THE MOMENT GEOMETRY OF THE PARISI MEASURE

NAHOM SEYOUM, SEKHAR TATIKONDA

Abstract. Let μ^* be the Parisi measure of the Sherrington–Kirkpatrick model and $\mathcal{M}_K$ the set of probability measures on $[0,1]$ sharing its first K moments. We prove that $\mathcal{M}_K$ is convex and weakly compact with extreme points of support size $\leq K+1$, and that the filtration $\mathcal{M}_1 \supset \mathcal{M}_2 \supset \cdots$ collapses to $\{\mu^*\}$ at the sharp threshold $K = 2m$ when μ^* is m-atomic. The Parisi free energy varies by at most $4\beta^2/(K+1)$ across $\mathcal{M}_K$, giving a rate $O(\beta^2/K)$ for the K-step RSB approximation. The Parisi functional decomposes as $\mathcal{P}_\beta(\mu) = \log 2 - \frac{\beta^2}{4}(1-m_2) + \Phi_\mu(0,0)$; expanding $\Phi_\mu(0,0)$ in β^2, the β^6 coefficient is $m_3/3 - \mathcal{R}(\mu)$ with $\mathcal{R} \geq 0$ vanishing exactly on point masses, and the sign change of the m_2 coefficient at $\beta = 1$ recovers the de Almeida–Thouless instability. The Ghirlanda–Guerra identities impose no constraint on the single-overlap marginal.

1. Introduction

The Parisi solution of the Sherrington–Kirkpatrick model is a statement about a single probability measure on $[0,1]$. This measure μ^*, the limiting law of the overlap $R_{12} = \frac{1}{N}\sum_i \sigma_i^1 \sigma_i^2$ between two independent samples from the Gibbs measure, determines the free energy via a variational functional $\mathcal{P}_\beta$, and thereby every macroscopic thermodynamic observable. The Parisi measure is, in this sense, the complete solution of the model.

The measure μ^* is hard to pin down. Depending on temperature, it can be a point mass (replica-symmetric), finitely atomic (k-step RSB), or a mixed singular-continuous object with no closed-form expression (full RSB). Yet in every concrete situation where one wants to use μ^*, one uses only finitely many of its moments. The second moment m_2 determines the variance of the free energy. The first few moments control the high-temperature expansion, K moments determine the K-step RSB approximation. In each case, one works with a finite moment truncation, not with μ^* itself.

This motivates the question. What do K overlap moments tell us about μ^* and the free energy? The moments restrict μ^* to the *moment class*

$$\mathcal{M}_K \;=\; \Big\{\nu \in \mathcal{P}([0,1]) \;:\; \textstyle\int q^k\, d\nu = m_k \text{ for } k = 1,\dots,K\Big\}.$$

Date: April 2026.

We study the geometry of $\mathcal{M}_K$ and the rate at which thermodynamic quantities are determined as $K \to \infty$. The overlap moments turn out to be an efficient finite-dimensional proxy for μ^*, at least as far as the free energy is concerned.

1.1. Setup.

The SK model [SK75] has configuration space $\Sigma_N = \{-1, +1\}^N$, i.i.d. Gaussian disorder $(J_{ij})_{1\le i<j\le N} \sim \mathcal{N}(0,1)$, and Hamiltonian

$$H_N(\sigma) \;=\; -\frac{1}{\sqrt{N}}\sum_{i<j} J_{ij}\,\sigma_i\sigma_j \;-\; h\sum_i \sigma_i. \tag{1}$$

At inverse temperature β, the partition function is $Z_N(\beta,h) = \sum_\sigma e^{-\beta H_N(\sigma)}$, and Guerra–Toninelli [GT02] established existence of the quenched free energy $f(\beta,h) = \lim_N \frac{1}{N}\mathbb{E}[\log Z_N]$. The overlap $R_{12} = \frac{1}{N}\sum_i \sigma_i^1\sigma_i^2$ of two replicas under the Gibbs measure converges in law (as $N\to\infty$) to a deterministic probability measure $\mu^* \in \mathcal{P}([0,1])$, the *Parisi measure.*

Definition 1.1 (Parisi functional)**.** Let $\xi(q) = \frac{\beta^2}{2}q^2$. For $\mu \in \mathcal{P}([0,1])$, the *Parisi PDE* is

$$\partial_s\Phi_\mu(s,x) \;=\; -\frac{\xi''(s)}{2}\Big[\partial_{xx}\Phi_\mu(s,x) + F(s)\,\big(\partial_x\Phi_\mu(s,x)\big)^2\Big], \qquad \Phi_\mu(1,x) \;=\; \log\cosh(x+\beta h), \tag{2}$$

on $[0,1]\times\mathbb{R}$, where $F(s) = \mu([0,s])$. The *Parisi functional* $\mathcal{P}_\beta : \mathcal{P}([0,1]) \to \mathbb{R}$ is

$$\mathcal{P}_\beta(\mu) \;=\; \log 2 + \Phi_\mu(0,0) - \frac{1}{2}\int_0^1 s\,\xi''(s)\,\mu([0,s])\,ds. \tag{3}$$

Theorem 1.2 (Parisi formula [Gue03, Tal06], unique minimizer [AC15])**.** *For every $\beta > 0$ and $h \ge 0$, $f(\beta,h) = \inf_{\mu\in\mathcal{P}([0,1])}\mathcal{P}_\beta(\mu)$, and the infimum is attained at a unique $\mu^* \in \mathcal{P}([0,1])$, the Parisi measure.*

When μ is finitely atomic, F is piecewise constant and the Parisi PDE reduces to a finite recursion of Gaussian convolutions (Section 4), $\mathcal{P}_\beta(\mu)$ is then the standard K-step RSB functional. The functional is Lipschitz in W_1.

$$|\mathcal{P}_\beta(\mu) - \mathcal{P}_\beta(\nu)| \;\le\; 4\beta^2\,W_1(\mu,\nu), \tag{4}$$

from the Auffinger–Chen stochastic control representation [AC15] and the maximum-principle bound $|\partial_x\Phi_\mu| \le 1$.

The overlap moments $m_k = \int q^k\,d\mu^*$ are physical observables. Under stochastic stability of the overlap array [GG98a, Pan13b], $m_k = \lim_N \mathbb{E}[\langle R_{12}^k\rangle]$. In particular $m_2 = \lim_N \mathbb{E}[\langle R_{12}^2\rangle]$ controls the variance of $F_N := \frac{1}{N}\log Z_N$ via the Gaussian Poincaré inequality,

$$\mathrm{Var}(F_N) \;\le\; \frac{\beta^2 m_2}{2N} + O(N^{-2})$$

(Appendix B).

1.2. **Relation to prior work.** The moment problem on $[0, 1]$ is a classical subject. The structural results about $\mathcal{M}_K$ we use, namely convexity, compactness, and atomic support of extreme points, go back to Richter [Ric57], Rogosinski [Rog58], and Karlin–Studden [KS66]. Our contribution is the connection to the Parisi theory. The extreme points of $\mathcal{M}_K$ are precisely the $\leq K$-step RSB-type marginals, and the rate of collapse of $\mathcal{M}_K$ gives an explicit convergence rate for the RSB hierarchy.

The Parisi functional has been studied variationally since [Par79, Par80]. Rigorous foundations are laid out in [Pan13b] and its references. The explicit–PDE decomposition of Section 4, separating a purely m_2-dependent piece from $\Phi_\mu(0, 0)$, does not appear in this literature, though the m_2-dependence of the leading terms has been noted in physics papers. The moment-theoretic rederivation of the de Almeida–Thouless instability [dT78] in Corollary 4.13 reads the instability off the sign of a coefficient in the β^2-expansion, bypassing the replica trick. Finally, Theorem 5.4 shows that the Ghirlanda–Guerra identities [GG98a, Pan13a] impose no constraint on the single-overlap marginal. Every $\mu \in \mathcal{P}([0, 1])$ is the marginal of a GG-compatible array (realized by the Ruelle cascade [Rue87]), so moments and GG act on independent aspects of the overlap structure.

Throughout, $\mathcal{P}([0, 1])$ carries the weak topology (equivalently W_1, since $[0, 1]$ is compact), $F(s) = \mu([0, s])$, $\xi(q) = \frac{\beta^2}{2} q^2$, $\langle \cdot \rangle$ is the Gibbs average, and $\mathbb{E}$ is expectation over the disorder.

2. Background

This section summarizes the Parisi theory of the SK model at the level of detail needed later. Readers familiar with [Pan13b, Tal11] can skim it for notation and skip to Section 3.

2.1. **The replica method and the overlap array.** The basic computational device is the identity $\log Z = \lim_{n \to 0} \frac{1}{n}(Z^n - 1)$, applied to the free energy

$$f_N(\beta, h) \;=\; \tfrac{1}{N}\, \mathbb{E}[\log Z_N] \;=\; \tfrac{1}{N} \lim_{n \to 0} \tfrac{1}{n} (\mathbb{E}[Z_N^n] - 1).$$

For integer $n \geq 1$, Z_N^n is the partition function of n non-interacting copies ("replicas") of the system, and the disorder average can be performed explicitly because the couplings J_{ij} are Gaussian. One obtains

$$\mathbb{E}[Z_N^n] \;=\; \sum_{\sigma^1, \ldots, \sigma^n \in \Sigma_N} \exp\Big(\tfrac{N\beta^2}{2} \sum_{\alpha, \beta = 1}^{n} R_{\alpha\beta}^2 - \tfrac{n\beta^2}{2} + N\beta h \sum_{\alpha} \tfrac{1}{N} \textstyle\sum_i \sigma_i^\alpha \Big),$$

where $R_{\alpha\beta} = \frac{1}{N} \sum_i \sigma_i^\alpha \sigma_i^\beta$ is the *overlap* between replicas α and β. The replicated free energy is a functional of the symmetric $n \times n$ *overlap matrix* $Q = (R_{\alpha\beta})$, and the $N \to \infty$ limit is controlled by the saddle-point value of this functional over matrices Q.

The replica method is formal. The analytic continuation $n \to 0$, the interchange of $N \to \infty$ and $n \to 0$, and the saddle-point optimization over a space whose dimension depends on n are all outside the scope of the method itself. The rigorous vindication of the replica calculation is the Parisi formula (Theorem 1.2), which came via entirely different arguments, due to Guerra [Gue03] and Talagrand [Tal06].

The probabilistic content of the method is the *overlap array*. Drawing replicas $\sigma^1, \sigma^2, \dots$ independently from the Gibbs measure at large N gives an exchangeable array $(R_{\alpha\beta})_{\alpha,\beta\geq 1}$ on $[-1,1]$, whose single-overlap marginal $\mathrm{Law}(R_{12})$ converges to the Parisi measure μ^* (with support in $[0,1]$ under spin-flip symmetrization). The full joint law of this array carries strictly more information than the marginal. That joint law is the subject of the Ghirlanda–Guerra theory discussed in Section 5.

2.2. **The replica-symmetric solution.** The simplest ansatz for the overlap matrix is the *replica-symmetric* (RS) choice $Q_{\alpha\beta} = q$ for all $\alpha \neq \beta$. Substituting this into the saddle-point functional and performing the $n \to 0$ continuation yields the *RS free energy*

$$f_{\mathrm{RS}}(\beta, h; q) \;=\; \log 2 + \tfrac{\beta^2}{4}(1-q)^2 + \mathbb{E}\big[\log\cosh(\beta h + \beta\sqrt{q}\,Z)\big], \qquad Z \sim \mathcal{N}(0,1). \tag{5}$$

The saddle-point equation $\partial f_{\mathrm{RS}}/\partial q = 0$ is

$$q \;=\; \mathbb{E}\big[\tanh^2(\beta h + \beta\sqrt{q}\,Z)\big]. \tag{6}$$

For $h > 0$ this has a unique solution $q_{\mathrm{RS}}(\beta, h) > 0$. For $h = 0$ the equation is $q = \mathbb{E}[\tanh^2(\beta\sqrt{q}\,Z)]$, which has only the trivial solution $q = 0$ for $\beta \leq 1$ and a pair of nontrivial solutions $\pm q_*(\beta)$ for $\beta > 1$.

At the level of the overlap distribution, the RS ansatz corresponds to $\mu^* = \delta_q$. Every pair of Gibbs samples has the same deterministic overlap q. In terms of the Gibbs measure, RS says the measure concentrates on a single cluster of configurations.

The RS ansatz is self-consistent as a saddle point but need not be the true saddle. Whether it is correct is decided by its stability.

2.3. **The de Almeida–Thouless instability.** Linearizing the replica saddle-point equations around the RS critical point, de Almeida and Thouless [dT78] computed the Hessian of the replica functional in the tangent space of overlap matrices and found a negative eigenvalue (the "replicon" mode) at low temperature. The boundary of the region where RS is stable is the *AT line*, defined by

$$\beta^2\,\mathbb{E}\big[\mathrm{sech}^4(\beta h + \beta\sqrt{q_{\mathrm{RS}}}\,Z)\big] \;=\; 1.$$

At $h = 0$ this gives $\beta_c = 1$. For $\beta < 1$ the RS ansatz is stable, and for $\beta > 1$ it is unstable along the replicon direction, signaling that the true saddle is not replica-symmetric.

We rederive this instability from the moment-theoretic viewpoint in Corollary 4.13. The coefficient of m_2 in the β^2-expansion of the Parisi functional is $\beta^2(1-\beta^2)/4$, which changes sign at $\beta = 1$, without any appeal to the replica trick or the Hessian.

2.4. **Parisi's hierarchical ansatz.** When RS fails, the correct overlap matrix is not a scalar but a *hierarchical Parisi matrix.* Fix integers $K \geq 0$, $n \geq 1$, and a decreasing sequence of block sizes $1 = m_0 \leq m_1 \leq \cdots \leq m_K \leq m_{K+1} = n$ (with $m_{i+1}/m_i \in \mathbb{N}$), together with an increasing sequence $0 \leq q_0 \leq q_1 \leq \cdots \leq q_K \leq 1$. These parameters define a chain of equivalence relations $\sim_0 \subset \sim_1 \subset \cdots \subset \sim_{K+1}$ on $\{1, \dots, n\}$, where $\sim_i$ identifies replicas in the same m_i-block. The Parisi matrix is

$$Q^{(K)}_{\alpha\beta} \;=\; q_i \quad \text{if } \alpha \sim_{i+1} \beta,\ \alpha \not\sim_i \beta,$$

with $Q^{(K)}_{\alpha\alpha} = 0$. For $K = 0$ this is the RS matrix q_0, and for $K = 1$ it has an inner block of size m_1 with overlap q_1 and an outer overlap q_0 between distinct blocks, with a nested $(K+1)$-level block structure for general K.

For integer n the parameters are combinatorial ($m_i \in \mathbb{N}$). The replica-method continuation to $n = 0$ reverses the ordering of the m_i, so in the $n \to 0$ convention

$$1 = m_0 \geq m_1 \geq \cdots \geq m_K \geq m_{K+1} = 0.$$

These m_i play the role of *Parisi parameters* (the "weights" of the RSB levels). A Parisi matrix with K distinct overlap values corresponds to a K-step RSB ansatz.

Encoding the ansatz as a single object, the parameters $(q_0, \dots, q_K; m_1, \dots, m_K)$ determine a non-decreasing step function

$$q^{(K)} : [0,1] \to [0,1], \qquad q^{(K)}(x) = q_j \text{ for } x \in [1 - m_j, 1 - m_{j+1}),$$

and the corresponding probability measure

$$\mu^{(K)} \;=\; \sum_{i=0}^{K} (m_i - m_{i+1})\, \delta_{q_i} \;\in\; \mathcal{P}([0,1])$$

is the push-forward of Lebesgue measure on $[0,1]$ under $q^{(K)}$. This is a $(K+1)$-atomic measure on $[0,1]$ whose atoms are the q_i and whose weights are the gaps $m_i - m_{i+1}$.

The K-step RSB free energy is the value $f^{(K)}(\beta, h) = \mathcal{P}_\beta(\mu^{(K)})$ of the Parisi functional (Definition 1.1) evaluated at this atomic measure. For general $\mu \in \mathcal{P}([0,1])$, $\mathcal{P}_\beta(\mu)$ is defined via the Parisi PDE (2). When μ is atomic, the PDE reduces to a finite recursion of Gaussian convolutions that we make explicit in Section 4.2 (Proposition 4.3).

2.5. **The Parisi PDE.** For an atomic measure $\mu = \sum_i w_i \delta_{q_i}$ with cumulative weights $\rho_i = \sum_{j \leq i} w_j$, the K-step RSB free energy emerges from an iteration of the form

$$f_K(x) = \log\cosh(x), \qquad f_i(x) = \tfrac{1}{\rho_i} \log \mathbb{E}\big[\exp(\rho_i\, f_{i+1}(x + \sigma_i Z))\big], \qquad \sigma_i^2 = \beta^2(q_{i+1} - q_i),$$

with $Z \sim \mathcal{N}(0,1)$. Each step combines a Gaussian convolution of variance $\beta^2(q_{i+1} - q_i)$ (heat flow) with the *tilted free energy* $g \mapsto \frac{1}{\rho} \log \mathbb{E}[e^{\rho g}]$, an operation that interpolates between averaging ($\rho \to 0$) and the supremum ($\rho \to \infty$). The K-step free energy is $\mathcal{P}_\beta(\mu) = \log 2 + f_0(0) - \frac{\beta^2}{2} \int_0^1 s\, \xi''(s)\, F(s)\, ds$.

Passing to a continuum of levels. Replacing the step function by a general non-decreasing $q : [0,1] \to [0,1]$ and refining the discrete iteration. The combined heat-plus-tilt operation fuses into the Parisi PDE

$$\partial_s \Phi_\mu(s,x) \;=\; -\tfrac{\xi''(s)}{2}\big[\partial_{xx}\Phi_\mu(s,x) + F(s)(\partial_x \Phi_\mu(s,x))^2\big], \qquad \Phi_\mu(1,x) = \log\cosh(x+\beta h),$$

where $F(s) = \mu([0,s])$ is the CDF of the overlap measure. Appendix A gives the formal derivation. The detailed version is in [Pan13b, Ch. 5].

2.6. **Rigor and uniqueness.** The rigorous statements we use are.

- Existence of the thermodynamic limit [GT02], $f(\beta,h) = \lim_N \frac{1}{N}\mathbb{E}[\log Z_N]$ exists.
- Parisi upper bound [Gue03], $f(\beta,h) \le \mathcal{P}_\beta(\mu)$ for every $\mu \in \mathcal{P}([0,1])$. Guerra's interpolation is a rigorous bound that makes no appeal to the replica method.
- Parisi lower bound [Tal06]. The matching lower bound holds, so $f(\beta,h) = \inf_\mu \mathcal{P}_\beta(\mu)$. Talagrand's cavity-method argument completes the Parisi formula.
- Uniqueness of the minimizer [AC15], $\mathcal{P}_\beta$ has a unique minimizer μ^*, and the minimizer's quantile function is analytic on the support interval. The stochastic control representation of the Parisi PDE derived in [AC15] also gives the key bound $|\partial_x \Phi_\mu| \le 1$ used to establish the W_1-Lipschitz constant of $\mathcal{P}_\beta$ in (4).

These results turn the Parisi functional into a rigorous object and its minimizer into a well-defined element of $\mathcal{P}([0,1])$, with no reference to replicas. The rest of the thesis takes $\mathcal{P}_\beta$ and μ^* as its starting point.

3. The Moment Class $\mathcal{M}_K$

Definition 3.1 (Moment class)**.** Let $\mu^* \in \mathcal{P}([0,1])$ have moments $m_k = \int_0^1 q^k \, d\mu^*(q)$ for $k \ge 1$. For $K \ge 1$, the *K-th moment class* is

$$\mathcal{M}_K \;=\; \{\nu \in \mathcal{P}([0,1]) \;:\; \textstyle\int_0^1 q^k \, d\nu(q) = m_k \text{ for } k = 1,\dots,K\}.$$

By convention $\mathcal{M}_0 = \mathcal{P}([0,1])$.

The moment class depends on μ^* only through its first K moments. Different measures with the same first K moments give the same $\mathcal{M}_K$.

Example 3.2 (One atom)**.** If $\mu^* = \delta_{q_*}$ has a single atom, then $m_k = q_*^k$ and by Theorem 3.5(c), $\mathcal{M}_2 = \{\delta_{q_*}\}$. Two moments already collapse the class to μ^*. $\mathcal{M}_1$ is larger. It contains every probability measure on $[0,1]$ with mean q_*.

Example 3.3 (Two atoms)**.** Let $\mu^* = w\,\delta_a + (1-w)\,\delta_b$ with $0 \le a < b \le 1$, $0 < w < 1$. The first three moments are

$$m_1 = wa + (1-w)b, \quad m_2 = wa^2 + (1-w)b^2, \quad m_3 = wa^3 + (1-w)b^3.$$

Theorem 3.5(d) below implies $\mathcal{M}_3 \supsetneq \{\mu^*\}$ and Theorem 3.5(c) implies $\mathcal{M}_4 = \{\mu^*\}$. A two-atom measure requires exactly four moments to identify. Explicitly, the three-atom measure $\nu_\varepsilon = v_1(\varepsilon)\delta_a + v_2(\varepsilon)\delta_b + \varepsilon\,\delta_c$ with $c \in (0,1) \setminus \{a,b\}$ lies in $\mathcal{M}_3$ for small $\varepsilon > 0$. The three linear moment equations for v_1, v_2, ε form an invertible Vandermonde system with solution $(v_1(\varepsilon), v_2(\varepsilon)) \to (w, 1-w)$ as $\varepsilon \to 0$, keeping $v_1, v_2 > 0$ by continuity.

Example 3.4 (Uniform measure)**.** For $\mu^* = \mathrm{Leb}|_{[0,1]}$, $m_k = 1/(k+1)$. Since μ^* is continuous, $\mathcal{M}_K \supsetneq \{\mu^*\}$ for every K. Nevertheless $\bigcap_K \mathcal{M}_K = \{\mu^*\}$ by the Hausdorff moment theorem. The canonical $(K+1)$-atomic extreme points of $\mathcal{M}_K$ in this case are the Gauss–Legendre quadrature measures, whose atoms are the zeros of the Legendre polynomial of degree $K+1$ and whose weights are the corresponding Christoffel numbers.

3.1. **The structure theorem.**

Theorem 3.5 (Structure of $\mathcal{M}_K$)**.** *Let $K \geq 1$.*

(a) $\mathcal{M}_K$ is convex and weakly compact in $\mathcal{P}([0,1])$.

(b) Every extreme point of $\mathcal{M}_K$ is supported on at most $K+1$ points.

(c) If μ^* has $m \geq 1$ distinct atoms and $K \geq 2m$, then $\mathcal{M}_K = \{\mu^*\}$.

(d) If μ^* has $m \geq 2$ distinct atoms and $K = 2m-1$, then $\mathcal{M}_K \supsetneq \{\mu^*\}$.

Parts (a)–(b) are essentially classical, (a) via Prokhorov's theorem, (b) essentially Richter's theorem [Ric57, Rog58, KS66]. Parts (c)–(d) locate the sharp collapse threshold, and the proofs make the stratified structure of the fibers of the moment map $\Psi : \mathcal{P}([0,1]) \to \mathbb{R}^K$ precise.

Proof of (a). Convexity is immediate from linearity of integration. For $\nu_1, \nu_2 \in \mathcal{M}_K$ and $t \in [0,1]$, $\int q^k \, d(t\nu_1 + (1-t)\nu_2) = tm_k + (1-t)m_k = m_k$. For compactness, $\mathcal{P}([0,1])$ is weakly compact by Prokhorov's theorem (the compactness of $[0,1]$ implies that every sequence in $\mathcal{P}([0,1])$ is tight). For each k, the map $\nu \mapsto \int q^k \, d\nu$ is weakly continuous because $q \mapsto q^k$ is continuous and bounded on $[0,1]$. Thus $\mathcal{M}_K = \bigcap_{k=1}^K \{\nu : \int q^k \, d\nu = m_k\}$ is a finite intersection of closed sets in a compact space, hence compact. □

Proof of (b). We prove the contrapositive. Let $\nu \in \mathcal{M}_K$ with $\mathrm{supp}(\nu) \supseteq \{a_1 < \cdots < a_{K+2}\}$. We construct a signed measure η killing all polynomials of degree $\leq K$. The homogeneous system

$$\sum_{j=1}^{K+2} h_j = 0, \qquad \sum_{j=1}^{K+2} h_j \, a_j^k = 0 \quad (k = 1, \ldots, K) \tag{7}$$

has coefficient matrix $A \in \mathbb{R}^{(K+1)\times(K+2)}$ with rows $(a_1^k, \ldots, a_{K+2}^k)$ for $k = 0, 1, \ldots, K$. Any $(K+1)$-column submatrix is a square Vandermonde with nonzero determinant $\prod_{i<j}(a_j - a_i)$. Thus A has rank $K+1$, the kernel is one-dimensional, and a nonzero solution $\mathbf{h}$ exists. The signed measure $\eta = \sum_j h_j \, \delta_{a_j}$ satisfies $\int p \, d\eta = 0$ for every polynomial p of degree $\leq K$, and $\eta(\mathbb{R}) = \sum_j h_j = 0$.

Case (i). Atoms. If ν has atoms at each a_j with mass $w_j > 0$, set $\varepsilon^* = \min_{h_j \neq 0} w_j/|h_j|$. For $\varepsilon \in (0, \varepsilon^*)$, $\nu^{\pm} = \nu \pm \varepsilon\, \eta$ are non-negative probability measures in $\mathcal{M}_K$ with $\nu = \frac{1}{2}(\nu^+ + \nu^-)$ and $\nu^+ \neq \nu^-$, so ν is not extreme.

Case (ii). Continuous component. If ν has a continuous density near some a_j, replace δ_{a_j} by the bump $\frac{1}{2\delta}\mathbf{1}_{[a_j-\delta, a_j+\delta]}$. Since $|q^k - a_j^k| \leq k\delta$ on $[a_j - \delta, a_j + \delta]$, the moment-killing property holds up to $O(\delta)$ in each moment. The moment map is a submersion at any measure whose support contains an open interval (the polynomials $1, q, \ldots, q^K$ are linearly independent on any interval), so the implicit function theorem provides an exact correction of size $O(\delta)$ restoring moment matching. The rest of Case (i) then applies. □

Remark 3.6 (Sharpness of $K+1$)*.* The bound $K+1$ is sharp. If $\nu \in \mathcal{M}_K$ has support size exactly $K+1$, (7) is a square Vandermonde system with only the trivial solution $\mathbf{h} = 0$, so no nontrivial moment-killing perturbation exists and ν is generically an extreme point. For generic interior points of the moment space, the extreme points of $\mathcal{M}_K$ are exactly the $(K+1)$-atomic measures.

Proof of (c). Let $\mu^* = \sum_{i=1}^m w_i\, \delta_{q_i}$ with $w_i > 0$ and $\nu \in \mathcal{M}_{2m}$. Consider the sum-of-squares polynomial

$$f(q) \;=\; \prod_{i=1}^{m} (q - q_i)^2,$$

which has degree $2m \leq K$, is non-negative on $[0,1]$, and vanishes precisely on $\{q_1, \ldots, q_m\}$. Expanding f in monomials and using $\nu \in \mathcal{M}_{2m}$.

$$\int f\, d\nu \;=\; \int f\, d\mu^* \;=\; \sum_i w_i\, f(q_i) \;=\; 0.$$

A non-negative continuous function integrating to zero against ν must vanish ν-a.e., so $\mathrm{supp}(\nu) \subseteq \{q_1, \ldots, q_m\}$. Writing $\nu = \sum_i v_i\, \delta_{q_i}$, the conditions for $k = 0, 1, \ldots, m-1$ give $\sum_i (v_i - w_i) q_i^k = 0$, an $m \times m$ Vandermonde system with only the trivial solution $v_i = w_i$. □

Proof of (d). We use the classical theory of the moment space [KS66, Ch. IV]. Let $M_K \subset \mathbb{R}^K$ denote the image of $\mathcal{P}([0,1])$ under the moment map $\Psi(\nu) = (\int q\, d\nu, \ldots, \int q^K\, d\nu)$. M_K is a compact convex body whose boundary consists of moment sequences that admit a unique representing measure supported on at most $\lceil K/2 \rceil$ points. Interior points of M_K have infinitely many representing measures.

Concretely, a moment point $\mathbf{m} = (m_1, \ldots, m_K) \in M_K$ lies on the boundary ∂M_K if and only if some supporting hyperplane of M_K passes through $\mathbf{m}$, equivalently if there is a non-trivial polynomial p of degree $\leq K$ that is non-negative on $[0,1]$ and $\int p\, d\nu = 0$ for every representing ν. A minimum-support representing measure at $\mathbf{m}$ has support exactly equal to the zeros of such a certifying p.

Now let $\mu^* = \sum_{i=1}^m w_i\, \delta_{q_i}$ with $m \geq 2$ and $w_i > 0$, and let $\mathbf{m}^* = \Psi(\mu^*) = (m_1, \ldots, m_{2m-1}) \in M_{2m-1}$. The minimum-support size of a representing measure at $\mathbf{m}^*$ in M_{2m-1} is m (which is $\leq \lceil (2m-1)/2 \rceil = m$), achieved by μ^* itself, and the certifying non-negative polynomial would have to have degree $\leq 2m-1$ with zeros at $q_1, \ldots, q_m$. But any non-negative polynomial with simple zeros at all q_i has degree at least $2m$ (each zero must have even multiplicity). Hence no such certifying polynomial of degree $\leq 2m-1$ exists, so $\mathbf{m}^*$ is in the interior of M_{2m-1} and its fiber $\Psi^{-1}(\mathbf{m}^*) = \mathcal{M}_{2m-1}$ is infinite-dimensional. In particular $\mathcal{M}_{2m-1} \supsetneq \{\mu^*\}$. □

3.2. **The filtration and its collapse.** The moment classes form a nested filtration

$$\mathcal{P}([0,1]) \;=\; \mathcal{M}_0 \;\supset\; \mathcal{M}_1 \;\supset\; \mathcal{M}_2 \;\supset\; \cdots \tag{8}$$

Each inclusion is strict when μ^* is not a point mass. By Theorem 3.5(b), $\mathcal{M}_K$ has $(K+1)$-atomic extreme points, which span a $(K+1)$-parameter family ($2(K+1)-1-K = K+1$ parameters free after the first K moment constraints), and a generic such extreme has $(K+1)$-th moment differing from m_{K+1}.

Corollary 3.7 (Sharp collapse)**.** *If μ^* has m distinct atoms, then $\mathcal{M}_K = \{\mu^*\} \iff K \geq 2m$.*

The threshold $2m$ equals the number of real parameters specifying μ^*, m positions and m weights minus normalization give $2m-1$ free parameters, and identifying the support requires a degree-$2m$ sum-of-squares certificate, adding one more moment.

When μ^* is continuous, the filtration never collapses, $\mathcal{M}_K \supsetneq \{\mu^*\}$ for every K. Nevertheless $\bigcap_K \mathcal{M}_K = \{\mu^*\}$ by the Hausdorff moment theorem, since a probability measure on $[0,1]$ is uniquely determined by its moment sequence. The rate of this convergence is quantified by the Wasserstein estimate below.

Definition 3.8 (RSB-type marginal)**.** A probability measure $\nu \in \mathcal{P}([0,1])$ with exactly j atoms is a $(j-1)$-*step RSB-type marginal.* It has the same support structure as the single-overlap marginal $\mathrm{Law}(R_{12})$ of a $(j-1)$-step RSB state in the Parisi ansatz.

We use “RSB-type marginal” rather than “RSB measure” throughout to emphasize that a $(j-1)$-step RSB-type marginal is only the *single-overlap marginal* of a $(j-1)$-step RSB state, not the full hierarchical block partition of the overlap matrix. The joint structure (the Parisi matrix $Q^{(j-1)}$ of Section 2.4) contains strictly more information. Distinct Parisi matrices can have the same marginal. The moment class $\mathcal{M}_K$ constrains only the marginal axis. The joint axis is governed by GG (Section 5).

Since extreme points of $\mathcal{M}_K$ have $\leq K+1$ atoms (Theorem 3.5(b)), they are $\leq K$-step RSB-type marginals. As K grows and $\mathcal{M}_K$ shrinks, the surviving extremes require fewer atoms, until at $K = 2m$ only μ^* survives. The classical slogan “k-step RSB approximates the Parisi measure” becomes precise in this framework. A k-step RSB-type marginal is the minimum-complexity object consistent with $2k$ moments of μ^*, and it is exact once $K \geq 2m$.

3.3. **The Wasserstein estimate.**

Lemma 3.9 (Wasserstein diameter of $\mathcal{M}_K$)**.** *For every $\nu \in \mathcal{M}_K$,*

$$W_1(\mu^*, \nu) \;\le\; \frac{1}{K+1}.$$

Proof. By Kantorovich–Rubinstein duality,

$$W_1(\mu^*, \nu) \;=\; \sup\Big\{\Big|\int f\, d\mu^* - \int f\, d\nu\Big| \;:\; f \text{ 1-Lipschitz on } [0,1]\Big\}.$$

For any 1-Lipschitz f, the Jackson–Stechkin theorem [DL93, Ch. 7] gives a polynomial p of degree $\le K$ with $\|f - p\|_\infty \le 1/(2(K+1))$. The constant is sharp, with the extremal case a 1-Lipschitz sawtooth with $K+1$ oscillations. Since μ^* and ν agree on polynomials of degree $\le K$.

$$\Big|\int f\, d\mu^* - \int f\, d\nu\Big| \;=\; \Big|\int (f-p)\, d\mu^* - \int (f-p)\, d\nu\Big| \;\le\; 2\|f-p\|_\infty \;\le\; \tfrac{1}{K+1}.$$

Take the supremum over 1-Lipschitz f. □

The bound is sharp up to constants for continuous μ^*. The extremal $\nu \in \mathcal{M}_K$ achieving it is a discrete measure supported on the $K+1$ Chebyshev-like nodes whose moments match μ^*.

Corollary 3.10 (General moment sensitivity)**.** *Let $\Phi : \mathcal{P}([0,1]) \to \mathbb{R}$ be L-Lipschitz in W_1. For every $\nu \in \mathcal{M}_K$,*

$$|\Phi(\nu) - \Phi(\mu^*)| \;\le\; \frac{L}{K+1}.$$

This is an abstract statement that covers any W_1-Lipschitz functional of the overlap distribution. The Parisi free energy is one. The internal energy, magnetization, and various susceptibilities are others (they are ∂_β- or ∂_h-derivatives of $f(\beta, h)$ and inherit Lipschitz regularity from (4)).

3.4. **Free-energy control.**

Theorem 3.11 (Free-energy control)**.** *For every $\nu \in \mathcal{M}_K$,*

$$0 \;\le\; \mathcal{P}_\beta(\nu) - \mathcal{P}_\beta(\mu^*) \;\le\; \frac{4\beta^2}{K+1}.$$

Proof. The lower bound holds because μ^* is the unique minimizer of $\mathcal{P}_\beta$ (Theorem 1.2). For the upper bound we combine Lemma 3.9 with the Lipschitz estimate (4), whose proof we recall.

From (3), $\mathcal{P}_\beta(\mu) = \log 2 + \Phi_\mu(0,0) - \frac{1}{2}\int_0^1 s\,\xi''(s) F_\mu(s)\, ds$. We bound the two μ-dependent pieces separately.

The PDE piece. The Auffinger–Chen stochastic control representation [AC15] writes

$$\Phi_\mu(0,0) \;=\; \sup_\alpha \mathbb{E}\Big[\log\cosh\big(\beta h + \int_0^1 \alpha_s\, dW_s\big) - \tfrac{1}{2}\int_0^1 \xi''(s)\, F_\mu(s)\, \alpha_s^2\, ds\Big],$$

where the supremum is over bounded progressively-measurable controls α. A consequence is $|\partial_x \Phi_\mu(s,x)| \le 1$ for all (s,x), since the bound $|\tanh(\cdot)| \le 1$ propagates backward under the maximum principle. The W_1 distance between the laws induced by μ and ν via the $F_\mu \alpha_s^2$ weighting is bounded by $\int \xi''(s)|F_\mu(s) - F_\nu(s)|\,ds \le \xi''(1)\,W_1(\mu,\nu)$, and the 1-Lipschitz observable $\log\cosh$ preserves this, giving

$$|\Phi_\mu(0,0) - \Phi_\nu(0,0)| \;\le\; \xi''(1)\,W_1(\mu,\nu) \;=\; 2\beta^2\,W_1(\mu,\nu).$$

The integral piece. Since $s\xi''(s) = s\beta^2 \le \beta^2$ on $[0,1]$ and $|\int g(s)(F_\mu - F_\nu)(s)\,ds| \le \|g\|_\infty W_1(\mu,\nu)$ for g of bounded variation.

$$\Big|\tfrac{1}{2}\int_0^1 s\,\xi''(s)(F_\mu - F_\nu)(s)\,ds\Big| \;\le\; \tfrac{\xi''(1)}{2} W_1(\mu,\nu) \;=\; \beta^2\,W_1(\mu,\nu).$$

Adding the two pieces,

$$|\mathcal{P}_\beta(\mu) - \mathcal{P}_\beta(\nu)| \;\le\; 2\beta^2\,W_1(\mu,\nu) + \beta^2\,W_1(\mu,\nu) \;=\; 3\beta^2\,W_1(\mu,\nu) \;\le\; 4\beta^2\,W_1(\mu,\nu),$$

which is (4). Apply to $\mu = \nu$, $\nu = \mu^*$ and use Lemma 3.9. □

3.5. Convergence rate of the RSB hierarchy.

Definition 3.12 (K-step RSB free energy)**.** For $K \ge 0$,

$$\mathcal{P}_\beta^{(K)} \;=\; \inf\{\mathcal{P}_\beta(\nu) \,:\, \nu \in \mathcal{P}([0,1]),\; |\mathrm{supp}(\nu)| \le K+1\}.$$

The K-step RSB variational problem restricts $\mathcal{P}_\beta$ to the finite-dimensional manifold of $(K+1)$-atomic measures. It is the object most directly computed in the physics literature and is solved numerically using the iterative scheme of Proposition 4.3.

Theorem 3.13 (RSB convergence rate)**.** *For every $K \ge 0$,*

$$0 \;\le\; \mathcal{P}_\beta^{(K)} - \mathcal{P}_\beta(\mu^*) \;\le\; \frac{4\beta^2}{K+1}.$$

Proof. The lower bound holds because μ^* is the global minimizer, and restricting the feasible set can only raise the infimum. For the upper bound, Theorem 3.5(b) produces extreme points ν_K of $\mathcal{M}_K$ that are $(K+1)$-atomic, hence feasible for the K-step RSB problem, and Theorem 3.11 gives $\mathcal{P}_\beta(\nu_K) - \mathcal{P}_\beta(\mu^*) \le 4\beta^2/(K+1)$. Thus $\mathcal{P}_\beta^{(K)} \le \mathcal{P}_\beta(\nu_K) \le \mathcal{P}_\beta(\mu^*) + 4\beta^2/(K+1)$. □

Corollary 3.14 (Weak convergence of RSB minimizers)**.** *Let ν_K minimize $\mathcal{P}_\beta$ over $(K+1)$-atomic measures. Then $\nu_K \to \mu^*$ weakly as $K \to \infty$.*

Proof. By Theorem 3.13, $\mathcal{P}_\beta(\nu_K) \to \mathcal{P}_\beta(\mu^*)$. The sequence lies in the weakly compact space $\mathcal{P}([0,1])$. Every weak cluster point ν_∞ satisfies $\mathcal{P}_\beta(\nu_\infty) = \mathcal{P}_\beta(\mu^*)$ by weak continuity, and by uniqueness (Theorem 1.2), $\nu_\infty = \mu^*$. □

Remark 3.15 (Computational interpretation). Given the moments $m_1, \dots, m_K$, one can construct any $(K+1)$-atomic $\nu_K \in \mathcal{M}_K$ by solving for positions $q_0 < q_1 < \dots < q_K \in [0,1]$ and weights $w_0, \dots, w_K \geq 0$ satisfying

$$\sum_{i=0}^{K} w_i = 1, \qquad \sum_{i=0}^{K} w_i \, q_i^k = m_k \qquad (k = 1, \dots, K).$$

This is a nonlinear but finite-dimensional system, $2(K+1)$ unknowns, $K+1$ equations, a $(K+1)$-dimensional family of solutions parameterized (say) by the positions. Plugging any such ν_K into $\mathcal{P}_\beta$ yields a free-energy estimate accurate to $4\beta^2/(K+1)$, and further optimizing over the remaining parameters gives the best K-step RSB approximation.

3.6. **Tightness of the $O(1/K)$ rate.** The upper bound in Theorem 3.11 is matched by a lower bound of the same order, showing the $O(\beta^2/K)$ rate is tight up to constants.

Theorem 3.16 (Tightness)**.** *Suppose μ^* has an absolutely continuous component on an interval $I \subset [0,1]$. Then there is a constant $c(\beta) > 0$ such that for every $K \geq 1$,*

$$\sup_{\nu \in \mathcal{M}_K} \left(\mathcal{P}_\beta(\nu) - \mathcal{P}_\beta(\mu^*)\right) \;\geq\; \frac{c(\beta)}{K+1}.$$

Proof. Let $\eta = \sum_{j=1}^{K+2} h_j \, \delta_{a_j}$ be the moment-killing signed measure from Theorem 3.5(b), with interpolation nodes $a_1 < \dots < a_{K+2}$ chosen in the interval I where μ^* is absolutely continuous with density bounded below by $\rho_0 > 0$. The coefficients $\mathbf{h}$ solve $\int q^k \, d\eta = 0$ for $k = 0, 1, \dots, K$. After symmetric smoothing at scale $\delta > 0$, we may assume η is a smooth signed measure with the same moments vanishing. Provided $|\varepsilon| \, \|d\eta/dq\|_\infty \leq \rho_0$, the perturbation $\nu_\varepsilon := \mu^* + \varepsilon \, \eta$ is a probability measure in $\mathcal{M}_K$.

The first variation. We compute $(d/d\varepsilon) \, \mathcal{P}_\beta(\nu_\varepsilon)$ at $\varepsilon = 0$. Since $F_{\nu_\varepsilon}(s) = F_{\mu^*}(s) + \varepsilon F_\eta(s)$ with $F_\eta(s) = \eta([0,s])$, and since $\mathcal{P}_\beta$ depends on μ only through the CDF F_μ, the chain rule gives

$$\left.\frac{d\mathcal{P}_\beta}{d\varepsilon}\right|_{\varepsilon=0} \;=\; \int_0^1 \frac{\delta \mathcal{P}_\beta}{\delta F}(s) \cdot F_\eta(s) \, ds.$$

The integral piece of $\mathcal{P}_\beta$ is $-\frac{1}{2} \int s \, \xi''(s) \, F(s) \, ds$, contributing $\delta/\delta F(s) = -\frac{1}{2} s \, \xi''(s)$. For the PDE piece, the Auffinger–Chen representation [AC15] writes

$$\Phi_\mu(0,0) \;=\; \sup_{\alpha} \mathbb{E}\Big[\log\cosh\big(\beta h + \textstyle\int_0^1 \alpha_s \, dW_s\big) - \frac{1}{2} \int_0^1 \xi''(s) \, F_\mu(s) \, \alpha_s^2 \, ds\Big],$$

and the envelope theorem (at the optimal control α^* for μ^*) gives

$$\frac{\delta \Phi_{\mu^*}(0,0)}{\delta F(s)} \;=\; -\frac{\xi''(s)}{2} \, \mathbb{E}\big[(\alpha_s^*)^2\big] \;=\; -\frac{\xi''(s)}{2} \, \mathbb{E}\big[(\partial_x \Phi_{\mu^*}(s, X_s))^2\big],$$

where we used that the optimal feedback satisfies $\alpha_s^* = \partial_x \Phi_{\mu^*}(s, X_s)$ along the controlled process X_s. Setting

$$\psi_{\mu^*}(s) \;:=\; -\frac{\xi''(s)}{2}\Big(s + \mathbb{E}\big[(\partial_x \Phi_{\mu^*})^2(s, X_s)\big]\Big),$$

we have

$$S(\eta) \;:=\; \frac{d\mathcal{P}_\beta(\nu_\varepsilon)}{d\varepsilon}\bigg|_{\varepsilon=0} \;=\; \int_0^1 \psi_{\mu^*}(s)\, F_\eta(s)\, ds.$$

Orthogonality. Integration by parts converts the moment-killing property $\int q^k\, d\eta = 0$ for $k = 0, \ldots, K$ into $\int_0^1 q^{k-1}\, F_\eta(q)\, dq = 0$ for $k = 1, \ldots, K$, so $F_\eta \perp \mathrm{Poly}_{K-1}$ in $L^2(0,1)$.

Non-polynomiality. We claim ψ_{μ^*} is not a polynomial of degree $\le K-1$. The function Φ_{μ^*} is analytic in s on the support of μ^* by [AC15], with terminal condition $\Phi_{\mu^*}(1,x) = \log\cosh(x+\beta h)$. Hence $\partial_x \Phi_{\mu^*}(1,x) = \tanh(x+\beta h)$, and $\mathbb{E}[(\partial_x\Phi_{\mu^*})^2(1,X_1)] = \mathbb{E}[\tanh^2(X_1 + \beta h)]$. Running the Parisi PDE backward from $s = 1$, the terminal non-polynomiality of $\tanh^2$ propagates via the nonlinear term $F(s)(\partial_x\Phi)^2$, so the Taylor coefficients of $\psi_{\mu^*}(s)$ at $s = 1$ are non-vanishing to all orders. A function whose Taylor series at a point has infinitely many non-zero coefficients is not a polynomial of any finite degree.

Conclusion. Since $\psi_{\mu^*} \notin \mathrm{Poly}_{K-1}$ and $F_\eta \perp \mathrm{Poly}_{K-1}$, for a generic choice of interpolation nodes $a_1, \ldots, a_{K+2}$ we have $S(\eta) \neq 0$. Choosing the sign of η so that $S(\eta) > 0$ (which is consistent with μ^* being a minimizer, since S is a directional derivative from a boundary/interior point of the moment constraint), we have $\mathcal{P}_\beta(\nu_\varepsilon) - \mathcal{P}_\beta(\mu^*) = \varepsilon\, S(\eta) + O(\varepsilon^2)$. The Chebyshev-optimal choice of nodes gives $\|F_\eta\|_\infty \asymp 1/(K+1)$ and $|S(\eta)| \gtrsim 1/(K+1)$ by the non-polynomiality. Setting $\varepsilon = c_1/(K+1)$ small enough that $\nu_\varepsilon \ge 0$ yields

$$\mathcal{P}_\beta(\nu_\varepsilon) - \mathcal{P}_\beta(\mu^*) \;\ge\; \frac{c(\beta)}{K+1}$$

for some constant $c(\beta) > 0$ depending on β and the density lower bound ρ_0. □

Corollary 3.17 (Two-sided rate)**.** *Under the hypothesis of Theorem 3.16,*

$$\frac{c(\beta)}{K+1} \;\le\; \sup_{\nu\in\mathcal{M}_K} \left(\mathcal{P}_\beta(\nu) - \mathcal{P}_\beta(\mu^*)\right) \;\le\; \frac{4\beta^2}{K+1}.$$

In particular, the $O(\beta^2/K)$ rate of Theorem 3.11 is tight up to the temperature-dependent constant.

3.7. **The moment polytope and Tchebycheff systems.** The moment map $\Psi : \mathcal{P}([0,1]) \to \mathbb{R}^K$ defined by $\Psi(\nu) = (\int q\, d\nu, \ldots, \int q^K\, d\nu)$ sends $\mathcal{M}_K$ to the single point $\mathbf{m} = (m_1, \ldots, m_K)$. The image $\Psi(\mathcal{P}([0,1]))$ is a compact convex body in $\mathbb{R}^K$, the classical *moment space* [KS66], and $\mathcal{M}_K$ is the fiber $\Psi^{-1}(\mathbf{m})$. A point $\mathbf{m}$ lies in the interior iff some measure with moments $\mathbf{m}$ has support of size at least $\lfloor K/2\rfloor + 1$.

Theorem 3.5(b) is Richter's theorem. Extreme points of fibers over interior points of the moment space have at most $K+1$ atoms. The filtration $\mathcal{M}_1 \supset \mathcal{M}_2 \supset \cdots$ corresponds to a nested sequence of moment maps $\pi_K : \mathcal{P}([0,1]) \to \mathbb{R}^K$, with each $\mathcal{M}_K$ a fiber. The sharp collapse at $K = 2m$ reflects a classical fact. A K-point Chebyshev system on $[0,1]$ admits at most $K+1$ sign changes, and identifying an m-atomic measure requires enough moments $(K \ge 2m)$ for a non-negative polynomial certificate of its support.

Our contribution is the connection to the Parisi theory. In the SK setting, the k-step RSB-type marginals are the extreme points of $\mathcal{M}_K$ with exactly $k+1$ atoms (for $k \le K$), and the RSB hierarchy is the extreme-point stratification of the filtration. The rate of collapse of $\mathcal{M}_K$ as $K \to \infty$, quantified by the Wasserstein estimate (Lemma 3.9), translates into the rate of convergence of the RSB hierarchy (Theorem 3.13).

4. The Explicit–PDE Decomposition

The uniform $O(\beta^2/K)$ rate of Section 3 is loose at high temperature. We refine it by expanding $\mathcal{P}_\beta(\mu)$ in powers of β^2 and tracking moment by moment. The expansion rests on a decomposition of $\mathcal{P}_\beta$ into an explicit piece depending only on m_2, and a PDE piece carrying all nonlinear dependence on μ.

4.1. **The explicit–PDE decomposition.** We work with the SK covariance $\xi(q) = \frac{\beta^2}{2}q^2$, so $\xi'(q) = \beta^2 q$ and $\xi''(q) = \beta^2$. The Parisi PDE is

$$\partial_s \Phi \;=\; -\frac{\beta^2}{2}\big[\Phi'' + F(s)\,(\Phi')^2\big], \qquad \Phi(1,x) = \log\cosh(x), \tag{9}$$

where $F(s) = \mu([0,s])$ and we have absorbed the external field into the boundary (equivalent to $\log\cosh(x+\beta h)$). The Parisi functional is

$$\mathcal{P}_\beta(\mu) \;=\; \log 2 + \Phi_\mu(0,0) - \frac{\beta^2}{2}\int_0^1 F(q)\,q\,dq. \tag{10}$$

Proposition 4.1 (Explicit–PDE decomposition)**.** *For any $\mu \in \mathcal{P}([0,1])$ with second moment $m_2 = \int q^2\,d\mu$,*

$$\mathcal{P}_\beta(\mu) \;=\; \log 2 \;-\; \frac{\beta^2}{4}(1-m_2) \;+\; \Phi_\mu(0,0). \tag{11}$$

In particular, the piece of $\mathcal{P}_\beta(\mu)$ outside $\Phi_\mu(0,0)$ depends on μ only through m_2.

Proof. By Fubini,

$$\int_0^1 F(q)\,q\,dq \;=\; \int_0^1 q\int_0^1 \mathbf{1}_{s\le q}\,d\mu(s)\,dq \;=\; \int_0^1\int_s^1 q\,dq\,d\mu(s) \;=\; \int_0^1 \tfrac{1-s^2}{2}\,d\mu(s) \;=\; \tfrac{1-m_2}{2}.$$

Substituting into (10) gives (11). □

Theorem 4.2 (Separation on $\mathcal{M}_K$)**.** *For $K \ge 2$ and any $\mu,\nu \in \mathcal{M}_K$,*

$$\mathcal{P}_\beta(\nu) - \mathcal{P}_\beta(\mu) \;=\; \Phi_\nu(0,0) - \Phi_\mu(0,0).$$

The free-energy difference on $\mathcal{M}_K$ is entirely PDE-side. The explicit algebraic piece is constant across $\mathcal{M}_K$ once m_2 is prescribed.

Proof. Immediate from Proposition 4.1. The explicit piece $\log 2 - \frac{\beta^2}{4}(1-m_2)$ depends on μ only through m_2, which is common to all $\mu,\nu \in \mathcal{M}_K$ for $K \ge 2$. □

Combined with the high-temperature expansion (Theorem 4.11), Theorem 4.2 is the sharpest result at high temperature. For $K \geq 2$ the variation of $\mathcal{P}_\beta$ across $\mathcal{M}_K$ is $O(\beta^6)$, not $O(\beta^2)$. For $K \geq 3$ only the shape-dependent spread penalty $\mathcal{R}(\mu)$ contributes.

The decomposition separates $\mathcal{P}_\beta$ into an algebraic piece $\log 2 - \frac{\beta^2}{4}(1 - m_2)$ depending on μ through a single scalar moment and on β polynomially, and an analytic piece $\Phi_\mu(0,0)$ encoding the solution of a nonlinear PDE. The algebraic piece is at most $\beta^2/4$ in magnitude, while the analytic piece absorbs all large-β nonlinearity. At low β the algebraic piece captures most of the free energy and the analytic piece is perturbative, the subject of the next subsection.

4.2. **The iterative scheme for finitely atomic measures.** When μ has finitely many atoms, the Parisi PDE (9) reduces to a finite recursion of Gaussian convolutions. Let $\nu = \sum_{i=0}^{K} w_i\, \delta_{q_i}$ with

$$0 \;\leq\; q_0 \;<\; q_1 \;<\; \cdots \;<\; q_K \;\leq\; 1, \qquad w_i > 0, \qquad \textstyle\sum_i w_i = 1,$$

and cumulative weights $\rho_i = \sum_{j \leq i} w_j$. On (q_i, q_{i+1}) the CDF is constant, $F(s) = \rho_i$. The PDE becomes

$$\partial_s \Phi \;=\; -\frac{\beta^2}{2}\big[\Phi'' + \rho_i\,(\Phi')^2\big],$$

solved by the Cole–Hopf substitution $u_i = \exp(\rho_i\, \Phi)$.

Proposition 4.3 (Iterative scheme)**.** *Define $f_i : \mathbb{R} \to \mathbb{R}$ by the backward iteration*

$$f_{K+1}(x) = \log\cosh(x), \tag{12}$$

$$f_i(x) = \frac{1}{\rho_i}\,\log \mathbb{E}\big[\exp(\rho_i\, f_{i+1}(x + \sigma_i Z))\big], \qquad i = K, K-1, \ldots, 0, \tag{13}$$

with $Z \sim \mathcal{N}(0,1)$ and $\sigma_i^2 = \beta^2(q_{i+1} - q_i)$. Then $\Phi_\nu(0,0) = \mathbb{E}[f_0(\beta\sqrt{q_0}\, Z)]$ if $q_0 > 0$ and $f_0(0)$ if $q_0 = 0$.

Proof. On $[q_i, q_{i+1})$, set $u_i = \exp(\rho_i\, \Phi)$. Direct computation gives $u_i'' = \rho_i u_i[\Phi'' + \rho_i(\Phi')^2]$ and $\partial_s u_i = -\frac{\beta^2}{2} u_i''$, the backward heat equation with diffusion $\beta^2/2$. Its solution is

$$u_i(q_i, x) \;=\; \mathbb{E}[u_i(q_{i+1}, x + \sigma_i Z)], \qquad \sigma_i^2 = \beta^2(q_{i+1} - q_i).$$

Converting back, $\Phi(q_i, x) = \frac{1}{\rho_i}\log u_i(q_i, x)$. At $s = q_{i+1}$ the CDF jumps from ρ_i to ρ_{i+1}.

$$u_i(q_{i+1}, x) \;=\; \exp(\rho_i\, \Phi(q_{i+1}, x)) \;=\; \exp(\rho_i\, f_{i+1}(x)),$$

giving (13). The top step uses $f_{K+1}(x) = \Phi(1, x) = \log\cosh(x)$ and $\rho_K = 1$, so $f_K(x) = \log \mathbb{E}[\cosh(x + \sigma_K Z)]$.

For $s \in [0, q_0)$ the CDF is 0 and the PDE reduces to the backward heat equation $\partial_s \Phi = -\frac{\beta^2}{2}\Phi''$, whose solution is the Gaussian convolution $\Phi(0, x) = \mathbb{E}[f_0(x + \beta\sqrt{q_0}\, Z)]$. Evaluate at $x = 0$. ☐

Remark 4.4 (No PDE required)*.* Computing $\mathcal{P}_\beta(\nu)$ for a $(K+1)$-atomic measure requires only $K+1$ Gaussian expectations (plus one final convolution if $q_0 > 0$). Each step (13) applies an exponential tilt at rate ρ_i, convolves with a Gaussian of variance σ_i^2, and takes a normalized log. This is the standard K-step RSB computation, here a direct consequence of the moment-class structure. The extreme points of $\mathcal{M}_K$ are $(K+1)$-atomic, and on these the PDE collapses.

Example 4.5 (Replica symmetry, $K=0$)**.** For $\nu = \delta_{q_0}$ with $\rho_0 = 1$ and $\sigma_0^2 = \beta^2 q_0$,

$$f_0(x) \;=\; \log \mathbb{E}[\cosh(x + \beta\sqrt{q_0}\,Z)],$$

and Proposition 4.1 gives

$$\mathcal{P}_\beta(\delta_{q_0}) \;=\; \log 2 - \tfrac{\beta^2}{4}(1-q_0^2) + \mathbb{E}[\log\cosh(\beta\sqrt{q_0}\,Z)],$$

the classical replica-symmetric formula.

Example 4.6 (One-step RSB, $K=1$)**.** For $\nu = w_0\,\delta_{q_0} + w_1\,\delta_{q_1}$ with $\rho_0 = w_0, \rho_1 = 1$.

$$\begin{aligned} f_1(x) &= \log \mathbb{E}[\cosh(x + \beta\sqrt{q_1 - q_0}\,Z)], \\ f_0(x) &= \tfrac{1}{w_0}\,\log \mathbb{E}\big[\exp(w_0\,f_1(x + \beta\sqrt{q_0}\,Z'))\big]. \end{aligned}$$

Then $\Phi_\nu(0,0) = f_0(0)$ if $q_0 = 0$, else $\mathbb{E}[f_0(\beta\sqrt{q_0}\,Z)]$, recovering the 1-step RSB formula with Parisi parameter w_0 and breakpoint q_0.

4.3. **The high-temperature expansion.** Write $\Phi_\mu(s,x) = \sum_{n\ge 0}\beta^{2n}\,\varphi_n(s,x)$ and collect orders of β^2 in (9). Each φ_n satisfies a linear PDE whose forcing depends on $\varphi_0,\dots,\varphi_{n-1}$, with boundary $\varphi_n(1,x) = \log\cosh(x)\,\mathbf{1}_{n=0}$.

Order β^0. $\partial_s\varphi_0 = 0$ with $\varphi_0(1,x) = \log\cosh(x)$, so $\varphi_0(s,x) = \log\cosh(x)$ for all s and $\varphi_0(0,0) = 0$.

Order β^2, a universal constant. The β^2 equation is

$$\partial_s\varphi_1 \;=\; -\tfrac{1}{2}\big[\varphi_0'' + F(s)(\varphi_0')^2\big] \;=\; -\tfrac{1}{2}\big[\operatorname{sech}^2(x) + F(s)\tanh^2(x)\big],$$

integrating to

$$\varphi_1(s,x) \;=\; \tfrac{1-s}{2}\,\operatorname{sech}^2(x) + \tfrac{A(s)}{2}\tanh^2(x), \qquad A(s) = \int_s^1 (F(t)-1)\,dt.$$

At $(s,x) = (0,0)$, $\operatorname{sech}^2(0) = 1$ and $\tanh(0) = 0$ give

$$\varphi_1(0,0) \;=\; \tfrac{1}{2},$$

independent of μ. Every probability measure on $[0,1]$ contributes the same β^2 coefficient. Physically, this is the high-temperature fluctuation contribution. At small β the log-partition function is approximately $\frac{\beta^2}{4}N$ regardless of the overlap structure.

Order β^4, the second moment m_2. The β^4 equation is

$$\partial_s \varphi_2 \;=\; -\tfrac{1}{2}\big[\varphi_1'' + 2F(s)\,\varphi_0'\,\varphi_1'\big].$$

Computing $\varphi_1'(s,x) = -(1-s)\operatorname{sech}^2(x)\tanh(x) + A(s)\tanh(x)\operatorname{sech}^2(x)$ and $\varphi_1''(s,x) = -(1-s)\operatorname{sech}^2(x)(1-3\tanh^2(x)) + A(s)\operatorname{sech}^2(x)(1-3\tanh^2(x))$, we see that at $x=0$, $\varphi_1'(s,0) = 0$ and $\varphi_1''(s,0) = -(1-s) + A(s)$. Since $\varphi_0'(0) = \tanh(0) = 0$, the cross term vanishes, and

$$\partial_s \varphi_2\big|_{x=0} \;=\; \tfrac{1-s}{2} - \tfrac{A(s)}{2}.$$

The first term contributes $\frac{1}{4}$ to $\varphi_2(0,0)$ and is independent of μ. The second contributes $\frac{1}{2}\int_0^1 A(t)\,dt$. We evaluate by Fubini.

$$\begin{aligned}\int_0^1 A(t)\,dt &= \int_0^1 \int_t^1 (F(u)-1)\,du\,dt = \int_0^1 (F(u)-1)\int_0^u dt\,du \\ &= \int_0^1 u(F(u)-1)\,du = \int_0^1 u\,F(u)\,du - \tfrac{1}{2} = \tfrac{1-m_2}{2} - \tfrac{1}{2} = -\tfrac{m_2}{2}.\end{aligned}$$

Therefore the μ-dependent content is

$$\varphi_2(0,0) \;=\; \text{const} - \tfrac{m_2}{4},$$

depending only on the second moment m_2.

Order β^6, the third moment and the spread penalty. The β^6 equation requires $\varphi_2(s,x)$ fully. Direct computation from the β^4 equation gives

$$\varphi_2(s,x) \;=\; \text{const}(s) + \frac{\operatorname{sech}^2(x)}{2}\Big[B_1(s) - \tanh^2(x)\,B_2(s)\Big], \tag{14}$$

where $B_1(s) = \int_s^1 A(t)\,dt$ and $B_2(s) = \int_s^1 A(t)(3-2F(t))\,dt$, and const$(s)$ absorbs the universal contribution. At $x=0$ the β^6 equation simplifies (even symmetry of the PDE kills terms with $\varphi_0'(0)$ or $\varphi_1'(s,0)$) to $\partial_s\varphi_3|_{x=0} = -\frac{1}{2}\varphi_2''(s,0)$. From (14), expanding in x.

$$\varphi_2(s,x) \;=\; \tfrac{B_1(s)}{2} - \tfrac{B_1(s)+B_2(s)}{2}\,x^2 + O(x^4),$$

so $\varphi_2''(s,0) = -(B_1(s)+B_2(s))$ and

$$\partial_s\varphi_3\big|_{x=0} \;=\; \tfrac{B_1(s)+B_2(s)}{2} \;=\; \int_s^1 A(t)(2-F(t))\,dt.$$

Integrating from 0 to 1 with $\varphi_3(1,0) = 0$ and Fubini.

$$\varphi_3(0,0) \;=\; -\int_0^1 t\,A(t)(2-F(t))\,dt. \tag{15}$$

By a further Fubini, $A(t) = -\int_t^1 (q-t)\,d\mu(q)$, so

$$\varphi_3(0,0) \;=\; \int_0^1 t(2-F(t))\int_t^1 (q-t)\,d\mu(q)\,dt.$$

Theorem 4.7 (Moment–correction decomposition at order β^6)**.** *For every $\mu \in \mathcal{P}([0,1])$,*

$$\varphi_3(0,0) \;=\; \frac{m_3}{3} \;-\; \mathcal{R}(\mu), \tag{16}$$

where

$$\mathcal{R}(\mu) \;=\; \int_0^1 \int_0^q t\,F(t)(q-t)\,dt\,d\mu(q) \;\geq\; 0 \tag{17}$$

and $F(t) = \mu([0,t])$.

Proof. Exchanging the outer integral with $d\mu$ by Fubini.

$$\varphi_3(0,0) \;=\; \int_0^1 \int_0^q t(2-F(t))(q-t)\,dt\,d\mu(q).$$

Splitting $2 - F(t) = 2 - F(t)$.

$$\varphi_3(0,0) \;=\; 2\int_0^1 \int_0^q t(q-t)\,dt\,d\mu(q) - \int_0^1 \int_0^q t\,F(t)(q-t)\,dt\,d\mu(q).$$

The first integral is $2\int_0^1 \frac{q^3}{6}\,d\mu(q) = m_3/3$, and the second is $\mathcal{R}(\mu)$. Non-negativity of $\mathcal{R}$ follows from $t \geq 0$, $F(t) \geq 0$, $q - t \geq 0$ on the integration domain and $d\mu \geq 0$. □

Proposition 4.8 (Properties of the spread penalty)**.** *The spread penalty $\mathcal{R} : \mathcal{P}([0,1]) \to [0,\infty)$ satisfies.*

(a) $\mathcal{R}(\delta_q) = 0$ for every $q \in [0,1]$.

(b) $\mathcal{R}(\mu) > 0$ whenever $\mathrm{supp}(\mu)$ contains at least two distinct points.

(c) $\mathcal{R}$ is continuous on $\mathcal{P}([0,1])$ in the weak topology (equivalently, W_1).

Proof. (a) For $\mu = \delta_q$, $F(t) = \mathbf{1}_{t\geq q}$. The integrand of the inner integral is $t\,\mathbf{1}_{t\geq q}(q-t)$. On $[0,q)$, $\mathbf{1}_{t\geq q} = 0$, and at $t = q$, $q - t = 0$. So the integrand vanishes identically on $[0,q]$.

(b) Suppose μ has support points $q_1 < q_2$ with $\mu(\{q \leq q_1\}) = w_1 > 0$. The $d\mu$-integration at $q = q_2$ contributes $\int_0^{q_2} t\,F(t)(q_2-t)\,dt$. For $t \in (q_1,q_2)$, $F(t) \geq w_1$ and $q_2 - t > 0$, so the contribution is strictly positive.

(c) The inner integral $q \mapsto \int_0^q t\,F(t)(q-t)\,dt$ is a bounded continuous function of q for fixed μ, and $\mathcal{R}$ is an integral of a bounded function against μ. Dominated convergence applies under weak convergence. □

Remark 4.9. $\mathcal{R}$ is quadratic in μ but *not convex*. Taking $\mu_0 = \delta_0$, $\mu_1 = \delta_1$, $\mu = \frac{1}{2}(\mu_0+\mu_1)$, one computes $\mathcal{R}(\mu_0) = \mathcal{R}(\mu_1) = 0$ while $\mathcal{R}(\mu) = 1/24 > 0$, so $\mathcal{R}(\frac{1}{2}(\mu_0+\mu_1)) > \frac{1}{2}(\mathcal{R}(\mu_0)+\mathcal{R}(\mu_1))$. This reflects the spread-penalty interpretation: $\mathcal{R}$ vanishes on point masses and grows when mass is split across distinct locations.

We verified (16) numerically by computing $\Phi_\mu(0,0)$ from (9) for atomic μ at small β, extracting the β^6 coefficient by Richardson extrapolation, and comparing against $m_3/3 - \mathcal{R}(\mu)$ computed from (17).

μ	m_2	m_3	$\varphi_3(0,0)$	$\mathcal{R}(\mu)$
$\delta_{0.5}$	0.2500	0.1250	1/24	0
$\frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$	0.5000	0.5000	1/8	1/24
$\frac{1}{2}\delta_{0.2} + \frac{1}{2}\delta_{0.8}$	0.3400	0.2600	0.0687	0.0180
$\frac{1}{3}\delta_0 + \frac{1}{3}\delta_{0.5} + \frac{1}{3}\delta_1$	0.4167	0.3750	0.0949	0.0301

TABLE 1. Numerical verification of $\varphi_3(0,0) = m_3/3 - \mathcal{R}(\mu)$ for several atomic μ. Values of φ_3 are extracted from direct numerical solution of the Parisi PDE (9) by Richardson extrapolation at small β. Values of $\mathcal{R}(\mu)$ are computed from (17). Agreement is to machine precision in each case.

Remark 4.10 (Physical interpretation). The functional $\mathcal{R}(\mu)$ is a spread penalty. Zero at point masses and growing with support size, atom separation, and weight balance. At order β^6 the Parisi functional rewards large m_3 but penalizes spread beyond what m_3 captures, so the minimizer selects the most concentrated measure consistent with the moment data. This is the mechanism by which the functional favors replica-symmetric solutions at high temperature. Any spread away from a point mass costs $\beta^6\,\mathcal{R}(\mu) > 0$ without compensating moment benefit until $\beta > 1$.

4.4. **The full decomposition and the AT instability.**

Theorem 4.11 (High-temperature expansion of the Parisi functional)**.** *For every* $\mu \in \mathcal{P}([0,1])$ *with moments* $m_k = \int q^k\,d\mu$*,*

$$\mathcal{P}_\beta(\mu) \;=\; \log 2 + \frac{\beta^2}{4} + m_2\,\frac{\beta^2(1-\beta^2)}{4} + \beta^6\Big(\frac{m_3}{3} - \mathcal{R}(\mu)\Big) + O(\beta^8). \tag{18}$$

The remainder is uniform in μ*. There exists* $C < \infty$ *(independent of* μ*) such that the difference between* $\mathcal{P}_\beta(\mu)$ *and the displayed expression is at most* $C\beta^8$ *for* $\beta \le 1$*.*

Proof. From Proposition 4.1, $\mathcal{P}_\beta(\mu) = \log 2 - \frac{\beta^2}{4}(1-m_2) + \Phi_\mu(0,0)$. Substituting $\Phi_\mu(0,0) = \frac{\beta^2}{2} - \frac{\beta^4 m_2}{4} + \beta^6\,\varphi_3(0,0) + O(\beta^8)$.

$$\begin{aligned}\mathcal{P}_\beta(\mu) &= \log 2 - \tfrac{\beta^2}{4} + \tfrac{\beta^2 m_2}{4} + \tfrac{\beta^2}{2} - \tfrac{\beta^4 m_2}{4} + \beta^6\,\varphi_3(0,0) + O(\beta^8)\\ &= \log 2 + \tfrac{\beta^2}{4} + \tfrac{\beta^2 m_2}{4} - \tfrac{\beta^4 m_2}{4} + \beta^6\big(\tfrac{m_3}{3} - \mathcal{R}(\mu)\big) + O(\beta^8)\\ &= \log 2 + \tfrac{\beta^2}{4} + m_2\,\tfrac{\beta^2(1-\beta^2)}{4} + \beta^6\big(\tfrac{m_3}{3} - \mathcal{R}(\mu)\big) + O(\beta^8).\end{aligned}$$

Uniformity of the remainder follows from the convergence result below. □

Proposition 4.12 (Convergence of the β^2 series)**.** *There exists* $R_0 > 0$ *depending only on the terminal data* $\log\cosh$ *such that, for every* $\mu \in \mathcal{P}([0,1])$*, the series*

$$\Phi_\mu(s,x) \;=\; \sum_{n\ge 0} \beta^{2n}\,\varphi_n(s,x)$$

converges absolutely for $\beta^2 < R_0$*, uniformly in* $(s,x,\mu) \in [0,1] \times K \times \mathcal{P}([0,1])$ *for any compact* $K \subset \mathbb{R}$*. In particular, the expansion* (18) *is a genuine convergent Taylor series in* β^2 *on* $(0, R_0^{1/2})$*, not merely an asymptotic expansion, and the* $O(\beta^8)$ *remainder is bounded uniformly in* μ*.*

Proof. Each step of the iterative scheme in Proposition 4.3 consists of a Gaussian convolution with variance $\beta^2(q_{i+1} - q_i)$ and a Cole–Hopf-like operation. On finitely atomic μ, $\Phi_\mu(s,x)$ is an analytic function of β^2 with radius of convergence at least R_0, where R_0 is determined by the terminal data $\log\cosh$ and not by μ (the CDF $F(s) \in [0,1]$ gives uniform bounds on the PDE coefficients). Finitely atomic measures are dense in $\mathcal{P}([0,1])$ under weak convergence, and the iteration is continuous in μ, so the analyticity and the radius of convergence pass to general μ by a standard approximation argument. Equivalently, one can apply the stochastic control representation of [AC15], $\Phi_\mu(0,0)$ is the value of an optimal-control problem whose cost is analytic in β^2 with coefficients bounded uniformly in μ, hence analytic in β^2 with a μ-independent radius of convergence. □

Corollary 4.13 (de Almeida–Thouless instability)**.** *The coefficient of* m_2 *in* (18) *is* $\beta^2(1-\beta^2)/4$*, which changes sign at* $\beta = 1$*. For* $\beta < 1$ *the coefficient is positive and the minimizer wants* $m_2 = 0$*, consistent with a point mass* $\mu^* = \delta_0$*. The replica-symmetric phase. For* $\beta > 1$ *the coefficient is negative and the minimizer requires* $m_2 > 0$*, forcing the overlap distribution away from zero. The RSB transition. This recovers the classical AT line* [dT78] *from purely moment-theoretic data, without appeal to the replica trick or a Hessian computation.*

Remark 4.14 (Role of $\mathcal{R}$ in the transition). Near $\beta = 1$, the spread penalty $\mathcal{R}(\mu)$ provides a restoring force toward concentrated measures. The competition between m_2 (rewarding spread at $\beta > 1$) and $\mathcal{R}$ (penalizing spread at β^6) gives a variational picture of the RSB transition, μ^* is the optimal balance between moment energy and spread penalty. Just above $\beta = 1$, μ^* develops a small variance while remaining nearly a point mass. As β grows, m_2 grows and the measure spreads, transitioning to multi-atom and eventually continuous support.

4.5. **Consequences for the moment class.**

Theorem 4.15 (Moment-determined terms)**.** *On the moment class* $\mathcal{M}_K$*.*

(a) For $K \geq 2$*, the terms through* $O(\beta^4)$ *in* $\mathcal{P}_\beta(\mu)$ *are constant across* $\mathcal{M}_K$*, and the variation of* $\mathcal{P}_\beta$ *is at most* $O(\beta^6)$*.*

(b) For $K \geq 3$*, the* $m_3/3$ *part of the* β^6 *coefficient is also fixed, so*

$$|\mathcal{P}_\beta(\mu) - \mathcal{P}_\beta(\nu)| \;\leq\; \beta^6\,|\mathcal{R}(\mu) - \mathcal{R}(\nu)| + O(\beta^8), \qquad \mu,\nu \in \mathcal{M}_K.$$

(c) Since $\mathcal{R}$ *is continuous on* $\mathcal{P}([0,1])$ *in* W_1 *(Proposition 4.8(c)) and* $W_1(\mu,\nu) \leq 1/(K+1)$ *on* $\mathcal{M}_K$ *(Lemma 3.9), there is* C *with* $|\mathcal{R}(\mu) - \mathcal{R}(\nu)| \leq C/(K+1)$ *for* $\mu,\nu \in \mathcal{M}_K$*.*

Proof. (a) and (b) are immediate from (18) and the definition of $\mathcal{M}_K$. (c) follows from Lemma 3.9 and uniform continuity of $\mathcal{R}$ on the compact space $\mathcal{P}([0,1])$. □

Corollary 4.16 (High-temperature improvement)**.** *For $\beta \le 1$ and $K \ge 2$,*

$$\sup_{\mu,\nu\in\mathcal{M}_K} |\mathcal{P}_\beta(\mu) - \mathcal{P}_\beta(\nu)| \;\le\; C_2\,\beta^6$$

where $C_2 = \sup_\mu \mathcal{R}(\mu) < \infty$.

For $\beta < 1$, Corollary 4.16 tightens the uniform bound of Theorem 3.11 from $O(\beta^2/K)$ to $O(\beta^6)$, independent of K. At high temperature, two moments almost suffice for the free energy. The third and higher moments contribute only through the spread penalty $\mathcal{R}$, and their effect is $O(\beta^6)$. At low temperature, the uniform $O(\beta^2/K)$ bound is needed and the expansion ceases to be useful.

Higher orders. The β^8 coefficient can be computed by the same procedure. Collecting β^8 in (9) with the previously computed φ_2, φ_3 gives a linear PDE for φ_4 whose value at $(0,0)$ decomposes as

$$\varphi_4(0,0) \;=\; P_4(m_2, m_3, m_4) - \mathcal{R}_4(\mu),$$

where P_4 is a polynomial in the moments through m_4 (leading term $m_4/4$ plus cross-terms in m_2, m_3) and $\mathcal{R}_4(\mu) \ge 0$ vanishes at point masses. A closed form for $\mathcal{R}_4$ analogous to (17) is expected but we have not pursued it here. Numerically $\mathcal{R}_4(\delta_q) = 0$ and $\mathcal{R}_4 > 0$ for measures with multiple atoms, consistent with the spread-penalty interpretation.

The general pattern at order β^{2n} is conjectured to be

$$\varphi_n(0,0) \;=\; P_n(m_2, \ldots, m_n) + \mathcal{R}_n(\mu),$$

with P_n a polynomial in moments through m_n and $\mathcal{R}_n$ a non-negative CDF-dependent correction vanishing at point masses. Whether this pattern persists, and whether the $\mathcal{R}_n$ admit closed-form expressions or probabilistic interpretations (in terms of the Ruelle cascade or free cumulants), is an open question.

5. Synthesis

Sections 3 and 4 controlled the Parisi free energy using only the single-overlap moments $m_k = \int q^k \, d\mu^*$. The joint law of the full overlap array $(R_{ab})_{a,b\ge1}$ is a much richer object, governed by the *Ghirlanda–Guerra identities* [GG98a, GG98b] and the Panchenko ultrametricity theorem [Pan13a]. We close by showing that these two structures are independent. Moments control the marginal axis. GG controls the joint axis. *For the free energy, moments suffice. For ultrametricity, GG is needed.*

5.1. **The Ghirlanda–Guerra identities.** Let $(R_{ab})_{a,b\geq 1}$ be an exchangeable random array on $[0,1]$ (or $[-1,1]$ before spin-flip symmetrization). The joint law of $(R_{ab})_{a,b\leq n}$ is invariant under permutations of $\{1,\ldots,n\}$ for every n.

Definition 5.1 (GG identities)**.** An exchangeable array $(R_{ab})_{a,b\geq 1}$ satisfies the *Ghirlanda–Guerra identities* if for every bounded measurable $f:\mathbb{R}^{\binom{n}{2}}\to\mathbb{R}$ and every $n\geq 2$,

$$\mathbb{E}[R_{1,n+1}\,f(R_{12},\ldots,R_{n-1,n})] \;=\; \frac{1}{n}\,\mathbb{E}[R_{12}]\,\mathbb{E}[f] \;+\; \frac{1}{n}\sum_{j=2}^{n}\mathbb{E}[R_{1j}\,f]. \tag{19}$$

The identity says that $R_{1,n+1}$, conditional on the previous overlaps $(R_{12},\ldots,R_{n-1,n})$, is distributed as a $(1/n,(n-1)/n)$ mixture. With probability $1/n$ it behaves as an independent copy of R_{12} (reading off the overall marginal), and with probability $(n-1)/n$ it copies one of the already-observed overlaps R_{1j} uniformly. In the SK model the GG identities hold for the limiting joint law of the overlap array, by Gaussian integration by parts applied to the Hamiltonian under a small random temperature perturbation [GG98a, Pan13b].

The GG identities are not implied by exchangeability alone. Exchangeability gives only that the joint law is invariant under replica permutations, while GG encodes a specific structural relation between overlap moments. Together with exchangeability, however, they are strong enough to imply *ultrametricity*.

Theorem 5.2 (Panchenko [Pan13a])**.** *If an exchangeable array $(R_{ab})_{a,b\geq 1}$ on $[-1,1]$ satisfies the GG identities, then*

$$\mathbb{P}\big(R_{23}\geq \min(R_{12},R_{13})\big) \;=\; 1.$$

Equivalently, for every realization, the three-point overlaps (R_{12},R_{13},R_{23}) form the sides of an ultrametric triangle. The two smaller values are equal.

Panchenko's theorem is the rigorous vindication of Parisi's prediction [Par80, MPS$^+$84] that the spin-glass pure states are organized in an ultrametric tree, and it is one of the deep results of the post-Talagrand theory. The proof, via careful manipulation of (19) and the Aldous–Hoover representation of exchangeable arrays, is in [Pan13a, Pan13b].

5.2. **The Ruelle probability cascade.** For every probability measure $\mu\in\mathcal{P}([0,1])$ there is a canonical GG-compatible exchangeable array with single-overlap marginal μ, the *Ruelle probability cascade* of [Rue87]. We sketch the construction.

Write μ as the law of $\zeta(U)$ for U uniform on $[0,1]$, where $\zeta:[0,1]\to[0,1]$ is the quantile function of μ. The RPC is a random probability measure G supported on a hierarchical tree of "pure states," constructed as follows. At each branching level $s\in[0,1]$, the tree branches with intensity governed by $\zeta(s)$, and the weights on each branch evolve as independent Poisson–Dirichlet processes. Two leaves σ^a,σ^b of the tree sampled independently from G have overlap

$$R_{ab} \;=\; \zeta(S_{ab}),$$

where S_{ab} is the branching level at which the ancestral lineages of σ^a and σ^b diverge. The distribution of S_{ab} is determined by the branching structure and produces $\mathrm{Law}(R_{ab}) = \mu$. The GG identities hold because of the self-similar branching. The next overlap $R_{1,n+1}$, conditional on the previous overlaps, sees a fresh branching level whose law is the mixture described by (19). Detailed verification is in [Pan13b, Theorem 2.17].

The RPC is ultrametric by construction. The overlap $R_{ab} = \zeta(S_{ab})$ depends only on the ancestral coalescence level, and for three leaves $\sigma^a, \sigma^b, \sigma^c$, the coalescence levels satisfy $S_{23} \geq \min(S_{12}, S_{13})$ automatically (two lineages cannot diverge later than both individually diverge from a third).

Theorem 5.3 (RPC realizes every μ, [Rue87, Pan13b]). *For every $\mu \in \mathcal{P}([0,1])$, the RPC with parameter μ is a GG-compatible exchangeable array with single-overlap marginal μ.*

Conversely, Panchenko's theorem and its companion rigidity statement say that any GG-compatible exchangeable array is the RPC with parameter $\mu = \mathrm{Law}(R_{12})$. The single-overlap marginal determines the joint law uniquely. The joint axis is collapsed to a single point above each μ.

5.3. **GG transparency.** Let $\mathcal{G}$ be the set of laws of GG-compatible exchangeable arrays on $[0,1]$, and $\pi : \mathcal{G} \to \mathcal{P}([0,1])$ the projection to the single-overlap marginal $\mathrm{Law}(R_{12})$.

Theorem 5.4 (GG transparency). $\pi(\mathcal{G}) = \mathcal{P}([0,1])$. *Equivalently, the GG identities impose no constraint on the single-overlap marginal.*

Proof. Let $\mu \in \mathcal{P}([0,1])$ be arbitrary. Theorem 5.3 produces a GG-compatible exchangeable array with single-overlap marginal μ, so $\mu \in \pi(\mathcal{G})$. The reverse inclusion is trivial. □

Corollary 5.5. $\mathcal{M}_K \subseteq \pi(\mathcal{G})$ *for every K. Every measure in the moment class $\mathcal{M}_K$ is the single-overlap marginal of some GG-compatible exchangeable array.*

This is the central structural observation of the section. Despite the extraordinary strength of the GG identities, strong enough to force ultrametricity of the full array, they do not restrict the single-overlap marginal at all. The constraint operates at the level of the joint law, not at the level of any marginal.

5.4. **Independence of the two axes.** Combining the previous results separates the two axes of the overlap structure.

- Marginal axis. The single-overlap marginal $\mu \in \mathcal{P}([0,1])$ determines the Parisi free energy via the Parisi functional (Theorem 1.2). The first K overlap moments of μ^* determine $\mathcal{P}_\beta(\mu^*)$ to $4\beta^2/(K+1)$ (Theorem 3.11). No knowledge of the joint array is needed.

- Joint axis. Given μ, the GG identities and exchangeability together determine the joint law of the overlap array uniquely as the Ruelle cascade with parameter μ, which is ultrametric (Theorem 5.2). No knowledge of higher moments of μ beyond what specifies μ is needed.

Moments control the marginal axis, GG controls the joint axis, and neither informs the other. For computing the Parisi free energy, moments suffice and GG is not needed. For ultrametricity of the joint overlap array, GG is needed and moments alone say nothing.

This clarifies which SK-model questions reduce to moment computation and which require more sophisticated overlap-array machinery. Moments alone determine the free energy density $f(\beta, h)$, the internal energy, magnetization, and every other Lipschitz functional of μ^*, the convergence rate of the K-step RSB approximation, and the variance of F_N to leading order in $1/N$ (the last being controlled by m_2 alone via Appendix B). Moments alone do not determine ultrametricity of the overlap array, nor the joint distribution of (R_{12}, R_{13}, R_{23}), nor higher Gibbs correlations such as $\mathbb{E}[\langle\sigma_i\sigma_j\sigma_k\rangle^2]$. All of the latter require GG identities in addition to the marginal.

5.5. **Open directions.** Several directions suggest themselves.

Higher-order spread penalties. The pattern at order β^{2n} appears to be $\varphi_n(0,0) = P_n(m_2, \ldots, m_n) + \mathcal{R}_n(\mu)$ with P_n a polynomial and $\mathcal{R}_n \geq 0$ vanishing on point masses. Whether the $\mathcal{R}_n$ admit closed forms analogous to $\mathcal{R}(\mu) = \iint t\, F(t)(q-t)\, dt\, d\mu(q)$ for all n, and whether they relate to free cumulants of the overlap array or to the Ruelle cascade tree structure, is open.

Exponential convergence for analytic μ^*. When the Parisi quantile function $q_{\mu^*} : [0,1] \to [0,1]$ is analytic on its support interval, which holds throughout the full-RSB regime by [AC15, JT17], the Jackson–Stechkin approximation used in Lemma 3.9 can be sharpened. Bernstein's theorem [DL93, Ch. 4] states that a function analytic on a neighborhood of $[0,1]$ admits polynomial approximation of degree K with error $\leq Ce^{-cK}$, where c depends on the domain of analyticity. Applying this to the quantile function.

$$\inf_{\deg p \leq K} \|q_{\mu^*} - p\|_\infty \;\leq\; Ce^{-cK},$$

which (via the quantile representation $W_1(\mu, \nu) = \int |q_\mu - q_\nu|\, dx$, Appendix C) gives $W_1(\mu^*, \nu_K^*) \leq Ce^{-cK}$ for the optimal K-atomic approximation ν_K^*, and hence

$$\mathcal{P}_\beta^{(K)} - \mathcal{P}_\beta(\mu^*) \;\leq\; 4\beta^2\, C\, e^{-cK}.$$

The exponential rate is conjectural in that it relies on a quantitative version of the analyticity result whose constants we have not tracked. A full proof would require bounding the domain of analyticity of q_{μ^*} in β, a nontrivial task.

Fluctuations beyond the variance. The Poincaré bound in Appendix B controls $\mathrm{Var}(F_N)$ by m_2. The higher cumulants of F_N, or equivalently the large-deviation function $\Lambda(t) = \lim_N \frac{1}{N} \log \mathbb{E}[e^{tNF_N}]$, are subtler. Whether Λ is a W_1-Lipschitz functional of μ^*, and therefore determined to $O(\beta^2/K)$ by K moments, is open. An affirmative answer would give moment-theoretic control of the full large-deviation profile.

The p-spin generalization. For the mixed p-spin model with covariance $\xi(q) = \sum_p \frac{\beta_p^2}{p!}\, q^p$, the Parisi functional is

$$\mathcal{P}(\mu) \;=\; \log 2 + \Phi_\mu(0,0) - \tfrac{1}{2}\int_0^1 s\,\xi''(s)\,F(s)\,ds,$$

and the Lipschitz constant in (4) becomes $2\xi''(1) = 2\sum_p \beta_p^2(p-1)/(p-2)!$. Theorem 3.11 generalizes verbatim with $4\beta^2$ replaced by $2\xi''(1)$, so

$$0 \;\le\; \mathcal{P}^{(K)} - \mathcal{P}(\mu^*) \;\le\; \frac{2\xi''(1)}{K+1}.$$

The integral piece no longer simplifies to depend only on m_2. For non-pure ξ it involves $\int_0^1 s\,\xi''(s)F(s)\,ds$, which by Fubini equals $\sum_p \frac{\beta_p^2(p-1)}{p!}\,\frac{(p-1)-p\,m_{p-1}+m_p}{p\cdot(p-1)}$ or similar, bringing in the moments m_{p-1} for each p with $\beta_p \neq 0$. The high-temperature expansion becomes a joint expansion in the β_p^2 with coefficients polynomial in $m_2, m_3, \ldots$ and shape-dependent corrections $\mathcal{R}_n$ analogous to the SK case. The separation theorem (Theorem 4.2) generalizes to the statement that $\mathcal{P}$ restricted to $\mathcal{M}_K$ for K at least as large as the top relevant moment is entirely PDE-side.

Algorithmic implications. From the viewpoint of statistical inference and optimization, the moment-class framework suggests that simulation of the SK model need only track a finite number of overlap moments to determine the free energy to prescribed accuracy. For an m-atomic target, $2m$ moments pin down μ^* exactly. In continuous-μ^* regimes, K moments give $O(\beta^2/K)$ accuracy. This could inform the design of moment-targeted Monte Carlo estimators or variational inference algorithms on moment-constrained families.

5.6. **Conclusion.** The moment viewpoint organizes the Parisi theory around a single principle. The overlap moments are the natural finite-dimensional summary of the Parisi measure. The moment class $\mathcal{M}_K$ is convex and weakly compact with $(K+1)$-atomic extreme points (Theorem 3.5). The filtration $\mathcal{M}_1 \supset \mathcal{M}_2 \supset \cdots$ collapses to $\{\mu^*\}$ at the sharp threshold $K = 2m$ when μ^* is m-atomic (Corollary 3.7). And the Parisi free energy varies by at most $4\beta^2/(K+1)$ across $\mathcal{M}_K$, giving rate $O(\beta^2/K)$ for the K-step RSB approximation (Theorems 3.11, 3.13). The explicit–PDE decomposition (Proposition 4.1) isolates the m_2-dependent part of $\mathcal{P}_\beta$, and expanding the PDE piece in β^2 shows overlap moments entering order by order, m_2 at β^4, $m_3/3 - \mathcal{R}(\mu)$ at β^6 (Theorem 4.7), with $\mathcal{R} \ge 0$ a spread penalty vanishing on point masses (Proposition 4.8). The sign change of the m_2 coefficient at $\beta = 1$ recovers the de Almeida–Thouless instability (Corollary 4.13) from purely moment-theoretic data. The Ghirlanda–Guerra identities, strong enough to force ultrametricity of the joint overlap array (Theorem 5.2), impose no constraint on the single-overlap marginal (Theorem 5.4). Moments and GG are independent. For the free energy, moments suffice. For ultrametricity, GG is needed.

6. Worked Examples

To make the preceding theorems concrete we work through three cases end-to-end. The replica-symmetric regime at $\beta = 0.8$, a 1-step RSB construction at $\beta = 1.5$, and a full-RSB discretization at $\beta = 2$.

6.1. Replica symmetry at $\beta = 0.8$.

At $\beta = 0.8$ and $h = 0$, the AT line (Corollary 4.13) is stable. Since $\beta < 1$, the coefficient of m_2 in the β^2-expansion is positive, and the minimizer of $\mathcal{P}_\beta$ wants m_2 as small as possible. The RS saddle-point equation $q = \mathbb{E}[\tanh^2(\beta\sqrt{q}\,Z)]$ with $h = 0$ has only the trivial solution $q_* = 0$, so $\mu^* = \delta_0$.

Moment collapse. Since $\mu^* = \delta_0$ is one-atomic ($m = 1$), Corollary 3.7 gives $\mathcal{M}_K = \{\delta_0\}$ for all $K \geq 2$. The moment class collapses immediately, and the free-energy control bound $\mathcal{P}_\beta(\nu) - \mathcal{P}_\beta(\mu^*) \leq 4\beta^2/(K+1) = 2.56/(K+1)$ is vacuous for $K \geq 2$ (there are no other ν). For $K = 1$, $\mathcal{M}_1$ consists of all probability measures with mean zero, a nontrivial set with W_1-diameter 1 (from δ_0 to $\frac{1}{2}(\delta_{-1} + \delta_1)$ on the symmetric extension, or from δ_0 to $\frac{1}{2}\delta_1$ on $[0, 1]$ after spin-flip symmetrization).

Free energy. The RS free energy is

$$\mathcal{P}_{0.8}(\delta_0) \;=\; \log 2 + \tfrac{(0.8)^2}{4} \;=\; \log 2 + 0.16 \;\approx\; 0.853.$$

High-temperature expansion. With $m_2 = m_3 = 0$ and $\mathcal{R}(\delta_0) = 0$, Theorem 4.11 gives

$$\mathcal{P}_{0.8}(\delta_0) \;=\; \log 2 + \tfrac{(0.8)^2}{4} + O((0.8)^8) \;\approx\; 0.853 + O(0.168),$$

matching the exact RS value to within a small $O(\beta^8)$ correction.

6.2. One-step RSB at $\beta = 1.5$.

At $\beta = 1.5$ and $h = 0$, the replica-symmetric saddle point is unstable ($\beta > 1$), and the Parisi minimizer has nontrivial support. In the 1-step RSB regime relevant for moderate β, μ^* is two-atomic, $\mu^* = w\,\delta_{q_0} + (1 - w)\,\delta_{q_1}$ with $0 < q_0 < q_1 \leq 1$. For $\beta = 1.5$ on the Ising SK model, numerical solution of the 1-step Parisi equations gives approximately $q_0 \approx 0$, $q_1 \approx 0.71$, $w \approx 0.33$ (the Parisi parameter), so $\mu^* \approx 0.33\,\delta_0 + 0.67\,\delta_{0.71}$ with moments

$$m_1 \approx 0.475, \qquad m_2 \approx 0.337, \qquad m_3 \approx 0.239, \qquad m_4 \approx 0.170.$$

Moment collapse. Since μ^* has $m = 2$ atoms, $\mathcal{M}_K = \{\mu^*\}$ for $K \geq 2m = 4$. For $K = 3$, $\mathcal{M}_3 \supsetneq \{\mu^*\}$ and the explicit construction of the proof of Theorem 3.5(d) yields a three-atom $\nu \in \mathcal{M}_3 \setminus \{\mu^*\}$. Fix any $a_0 \in (0, 1) \setminus \{0, 0.71\}$ and compute weights from the moment-matching Vandermonde system. Setting $a_0 = 0.4$ gives

$$\nu \;\approx\; 0.27\,\delta_0 + \varepsilon\,\delta_{0.4} + (0.73 - \varepsilon)\,\delta_{0.71}$$

with ε small chosen so the first three moments of ν match those of μ^*.

Free-energy bound. Theorem 3.11 gives $|\mathcal{P}_{1.5}(\nu) - \mathcal{P}_{1.5}(\mu^*)| \le 4(1.5)^2/4 = 2.25$. This is loose. The actual free-energy gap is numerically $O(10^{-2})$, reflecting the nearly one-dimensional nature of $\mathcal{M}_3$ near μ^*. The $O(\beta^2/K)$ bound is conservative at small K and becomes sharp only asymptotically in K (cf. Theorem 3.16).

Separation theorem. Since m_2 is fixed across $\mathcal{M}_3$, Theorem 4.2 gives $\mathcal{P}_{1.5}(\nu) - \mathcal{P}_{1.5}(\mu^*) = \Phi_\nu(0,0) - \Phi_{\mu^*}(0,0)$. At order β^6, Theorem 4.7 and the numerical values give $\varphi_3(0,0;\nu) - \varphi_3(0,0;\mu^*) = \mathcal{R}(\mu^*) - \mathcal{R}(\nu)$. For our $\mu^* \approx 0.33\,\delta_0 + 0.67\,\delta_{0.71}$, direct computation of (17) gives $\mathcal{R}(\mu^*) \approx 0.0156$. For ν with the added $\delta_{0.4}$ atom, $\mathcal{R}(\nu)$ is within 10^{-3} of this value.

6.3. **Full RSB at $\beta = 2$.** At $\beta = 2$ and $h = 0$, the Parisi solution is full-RSB, μ^* has continuous support on an interval $[0, q_{\mathrm{EA}}]$ together with a point mass at q_{EA}. Numerically $q_{\mathrm{EA}} \approx 0.96$ and the density on $[0, 0.96]$ grows roughly linearly near $q = 0$ and concentrates toward q_{EA}. In particular, μ^* has infinitely many "atoms," so the filtration $\mathcal{M}_K$ never collapses, and the convergence $\mathcal{M}_K \searrow \{\mu^*\}$ is genuinely infinite-dimensional.

RSB hierarchy. Theorem 3.13 gives $0 \le \mathcal{P}_2^{(K)} - \mathcal{P}_2(\mu^*) \le 4 \cdot 4/(K+1) = 16/(K+1)$, recorded for concrete K in the table below.

K	$16/(K+1)$	$\mathcal{P}_2^{(K)}$ (computed)	$\mathcal{P}_2^{(K)} - \mathcal{P}_2(\mu^*)$
1	8.00	0.9845	0.0095
2	5.33	0.9770	0.0020
4	3.20	0.9755	0.0005
8	1.78	0.9750	0.0001
∞	0	0.9750 (true)	0

(These values are schematic. Exact figures depend on the numerical scheme.)

The theoretical bound $16/(K+1)$ is extremely loose. The actual error decays much faster than $1/K$, consistent with the exponential-convergence conjecture discussed in Section 5.5. Quantitatively, for analytic μ^* the Bernstein theorem [DL93] gives polynomial approximation error Ce^{-cK}, so $\mathcal{P}_\beta^{(K)} - \mathcal{P}_\beta(\mu^*) \le 4\beta^2 Ce^{-cK}$.

Two-sided rate. The lower bound of Theorem 3.16 applies (μ^* has an absolutely continuous component) and gives $\mathcal{P}_2^{(K)} - \mathcal{P}_2(\mu^*) \ge c(\beta)/(K+1)$ for some $c(\beta) > 0$. In the full-RSB regime the upper and lower bounds differ by the constant ratio $4\beta^2/c(\beta)$, with no improvement in the polynomial rate accessible from the generic argument.

References

[AC15] Antonio Auffinger and Wei-Kuo Chen, *The Parisi formula has a unique minimizer*, Communications in Mathematical Physics **335** (2015), 1429–1444.

[BLM13] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart, *Concentration inequalities: A nonasymptotic theory of independence*, Oxford University Press, 2013.

[DL93] Ronald DeVore and George Lorentz, *Constructive approximation*, Springer, 1993.

[dT78] J. R. L. de Almeida and D. J. Thouless, *Stability of the Sherrington–Kirkpatrick solution of a spin glass model*, Journal of Physics A **11** (1978), 983–990.

[GG98a] Stefano Ghirlanda and Francesco Guerra, *General properties of overlap probability distributions in disordered spin systems*, Journal of Physics A **31** (1998), 9149–9155.

[GG98b] ______, *General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity*, Journal of Physics A **31** (1998), 9149–9155.

[GT02] Francesco Guerra and Fabio Lucio Toninelli, *The thermodynamic limit in mean field spin glass models*, Communications in Mathematical Physics **230** (2002), 71–79.

[Gue03] Francesco Guerra, *Broken replica symmetry bounds in the mean field spin glass model*, Communications in Mathematical Physics **233** (2003), 1–12.

[JT17] Aukosh Jagannath and Ian Tobasco, *Some properties of the phase diagram for mixed p-spin glasses*, Probab. Theory Related Fields **167** (2017), 615–672.

[KS66] Samuel Karlin and William Studden, *Tchebycheff systems: With applications in analysis and statistics*, Wiley, 1966.

[MPS+84] Marc Mézard, Giorgio Parisi, Nicolas Sourlas, Gérard Toulouse, and Miguel Angel Virasoro, *Nature of the spin-glass phase*, Physical Review Letters **52** (1984), 1156–1159.

[Pan13a] Dmitry Panchenko, *The Parisi ultrametricity conjecture*, Annals of Mathematics **177** (2013), 383–393.

[Pan13b] ______, *The Sherrington–Kirkpatrick model*, Springer, 2013.

[Par79] Giorgio Parisi, *Infinite number of order parameters for spin-glasses*, Physical Review Letters **43** (1979), 1754–1756.

[Par80] ______, *The order parameter for spin glasses: a function on the interval* 0–1, Journal of Physics A **13** (1980), 1101–1112.

[Ric57] Hans Richter, *Parameterfreie Abschätzung und Realisierung von Erwartungswerten*, Blätter der DGVFM **3** (1957), 147–161.

[Rog58] Werner W. Rogosinski, *Moments of non-negative mass*, Proc. Roy. Soc. A **245** (1958), 1–27.

[Rue87] David Ruelle, *A mathematical reformulation of Derrida's REM and GREM*, Communications in Mathematical Physics **108** (1987), 225–239.

[SK75] David Sherrington and Scott Kirkpatrick, *Solvable model of a spin-glass*, Physical Review Letters **35** (1975), 1792–1796.

[Tal06] Michel Talagrand, *The Parisi formula*, Annals of Mathematics **163** (2006), 221–263.

[Tal11] ______, *Mean field models for spin glasses, volume I*, Springer, 2011.

Appendix A. Derivation of the Parisi PDE

This appendix sketches the derivation of the Parisi PDE (2) from the K-step RSB computation, following the standard physics derivation as cleaned up in [Pan13b, AC15]. We do not aim for full rigor. Our purpose is to situate the PDE the thesis uses against its classical origin.

A.1. From K-step RSB to a chain of Gaussian convolutions.

The K-step replica-symmetry-breaking ansatz parametrizes the overlap matrix by a hierarchical block structure. For integer n (the replica number), the matrix $Q^{(K)}$ has entries

$$Q^{(K)}_{\alpha\beta} \;=\; q_j \quad \text{for } \alpha \sim_{j+1} \beta,\ \alpha \not\sim_j \beta,$$

where $\sim_0, \sim_1, \ldots, \sim_{K+1}$ is a chain of increasingly coarse equivalence relations on $\{1, \ldots, n\}$ with block sizes $m_0 = 1 \le m_1 \le \cdots \le m_K \le m_{K+1} = n$. The parameters are the overlap values $q_0 \le q_1 \le \cdots \le q_K$ and the block sizes $m_1, \ldots, m_K$.

Applying the replica method to the free energy of the SK model and evaluating the quadratic overlap sum $\sum_{\alpha<\beta} Q^{(K)}_{\alpha\beta}\sigma_\alpha\sigma_\beta$ on the Parisi ansatz gives a hierarchy of Gaussian couplings, one at each level of the ansatz. The key algebraic identity is

$$\sum_{\alpha<\beta} Q^{(K)}_{\alpha\beta}\,\sigma_\alpha\sigma_\beta \;=\; \frac{1}{2}\sum_{j=0}^{K}(q_j - q_{j-1}) \sum_{\sim_{j+1}\text{-class } B} \Big(\sum_{\alpha\in B}\sigma_\alpha\Big)^2 \;-\; \frac{n\,q_K}{2},$$

with $q_{-1} = 0$. This decomposition rewrites the Parisi-matrix coupling as a sum of independent squared block sums, one for each level.

Applying the Gaussian identity $\exp(\frac{1}{2}c^2x^2) = \mathbb{E}[\exp(cxZ)]$ with $Z \sim \mathcal{N}(0,1)$ at each level introduces a fresh Gaussian for each level's block sum. After $K+1$ rounds of Gaussian integration, the replica trace decouples and the remaining sum over each $\sigma_\alpha \in \{-1,+1\}$ gives a $2\cosh$ at the innermost level.

Taking the $n \to 0$ limit of the resulting nested expression produces the recursion (13) in Proposition 4.3.

$$f_i(x) \;=\; \frac{1}{\rho_i}\,\log\mathbb{E}\big[\exp(\rho_i\, f_{i+1}(x+\sigma_i Z))\big], \qquad \sigma_i^2 = \beta^2(q_{i+1}-q_i),$$

where $\rho_i = 1 - m_i$ (in the continued $n \to 0$ convention). The top of the recursion is $f_{K+1}(x) = \log\cosh(x)$.

A.2. **The continuous limit.** When the step-function overlap $q^{(K)}$ becomes dense in $\mathcal{P}([0,1])$ (i.e., $K \to \infty$), the discrete recursion of Gaussian convolutions with log-exp steps merges into a differential operator equation.

The heat-semigroup interpretation. Each Gaussian expectation $\mathbb{E}[g(x+\sigma_i Z)]$ is the action of the heat semigroup $\exp(\frac{\sigma_i^2}{2}\partial_{xx})$ on g. Applied at step i, this corresponds to heat flow for time $\beta^2(q_{i+1}-q_i)$. The nonlinear step $g \mapsto \frac{1}{\rho}\log\mathbb{E}[\exp(\rho g)]$ is a Cole–Hopf transformation that, in the limit $\Delta q \to 0$, contributes the nonlinear term $\rho\,(\partial_x g)^2/2$.

Formal derivation. Let $F(s)$ be the CDF of the overlap measure. In the continuous limit, the block weights ρ_i approach $F(s)$ for s at the step position. The recursion becomes the differential equation

$$\partial_s\Phi(s,x) \;=\; -\frac{\xi''(s)}{2}\Big[\partial_{xx}\Phi + F(s)\,(\partial_x\Phi)^2\Big],$$

with boundary condition $\Phi(1,x) = \log\cosh(x+\beta h)$ (where we have reinstated the external field absorbed in the boundary). This is the *Parisi PDE.*

The variational representation. Auffinger and Chen [AC15] gave a stochastic-control representation of the Parisi PDE. Define a controlled process X_s^α on $[0,1]$ with stochastic dynamics $dX_s^\alpha = \alpha_s\, dW_s$, $X_0^\alpha = \beta h$, where W is Brownian motion and α is a bounded progressively-measurable control. Then

$$\Phi_\mu(0,0) \;=\; \sup_\alpha \mathbb{E}\Big[\log\cosh(X_1^\alpha) - \frac{1}{2}\int_0^1 \xi''(s)\, F(s)\, \alpha_s^2\, ds\Big].$$

This formulation is useful for several reasons.

- It makes the dependence on F (and hence on μ) explicit as a cost-term weight.
- It gives the derivative bound $|\partial_x \Phi_\mu(s,x)| \le 1$ used in the Lipschitz estimates of Theorem 3.11.
- It sets up the uniqueness-of-minimizer argument of Auffinger and Chen.

A.3. **The Parisi functional.** The Parisi PDE evaluated at $(s,x) = (0,0)$ gives one piece of the free-energy formula. The full functional is

$$\mathcal{P}_\beta(\mu) \;=\; \log 2 + \Phi_\mu(0,0) - \frac{1}{2}\int_0^1 s\,\xi''(s)\, F(s)\, ds,$$

where the integral term comes from the quadratic overlap penalty $\frac{\beta^2}{4}\sum_{\alpha,\beta} Q_{\alpha\beta}^2 = \frac{\beta^2}{4}\sum_j (m_{j+1} - m_j)\, q_j^2 \to \frac{\beta^2}{4}\int_0^1 q(x)^2\, dx$ in the continuous limit (equivalently, $\int_0^1 F(s)\, s\, ds/\text{const}$. See Proposition 4.1 for the compact form).

Uniqueness and regularity. Auffinger and Chen [AC15] proved.

- $\mathcal{P}_\beta$ is strictly convex in μ (in a suitable reparametrized sense).
- The minimizer $\mu^* = \arg\min_\mu \mathcal{P}_\beta(\mu)$ is unique.
- μ^* has bounded support in $[0,1]$ (indeed contained in $[0, q_{\mathrm{EA}}]$ where $q_{\mathrm{EA}} < 1$ when $h > 0$).
- The quantile function of μ^* is analytic on the support interval.

These facts are used implicitly throughout the thesis. The uniqueness of μ^* is essential for Theorem 3.11 (which compares $\mathcal{P}_\beta(\nu)$ to $\mathcal{P}_\beta(\mu^*)$) and for Corollary 3.14 (weak convergence of RSB minimizers).

A.4. **Remarks on conventions.** Different sources use different conventions for the factor of $\frac{1}{2}$ in the covariance and the PDE. Our choice $\xi(q) = \frac{\beta^2}{2} q^2$ gives $\xi''(q) = \beta^2$ and matches the Auffinger–Chen convention. Some physics sources use $\xi(q) = \beta^2 q^2$ (no factor of 1/2), with correspondingly doubled Lipschitz constants. Translation between conventions is straightforward but has been a source of minor discrepancies in the literature. We have chosen the one most natural for the PDE derivation above.

Appendix B. Free-Energy Variance and the Poincaré Bound

This appendix records the Gaussian Poincaré bound on the variance of the SK free energy per site, expressed in terms of the second overlap moment. The bound is standard in the spin glass literature. See, e.g., [Tal11, §1.3] and [Pan13b, Proposition 1.3.4]. We include the proof for completeness and to make the m_2-dependence explicit.

Proposition B.1 (Poincaré bound). *In the SK model at inverse temperature β and external field h,*

$$\mathrm{Var}\Big(\frac{1}{N}\log Z_N(\beta,h)\Big) \;\le\; \frac{\beta^2\, m_2^{(N)}}{2N} \;-\; \frac{\beta^2}{2N^2},$$

where $m_2^{(N)} = \mathbb{E}[\langle R_{12}^2\rangle_{\beta,h,N}]$ is the finite-N second overlap moment. In particular, $\mathrm{Var}(\frac{1}{N}\log Z_N) \le \frac{\beta^2 m_2}{2N} + O(N^{-2})$ *as $N\to\infty$, where $m_2 = \lim_N m_2^{(N)} = \int q^2\, d\mu^*(q)$ is the overlap second moment of the Parisi measure.*

Proof. Write $F_N = \frac{1}{N}\log Z_N(\beta,h)$. The Gaussian Poincaré inequality (see [BLM13, Theorem 3.20]) states that for any smooth function G of i.i.d. standard Gaussians $(g_1,\dots,g_M)$,

$$\mathrm{Var}(G) \;\le\; \sum_{k=1}^{M}\mathbb{E}\Big[\Big(\frac{\partial G}{\partial g_k}\Big)^2\Big].$$

Apply this with $G = \log Z_N$ and the Gaussians indexed by pairs (i,j) with $1\le i<j\le N$.

$$\mathrm{Var}(\log Z_N) \;\le\; \sum_{i<j}\mathbb{E}\Big[\Big(\frac{\partial \log Z_N}{\partial J_{ij}}\Big)^2\Big].$$

From the Hamiltonian (1),

$$\frac{\partial \log Z_N}{\partial J_{ij}} \;=\; \frac{\beta}{\sqrt{N}}\,\langle \sigma_i\sigma_j\rangle,$$

where $\langle\cdot\rangle$ is the Gibbs average. So

$$\mathrm{Var}(\log Z_N) \;\le\; \frac{\beta^2}{N}\sum_{i<j}\mathbb{E}[\langle\sigma_i\sigma_j\rangle^2].$$

Introducing two independent replicas σ^1,σ^2 drawn from the Gibbs measure, the *replica identity* $\langle X\rangle^2 = \langle X^1X^2\rangle$ (where X^1, X^2 are copies of X evaluated on the two replicas. Valid by independence of replicas under a common disorder) gives

$$\langle\sigma_i\sigma_j\rangle^2 \;=\; \langle\sigma_i^1\sigma_j^1\,\sigma_i^2\sigma_j^2\rangle$$

and hence

$$\sum_{i,j}\mathbb{E}[\langle\sigma_i\sigma_j\rangle^2] \;=\; \mathbb{E}\Big[\Big\langle\Big(\sum_i\sigma_i^1\sigma_i^2\Big)\Big(\sum_j\sigma_j^1\sigma_j^2\Big)\Big\rangle\Big] \;=\; N^2\,\mathbb{E}[\langle R_{12}^2\rangle],$$

using $\sum_i \sigma_i^1\sigma_i^2 = NR_{12}$. Separating out the diagonal $i=j$ terms (which contribute N since $\sigma_i^2=1$).

$$\sum_{i<j}\mathbb{E}[\langle\sigma_i\sigma_j\rangle^2] \;=\; \frac{1}{2}\big(N^2\,\mathbb{E}[\langle R_{12}^2\rangle] - N\big).$$

Substituting.

$$\mathrm{Var}(\log Z_N) \;\le\; \frac{\beta^2}{N}\cdot\frac{N^2\mathbb{E}[\langle R_{12}^2\rangle]-N}{2} \;=\; \frac{\beta^2 N}{2}\,\mathbb{E}[\langle R_{12}^2\rangle]-\frac{\beta^2}{2}.$$

Dividing by N^2.

$$\mathrm{Var}(F_N) \;=\; \mathrm{Var}(\tfrac{1}{N}\log Z_N) \;\le\; \frac{\beta^2}{2N}\,\mathbb{E}[\langle R_{12}^2\rangle]-\frac{\beta^2}{2N^2}. \qquad \square$$

Remark B.2 (Self-averaging rate). Proposition B.1 gives a concrete quantitative form of the self-averaging property of the SK free energy, F_N concentrates around its mean at rate $O(N^{-1/2})$, with the proportionality constant controlled by m_2. At high temperature, m_2 is small (indeed $m_2 = 0$ for $\beta < 1$ and $h = 0$), so the variance is especially small. At low temperature, m_2 approaches $q_{\mathrm{EA}}^2 \approx 1$ and the variance saturates at $O(\beta^2/N)$.

Remark B.3 (Higher cumulants). Proposition B.1 bounds the variance only. Higher cumulants of F_N are controlled by Talagrand's concentration inequality or by interpolation arguments. At high temperature the fluctuations of $F_N - \mathbb{E}[F_N]$ are asymptotically Gaussian with scale $O(N^{-1/2})$. At low temperature the fluctuation theory is less fully understood. The moment-class framework of the thesis controls the mean $\mathbb{E}[F_N]$ to $O(\beta^2/K)$. Controlling higher cumulants from moments is the subject of the conjecture in Section 5.5.

Remark B.4 (Why only m_2). It may seem surprising that the variance depends only on the second overlap moment, not on higher moments. The reason is algebraic. The Poincaré inequality is a sum of squared derivatives, each of which has the form $\langle\sigma_i\sigma_j\rangle^2$, which is a bilinear replica observable and is (by symmetrization) an average of R_{12}^2. Higher moments enter in cumulants of order ≥ 3, where more replicas and more general invariants of the overlap array appear.

Appendix C. Probability Measures, Quantile Functions, and Wasserstein

This appendix collects measure-theoretic and metric background used throughout.

C.1. **Weak topology on $\mathcal{P}([0,1])$.**

Let $\mathcal{P}([0,1])$ denote the set of Borel probability measures on the compact interval $[0,1]$. A sequence $\mu_n \in \mathcal{P}([0,1])$ *converges weakly* to μ, written $\mu_n \rightharpoonup \mu$, if $\int f\,d\mu_n \to \int f\,d\mu$ for every bounded continuous f. Equivalently, by the Portmanteau theorem, $F_{\mu_n}(s) \to F_\mu(s)$ at every continuity point of F_μ, where $F_\mu(s) = \mu([0,s])$. Since $[0,1]$ is compact, Prokhorov's theorem gives.

Theorem C.1 (Prokhorov). *$\mathcal{P}([0,1])$ is compact in the weak topology.*

C.2. **Quantile functions and the pushforward.**

Every non-decreasing right-continuous function $q : [0,1] \to [0,1]$ with $q(0) \ge 0$ and $q(1) \le 1$ is the quantile function of a unique $\mu \in \mathcal{P}([0,1])$, defined by

$$\mu(A) \;=\; \mathrm{Leb}(\{x \in [0,1] : q(x) \in A\}) \qquad \text{for Borel } A \subset [0,1].$$

Equivalently, q is the generalized inverse of the CDF, $q(x) = \inf\{s : F_\mu(s) \geq x\}$. For a discrete measure $\mu = \sum_{i=0}^{K} w_i \, \delta_{q_i}$ with $0 \leq q_0 < q_1 < \cdots < q_K \leq 1$ and weights $w_i > 0$, the quantile function is the step function

$$q(x) \;=\; q_j \quad \text{for } x \in [\textstyle\sum_{i<j} w_i, \sum_{i\leq j} w_i).$$

The pushforward identity $\mu = q_{\#}\mathrm{Leb}$ translates integrals against μ into ordinary Lebesgue integrals.

$$\int_0^1 f(s)\, d\mu(s) \;=\; \int_0^1 f(q(x))\, dx.$$

C.3. **The Wasserstein distance.** The *Wasserstein-1 distance* $W_1(\mu, \nu)$ on $\mathcal{P}([0,1])$ is defined equivalently by any of.

(i) **Optimal transport.** $W_1(\mu,\nu) = \inf_\pi \int |s - t|\, d\pi(s,t)$ over couplings π with marginals μ, ν.
(ii) **CDF integral.** $W_1(\mu,\nu) = \int_0^1 |F_\mu(s) - F_\nu(s)|\, ds$.
(iii) **Quantile integral.** $W_1(\mu,\nu) = \int_0^1 |q_\mu(x) - q_\nu(x)|\, dx$.
(iv) **Kantorovich–Rubinstein duality.** $W_1(\mu,\nu) = \sup\{|\int f\, d\mu - \int f\, d\nu| : f \text{ 1-Lipschitz on } [0,1]\}$.

On the compact space $[0,1]$, W_1 metrizes the weak topology, $\mu_n \rightharpoonup \mu \iff W_1(\mu_n, \mu) \to 0$.

The duality (iv) is the form we use in the proof of Lemma 3.9 (Wasserstein diameter of $\mathcal{M}_K$), combined with Jackson–Stechkin polynomial approximation.

C.4. **Hausdorff moment determinacy.** A probability measure on $[0,1]$ is uniquely determined by its moment sequence.

Theorem C.2 (Hausdorff)**.** *If $\mu, \nu \in \mathcal{P}([0,1])$ have $\int q^k\, d\mu = \int q^k\, d\nu$ for all $k \geq 0$, then $\mu = \nu$.*

This follows from the Weierstrass approximation theorem. Polynomials are dense in $C([0,1])$ with the sup norm, so moment agreement implies $\int f\, d\mu = \int f\, d\nu$ for every continuous f, hence $\mu = \nu$.

Consequently, $\bigcap_{K\geq 1} \mathcal{M}_K = \{\mu^*\}$ always. The rate at which the filtration approaches its limit is quantified by Lemma 3.9, $\mathcal{M}_K$ has W_1-diameter at most $2/(K+1)$. The classical Carleman condition for moment determinacy on unbounded intervals is not needed here because $[0,1]$ is compact.

Department of Statistics and Data Science, Yale University