P}_\beta(\alpha) \text{ a.s.}$$

The infinite-dimensional variational problem on the right-hand side of (1.2) has a unique minimizer [6], denoted by $\alpha_{P,\beta}$. The measure $\alpha_{P,\beta}(dt)$ induced by $\alpha_{P,\beta}$ is known as the Parisi measure [21].[1] In Parisi's solution, it is predicted that the Parisi measure is the limiting distribution of the overlap $R(\sigma^1, \sigma^2)$ under the measure $\mathbb{E} G_N^{\otimes 2}$. More importantly, this together with the asymptotic ultrametricity implies that the spin configurations under the Gibbs measure form a hierarchical clustering structure and the number of levels therein is determined by the number

---

[1] The Parisi measure is the inverse of the functional order parameter $q(x)$ in [29], sometimes written as $x(q)$.

of points in the support of the Parisi measure $\alpha_{P,\beta}(dt)$. The importance of (iii) lies in the fact that it indicates the phase transition of the model between high- and low-temperature regimes. While at high temperature the clusters contain no layers, at low temperature they begin to possess clustering structures with multiple layers. The statement (iii) advocates the existence of infinitely many layers within the clusters at low enough temperature. In a nutshell, the Parisi measure is the key ingredient of the matter that describes the structure of the Gibbs measure as well as the free energy of the system. See [21, 26] for a detailed discussion.

The importance of the Parisi measure leads to the following classification. If a Parisi measure $\alpha_{P,\beta}(dt)$ is a Dirac measure, we say that the model is replica symmetric (RS). For $k \geq 1$, we say that the model has k levels of replica symmetry breaking (k-RSB) if the Parisi measure is atomic and has exactly $k + 1$ jumps. If the Parisi measure is neither RS nor k-RSB for some $k \geq 1$, then the model has infinite levels of replica symmetry breaking (∞-RSB). We will also say that the model is at least k-RSB if the Parisi measure contains at least $k + 1$ distinct values in its support.

It is expected that when ∞-RSB happens, the functional order parameter has a nonempty interval in its support. This prediction in the physics literature is named as full-step replica symmetry breaking (FRSB). This prediction plays an inevitable role in Parisi's original solution of the SK model. It can be written as follows:

PREDICTION (Parisi). *For any ξ and h, there exists a critical inverse temperature $\beta_c > 0$ such that for any $\beta > \beta_c$, the mixed p-spin model is FRSB.*

1.2 Main Results

In this paper, we establish this prediction at zero temperature. To prepare for the statement of our main result, we recall the Parisi formula for the ground state energy of H_N as follows. First of all, the Parisi formula allows us to compute the ground state energy of the model by sending the temperature T to 0,

$$(1.3) \qquad \text{GSE} := \lim_{N \to \infty} \max_{\sigma \in \Sigma_N} \frac{H_N(\sigma)}{N} = \lim_{\beta \to \infty} F_\beta = \lim_{\beta \to \infty} \inf_{\alpha \in \mathcal{M}} \mathscr{P}_\beta(\alpha),$$

where the validity of the first equality can be found, for instance, in Panchenko's book [26, chap. 1]. In the physics literature, it is a convention that the ground state energy is defined as the minimum of the Hamiltonian. This is indeed equivalent to our formulation here since the disorder coefficients $g_{i_1,\dots,i_p}$ are symmetric with respect to the origin. Recently the analysis of the β-limit of the second equality was carried out in Auffinger-Chen [8], and it was discovered that the ground state energy can be written as a Parisi-type formula. Let $\mathscr{U}$ denote the collection of all cumulative distribution functions γ on $[0, 1)$ induced by any measures on $[0, 1)$ and satisfying $\int_0^1 \gamma(t)dt < \infty$. Denote by $\gamma(dt)$ the measure that induces γ and endow $\mathscr{U}$ with the $L^1(dt)$-distance. For each $\gamma \in \mathscr{U}$, consider the weak solution

to the Parisi PDE,

$$\partial_t \Psi_\gamma(t, x) = -\frac{\xi''(t)}{2}\big(\partial_{xx}\Psi_\gamma(t, x) + \gamma(t)(\partial_x\Psi_\gamma(t, x))^2\big)$$

for $(t, x) \in [0, 1) \times \mathbb{R}$ with boundary condition

$$\Psi_\gamma(1, x) = |x|.$$

One may find the existence and regularity properties of this PDE solution in [12]. The Parisi functional at zero temperature is given by

$$\mathscr{P}(\gamma) = \Psi_\gamma(0, h) - \frac{1}{2}\int_0^1 t\xi''(t)\gamma(t)dt.$$

Auffinger and Chen [8] proved that the maximum energy can be computed through

(1.4) $$\text{GSE} = \inf_{\gamma \in \mathscr{U}} \mathscr{P}(\gamma) \text{ a.s.}$$

We call this variational representation the Parisi formula at zero temperature. It was proved in [12] that this formula has a unique minimizer denoted by γ_P. We call $\gamma_P(dt)$ the Parisi measure at zero temperature. We say that the model is ∞-RSB at zero temperature if $\gamma_P(dt)$ contains infinitely many points in its support. Our first main result is a proof of Parisi's ∞-RSB prediction at zero temperature.

THEOREM 1.1. *For any ξ and h, the mixed p-spin model at zero temperature is ∞-RSB.*

Remark 1.2. The above theorem is a first step towards the validation of Parisi's FRSB prediction. At zero temperature (and at sufficiently low temperatures), it is expected that the support of the Parisi measure contains an interval. This remains as an important open question.

Similar to the role of the Parisi measure at positive temperature played in describing the behavior of the model, the Parisi measure at zero temperature also has its own relevance in understanding the energy landscape of the Hamiltonian around the maximum energy.

PROPOSITION 1.3. *For any ξ and h, if u lies in the support of $\gamma_P(dt)$, then for any $\varepsilon, \eta > 0$, there exists some constant $K > 0$ independent of N such that*

(1.5) $$\mathbb{P}\bigg(\exists \sigma^1, \sigma^2 \text{ such that } R_{1,2} \in (u - \varepsilon, u + \varepsilon)$$
$$\text{and } \frac{H_N(\sigma^1)}{N}, \frac{H_N(\sigma^2)}{N} \geq \text{GSE} - \eta\bigg) \geq 1 - Ke^{-\frac{N}{K}}$$

for all $N \geq 1$.

This means that for any $u \in \text{supp}\,\gamma_P$, one can always find two spin configurations around the maximum energy such that their overlap is near u with overwhelming probability. Recently it was proved in [12] that when $h = 0$ and the model is even, i.e., $c_p = 0$ for all odd p, the system exhibits the so-called multiple

peaks feature in the sense that there exist exponentially many near maximizers, and they are nearly orthogonal to each other. Theorem 1.1 and Proposition 1.3 together add a more fruitful structure to the landscape of the Hamiltonian. In view of Proposition 1.3, it would be of great interest to understand the size of the number of the pairs (σ^1, σ^2) with $R_{1,2} \in (u - \varepsilon, u + \varepsilon)$ and $H_N(\sigma^1), H_N(\sigma^2) \geq N(\mathrm{GSE} - \eta)$ for a typical realization.

Remark 1.4. The relationship between the Parisi solution and combinatorial optimization problems has been investigated since the 1980s; see [21] for references. The problem of computing the maximum energy is also generally known as the Dean's problem and is frequently used to motivate the theory of mean field spin glasses; see [21, 26]. More recently the formula (1.3) at zero temperature has appeared in other optimization problems related to theoretical computer science such as extremal cuts on sparse random graphs; see [13, 33] and the references therein. In the physics literature, the FRSB prediction is also discussed in [22–24].

Remark 1.5. For the mixed p-spin model defined on the sphere, $\{\sigma \in \mathbb{R}^N : \sum_{i=1}^N \sigma_i^2 = N\}$, the corresponding Parisi-type formula at both positive and zero temperatures is much simpler than (1.1) and (1.3) as there is no PDE involved. In this setting, it is known in [5, 9, 10, 18, 38] that the Parisi variational formulas can be explicitly solved for some choices of the mixture parameters $(c_p)_{p \geq 2}$, and algebraic conditions on ξ were presented so that the model has Parisi measures of RS, 1RS, 2RSB, and FRSB. The study of the energy landscapes has also appeared in [3, 4, 9, 11, 18, 35, 36].

Theorem 1.1 and Proposition 1.3 indicate that the spin configurations around the maximum energy are not simply clustered into equidistant groups. This is in sharp contrast to the energy landscape of the spherical version of the mixed p-spin model, where in the pure p-spin model, i.e., $\xi(t) = t^p$ for $p \geq 3$, it was shown by Subag [35] that around the maximum energy, the spin configurations are essentially orthogonally structured. This structure was also presented in more general mixtures of the spherical model in the recent work of Auffinger and Chen [9].

We now return to the positive temperature case. Recall the Parisi measure $\alpha_{P,\beta}$ introduced in (1.2). Our second main result, as a consequence of Theorem 1.1, shows that for any mixture parameter ξ and external field h, the number of levels of replica symmetry breaking must diverge as β goes to infinity.

THEOREM 1.6. *Let $k \geq 1$. For any ξ and h, there exists β_k such that the mixed p-spin model is at least k-RSB for all $\beta > \beta_k$.*

For the SK model without external field, $\xi(s) = s^2/2$ and $h = 0$, Aizenman, Lebowitz, and Ruelle [1] showed that the model is RS, i.e., the limiting distribution of $R_{1,2}$ is a Dirac mass if β is in the high-temperature regime $\beta < 1$. Later, it was also understood by Toninelli [42] that the model is not RS in the low-temperature region $\beta > 1$. It is conjectured that the whole region $\beta > 1$ is expected to be of

FRSB. One may find recent progress on the phase transition from RS to RSB for more general mixtures in [5, 19, 42]. However, the existence of FRSB in the mixed p-spin models remains unknown.

1.3 Our Approach

The main novelty of our approach to Theorem 1.1 is to explore the Parisi formula for the ground state energy (1.4) by considering a perturbation of the Parisi functional around the point 1. In short, we show that it is always possible to lower the value of the Parisi functional of any atomic measure with finite atoms by adding a large enough jump near 1. The argument is quite delicate and depends on some observations that may look counterintuitive at first glance; see Remark 2.9 below. At finite temperature, since the Parisi measure is a probability measure, the idea of adding a large jump is not feasible. Theorem 1.6 follows from Theorem 1.1 after some weak convergence considerations. Recently there have been many papers investigating the properties of the Parisi functional via the stochastic optimal control theory for the Parisi PDE, e.g., [6, 18, 19]. As our analysis here is quite subtle and relies on many identities related to the Gaussian expectations as well as some technical calculations, we prefer adapting the setting and framework as in Talagrand's book [39] in order to keep track of our control with clarity.

2 Lowering the Value of the Parisi Functional

In this section, we show that for any atomic $\gamma(ds)$ with finitely many jumps, one can always lower the value of the Parisi functional by a perturbation of γ around 1. Let $\gamma \in \mathcal{U}$ be fixed. Suppose that $\gamma(dt)$ is atomic and consists of finitely many jumps, that is,

$$\gamma(t) = \sum_{i=0}^{n-1} m_i \mathbb{1}_{[q_i, q_{i+1})}(t) + m_n \mathbb{1}_{[q_n, 1)}(t),$$

where $(q_i)_{0 \le i \le n}$ and $(m_i)_{0 \le i \le n}$ satisfy

(2.1)
$$0 = q_0 < q_1 < \cdots < q_n < 1,$$
$$0 \le m_0 < m_1 < \cdots < m_n < \infty.$$

Here and in what follows, $\mathbb{1}_B(t) = \mathbb{1}_{[t \in B]}$ is the indicator function of the set $B \subset \mathbb{R}$. Let m_{n+1} be any number greater than m_n. For any $q \in (q_n, 1)$, consider the following perturbation of γ,

(2.2) $$\gamma_q(t) = \sum_{i=0}^{n-1} m_i \mathbb{1}_{[q_i, q_{i+1})}(t) + m_n \mathbb{1}_{[q_n, q)}(t) + m_{n+1} \mathbb{1}_{[q, 1)}(t),$$

where q_i and m_i are from (2.1). In other words, we add a jump to the top of γ. Our main result is the following theorem. It says that if m_{n+1} is large enough, then the Parisi functional evaluated at perturbed measure $\gamma_q(dt)$ has a smaller value than $\mathscr{P}(\gamma)$ locally for q near 1.

THEOREM 2.1. *There exist $m_{n+1} > m_n$ and $\eta \in (q_n, 1)$ such that*

$$\mathscr{P}(\gamma_q) < \mathscr{P}(\gamma)$$

for all $\eta \le q < 1$.

The following three subsections are devoted to the proof of Theorem 2.1.

2.1 Probabilistic Representation of $\mathscr{P}$

We start by observing that the Parisi functional at γ admits a probabilistic expression by an application of the Cole-Hopf transformation to the Parisi PDE. Indeed, let $z_0, \ldots, z_n$ be i.i.d. standard Gaussian random variables. For q_i, $i = 0, \ldots, n$, given in (2.1), we denote

$$J = h + \sum_{i=0}^{n-1} z_i \sqrt{\xi'(q_{i+1}) - \xi'(q_i)} + z_n \sqrt{\xi'(1) - \xi'(q_n)}.$$

Set

$$X_{n+1} = |J|.$$

Define iteratively, for m_i, $i = 0, \ldots, n$, given in (2.1),

$$X_i = \frac{1}{m_i} \log \mathbb{E}_{z_i} \exp m_i X_{i+1},$$

where $\mathbb{E}_{z_i}$ stands for the expectation for z_i. Here X_0 is defined as $\mathbb{E}_{z_0} X_1$ if $m_0 = 0$. Then by applying the Cole-Hopf transformation, $\Psi_\gamma(0, h) = X_0$, and thus

$$\mathscr{P}(\gamma) = X_0 - \frac{1}{2} \sum_{i=0}^{n-1} m_i \int_{q_i}^{q_{i+1}} t\xi''(t)dt - \frac{m_n}{2} \int_{q_n}^{1} t\xi''(t)dt.$$

Recall the perturbation γ_q from (2.2). Clearly $\gamma_q = \gamma$ on $[0, q)$ for all $q_n < q < 1$. For notational convenience, we denote

$$(2.3) \qquad\qquad q_{n+1} = q, \quad q_{n+2} = 1.$$

In a similar manner, we can express $\Psi_{\gamma_q}(0, h)$ as follows. Let z_{n+1} be a standard Gaussian random variables independent of $z_1, \ldots, z_n$. Define

$$Y_{n+2} = \left| h + \sum_{j=0}^{n+1} z_j \sqrt{\xi'(q_{j+1}) - \xi'(q_j)} \right|$$

and iteratively, for $0 \le i \le n + 1$,

$$(2.4) \qquad\qquad Y_i = \frac{1}{m_i} \log \mathbb{E}_{z_i} \exp m_i Y_{i+1}.$$

Here again we let $Y_0 = \mathbb{E}_{z_0} Y_1$ whenever $m_0 = 0$. Thus, $\Psi_{\gamma_q}(0, h) = Y_0$ for any $q \in (q_n, 1)$. As a result,

$$
\mathscr{P}(\gamma_q) = Y_0 - \frac{1}{2} \sum_{i=0}^{n-1} m_i \int_{q_i}^{q_{i+1}} t\xi''(t)dt - \frac{m_n}{2} \int_{q_n}^{q} t\xi''(t)dt
$$

(2.5)

$$
- \frac{m_{n+1}}{2} \int_{q}^{1} t\xi''(t)dt.
$$

In particular, we have $\lim_{q\to 1-} \Psi_{\gamma_q}(0, h) = \Psi_{\gamma}(0, h)$ and $\lim_{q\to 1-} \mathscr{P}(\gamma_q) = \mathscr{P}(\gamma)$.

2.2 Some Auxiliary Lemmas

We state two propositions that will be heavily used in our main proof in the next subsection. Let $0 \le a < t < b$ and $0 < m < m'$. Denote by z a standard normal random variable. Define

$$
A(t, x) = \frac{1}{m'} \log \mathbb{E} \exp m' |x + z\sqrt{b - t}|,
$$

(2.6)

$$
B(t, x) = \frac{1}{m} \log \mathbb{E} \exp m A(t, x + z\sqrt{t - a}),
$$

$$
C(t, x) = \mathbb{E} \big(\partial_x A(t, x + z\sqrt{t - a}) \big)^2 V(t, x, x + z\sqrt{t - a}),
$$

where

$$
V(t, x, y) = e^{m(A(t,y) - B(t,x))}.
$$

For any $(t, x) \in [a, b) \times \mathbb{R}$, define random variables:

$$
\widetilde{V}(t, x) = V(t, x, x + z\sqrt{t - a}),
$$

$$
A_t(t, x) = \partial_t A(t, x + z\sqrt{t - a}),
$$

$$
A_{tx}(t, x) = \partial_{tx} A(t, x + z\sqrt{t - a}),
$$

$$
A_x(t, x) = \partial_x A(t, x + z\sqrt{t - a}),
$$

$$
A_{xx}(t, x) = \partial_{xx} A(t, x + z\sqrt{t - a}),
$$

$$
A_{xxx}(t, x) = \partial_{xxx} A(t, x + z\sqrt{t - a}).
$$

We stress that $A_t(t, x) \ne \partial_t A(t, x)$, $A_x(t, x) \ne \partial_x A(t, x)$, etc., in our notation. Note that $\mathbb{E}\widetilde{V}(t, x) = 1$. The main results of this subsection are the following two propositions.

PROPOSITION 2.2. *For any* $(t, x) \in [a, b) \times \mathbb{R}$*, we have that*

(2.7)
$$
\partial_t B(t, x) = \frac{(m - m')}{2} C(t, x)
$$

and

$$\partial_t C(t, x) = \mathbb{E}\big(A_{xx}(t, x)^2 + 2(m - m')A_{xx}(t, x)A_x(t, x)^2\big)\tilde{V}(t, x)$$

(2.8)
$$+ \frac{(m - m')m}{2}\big(\mathbb{E}A_x(t, x)^4\tilde{V}(t, x) - \big(\mathbb{E}A_x(t, x)^2\tilde{V}(t, x)\big)^2\big).$$

Remark 2.3. Functions (2.6) and the formula (2.7) also appeared in [41, sec. 14.7] in a similar manner, where in the exponent of A, the author used the random variable $\beta^{-1}\log\cosh(\beta(x + z\sqrt{b - t}))$ instead of $|x + z\sqrt{b - t}|$.

PROPOSITION 2.4. *For* $(t, x) \in [a, b) \times \mathbb{R}$, *we have that*

(2.9)
$$\lim_{t \to b-} C(t, x) = 1$$

and

$$\liminf_{t \to b-} \partial_t C(t, x) \geq \frac{2(m + m')}{3}\Delta(x),$$

where

(2.10)
$$\Delta(x) = \frac{2}{\sqrt{2\pi(b - a)}}\frac{e^{-\frac{x^2}{2(b-a)}}}{\mathbb{E}e^{m|x+z\sqrt{b-a}|}}.$$

Before we turn to the proofs of Propositions 2.2 and 2.4, we first gather some fundamental properties of the function A.

LEMMA 2.5. *A is the classical solution to the following PDE with boundary condition* $A(b, x) = |x|$:

(2.11)
$$\partial_t A(t, x) = -\frac{1}{2}\big(\partial_{xx}A(t, x) + m'\big(\partial_x A(t, x)\big)^2\big)$$

for $(t, x) \in [a, b) \times \mathbb{R}$. *In addition,*

(2.12)
$$|\partial_x A(t, x)| \leq 1, \quad (t, x) \in [a, b) \times \mathbb{R},$$

(2.13)
$$\lim_{t \to b-} \partial_x A(t, x) = \text{sign}(x) \quad \forall x \in \mathbb{R} \setminus \{0\},$$

$$\lim_{t \to b-} \mathbb{E}A_x(t, x)^{2k}\tilde{V}(t, x) = 1$$

(2.14)
$$\forall(t, x) \in [a, b) \times \mathbb{R}, \ \forall k \in \mathbb{N}, \ \forall 0 < m < m',$$

where $\text{sign}(x) = 1$ *if* $x > 0$ *and* $= -1$ *if* $x < 0$.

PROOF. Define

$$g(t, x) = e^{\frac{(b-t)m'^2}{2} + m'x}\Phi\left(m'\sqrt{b - t} + \frac{x}{\sqrt{b - t}}\right),$$

where Φ is the cumulative distribution function of a standard normal random variable. Note that a direct computation gives

$$\mathbb{E}e^{m'|x+z\sqrt{b-t}|} = g(t, x) + g(t, -x).$$

Thus,

$$A(t, x) = \frac{1}{m'} \log\big(g(t, x) + g(t, -x)\big).$$

From this expression, we can compute that

$$\partial_x A(t, x) = \frac{g(t, x) - g(t, -x)}{g(t, x) + g(t, -x)},$$

(2.15) $$\partial_{xx} A(t, x) = m'\left(1 - \left(\frac{g(t, x) - g(t, -x)}{g(t, x) + g(t, -x)}\right)^2\right) + 2\Gamma(t, x),$$

$$\partial_t A(t, x) = -\frac{m'}{2} - \Gamma(t, x),$$

where

$$\Gamma(t, x) := \frac{1}{\sqrt{2\pi(b - t)}} \frac{e^{-\frac{x^2}{2(b-t)}}}{g(t, x) + g(t, -x)}.$$

Therefore, these equations together validate (2.11). From the first equation, we can also conclude (2.12) and (2.13). Note that

$$\lim_{t \to b-} V(t, x, y) = V(b, x, y) = \frac{e^{m|y|}}{\mathbb{E} e^{m|x+z\sqrt{b-a}|}}$$

and $\ln V(t, \cdot, \cdot)$ is at most of linear growth. From (2.12) and (2.13), the dominated convergence theorem implies (2.14). $\qquad\square$

PROOF OF PROPOSITION 2.2. To lighten our notation, we will drop the dependence on (t, x) in this proof. Recall that the Gaussian integration by parts states that for a standard normal random variable z, $\mathbb{E} z f(z) = \mathbb{E} f'(z)$ for all absolutely continuous functions f satisfying that $\ln |f|$ is at most of linear growth at infinity; see, e.g., [41, sec. A.4]. From this formula and the PDE (2.11),

$$\partial_t B = \mathbb{E}\left(A_t + \frac{z}{2\sqrt{t-a}} A_x\right)\tilde{V}$$

$$= \mathbb{E}\left(-\frac{1}{2}(A_{xx} + m' A_x^2) + \frac{1}{2}(A_{xx} + m A_x^2)\right)\tilde{V}$$

$$= \frac{m - m'}{2} \mathbb{E} A_x^2 \tilde{V},$$

which gives (2.7). To compute the partial derivative of C in t, write $\partial_t C = \mathrm{I} + \mathrm{II}$ for

$$\mathrm{I} := 2\mathbb{E}\left(A_{tx} + \frac{z}{2\sqrt{t-a}} A_{xx}\right) A_x \tilde{V},$$

$$\mathrm{II} := m\mathbb{E} A_x^2\left(A_t + \frac{z}{2\sqrt{t-a}} A_x - \partial_t B\right)\tilde{V}.$$

Here, from (2.11), since

$$\partial_{tx} A = -\frac{1}{2}\big(\partial_{xxx} A + 2m'\partial_{xx} A \cdot \partial_x A\big)$$

for any $(t, x) \in [a, b) \times \mathbb{R}$, using the Gaussian integration by parts again gives

$$
\begin{aligned}
\mathrm{I} &= 2\mathbb{E}\left(A_{tx} A_x + \frac{1}{2}\big(A_{xxx} A_x + A_{xx}^2 + m A_{xx} A_x^2\big)\right)\widetilde{V}\\
&= \mathbb{E}\big(-A_x\big(A_{xxx} + 2m' A_{xx} A_x\big) + \big(A_{xxx} A_x + A_{xx}^2 + m A_{xx} A_x^2\big)\big)\widetilde{V}\\
&= \mathbb{E}\big(A_{xx}^2 + (m - 2m') A_{xx} A_x^2\big)\widetilde{V}.
\end{aligned}
$$

In addition, from (2.11) and the Gaussian integration by parts,

$$
\begin{aligned}
\mathrm{II} &= m\mathbb{E}\left(-\frac{1}{2}\big(A_{xx} A_x^2 + m' A_x^4\big) + \frac{1}{2}\big(3 A_{xx} A_x^2 + m A_x^4\big) - A_x^2 \partial_t B\right)\widetilde{V}\\
&= m\mathbb{E} A_{xx} A_x^2 \widetilde{V} + \frac{(m - m')m}{2}\big(\mathbb{E} A_x^4 \widetilde{V} - \big(\mathbb{E} A_x^2 \widetilde{V}\big)^2\big).
\end{aligned}
$$

From the above, (2.8) follows. $\square$

To handle the limits in Proposition 2.4, we need two lemmas.

LEMMA 2.6. *For any odd $k \geq 1$, there exists a constant K independent of t such that*

$$(2.16) \qquad\qquad \mathbb{E} A_x(t, x)^{k-1} A_{xx}(t, x) \widetilde{V}(t, x) \leq \frac{K e^{m|x|}}{\sqrt{t - a}}$$

for all $t \in [a, b)$ and $x \in \mathbb{R}$. Moreover,

$$(2.17) \qquad\qquad \lim_{t \to b-} \mathbb{E} A_x(t, x)^{k-1} A_{xx}(t, x) \widetilde{V}(t, x) = \frac{1}{k}\Delta(x),$$

where $\Delta(x)$ is defined in Proposition 2.4.

PROOF. Define

$$D(t, x) = \mathbb{E} z A_x^k(t, x) \widetilde{V}(t, x).$$

Note that $|\partial_x A(t, x)| \leq 1$ and $B(t, x) \geq 0$. We have

$$V(t, x, y) = e^{m(A(t,y) - B(t,x))} \leq e^{mA(t,0) + m|y|}.$$

Using the Gaussian integration by parts, we can write

$$(2.18) \quad D(t, x) = \sqrt{t - a}\,\mathbb{E}\big(k A_x(t, x)^{k-1} A_{xx}(t, x) + m A_x(t, x)^{k+1}\big)\widetilde{V}(t, x).$$

This and the previous inequality together imply (2.16) since

$$
\begin{aligned}
k\sqrt{t - a}\,\mathbb{E} A_x(t, x)^{k-1} A_{xx}(t, x) \widetilde{V}(t, x) &\leq D(t, x)\\
&\leq e^{mA(a,0) + m|x|}\mathbb{E}|z| e^{m|z|\sqrt{b - a}},
\end{aligned}
$$

where the first inequality used the fact that $k + 1$ is even.

Next, we verify (2.17). Note that from (2.12) and (2.13), the dominated convergence theorem implies

$$\lim_{t \to b-} D(t, x) = \frac{\mathbb{E} z \, \text{sign}(x + z\sqrt{b-a}) e^{m|x+z\sqrt{b-a}|}}{\mathbb{E} e^{m|x+z\sqrt{b-a}|}} = \sqrt{b-a}\big(\Delta(x) + m\big),$$

where the second equation used the fact that

(2.19)
$$\mathbb{E} z \, \text{sign}(x + z\sqrt{b-a}) e^{m|x+z\sqrt{b-a}|}$$
$$= \frac{2}{\sqrt{2\pi}} e^{-\frac{x^2}{2(b-a)}} + m\sqrt{b-a}\, \mathbb{E} e^{m|x+z\sqrt{b-a}|}.$$

See the verification of this equation in Lemma A.1 in the appendix. In addition, since $k + 1$ is even, (2.14) yields

$$\lim_{t \to b-} \mathbb{E} A_x(t, x)^{k+1} \widetilde{V}(t, x) = 1.$$

Thus, from (2.18) and the last two limits,

$$\Delta(x)\sqrt{b-a} + m\sqrt{b-a}$$
$$= \lim_{t \to b-} D(t, x) = \sqrt{b-a}\big(k \lim_{t \to b-} \mathbb{E} A_x(t, x)^{k-1} A_{xx}(t, x) \widetilde{V}(t, x) + m\big),$$

from which (2.17) follows. □

LEMMA 2.7. *We have that*

$$\liminf_{t \to b-} \mathbb{E} A_{xx}(t, x)^2 \widetilde{V}(t, x) \geq \frac{4m'}{3}\Delta(x).$$

PROOF. Recall the middle equation of (2.15). We see that

(2.20)
$$A_{xx}(t, x) = m'(1 - A_x(t, x)^2) + 2\widetilde{\Gamma}(t, x)$$

on $[a, b) \times \mathbb{R}$, where $\widetilde{\Gamma}(t, x) := \Gamma(t, x + z\sqrt{b-a})$. Using (2.14), (2.20), and Lemma 2.6 with $k = 1$ gives

$$\lim_{t \to b-} \mathbb{E} \widetilde{\Gamma}(t, x) \widetilde{V}(t, x) = \frac{1}{2}\Delta(x).$$

Also, multiplying both sides of (2.20) by $A_x(t, x)^2$ and applying (2.14) and Lemma 2.6 with $k = 3$ yields

$$\lim_{t \to b-} \mathbb{E} A_x(t, x)^2 \widetilde{\Gamma}(t, x) \widetilde{V}(t, x) = \frac{1}{6}\Delta(x).$$

From (2.20), since

$$A_{xx}(t, x)^2 = \big(m'(1 - A_x(t, x)^2) + 2\widetilde{\Gamma}(t, x)\big)^2$$
$$\geq [m'(1 - A_x(t, x)^2)]^2 + 4m'(1 - A_x(t, x)^2)\widetilde{\Gamma}(t, x),$$

the announced result follows by the last two limits. □

PROOF OF PROPOSITION 2.4. Statement (2.9) follows from (2.6) and (2.14). From (2.8), (2.14), Lemma 2.7, and (2.17), we find

$$\liminf_{t \to b-} \partial_t C(t, x)$$

$$\geq \liminf_{t \to b-} \mathbb{E}\big(A_{xx}(t, x)^2 + 2(m - m')A_{xx}(t, x)A_x(t, x)^2\big)\tilde{V}(t, x)$$

$$\geq \liminf_{t \to b-} \mathbb{E}A_{xx}(t, x)^2 \tilde{V}(t, x) + 2(m - m') \lim_{t \to b-} \mathbb{E}A_{xx}(t, x)A_x(t, x)^2 \tilde{V}(t, x)$$

$$= \frac{4m'}{3}\Delta(x) + \frac{2(m - m')}{3}\Delta(x)$$

$$= \frac{2(m + m')}{3}\Delta(x),$$

which completes the proof.   □

2.3 Proof of Theorem 2.1

Recall the sequences $(q_i)_{0 \leq i \leq n+2}$ and $(m_i)_{0 \leq i \leq n+1}$ from (2.1) and (2.3). Recall the quantities m, m', a, b and the functions A, B, C, V from (2.6). From now on, we take

$$m = m_n, \qquad m' = m_{n+1},$$
$$a = \xi'(q_n), \qquad b = \xi'(1),$$

and let

$$\hat{A}(q, x) = A(\xi'(q), x), \qquad \hat{B}(q, x) = B(\xi'(q), x),$$
$$\hat{C}(q, x) = C(\xi'(q), x), \qquad \hat{V}(q, x, y) = V(\xi'(q), x, y).$$

For $0 \leq i \leq n$, set

$$W_i = \exp m_i(Y_{i+1} - Y_i),$$

where Y_i are given in (2.4). Denote

$$Z = h + \sum_{j=0}^{n-1} z_j \sqrt{\xi'(q_{j+1}) - \xi'(q_j)},$$

and

$$\phi(q) = \mathbb{E}W_0 \cdots W_{n-1}\hat{C}(q, Z).$$

LEMMA 2.8. *We have that*

$$(2.21) \qquad \partial_q \mathscr{P}(\gamma_q) = \frac{\xi''(q)}{2}(m_{n+1} - m_n)\big(q - \phi(q)\big)$$

and

$$\phi'(q) = \frac{\xi''(q)(m_n - m_{n+1})}{2}$$

$$\cdot \sum_{i=0}^{n-1} m_i \mathbb{E}\Big[W_0 \cdots W_i \big(\mathbb{E}_{i+1}\big[W_{i+1} \cdots W_{n-1}\widehat{C}(q, Z)\big]\big)^2\Big]$$

(2.22)
$$- \frac{\xi''(q)(m_n - m_{n+1})}{2}$$

$$\cdot \sum_{i=0}^{n-1} m_i \mathbb{E}\Big[W_0 \cdots W_i \mathbb{E}_{z_i}\big[W_i \big(\mathbb{E}_{i+1}\big[W_{i+1} \cdots W_{n-1}\widehat{C}(q, Z)\big]\big)^2\big]\Big]$$

$$+ \mathbb{E}\, W_0 \cdots W_{n-1}\widehat{C}_q(q, Z),$$

where $\mathbb{E}_i$ is the expectation with respect to $z_i, \ldots, z_{n-1}$ and $\widehat{C}_q$ is the partial derivative with respect to q.

PROOF. Observe that for $0 \le i \le n - 1$,

$$\partial_q Y_i = \mathbb{E}_{z_i} W_i \partial_q Y_{i+1}.$$

An induction argument yields

$$\partial_q Y_i = \mathbb{E}_i W_i \cdots W_{n-1} \partial_q Y_n$$

for $0 \le i \le n - 1$. Since $Y_n = \widehat{B}(q, Z)$, the equation (2.7) leads to

(2.23)
$$\partial_q Y_i = \frac{\xi''(q)(m_n - m_{n+1})}{2} \mathbb{E}_i W_i \cdots W_{n-1}\widehat{C}(q, Z).$$

From (2.5), since

$$\partial_q \mathscr{P}(\gamma_q) = \partial_q Y_0 + \frac{q\xi''(q)}{2}(m_{n+1} - m_n),$$

this and (2.23) with $i = 0$ yield (2.21). On the other hand, for $0 \le i \le n - 1$, from (2.23),

$$\partial_q W_i = m_i \big(\partial_q Y_{i+1} - \partial_q Y_i\big) W_i$$

$$= \frac{\xi''(q)}{2}(m_n - m_{n+1})m_i W_i \big(\mathbb{E}_{i+1} W_{i+1} \cdots W_{n-1}\widehat{C}(q, Z)$$

$$- \mathbb{E}_i W_i \cdots W_{n-1}\widehat{C}(q, Z)\big)$$

where $\mathbb{E}_{i+1} W_{i+1} \cdots W_{n-1}\widehat{C}(q, Z) = \widehat{C}(q, Z)$ if $i = n - 1$. Finally, since

$$\phi'(q) = \sum_{i=0}^{n-1} \mathbb{E}\, W_0 \cdots W_{i-1}(\partial_q W_i) W_{i+1} \cdots W_{n-1}\widehat{C}(q, Z)$$

$$+ \mathbb{E}\, W_0 \cdots W_{n-1}\widehat{C}_q(q, Z),$$

plugging the last equation into this derivative yields (2.22). □

PROOF OF THEOREM 2.1. Recall $\phi'(q)$ from (2.22). Let $\widetilde{W}_0, \ldots, \widetilde{W}_{n-1}$ be $W_0, \ldots, W_{n-1}$ evaluated at $q = 1$. Note that $\mathbb{E}_{z_i} W_i = 1$ for all $0 \le i \le n-1$, $\lim_{q \to 1-} \widehat{C}(q, Z) = 1$ by (2.9), and $|\widehat{C}(q, Z)| \le 1$ by (2.12). Applying Fatou's lemma and conditional expectation yields that the first two lines of (2.22) cancel each other and as a result of Proposition 2.4,

$$
\begin{aligned}
\liminf_{q \to 1-} \phi'(q) &= \liminf_{q \to 1-} \mathbb{E} W_0 \cdots W_{n-1} \widehat{C}_q(q, Z) \\
&\ge \mathbb{E} \widetilde{W}_0 \cdots \widetilde{W}_{n-1} \liminf_{q \to 1-} \widehat{C}_q(q, Z) \\
&\ge \frac{2\xi''(1)(m_n + m_{n+1})}{3} \mathbb{E} \widetilde{W}_0 \cdots \widetilde{W}_{n-1} \Delta(Z),
\end{aligned}
$$

(2.24)

where $\Delta(Z)$ is defined through (2.10) with $a = \xi'(q_n)$, $b = \xi'(1)$, and $m = m_n$. We emphasize that although we do not know whether $\widehat{C}_q$ is nonnegative (see (2.8)), the use of Fatou's lemma remains justifiable. Indeed, note that $|\widehat{A}_x| \le 1$, $\mathbb{E}_{z_n} \widehat{V} = 1$, and by (2.16),

$$
\begin{aligned}
0 &\le \mathbb{E}_{z_n} \widehat{A}_{xx}(q, Z) \widehat{A}_x^2(q, Z) \widehat{V}(q, Z, Z + z_n \sqrt{\xi'(q) - \xi'(q_n)}) \\
&\le \frac{K e^{m_n |Z|}}{\sqrt{\xi'(q) - \xi'(q_n)}},
\end{aligned}
$$

where K is a constant independent of q. From (2.8),

$$
\widehat{C}_q(q, Z) \ge -(m_{n+1} - m_n)\xi''(q) \left(\frac{2K e^{m_n |Z|}}{\sqrt{\xi'(q) - \xi'(q_n)}} + \frac{m_n}{2} \right).
$$

In addition, it can be shown that each $\ln(W_0 W_1 \cdots W_{n-1})$ is at most of linear growth in $z_0, \ldots, z_{n-1}$ following from the fact that each Y_i is uniformly Lipschitz in the variable z_i for all $q \in [q_n, 1]$. This and the last inequality together validate (2.24).

Next, from (2.24), we can choose m_{n+1} large enough in the beginning such that

$$
\liminf_{q \to 1-} \phi'(q) > 1.
$$

Note that $\lim_{q \to 1-} \phi(q) = 1$. From (2.21), the above inequality implies that $\partial_q \mathscr{P}(\gamma_q) < 0$ for our choice of m_{n+1} as long as q is sufficiently close to 1. This completes our proof. □

Remark 2.9. The validity of (2.24) and Theorem 2.1 relies on the positive lower bound of $\partial_t C$ coming from Proposition 2.4. When one looks at (2.8) together with the fact $\lim_{t \to b-} (\partial_x A)^2 = 1$, it is tempting to think that $\partial_t C$ is actually negative since $m' = m_{n+1}$ is taken to be large. As a result, Proposition 2.4 may look counterintuitive. The remedy for this puzzle is the fact that $\partial_{xx} A(t, x)$ is singular in the limit $t \to b-$ and the dominated convergence theorem does not apply. These

"singular expectations" are one of the major difficulties in proving Theorem 2.1. They are handled by the exact computations coming from Lemmas 2.6 and 2.7.

Remark 2.10. In the work [42], Toninelli proved that in the SK model, the Parisi formula for the free energy possesses RSB Parisi measure when the temperature stays below the de Almeida–Thouless transition line. In view of the argument therein, he considered the first derivative of the Parisi functional $\mathscr{P}_\beta(\gamma_q)$ in the variable m_{n+1} and then sent $m_{n+1} \to 1$ for $n = 0$ to obtain a quantity $K(\beta, h, q)$. He then concluded the proof by calculating the first and second derivatives of $K(\beta, h, q)$ w.r.t. q near $\overline{q}$, where $\overline{q}$ is the unique solution of $\overline{q} = \int d\mu(z) \tanh^2(z\beta\sqrt{\overline{q}} + \beta h)$ for $z \sim N(0, 1)$. In our situation, the maximum value of m_{n+1} is unclear since γ is not a probability measure. Our argument considers the first and second derivatives of the perturbed Parisi functional $\mathscr{P}(\gamma_q)$ in q and evaluated at $q = 1$, which will not work for the positive temperature case because $\alpha_{P,\beta}([0, x])$ is less than or equal to 1 for x being close to 1.

3 Proofs of Main Results

PROOF OF THEOREM 1.1. We prove Theorem 1.1 by contradiction. First, note that it is known by [12, theorem 6] that the Parisi measure γ_P is not constantly 0. Suppose that the support of γ_P consists of only $n \geq 1$ points. Then from Theorem 2.1, we can lower the value of the Parisi functional by a perturbation of γ_P at 1 defined in (2.2). This leads to a contradiction of the minimality of $\mathscr{P}(\gamma_P)$. Hence, the support of γ_P must contain infinitely many points. □

Remark 3.1. The statement of Theorem 1.1 can be strengthened to the fact that the Parisi measure γ_P cannot be "flat" near 1, i.e., $\gamma_P(t) < \gamma_P(1-)$ for any $0 < t < 1$. In fact, if this is not true, then γ_P is a constant function on $[a, 1)$ for some a. One can then apply essentially the same argument as Proposition 2.1 to lower the Parisi functional. The only difference is that since γ_P is not necessarily a step function on $[0, a)$, the terms $W_1 \cdots W_{n-1}$ in Lemma 2.8 have to be replaced by a continuous modification using the optimal stochastic control representation for Ψ_γ in [12]. We omit the details of the argument.

Remark 3.2. Our argument of Theorem 1.1 does not rely on the uniqueness of the Parisi measure. All we need is the existence of *a* Parisi measure, which was proved in [8].

PROOF OF PROPOSITION 1.3. For any $u \in [-1, 1]$ and $\varepsilon > 0$, consider the coupled free energy and the maximum energy

$$CF_{N,\beta}(u, \varepsilon) := \frac{1}{\beta N} \log \sum_{|R_{1,2}-u|<\varepsilon} e^{\beta H_N(\sigma^1)+\beta H_N(\sigma^2)}$$

and

$$CM_N(u, \varepsilon) := \frac{1}{N} \max_{|R_{1,2}-u|<\varepsilon} \left(H_N(\sigma^1) + H_N(\sigma^2)\right).$$

First we claim that if u is in the support of the Parisi measure $\alpha_{P,\beta}(dt)$, then for any $\varepsilon > 0$,

$$(3.1) \qquad \lim_{N \to \infty} \mathbb{E}CF_{N,\beta}(u, \varepsilon) = 2F_\beta.$$

Note that the expectation of the free energy $F_{N,\beta}$ and the coupled free energy $CF_{N,\beta}$ are Lipschitz functions in the mixture parameters $(c_p)_{p \geq 2}$ with respect to the $\ell_2(\mathbb{R})$-norm. In addition, the Parisi measure is continuous in β as well as the mixture parameters $(c_p)_{p \geq 2}$ (see, e.g., [7]). From these, we may assume without loss of generality that $c_p > 0$ for all $p \geq 2$. This condition guarantees that the Parisi measure $\alpha_{P,\beta}$ is the limiting distribution of the overlap $R_{1,2}$ between two i.i.d. samplings σ^1 and σ^2 from G_N; see [26, 39]. Namely, we have

$$\lim_{N \to +\infty} \mathbb{E}G_N^{\otimes 2}(R_{1,2} \in B) = \alpha_{P,\beta}(B)$$

for any Borel set $B \subset [0, 1]$. To see why the claim (3.1) holds, observe that if there exists a positive δ such that

$$\limsup_{N \to \infty} \mathbb{E}CF_{N,\beta}(u, \varepsilon) \leq 2F_\beta - \delta,$$

it follows by the Gaussian concentration of measure that there exists a constant $K > 0$ such that for any $N \geq 1$, with probability at least $1 - Ke^{-N/K}$,

$$CF_{N,\beta}(u, \varepsilon) \leq 2F_{N,\beta} - \frac{\delta}{2}.$$

Multiplying N and taking the exponential leads to

$$\mathbb{E}G_N^{\otimes 2}\big(R_{1,2} \in (u - \varepsilon, u + \varepsilon)\big) \leq Ke^{-N/K} + e^{-N\beta\delta/2},$$

which implies, by the weak convergence of $R_{1,2}$,

$$\int_{(u-\varepsilon, u+\varepsilon)} \alpha_{P,\beta}(dt) = 0.$$

This contradicts the fact that u is in the support of the Parisi measure and completes the proof of our claim.

Next, note that evidently

$$\mathbb{E}CM_N(u, \varepsilon) \leq \mathbb{E}CF_{N,\beta}(u, \varepsilon) \leq \frac{\log 4}{\beta} + \mathbb{E}CM_N(u, \varepsilon).$$

This together with our claim implies that

$$\lim_{N \to \infty} \mathbb{E}CM_N(u, \varepsilon) = 2 \lim_{\beta \to \infty} F_\beta = 2\text{GSE}.$$

From the Gaussian concentration of measure inequality, for any $\eta > 0$, there exists a constant $K' > 0$ such that for any $N \geq 1$ with probability at least $1 - K'e^{-N/K'}$,

$$CM_N(u, \varepsilon) \geq 2\text{GSE} - \frac{\eta}{2}$$

and

$$\max_{\sigma \in \Sigma_N} \frac{H_N(\sigma)}{N} \le \text{GSE} + \frac{\eta}{4}.$$

From the first inequality, there exist σ^1 and σ^2 with $|R_{1,2} - u| < \varepsilon$ and

$$\frac{1}{N}\big(H_N(\sigma^1) + H_N(\sigma^2)\big) \ge 2\text{GSE} - \frac{\eta}{2}.$$

If either $H_N(\sigma^1) < \text{GSE} - \eta$ or $H_N(\sigma^2) < \text{GSE} - \eta$, then we arrive at a contradiction,

$$2\text{GSE} - \frac{\eta}{2} \le \frac{1}{N}\big(H_N(\sigma^1) + H_N(\sigma^2)\big) \le \text{GSE} - \eta + \text{GSE} + \frac{\eta}{4} = 2\text{GSE} - \frac{3\eta}{4}.$$

Therefore, the inequality (1.5) must hold and this finishes our proof. □

PROOF OF THEOREM 1.6. Recall the Parisi measure $\alpha_{P,\beta}$ for the free energy from (1.2). We first claim that $(\beta\alpha_{P,\beta})_{\beta>0}$ converges to γ_P vaguely on $[0, 1)$. Suppose there exists an infinite sequence $(\beta_l)_{l\ge 1}$ such that $(\beta_l\alpha_{P,\beta_l})_{l\ge 1}$ does not converge to γ_P vaguely on $[0, 1)$. By an identical argument to [8, eq. (16)], we can further pass to a subsequence of $(\beta_l\alpha_{P,\beta_l})_{l\ge 1}$ such that it vaguely converges to some γ on $[0, 1)$. To ease our notation, we use $(\beta_l\alpha_{P,\beta_l})_{l\ge 1}$ to stand for this subsequence. It was established in [8, lemma 3] that

$$\lim_{l\to\infty} F_{\beta_l} \ge \mathscr{P}(\gamma).$$

From this,

$$\mathscr{P}(\gamma_P) = \lim_{\beta\to\infty} F_\beta \ge \mathscr{P}(\gamma).$$

From the uniqueness of γ_P shown in [12, theorem 4], it follows that $\gamma_P = \gamma$, a contradiction. Thus, $(\beta\alpha_{P,\beta})_{\beta>0}$ converges to γ_P vaguely on $[0, 1)$. This completes the proof of our claim.

Next, if Theorem 1.6 does not hold, then from the above claim, there exists some $k \ge 1$ such that the support of $\alpha_{P,\beta}$ contains at most k points for all sufficiently large β. This implies that the support of γ_P contains at most k points. This contradicts Theorem 1.1. □

Appendix

Denote by $\Phi(x)$ the c.d.f. of the standard normal distribution. The following lemma follows a standard Gaussian computation and is used in (2.19).

LEMMA A.1. *Suppose that z is a standard normal random variable. For any $x, m \in \mathbb{R}$ and $a > 0$,*

$$\mathbb{E}z e^{m|x+az|}\text{sign}(x + az) = \sqrt{\frac{2}{\pi}} e^{-\frac{x^2}{2a^2}} + ma\mathbb{E}e^{m|x+az|}.$$

PROOF. Define

$$f(y) = e^{my} \Phi\left(\frac{y}{a} + ma\right),$$

for $y \in \mathbb{R}$, where $\Phi(x)$ is the c.d.f. of the standard Gaussian random variable. Note that

$$amz - \frac{z^2}{2} = -\frac{1}{2}(z - ma)^2 + \frac{m^2a^2}{2}.$$

Computing directly gives

$$\mathbb{E}ze^{m(x+az)}\mathbb{1}_{\{x+az>0\}} = \frac{e^{mx}}{\sqrt{2\pi}} \int_{-x/a}^{\infty} ze^{amz-\frac{z^2}{2}} \, dz.$$

$$= \frac{e^{mx+\frac{m^2a^2}{2}}}{\sqrt{2\pi}} \int_{-x/a}^{\infty} (z - ma)e^{-\frac{(z-ma)^2}{2}} \, dz$$

$$+ \frac{mae^{mx+\frac{m^2a^2}{2}}}{\sqrt{2\pi}} \int_{-x/a}^{\infty} e^{-\frac{(z-ma)^2}{2}} \, dz$$

$$= \frac{e^{-\frac{x^2}{2a^2}}}{\sqrt{2\pi}} + mae^{\frac{m^2a^2}{2}} f(x).$$

On the other hand, since $z' := -z$ is a standard Gaussian random variable, we may apply the above formula to obtain

$$\mathbb{E}ze^{-m(x+az)}\mathbb{1}_{\{x+az<0\}} = -\mathbb{E}(-z)e^{m((-x)+a(-z))}\mathbb{1}_{\{-x+a(-z)>0\}}$$

$$= -\mathbb{E}z'e^{m((-x)+az')}\mathbb{1}_{\{-x+az'>0\}}$$

$$= -\frac{e^{-\frac{x^2}{2a^2}}}{\sqrt{2\pi}} - mae^{\frac{m^2a^2}{2}} f(-x).$$

Combining these two equations leads to

$$\mathbb{E}ze^{m(x+az)}\mathbb{1}_{\{x+az>0\}} - \mathbb{E}ze^{-m(x+az)}\mathbb{1}_{\{x+az<0\}}$$

$$= \sqrt{\frac{2}{\pi}}e^{-\frac{x^2}{2a^2}} + mae^{\frac{m^2a^2}{2}} \big(f(x) + f(-x)\big).$$

Here, note that

$$\mathbb{E}e^{m|x+az|} = e^{\frac{m^2a^2}{2}+xm} \Phi\left(\frac{x}{a} + am\right)$$

$$+ e^{\frac{m^2a^2}{2}-xm} \Phi\left(-\frac{x}{a} + am\right) = e^{\frac{m^2a^2}{2}} \big(f(x) + f(-x)\big).$$

This and the last equation together imply the announced result. $\square$

Bibliography

[1]   Aizenman, M.; Lebowitz, J. L.; Ruelle, D. Some rigorous results on the Sherrington-Kirkpatrick spin glass model. *Comm. Math. Phys.* **112** (1987), no. 1, 3–20.

[2]   Aizenman, M.; Sims, R.; Starr, S. L. An extended variational principle for the SK spin-glass model. *Phys. Rev. B.* **68** (2003), no. 21, 214403. doi:10.1103/PhysRevB.68.214403

[3]   Auffinger, A.; Ben Arous, G. Complexity of random smooth functions on the high-dimensional sphere. *Ann. Probab.* **41** (2013), no. 6, 4214–4247. doi:10.1214/13-AOP862

[4]   Auffinger, A.; Ben Arous, G.; Černý, J. Random matrices and complexity of spin glasses. *Comm. Pure Appl. Math.* **66** (2013), no. 2, 165–201. doi:10.1002/cpa.21422

[5]   Auffinger, A.; Chen, W.-K. On properties of Parisi measures. *Probab. Theory Related Fields* **161** (2015), no. 3-4, 817–850. doi:10.1007/s00440-014-0563-y

[6]   Auffinger, A.; Chen, W.-K. The Parisi formula has a unique minimizer. *Comm. Math. Phys.* **335** (2015), no. 3, 1429–1444. doi:10.1007/s00220-014-2254-z

[7]   Auffinger, A.; Chen, W.-K. The Legendre structure of the Parisi formula. *Comm. Math. Phys.* **348** (2016), no. 3, 751–770. doi:10.1007/s00220-016-2673-0

[8]   Auffinger, A.; Chen, W.-K. Parisi formula for the ground state energy in the mixed p-spin model. *Ann. Probab.* **45** (2017), no. 6B, 4617–4631. doi:10.1214/16-AOP1173

[9]   Auffinger, A.; Chen, W.-K. On the energy landscape of spherical spin glasses. *Adv. Math.* **330** (2018), 553–588. doi:10.1016/j.aim.2018.03.028

[10]   Auffinger, A.; Zeng, Q. Existence of two-step replica symmetry breaking for the spherical mixed p-spin glass at zero temperature. *Comm. Math. Phys.* **370** (2019), no. 1, 377–402. doi:10.1007/s00220-018-3252-3

[11]   Ben Arous, G.; Subag, E.; Zeitouni, O. Geometry and temperature chaos in mixed spherical spin glasses at low temperature—the perturbative regime. Preprint, 2018. arXiv:1804.10573 [math.PR]

[12]   Chen, W.-K.; Handschy, M.; Lerman, G. On the energy landscape of the mixed even p-spin model. *Probab. Theory Related Fields* **171** (2018), no. 1-2, 53–95. doi:10.1007/s00440-017-0773-1

[13]   Dembo, A.; Montanari, A.; Sen, S. Extremal cuts of sparse random graphs. *Ann. Probab.* **45** (2017), no. 2, 1190–1217. doi:10.1214/15-AOP1084

[14]   Edwards, S. F.; Anderson, P. W. Theory of spin glasses. *J. Phys. F: Metal Physics* **5** (1975), no. 5, 965. doi:10.1088/0305-4608/5/5/017

[15]   Ghirlanda, S.; Guerra, F. General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity. *J. Phys. A* **31** (1998), no. 46, 9149–9155. doi:10.1088/0305-4470/31/46/006

[16] Guerra, F. Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.* **233** (2003), no. 1, 1–12. doi:10.1007/s00220-002-0773-5

[17] Jagannath, A.; Tobasco, I. A dynamic programming approach to the Parisi functional. *Proc. Amer. Math. Soc.* **144** (2016), no. 7, 3135–3150. doi:10.1090/proc/12968

[18] Jagannath, A.; Tobasco, I. Low temperature asymptotics of spherical mean field spin glasses. *Comm. Math. Phys.* **352** (2017), no. 3, 979–1017. doi:10.1007/s00220-017-2864-3

[19] Jagannath, A.; Tobasco, I. Some properties of the phase diagram for mixed *p*-spin glasses. *Probab. Theory Related Fields* **167** (2017), no. 3-4, 615–672. doi:10.1007/s00440-015-0691-z

[20] Mézard, M.; Montanari, A. *Information, physics, and computation*. Oxford University Press, Oxford, 2009..

[21] Mézard, M.; Parisi, G.; Virasoro, M. *Spin glass theory and beyond.* World Scientific Lecture Notes in Physics, 9. World Scientific, Teaneck, N.J., 1987.

[22] Montanari, A.; Ricci-Tersenghi, F. On the nature of the low-temperature phase in discontinuous mean-field spin glasses. *Euro. Phys. J. B* **33** (2003), no. 3, 339–346. doi:10.1140/epjb/e2003-00174-7

[23] Oppermann, R.; Schmidt, M. J.; Sherrington, D. Double criticality of the Sherrington-Kirkpatrick model at $T = 0$. *Phys. Rev. Lett.* **98** (2007), no. 12, 127201. doi:10.1103/PhysRevLett.98.127201

[24] Oppermann, R.; Sherrington, D. Scaling and renormalization group in replica-symmetry-breaking space: evidence for a simple analytical solution of the Sherrington-Kirkpatrick model at zero temperature. *Phys. Rev. Lett.* **95** (2005), no. 19, 197203. doi:10.1103/PhysRevLett.95.197203

[25] Panchenko, D. The Parisi ultrametricity conjecture. *Ann. of Math. (2)* **177** (2013), no. 1, 383–393. doi:10.4007/annals.2013.177.1.8

[26] Panchenko, D. *The Sherrington-Kirkpatrick model.* Springer Monographs in Mathematics. Springer, New York, 2013. doi:10.1007/978-1-4614-6289-7

[27] Panchenko, D. The Parisi formula for mixed *p*-spin models. *Ann. Probab.* **42** (2014), no. 3, 946–958. doi:10.1214/12-AOP800

[28] Parisi, G. Infinite number of order parameters for spin-glasses. *Phys. Rev. Lett.* **43** (1979), no. 23, 1754–1756. doi:10.1103/PhysRevLett.43.1754

[29] Parisi, G. A sequence of approximate solutions to the SK model for spin glasses. *J. Phys. A* **13** (1980), no. 4, L–115.

[30] Parisi, G. The order parameter for spin glasses: A function on the interval 0-1. *J. Phys. A* **13** (1980), no. 3, 1101. doi:10.1088/0305-4470/13/3/042

[31] Parisi, G. Order parameter for spin-glasses. *Phys. Rev. Lett.* **50** (1983), no. 24, 1946–1948. doi:10.1103/PhysRevLett.50.1946

[32] Parisi, G. *Field theory, disorder and simulations.* World Scientific Lecture Notes in Physics, 49. World Scientific, River Edge, N.J., 1992. doi:10.1142/9789812799050

[33] Sen, S. Optimization on sparse random hypergraphs and spin glasses. *Random Structures Algorithms* **53** (2018), no. 3, 504–536. doi:10.1002/rsa.20774

[34] Sherrington, D.; Kirkpatrick, S. Solvable model of a spin-glass. *Phys. Rev. Lett.* **35** (1975), no. 26, 1792–1796. doi:10.1103/PhysRevLett.35.1792

[35] Subag, E. The geometry of the Gibbs measure of pure spherical spin glasses. *Invent. Math.* **210** (2017), no. 1, 135–209. doi:10.1007/s00222-017-0726-4

[36] Subag, E. Free energy landscapes in spherical spin glasses. Preprint, 2018. arXiv:1804.10576 [math.PR]

[37] Talagrand, M. *Spin glasses: a challenge for mathematicians. Cavity and mean field models*. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics, 46. Springer, Berlin, 2003.

[38] Talagrand, M. Free energy of the spherical mean field model. *Probab. Theory Related Fields* **134** (2006), no. 3, 339–382. doi:10.1007/s00440-005-0433-8

[39] Talagrand, M. Parisi measures. *J. Funct. Anal.* **231** (2006), no. 2, 269–286. doi:10.1016/j.jfa.2005.03.001

[40] Talagrand, M. The Parisi formula. *Ann. of Math. (2)* **163** (2006), no. 1, 221–263. doi:10.4007/annals.2006.163.221

[41] Talagrand, M. *Mean field models for spin glasses. Volume II. Advanced replica-symmetry and low temperature.* Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics, 55. Springer, Heidelberg, 2011.

[42] Toninelli, F. About the Almeida-Thouless transition line in the Sherrington-Kirkpatrick mean field spin glass model. *Europhysics Letters* **60** (2002), no. 5, 764–767. doi:10.1209/epl/i2002-00667-5

ANTONIO AUFFINGER
Department of Mathematics
Northwestern University
2033 N. Sheridan Rd
Evanston, IL 60613
USA
E-mail: `tuca@northwestern.edu`

WEI-KUO CHEN
School of Mathematics
University of Minnesota
USA
E-mail: `wkchen@umn.edu`

QIANG ZENG
Department of Mathematics
Northwestern University
2033 N. Sheridan Rd
Evanston, IL 60613
USA
E-mail: `qzeng@`
`math.northwestern.edu`