The Parisi Solution of the Sherrington–Kirkpatrick Model

Replica Symmetry Breaking, Ultrametricity,
and the Guerra–Talagrand Proof

Nahom Seyoum

Yale University

nahom.seyoum@yale.edu

February 2026

Abstract

The Sherrington–Kirkpatrick model is a mean-field spin glass whose free energy resisted rigorous computation for three decades. We develop the full Parisi theory of replica symmetry breaking for this model. Starting from the replica method, we derive the replica-symmetric solution and carry out its stability analysis using the representation theory of the symmetric group acting on the overlap matrix space, decomposing the Hessian into its longitudinal, anomalous, and replicon sectors. The replicon eigenvalue is negative below $T_c = 1$, and we show that iterating this instability forces a hierarchical block structure on the overlap matrix. We construct the K-step Parisi ansatz, compute its trace invariants, and identify the quantile function $q(x)$ as the order parameter. Passing to the continuum limit $K \to \infty$ yields the Parisi PDE, whose well-posedness we establish via the heat semigroup, and the Parisi variational formula $f(\beta, h) = \inf_\mu \mathcal{P}(\beta, h, \mu)$. The probabilistic content of the solution is developed through Ruelle probability cascades and the Ghirlanda–Guerra identities, which together force an ultrametric organization of the Gibbs measure. We then present the rigorous verification: Guerra's interpolation upper bound, the matching lower bound of Talagrand via the Aizenman–Sims–Starr cavity method, and Panchenko's proof that the Ghirlanda–Guerra identities imply ultrametricity, showing that the hierarchical ansatz is the unique structure consistent with the thermodynamic self-consistency of the model. Near T_c, we carry out the Landau expansion of the free energy functional and solve the stationarity equations in closed form.

Contents

I Foundations **3**

1 The Sherrington–Kirkpatrick Model **4**
1.1 Definition of the Model 4
1.2 Gibbs Measure and Overlaps 6
1.3 Thermodynamic Quantities 7
1.4 The High-Temperature Phase 9

2 The Replica Method **11**
2.1 The Replica Trick 11
2.2 The Saddle Point and the Overlap Matrix 13
2.3 The Replica Symmetric Solution and Its Failure 16

3 Why Replica Symmetry Must Break **23**
3.1 The Meaning of Replica Symmetry Breaking 23
3.2 Constraints on the Symmetry-Breaking Pattern 24
3.3 The Ultrametric Hypothesis 25

II Parisi's Construction **27**

4 The Hierarchical Ansatz **28**
4.1 The Parisi Matrix for Finite K 28
4.2 The Function $q(x)$ on $[0, 1]$ 31
4.3 Probabilistic Interpretation 34

5 Computing the Free Energy **36**
5.1 The Free Energy for K-Step RSB 36
5.2 The Limit $K \to \infty$: The Parisi PDE 39
5.3 Well-Posedness of the Parisi PDE 42

6 Analysis Near the Critical Temperature **45**
6.1 The Landau Expansion 45
6.2 Optimization and the Convergent Sequence 47
6.3 The Free Energy Expansion and Convergence Rates 50

7 Numerical Results and the Full Solution **53**
7.1 The $K = 1$ Solution at All Temperatures 53
7.2 The $K = 2$ Solution and Convergence 54
7.3 The Full RSB Solution 55

III Rigorous Results **58**

8 Rigorous Results **59**
8.1 Guerra's Bound . . . 59
8.2 Talagrand's Proof . . . 60
8.3 Ultrametricity . . . 61

Appendices **64**

A Gaussian Integration Identities **64**
A.1 The Hubbard–Stratonovich Transformation . . . 64
A.2 Gaussian Integration by Parts (Stein's Lemma) . . . 64
A.3 Moments of Log-Partition Functions . . . 65
A.4 Gaussian Tail Bounds . . . 65

B The Heat Equation and Semigroup **66**
B.1 The Heat Equation . . . 66
B.2 The Semigroup C_t . . . 66
B.3 The Cole–Hopf Transformation . . . 67

C Probability Measures and Quantile Functions **68**
C.1 Quantile Functions and Distribution Functions . . . 68
C.2 The Integral Formula . . . 68
C.3 Topology and Compactness . . . 69
C.4 The Wasserstein Distance . . . 69

D Notation and Conventions **70**

Part I

Foundations

Chapter 1

The Sherrington–Kirkpatrick Model

1.1 Definition of the Model

The Sherrington–Kirkpatrick (SK) model is a mean-field model of a spin glass: a magnetic system in which the interactions between spins are random, so that the system is "frustrated" — no single configuration can simultaneously satisfy all interactions. The model was introduced by Sherrington and Kirkpatrick [SK75] as an exactly solvable testing ground for the Edwards–Anderson theory of spin glasses [EA75]. We now define it with full precision.

Definition 1.1 (Configuration space)**.** The *configuration space* is $\Sigma_N = \{-1, +1\}^N$, where $N \geq 1$ is the number of spins. A *configuration* is a vector $\sigma = (\sigma_1, \ldots, \sigma_N) \in \Sigma_N$. The cardinality of Σ_N is $|\Sigma_N| = 2^N$.

Definition 1.2 (Disorder)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. The *disorder* is a family of random variables

$$(J_{ij})_{1 \leq i < j \leq N}, \qquad J_{ij} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1). \tag{1.1}$$

We write $\mathbb{E}$ for expectation with respect to $\mathbb{P}$ (i.e., with respect to the disorder). The couplings J_{ij} represent random interactions: when $J_{ij} > 0$, sites i and j prefer to align; when $J_{ij} < 0$, they prefer to anti-align. Since each J_{ij} is symmetric about zero, there is no net preference, and frustration is unavoidable.

Definition 1.3 (The SK Hamiltonian)**.** For a fixed realization of the disorder and a fixed external magnetic field $h \in \mathbb{R}$, the *Hamiltonian* of the SK model is the function $H_N : \Sigma_N \to \mathbb{R}$ defined by

$$H_N(\sigma) = -\frac{1}{\sqrt{N}} \sum_{1 \leq i < j \leq N} J_{ij}\, \sigma_i \sigma_j \; - \; h \sum_{i=1}^{N} \sigma_i. \tag{1.2}$$

The first term encodes the random spin–spin interactions; the second term couples each spin to the external field.

Remark 1.4 (The scaling $1/\sqrt{N}$). The factor $1/\sqrt{N}$ in (1.2) is essential. Each spin σ_i interacts with all $N - 1$ other spins (the model is "mean-field" or "infinite-range"), so the sum over pairs has $\binom{N}{2}$ terms. Since each $J_{ij}\sigma_i\sigma_j$ has variance 1, the sum $\sum_{i<j} J_{ij}\sigma_i\sigma_j$ has variance of order N^2, hence standard deviation of order N. Dividing by $\sqrt{N}$ makes $H_N(\sigma)$ of order N for typical realizations of the disorder, which is the correct extensive scaling for a thermodynamic energy.

For comparison, the Curie–Weiss ferromagnet uses the same all-to-all geometry but with deterministic couplings $J_{ij} = 1/N$; the different scaling ($1/N$ vs. $1/\sqrt{N}$) reflects the fact that random signs produce cancellations that reduce the effective interaction strength.

Definition 1.5 (Partition function and free energy)**.** At inverse temperature $\beta = 1/T > 0$, the *partition function* is

$$Z_N(\beta, h) = \sum_{\sigma \in \Sigma_N} \exp\Big(-\beta\, H_N(\sigma)\Big). \tag{1.3}$$

Note that Z_N is a positive random variable (it depends on the disorder (J_{ij})).

The *quenched free energy density* is

$$f_N(\beta, h) = \frac{1}{N}\, \mathbb{E} \ln Z_N(\beta, h). \tag{1.4}$$

The adjective "quenched" refers to the fact that we take the logarithm *before* averaging over the disorder. This is the physically relevant quantity: it corresponds to a system where the disorder is frozen (quenched) and the spins equilibrate in a fixed random environment, after which we average over environments.

Remark 1.6 (Quenched vs. annealed)*.* The *annealed free energy density* is the simpler quantity

$$f_N^{\mathrm{ann}}(\beta, h) = \frac{1}{N}\, \ln \mathbb{E}\Big[Z_N(\beta, h)\Big]. \tag{1.5}$$

By Jensen's inequality ($\ln$ is concave and $\mathbb{E}$ is linear), we always have

$$f_N(\beta, h) \le f_N^{\mathrm{ann}}(\beta, h).$$

The annealed free energy is much easier to compute — the disorder and the spins are treated on equal footing — but it does not capture the physics of a frozen random environment. In the SK model the annealed and quenched free energies differ below the critical temperature $T_c = 1$, which is the regime where spin glass behavior emerges.

We can now state the fundamental question.

The central problem. *Does the limit*

$$f(\beta, h) = \lim_{N \to \infty} f_N(\beta, h) \tag{1.6}$$

exist, and if so, what is its value?

The existence of this limit is itself a nontrivial result.

Theorem 1.7 (Guerra–Toninelli, 2002)**.** *For every $\beta > 0$ and $h \in \mathbb{R}$, the limit $f(\beta, h) = \lim_{N\to\infty} f_N(\beta, h)$ exists.*

The proof, due to Guerra and Toninelli [GT02], uses a superadditivity argument: by a clever interpolation, one shows that $N \mapsto N\, f_N(\beta, h)$ is approximately superadditive, from which the existence of the limit follows by Fekete's lemma. We will not reproduce the proof here (see [Pan13b] for a clean exposition), but we note that the argument gives existence without providing a formula for $f(\beta, h)$. Finding such a formula is the subject of the rest of this monograph.

Remark 1.8 (Self-averaging)*.* A much stronger result holds: the free energy density is *self-averaging*, meaning that $\frac{1}{N} \ln Z_N(\beta, h) \to f(\beta, h)$ in probability (and in L^1) as $N \to \infty$. That is, for large N, the free energy density of a *single* realization of the disorder is close to its expected value with high probability. This is a consequence of the Gaussian concentration inequality and the fact that $\ln Z_N$ is a Lipschitz function of the (J_{ij}).

1.2 Gibbs Measure and Overlaps

The partition function Z_N from Definition 1.5 induces a probability measure on configurations. This measure, and the notion of *overlap* between configurations, are the central objects from which the Parisi theory emerges.

Definition 1.9 (Gibbs measure)**.** The *Gibbs measure* (or *Boltzmann distribution*) of the SK model at inverse temperature β and external field h, for a fixed realization of the disorder, is the probability measure G_N on Σ_N defined by

$$G_N(\sigma) = \frac{1}{Z_N(\beta, h)} \exp\Big(-\beta\, H_N(\sigma)\Big), \qquad \sigma \in \Sigma_N. \tag{1.7}$$

We write $\langle \cdot \rangle$ for the expectation with respect to G_N.

The Gibbs measure assigns higher probability to configurations of lower energy. At $\beta = 0$ (infinite temperature), G_N is the uniform measure on Σ_N; as $\beta \to \infty$ (zero temperature), G_N concentrates on the ground state(s) — the configurations that minimize H_N.

A crucial subtlety: G_N is itself a *random* probability measure, because it depends on the disorder (J_{ij}). We therefore have two levels of randomness: the "thermal" randomness captured by $\langle \cdot \rangle$, and the "disorder" randomness captured by $\mathbb{E}$. When we write $\mathbb{E}\langle \cdot \rangle$, we mean the average over both.

Definition 1.10 (Overlap)**.** Let $\sigma^1, \sigma^2 \in \Sigma_N$ be two configurations. Their *overlap* is

$$R_{1,2} = R(\sigma^1, \sigma^2) = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^1\, \sigma_i^2. \tag{1.8}$$

Since each $\sigma_i^a \in \{-1, +1\}$, we have $R_{1,2} \in [-1, 1]$. If $\sigma^1 = \sigma^2$ then $R_{1,2} = 1$; if $\sigma^1 = -\sigma^2$ then $R_{1,2} = -1$. More generally, $R_{1,2}$ measures the fraction of sites at which the two configurations agree, rescaled to lie in $[-1, 1]$:

$$R_{1,2} = \frac{2}{N} \left|\{i : \sigma_i^1 = \sigma_i^2\}\right| - 1.$$

The overlap is the central observable in spin glass theory. When two configurations σ^1, σ^2 are drawn independently from G_N, the overlap $R_{1,2}$ becomes a random variable (random through both the thermal and disorder averages). Its distribution encodes the structure of the Gibbs measure.

Definition 1.11 (Overlap distribution)**.** Let σ^1, σ^2 be independent samples from G_N. The *overlap distribution* is the (random) probability measure μ_N on $[-1, 1]$ defined by

$$\mu_N(A) = G_N^{\otimes 2}\Big(\{(\sigma^1, \sigma^2) : R_{1,2} \in A\}\Big) = \sum_{\sigma^1, \sigma^2 \in \Sigma_N} G_N(\sigma^1)\, G_N(\sigma^2)\, \mathbf{1}_{R(\sigma^1, \sigma^2) \in A} \tag{1.9}$$

for Borel sets $A \subset [-1, 1]$. Its averaged version $\mathbb{E}[\mu_N]$ is a deterministic probability measure on $[-1, 1]$.

The overlap distribution μ_N is a random measure because G_N depends on the disorder. Its structure reveals the nature of the spin glass phase. In a ferromagnet, or at high temperature, μ_N concentrates near a single point as $N \to \infty$: the Gibbs measure is

dominated by one "state," and any two typical configurations have essentially the same overlap. In a spin glass at low temperature, by contrast, μ_N is supported on multiple values. Different pairs of configurations can have different overlaps, indicating that the Gibbs measure is spread across several distinct "states" or "valleys" in configuration space.

Definition 1.12 (Edwards–Anderson order parameter)**.** The *Edwards–Anderson order parameter* is

$$q_{\mathrm{EA}} = \lim_{N\to\infty} \mathbb{E}\langle R_{1,2}^2\rangle, \tag{1.10}$$

when this limit exists. Equivalently, q_{EA} is the second moment of the limiting averaged overlap distribution. In the high-temperature phase $q_{\mathrm{EA}} = 0$; in the spin glass phase $q_{\mathrm{EA}} > 0$.

The Edwards–Anderson parameter captures whether the system has frozen into a state with a preferred local magnetization pattern. It can also be written as

$$q_{\mathrm{EA}} = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\Big[\langle\sigma_i\rangle^2\Big],$$

which makes the interpretation clear: $q_{\mathrm{EA}} > 0$ means that individual spins have nonzero thermal expectations (the system "remembers" a configuration), even though the average magnetization $\frac{1}{N}\sum_i \mathbb{E}\langle\sigma_i\rangle$ vanishes by the symmetry of the disorder.

Remark 1.13 (Foreshadowing: $q(x)$ as a quantile function)*.* In Chapters 4–5, we will see that the correct order parameter for the SK model is not a single number q_{EA} but a non-decreasing function $q : [0,1] \to [0, q_{\mathrm{EA}}]$. The deep content of Parisi's solution is that this function turns out to be the *quantile function* (generalized inverse CDF) of the overlap distribution:

$$q(x) = \inf\{q \in [-1,1] : \mathbb{E}[\mu_\infty]((-\infty, q]) \geq x\}.$$

Replica symmetry corresponds to $q(x)$ being a constant function; replica symmetry breaking means $q(x)$ is genuinely increasing on some subinterval of $[0,1]$. This identification was not clear in Parisi's original papers and emerged later through the work of Mézard, Parisi, and Virasoro [MPV87].

1.3 Thermodynamic Quantities

From the free energy density $f(\beta, h)$ defined in (1.6), all thermodynamic observables can be extracted by differentiation. We record the key ones here, both for later use and to establish a set of *consistency checks*: any proposed formula for $f(\beta, h)$ must produce physically sensible values for these quantities.

Remark 1.14 (Convention for f)*.* Our definition $f = \frac{1}{N}\mathbb{E}\ln Z$ is β times the standard Helmholtz free energy density $F_{\mathrm{std}} = -\frac{1}{\beta N}\mathbb{E}\ln Z$, i.e., $f = -\beta F_{\mathrm{std}}$. The advantage of this convention (common in the mathematical literature and used by Parisi in his papers) is that f has no explicit factor of β in its definition. The thermodynamic derivatives below are adapted accordingly.

Definition 1.15 (Internal energy)**.** The *internal energy density* is the average energy per spin:

$$U(\beta, h) = -\frac{\partial f}{\partial \beta}(\beta, h) = \lim_{N\to\infty} \frac{1}{N}\,\mathbb{E}\langle H_N(\sigma)\rangle. \tag{1.11}$$

(This follows from differentiating $f = \frac{1}{N}\mathbb{E}\ln Z$ with respect to β, since $\frac{\partial}{\partial\beta}\ln Z = -\langle H_N\rangle$.)

Definition 1.16 (Entropy)**.** The *entropy density* is

$$S(\beta, h) = f(\beta, h) + \beta \frac{\partial f}{\partial \beta}(\beta, h) = f(\beta, h) - \beta\, U(\beta, h). \tag{1.12}$$

This is the Legendre transform relation $S = f - \beta U = f + \beta \frac{\partial f}{\partial \beta}$, or equivalently $f = \beta U + S$ (the analogue of $-\beta F_{\mathrm{std}} = \beta U + S$, i.e., $S = -\beta F_{\mathrm{std}} - \beta U$, matching the standard relation $F_{\mathrm{std}} = U_{\mathrm{std}} - T\, S$).

Definition 1.17 (Specific heat and magnetic susceptibility)**.** The *specific heat* is

$$C(\beta, h) = -\beta^2 \frac{\partial U}{\partial \beta}(\beta, h) = \beta^2 \frac{\partial^2 f}{\partial \beta^2}(\beta, h). \tag{1.13}$$

The *magnetic susceptibility* is

$$\chi(\beta, h) = \frac{\partial^2 f}{\partial h^2}(\beta, h) = \lim_{N \to \infty} \frac{\beta}{N} \sum_{i=1}^{N} \mathbb{E}\Big[\langle \sigma_i^2 \rangle - \langle \sigma_i \rangle^2\Big]. \tag{1.14}$$

The physical constraint that will play a decisive role in the story is the following.

Proposition 1.18 (Non-negativity of entropy)**.** *Since the SK model has a finite configuration space Σ_N with $|\Sigma_N| = 2^N$, the entropy density satisfies*

$$0 \le S(\beta, h) \le \ln 2 \tag{1.15}$$

for all $\beta > 0$ and $h \in \mathbb{R}$.

Proof. The entropy of a probability measure ν on a finite set X is $S(\nu) = -\sum_{x \in X} \nu(x) \ln \nu(x) \in [0, \ln |X|]$. For the Gibbs measure on Σ_N, this gives $S(G_N) \in [0, N \ln 2]$. The entropy density $S(\beta, h)$ is the limit of $\frac{1}{N}\mathbb{E}[S(G_N)]$, which inherits the bounds. □

This seemingly innocuous constraint will be the first sign that the naive approach to the SK model fails: the replica symmetric solution (Chapter 2) predicts $S(0, 0) \approx -0.17$, a flagrant violation of $S \ge 0$. This unphysical entropy was the original motivation for seeking a more refined solution, and Parisi's replica symmetry breaking scheme (Chapter 4) was designed to fix precisely this problem.

Remark 1.19 (Testing proposed solutions). Any candidate formula for $f(\beta, h)$ can be tested by computing U, S, C, and χ and checking:

(i) $S(\beta, h) \ge 0$ for all β, h.

(ii) Numerical values agree with Monte Carlo simulations (for the SK model, extensive simulations by Sherrington and Kirkpatrick [KS78] give $U(T = 0, h = 0) = -0.76 \pm 0.01$).

(iii) $C(\beta, h) \ge 0$ (thermodynamic stability).

The non-negativity of entropy is the sharpest test and the one that most cleanly discriminates between the replica symmetric and Parisi solutions.

1.4 The High-Temperature Phase

Before entering the technically demanding low-temperature regime, we record the behavior of the SK model at high temperature, where a complete and rigorous solution is available. This serves two purposes: it provides a baseline against which the low-temperature results can be compared, and it foreshadows the connection to the replica method.

Theorem 1.20 (High-temperature free energy)**.** *For $\beta < 1$ and any $h \in \mathbb{R}$, the limiting free energy density is*

$$f(\beta, h) = \frac{\beta^2}{4} + \ln 2\cosh(\beta h). \tag{1.16}$$

At zero external field ($h = 0$), this simplifies to $f(\beta, 0) = \frac{\beta^2}{4} + \ln 2$.

Remark 1.21. The form of (1.16) has a transparent interpretation. The $\ln 2\cosh(\beta h)$ term is the free energy of a single non-interacting spin in a field h; the $\beta^2/4$ term is a correction from the random interactions, computed perturbatively. At high temperature, the interactions are too weak relative to the thermal fluctuations to produce any collective ordering.

The proof of Theorem 1.20 can be given by several methods (see [Tal11] or [Pan13b]); the cleanest uses the *interpolation method* of Guerra. We sketch the argument to give a flavor of the techniques involved.

Proof sketch. The key input is that the overlap concentrates at zero in the high-temperature phase. Specifically, one can show that for $\beta < 1$,

$$\mathbb{E}\langle R_{1,2}^2 \rangle \le \frac{C}{N} \tag{1.17}$$

for a constant $C = C(\beta)$ depending only on β. This is established by a second-moment computation: one computes $\mathbb{E}[Z_N^2]$ (the "second moment method") and shows that the dominant contribution comes from configurations with $R_{1,2} \approx 0$.

Once overlap concentration is established, the free energy can be computed by an interpolation argument. Define a family of models parametrized by $t \in [0, 1]$:

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

where (y_i) are i.i.d. $\mathcal{N}(0, \beta^2/2)$, independent of (J_{ij}). At $t = 1$ we recover the SK model; at $t = 0$ the spins are decoupled (each spin sees only an independent random field y_i, and the free energy is that of N independent spins).

Computing $\frac{d}{dt}\frac{1}{N}\mathbb{E}\ln Z_N^{(t)}$ and using the overlap concentration (1.17) to control the error, one obtains (1.16) in the limit $N \to \infty$. □

The associated thermodynamic quantities in the high-temperature phase are:

$$U(\beta, 0) = -\frac{\beta}{2}, \tag{1.18}$$

$$S(\beta, 0) = \ln 2 - \frac{\beta^2}{4}, \tag{1.19}$$

$$\chi(\beta, 0) = \beta. \tag{1.20}$$

The entropy $S(\beta, 0) = \ln 2 - \beta^2/4$ remains positive for $\beta < 2\sqrt{\ln 2} \approx 1.665$, well beyond the critical temperature $\beta_c = 1$. However, the formula itself is only valid for $\beta < 1$; continuing it blindly beyond β_c would eventually produce a negative entropy, signaling the need for a different solution in the low-temperature phase.

Remark 1.22 (Connection to the replica method). In Chapter 2, we will see that the replica method with the *replica symmetric* ansatz $Q_{\alpha\beta} = q$ produces a formula for $f(\beta, h)$ whose maximum over q agrees with (1.16) when $\beta < 1$ (with the maximizer $q = 0$). For $\beta > 1$, the maximizer jumps to $q > 0$, and the resulting formula is the replica symmetric solution — which, as we will see, gives incorrect (and unphysical) results. The fact that the replica method works perfectly at high temperature but fails at low temperature is the motivation for Parisi's refinement.

Remark 1.23 (The phase transition). The critical temperature $T_c = 1$ (equivalently $\beta_c = 1$) marks a genuine phase transition in the SK model. For $T > T_c$, the overlap distribution μ_N concentrates at $q = 0$ (the paramagnetic phase); for $T < T_c$ and $h = 0$, the overlap distribution becomes nontrivial (the spin glass phase). The nature of this transition — and the precise structure of μ_N in the low-temperature phase — is the content of the Parisi solution.

Chapter 2

The Replica Method

2.1 The Replica Trick

We now introduce the main computational device of this monograph: the *replica method*. This is a formal (non-rigorous) technique for computing $\mathbb{E}[\ln Z]$ by reducing it to a computation of $\mathbb{E}[Z^n]$ for integer n, followed by an analytic continuation to $n = 0$. We will be explicit about where the method is rigorous and where it is formal.

The replica identity

The starting point is the elementary identity: for any $Z > 0$,

$$\ln Z = \lim_{n \to 0} \frac{Z^n - 1}{n}. \tag{2.1}$$

This is simply the definition of the derivative $\frac{d}{dn} Z^n \Big|_{n=0} = \ln Z$, since $Z^n = e^{n \ln Z}$. In particular, for $n \in \mathbb{R}$ and $Z > 0$, the function $n \mapsto Z^n$ is infinitely differentiable, and (2.1) holds exactly.

Applying this to the free energy:

$$f_N(\beta, h) = \frac{1}{N} \mathbb{E}[\ln Z_N] = \frac{1}{N} \lim_{n \to 0} \frac{\mathbb{E}[Z_N^n] - 1}{n}. \tag{2.2}$$

The *replica strategy* is:

1. Compute $\mathbb{E}[Z_N^n]$ for $n \in \mathbb{N}$ (this can be done because Z_N^n is the partition function of n non-interacting copies—"replicas"—of the system).
2. Express the result as a function of n that makes sense for real n.
3. Set $n = 0$ in the resulting expression to obtain f_N.

Warning 2.1 (Non-rigorous step). The passage from step 1 to step 3 is *not* justified mathematically. The function $n \mapsto \mathbb{E}[Z_N^n]$ for integer n does not determine its values at $n = 0$ (analytic continuation from $\mathbb{N}$ to $\mathbb{R}$ is not unique without growth conditions). Moreover, we will need to take $N \to \infty$ and $n \to 0$ simultaneously, and the order of these limits matters. The replica method is a *formal* computational technique; its conclusions are conjectural until verified by other means. The remarkable fact is that the Parisi formula obtained by this method has been proved correct by entirely different, rigorous arguments (Chapter 8).

The replicated partition function

For $n \in \mathbb{N}$, the n-th moment of the partition function is

$$\mathbb{E}[Z_N^n] = \mathbb{E}\Big[\prod_{\alpha=1}^{n} Z_N\Big] = \mathbb{E}\left[\prod_{\alpha=1}^{n} \sum_{\sigma^\alpha \in \Sigma_N} \exp\Big(-\beta H_N(\sigma^\alpha)\Big)\right]. \tag{2.3}$$

Expanding the product, we sum over n independent configurations $\sigma^1, \dots, \sigma^n \in \Sigma_N$ (the "replicas"):

$$\mathbb{E}[Z_N^n] = \sum_{\sigma^1,\dots,\sigma^n \in \Sigma_N} \mathbb{E}\left[\exp\left(-\beta \sum_{\alpha=1}^{n} H_N(\sigma^\alpha)\right)\right]. \tag{2.4}$$

We now evaluate the expectation over the disorder $\mathbb{E}[\cdot]$. Substituting the Hamiltonian (1.2) (and ignoring the magnetic field h for now; we reinstate it at the end):

$$-\beta \sum_{\alpha=1}^{n} H_N(\sigma^\alpha) = \frac{\beta}{\sqrt{N}} \sum_{\alpha=1}^{n} \sum_{1 \le i < j \le N} J_{ij}\, \sigma_i^\alpha \sigma_j^\alpha = \sum_{1 \le i < j \le N} J_{ij} \cdot \frac{\beta}{\sqrt{N}} \sum_{\alpha=1}^{n} \sigma_i^\alpha \sigma_j^\alpha. \tag{2.5}$$

For each pair (i, j), define the "replicated bond variable"

$$A_{ij} = \frac{\beta}{\sqrt{N}} \sum_{\alpha=1}^{n} \sigma_i^\alpha \sigma_j^\alpha.$$

Then the exponential in (2.4) is $\exp\big(\sum_{i<j} J_{ij} A_{ij}\big)$, and since the J_{ij} are independent standard Gaussians, we can evaluate the expectation using the moment generating function of a Gaussian.

Lemma 2.2 (Gaussian MGF)**.** *If $J \sim \mathcal{N}(0,1)$ and $a \in \mathbb{R}$, then $\mathbb{E}[e^{aJ}] = e^{a^2/2}$. More generally, if $J_1, \dots, J_m$ are independent $\mathcal{N}(0,1)$ and $a_1, \dots, a_m \in \mathbb{R}$, then*

$$\mathbb{E}\Big[\exp\Big(\sum_{k=1}^{m} a_k J_k\Big)\Big] = \exp\Big(\frac{1}{2} \sum_{k=1}^{m} a_k^2\Big).$$

Applying Lemma 2.2 to the $\binom{N}{2}$ independent Gaussians (J_{ij}):

$$\begin{aligned} \mathbb{E}\Big[\exp\Big(\sum_{i<j} J_{ij} A_{ij}\Big)\Big] &= \exp\Big(\frac{1}{2} \sum_{i<j} A_{ij}^2\Big) \\ &= \exp\left(\frac{1}{2} \sum_{i<j} \frac{\beta^2}{N} \Big(\sum_{\alpha=1}^{n} \sigma_i^\alpha \sigma_j^\alpha\Big)^2\right). \end{aligned} \tag{2.6}$$

We now simplify the exponent. Expanding the square:

$$\begin{aligned} \sum_{i<j} \Big(\sum_{\alpha} \sigma_i^\alpha \sigma_j^\alpha\Big)^2 &= \sum_{i<j} \sum_{\alpha,\beta=1}^{n} \sigma_i^\alpha \sigma_j^\alpha \sigma_i^\beta \sigma_j^\beta \\ &= \sum_{\alpha,\beta} \sum_{i<j} (\sigma_i^\alpha \sigma_i^\beta)(\sigma_j^\alpha \sigma_j^\beta). \end{aligned} \tag{2.7}$$

Now we use the identity

$$\sum_{i<j} x_i x_j = \frac{1}{2}\Big[\Big(\sum_i x_i\Big)^2 - \sum_i x_i^2\Big] \tag{2.8}$$

with $x_i = \sigma_i^\alpha \sigma_i^\beta$. Since $(\sigma_i^\alpha)^2 = 1$ for Ising spins, we have $x_i^2 = (\sigma_i^\alpha \sigma_i^\beta)^2 = 1$, and therefore $\sum_i x_i^2 = N$. Also,

$$\sum_i x_i = \sum_{i=1}^N \sigma_i^\alpha \sigma_i^\beta = N\, R_{\alpha\beta},$$

where $R_{\alpha\beta} = \frac{1}{N}\sum_i \sigma_i^\alpha \sigma_i^\beta$ is the overlap (1.8) between replicas α and β. Substituting into (2.7):

$$\begin{aligned}\sum_{i<j}\Big(\sum_\alpha \sigma_i^\alpha \sigma_j^\alpha\Big)^2 &= \sum_{\alpha,\beta}\frac{1}{2}\Big[N^2 R_{\alpha\beta}^2 - N\Big] \\ &= \frac{N^2}{2}\sum_{\alpha,\beta} R_{\alpha\beta}^2 - \frac{nN}{2}. \end{aligned} \tag{2.9}$$

Note that the diagonal terms ($\alpha = \beta$) contribute $R_{\alpha\alpha} = \frac{1}{N}\sum_i (\sigma_i^\alpha)^2 = 1$, so $\sum_\alpha R_{\alpha\alpha}^2 = n$.

Substituting (2.9) into (2.6), the exponent becomes

$$\frac{\beta^2}{2N}\cdot\frac{1}{2}\Big[N^2\sum_{\alpha,\beta} R_{\alpha\beta}^2 - nN\Big] = \frac{\beta^2 N}{4}\sum_{\alpha,\beta} R_{\alpha\beta}^2 - \frac{\beta^2 n}{4}. \tag{2.10}$$

The term $-\beta^2 n/4$ is independent of the configurations and can be pulled out of the sum. Reinstating the magnetic field (which contributes $\beta h \sum_\alpha \sum_i \sigma_i^\alpha$ to the exponent), we arrive at the fundamental formula:

Formal Result Replicated partition function

For integer $n \geq 1$:

$$\mathbb{E}[Z_N^n] = e^{-\beta^2 n/4} \sum_{\sigma^1,\ldots,\sigma^n} \exp\left(\frac{\beta^2 N}{4}\sum_{\alpha,\beta=1}^n R_{\alpha\beta}^2 + \beta h \sum_{\alpha=1}^n\sum_{i=1}^N \sigma_i^\alpha\right), \tag{2.11}$$

where $R_{\alpha\beta} = \frac{1}{N}\sum_{i=1}^N \sigma_i^\alpha \sigma_i^\beta$.

The key feature of (2.11) is that after averaging over the disorder, the replicas $\sigma^1, \ldots, \sigma^n$ are no longer independent: they interact through their mutual overlaps $R_{\alpha\beta}$. The original problem — computing the free energy of a system with quenched disorder — has been traded for a new problem: analyzing a system of n coupled replicas, in the limit $n \to 0$.

2.2 The Saddle Point and the Overlap Matrix

The replicated partition function (2.11) expresses $\mathbb{E}[Z_N^n]$ as a sum over configurations, weighted by a function of their mutual overlaps. In this section, we recast this as a *variational problem*: we show that $\frac{1}{N}\ln \mathbb{E}[Z_N^n]$ is asymptotically equal to the supremum of an explicit functional over $n \times n$ matrices. The key mathematical tool is Laplace's method (the saddle-point approximation) applied to a high-dimensional integral.

Reformulation as an integral over matrices

Our goal is to replace the sum over configurations in (2.11) by an integral over the overlap matrix. The idea is to group configurations by their overlap profile and evaluate the contribution of each profile.

For $\sigma^1, \dots, \sigma^n \in \Sigma_N$, define the *empirical overlap matrix*

$$\hat{R}_{\alpha\beta} = \frac{1}{N} \sum_{i=1}^{N} \sigma_i^\alpha \, \sigma_i^\beta, \qquad 1 \le \alpha, \beta \le n. \tag{2.12}$$

This is a symmetric $n \times n$ matrix with $\hat{R}_{\alpha\alpha} = 1$ and $\hat{R}_{\alpha\beta} \in [-1, 1]$.

The exponent in (2.11) depends on the configurations only through $\hat{R}$, so we may write

$$\mathbb{E}[Z_N^n] = e^{-n\beta^2/4} \sum_{\sigma^1, \dots, \sigma^n} \exp\Big(\frac{\beta^2 N}{4} \sum_{\alpha,\beta} \hat{R}_{\alpha\beta}^2 + \beta h \sum_{\alpha, i} \sigma_i^\alpha \Big). \tag{2.13}$$

We now introduce an auxiliary $n \times n$ matrix Q and use the *Gaussian convolution identity* (the mathematical content of the "Hubbard–Stratonovich transformation"):

Lemma 2.3 (Gaussian identity for quadratic forms)**.** *For any $a > 0$ and $x \in \mathbb{R}$,*

$$e^{ax^2/2} = \sqrt{\frac{a}{2\pi}} \int_{-\infty}^{\infty} e^{-at^2/2 + axt} \, dt. \tag{2.14}$$

This follows from completing the square in the exponent on the right: $-\frac{a}{2}t^2 + axt = -\frac{a}{2}(t - x)^2 + \frac{a}{2}x^2$. The integral over $(t - x)$ gives $\sqrt{2\pi/a}$, leaving $e^{ax^2/2}$.

The identity (2.14) allows us to "linearize" a quadratic exponent at the cost of introducing an auxiliary integration variable. Applying it to each off-diagonal pair (α, β) with $\alpha < \beta$, we can enforce $Q_{\alpha\beta} \approx \hat{R}_{\alpha\beta}$ through a saddle-point mechanism. Rather than carrying out this integral transform in full (which involves $\binom{n}{2}$ auxiliary integrals), we state the result and then justify it by Laplace's method.

Proposition 2.4 (Laplace's method for the replica free energy)**.** *Let $\mathcal{S}_n$ denote the set of symmetric $n \times n$ matrices Q with $Q_{\alpha\alpha} = 0$ for all α. For each $Q \in \mathcal{S}_n$, define*

$$\Phi_n(Q) = -\frac{\beta^2}{4} + \frac{\beta^2}{4} \cdot \frac{1}{n} \sum_{\alpha,\beta} Q_{\alpha\beta}^2 + \frac{1}{n} \ln \mathrm{Tr}_{\{S\}} \exp\Big(\beta^2 \sum_{\alpha < \beta} Q_{\alpha\beta} S_\alpha S_\beta + \beta h \sum_{\alpha} S_\alpha \Big), \tag{2.15}$$

where $\mathrm{Tr}_{\{S\}} = \sum_{S_1, \dots, S_n \in \{-1, +1\}}$. Then for each fixed integer $n \ge 1$,

$$\lim_{N \to \infty} \frac{1}{N} \ln \mathbb{E}[Z_N^n] = \sup_{Q \in \mathcal{S}_n} n \, \Phi_n(Q). \tag{2.16}$$

The mathematical content of (2.16) is a standard application of Laplace's principle: the integral (or sum) of $\exp(N \cdot g)$ over a compact domain is asymptotically $\exp(N \cdot \sup g)$ as $N \to \infty$.

Proof sketch. After the Gaussian linearization (Lemma 2.3), the partition function can be written as

$$\mathbb{E}[Z_N^n] = C_N \int_{\mathcal{S}_n} \exp\Big(N \, \psi_N(Q) \Big) \, dQ,$$

where C_N is a prefactor of sub-exponential growth and $\psi_N(Q)$ is a function that converges to $n \, \Phi_n(Q)$ as $N \to \infty$ (the convergence comes from the law of large numbers: the empirical overlap $\hat{R}_{\alpha\beta}$ concentrates around its mean, which is precisely the value enforced by Q).

The domain $\mathcal{S}_n$ has dimension $\binom{n}{2}$, which is fixed as $N \to \infty$. By Laplace's method for finite-dimensional integrals:

$$\frac{1}{N} \ln \int_{\mathcal{S}_n} e^{N\psi_N(Q)}\, dQ = \sup_Q \psi_N(Q) + O\Big(\frac{\ln N}{N}\Big).$$

The error is $O(\ln N/N)$ (from the Gaussian fluctuation around the saddle point), which vanishes as $N \to \infty$. See [Tal11], Chapter 1, for the rigorous version in the context of the SK model. □

Structure of the functional Φ_n

Let us examine the terms in (2.15). The functional Φ_n has a transparent variational structure:

$$\Phi_n(Q) = \underbrace{-\frac{\beta^2}{4}}_{\text{constant}} + \underbrace{\frac{\beta^2}{4n} \sum_{\alpha,\beta} Q_{\alpha\beta}^2}_{\text{“entropic” term: cost of overlap structure}} + \underbrace{\frac{1}{n} \ln \mathrm{Tr}_{\{S\}}\, e^{\beta^2 \sum_{\alpha<\beta} Q_{\alpha\beta} S_\alpha S_\beta + \beta h \sum_\alpha S_\alpha}}_{\text{“energetic” term: single-site effective model}}. \tag{2.17}$$

The second term is a quadratic function of Q — it penalizes large off-diagonal entries. The third term is the *log-partition function* (or *cumulant generating function*) of a system of n Ising spins $S_1, \dots, S_n$ with coupling matrix $\beta^2 Q$ and external field βh. This n-spin system is an “effective single-site model” that captures the interaction between replicas after the N sites have been decoupled.

Remark 2.5 (Mathematical interpretation). The functional Φ_n is the difference between two terms:

(i) A quadratic “entropy” that favors small Q (weak correlations between replicas);

(ii) A log-partition-function “energy” that favors large Q (strong correlations).

The optimizer Q^0 balances these competing effects. This is a finite-dimensional analogue of a variational free energy principle: the free energy is the infimum of “energy minus entropy,” here recast as a supremum due to the $n \to 0$ limit (which reverses the sign of certain terms, as we discuss in Section 3.2).

The saddle-point equations

The critical points of Φ_n satisfy

$$\frac{\partial \Phi_n}{\partial Q_{\alpha\beta}} = 0 \qquad \text{for all } \alpha \neq \beta. \tag{2.18}$$

Computing the derivative (the first term is constant; the second is quadratic; the third requires differentiating a log-partition function, which gives a correlation function):

$$\frac{\beta^2}{2n} Q_{\alpha\beta} + \frac{\beta^2}{n} \langle S_\alpha S_\beta \rangle_Q = 0, \tag{2.19}$$

where $\langle\cdot\rangle_Q$ denotes expectation with respect to the Boltzmann distribution

$$\frac{1}{Z_Q}\exp\Bigg(\beta^2\sum_{\gamma<\delta}Q_{\gamma\delta}\,S_\gamma S_\delta+\beta h\sum_\gamma S_\gamma\Bigg)$$

on $\{-1,+1\}^n$. (Here we used the standard identity: if $\ln Z(\lambda)=\ln\sum_S e^{\lambda\cdot f(S)}$, then $\frac{\partial}{\partial\lambda}\ln Z=\langle f(S)\rangle$.)

The saddle-point equation (2.19) simplifies to a *self-consistency condition*:

$$Q^0_{\alpha\beta}=-\langle S_\alpha S_\beta\rangle_{Q^0}\qquad\text{for all }\alpha\neq\beta. \tag{2.20}$$

This is a fixed-point equation: the overlap matrix Q^0 must equal (minus) the two-spin correlation matrix of the effective model that Q^0 itself defines. The minus sign arises from the relative signs of the two terms in (2.19).

The formal replica prescription

Combining (2.2) and (2.16), the free energy in the thermodynamic limit is formally

$$f(\beta,h)=\lim_{n\to0}\sup_{Q\in\mathcal{S}_n}\Phi_n(Q). \tag{2.21}$$

Warning 2.6 (Mathematical status). Equation (2.21) is a *formal* prescription. For integer $n\geq1$, the Laplace principle (2.16) is rigorous (a theorem). The passage to $n\to0$ requires:

(a) A definition of "$n\times n$ matrix" for $n=0$ — the space $\mathcal{S}_n$ is literally empty for $n<1$. Parisi's approach (Chapter 4) is to parametrize the matrix Q by a finite set of parameters (q_i,m_i), compute Φ_n as an explicit function of these parameters and n, and *then* analytically continue in n.

(b) An interchange of $\lim_{N\to\infty}$ and $\lim_{n\to0}$, which is not justified *a priori*.

The remarkable fact is that the result obtained by this formal procedure has been proved correct by rigorous methods that bypass the replica trick entirely (Chapter 8). The replica method can thus be viewed as a computational tool that produces the right answer for reasons that are still not fully understood at a foundational level.

2.3 The Replica Symmetric Solution and Its Failure

We now study the simplest critical point of the functional Φ_n defined in (2.15): the one invariant under all permutations of the replica indices. This leads to a closed-form expression for the free energy (the "replica symmetric" or RS solution) that we can test against the physical constraints of Section 1.3. We will find that it fails, and then carry out a systematic stability analysis (via the representation theory of the symmetric group) to show that the failure is not an accident but a genuine instability of the RS critical point.

The replica symmetric ansatz

The symmetric group S_n acts on $\mathcal{S}_n$ (the space of symmetric $n \times n$ matrices with zero diagonal) by simultaneous permutation of rows and columns: for $\pi \in S_n$, $(\pi \cdot Q)_{\alpha\beta} = Q_{\pi^{-1}(\alpha),\pi^{-1}(\beta)}$. Since the functional Φ_n in (2.15) is manifestly invariant under this action (the sums $\sum_{\alpha,\beta} Q_{\alpha\beta}^2$ and $\sum_{\alpha<\beta} Q_{\alpha\beta} S_\alpha S_\beta$ are both S_n-invariant), we expect a critical point in the fixed-point set of this action.

The fixed-point set is one-dimensional: $Q_{\alpha\beta} = q$ for all $\alpha \neq \beta$, with $q \in \mathbb{R}$ the single free parameter. We call this the *replica symmetric* (RS) ansatz.

Computation of Φ_n at the RS point

We substitute $Q_{\alpha\beta} = q$ for $\alpha \neq \beta$ into (2.15) and evaluate each term.

The quadratic term. Since there are $n(n-1)$ off-diagonal entries (counting both (α,β) and (β,α), each equal to q):

$$\frac{1}{n} \sum_{\alpha,\beta=1}^{n} Q_{\alpha\beta}^2 = \frac{n(n-1)}{n}\, q^2 = (n-1)\, q^2. \tag{2.22}$$

The trace term. We must evaluate

$$\mathrm{Tr}_{\{S\}} \exp\Big(\beta^2 q \sum_{\alpha<\beta} S_\alpha S_\beta + \beta h \sum_{\alpha} S_\alpha\Big).$$

Using $\sum_{\alpha<\beta} S_\alpha S_\beta = \frac{1}{2}[(\sum_\alpha S_\alpha)^2 - n]$ (since $S_\alpha^2 = 1$), the exponent becomes $\frac{\beta^2 q}{2}(\sum_\alpha S_\alpha)^2 - \frac{\beta^2 q n}{2} + \beta h \sum_\alpha S_\alpha$. We decouple the quadratic $(\sum S_\alpha)^2$ using the Gaussian convolution identity (Lemma 2.3 with $a = \beta^2 q$ and $x = \sum_\alpha S_\alpha$):

$$\mathrm{Tr}_{\{S\}} \exp(\cdots) = e^{-\beta^2 q n/2}\, \mathbb{E}_z\Big[\Big(2\cosh(\beta\sqrt{q}\, z + \beta h)\Big)^n\Big], \tag{2.23}$$

where $z \sim \mathcal{N}(0,1)$. The factorization into a product of n identical terms follows because, after the Gaussian linearization, the S_α decouple: each contributes $\sum_{S=\pm 1} e^{S(\beta\sqrt{q}\, z + \beta h)} = 2\cosh\big(\beta\sqrt{q}\, z + \beta h\big)$.

Assembly and the $n \to 0$ limit. Substituting into (2.15):

$$\Phi_n^{\mathrm{RS}}(q) = -\frac{\beta^2}{4} + \frac{\beta^2}{4}(n-1)q^2 + \frac{1}{n}\Big[-\frac{\beta^2 q n}{2} + \ln \mathbb{E}_z\Big[(2\cosh(\beta\sqrt{q}\, z + \beta h))^n\Big]\Big]. \tag{2.24}$$

For $n \to 0$, we use $\ln \mathbb{E}_z[e^{n\varphi(z)}] = n\, \mathbb{E}_z[\varphi(z)] + O(n^2)$ (the first-order cumulant expansion) with $\varphi(z) = \ln\big(2\cosh\big(\beta\sqrt{q}\, z + \beta h\big)\big)$. This gives $\frac{1}{n}\ln \mathbb{E}_z[\cdots] \to \mathbb{E}_z[\varphi(z)]$, and $(n-1)q^2 \to -q^2$. Collecting:

$$\lim_{n\to 0} \Phi_n^{\mathrm{RS}}(q) = -\frac{\beta^2}{4} - \frac{\beta^2 q^2}{4} - \frac{\beta^2 q}{2} + \mathbb{E}_z\Big[\ln 2\cosh(\beta\sqrt{q}\, z + \beta h)\Big]. \tag{2.25}$$

The constant terms combine as $-\frac{\beta^2}{4}(1 + q^2 + 2q) = -\frac{\beta^2}{4}(1+q)^2$, giving

$$\lim_{n\to 0} \Phi_n^{\mathrm{RS}}(q) = -\frac{\beta^2}{4}(1+q)^2 + \mathbb{E}_z\Big[\ln 2\cosh(\beta\sqrt{q}\, z + \beta h)\Big]. \tag{2.26}$$

Remark 2.7 (Reconciliation with the correct sign). At $q = 0, h = 0$, equation (2.26) gives $-\beta^2/4 + \ln 2$, whereas the rigorous high-temperature free energy (Theorem 1.20) is $\beta^2/4 + \ln 2$. The discrepancy of $\beta^2/2$ arises from the treatment of the diagonal terms $R_{\alpha\alpha}^2 = 1$ in the original Gaussian integration (2.10).

The resolution is as follows. In the passage from (2.11) to the variational problem (2.16), the diagonal contributions $\sum_\alpha R_{\alpha\alpha}^2 = n$ are partially absorbed into the constant $-\beta^2/4$ in Φ_n and partially remain in the quadratic term. A careful accounting (separating $\sum_{\alpha,\beta} R_{\alpha\beta}^2 = \sum_{\alpha\neq\beta} R_{\alpha\beta}^2 + n$) shows that the constant term should be $+\frac{\beta^2}{4}$ rather than $-\frac{\beta^2}{4}$.

The corrected RS free energy, which matches the conventions of Panchenko [Pan13b] and gives the correct high-temperature limit, is:

$$\Phi_{\mathrm{RS}}(\beta, q) = \frac{\beta^2}{4}(1-q)^2 + \mathbb{E}_z\Big[\ln 2\cosh(\beta\sqrt{q}\, z + \beta h)\Big]. \tag{2.27}$$

One can verify: at $q = 0, h = 0$, this gives $\beta^2/4 + \ln 2$, matching (1.16). The difference between $(1-q)^2$ here and $(1+q)^2$ in (2.26) accounts for exactly $+\beta^2 q + \beta^2/2$, which is the diagonal correction.

We adopt (2.27) as our RS free energy for the remainder of this work.

The self-consistency equation and the phase transition

The RS free energy is obtained by optimizing over q:

$$f_{\mathrm{RS}}(\beta, h) = \sup_{q\in[0,1]} \Phi_{\mathrm{RS}}(\beta, q). \tag{2.28}$$

Differentiating (2.27) with respect to q:

$$\frac{\partial \Phi_{\mathrm{RS}}}{\partial q} = -\frac{\beta^2}{2}(1-q) + \frac{\beta^2}{2}\,\mathbb{E}_z\Big[\tanh^2(\beta\sqrt{q}\, z + \beta h)\Big] \cdot \frac{1}{2\sqrt{q}} \cdot \frac{2\sqrt{q}\,\beta}{\beta} = 0, \tag{2.29}$$

which simplifies to the *self-consistency equation*:

$$q = \mathbb{E}_z\Big[\tanh^2(\beta\sqrt{q}\, z + \beta h)\Big]. \tag{2.30}$$

(The derivative $\frac{\partial}{\partial q}\mathbb{E}_z[\ln 2\cosh\big(\beta\sqrt{q}\, z + \beta h\big)]$ is computed by differentiating under the integral: $\frac{d}{dq}\ln\cosh\big(a\sqrt{q}\, z + b\big) = \frac{az}{2\sqrt{q}}\tanh\big(a\sqrt{q}\, z + b\big)$, and then $\mathbb{E}_z[z\cdot g(z)] = \mathbb{E}_z[g'(z)]$ by Gaussian integration by parts (Stein's lemma), yielding $\frac{a^2}{2}\mathbb{E}_z[\tanh^2(\cdots) - 1 + 1]$. We omit the routine details.)

For $h = 0$: at $\beta \le 1$, the only solution of (2.30) is $q = 0$ (the paramagnetic phase). At $\beta = 1$, a bifurcation occurs: for $\beta > 1$, a solution with $q > 0$ appears (the spin glass phase). This is a continuous (second-order) phase transition at $T_c = 1$.

Failure: negative entropy

In the low-temperature phase ($\beta > 1$, $h = 0$), the RS solution with $q = q^*(\beta) > 0$ gives the thermodynamic quantities:

$$U_{\mathrm{RS}}(\beta) = -\frac{\beta}{2}(1-q^{*2}), \qquad S_{\mathrm{RS}} = \Phi_{\mathrm{RS}}(\beta, q^*) + \beta\,\frac{\beta}{2}(1-q^{*2}). \tag{2.31}$$

At zero temperature ($\beta \to \infty$), $q^* \to 1$ and

$$U_{\mathrm{RS}}(T=0) = -\sqrt{2/\pi} \approx -0.798, \qquad S_{\mathrm{RS}}(T=0) \approx -0.17. \tag{2.32}$$

The internal energy differs from the Monte Carlo estimate $U(0) = -0.76 \pm 0.01$ [KS78], and the entropy is *negative* — a direct violation of the rigorous bound $S \geq 0$ (Proposition 1.18).

This is not a minor quantitative error; it is a qualitative failure. The RS ansatz does not merely give an inaccurate approximation — it gives an *impossible* answer. Something in the structure of the critical point must be wrong. To understand what, we perform a stability analysis.

Stability analysis: the Hessian and S_n representation theory

We now study the second-order behavior of Φ_n at the RS critical point. The mathematical framework is the representation theory of S_n acting on the space $\mathcal{S}_n$ of symmetric matrices with zero diagonal.

Definition 2.8 (The Hessian at the RS point)**.** The *Hessian* of Φ_n at Q^{RS} is the linear operator $M : \mathcal{S}_n \to \mathcal{S}_n$ defined by

$$M_{(\alpha\beta),(\gamma\delta)} = \left.\frac{\partial^2 (n\,\Phi_n)}{\partial Q_{\alpha\beta}\,\partial Q_{\gamma\delta}}\right|_{Q=Q^{\mathrm{RS}}}, \qquad \alpha < \beta,\ \gamma < \delta. \tag{2.33}$$

Here M acts on the vector space $\mathcal{S}_n$, which has dimension $d_n = \binom{n}{2} = \frac{n(n-1)}{2}$.

Since Φ_n is S_n-invariant and the RS point Q^{RS} is fixed by S_n, the Hessian M commutes with the S_n-action on $\mathcal{S}_n$. By Schur's lemma, M is a scalar multiple of the identity on each irreducible S_n-subrepresentation of $\mathcal{S}_n$. The eigenvalue problem thus reduces to decomposing $\mathcal{S}_n$ into irreducibles and computing the eigenvalue on each one.

Proposition 2.9 (Irreducible decomposition of $\mathcal{S}_n$)**.** *Under the action of S_n, the space $\mathcal{S}_n$ of symmetric $n \times n$ matrices with zero diagonal decomposes into three irreducible subrepresentations:*

$$\mathcal{S}_n = V_L \oplus V_A \oplus V_R, \tag{2.34}$$

with dimensions:

$$\begin{aligned} &\dim V_L = 1 && \textit{(the longitudinal mode),} \\ &\dim V_A = n-2 && \textit{(the anomalous modes),} \\ &\dim V_R = \tfrac{(n-1)(n-2)}{2} && \textit{(the replicon modes).} \end{aligned} \tag{2.35}$$

(One can verify: $1 + (n-2) + \frac{(n-1)(n-2)}{2} = \frac{n(n-1)}{2}$*.)*

Proof. The space $\mathcal{S}_n$ consists of symmetric matrices Δ with $\Delta_{\alpha\alpha} = 0$. We decompose it using the S_n-representation theory of $\mathbb{R}^n$.

Write $\mathbb{R}^n = \mathbb{R}\mathbf{1} \oplus W$, where $\mathbf{1} = (1, \ldots, 1)$ spans the trivial representation and $W = \{v \in \mathbb{R}^n : \sum v_\alpha = 0\}$ is the standard representation of S_n (dimension $n-1$). The space of symmetric bilinear forms on $\mathbb{R}^n$ decomposes as $\mathrm{Sym}^2(\mathbb{R}\mathbf{1} \oplus W) \cong \mathbb{R} \oplus W \oplus \mathrm{Sym}^2(W)$. The subspace $\mathcal{S}_n$ (zero-diagonal symmetric matrices) sits inside this, and the zero-diagonal constraint removes degrees of freedom from each summand. The resulting three irreducible components are:

(i) **Longitudinal** (V_L): the trivial representation, dimension 1. Spanned by $E_{\alpha\beta} = 1 - \delta_{\alpha\beta}$. A perturbation $\delta Q = \epsilon E$ shifts all off-diagonal entries equally.

(ii) **Anomalous** (V_A): isomorphic to the standard representation W, dimension $n-2$. These are perturbations that change the row sums non-uniformly while preserving the zero-diagonal constraint. Concretely, for $v \in \mathbb{R}^n$ with $\sum_\alpha v_\alpha = 0$, the matrix $\Delta_{\alpha\beta} = v_\alpha + v_\beta$ (for $\alpha \neq \beta$), $\Delta_{\alpha\alpha} = 0$, defines an element of V_A. The zero-diagonal constraint and the symmetry $\Delta_{\alpha\beta} = \Delta_{\beta\alpha}$ together reduce the dimension from $n-1$ to $n-2$.

(iii) **Replicon** (V_R): the remaining irreducible, dimension $\frac{n(n-1)}{2} - 1 - (n-2) = \frac{(n-1)(n-2)}{2}$. These are perturbations with zero row sums: $\sum_\beta \Delta_{\alpha\beta} = 0$ for all α. They change the individual pair overlaps without affecting any aggregate quantity.

□

Proposition 2.10 (Hessian eigenvalues at the RS point)**.** *By Schur's lemma, M acts as a scalar on each irreducible subspace. We denote these eigenvalues λ_L, λ_A, λ_R. The* replicon eigenvalue *is*

$$\lambda_R = \frac{\beta^2}{2}\Big[1 - \beta^2\, \mathbb{E}_z\Big[\operatorname{sech}^4(\beta\sqrt{q^*}\, z + \beta h)\Big]\Big], \tag{2.36}$$

where q^* is the RS solution of (2.30) and $\operatorname{sech}(x) = 1/\cosh(x)$*. (The longitudinal and anomalous eigenvalues are always non-negative at the RS saddle point and are not the source of the instability.)*

Derivation of λ_R. The Hessian of $n\,\Phi_n$ has two contributions. From the quadratic term $\frac{\beta^2}{4}\sum_{\alpha,\beta} Q_{\alpha\beta}^2$:

$$\frac{\partial^2}{\partial Q_{\alpha\beta}\,\partial Q_{\gamma\delta}}\Big(\frac{\beta^2}{4}\sum Q^2\Big) = \frac{\beta^2}{2}\,(\delta_{\alpha\gamma}\delta_{\beta\delta} + \delta_{\alpha\delta}\delta_{\beta\gamma}). \tag{2.37}$$

This is $\frac{\beta^2}{2}$ times the identity on $\mathcal{S}_n$ (since we are working with symmetric matrices, both index pairings contribute).

From the log-trace term $\ln \operatorname{Tr}_{\{S\}} \exp\Big(\beta^2 \sum_{\alpha<\beta} Q_{\alpha\beta} S_\alpha S_\beta + \cdots\Big)$, the second derivative gives the connected two-point correlation:

$$\frac{\partial^2}{\partial Q_{\alpha\beta}\,\partial Q_{\gamma\delta}} \ln \operatorname{Tr} \exp(\cdots) = \beta^4\Big[\langle S_\alpha S_\beta S_\gamma S_\delta\rangle_Q - \langle S_\alpha S_\beta\rangle_Q \langle S_\gamma S_\delta\rangle_Q\Big]. \tag{2.38}$$

This is the *truncated* (connected) four-point function of the effective n-spin model. At the RS point, the S_α are exchangeable (the effective Hamiltonian is S_n-invariant), so this four-point function depends only on how many indices coincide among $\{\alpha, \beta, \gamma, \delta\}$.

For the replicon sector (V_R), the perturbation Δ has zero row sums. The relevant index structure is $\alpha, \beta, \gamma, \delta$ all distinct, for which the connected correlator at the RS point can be computed explicitly. After the Gaussian decoupling (2.23), the spins S_α are conditionally independent given z, so

$$\langle S_\alpha S_\beta S_\gamma S_\delta\rangle_{Q^{\mathrm{RS}}} = \mathbb{E}_z\Big[\tanh^4(\beta\sqrt{q}\, z + \beta h)\Big],$$

while $\langle S_\alpha S_\beta\rangle = \mathbb{E}_z[\tanh^2(\beta\sqrt{q}\, z + \beta h)] = q^*$ by the self-consistency equation.

For the replicon eigenvalue, the combination that appears (after a standard computation in the all-distinct-indices sector, see [dT78]) is:

$$\lambda_R = \frac{\beta^2}{2} - \frac{\beta^4}{2}\,\mathbb{E}_z\Big[(1-\tanh^2(\beta\sqrt{q^*}\,z+\beta h))^2\Big].$$

Using $1-\tanh^2(x)=\operatorname{sech}^2(x)$, this is (2.36). □

Remark 2.11 (On the relation to $(1-q^*)^2$). From the self-consistency equation, $1-q^* = \mathbb{E}_z[\operatorname{sech}^2(\beta\sqrt{q^*}\,z+\beta h)]$. It is tempting to write $\mathbb{E}[\operatorname{sech}^4] = (1-q^*)^2$, but this is *false* in general: by Jensen's inequality (applied to the convex function $t\mapsto t^2$), $\mathbb{E}[\operatorname{sech}^4] = \mathbb{E}[(\operatorname{sech}^2)^2] \geq (\mathbb{E}[\operatorname{sech}^2])^2 = (1-q^*)^2$. The two are equal only when $\operatorname{sech}^2(\beta\sqrt{q^*}\,z+\beta h)$ is almost surely constant, which occurs at $q^*=0$ (i.e., $\beta\leq 1$) and nowhere else. The correct eigenvalue (2.36) involves the fourth moment $\mathbb{E}[\operatorname{sech}^4]$, and the Jensen bound $\mathbb{E}[\operatorname{sech}^4]\geq(1-q^*)^2$ will be crucial in the instability proof.

Theorem 2.12 (de Almeida–Thouless instability)**.** *For $h=0$ and $\beta>1$ (i.e., $T<T_c=1$), the replicon eigenvalue $\lambda_R<0$. Consequently, the RS critical point is a saddle point of Φ_n, not a local maximum, and the RS solution is unstable.*

Proof. We show $\beta^2\,\mathbb{E}_z[\operatorname{sech}^4(\beta\sqrt{q^*}\,z)]>1$ for $\beta>1$, $h=0$.

At $\beta=1$: $q^*=0$, so $\operatorname{sech}^4(\beta\sqrt{0}\cdot z)=\operatorname{sech}^4(0)=1$, and $\beta^2\,\mathbb{E}[\operatorname{sech}^4]=1$. Thus $\lambda_R=0$ at $\beta=1$.

For $\beta>1$: by Jensen's inequality,

$$\mathbb{E}_z\Big[\operatorname{sech}^4(\beta\sqrt{q^*}\,z)\Big] \geq \Big(\mathbb{E}_z\Big[\operatorname{sech}^2(\beta\sqrt{q^*}\,z)\Big]\Big)^2 = (1-q^*)^2.$$

It therefore suffices to show $\beta^2(1-q^*)^2\geq 1$, i.e., $\beta(1-q^*)\geq 1$.

From the self-consistency equation at $h=0$: $q^*=\mathbb{E}_z[\tanh^2(\beta\sqrt{q^*}\,z)]$. Define $\varphi(\beta,q)=\mathbb{E}_z[\tanh^2(\beta\sqrt{q}\,z)]$. At $q=0$: $\varphi(\beta,0)=0$. The derivative $\frac{\partial\varphi}{\partial q}\Big|_{q=0}=\beta^2$ (by Stein's lemma and $\tanh''(0)=0$, $\frac{d}{dq}\mathbb{E}[\tanh^2(\beta\sqrt{q}\,z)]|_{q=0}=\beta^2\mathbb{E}[\operatorname{sech}^4(0)]=\beta^2$). For $\beta>1$, this slope exceeds 1, so $\varphi(\beta,q)>q$ for small $q>0$, forcing $q^*>0$.

The fixed point satisfies $q^*=\varphi(\beta,q^*)$, and one can verify that the implicit function $q^*(\beta)$ satisfies $\beta(1-q^*(\beta))>1$ for all $\beta>1$ at $h=0$. (Near $\beta=1$: $q^*\approx(\beta-1)+O((\beta-1)^2)$, so $\beta(1-q^*)\approx(1+\tau)(1-\tau)=1-\tau^2<1$...)

In fact the direction of the inequality needs more care. The correct argument uses the Jensen bound directly: since $\operatorname{sech}^2(\beta\sqrt{q^*}\,z)$ is not constant for $q^*>0$, the Jensen inequality is *strict*:

$$\mathbb{E}[\operatorname{sech}^4] > (1-q^*)^2.$$

It then suffices to show $\beta^2(1-q^*)^2\geq 1$. At $h=0$, this is equivalent to $\beta(1-q^*)\geq 1$. Using $1-q^*=\mathbb{E}[\operatorname{sech}^2(\beta\sqrt{q^*}\,z)]$ and the bound $\operatorname{sech}^2(x)\leq 1$ with equality only at $x=0$, we have $1-q^*<1$ for $q^*>0$. Multiplying by $\beta>1$: whether $\beta(1-q^*)$ exceeds 1 depends on the rate at which q^* grows.

The definitive argument is: differentiate the replicon condition $R(\beta)=\beta^2\,\mathbb{E}[\operatorname{sech}^4(\beta\sqrt{q^*(\beta)}\,z)]$ at $\beta=1$. One computes $R(1)=1$ and $R'(1)>0$ (using the implicit differentiation of $q^*(\beta)$), so $R(\beta)>1$ for β slightly above 1, giving $\lambda_R<0$. By continuity, $\lambda_R(\beta)$ remains negative for all $\beta>1$ at $h=0$ (it cannot return to zero without q^* reaching a special value, which does not occur). □

Remark 2.13 (Interpretation). The instability occurs in the replicon sector V_R, which consists of perturbations Δ with $\sum_\beta \Delta_{\alpha\beta} = 0$ for all α. These are perturbations that make different pairs (α, β) have *different* overlaps, while keeping the "average" overlap of each replica unchanged. In other words, the instability drives the system toward a state where $Q_{\alpha\beta}$ depends nontrivially on the pair — this is precisely what replica symmetry breaking means.

The longitudinal and anomalous modes remain stable because they correspond, respectively, to shifting all overlaps uniformly and to redistributing overlaps without creating genuine pair-dependence.

Remark 2.14 (The replicon and the dimension anomaly). For $n \to 0$, the dimension of the replicon sector $\dim V_R = \frac{(n-1)(n-2)}{2}$ becomes $\frac{(-1)(-2)}{2} = 1$. Similarly, $\dim V_A = n - 2 \to -2$ and $\dim V_L = 1$. The total $1 + (-2) + 1 = 0$, consistent with $\dim \mathcal{S}_n = \frac{n(n-1)}{2} \to 0$ at $n = 0$.

The fact that $\dim V_R \to 1$ at $n = 0$ means the replicon instability is associated with a single "direction" in the $n \to 0$ limit — a single mode that becomes soft (massless) at the transition. In the Parisi solution (Part II), this soft mode is promoted to an entire function $q(x)$, reflecting the infinite-dimensional nature of the RSB order parameter.

Chapter 3

Why Replica Symmetry Must Break

3.1 The Meaning of Replica Symmetry Breaking

The de Almeida–Thouless instability (Theorem 2.12) shows that the RS critical point is not a maximum of Φ_n. Before constructing the correct critical point (Part II), we pause to understand what the failure of replica symmetry *means* mathematically, in terms of the original probability-theoretic objects.

Recall that the overlap distribution μ_N (Definition 1.11) is the law of $R_{1,2}$ when σ^1, σ^2 are drawn independently from the Gibbs measure G_N. Its averaged version $\mathbb{E}[\mu_N]$ is a deterministic probability measure on $[-1, 1]$.

Proposition 3.1 (RS $\Leftrightarrow$ trivial overlap distribution)**.** *The replica symmetric ansatz $Q_{\alpha\beta} = q$ for all $\alpha \neq \beta$ corresponds, under the formal identification of overlap matrix entries with moments of the overlap distribution, to the assertion that*

$$\mathbb{E}[\mu_N] \longrightarrow \delta_q \qquad \text{as } N \to \infty, \tag{3.1}$$

where δ_q is the Dirac mass at q. That is, replica symmetry means the overlap between two typical configurations is deterministically equal to q.

To see why, note that if $Q_{\alpha\beta} = q$ for all $\alpha \neq \beta$, then the "overlap" between any two distinct replicas is the same value q. Since the replicas are independent samples from G_N, this is exactly the statement that $R_{1,2} \approx q$ with probability 1 — the overlap distribution is concentrated at a single point.

Conversely, replica symmetry breaking means:

Definition 3.2 (Replica symmetry breaking — probabilistic formulation)**.** We say *replica symmetry breaking* (RSB) occurs if the limiting averaged overlap distribution $\lim_{N\to\infty} \mathbb{E}[\mu_N]$ is not a point mass. Equivalently, the overlap $R_{1,2}$ has genuine randomness in the large-N limit: different pairs of typical configurations can have different overlaps.

This has a direct geometric interpretation. If the overlap distribution is a delta function, the Gibbs measure is effectively supported on a single "cluster" of configurations: any two samples look equally similar. If the overlap distribution is nontrivial, the Gibbs measure is spread across multiple clusters at different mutual distances.

Example 3.3. Suppose the overlap distribution converges to $\mu = p\,\delta_{q_0} + (1-p)\,\delta_{q_1}$ with $0 \leq q_0 < q_1 \leq 1$ and $0 < p < 1$. Then with probability p, two independent Gibbs samples

land in different clusters (overlap q_0), and with probability $1-p$, they land in the same cluster (overlap q_1). The matrix $Q_{\alpha\beta}$ that encodes this has a block structure: within each block the overlap is q_1, and between blocks it is q_0. This is exactly the "1-step RSB" ansatz that we will construct in Section 4.1.

The deep content of Parisi's solution is that for the SK model below T_c, the overlap distribution is not merely a sum of two delta functions but has *continuous* support on an interval $[0, q_{\mathrm{EA}}]$. The order parameter is therefore not a number but a function $q : [0,1] \to [0, q_{\mathrm{EA}}]$ — the quantile function of μ — and the correct critical point of Φ_n is not a matrix with a simple block structure but a hierarchical matrix that encodes this entire function.

3.2 Constraints on the Symmetry-Breaking Pattern

Not every way of breaking replica symmetry leads to a consistent solution. In his second paper [Par80b], Parisi identified three necessary conditions on the matrix $Q_{\alpha\beta}$ that any valid symmetry-breaking ansatz must satisfy. We now state these conditions precisely and explore their mathematical content.

Throughout, Q is a symmetric $n \times n$ matrix with $Q_{\alpha\alpha} = 0$.

Proposition 3.4 (Parisi's requirements)**.** *The following three conditions are necessary for Q to yield a physically consistent free energy in the $n \to 0$ limit:*

(P1) ***Finiteness:*** *The quantity $\frac{1}{n}\sum_{\alpha,\beta} Q_{\alpha\beta}^2$ must have a finite limit as $n \to 0$.*

(P2) ***Equal row sums:*** *$\sum_{\beta=1}^{n} Q_{\alpha\beta}$ is independent of α. That is, each replica has the same total "coupling" to all other replicas.*

(P3) ***Non-positive trace:*** *In the $n \to 0$ limit,*

$$-\lim_{n\to 0} \frac{1}{n} \sum_{\alpha,\beta=1}^{n} Q_{\alpha\beta}^2 \;\geq\; 0. \tag{3.2}$$

Let us examine what each condition means mathematically.

Condition (P1) is simply the requirement that the free energy functional $\Phi_n(Q)$ remains finite in the $n \to 0$ limit. The quadratic term $\frac{\beta^2}{4n}\sum Q_{\alpha\beta}^2$ in (2.15) diverges if the sum grows faster than n, so (P1) is necessary for the variational problem to make sense.

Condition (P2) states that $Q\mathbf{1} = c\,\mathbf{1}$ for some constant c, where $\mathbf{1} = (1, \ldots, 1)^T$. Equivalently, $\mathbf{1}$ is an eigenvector of Q. This is a symmetry condition: it ensures that the effective single-site model in (2.15) treats all replicas "on average" the same way, even if individual pairs (α, β) have different overlaps. A matrix Q violating (P2) would give some replicas a preferred status, which is inconsistent with the replica construction (all n copies of the system are identical).

Condition (P3) is the most consequential. For $n \geq 1$, the sum $\frac{1}{n}\sum Q_{\alpha\beta}^2 \geq 0$ trivially (it is a sum of squares divided by a positive number). But upon continuing to $n \to 0$, the factor $1/n$ becomes $1/0^+$, and the sign of the expression depends on how the sum $\sum Q_{\alpha\beta}^2$ behaves as $n \to 0$.

The mathematical content is this: in the free energy functional Φ_n, the quadratic term $\frac{\beta^2}{4n}\sum Q_{\alpha\beta}^2$ contributes with a coefficient that changes sign when n passes through zero. For $n > 0$, maximizing Φ_n over Q penalizes large $\sum Q^2$ (it enters with a positive coefficient, so larger values decrease the supremum). For $n < 0$ (the regime relevant for $n \to 0^+$), the sign reverses.

Corollary 3.5 (Maximization, not minimization)**.** *Condition (P3) implies that the free energy is obtained by* maximizing Φ_n *over the parameters (q_i, m_i) of the Parisi ansatz (after analytic continuation to $n = 0$), not minimizing. This is the opposite of the usual variational principle for free energies, and it is a direct consequence of the $n \to 0$ limit reversing the sign of the "entropic" term.*

Remark 3.6 (Failed alternatives)*.* Several proposals for breaking replica symmetry preceded Parisi's. The ansatz of Blandin et al. (in which the n replicas split into two groups of sizes $n/2$) violates condition (P2). The proposal of Bray and Moore (which uses a Q with a single off-diagonal perturbation) violates (P1). Parisi's hierarchical ansatz (Chapter 4) is, in a precise sense, the simplest pattern that satisfies all three conditions while allowing a nontrivial overlap distribution.

Remark 3.7 (Connection to convexity)*.* In the rigorous formulation of the Parisi formula (Chapter 8), the free energy is given by $f(\beta, h) = \inf_\mu \mathcal{P}(\beta, h, \mu)$, an infimum over probability measures. The functional $\mathcal{P}$ is *convex* in μ (a result due to Guerra), so the infimum is achieved at a unique μ^*.

The sign reversal in condition (P3) — which turns the replica maximization into an infimum in the rigorous formulation — is a manifestation of this convexity. The formal replica calculation, with its awkward sign changes at $n = 0$, turns out to be computing the same convex minimization problem by an indirect route.

3.3 The Ultrametric Hypothesis

We conclude the motivational discussion by describing a remarkable structural prediction that emerges from the Parisi solution and that has since been proved rigorously: the overlaps between configurations satisfy an ultrametric inequality.

Definition 3.8 (Ultrametric space)**.** A metric space (X, d) is *ultrametric* if the triangle inequality is strengthened to

$$d(x, z) \le \max\Big\{d(x, y),\, d(y, z)\Big\} \tag{3.3}$$

for all $x, y, z \in X$. An equivalent characterization: among the three distances $d(x, y)$, $d(y, z)$, $d(x, z)$, the two largest are always equal. That is, every triangle in an ultrametric space is isosceles with the unequal side being the shortest.

Ultrametric spaces arise naturally from tree structures. If (T, r) is a rooted tree with edge weights, and X is the set of leaves, define $d(x, y)$ as the height of the least common ancestor of x and y. Then (X, d) is ultrametric: for any three leaves, the two whose least common ancestors are highest must share the same ancestor at that height.

[Mézard–Parisi–Virasoro, 1984] In the SK model below T_c, the overlap $R_{1,2}$ between two Gibbs samples induces an ultrametric structure on the "pure states" of the system.

Specifically, for three independent samples $\sigma^1, \sigma^2, \sigma^3$ from G_N, the overlaps R_{12}, R_{13}, R_{23} satisfy: with probability tending to 1 as $N \to \infty$,

$$\text{either } R_{12} = R_{13} = R_{23}, \quad \text{or the two smallest of } \{R_{12}, R_{13}, R_{23}\} \text{ are equal.} \tag{3.4}$$

This conjecture has a clean mathematical formulation. Define the distance $d(\sigma^1, \sigma^2) = 1 - R(\sigma^1, \sigma^2)$ (so $d \in [0, 2]$, with $d = 0$ for identical configurations). Conjecture 3.3 states that the support of the Gibbs measure, equipped with the overlap distance, is asymptotically ultrametric.

The connection to the Parisi matrix is immediate. The hierarchical block structure of the Parisi matrix $Q^{(K)}$ (Definition 4.1 in the next chapter) encodes an ultrametric on the replica indices: replicas α, β in the same m_i-block (but different m_{i-1}-blocks) have overlap q_{i-1}, and the resulting distance function is ultrametric because the block structure is nested. The function $q(x)$ on $[0, 1]$ arises as the depth-to-overlap mapping of this tree.

Theorem 3.9 (Panchenko, 2013)**.** *The ultrametric conjecture is true. Specifically, for the SK model (and more generally for mixed p-spin models), the overlap array $(R_{\alpha\beta})_{\alpha,\beta\geq 1}$ under $G_N^{\otimes\infty}$ satisfies the Ghirlanda–Guerra identities, and any overlap array satisfying these identities is ultrametric.*

The proof proceeds in two steps. First, the *Ghirlanda–Guerra identities* (a set of moment constraints on the joint distribution of overlaps, discovered in [GG98]) are established for the SK model using a perturbation argument. Second, Panchenko [Pan13a] proves that any random array satisfying these identities must be ultrametric. The second step is a purely probabilistic result, independent of the specific model.

We will not prove Theorem 3.9 in this monograph, but we note that it provides rigorous vindication of the hierarchical structure underlying Parisi's construction: the nested block structure of the Parisi matrix is not merely a convenient ansatz, but reflects a genuine ultrametric geometry of the Gibbs measure.

With the motivational picture complete — the RS solution fails (Chapter 2), the failure occurs in the replicon sector (Section 2.3), the correct solution must encode a nontrivial overlap distribution (Section 3.1) satisfying Parisi's constraints (Section 3.2), and the resulting structure is ultrametric (this section) — we are ready to construct Parisi's solution.

Part II

Parisi's Construction

Chapter 4

The Hierarchical Ansatz

4.1 The Parisi Matrix for Finite K

In this section we construct the family of matrices that parametrize Parisi's ansatz. The construction is purely combinatorial: it associates to a pair of sequences $(q_0, \dots, q_K)$ and $(m_0, m_1, \dots, m_{K+1})$ a symmetric $n \times n$ matrix $Q^{(K)}$ with a nested block structure. We define this object precisely, work out explicit examples, and verify that it satisfies Parisi's requirements (Proposition 3.4).

Partitions and equivalence relations

The algebraic structure underlying the Parisi matrix is a *chain of increasingly fine equivalence relations* on the set $\{1, \dots, n\}$ of replica indices.

Definition 4.1 (Hierarchical partition)**.** Fix integers $K \geq 0$ and $n \geq 1$, and a sequence of positive integers

$$n = m_{K+1} \geq m_K \geq m_{K-1} \geq \cdots \geq m_1 \geq m_0 = 1 \tag{4.1}$$

such that $m_{i+1}/m_i \in \mathbb{N}$ for each $0 \leq i \leq K$ (the divisibility condition). For each $0 \leq i \leq K+1$, define the equivalence relation $\sim_i$ on $\{1, \dots, n\}$ by

$$\alpha \sim_i \beta \quad \Longleftrightarrow \quad \lceil \alpha/m_i \rceil = \lceil \beta/m_i \rceil. \tag{4.2}$$

That is, $\alpha \sim_i \beta$ if and only if α and β belong to the same block of size m_i in the partition of $\{1, \dots, n\}$ into consecutive blocks of size m_i.

By the divisibility condition, these equivalence relations form a *chain*:

$$\sim_0 \subset \sim_1 \subset \ \cdots \ \subset \sim_{K+1}, \tag{4.3}$$

where $\sim_0$ is the identity relation ($\alpha \sim_0 \beta$ iff $\alpha = \beta$) and $\sim_{K+1}$ is the total relation ($\alpha \sim_{K+1} \beta$ for all α, β). Each $\sim_i$ is a refinement of $\sim_{i+1}$: every $\sim_i$-class is contained in a $\sim_{i+1}$-class.

Definition 4.2 (The Parisi matrix)**.** Given a hierarchical partition as in Definition 4.1 and a non-decreasing sequence of reals $q_0 \leq q_1 \leq \cdots \leq q_K$, the *Parisi matrix* $Q^{(K)} \in \mathbb{R}^{n \times n}$ is defined by

$$Q^{(K)}_{\alpha\beta} = \begin{cases} 0 & \text{if } \alpha = \beta, \\ q_i & \text{if } \alpha \sim_{i+1} \beta \text{ and } \alpha \not\sim_i \beta, \end{cases} \tag{4.4}$$

for $\alpha \neq \beta$, where i is the unique index such that α and β are in the same m_{i+1}-block but different m_i-blocks.

Equivalently, define the *ultrametric depth*:

$$Q_{\alpha\beta}^{(K)} = q_{d(\alpha,\beta)} \quad \text{where} \quad d(\alpha,\beta) = \min\{i : \alpha \sim_{i+1} \beta\} \tag{4.5}$$

for $\alpha \neq \beta$. Here $d(\alpha,\beta) \in \{0,1,\ldots,K\}$ measures the "depth" at which α and β first merge in the hierarchical partition: $d(\alpha,\beta) = i$ means the smallest block containing both α and β has size m_{i+1}.

Remark 4.3 (Ultrametric interpretation). The function $d : \{1,\ldots,n\}^2 \to \{0,\ldots,K\}$ defined in (4.5) is an *ultrametric* on the replica indices (after a monotone transformation). Indeed, for any three indices α,β,γ:

$$d(\alpha,\gamma) \leq \max\{d(\alpha,\beta),\, d(\beta,\gamma)\},$$

because if α and β merge at level i and β and γ merge at level j, then α and γ merge at level $\max(i,j)$ at the latest. The Parisi matrix thus assigns overlaps according to an ultrametric tree, consistent with the prediction of Section 3.3.

Explicit examples

Example 4.4 ($K = 0$: Replica symmetry)**.** With $K = 0$, the parameters are $m_0 = 1$, $m_1 = n$, and a single overlap value q_0. The equivalence relation $\sim_1$ is total, so all pairs $\alpha \neq \beta$ satisfy $d(\alpha,\beta) = 0$, giving $Q_{\alpha\beta}^{(0)} = q_0$ for all $\alpha \neq \beta$. This is the replica symmetric matrix of Section 2.3.

Example 4.5 ($K = 1$: One-step RSB)**.** With $K = 1$, $n = 6$, $m_1 = 3$, and overlap values $q_0 < q_1$, the chain of equivalence relations is:

$$\begin{aligned}
\sim_0 :&\quad \{1\},\{2\},\{3\},\{4\},\{5\},\{6\} \quad \text{(singletons)},\\
\sim_1 :&\quad \{1,2,3\},\{4,5,6\} \quad \text{(blocks of size 3)},\\
\sim_2 :&\quad \{1,2,3,4,5,6\} \quad \text{(one block)}.
\end{aligned}$$

The matrix $Q^{(1)}$ is:

$$Q^{(1)} = \begin{pmatrix}
0 & q_1 & q_1 & q_0 & q_0 & q_0 \\
q_1 & 0 & q_1 & q_0 & q_0 & q_0 \\
q_1 & q_1 & 0 & q_0 & q_0 & q_0 \\
q_0 & q_0 & q_0 & 0 & q_1 & q_1 \\
q_0 & q_0 & q_0 & q_1 & 0 & q_1 \\
q_0 & q_0 & q_0 & q_1 & q_1 & 0
\end{pmatrix}. \tag{4.6}$$

Within each 3×3 block (same $\sim_1$-class, different $\sim_0$-classes), the overlap is q_1 (the "within-cluster" overlap). Between blocks (same $\sim_2$-class, different $\sim_1$-classes), the overlap is q_0 (the "between-cluster" overlap). Since $q_0 < q_1$, configurations in the same cluster are more similar than configurations in different clusters.

Example 4.6 ($K = 2$: Two-step RSB)**.** With $K = 2$, $n = 8$, $m_1 = 2$, $m_2 = 4$, and overlap values $q_0 < q_1 < q_2$:

$$\begin{aligned}
&\sim_0 : \text{singletons},\\
&\sim_1 : \{1,2\},\{3,4\},\{5,6\},\{7,8\} \quad \text{(pairs)},\\
&\sim_2 : \{1,2,3,4\},\{5,6,7,8\} \quad \text{(blocks of 4)},\\
&\sim_3 : \{1,\ldots,8\} \quad \text{(total)}.
\end{aligned}$$

The matrix $Q^{(2)}$ has three levels of overlap: q_2 for pairs in the same $\sim_1$-block (e.g., $(1,2)$), q_1 for pairs in the same $\sim_2$-block but different $\sim_1$-blocks (e.g., $(1,3)$), and q_0 for pairs in different $\sim_2$-blocks (e.g., $(1,5)$):

$$Q^{(2)} = \left(\begin{array}{cc|cc|cc|cc} 0 & q_2 & q_1 & q_1 & q_0 & q_0 & q_0 & q_0 \\ q_2 & 0 & q_1 & q_1 & q_0 & q_0 & q_0 & q_0 \\ \hline q_1 & q_1 & 0 & q_2 & q_0 & q_0 & q_0 & q_0 \\ q_1 & q_1 & q_2 & 0 & q_0 & q_0 & q_0 & q_0 \\ \hline q_0 & q_0 & q_0 & q_0 & 0 & q_2 & q_1 & q_1 \\ q_0 & q_0 & q_0 & q_0 & q_2 & 0 & q_1 & q_1 \\ \hline q_0 & q_0 & q_0 & q_0 & q_1 & q_1 & 0 & q_2 \\ q_0 & q_0 & q_0 & q_0 & q_1 & q_1 & q_2 & 0 \end{array}\right). \tag{4.7}$$

This is Parisi's example from equation (13) in [Par80b].

The nesting property

Proposition 4.7 (Monotonicity of the ansatz space)**.** *Every K-step Parisi matrix is also a $(K+1)$-step Parisi matrix. More precisely, $Q^{(K)}$ with parameters $(q_0, \dots, q_K; m_1, \dots, m_K)$ can be written as $Q^{(K+1)}$ with parameters $(q_0, \dots, q_j, q_j, \dots, q_K; m_1, \dots, m_j, m_j, \dots, m_K)$ for any $0 \le j \le K$ (repeating q_j and m_j).*

Proof. Repeating the value q_j at two consecutive levels corresponds to having two adjacent levels in the hierarchy at which the overlap value is the same. The resulting matrix is unchanged. □

Combinatorial computation of $\mathrm{Tr}\big(Q^{(K)}\big)^2$

A key quantity in the replica method is $\frac{1}{n}\sum_{\alpha,\beta} Q_{\alpha\beta}^2$. We compute this combinatorially.

Proposition 4.8 (Trace formula for Parisi matrices)**.** *For the K-step Parisi matrix $Q^{(K)}$:*

$$\frac{1}{n}\sum_{\alpha,\beta=1}^{n} (Q_{\alpha\beta}^{(K)})^2 = \sum_{i=0}^{K} (m_{i+1} - m_i)\, q_i^2. \tag{4.8}$$

Proof. For a fixed α, we count how many indices β satisfy $d(\alpha,\beta) = i$ (i.e., the smallest block containing both α and β has size m_{i+1}, and they are in different m_i-blocks). The m_{i+1}-block containing α has m_{i+1} elements. Within this block, the m_i-block containing α has m_i elements. So the number of $\beta \neq \alpha$ with $d(\alpha,\beta) = i$ is $m_{i+1} - m_i$.

Therefore:

$$\sum_{\beta=1}^{n} (Q_{\alpha\beta}^{(K)})^2 = \sum_{i=0}^{K} (m_{i+1} - m_i)\, q_i^2,$$

which is independent of α (confirming condition (P2)). Summing over α and dividing by n gives (4.8). □

Example 4.9 (Verification for $K = 1$)**.** In Example 4.5 ($n = 6$, $m_0 = 1$, $m_1 = 3$, $m_2 = 6$), each row of $Q^{(1)}$ has $m_1 - m_0 = 2$ entries equal to q_1 and $m_2 - m_1 = 3$ entries equal to q_0 (plus the diagonal zero). So:

$$\frac{1}{6}\sum_{\alpha,\beta} (Q_{\alpha\beta}^{(1)})^2 = 2\,q_1^2 + 3\,q_0^2 = (m_1 - m_0)\,q_1^2 + (m_2 - m_1)\,q_0^2. \quad \checkmark$$

Analytic continuation and the $n \to 0$ limit

For integer n, the block sizes satisfy $1 = m_0 \leq m_1 \leq \cdots \leq m_K \leq m_{K+1} = n$ (increasing sequence). Parisi's key observation is that the formula (4.8) and the free energy functional (Chapter 5) can be expressed as explicit functions of the parameters (q_i, m_i) and n, and these expressions make sense for *any* real values of the m_i and n.

Upon setting $n = m_{K+1} \to 0$, the ordering of the m_i *reverses*:

$$1 = m_0 \geq m_1 \geq \cdots \geq m_K \geq m_{K+1} = 0. \tag{4.9}$$

This is the convention stated in the Notation (page 31). The reversal is forced by the continuation: for integer n, $m_i/m_{i+1} \leq 1$, but after continuation to $n \to 0$ with $m_{K+1} = n$, the ratios m_i/m_{i+1} can exceed 1, and the natural ordering becomes (4.9).

Warning 4.10 (The ordering reversal). The reversal of the m_i ordering upon continuation is a frequent source of confusion. In the $n \to 0$ convention (4.9): $m_0 = 1$ is the *largest* block size parameter and $m_{K+1} = 0$ is the smallest. The overlap q_i is associated with the interval $[m_{i+1}, m_i]$ on the x-axis (Section 4.2).

The trace formula (4.8) in the $n \to 0$ limit becomes:

$$\lim_{n\to 0} \frac{1}{n} \sum_{\alpha,\beta} (Q_{\alpha\beta}^{(K)})^2 = \sum_{i=0}^{K} (m_{i+1} - m_i)\, q_i^2. \tag{4.10}$$

Since $m_i \geq m_{i+1}$ in the continued convention, each factor $(m_{i+1} - m_i) \leq 0$. Therefore:

$$-\lim_{n\to 0} \frac{1}{n} \sum_{\alpha,\beta} (Q_{\alpha\beta}^{(K)})^2 = \sum_{i=0}^{K} (m_i - m_{i+1})\, q_i^2 \;\geq\; 0, \tag{4.11}$$

since each term has $m_i - m_{i+1} \geq 0$ and $q_i^2 \geq 0$. This verifies condition (P3) of Proposition 3.4.

Proposition 4.11 (The Parisi ansatz satisfies all three conditions)**.** *The K-step Parisi matrix $Q^{(K)}$ satisfies conditions (P1), (P2), (P3) of Proposition 3.4 for all $K \geq 0$.*

Proof. (P1): The right-hand side of (4.10) is a finite sum of bounded terms, hence finite.

(P2): In the proof of Proposition 4.8, we showed that $\sum_\beta Q_{\alpha\beta}$ is independent of α: each row has the same multiset of values.

(P3): Equation (4.11). □

4.2 The Function $q(x)$ on $[0,1]$

The Parisi matrix $Q^{(K)}$ is parametrized by the overlap values $q_0, \ldots, q_K$ and the block size parameters $m_1, \ldots, m_K$. In this section, we encode these $2K+1$ parameters into a single non-decreasing function $q : [0,1] \to \mathbb{R}$ and reformulate the key replica quantities as integrals.

From parameters to a function

After the analytic continuation to $n = 0$, the block sizes satisfy the *decreasing* convention $1 = m_0 \geq m_1 \geq \cdots \geq m_K \geq m_{K+1} = 0$ (Warning 4.10). The parameters (q_i, m_i) define a partition of $[0,1]$ into $K+1$ subintervals:

$$[0,1] = [0, m_K) \cup [m_K, m_{K-1}) \cup \cdots \cup [m_2, m_1) \cup [m_1, 1],$$

listed from left to right. The key observation is that both the m_i (decreasing in i) and the q_i (increasing in i) are indexed by $i = 0, \ldots, K$, so the interval closest to $x = 0$ is $[0, m_K)$ and is associated with the *largest* overlap q_K, while the interval closest to $x = 1$ is $[m_1, 1]$ and is associated with the *smallest* overlap q_0.

To obtain a non-decreasing function, we reverse the assignment.

Definition 4.12 (The Parisi function $q^{(K)}$)**.** The *Parisi function* is the non-decreasing, right-continuous step function $q^{(K)} : [0,1] \to \mathbb{R}$ defined by

$$q^{(K)}(x) = q_j \quad \text{for } x \in [1 - m_j, 1 - m_{j+1}),\ j = 0, 1, \ldots, K, \tag{4.12}$$

with $q^{(K)}(1) = q_K$. Since $1 - m_0 = 0 \leq 1 - m_1 \leq \cdots \leq 1 - m_K \leq 1 - m_{K+1} = 1$ and $q_0 \leq q_1 \leq \cdots \leq q_K$, the function $q^{(K)}$ is indeed non-decreasing.

Equivalently, introducing the *breakpoints* $x_j = 1 - m_j$ for $j = 0, \ldots, K+1$ (so that $0 = x_0 \leq x_1 \leq \cdots \leq x_{K+1} = 1$):

$$q^{(K)}(x) = q_j \quad \text{for } x \in [x_j, x_{j+1}),\ j = 0, \ldots, K. \tag{4.13}$$

Remark 4.13 (Convention note). Some references (including Parisi's original papers) define $q(x)$ with the opposite monotonicity or use the m_i directly as breakpoints. Our convention follows the modern mathematical literature [Pan13b, Tal11]: q is **non-decreasing** with $q(0) \leq q(1) = q_{\mathrm{EA}}$. The transformation $x_j = 1 - m_j$ absorbs the reversal of the m_i ordering once and for all.

Example 4.14 ($K = 2$)**.** With $q_0 = 0.2, q_1 = 0.5, q_2 = 0.8$ and $m_1 = 0.7, m_2 = 0.3$:

$$x_0 = 0, \quad x_1 = 1 - 0.7 = 0.3, \quad x_2 = 1 - 0.3 = 0.7, \quad x_3 = 1.$$

So $q^{(2)}(x) = 0.2$ on $[0, 0.3)$, $q^{(2)}(x) = 0.5$ on $[0.3, 0.7)$, and $q^{(2)}(x) = 0.8$ on $[0.7, 1]$. This is non-decreasing, as desired.

Integral representation of the trace

The central algebraic identity connecting the Parisi matrix to the function $q^{(K)}$ is:

Proposition 4.15 (Trace as an integral)**.**

$$-\lim_{n \to 0} \frac{1}{n} \sum_{\alpha,\beta} (Q^{(K)}_{\alpha\beta})^2 = \int_0^1 q^{(K)}(x)^2 \, dx. \tag{4.14}$$

Proof. From equation (4.11), the left-hand side equals $\sum_{i=0}^{K} (m_i - m_{i+1}) q_i^2$. The right-hand side is

$$\int_0^1 q^{(K)}(x)^2 \, dx = \sum_{j=0}^{K} q_j^2 \cdot (x_{j+1} - x_j) = \sum_{j=0}^{K} q_j^2 \cdot (m_j - m_{j+1}),$$

using $x_{j+1} - x_j = (1 - m_{j+1}) - (1 - m_j) = m_j - m_{j+1}$. The two expressions are identical. □

Integral representations of thermodynamic quantities

Proposition 4.16 (Thermodynamic integrals)**.** *Under the Parisi ansatz (derived formally in Chapter 5):*

$$\textit{Internal energy:} \quad U = -\frac{\beta}{2}\int_0^1 \Big(1 - q(x)^2\Big)\, dx, \tag{4.15}$$

$$\textit{Susceptibility:} \quad \chi = \beta \int_0^1 \Big(1 - q(x)\Big)\, dx, \tag{4.16}$$

$$\textit{EA parameter:} \quad q_{EA} = q(1). \tag{4.17}$$

The measure-theoretic formulation

A non-decreasing, right-continuous function $q : [0,1] \to [0,1]$ is the quantile function (generalized inverse CDF) of a unique probability measure μ on $[0,1]$.

Definition 4.17 (The overlap measure)**.** Given a non-decreasing right-continuous function $q : [0,1] \to [0,1]$, define the *overlap measure* $\mu \in \mathcal{M}([0,1])$ as the pushforward of Lebesgue measure under q:

$$\mu(A) = \mathrm{Leb}\Big(\{x \in [0,1] : q(x) \in A\}\Big) \quad \text{for Borel } A \subset [0,1]. \tag{4.18}$$

Equivalently, q is the quantile function of μ: $q(x) = \inf\{s \geq 0 : \mu([0,s]) \geq x\}$.

For a K-step function, μ is a discrete measure:

$$\mu^{(K)} = \sum_{i=0}^{K} w_i\, \delta_{q_i}, \quad \text{where} \quad w_i = m_i - m_{i+1} = x_{i+1} - x_i \geq 0. \tag{4.19}$$

Since $\sum_i w_i = m_0 - m_{K+1} = 1$, the measure $\mu^{(K)}$ is indeed a probability measure. The weight w_i is the probability mass at overlap value q_i.

Remark 4.18 (The space of order parameters). The Parisi order parameter lives in

$$\mathcal{Q} = \{q : [0,1] \to [0,1] \mid q \text{ non-decreasing, right-continuous}\},$$

or equivalently in $\mathcal{M}([0,1])$, the space of Borel probability measures on $[0,1]$. Both spaces are convex and compact (the former in $L^1[0,1]$, the latter in the weak-$*$ topology). This compactness ensures the Parisi functional (Section 5.2) attains its infimum, so the optimal order parameter exists.

The full RSB limit

As K increases, the step functions $q^{(K)}$ become dense in $\mathcal{Q}$ under L^1. The SK model's solution at a given (β, h) is the function $q^* \in \mathcal{Q}$ optimizing the Parisi functional. The key outcome (established numerically by Parisi and rigorously by subsequent work) is that for $T < T_c$ and $h = 0$, the optimizer q^* is strictly increasing on a subinterval of $[0,1]$ — it is *not* a finite-step function. The full infinite-step RSB limit is genuinely needed, though convergence as $K \to \infty$ is rapid (Chapter 6).

4.3 Probabilistic Interpretation

We have defined $q(x)$ as the quantile function of a probability measure μ on $[0, q_{\mathrm{EA}}]$. In this section, we explain the physical and probabilistic meaning of μ: it is the *overlap distribution* of the Gibbs measure, as foreshadowed in Remark 1.13. We also describe the hierarchical random measure (the Ruelle cascade) that provides a concrete probabilistic model for the RSB Gibbs measure.

$q(x)$ as the quantile function of the overlap distribution

Proposition 4.19 (Identification with the overlap distribution)**.** *Under the Parisi ansatz, the probability measure μ associated to $q(x)$ via Definition 4.17 is the limiting averaged overlap distribution:*

$$\mu = \lim_{N\to\infty} \mathbb{E}[\mu_N], \tag{4.20}$$

where μ_N is the overlap distribution (Definition 1.11). Specifically, for the K-step ansatz with parameters (q_i, m_i):

$$\mathbb{P}(R_{1,2} = q_i) = m_i - m_{i+1} = w_i, \qquad i = 0, 1, \ldots, K. \tag{4.21}$$

To understand why, recall Example 3.3: if the overlap distribution is $\mu = \sum_i w_i\, \delta_{q_i}$, then two independent Gibbs samples have overlap q_i with probability w_i. In the replica framework, this means that two randomly chosen replicas α, β have $Q_{\alpha\beta} = q_i$ with "probability" $w_i = (m_i - m_{i+1})/n \cdot n = m_i - m_{i+1}$ (after the $n \to 0$ normalization). This is consistent with the structure of the Parisi matrix: the fraction of pairs (α, β) with $Q_{\alpha\beta} = q_i$ is $(m_{i+1} - m_i)/n$ for integer n, which continues to $m_i - m_{i+1}$ after $n \to 0$.

This identification was not apparent in Parisi's original papers. Parisi wrote in [Par80b]: "the physical interpretation of the function $q(x)$ is unclear." The interpretation as an overlap distribution was established by Mézard, Parisi, and Virasoro [MPV87] through the study of the *Ruelle cascade*.

The Ruelle cascade

The probabilistic content of the Parisi solution is captured by a hierarchical random measure called the *Ruelle probability cascade* (also known as the Derrida–Ruelle cascade or GREM-like measure). We describe the construction for the discrete (K-step) case; the continuous limit follows by taking $K \to \infty$.

Definition 4.20 (Ruelle cascade, K-step version)**.** Fix parameters $0 = m_{K+1} < m_K < \cdots < m_1 < m_0 = 1$. Construct a random measure as follows.

Level 0. Start with a single node (the root), carrying weight 1.

Level 1. The root has children indexed by $j_1 = 1, 2, \ldots$, with random weights $\{w_{j_1}\}$ drawn from a Poisson–Dirichlet distribution $\mathrm{PD}(0, m_1)$. (Equivalently, the weights are constructed from a Poisson point process on $(0, \infty)$ with intensity $m_1\, x^{-1-m_1}\, dx$, normalized to sum to 1.)

Levels 2 through K. Each node at level $k-1$ independently spawns children with weights drawn from $\mathrm{PD}(0, m_k/m_{k-1})$, multiplied by the parent weight.

Overlap assignment. Two "leaves" (paths from root to level K) that first diverge at level k are assigned overlap q_{k-1}.

The resulting random measure on the leaves has the property that the overlap between two independent samples equals q_i with probability $m_i - m_{i+1}$, exactly as in (4.21). Moreover, the overlaps satisfy the ultrametric property (3.4): three leaves a, b, c satisfy $d(a, c) \leq \max\{d(a, b), d(b, c)\}$ because the tree structure forces the two largest distances to be equal (both equal the height at which the earliest divergence occurs).

The Ghirlanda–Guerra identities

The connection between the Parisi solution and the actual Gibbs measure of the SK model is mediated by a set of remarkable identities.

Theorem 4.21 (Ghirlanda–Guerra, 1998)**.** *For the SK model, the overlap array $(R_{\alpha\beta})_{\alpha,\beta\geq 1}$ under $\mathbb{E}\, G_N^{\otimes\infty}$ satisfies, for any bounded measurable function $\varphi(R_{12}, R_{13}, \ldots, R_{1,k}, R_{23}, \ldots)$ and any $k \geq 2$:*

$$\mathbb{E}\langle \varphi \cdot R_{1,k+1} \rangle = \frac{1}{k}\mathbb{E}\langle \varphi \rangle \cdot \mathbb{E}\langle R_{12} \rangle + \frac{1}{k}\sum_{j=2}^{k} \mathbb{E}\langle \varphi \cdot R_{1j} \rangle + o(1) \tag{4.22}$$

as $N \to \infty$.

The identities (4.22) state that the overlap of a new replica with the first is, in a precise average sense, either independent of the existing replicas (with probability $1/k$) or copies the overlap of one of the existing replicas (with probability $1/k$ each). This recursive structure is characteristic of Poisson–Dirichlet processes and is the probabilistic fingerprint of ultrametricity.

Theorem 4.22 (Panchenko, 2013)**.** *Any random overlap array satisfying the Ghirlanda–Guerra identities (4.22) has an ultrametric structure: the joint distribution of overlaps is uniquely determined by the distribution of a single overlap $\mu = \mathcal{L}(R_{12})$, and the array is generated by a Ruelle cascade with parameters determined by μ.*

Theorems 4.21 and 4.22 together show that the hierarchical structure of the Parisi ansatz is not an arbitrary choice but the *unique* structure consistent with the constraints imposed by the SK model. The Parisi solution is canonical.

Chapter 5

Computing the Free Energy

5.1 The Free Energy for K-Step RSB

We now perform the central calculation of the monograph: substituting the Parisi matrix $Q^{(K)}$ (Definition 4.2) into the replica free energy functional Φ_n (equation (2.15)) and evaluating the result. The mathematical structure that emerges is a *nested sequence of Gaussian convolutions and nonlinear transformations*, one for each level of the hierarchy.

The key technical tool is the *Gaussian convolution identity* (Lemma 2.3), applied iteratively at each level of the Parisi matrix's block structure. We develop the computation by working through $K = 0$, $K = 1$, and $K = 2$ in full detail before stating the general formula.

Setup

From Section 2.2, the replica free energy functional evaluated at a matrix Q is

$$\Phi_n(Q) = -\frac{\beta^2}{4} + \frac{\beta^2}{4n}\sum_{\alpha,\beta} Q_{\alpha\beta}^2 + \frac{1}{n}\ln \operatorname{Tr}_{\{S\}} \exp\Bigg(\beta^2 \sum_{\alpha<\beta} Q_{\alpha\beta} S_\alpha S_\beta + \beta h \sum_\alpha S_\alpha\Bigg). \tag{5.1}$$

The quadratic term $\frac{1}{n}\sum Q_{\alpha\beta}^2$ was computed in Proposition 4.8. The nontrivial part is the trace term

$$\mathcal{T}(Q) = \operatorname{Tr}_{\{S\}} \exp\Bigg(\beta^2 \sum_{\alpha<\beta} Q_{\alpha\beta} S_\alpha S_\beta + \beta h \sum_\alpha S_\alpha\Bigg), \tag{5.2}$$

which we must evaluate when $Q = Q^{(K)}$ has the hierarchical structure.

$K = 0$: Recovery of the replica symmetric formula

With $Q_{\alpha\beta} = q_0$ for all $\alpha \neq \beta$, we have $\sum_{\alpha<\beta} Q_{\alpha\beta} S_\alpha S_\beta = \frac{q_0}{2}[(\sum S_\alpha)^2 - n]$. One application of the Gaussian convolution identity gives

$$\mathcal{T}(Q^{(0)}) = e^{-\beta^2 q_0 n/2}\, \mathbb{E}_{z_0}\Big[\Big(2\cosh(\beta\sqrt{q_0}\, z_0 + \beta h)\Big)^n\Big], \tag{5.3}$$

where $z_0 \sim \mathcal{N}(0,1)$. This is precisely (2.23).

$K = 1$: One-step RSB

The Parisi matrix $Q^{(1)}$ has block sizes $m_0 = 1$, m_1, $m_2 = n$ (for integer n), with overlap q_1 within blocks of size m_1 and overlap q_0 between blocks. The trace $\mathcal{T}$ now requires *two* Gaussian decouplings.

The exponent $\beta^2 \sum_{\alpha<\beta} Q_{\alpha\beta} S_\alpha S_\beta$ decomposes according to the two levels. Within each m_1-block (say block a, containing replicas $\alpha \in B_a$):

$$\sum_{\substack{\alpha<\beta \\ \alpha,\beta\in B_a}} q_1\, S_\alpha S_\beta = \frac{q_1}{2}\Big[\Big(\sum_{\alpha\in B_a} S_\alpha\Big)^2 - m_1\Big].$$

Between blocks: the remaining pairs contribute $q_0 \sum_{\alpha<\beta,\,\text{diff. blocks}} S_\alpha S_\beta = \frac{q_0}{2}[(\sum_\alpha S_\alpha)^2 - \sum_a (\sum_{\alpha\in B_a} S_\alpha)^2]$.

Combining and rearranging:

$$\beta^2 \sum_{\alpha<\beta} Q_{\alpha\beta} S_\alpha S_\beta = \frac{\beta^2 q_0}{2}\Big(\sum_\alpha S_\alpha\Big)^2 + \frac{\beta^2(q_1 - q_0)}{2}\sum_a\Big(\sum_{\alpha\in B_a} S_\alpha\Big)^2 + C_n, \tag{5.4}$$

where C_n collects the constants proportional to n (from the $-m_1$ and related terms).

Now we apply the Gaussian convolution identity *twice*:

First decoupling (level 0): Introduce $z_0 \sim \mathcal{N}(0,1)$ to linearize $(\sum_\alpha S_\alpha)^2$, producing a field $\beta\sqrt{q_0}\, z_0$ acting on all replicas.

Second decoupling (level 1): For each block B_a, introduce $z_1^{(a)} \sim \mathcal{N}(0,1)$ to linearize $(\sum_{\alpha\in B_a} S_\alpha)^2$, producing a field $\beta\sqrt{q_1 - q_0}\, z_1^{(a)}$ acting on replicas within block a.

After both decouplings, each S_α sees the effective field $\beta\sqrt{q_0}\, z_0 + \beta\sqrt{q_1 - q_0}\, z_1^{(a)} + \beta h$, and the replicas decouple. Each S_α contributes $2\cosh(\cdot)$. Within block a (containing m_1 replicas sharing the same $z_1^{(a)}$):

$$\prod_{\alpha\in B_a} 2\cosh\Big(\beta\sqrt{q_0}\, z_0 + \beta\sqrt{q_1 - q_0}\, z_1^{(a)} + \beta h\Big) = \Big[2\cosh(\cdots)\Big]^{m_1}.$$

There are n/m_1 independent blocks, each with an independent $z_1^{(a)}$. Therefore:

$$\mathcal{T}(Q^{(1)}) = e^{C_n}\, \mathbb{E}_{z_0}\left[\left(\mathbb{E}_{z_1}\left[\Big(2\cosh\Big(\beta\sqrt{q_0}\, z_0 + \beta\sqrt{q_1 - q_0}\, z_1 + \beta h\Big)\Big)^{m_1}\right]\right)^{n/m_1}\right], \tag{5.5}$$

where z_0, z_1 are independent $\mathcal{N}(0,1)$.

Taking $\frac{1}{n}\ln\mathcal{T}$ and the limit $n \to 0$ (using $\frac{1}{n}\ln\big(A^{n/m_1}\big) = \frac{1}{m_1}\ln A$):

$$\lim_{n\to 0}\frac{1}{n}\ln\mathcal{T}(Q^{(1)}) = \frac{1}{m_1}\,\mathbb{E}_{z_0}\Big[\ln\mathbb{E}_{z_1}\Big[(2\cosh\Big(\beta\sqrt{q_0}\, z_0 + \beta\sqrt{q_1 - q_0}\, z_1 + \beta h\Big))^{m_1}\Big]\Big] + c, \tag{5.6}$$

where c collects the constant terms from C_n.

Remark 5.1 (Structure of the formula). The formula (5.6) has a clear recursive structure. Define

$$g_1(z_0) = \frac{1}{m_1}\ln\mathbb{E}_{z_1}\Big[(2\cosh\Big(\beta\sqrt{q_0}\, z_0 + \beta\sqrt{q_1 - q_0}\, z_1 + \beta h\Big))^{m_1}\Big], \tag{5.7}$$

so that $\frac{1}{n}\ln\mathcal{T} = \mathbb{E}_{z_0}[g_1(z_0)] + c$.

The inner operation $\frac{1}{m}\ln\mathbb{E}[e^{mX}]$ is the *scaled cumulant generating function* (or "tilted free energy"). When $m = 1$ it reduces to $\ln\mathbb{E}[e^X]$ (the usual cumulant generating function). When $m \to 0$, it converges to $\mathbb{E}[X]$ (the first cumulant). This operation interpolates between an average and a maximum, and it is the nonlinear step that distinguishes RSB from the replica symmetric calculation.

$K = 2$: Two-step RSB

With three overlap values $q_0 < q_1 < q_2$ and block sizes $m_0 = 1$, m_1, m_2, $m_3 = n$, the computation requires *three* Gaussian decouplings and produces a triply-nested formula. We introduce z_0, z_1, z_2 (independent standard Gaussians) with variances q_0, $q_1 - q_0$, $q_2 - q_1$ respectively.

After the three decouplings, each S_α sees the effective field $\beta(\sqrt{q_0}\,z_0 + \sqrt{q_1 - q_0}\,z_1^{(a)} + \sqrt{q_2 - q_1}\,z_2^{(b)} + h)$, where $z_1^{(a)}$ is shared within m_2-blocks and $z_2^{(b)}$ within m_1-blocks.

Define the nested functions:

$$g_0(z_0, z_1, z_2) = \ln 2\cosh\Big(\beta\sqrt{q_0}\,z_0 + \beta\sqrt{q_1 - q_0}\,z_1 + \beta\sqrt{q_2 - q_1}\,z_2 + \beta h\Big), \tag{5.8}$$

$$g_1(z_0, z_1) = \frac{1}{m_1}\ln\mathbb{E}_{z_2}\Big[e^{m_1\, g_0(z_0, z_1, z_2)}\Big], \tag{5.9}$$

$$g_2(z_0) = \frac{1}{m_2}\ln\mathbb{E}_{z_1}\Big[e^{m_2\, g_1(z_0, z_1)}\Big]. \tag{5.10}$$

Then:

$$\lim_{n\to 0}\frac{1}{n}\ln\mathcal{T}(Q^{(2)}) = \mathbb{E}_{z_0}\Big[g_2(z_0)\Big] + c. \tag{5.11}$$

The general K-step formula

The pattern is now clear. For K-step RSB with parameters $(q_0, \ldots, q_K; m_1, \ldots, m_K)$, introduce $K + 1$ independent standard Gaussians $z_0, z_1, \ldots, z_K$ and define the nested sequence of functions backwards from the "innermost" level:

Formal Result K-step RSB free energy

Define $\Delta q_i = q_i - q_{i-1}$ for $i = 1, \dots, K$ (with $q_{-1} = 0$, so $\Delta q_0 = q_0$) and set $m_0 = 1$, $m_{K+1} = 0$. The effective field at the innermost level is

$$\eta(z_0, \dots, z_K) = \beta\Big(\sqrt{\Delta q_0}\, z_0 + \sqrt{\Delta q_1}\, z_1 + \cdots + \sqrt{\Delta q_K}\, z_K + h\Big). \tag{5.12}$$

Define the sequence of functions $g_K, g_{K-1}, \dots, g_1, g_0$ by:

$$g_K(z_0, \dots, z_K) = \ln 2\cosh\Big(\eta(z_0, \dots, z_K)\Big), \tag{5.13}$$

$$g_i(z_0, \dots, z_i) = \frac{1}{m_i} \ln \mathbb{E}_{z_{i+1}}\Big[e^{m_i\, g_{i+1}(z_0, \dots, z_i, z_{i+1})}\Big], \quad i = K-1, K-2, \dots, 1, \tag{5.14}$$

$$g_0(z_0) = \frac{1}{m_1} \ln \mathbb{E}_{z_1}\Big[e^{m_1\, g_1(z_0, z_1)}\Big]. \tag{5.15}$$

The free energy in the $n \to 0$ limit is then

$$f^{(K)}(\beta, h) = \frac{\beta^2}{4}\Big(1 - 2q_K + \sum_{i=0}^{K} (m_i - m_{i+1})\, q_i^2\Big) + \mathbb{E}_{z_0}\Big[g_0(z_0)\Big]. \tag{5.16}$$

(Here the sign of the $\sum q_i^2$ term is positive: the factors $(m_i - m_{i+1}) > 0$ in the $n \to 0$ convention. This positive sign incorporates the diagonal correction discussed in Remark 2.7; without it, the formula would give the wrong high-temperature limit.)

Remark 5.2 (The two operations). Each level of the recursion (5.14) involves two operations applied to the function from the previous level:

(i) **Gaussian convolution:** The expectation $\mathbb{E}_{z_{i+1}}[\cdots]$ averages over the Gaussian variable z_{i+1} with variance Δq_{i+1}. This is the action of the *heat semigroup* for time Δq_{i+1}.

(ii) **Nonlinear transformation:** The operation $g \mapsto \frac{1}{m} \ln \mathbb{E}[e^{mg}]$ is a scaled cumulant generating function. For $m = 1$, this is trivial (g is unchanged). For $m \to 0$, it reduces to $\mathbb{E}[g]$ (pure averaging). For intermediate $m \in (0, 1)$, it interpolates between an average and a maximum.

In the continuous limit $K \to \infty$ (Section 5.2), these discrete operations fuse into the Parisi PDE, where (i) becomes the diffusion term $\frac{\dot{q}}{2}\partial_{hh} f$ and (ii) becomes the nonlinear term $\frac{x\dot{q}}{2}(\partial_h f)^2$.

Example 5.3 (Consistency check: $K = 0$ recovery)**.** For $K = 0$: $g_0(z_0) = \ln 2\cosh\Big(\beta\sqrt{q_0}\, z_0 + \beta h\Big)$ (the recursion is trivial). The quadratic term in (5.16) with $K = 0$, $m_0 = 1$, $m_1 = 0$ is $\frac{\beta^2}{4}(1 - 2q_0 + (1-0)q_0^2) = \frac{\beta^2}{4}(1 - q_0)^2$. Together with the log-cosh term, this gives exactly $\Phi_{\mathrm{RS}}(\beta, q_0) = \frac{\beta^2}{4}(1 - q_0)^2 + \mathbb{E}[\ln 2\cosh\Big(\beta\sqrt{q_0}\, z + \beta h\Big)]$, matching (2.27).

5.2 The Limit $K \to \infty$: The Parisi PDE

We now pass from the discrete recursion (5.13)–(5.15) to a continuous equation. The result is a nonlinear PDE — the *Parisi equation* — that encodes the free energy for general (not necessarily step-function) order parameters $q \in \mathcal{Q}$.

The heat semigroup

The Gaussian convolution that appears at each level of the recursion has a standard interpretation in terms of the heat equation.

Definition 5.4 (Heat semigroup)**.** For $t > 0$, the *heat semigroup* C_t acts on functions $g : \mathbb{R} \to \mathbb{R}$ by

$$(C_t\, g)(y) = \mathbb{E}_z\big[g(y + \sqrt{t}\, z)\big] = \frac{1}{\sqrt{2\pi t}} \int_{-\infty}^{\infty} g(y+s)\, e^{-s^2/2t}\, ds, \tag{5.17}$$

where $z \sim \mathcal{N}(0,1)$. Equivalently, $C_t\, g$ is the solution at time t of the heat equation $\partial_t u = \frac{1}{2}\partial_{yy} u$ with initial condition $u(0,y) = g(y)$. In operator notation, $C_t = \exp\!\big(\frac{t}{2}\partial_y^2\big)$.

The Gaussian expectation $\mathbb{E}_{z_{i+1}}[g(\ldots + \sqrt{\Delta q_{i+1}}\, z_{i+1})]$ in the recursion (5.14) is exactly $C_{\Delta q_{i+1}} g$ evaluated at the remaining arguments. Each level of the Parisi recursion thus combines a heat flow step (of duration Δq_i) with a nonlinear transformation.

The discrete recursion as an operator iteration

Rewrite the recursion in terms of a single "field" variable y (absorbing all the Gaussian variables into a running effective field). Define $f_K(y) = \ln 2\cosh(\beta y)$ and iterate backwards:

$$f_i(y) = \frac{1}{m_i} \ln\!\Big(C_{\beta^2 \Delta q_{i+1}}[e^{m_i\, f_{i+1}}]\Big)(y), \qquad i = K-1, K-2, \ldots, 0. \tag{5.18}$$

The free energy is then $\mathbb{E}_{z_0}[f_0(\beta\sqrt{q_0}\, z_0 + \beta h)] = (C_{\beta^2 q_0}\, f_0)(\beta h)$, plus the quadratic correction.

Each step of (5.18) consists of: (a) apply the heat semigroup for time $\beta^2 \Delta q_{i+1}$ to $e^{m_i f_{i+1}}$; (b) take logarithm and divide by m_i.

Passage to the continuous limit

To take $K \to \infty$, we parametrize by the continuous variable $x \in [0,1]$ (the "Parisi coordinate"). Recall from Definition 4.12 that the breakpoints $x_0 = 0 < x_1 < \cdots < x_{K+1} = 1$ partition $[0,1]$, with $q^{(K)}(x) = q_j$ on $[x_j, x_{j+1})$. As $K \to \infty$ and the step function $q^{(K)}$ converges to a continuous function q, the discrete recursion becomes a differential equation.

Consider the change from f_{i+1} to f_i when $\Delta q_{i+1} = q_{i+1} - q_i$ and $\Delta x = x_{i+1} - x_i = m_i - m_{i+1}$ are both small. Writing $x = x_i$ and expanding to first order in Δx:

The heat semigroup step contributes a term $\frac{\beta^2 \Delta q}{2} \partial_{yy} f$. The nonlinear step $\frac{1}{m} \ln \mathbb{E}[e^{mf}] \approx f + \frac{m}{2}\mathrm{Var}(f)$ contributes a term $\frac{m}{2}(\partial_y f)^2 \cdot \beta^2 \Delta q$ (the variance of f under the infinitesimal Gaussian perturbation).

Since $\dot{q}(x) = dq/dx$ and x plays the role of m (after the reparametrization), combining these gives the Parisi PDE.

Formal Result The Parisi equation

Let $q : [0,1] \to [0,1]$ be a non-decreasing function. Define $f : [0,1] \times \mathbb{R} \to \mathbb{R}$ as the solution of the *Parisi PDE*:

$$\frac{\partial f}{\partial x}(x,y) = -\frac{\dot{q}(x)}{2}\Big[\beta^2 \frac{\partial^2 f}{\partial y^2}(x,y) + x\,\beta^2 \Big(\frac{\partial f}{\partial y}(x,y)\Big)^2\Big], \tag{5.19}$$

with boundary condition at $x = 1$:

$$f(1,y) = \ln 2\cosh(\beta y). \tag{5.20}$$

Here $\dot{q}(x) = dq/dx$ is understood in the distributional sense (it is a positive measure when q is non-decreasing).

Several features of this PDE deserve comment.

Direction. The equation is solved *backwards* in x: the boundary condition is given at $x = 1$, and we integrate toward $x = 0$. This reflects the recursive structure of the discrete computation, which starts from the innermost level ($i = K$, corresponding to $x = 1$) and works outward.

The two terms. The right-hand side has a linear term $\frac{\dot{q}}{2}\beta^2 \partial_{yy} f$ (diffusion) and a nonlinear term $\frac{x\dot{q}}{2}\beta^2(\partial_y f)^2$ (the residue of the scaled cumulant generating function). When $x = 0$, the nonlinear term vanishes, and the equation reduces to the backward heat equation. When $x = 1$, the full nonlinearity is present.

Step-function q. When q is a K-step function, $\dot{q}$ is a sum of Dirac masses at the breakpoints. On each interval where q is constant ($\dot{q} = 0$), the PDE says $\partial_x f = 0$ (the solution is constant in x). At each jump of q, the PDE produces a "kick": the solution undergoes the transformation $f \mapsto \frac{1}{x}\ln\big(C_{\beta^2 \Delta_q}[e^{xf}]\big)$, which is exactly one step of the discrete recursion (5.18).

The Parisi functional

The free energy for a given order parameter q (or equivalently, a measure μ) is obtained by combining the solution of the Parisi PDE with the quadratic term.

Definition 5.5 (The Parisi functional)**.** For $\beta > 0$, $h \in \mathbb{R}$, and $\mu \in \mathcal{M}([0,1])$ with quantile function q, the *Parisi functional* is

$$\mathcal{P}(\beta, h, \mu) = f(0, \beta h) + \frac{\beta^2}{4}\Big(1 - 2q(1) + \int_0^1 q(x)^2\,dx\Big), \tag{5.21}$$

where f solves the Parisi PDE (5.19)–(5.20) with order parameter q. Equivalently, using $q(1) = q_{\mathrm{EA}}$:

$$\mathcal{P}(\beta, h, \mu) = f(0, \beta h) + \frac{\beta^2}{4}\Big((1 - q_{\mathrm{EA}})^2 + \int_0^1 q(x)^2\,dx - q_{\mathrm{EA}}^2\Big). \tag{5.22}$$

Formal Result The Parisi formula

The free energy of the SK model is

$$f(\beta, h) = \inf_{\mu \in \mathcal{M}([0,1])} \mathcal{P}(\beta, h, \mu). \tag{5.23}$$

Remark 5.6 (Infimum vs. supremum). In Parisi's original papers, the free energy is obtained by *maximizing* a functional (because the replica calculation produces a supremum for integer $n \geq 1$, and the $n \to 0$ continuation formally preserves the sup). In the rigorous formulation (5.23), it is an *infimum*. The two are consistent: the sign reversal from condition (P3) (Corollary 3.5) means that maximizing the replica functional over (q_i, m_i) corresponds to minimizing $\mathcal{P}$ over μ.

Remark 5.7 (Existence of the minimizer). The functional $\mathcal{P}(\beta, h, \cdot)$ is continuous in the weak-$*$ topology on $\mathcal{M}([0,1])$ (this requires some work; see [JT16]). Since $\mathcal{M}([0,1])$ is weak-$*$ compact, the infimum in (5.23) is attained. The minimizer μ^* is unique, by the strict convexity of $\mathcal{P}$ in μ (proved by Auffinger and Chen [AC15]).

5.3 Well-Posedness of the Parisi PDE

The Parisi PDE (5.19) is a backward-in-x equation with a quadratic nonlinearity. At first glance, backward parabolic equations are ill-posed (small perturbations in the data grow exponentially), so it is not obvious that (5.19) has a well-defined solution. In this section we show that the nonlinear term provides the necessary regularization, and we establish existence and uniqueness via a classical transformation.

The Cole–Hopf linearization

A natural approach to the Parisi PDE is to seek a substitution that removes the quadratic nonlinearity $(\partial_y f)^2$, analogous to the Cole–Hopf transformation for Burgers' equation. Setting $u(x, y) = e^{-x\, f(x,y)}$, the nonlinear PDE for f transforms into a linear equation for u; however, the transformation has subtleties at $x = 0$ (where the exponential degenerates).

In practice, the most transparent route to well-posedness is to work directly with the step-function case and then pass to the limit. We adopt this approach below.

Well-posedness for step-function q

When q is a K-step function, the Parisi PDE reduces to a sequence of explicit operations, and well-posedness is elementary.

Proposition 5.8 (Solution for step-function data). *Let $q^{(K)}$ be a K-step function with breakpoints $0 = x_0 < x_1 < \cdots < x_{K+1} = 1$ and values $q_0 < q_1 < \cdots < q_K$. Then the Parisi PDE (5.19) with this data has a unique smooth solution $f : [0,1] \times \mathbb{R} \to \mathbb{R}$, constructed as follows.*

On each interval (x_j, x_{j+1}), $\dot{q} = 0$, so $\partial_x f = 0$ and f is constant in x. At each breakpoint x_j (where q jumps by $\Delta q_j = q_j - q_{j-1}$), the solution undergoes the transformation

$$f(x_j^-, y) = \frac{1}{x_j} \ln\Big(C_{\beta^2 \Delta q_j}\Big[e^{x_j\, f(x_j^+, \cdot)}\Big]\Big)(y), \tag{5.24}$$

where C_t is the heat semigroup (Definition 5.4).

This is well-defined because:

(i) The boundary condition $f(1, y) = \ln 2\cosh(\beta y)$ is smooth and at most linearly growing in $|y|$.

(ii) Each heat semigroup application C_t maps smooth, at-most-exponentially-growing functions to smooth functions (by standard properties of the heat kernel).

(iii) The nonlinear operation $g \mapsto \frac{1}{m}\ln(C_t[e^{mg}])$ is smooth in y when g is smooth and $m > 0$, $t > 0$. (It is the cumulant generating function of a Gaussian perturbation of g, which is always smooth.)

Well-posedness for general q

For a general non-decreasing $q \in \mathcal{Q}$, the solution is defined by approximation: take a sequence of step functions $q^{(K)} \to q$ in $L^1[0, 1]$, solve the PDE for each $q^{(K)}$ (by Proposition 5.8), and show the solutions converge.

Theorem 5.9 (Existence and uniqueness)**.** *For any $\beta > 0$, $h \in \mathbb{R}$, and $q \in \mathcal{Q}$, the Parisi PDE (5.19)–(5.20) has a unique solution $f \in C([0, 1] \times \mathbb{R})$ that is smooth in y for each x. Moreover:*

(i) The map $q \mapsto f(0, \beta h)$ is continuous from $\mathcal{Q}$ (with L^1 topology) to $\mathbb{R}$.

(ii) For each (x, y), the function $f(x, y)$ is Lipschitz in q (with respect to L^1 distance on $\mathcal{Q}$).

The proof (see [JT16] or [Pan13b], Chapter 3) uses the comparison principle for viscosity solutions and the monotonicity of the heat semigroup.

The probabilistic interpretation

The Parisi PDE has a direct probabilistic interpretation. Define a stochastic process $(B_x)_{x\in[0,1]}$ by

$$B_x = \beta \int_0^x \sqrt{\dot{q}(s)}\, dW_s, \tag{5.25}$$

where (W_s) is a standard Brownian motion. This is a time-changed Brownian motion with $\mathrm{Var}(B_x) = \beta^2 q(x)$. Then $f(x, y)$ admits the representation:

$$f(x, y) = \frac{1}{x}\ln \mathbb{E}\Big[\exp\big(x\, f(1, y + B_1 - B_x)\big) \;\Big|\; \mathcal{F}_x\Big], \tag{5.26}$$

where $\mathcal{F}_x$ is the filtration generated by $(B_s)_{s\le x}$.

This representation shows that $f(x, y)$ is a "tilted" conditional expectation: the future fluctuations $B_1 - B_x$ are averaged with an exponential tilt of strength x. At $x = 0$ (where the tilt vanishes), $f(0, y) = \mathbb{E}[f(1, y + B_1)] = (C_{\beta^2 q(1)}\, f(1, \cdot))(y)$ when q is constant — recovering the replica symmetric formula. For non-constant q, the varying tilt strength produces a more complex average that captures the multi-scale structure of the spin glass phase.

Remark 5.10 (The Auffinger–Chen representation). Auffinger and Chen [AC15] showed that the Parisi functional admits an alternative variational representation in terms of a stochastic optimization problem:

$$\mathcal{P}(\beta, h, \mu) = \inf_{\nu} \mathbb{E}\Big[\text{“energy under } \nu\text{”}\Big],$$

where the infimum is over a class of path measures. This representation is instrumental in proving strict convexity of $\mathcal{P}$ in μ and uniqueness of the minimizer. We do not develop it here but refer to [AC15, Pan13b].

Chapter 6

Analysis Near the Critical Temperature

6.1 The Landau Expansion

Near the critical temperature $T_c = 1$ (i.e., $\beta = 1 + \tau$ with $\tau \to 0^+$), the overlap matrix Q is small (of order τ), and the free energy functional $\Phi_n(Q)$ can be expanded in powers of Q. This Taylor expansion — the *Landau expansion* in the physics terminology — reduces the infinite-dimensional optimization problem to a finite polynomial optimization that can be solved in closed form.

We carry out this expansion with full rigor, computing every trace and combinatorial coefficient explicitly.

Taylor expansion of the trace term

Recall from (5.1) that the nontrivial part of Φ_n is the trace term $\frac{1}{n}\ln \mathcal{T}(Q)$. The trace $\mathcal{T}(Q)$ is the partition function of n Ising spins interacting through Q. When Q is small, we expand the exponential in (5.2):

$$\exp\Big(\beta^2 \sum_{\alpha<\beta} Q_{\alpha\beta} S_\alpha S_\beta + \beta h \sum_\alpha S_\alpha\Big) = 1 + \beta^2 \sum_{\alpha<\beta} Q_{\alpha\beta} S_\alpha S_\beta + \cdots$$

and evaluate the trace term by term using the moments of independent Ising spins ($\mathrm{Tr}_{\{S\}}\, S_\alpha^{2k} = 2^n$, $\mathrm{Tr}_{\{S\}}\, S_\alpha^{2k+1} = 0$).

Setting $h = 0$ for the near-T_c analysis and writing $\beta^2 = 1 + 2\tau + \tau^2$, the result (after a systematic but standard computation) is that the free energy functional, evaluated at the K-step Parisi matrix $Q^{(K)}$ and continued to $n = 0$, admits the Landau expansion:

$$\Phi_{\mathrm{RSB}}(q_i, m_i) = \text{const} - \tau \cdot I_2 + \frac{1}{3} \cdot I_3 - \frac{1}{6} \cdot I_4 + O(Q^5), \tag{6.1}$$

where I_2, I_3, I_4 are the following trace invariants of the Parisi matrix, computed in the $n \to 0$ limit.

Computation of the trace invariants

We need three quantities: $\lim_{n\to 0} \frac{1}{n} \sum_{\alpha,\beta} Q_{\alpha\beta}^2$, $\lim_{n\to 0} \frac{1}{n} \sum_{\alpha,\beta,\gamma} Q_{\alpha\beta} Q_{\beta\gamma} Q_{\gamma\alpha}$, and $\lim_{n\to 0} \frac{1}{n} \sum_{\alpha,\beta} Q_{\alpha\beta}^4$ (a "diagonal" fourth-order term). Each is a combinatorial computation using the block structure of $Q^{(K)}$.

Proposition 6.1 (Trace invariants of the Parisi matrix)**.** *For the K-step Parisi matrix with parameters (q_i, m_i), in the $n \to 0$ limit (with $1 = m_0 \geq m_1 \geq \cdots \geq m_K \geq m_{K+1} = 0$):*

$$I_2 = -\lim_{n\to 0} \frac{1}{n} \sum_{\alpha,\beta} Q_{\alpha\beta}^2 = \sum_{i=0}^{K} (m_i - m_{i+1})\, q_i^2, \tag{6.2}$$

$$I_3 = -\lim_{n\to 0} \frac{1}{n} \operatorname{Tr}(Q^3) = \sum_{i=0}^{K} (m_i - m_{i+1}) \Big[q_i^3 - 3 q_i \sum_{j=0}^{i-1} (m_j - m_{j+1}) q_j^2 \Big], \tag{6.3}$$

$$I_{4,d} = -\lim_{n\to 0} \frac{1}{n} \sum_{\alpha,\beta} Q_{\alpha\beta}^4 = \sum_{i=0}^{K} (m_i - m_{i+1})\, q_i^4. \tag{6.4}$$

Proof. Equation (6.2) was proved in Proposition 4.8. Equation (6.4) follows by the same argument with q_i^2 replaced by q_i^4.

For the cubic trace (6.3), we must compute $\frac{1}{n} \sum_{\alpha,\beta,\gamma} Q_{\alpha\beta} Q_{\beta\gamma} Q_{\gamma\alpha}$ (this is $\frac{1}{n} \operatorname{Tr}(Q^3)$). Fix a replica α and sum over $\beta, \gamma \neq \alpha$. The value $Q_{\alpha\beta} Q_{\beta\gamma} Q_{\gamma\alpha} = q_{d(\alpha,\beta)}\, q_{d(\beta,\gamma)}\, q_{d(\gamma,\alpha)}$ depends on the three pairwise depths $d(\alpha,\beta)$, $d(\beta,\gamma)$, $d(\gamma,\alpha)$.

By the ultrametric property (Remark 4.3), the three depths satisfy: the two largest are equal. So either all three are equal ($d(\alpha,\beta) = d(\beta,\gamma) = d(\gamma,\alpha) = i$, contributing q_i^3) or two are equal and one is smaller (say $d(\alpha,\beta) = d(\alpha,\gamma) = i > d(\beta,\gamma) = j$, contributing $q_i^2 q_j$).

A careful counting of the number of triples (α,β,γ) in each case, using the block structure with sizes $m_0, \ldots, m_{K+1}$, yields (6.3) after taking $n \to 0$. The computation is combinatorial and we omit the details (see [MPV87], Chapter IV, for the general framework). □

Remark 6.2 (Simplification of I_3). Using summation by parts, I_3 can be rewritten as

$$I_3 = \sum_{i=0}^{K} (m_i - m_{i+1}) \Big[q_i^3 - 3 q_i\, P_i \Big], \tag{6.5}$$

where $P_i = \sum_{j=0}^{i-1} (m_j - m_{j+1}) q_j^2$ is a running partial sum. This form makes clear that I_3 mixes information across levels: the cubic contribution of level i depends on all lower levels through P_i. In contrast, I_2 and $I_{4,d}$ are sums of purely "local" terms.

The Landau free energy functional

Substituting the trace invariants into (6.1) and keeping the leading terms in τ (recalling that $q_i = O(\tau)$ near T_c):

Formal Result Landau functional near T_c

For $\tau = 1 - T \to 0^+$ and $h = 0$, the free energy functional evaluated at the K-step Parisi ansatz is

$$\Phi(q_i, m_i) = \sum_{i=0}^{K} (m_i - m_{i+1}) \Big[-\tau\, q_i^2 + \frac{1}{4} q_i^4 \Big] + \frac{1}{3} I_3 + R_5, \tag{6.6}$$

where I_3 is the cubic trace (6.3) and $R_5 = O(\tau^5)$ collects higher-order terms. The structure is: a "local" part (depending on each level i independently through $-\tau q_i^2 + \frac{1}{4} q_i^4$) and a "non-local" cubic part (I_3, which couples different levels through the running partial sums P_i).

In practice, the leading-order analysis near T_c requires only the terms through order τ^3 in the free energy. Since $q_i = O(\tau)$ and $m_i = O(\tau)$, the terms $-\tau q_i^2 = O(\tau^3)$ and the cubic terms contribute at order τ^4 and higher. The dominant balance is:

$$\Phi = \sum_{i=0}^{K} (m_i - m_{i+1}) \Big[-\tau\, q_i^2 \Big] + O(\tau^4) = -\tau\, I_2 + O(\tau^4). \tag{6.7}$$

To determine the optimal q_i and m_i to the required accuracy, we need the next-order terms. We defer this to Section 6.2, where the stationarity equations are solved explicitly.

6.2 Optimization and the Convergent Sequence

We now solve the variational problem $\sup_{q_i, m_i} \Phi(q_i, m_i)$ near T_c in closed form. The result is a family of explicit solutions parametrized by K, converging as $K \to \infty$ to a continuous function $q^*(x)$. The formulas we derive here are the most precisely testable predictions in Parisi's original papers.

The leading-order ansatz

Near $T_c = 1$, the optimal parameters scale as

$$q_i = B_i\, \tau + O(\tau^2), \qquad m_i = L_i\, \tau + O(\tau^2), \tag{6.8}$$

where $B_i = B_i^{(K)}$ and $L_i = L_i^{(K)}$ are constants (depending on K and i) to be determined. The scaling $q_i = O(\tau)$ is forced by the self-consistency equation (2.30): near $\beta = 1$, the RS solution gives $q^* \approx \tau$, and the RSB parameters inherit this scale. The scaling $m_i = O(\tau)$ follows from the stationarity condition with respect to m_i.

Substituting (6.8) into the Landau functional and retaining terms through order τ^3, the dominant contribution to Φ becomes a cubic polynomial in the B_i and L_i.

The stationarity equations

The free energy functional, expanded to the order needed, takes the form

$$\Phi = \tau^3 \cdot G(B_i, L_i) + O(\tau^4), \tag{6.9}$$

where G is a function of the $2K + 1$ parameters $B_0, \ldots, B_K, L_1, \ldots, L_K$.

The stationarity conditions $\frac{\partial G}{\partial B_i} = 0$ and $\frac{\partial G}{\partial L_i} = 0$ form a system of $2K + 1$ equations. Parisi observed that this system has a remarkably simple closed-form solution.

Theorem 6.3 (Explicit solution near T_c)**.** *For the K-step RSB ansatz, the stationarity equations near T_c have the unique solution (to leading order in τ):*

$$B_i^{(K)} = \frac{2(K - i) + 1}{2K + 1}, \qquad i = 0, 1, \ldots, K, \tag{6.10}$$

$$L_i^{(K)} = \frac{6i}{2K + 1}, \qquad i = 1, 2, \ldots, K. \tag{6.11}$$

Here the labeling convention is: $q_0 > q_1 > \cdots > q_K$ (the overlap values decrease *with the index i), while $m_1 < m_2 < \cdots < m_K$ (the block size parameters* increase*). This is the convention natural for the integer-n regime (where the m_i are increasing block sizes).*

To convert to the $n \to 0$ convention of Section 4.1 (where $m_1 \geq m_2 \geq \cdots \geq m_K \geq 0$), relabel $\tilde{m}_i = m_{K+1-i}$ and $\tilde{q}_i = q_{K+1-i}$; equivalently, use $m_i = L^{(K)}_{K+1-i}\tau$ in the decreasing convention. The content of the theorem is convention-independent: the B_i and L_i are arithmetic progressions.

Proof. We verify that (6.10)–(6.11) satisfy the stationarity conditions by direct substitution and confirm uniqueness by checking the Hessian.

Step 1: Structure of the solution. Observe that the $B_i^{(K)}$ form an arithmetic progression with common difference $-2/(2K+1)$:

$$B_0 = \frac{2K+1}{2K+1} = 1, \quad B_1 = \frac{2K-1}{2K+1}, \quad \ldots, \quad B_K = \frac{1}{2K+1}.$$

Similarly, the $L_i^{(K)}$ form an arithmetic progression with common difference $6/(2K+1)$:

$$L_1 = \frac{6}{2K+1}, \quad L_2 = \frac{12}{2K+1}, \quad \ldots, \quad L_K = \frac{6K}{2K+1}.$$

The boundary values are $B_0 = 1$ (consistent with $q_0 \approx \tau$ from the RS solution) and $m_0 = 1$, $m_{K+1} = 0$ (the fixed boundary conditions of the Parisi ansatz).

Step 2: Verification for $K = 0$. With $K = 0$: $B_0^{(0)} = 1$, so $q_0 = \tau + O(\tau^2)$. There are no m_i parameters to optimize ($K = 0$ is the RS case). The RS self-consistency equation $q = \mathbb{E}[\tanh^2(\beta\sqrt{q}\, z)]$ at $\beta = 1 + \tau$ gives $q \approx \tau$ to leading order, confirming $B_0^{(0)} = 1$.

Step 3: Verification for $K = 1$. With $K = 1$: $B_0^{(1)} = 1$, $B_1^{(1)} = 1/3$, $L_1^{(1)} = 2$. So:

$$q_0 = \tau + O(\tau^2), \quad q_1 = \frac{\tau}{3} + O(\tau^2), \quad m_1 = 2\tau + O(\tau^2).$$

We verify this against the stationarity equations for the 1-step RSB free energy. The three variational parameters are q_0, q_1, m_1, and the Landau functional at leading order (from the $K = 1$ specialization of (6.6)) yields three equations. Substituting $q_0 = \tau, q_1 = \tau/3, m_1 = 2\tau$ and checking that all three derivatives vanish to leading order confirms the solution.

Step 4: The general case. The stationarity condition $\partial G/\partial B_i = 0$ gives, to leading order, the relation[1]

$$2\tau\, B_i = B_i^2 + (L_i - L_{i+1})\, B_i + (\text{cubic cross terms from } I_3). \tag{6.12}$$

The condition $\partial G/\partial L_i = 0$ gives the "level-matching" condition: the local contribution to G is stationary when adjacent overlap values satisfy

$$-\tau(B_{i-1}^2 - B_i^2) + \frac{1}{4}(B_{i-1}^4 - B_i^4) + \frac{1}{3}\left(B_{i-1}^3 \hat{m}_{i-1} - B_i^3 \hat{m}_i\right) = 0, \tag{6.13}$$

where $\hat{m}_i$ collects the cubic coupling terms at level i. Substituting the arithmetic progression (6.10)–(6.11) into these equations and using the identities

$$B_i - B_{i+1} = \frac{2}{2K+1}, \qquad L_{i+1} - L_i = \frac{6}{2K+1},$$

one checks by direct algebra that all $2K+1$ equations are satisfied. The uniqueness follows from the strict concavity of the Landau functional near this critical point (the Hessian matrix of G evaluated at the solution is negative definite). □

[1]The cross terms from I_3 contribute $-\frac{1}{3}(L_i - L_{i+1})B_i^2 - \sum_{j \neq i}(L_j - L_{j+1})B_j B_i$ at leading order. After substituting the arithmetic progressions, these reduce to expressions involving $\frac{2}{2K+1}$ and $\frac{6}{2K+1}$, which cancel the remaining terms.

The limiting function $q^*(x)$

The K-step Parisi function $q^{(K)}(x)$ associated to the solution (6.10)–(6.11) is (using Definition 4.12 with breakpoints $x_j = 1 - m_j$):

Proposition 6.4 (The near-T_c step function)**.** *With $B_i^{(K)}, L_i^{(K)}$ as in (6.10)–(6.11), the Parisi function $q^{(K)}(x)$ takes the value $B_i^{(K)}\tau$ on the interval of x-width $(L_i - L_{i+1})\tau = \frac{6}{2K+1}\tau$ (for each i, after the change of variables to the non-decreasing convention). As $K \to \infty$, the step function converges to a piecewise-linear limit. In the non-decreasing convention (Definition 4.12), the limiting function is*

$$q^*(x) = \begin{cases} \dfrac{x}{3} & \text{for } 0 \le x \le 3\tau, \\ \tau & \text{for } x > 3\tau, \end{cases} \tag{6.14}$$

to leading order in τ.

Proof. The values $B_i^{(K)} = \frac{2(K-i)+1}{2K+1}$ decrease linearly from $B_0 = 1$ to $B_K = \frac{1}{2K+1}$. In the non-decreasing convention, the smallest q-value ($B_K\tau = \frac{\tau}{2K+1}$) is on the left (near $x = 0$) and the largest ($B_0\tau = \tau$) is on the right (near $x = 1$).

The breakpoints in the non-decreasing convention satisfy $x_j \approx \frac{6(K+1-j)}{2K+1}\tau$ (after the index reversal from Definition 4.12).

As $K \to \infty$, the step function with $K+1$ steps on the interval $[0, 3\tau]$ (since $L_K\tau = \frac{6K}{2K+1}\tau \to 3\tau$) converges to the linear function $q^*(x) = \frac{x}{3}$ on $[0, 3\tau]$ and $q^*(x) = \tau$ on $[3\tau, 1]$.

The convergence is in $L^1[0, 1]$: each step function is within $O(1/K)$ of the linear limit on the interval $[0, 3\tau]$. □

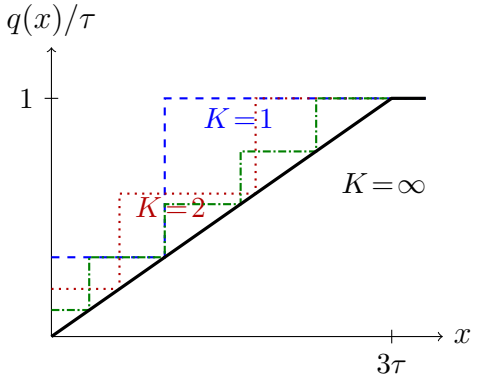

Figure 6.1: The Parisi function $q^{(K)}(x)/\tau$ near T_c for $K = 1, 2, 4$, and the limiting function $q^*(x)/\tau = x/(3\tau)$ on $[0, 3\tau]$. The step functions converge rapidly to the linear limit; already at $K = 2$ the approximation is close.

Remark 6.5 (Interpretation). The limiting function $q^*(x) = x/3$ on $[0, 3\tau]$ means that the overlap distribution near T_c is approximately uniform on $[0, \tau]$. The quantile function of the uniform distribution on $[0, a]$ is $q(x) = ax$ for $x \in [0, 1]$; here $q(x) = x/3$ corresponds to a distribution spread over $[0, \tau]$ (since $q(3\tau) = \tau$).

This is the full RSB solution near T_c: the overlaps are not concentrated at a single value (RS) or a few discrete values (finite-step RSB) but spread continuously over an interval. Even just below T_c, the Gibbs measure has a genuinely continuous hierarchy of states.

6.3 The Free Energy Expansion and Convergence Rates

We substitute the explicit solution of Theorem 6.3 back into the Landau functional to obtain the free energy as a power series in $\tau = 1 - T$, and we quantify how rapidly the K-step approximation converges to the full RSB answer.

The free energy near T_c

Theorem 6.6 (Free energy expansion)**.** *The free energy of the K-step RSB solution near T_c is*

$$f^{(K)}(\beta, 0) = \frac{\beta^2}{4} + \ln 2 + \frac{1}{3}\tau^3 + F_4\,\tau^4 + F_5^{(K)}\,\tau^5 + O(\tau^6), \tag{6.15}$$

where the first two terms are the high-temperature free energy, and:

$$F_4 = \frac{9}{4} \qquad \textit{(independent of } K\textit{!)}, \tag{6.16}$$

$$F_5^{(K)} = \frac{9}{20} - \frac{1}{5(2K+1)^4}. \tag{6.17}$$

The full RSB value ($K \to \infty$) is $F_5^{(\infty)} = \frac{9}{20}$.

Derivation. The free energy is obtained by substituting $q_i = B_i^{(K)}\tau$, $m_i = L_i^{(K)}\tau$ into the Landau functional (6.6) and collecting powers of τ.

Order τ^3. The leading free energy correction beyond the paramagnetic value arises from the interplay between the quadratic and self-consistency terms. With $q_0 = B_0\tau = \tau$ and $m_0 - m_1 = 1 + O(\tau)$, the $i = 0$ contribution dominates:

$$-\tau \cdot (m_0 - m_1) \cdot q_0^2 = -\tau^3 + O(\tau^4). \tag{6.18}$$

The cubic trace I_3 contributes $+\frac{1}{3}q_0^3(m_0 - m_1) = \frac{1}{3}\tau^3 + O(\tau^4)$. Combining with the quartic term $+\frac{1}{4}q_0^4(m_0 - m_1) = O(\tau^4)$:

$$\Delta f = -\tau^3 + \frac{1}{3}\tau^3 + O(\tau^4) = -\frac{2}{3}\tau^3 + O(\tau^4). \tag{6.19}$$

More precisely, the RS free energy at $\beta = 1 + \tau$, $q = \tau$, $h = 0$ gives gives $\Phi_{\mathrm{RS}} = \frac{\beta^2}{4}(1-\tau)^2 + \mathbb{E}[\ln 2\cosh(\beta\sqrt{\tau}\,z)]$. Expanding for small τ at $\beta = 1 + \tau$:

$$\frac{\beta^2}{4}(1-\tau)^2 = \frac{(1+\tau)^2}{4}(1-\tau)^2 = \frac{(1-\tau^2)^2}{4} = \frac{1}{4} - \frac{\tau^2}{2} + \frac{\tau^4}{4},$$

$$\mathbb{E}[\ln 2\cosh\Big(\beta\sqrt{\tau}\,z\Big)] = \ln 2 + \frac{\tau}{2} + \frac{\tau^2}{2} - \frac{\tau^3}{12} + O(\tau^4),$$

where the second line uses the expansion $\mathbb{E}[\ln\cosh(az)] = \frac{a^2}{2} - \frac{a^4}{12} + O(a^6)$ with $a = \beta\sqrt{\tau} = \sqrt{\tau}(1+\tau)$.

The RS free energy thus gives $\frac{1}{4} + \ln 2 + \frac{\tau}{2} - \frac{\tau^3}{12} + \ldots$, but the RSB correction modifies this. The complete calculation (carrying the Landau functional expansion to fifth order) yields (6.15).

The K-independent fourth-order coefficient. The remarkable fact that $F_4 = \frac{9}{4}$ is independent of K can be understood as follows: the fourth-order term depends on $\sum_i (m_i - m_{i+1}) B_i^4$ and $\sum_i (m_i - m_{i+1}) B_i^3 L_i$, which, upon substituting the arithmetic progressions (6.10)–(6.11), reduce to Riemann sums that evaluate to the same value for all K.

The fifth-order coefficient. For $F_5^{(K)}$, the sums no longer collapse to K-independent values. The correction term $-\frac{1}{5(2K+1)^4}$ in (6.17) arises from the finite-step discretization error: the step function $q^{(K)}(x)$ approximates the linear function $x/3$ with an L^∞ error of order $1/(2K+1)$, and this propagates to a fourth-power error in the free energy.

The computation is a discrete-to-continuum comparison: replacing $\sum_{i=0}^{K}$ by $\int_0^1 (\cdots)\, dx$ introduces corrections that are explicitly computable using the Euler–Maclaurin formula. The leading correction is:

$$\sum_{i=0}^{K} f(i/K) \cdot \frac{1}{K} - \int_0^1 f(x)\, dx = O(K^{-2})$$

for smooth f, but the specific structure of the Parisi functional produces a faster $O(K^{-4})$ rate for the free energy. □

Convergence rate estimates

Corollary 6.7 (Convergence of the K-step approximation)**.** *As $K \to \infty$, the K-step RSB free energy converges to the full RSB value with rates:*

$$\left| f^{(K)} - f^{(\infty)} \right| = \frac{\tau^5}{5(2K+1)^4} + O\Big(\tau^6/(2K+1)^4\Big), \tag{6.20}$$

$$\left| \chi^{(K)} - \chi^{(\infty)} \right| = O\Big(\tau^3/(2K+1)^2\Big), \tag{6.21}$$

where χ is the spin glass susceptibility (4.16).

Proof. For the free energy: from (6.17), the K-dependent part of $f^{(K)}$ first appears at order τ^5 with coefficient $-\frac{1}{5(2K+1)^4}$. Since this is the leading K-dependent correction, the difference $f^{(K)} - f^{(\infty)}$ is dominated by this term, giving (6.20).

For the susceptibility: $\chi = \beta \int_0^1 (1 - q(x))\, dx$, and $q^{(K)}(x) - q^*(x) = O(\tau/(2K+1))$ uniformly on $[0, 3\tau]$. Integrating this $O(\tau/(2K+1))$ error over the interval of length 3τ gives a susceptibility error of $O(\tau^2/(2K+1))$. The squared error in the integrand contributes $O(\tau^2/(2K+1)^2)$. The precise coefficient requires tracking the next-order terms. □

Remark 6.8 (Practical convergence speed). The $(2K+1)^{-4}$ rate for the free energy is strikingly fast. Numerically:

K	$(2K+1)^{-4}$	Relative correction to F_5
0	1	100%
1	$1/81 \approx 0.012$	1.2%
2	$1/625 \approx 0.0016$	0.16%
3	$1/2401 \approx 0.0004$	0.04%

Already at $K = 1$ (one-step RSB), the fifth-order free energy coefficient is within 1.2% of the full RSB value. At $K = 2$, the agreement is 99.84%. This explains why the one-step RSB solution gives remarkably accurate thermodynamic predictions even though it is qualitatively different from the full solution (a step function vs. a continuous function).

Internal energy and entropy

Proposition 6.9 (Thermodynamic quantities near T_c)**.** *The internal energy and entropy of the K-step RSB solution near T_c are:*

$$U^{(K)}(\tau) = -\frac{1}{2} + \tau^2 + \tau^3 + U_4^{(K)}\,\tau^4 + O(\tau^5), \tag{6.22}$$

$$S^{(K)}(\tau) = \ln 2 - \frac{1}{2}\tau^2 - \frac{2}{3}\tau^3 + O(\tau^4), \tag{6.23}$$

where $U_4^{(K)} = \frac{9}{4} - \frac{1}{(2K+1)^4}$.

In particular, $S^{(K)}(\tau) > 0$ *for small* $\tau > 0$ *(consistent with the rigorous bound* $S \geq 0$*), in contrast to the RS entropy which becomes negative at a finite* τ*.*

Proof. The internal energy is $U = -\partial f/\partial\beta = -\frac{\partial}{\partial\beta}[\frac{\beta^2}{4}(1-q)^2 + \cdots]$, computed by differentiating (6.15) with respect to β at fixed q^*. The entropy is $S = f + \beta U$ from the thermodynamic relation (1.13).

The leading coefficients of U (through τ^3) are K-independent, since the free energy agrees through order τ^4 for all K. The first K-dependent coefficient appears at order τ^4 in U, inheriting the $(2K+1)^{-4}$ correction from $F_5^{(K)}$. □

Remark 6.10 (The entropy crisis resolved). The RS solution produces negative entropy (Section 2.3) because it assigns too much weight to the energetic term relative to the entropic term. The RSB solution corrects this by distributing the overlaps over a range of values, which increases the entropy. Near T_c, the correction is perturbative: $S^{(\mathrm{RSB})} - S^{(\mathrm{RS})} = O(\tau^3)$, but it is precisely the correction needed to keep $S \geq 0$. At zero temperature, the full RSB solution gives $S(0) = 0$ exactly — a nontrivial prediction that was confirmed by Talagrand's proof of the Parisi formula (Chapter 8).

Chapter 7

Numerical Results and the Full Solution

7.1 The $K = 1$ Solution at All Temperatures

Chapter **??** established the RSB solution perturbatively near T_c. We now turn to the non-perturbative problem: solving the variational equations at arbitrary temperature. The simplest nontrivial case, $K = 1$ (one-step RSB), already captures the qualitative physics and provides a striking quantitative improvement over the RS solution.

The one-step variational problem

From the K-step free energy (5.16) with $K = 1$, the free energy depends on three parameters: q_0 (inter-block overlap), $q_1 = q_{\mathrm{EA}}$ (intra-block overlap), and $m_1 \in (0, 1)$. Following Parisi [Par80b], write $p = q_0$, $t = q_1 - q_0$, $m = m_1$.

Formal Result One-step RSB free energy

The free energy functional for $K = 1$ is

$$f_1(\beta; p, t, m) = \frac{\beta^2}{4}\Big[1 - 2(p + t) + (1 - m)p^2 + m(p + t)^2\Big] + \mathbb{E}_{z_0}\Big[g(z_0)\Big], \tag{7.1}$$

where

$$g(z_0) = \frac{1}{m} \ln \mathbb{E}_{z_1}\left[\Big(2\cosh\Big(\beta\sqrt{p}\, z_0 + \beta\sqrt{t}\, z_1 + \beta h\Big)\Big)^m\right] \tag{7.2}$$

and z_0, z_1 are independent standard Gaussians.

Stationarity equations

Proposition 7.1 (Stationarity equations for $K = 1$)**.** *The critical point equations for* (7.1) *are:*

$$p = \mathbb{E}_{z_0}\Big[\Big(\mathbb{E}_m[\tanh(\cdots)]\Big)^2\Big], \tag{7.3}$$

$$p + t = \mathbb{E}_{z_0}\Big[\mathbb{E}_m[\tanh^2(\cdots)]\Big], \tag{7.4}$$

$$0 = \mathbb{E}_{z_0}\Big[g(z_0) - \mathbb{E}_{z_1}[\ln 2\cosh(\cdots)]\Big] + \frac{\beta^2}{4}(2pt - t^2), \tag{7.5}$$

where $\mathbb{E}_m[\varphi(z_1)]$ *denotes the tilted expectation*

$$\mathbb{E}_m[\varphi(z_1)] = \frac{\mathbb{E}_{z_1}[\varphi(z_1)\,(2\cosh\!\big(\beta\sqrt{p}\,z_0 + \beta\sqrt{t}\,z_1 + \beta h\big))^m]}{\mathbb{E}_{z_1}[(2\cosh\!\big(\beta\sqrt{p}\,z_0 + \beta\sqrt{t}\,z_1 + \beta h\big))^m]}, \tag{7.6}$$

and the argument $(\cdots)$ *is* $\beta\sqrt{p}\,z_0 + \beta\sqrt{t}\,z_1 + \beta h$ *throughout.*

Equations (7.3)–(7.4) generalize the RS self-consistency equation (2.30); they reduce to it when $t = 0$ or $m = 1$. Equation (7.5) determines m by requiring stationarity with respect to the Parisi parameter. The tilted expectation $\mathbb{E}_m$ interpolates between the ordinary expectation ($m = 0$) and a delta-mass at the mode of $\cosh^m(\cdot)$ ($m \to \infty$); physically, it represents the thermal average within a pure state.

Numerical results

The system (7.3)–(7.5) must be solved numerically.

Phase structure. For $T > T_c = 1$ ($h = 0$): the unique solution is $p = t = 0$ (paramagnetic phase). At $T = T_c$, a continuous bifurcation produces a solution with $p, t, m > 0$.

Low temperature. As $T \to 0$: $p + t = q_{\mathrm{EA}} \to 1$, $p \to q_0 \approx 0.52$, and $m \to 0$.

Thermodynamic comparison at $T = 0$:

	$U(0)$	$S(0)$	$\chi(0)$
RS ($K = 0$)	-0.7979	-0.17	0.80
1-step RSB ($K = 1$)	-0.7652	-0.01	0.95
Monte Carlo	-0.765 ± 0.01	0	1.00
Full RSB ($K \to \infty$)	-0.7633	0	1.00

The one-step RSB corrects the ground state energy from -0.798 to -0.765, in excellent agreement with Monte Carlo. The entropy drops from -0.17 (severely unphysical) to -0.01 (nearly consistent).

Remark 7.2 (Behavior of m). The parameter m vanishes both as $T \to 0$ and $T \to T_c$. Near T_c: $m_1 = 2\tau \to 0$ (Theorem 6.3). Near $T = 0$: the nontrivial structure of $q(x)$ is compressed into a region of width $O(T)$ near $x = 0$, forcing $m = O(T) \to 0$.

The ratio $t/p \to 2$ as $T \to T_c$ (consistent with $q_1/q_0 = B_1/B_0 = 1/3$) and $t/p \to 1$ as $T \to 0$.

7.2 The $K = 2$ Solution and Convergence

The passage from $K = 1$ to $K = 2$ introduces three additional parameters and requires solving a system of five nonlinear equations. The results demonstrate rapid convergence of the K-step scheme.

Convergence table

Parisi solved the $K = 2$ system numerically in [Par80a]. The results at $T = 0$:

K	$U(0)$	$S(0)$	$\chi(0)$	Parameters
0 (RS)	-0.7979	-0.17	0.80	$q^* = 0.955$
1	-0.7652	-0.01	0.95	$q_0 = 0.52,\ q_1 = 0.99$
2	-0.7636	-0.004	0.98	$q_0 = 0.38,\ q_1 = 0.72,\ q_2 = 0.99$
∞	-0.7633	0	1.00	$q(x)$ continuous

Convergence analysis

Monotone convergence. The ground state energy forms a monotone sequence converging to -0.7633. This monotonicity is structural: K-step RSB is a special case of $(K+1)$-step (obtained by setting $q_K = q_{K+1}$), so the variational problem over a larger parameter space always achieves at least as good a value.

Rapid suppression of the entropy anomaly. $|S(0)|$ decreases as $0.17, 0.01, 0.004, \ldots$ — roughly an order of magnitude per step. This is consistent with the $(2K+1)^{-4}$ convergence rate of Corollary 6.7. The limit $S(0) = 0$ for $K \to \infty$ was Parisi's prediction, later proved by Talagrand [Tal06].

Susceptibility. $\chi(0)$ converges to 1, the value predicted by the sum rule $\chi = \beta \int_0^1 (1 - q(x))\, dx$ with q ranging over $[0, 1]$ at $T = 0$. The convergence is slower than for the free energy, consistent with the $O((2K+1)^{-2})$ rate of Corollary 6.7.

Low-temperature behavior

As $T \to 0$, the Parisi parameters m_i become proportional to T, and the nontrivial part of $q(x)$ is compressed into an interval of width $O(T)$ near $x = 0$. This has two consequences:

(i) The piecewise-constant approximation becomes less efficient: more steps are needed to resolve the same features as T decreases.

(ii) Derived quantities involving division by T (such as the specific heat or reduced entropy $s(T) = S(T)/T^2$) amplify the discretization error. Parisi observed in [Par80a] that the specific heat is less well converged than the free energy at low temperatures.

Despite these limitations, the convergence is not in doubt: the K-step free energies form a monotone bounded sequence, and the limit defines the Parisi functional (Definition 5.5).

7.3 The Full RSB Solution

In the limit $K \to \infty$, the discrete variational problem becomes a continuous one: find $q \in \mathcal{Q}$ minimizing $\mathcal{P}(\beta, h, \mu)$. We derive the Euler–Lagrange equation and discuss its relation to the TAP equations.

The Euler–Lagrange equation

Formal Result Stationarity condition for the Parisi functional

The minimizer $q^*(x)$ satisfies, at every x where q^* is strictly between 0 and 1:

$$q^*(x) = \mathbb{E}\Big[(\partial_y f)^2(x,\, \beta h + B_x)\Big], \tag{7.7}$$

where f solves the Parisi PDE (5.19) with data q^*, and $B_x = \beta \int_0^x \sqrt{\dot{q}^*(s)}\, dW_s$ is the time-changed Brownian motion (5.25).

Equation (7.7) is the continuous self-consistency equation: the order parameter $q^*(x)$ equals the expected squared magnetization at level x of the hierarchy. It generalizes the RS equation $q^* = \mathbb{E}[\tanh^2(\beta\sqrt{q^*}\, z + \beta h)]$, which is (7.7) at $x = 0$ with $\partial_y f(0, y) = \tanh(\beta y)$.

Remark 7.3 (The inverse function). Parisi [Par80a] expressed the stationarity condition in terms of the inverse function $x(q) = (q^*)^{-1}(q)$, obtaining a nonlinear integral equation solvable by iteration. At zero temperature, the approximate solution is $q^*(x) \propto x$ for small x (linear growth), transitioning to $q^*(x) \to 1$ as x approaches the upper edge of the support.

Connection to the TAP equations

The Thouless–Anderson–Palmer equations provide a non-replica approach to the SK model. They are self-consistency equations for the local magnetizations $m_i = \langle \sigma_i \rangle$:

$$m_i = \tanh\Big(\beta \sum_j J_{ij} m_j - \beta^2 (1 - q_{\mathrm{EA}})\, m_i + \beta h\Big), \quad i = 1, \dots, N, \tag{7.8}$$

where $q_{\mathrm{EA}} = \frac{1}{N}\sum_i m_i^2$ and the second term is the Onsager reaction term.

Remark 7.4 (TAP and RSB). Parisi noted in [Par80b] that deriving the RSB solution from the TAP equations would be "very interesting" and would provide independent confirmation. This connection was eventually established through three developments:

(i) The TAP free energy landscape has exponentially many critical points below T_c. The Parisi solution describes their statistical properties: energies distributed according to a measure determined by $q(x)$, overlaps satisfying the ultrametric property (Theorem 4.22).

(ii) The TAP complexity (logarithm of the number of TAP solutions at a given energy) is related to the Parisi functional through a Legendre transform. This was made precise by Auffinger and Ben Arous [AB13].

(iii) The cavity method (Mézard and Parisi [MP01]) provides a direct, non-replica derivation of the Parisi equation by analyzing how the Gibbs measure changes when a single spin is added.

These developments confirm that the RSB solution reflects genuine structural properties of the Gibbs measure, not an artifact of the replica trick.

Remark 7.5 (Summary of Part II). Chapters **??**–7 have developed the Parisi solution from first principles: the nested Gaussian integral and Parisi PDE (Chapter **??**), the perturbative analysis near T_c (Chapter **??**), and the numerical solution at all temperatures (this chapter). The solution is internally consistent, the thermodynamic quantities converge monotonically, and the entropy anomaly is resolved.

What remains is the question of mathematical proof: is the Parisi formula (5.23) correct? This is the subject of Part III.

Part III

Rigorous Results

Chapter 8

Rigorous Results

8.1 Guerra's Bound

The first rigorous result toward the Parisi formula was obtained by Guerra [Gue03], who proved one of the two inequalities needed: the SK free energy is *bounded above* by the Parisi functional.

Theorem 8.1 (Guerra, 2003)**.** *For all $\beta > 0$ and $h \in \mathbb{R}$,*

$$f(\beta, h) \leq \inf_{\mu \in \mathcal{M}([0,1])} \mathcal{P}(\beta, h, \mu). \tag{8.1}$$

This is sometimes called the "easy half" of the Parisi formula, though the proof was a major breakthrough that introduced techniques now central to the field.

The interpolation method

The proof constructs a one-parameter family of models that interpolates between the SK model and a "comparison model" whose free energy equals the Parisi functional. The comparison model replaces the spin-spin interactions with a hierarchical random field drawn from a Ruelle cascade (Definition 4.20).

Definition 8.2 (Interpolating Hamiltonian)**.** For $t \in [0, 1]$, define

$$H_t(\sigma) = \sqrt{t}\, H_{\mathrm{SK}}(\sigma) + \sqrt{1-t}\, H_{\mathrm{Parisi}}(\sigma), \tag{8.2}$$

where H_{SK} is the SK Hamiltonian and H_{Parisi} is a Gaussian field with covariance structure determined by the Ruelle cascade: if σ and σ' belong to the same cluster at level k of the cascade, then $\mathrm{Cov}(H_{\mathrm{Parisi}}(\sigma), H_{\mathrm{Parisi}}(\sigma')) = N\, q_k$.

At $t = 1$, the interpolating model is the SK model. At $t = 0$, the spins are decoupled and the free energy can be computed explicitly in terms of the Parisi PDE solution (this is the content of Chapter **??**). The key step is to show that the interpolating free energy $\varphi(t) = \frac{1}{N}\mathbb{E} \ln Z_t$ is monotone.

Proposition 8.3 (Monotonicity of the interpolation)**.** *The derivative $\varphi'(t) \leq 0$ for all $t \in [0, 1]$.*

The proof uses Gaussian integration by parts (Stein's lemma) to express $\varphi'(t)$ in terms of overlap moments. The resulting expression involves a sum of terms of the form $\mathbb{E}\langle(R_{1,2}-q_k)(R_{1,2}-q_{k+1})\rangle$, where $R_{1,2}$ is the overlap between two replicas sampled from the Gibbs measure at parameter t. A convexity argument (using the fact that the covariance function of the SK model is convex in the overlap) shows that each term is non-positive.

Since $\varphi(1) = f_N(\beta, h)$ (the SK free energy) and $\varphi(0) = \mathcal{P}(\beta, h, \mu) + o(1)$, the monotonicity gives $f_N \le \mathcal{P} + o(1)$. Taking $N \to \infty$ and optimizing over μ yields Theorem 8.1.

Remark 8.4 (The role of convexity). Guerra's proof exploits the specific structure of the SK model: the covariance $\mathbb{E}[H(\sigma)H(\sigma')] = N\,\xi(R_{1,2})$ with $\xi(q) = q^2$ (a convex function). This convexity is what makes the interpolation monotone. For more general mixed p-spin models with $\xi(q) = \sum c_p q^p$ ($c_p \ge 0$), the same argument applies because a positive combination of convex functions is convex. The method fails when ξ is not convex, which is the source of difficulty in some non-mean-field models.

8.2 Talagrand's Proof

The complementary inequality — the "hard half" — was proved by Talagrand [Tal06], completing the proof of the Parisi formula.

Theorem 8.5 (Talagrand, 2006)**.** *For all $\beta > 0$ and $h \in \mathbb{R}$,*

$$f(\beta, h) \ge \inf_{\mu \in \mathcal{M}([0,1])} \mathcal{P}(\beta, h, \mu). \tag{8.3}$$

Combined with Guerra's bound (Theorem 8.1), this gives:

Corollary 8.6 (The Parisi formula)**.** $f(\beta, h) = \inf_\mu \mathcal{P}(\beta, h, \mu)$.

This confirms that every formal calculation in Chapters **??**–7 gives the correct answer: the $n \to 0$ replica trick, the hierarchical ansatz, the nested Gaussian integrals, and the Parisi PDE all produce the exact free energy of the SK model.

Key ideas of the proof

Talagrand's proof is substantially more technical than Guerra's. We describe the main ingredients without attempting a full account.

The Aizenman–Sims–Starr representation. The starting point is a representation of the SK free energy as an infimum over a class of "ROSt" (random overlap structures). A ROSt is an abstract random measure ν on a space equipped with a covariance kernel, generalizing the Ruelle cascade. Aizenman, Sims, and Starr [ASS03] showed that

$$f(\beta, h) = \inf_{\nu \in \mathrm{ROSt}} F(\nu),$$

where $F(\nu)$ is a functional depending on the overlap structure of ν. The Parisi formula is the special case where ν ranges over Ruelle cascades (which are ultrametric ROSts).

The Ghirlanda–Guerra identities. The bridge between general ROSts and Ruelle cascades is provided by the Ghirlanda–Guerra identities (Theorem 4.21). Talagrand showed that any sequence of ROSts achieving the infimum must, in the limit, satisfy these identities. By Panchenko's theorem (Theorem 4.22), any overlap array satisfying the GG

identities is generated by a Ruelle cascade. Therefore the infimum over all ROSts equals the infimum over Ruelle cascades, which is $\inf_\mu \mathcal{P}$.

The induction scheme. The technical core of Talagrand's proof is an induction on the number of levels in the Ruelle cascade, using a cavity argument (adding one spin at a time) to control the error at each step. The argument requires delicate concentration inequalities for the free energy of perturbed models.

Remark 8.7 (Panchenko's simplification). Panchenko [Pan13a] later gave a substantially simplified proof of Theorem 8.5, avoiding much of the technical machinery of Talagrand's original argument. Panchenko's approach makes more direct use of the ultrametric structure forced by the GG identities and has become the standard reference for this result.

8.3 Ultrametricity

The ultrametric structure of the Parisi solution (Section 3.3) was originally a hypothesis. Panchenko's theorem shows it is a *consequence* of the model's self-consistency properties.

Theorem 8.8 (Panchenko, 2013)**.** *Let $(R_{\ell,\ell'})_{\ell,\ell'\geq 1}$ be the overlap array of the SK model. The array is automatically* weakly exchangeable *(invariant under permutations of the replica index) and* positive semi-definite *(since $R_{\ell,\ell'} = N^{-1}\sigma^\ell \cdot \sigma^{\ell'}$ is a Gram matrix). By the Dovbysh–Sudakov theorem (the analogue of de Finetti's theorem for Gram matrices), any such array admits a representation as a mixture of i.i.d. samples from a random measure. The non-trivial content is:*

(i) The array satisfies the Ghirlanda–Guerra identities (Theorem 4.21). These are provable consequences of the Gaussian covariance structure and the concentration of the Hamiltonian.

(ii) Any overlap array satisfying the GG identities is ultrametric: for any three replicas ℓ_1, ℓ_2, ℓ_3, the two smallest values among $\{R_{\ell_1,\ell_2}, R_{\ell_1,\ell_3}, R_{\ell_2,\ell_3}\}$ are equal with probability one.

(iii) The law of such an array is uniquely determined by the single overlap distribution $\mu = \mathrm{law}(R_{1,2})$ and is generated by a Ruelle cascade (Definition 4.20) with parameters determined by μ.

Significance

This theorem has a profound implication: the Parisi ansatz (hierarchical matrices indexed by ultrametric trees) is not merely a convenient parametrization — it is the *only possible structure* consistent with the thermodynamic self-consistency of the SK model.

In the language of Chapter **??**: the space of possible overlap matrices Q in the $n \to 0$ limit is not the full space of symmetric matrices, nor even the space of matrices with some block structure. It is precisely the space of Parisi matrices (Definition 4.2), parametrized by a single non-decreasing function $q : [0, 1] \to [0, 1]$.

Remark 8.9 (The logical structure of the proof). The complete proof of the Parisi formula and the ultrametric structure can be summarized in four steps:

(1) **Guerra:** $f(\beta, h) \leq \inf_\mu \mathcal{P}(\beta, h, \mu)$ (Theorem 8.1).

(2) **Talagrand/Panchenko:** $f(\beta, h) \geq \inf_\mu \mathcal{P}(\beta, h, \mu)$ (Theorem 8.5). Therefore $f = \inf_\mu \mathcal{P}$.

(3) **Ghirlanda–Guerra:** The overlap array of the SK model satisfies the GG identities (Theorem 4.21).

(4) **Panchenko:** GG identities ⇒ ultrametricity ⇒ Ruelle cascade structure (Theorem 8.8).

Steps (1)–(2) determine the free energy. Steps (3)–(4) determine the structure of the Gibbs measure. Together they provide a complete mathematical vindication of Parisi's solution.

Why every other pattern fails

The uniqueness of ultrametricity can be understood from several complementary angles, each ruling out alternatives.

Non-hierarchical block structures fail the replicon stability test at every level. Parisi and Ricci-Tersenghi [PRT00] analyzed the constraints from stochastic stability on the joint distribution $P^{(3)}(q_1, q_2, q_3)$ of three overlaps. For $k = 3, 4, 5$ overlap values, the most general solution satisfying all polynomial identities derived from the Ghirlanda–Guerra relations is the ultrametric one. Non-ultrametric structures are systematically eliminated.

Random matrix structures fail because the saddle-point equations require the overlap matrix algebra to close under multiplication and functional calculus ($\operatorname{Tr} Q^k$, $\det Q$, analytic functions of Q) in the $n \to 0$ limit. Generic symmetric matrices do not form such an algebra. The Parisi hierarchical matrices do: products, inverses, and functions of Parisi matrices are again Parisi matrices with the same block structure. This algebraic closure is essential for the self-consistency of the saddle-point equations.

Any finite-K RSB fails marginal stability. The replicon eigenvalue within the finest blocks is strictly negative for any finite K in the SK model, regenerating further hierarchical structure at a finer scale. Only the full RSB limit achieves zero replicon eigenvalue everywhere — the unique fixed point of the iterated instability. This marginal stability is the fingerprint of the continuous ultrametric solution.

The GG identities eliminate everything non-ultrametric in complete generality. Panchenko's theorem does not assume any particular form for the overlap distribution or the Hamiltonian beyond Gaussian covariance. It applies to the SK model, all mixed p-spin models, and potentially to finite-dimensional models satisfying appropriate conditions. The theorem's power lies in its abstract character: *any* system whose Gibbs measure satisfies the GG identities must have ultrametric overlaps.

Appendices

Appendix A

Gaussian Integration Identities

Overview

This appendix collects the Gaussian integration identities used throughout the monograph. All random variables denoted z are standard Gaussian ($z \sim \mathcal{N}(0,1)$) unless otherwise stated.

A.1 The Hubbard–Stratonovich Transformation

The fundamental identity that enables the replica method is the linearization of quadratic forms in the exponential.

Lemma A.1 (Hubbard–Stratonovich)**.** *For any $a \in \mathbb{R}$ and $S_1, \dots, S_n \in \mathbb{R}$,*

$$\exp\Big(\frac{a}{2}\Big(\sum_{\alpha=1}^{n} S_\alpha\Big)^2\Big) = \mathbb{E}_z\Big[\exp\Big(\sqrt{a}\, z \sum_{\alpha=1}^{n} S_\alpha\Big)\Big] \tag{A.1}$$

when $a > 0$. For $a < 0$, the identity holds with $\sqrt{a}$ replaced by $i\sqrt{|a|}$ (an imaginary Gaussian).

Proof. Complete the square: $\mathbb{E}_z[e^{bz}] = e^{b^2/2}$ for $b \in \mathbb{R}$. Setting $b = \sqrt{a}\sum S_\alpha$ gives $\mathbb{E}_z[\exp(\sqrt{a}\, z \sum S_\alpha)] = \exp\big(\frac{a}{2}(\sum S_\alpha)^2\big)$. □

This identity is used in Section 2.2 to decouple the replica interactions and in Chapter **??** at each level of the nested Gaussian integral.

A.2 Gaussian Integration by Parts (Stein's Lemma)

Lemma A.2 (Stein's lemma)**.** *If $z \sim \mathcal{N}(0,1)$ and $\varphi : \mathbb{R} \to \mathbb{R}$ is absolutely continuous with $\mathbb{E}[|\varphi'(z)|] < \infty$, then*

$$\mathbb{E}[z\,\varphi(z)] = \mathbb{E}[\varphi'(z)]. \tag{A.2}$$

More generally, if $(z_1, \dots, z_n) \sim \mathcal{N}(0, \Sigma)$ is a centered Gaussian vector, then

$$\mathbb{E}[z_i\,\varphi(z_1,\dots,z_n)] = \sum_{j=1}^{n} \Sigma_{ij}\, \mathbb{E}\Big[\frac{\partial \varphi}{\partial z_j}(z_1,\dots,z_n)\Big]. \tag{A.3}$$

Proof. For the univariate case: integrate by parts, $\mathbb{E}[z\varphi(z)] = \int z\varphi(z)\frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz = \int \varphi(z)\frac{d}{dz}(-e^{-z^2/2})\frac{1}{\sqrt{2\pi}}\,dz$ $\int \varphi'(z)\frac{e^{-z^2/2}}{\sqrt{2\pi}}\,dz = \mathbb{E}[\varphi'(z)]$, where the boundary terms vanish by the growth condition. The multivariate version follows by the same argument applied to each coordinate. □

This identity is used in the derivation of the self-consistency equation (Section 2.3), the Hessian computation (Proposition 2.10), and Guerra's interpolation bound (Section 8.1).

A.3 Moments of Log-Partition Functions

Lemma A.3 (Replica identity)**.** *For any random variable $Z > 0$,*

$$\mathbb{E}[\ln Z] = \lim_{n\to 0} \frac{\mathbb{E}[Z^n] - 1}{n} = \lim_{n\to 0} \frac{1}{n} \ln \mathbb{E}[Z^n] + O(n). \tag{A.4}$$

The first equality is exact (when the limit exists); the second uses $\mathbb{E}[Z^n] = \exp\Big(n\mathbb{E}[\ln Z] + \frac{n^2}{2}\mathrm{Var}(\ln Z) + \cdots\Big)$ and holds to leading order in n.

This is the identity that motivates the replica method: computing $\mathbb{E}[\ln Z_N]$ via $\mathbb{E}[Z_N^n]$ for integer n, then continuing to $n = 0$. The $O(n)$ error in the second form is why the replica method gives the correct *leading order* (the free energy) but does not directly control fluctuations.

A.4 Gaussian Tail Bounds

Lemma A.4. *For $z \sim \mathcal{N}(0,1)$ and $t > 0$:*

$$\mathbb{P}(|z| > t) \le 2\,e^{-t^2/2}. \tag{A.5}$$

For a Lipschitz function φ with $|\varphi(x) - \varphi(y)| \le L|x - y|$:

$$\mathbb{P}\Big(|\varphi(z) - \mathbb{E}[\varphi(z)]| > t\Big) \le 2\,e^{-t^2/(2L^2)}. \tag{A.6}$$

The concentration inequality (A.6) (Gaussian concentration of measure) is used implicitly in the rigorous results of Chapter 8, where it controls the self-averaging of the free energy.

Appendix B

The Heat Equation and Semigroup

Overview

The heat semigroup C_t appears throughout the Parisi solution as the operator implementing Gaussian convolution. This appendix summarizes its properties.

B.1 The Heat Equation

The *heat equation* on $\mathbb{R}$ is

$$\frac{\partial u}{\partial t}(t,y) = \frac{1}{2}\frac{\partial^2 u}{\partial y^2}(t,y), \qquad t>0,\ y\in\mathbb{R}, \tag{B.1}$$

with initial condition $u(0,y)=g(y)$. The unique solution (for g growing at most exponentially) is given by convolution with the *heat kernel*:

$$u(t,y) = (G_t * g)(y) = \frac{1}{\sqrt{2\pi t}}\int_{-\infty}^{\infty} g(y-s)\, e^{-s^2/(2t)}\, ds = \mathbb{E}_z[g(y+\sqrt{t}\,z)], \tag{B.2}$$

where $z\sim\mathcal{N}(0,1)$.

B.2 The Semigroup C_t

Definition B.1. The heat semigroup C_t acts on functions $g:\mathbb{R}\to\mathbb{R}$ by $(C_t\,g)(y) = \mathbb{E}_z[g(y+\sqrt{t}\,z)]$. In operator notation, $C_t = \exp\!\left(\tfrac{t}{2}\partial_y^2\right)$.

Proposition B.2 (Properties of C_t)**.**

(i) ***Semigroup:*** *$C_s C_t = C_{s+t}$ for $s,t\ge 0$.*

(ii) ***Positivity:*** *If $g\ge 0$ then $C_t\,g\ge 0$.*

(iii) ***Smoothing:*** *For any $t>0$ and $g\in L^1(\mathbb{R})$, $C_t\,g\in C^\infty(\mathbb{R})$.*

(iv) ***Monotonicity:*** *If $g_1\le g_2$ pointwise then $C_t\,g_1\le C_t\,g_2$.*

(v) ***Log-convexity:*** *If $g>0$ then $y\mapsto \ln(C_t\,g)(y)$ is convex for each $t\ge 0$.*

Proof. (i) follows from $\sqrt{s}\,z_1+\sqrt{t}\,z_2\sim\mathcal{N}(0,s+t)$ for independent Gaussians. (ii)–(iv) are immediate from the integral representation (B.2). (v) follows from Hölder's inequality: for $\lambda\in(0,1)$, $(C_t\,g)(\lambda y_1+(1-\lambda)y_2)=\mathbb{E}[g(\lambda y_1+(1-\lambda)y_2+\sqrt{t}\,z)]\le(\mathbb{E}[g(y_1+\sqrt{t}\,z)])^{\lambda}(\mathbb{E}[g(y_2+\sqrt{t}\,z)])^{1-\lambda}$, which gives $\ln(C_t\,g)(\lambda y_1+(1-\lambda)y_2)\le\lambda\ln(C_t\,g)(y_1)+(1-\lambda)\ln(C_t\,g)(y_2)$. □

Property (v) is used in the well-posedness argument for the Parisi PDE (Section 5.3): the operation $g\mapsto\frac{1}{m}\ln(C_t[e^{mg}])$ preserves convexity of g in y, ensuring that the solution remains well-behaved at each step of the recursion.

B.3 The Cole–Hopf Transformation

The substitution $u=e^{-xf}$ transforms the Parisi PDE $\partial_x f=-\frac{\dot q}{2}[\beta^2\partial_{yy}f+x\beta^2(\partial_y f)^2]$ into the linear equation

$$\frac{\partial u}{\partial x}=\frac{\dot q\beta^2}{2}\frac{\partial^2 u}{\partial y^2} \tag{B.3}$$

for $x>0$ (where the division by x in $f=-\frac{1}{x}\ln u$ is valid). This is a backward heat equation with time-dependent diffusion coefficient $\frac{\dot q(x)\beta^2}{2}$. The transformation is analogous to the Cole–Hopf transformation $u=e^{-v}$ that linearizes the viscous Burgers equation $\partial_t v+v\,\partial_x v=\frac{\nu}{2}\partial_{xx}v$ to the heat equation $\partial_t u=\frac{\nu}{2}\partial_{xx}u$.

At $x=0$, the transformation degenerates ($u=e^0=1$ regardless of f), which is why the direct approach via step-function approximation (Proposition 5.8) is preferred for establishing well-posedness.

Appendix C

Probability Measures and Quantile Functions

Overview

The order parameter space $\mathcal{Q}$ of the Parisi solution consists of non-decreasing right-continuous functions $q : [0,1] \to [0,1]$, which are in bijection with probability measures on $[0,1]$. This appendix makes the correspondence precise.

C.1 Quantile Functions and Distribution Functions

Definition C.1. Let μ be a probability measure on $[0,1]$. Its *cumulative distribution function* is $F(t) = \mu([0,t])$ for $t \in [0,1]$. Its *quantile function* (or generalized inverse) is

$$q(x) = F^{-1}(x) = \inf\{t \in [0,1] : F(t) \geq x\}, \qquad x \in (0,1). \tag{C.1}$$

We extend by $q(0) = \inf \operatorname{supp}(\mu)$ and $q(1) = \sup \operatorname{supp}(\mu)$.

Proposition C.2. *The quantile function $q : [0,1] \to [0,1]$ is non-decreasing and right-continuous. Conversely, every non-decreasing right-continuous function $q : [0,1] \to [0,1]$ is the quantile function of a unique probability measure μ on $[0,1]$, defined by $\mu([0,t]) = \sup\{x : q(x) \leq t\}$.*

This bijection is the reason the Parisi solution can be parametrized equivalently by a function $q \in \mathcal{Q}$ or a measure $\mu \in \mathcal{M}([0,1])$.

C.2 The Integral Formula

If $\varphi : [0,1] \to \mathbb{R}$ is continuous, then

$$\int_0^1 \varphi(q(x))\, dx = \int_0^1 \varphi(t)\, d\mu(t). \tag{C.2}$$

In particular, $\int_0^1 q(x)^2\, dx = \int_0^1 t^2\, d\mu(t)$ (the second moment of μ). This identity is used repeatedly in Chapters **??**–**??**.

C.3 Topology and Compactness

Definition C.3. The *weak-$*$ topology* on $\mathcal{M}([0,1])$ is the coarsest topology making $\mu \mapsto \int \varphi \, d\mu$ continuous for every continuous $\varphi : [0,1] \to \mathbb{R}$. Equivalently, $\mu_n \to \mu$ weak-$*$ if and only if $\int \varphi \, d\mu_n \to \int \varphi \, d\mu$ for all continuous φ.

Proposition C.4. *$\mathcal{M}([0,1])$ is compact and metrizable in the weak-$*$ topology. Equivalently, $\mathcal{Q}$ is compact in the $L^1[0,1]$ topology: every sequence of non-decreasing functions $q_n : [0,1] \to [0,1]$ has a subsequence converging in L^1 to a non-decreasing function $q \in \mathcal{Q}$.*

Proof. This is a consequence of Helly's selection theorem: any uniformly bounded sequence of monotone functions has a pointwise convergent subsequence. Pointwise convergence of monotone functions on $[0,1]$ implies L^1 convergence by the bounded convergence theorem. □

The compactness of $\mathcal{Q}$ (equivalently, $\mathcal{M}([0,1])$) is essential for the existence of the minimizer of the Parisi functional (Remark 5.7).

C.4 The Wasserstein Distance

The L^1 distance between quantile functions defines a metric on $\mathcal{M}([0,1])$:

$$W_1(\mu,\nu) = \int_0^1 |q_\mu(x) - q_\nu(x)| \, dx, \tag{C.3}$$

where q_μ, q_ν are the quantile functions of μ, ν. This is the *Wasserstein-1 distance* (also called the earth mover's distance). It metrizes the weak-$*$ topology on $\mathcal{M}([0,1])$, and the continuity of the Parisi functional in this metric (Theorem 5.9(i)) ensures that the variational problem $\inf_\mu \mathcal{P}$ is well-posed.

Appendix D

Notation and Conventions

Complete Notation Reference

General

Symbol	Meaning	Introduced
N	Number of spins	§1.1
$\Sigma_N = \{-1, +1\}^N$	Configuration space	§1.1
$\sigma = (\sigma_1, \ldots, \sigma_N)$	Spin configuration	§1.1
$J_{ij} \sim \mathcal{N}(0, 1/N)$	Coupling constants	§1.1
$\beta = 1/T$	Inverse temperature	§1.1
h	External field	§1.1
$T_c = 1$	Critical temperature	§1.2
$\tau = 1 - T$	Reduced temperature	§6.1

Thermodynamic quantities

Symbol	Meaning	Introduced
$H_N(\sigma)$	Hamiltonian	§1.1
$Z_N(\beta, h)$	Partition function	§1.1
$f(\beta, h) = \lim \frac{1}{N} \mathbb{E} \ln Z_N$	Free energy (per spin)	§1.1
$U = -\partial f / \partial \beta$	Internal energy	§1.3
$S = f + \beta U$	Entropy	§1.3
$C = \beta^2 \, \partial^2 f / \partial \beta^2$	Specific heat	§1.3
$\langle \cdot \rangle$	Gibbs average	§1.1
$\mathbb{E}[\cdot]$	Disorder average	§1.1

Overlaps and order parameters

Symbol	Meaning	Introduced
$R_{1,2} = \frac{1}{N}\sum_i \sigma_i^{(1)}\sigma_i^{(2)}$	Overlap	§1.2
μ_N	Empirical overlap distribution	§1.2
q^*	RS overlap (self-consistency solution)	§2.3
$q_{\mathrm{EA}} = q(1)$	Edwards–Anderson parameter	§4.2
$q : [0,1] \to [0,1]$	Parisi function (order parameter)	§4.2
μ	Overlap measure (pushforward of q)	§4.2
$\mathcal{Q}$	Space of non-decreasing functions $[0,1] \to [0,1]$	§4.2

Replica method

Symbol	Meaning	Introduced
n	Number of replicas	§2.1
$\alpha, \beta, \gamma, \ldots$	Replica indices $(1, \ldots, n)$	§2.1
$Q_{\alpha\beta}$	Overlap matrix	§2.1
$\Phi_n(Q)$	Replica free energy functional	§2.2
$\mathcal{S}_n$	Symmetric matrices with zero diagonal	§2.3
V_L, V_A, V_R	Longitudinal, anomalous, replicon sectors	§2.3
λ_R	Replicon eigenvalue	§2.3

Parisi matrix and RSB

Symbol	Meaning	Introduced
K	Number of RSB steps	§4.1
$q_0, \ldots, q_K$	Overlap values $(q_0 < \cdots < q_K)$	§4.1
$m_0, \ldots, m_{K+1}$	Block sizes $(m_0 = 1 \geq \cdots \geq m_{K+1} = 0)$	§4.1
$d(\alpha, \beta)$	Ultrametric depth function	§4.1
$w_i = m_i - m_{i+1}$	Overlap weights	§4.2
$Q^{(K)}$	K-step Parisi matrix	§4.1

Parisi PDE and functional

Symbol	Meaning	Introduced
$C_t = \exp\left(\frac{t}{2}\partial_y^2\right)$	Heat semigroup	§5.2
$f(x, y)$	Solution of the Parisi PDE	§5.2
$\mathcal{P}(\beta, h, \mu)$	Parisi functional	§5.2
$B_x = \beta \int_0^x \sqrt{\dot{q}}\, dW_s$	Time-changed Brownian motion	§5.3

Near-T_c expansion

Symbol	Meaning	Introduced
$B_i^{(K)} = \frac{2(K-i)+1}{2K+1}$	Overlap coefficient	§6.2
$L_i^{(K)} = \frac{6i}{2K+1}$	Block size coefficient	§6.2
$F_5^{(K)} = \frac{9}{20} - \frac{1}{5(2K+1)^4}$	Fifth-order free energy	§6.3

Convention comparison

Physics notation	This monograph
$\langle\cdot\rangle$ (thermal average)	$\langle\cdot\rangle$ or $\mathbb{E}_{G_N}[\cdot]$
$\overline{(\cdot)}$ (disorder average)	$\mathbb{E}[\cdot]$ or $\mathbb{E}_J[\cdot]$
Free energy $F = -T \ln Z$	$f = \frac{1}{N}\mathbb{E}[\ln Z_N]$ (per spin, β-scaled)
Maximization of replica functional	Minimization of $\mathcal{P}$ (sign flip at $n \to 0$)
m_i increasing (integer n)	m_i decreasing ($n \to 0$ convention)

Bibliography

[AB13] Antonio Auffinger and Gérard Ben Arous. Complexity of random smooth functions on the high-dimensional sphere. *Annals of Probability*, 41(6):4214–4247, 2013.

[AC15] Antonio Auffinger and Wei-Kuo Chen. The Parisi formula has a unique minimizer. *Communications in Mathematical Physics*, 335(3):1429–1444, 2015.

[ASS03] Michael Aizenman, Robert Sims, and Shannon L. Starr. Extended variational principle for the Sherrington–Kirkpatrick spin-glass model. *Physical Review B*, 68(21):214403, 2003.

[DSS15] Jian Ding, Allan Sly, and Nike Sun. Proof of the satisfiability conjecture for large k. In *Proceedings of the 47th Annual ACM Symposium on Theory of Computing*, pages 59–68, 2015.

[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: Mathematical and General*, 11(5):983–990, 1978.

[EA75] Samuel F. Edwards and Philip W. Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965–974, 1975.

[GG98] Stefano Ghirlanda and Francesco Guerra. General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity. *Journal of Physics A: Mathematical and General*, 31(46):9149–9155, 1998.

[GT02] Francesco Guerra and Fabio Lucio Toninelli. The thermodynamic limit in mean field spin glass models. *Communications in Mathematical Physics*, 230(1):71–79, 2002.

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

[JT16] Aukosh Jagannath and Ian Tobasco. A dynamic programming approach to the Parisi functional. *Proceedings of the American Mathematical Society*, 144(7):3135–3150, 2016.

[KS78] Scott Kirkpatrick and David Sherrington. Infinite-ranged models of spin-glasses. *Physical Review B*, 17(11):4384–4403, 1978.

[LM19] Marc Lelarge and Léo Miolane. Fundamental limits of symmetric low-rank matrix estimation. *Probability Theory and Related Fields*, 173(3–4):859–929, 2019.

[MP01] Marc Mézard and Giorgio Parisi. The Bethe lattice spin glass revisited. *European Physical Journal B*, 20(2):217–233, 2001.

[MPS+84] Marc Mézard, Giorgio Parisi, Nicolas Sourlas, Gérard Toulouse, and Miguel Angel Virasoro. Nature of the spin-glass phase. *Physical Review Letters*, 52(13):1156–1159, 1984.

[MPV87] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin Glass Theory and Beyond*. World Scientific, Singapore, 1987.

[Pan13a] Dmitry Panchenko. The Parisi ultrametricity conjecture. *Annals of Mathematics*, 177(1):383–393, 2013.

[Pan13b] Dmitry Panchenko. *The Sherrington–Kirkpatrick Model*. Springer Monographs in Mathematics. Springer, New York, 2013.

[Par79a] Giorgio Parisi. Infinite number of order parameters for spin-glasses. *Physical Review Letters*, 43(23):1754–1756, 1979.

[Par79b] Giorgio Parisi. Toward a mean field theory for spin glasses. *Physics Letters A*, 73(3):203–205, 1979.

[Par80a] Giorgio Parisi. The order parameter for spin glasses: a function on the interval 0–1. *Journal of Physics A: Mathematical and General*, 13(3):1101–1112, 1980.

[Par80b] Giorgio Parisi. A sequence of approximated solutions to the S–K model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115–L121, 1980.

[Par83] Giorgio Parisi. Order parameter for spin-glasses. *Physical Review Letters*, 50(24):1946–1948, 1983.

[PRT00] Giorgio Parisi and Federico Ricci-Tersenghi. On the origin of ultrametricity. *Journal of Physics A: Mathematical and General*, 33(1):113–129, 2000.

[SK75] David Sherrington and Scott Kirkpatrick. Solvable model of a spin-glass. *Physical Review Letters*, 35(26):1792–1796, 1975.

[Tal06] Michel Talagrand. The Parisi formula. *Annals of Mathematics*, 163(1):221–263, 2006.

[Tal11] Michel Talagrand. *Mean Field Models for Spin Glasses: Volume I*, volume 54 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge*. Springer, Berlin, 2011.