

The Sherrington-Kirkpatrick model

The Sherrington-Kirkpatrick (SK) model was introduced by David Sherrington and Scott Kirkpatrick in 1975 as a simple ‘solvable’ (in their words) model for spin-glasses. Spin-glasses are some type of magnetic alloys, and ‘solvable’ meant that the asymptotic free entropy density could be computed exactly.

It turns out that the original SK solution was incorrect and in fact inconsistent (the authors knew this). A consistent conjecture for the asymptotic free energy per spin was put forward by Giorgio Parisi in 1982, and derived through the non-rigorous replica method.

It took 24 years to prove this conjecture. The final proof is due to Michel Talagrand (2006) and is a real tour de force. In these two lectures we will prove that the asymptotic free entropy density exists and that it is upper bounded by the Parisi formula. The first result is due to Francesco Guerra and Fabio Toninelli [GT02], and the second to Guerra [Gue03]. They are based on an interpolation trick that was a authentic breakthrough eventually leading to Talagrand’s proof.

1 Definitions and the Parisi formula

Let $\{J_{ij}\}_{i,j \in [n]}$ be a collection of i.i.d. $\mathbf{N}(0, 1/(2n))$ random variables. The SK model is the random measure over $\underline{x} \in \{+1, -1\}^n$ defined by

$$\mu_{J,\beta,B}(\underline{x}) = \frac{1}{Z_n(\beta, B)} \exp \left\{ \beta \sum_{i,j=1}^n J_{ij} x_i x_j + B \sum_{i=1}^n x_i \right\}, \quad (1)$$

Here $Z_n(\beta, B)$ is defined by the normalization condition $\sum_{\underline{x}} \mu_{J,\beta,B}(\underline{x}) = 1$. It is of course a random variable. An alternative but sometime useful formulation of the model consists in saying that

$$\mu_{\beta,B}(\underline{x}) = \frac{1}{Z_n(\beta, B)} \exp \left\{ -H(\underline{x}) \right\}, \quad (2)$$

where $H(\underline{x})$ is a gaussian process indexed by $\underline{x} \in \{+1, -1\}^n$ with mean and covariance

$$\mathbb{E} H(\underline{x}) = -n B M_{\underline{x}}, \quad \text{Cov}(H(\underline{x}), H(\underline{y})) = \frac{1}{2} n \beta^2 Q_{\underline{x}, \underline{y}}^2. \quad (3)$$

Here $M_{\underline{x}}$, and $Q_{\underline{x}, \underline{y}}$ denote the empirical magnetization and empirical overlap, namely

$$M_{\underline{x}} \equiv \frac{1}{n} \sum_{i=1}^n x_i, \quad Q_{\underline{x}, \underline{y}} \equiv \frac{1}{n} \sum_{i=1}^n x_i y_i. \quad (4)$$

Throughout these lectures we will be interested in the free entropy density

$$\phi_n(\beta, B) \equiv \frac{1}{n} \log Z_n(\beta, B). \quad (5)$$

Parisi formula provides a surprising prediction for this quantity.

Definition 1. Let \mathcal{D} be the space of non-decreasing functions $x : [0, 1] \rightarrow [0, 1]$. The Parisi functional is the function $\mathbb{P} : \mathcal{D} \times \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$ defined by

$$\mathbb{P}[x; \beta, B] = \log 2 + f(0, B; x) - \frac{\beta^2}{2} \int_0^1 q x(q) dq, \quad (6)$$

where $f : [0, 1] \times \mathbb{R} \times \mathcal{D} \rightarrow \mathbb{R}$, $(q, y, x) \mapsto f(q, y; x)$ is the unique solution of the partial differential equation

$$\frac{\partial f}{\partial q} + \frac{1}{2} \frac{\partial^2 f}{\partial y^2} + \frac{1}{2} x(q) \left(\frac{\partial f}{\partial y} \right)^2 = 0, \quad (7)$$

with boundary condition $f(1, y; x) = \log \cosh \beta y$.

The relation between Parisi formula and the SK free entropy is given by the following theorem.

Theorem 2. *Let $\phi_n(\beta, B)$ be the free entropy density of a Sherrington-Kirkpatrick model with n variables. Then, almost surely*

$$\lim_{n \rightarrow \infty} \phi_n(\beta, B) = P_*(\beta, B) \equiv \inf_{x \in \mathcal{D}} P[x; \beta, B]. \quad (8)$$

At first sight, this result appears analogous to the one we proved in the first lecture for the Curie-Weiss model, that we reproduce here for convenience:

$$\lim_{n \rightarrow \infty} \phi_n^{\text{CW}}(\beta, B) = \sup_{m \in [-1, 1]} \varphi_{\beta, B}(m), \quad (9)$$

$$\varphi_{\beta, B}(m) \equiv H\left(\frac{1+m}{2}\right) + \frac{1}{2} \beta m^2 + Bm. \quad (10)$$

Notice however two important and surprising differences: (i) The supremum in the last expression is replaced by an infimum in Eq. (8); (ii) The optimum is taken over a single real parameter in the Curie-Weiss model, and over a function in the SK case.

2 Existence of the limit

In this section we will prove that the limit $n \rightarrow \infty$ exists almost surely.

Theorem 3. *Let $\phi_n(\beta, B)$ be the free entropy density of a Sherrington-Kirkpatrick model with n variables. Then, almost surely*

$$\lim_{n \rightarrow \infty} \phi_n(\beta, B) = \phi_*(\beta, B), \quad (11)$$

for some non-random quantity $\phi_*(\beta, B)$.

Proof Let $\phi_n^{\text{av}}(\beta, B) = \mathbb{E} \phi_n(\beta, B)$ be the expected free entropy density. (We will often drop the dependence on β, B in the following.) The proof is proceed in two steps. First one shows that ϕ_n concentrates around ϕ_n^{av} , and therefore is sufficient to show that the latter (deterministic) sequence converges. This is proved by showing that the sequence $\{\phi_n^{\text{av}}\}_{n \geq 0}$ is superadditive, and applying Fekete's lemma.

Concentration follows from gaussian isoperimetry. Recall indeed that if $F : \mathbb{R}^k \rightarrow \mathbb{R}$ is Lipschitz continuous with modulus L (i.e. if $|F(a) - F(b)| \leq L \|a - b\|$) and $Z = (Z_1, \dots, Z_k)$ is a vector of i.i.d. standard $\mathbf{N}(0, \sigma^2)$ random variables, then

$$\mathbb{P}\{F(Z) - \mathbb{E} F(Z) \geq u\} \leq e^{-u^2/2(L\sigma)^2}. \quad (12)$$

It is easy to check that ϕ_n is a Lipschitz continuous function with modulus β of the n^2 random variables J_{ij} , whence

$$\mathbb{P}\{|\phi_n - \phi_n^{\text{av}}| \geq u\} \leq e^{-nu^2/\beta^2}. \quad (13)$$

By Borel-Cantelli it is therefore sufficient to prove that ϕ_n^{av} has a limit.

As mentioned, existence of the limit follows from the superadditivity of the sequence $\{n\phi_n^{\text{av}}\}_{n \geq 0}$, i.e. from the observation that, for any $n_1, n_2 \in \mathbb{N}$, letting $n = n_1 + n_2$, we have

$$n\phi_n^{\text{av}} \geq n_1\phi_{n_1}^{\text{av}} + n_2\phi_{n_2}^{\text{av}}. \quad (14)$$

In order to prove this inequality, define, for $t \in [0, 1]$, the gaussian process on the hypercube $\{H_t(\underline{x})\}_{\underline{x} \in \{+1, -1\}^n}$ with mean and covariance

$$\mathbb{E} H_t(\underline{x}) = -n B M = -n_1 B M_1 - n_2 B M_2, \quad (15)$$

$$\text{Cov}(H_t(\underline{x}), H_t(\underline{y})) = \frac{1}{2} \beta^2 \{t n Q^2 + (1-t)[n_1 Q_1^2 + n_2 Q_2^2]\}. \quad (16)$$

where, letting $[n] = V_1 \cup V_2$ with $V_1 = \{1, \dots, n_1\}$, $V_2 = \{n_1 + 1, \dots, n_1 + n_2 = n\}$, we defined, for $a \in \{1, 2\}$,

$$M_a \equiv \frac{1}{n_a} \sum_{i \in V_a} x_i, \quad (17)$$

$$Q_a \equiv \frac{1}{n_a} \sum_{i \in V_a} x_i y_i. \quad (18)$$

Further $M = (n_1/n)M_1 + (n_2/n)M_2$ and $Q = (n_1/n)Q_1 + (n_2/n)Q_2$ are the magnetization and overlap defined in Eq. (4).

Define

$$\Phi(t) \equiv \mathbb{E} \log \left\{ \sum_{\underline{x}} e^{-H_t(\underline{x})} \right\}. \quad (19)$$

It is obvious that $\Phi(1) = n\phi_n^{\text{av}}$. Further, for $t = 0$, one can represent the gaussian process as

$$\begin{aligned} H_0(\underline{x}) &= -\beta \sum_{i,j \in V_1} J_{ij}^1 x_i x_j - B \sum_{i \in V_1} x_i \\ &\quad -\beta \sum_{i,j \in V_2} J_{ij}^2 x_i x_j - B \sum_{i \in V_2} x_i, \end{aligned} \quad (20)$$

with $J_{ij}^1 \sim \mathcal{N}(0, 1/(2n_1))$ i.i.d. and $J_{ij}^2 \sim \mathcal{N}(0, 1/(2n_2))$. It follows that $\Phi(0) = n_1\phi_{n_1}^{\text{av}} + n_2\phi_{n_2}^{\text{av}}$.

The superadditivity is proved by showing that the first derivative $\Phi'(t)$ is non-negative. In order to compute this derivative, it is convenient to use the following Lemma.

Lemma 4. *For $t \in [0, 1]$, let $\{X_k\}_{k \in S}$ be a finite collection of normal random variables, with covariance $C_{kl}(t) = \mathbb{E}_t[X_k X_l] - \mathbb{E}_t[X_k] \mathbb{E}_t[X_l]$, and mean $a_k = \mathbb{E}_t[X_k]$ independent of t . Then, for any polynomially bounded function $F : \mathbb{R}^S \rightarrow \mathbb{R}$,*

$$\frac{d}{dt} \mathbb{E}_t \{F(X)\} = \frac{1}{2} \sum_{k,l \in S} \frac{dC_{k,l}}{dt}(t) \mathbb{E}_t \left\{ \frac{\partial^2 F}{\partial x_k \partial x_l}(X) \right\} \quad (21)$$

In order to use this formula for computing $\Phi'(t)$, we use

$$\frac{\partial^2}{\partial H(\underline{x}) \partial H(\underline{y})} \log \left\{ \sum_{\underline{x}} e^{-H(\underline{x})} \right\} = \mu(\underline{x}) \mathbb{I}_{\underline{x}=\underline{y}} - \mu(\underline{x}) \mu(\underline{y}), \quad (22)$$

and

$$\frac{d}{dt} \text{Cov}(H_t(\underline{x}), H_t(\underline{y})) = \frac{1}{2} \beta^2 n \left\{ \left(\frac{n_1}{n} Q_1(\underline{x}, \underline{y}) + \frac{n_2}{n} Q_2(\underline{x}, \underline{y}) \right)^2 - \frac{n_1}{n} Q_1(\underline{x}, \underline{y})^2 - \frac{n_2}{n} Q_2(\underline{x}, \underline{y})^2 \right\}. \quad (23)$$

Therefore the first term in Eq. (22) does not give any contribution to the derivative, and

$$\sum_{\underline{x}, \underline{y}} \frac{d}{dt} \text{Cov}(H_t(\underline{x}), H_t(\underline{y})) \frac{\partial^2 \log \left\{ \sum_{\underline{x}} e^{-H(\underline{x})} \right\}}{\partial H(\underline{x}) \partial H(\underline{y})} = -\frac{1}{2} \beta^2 n \mathbb{E}_{\mu \times \mu} \left\{ \left(\frac{n_1}{n} Q_1 + \frac{n_2}{n} Q_2 \right)^2 - \frac{n_1}{n} Q_1^2 - \frac{n_2}{n} Q_2^2 \right\},$$

which is non-negative by convexity of $x \mapsto x^2$. The derivative $\Phi'(t)$ is obtained by taking the expectation of the above and hence is non-negative as well. \square

3 RSB bounds

Theorem 5. *Let $\phi_n^{\text{av}}(\beta, B)$ be the expected free entropy density of a Sherrington-Kirkpatrick model with n variables. Then*

$$\phi_n^{\text{av}}(\beta, B) \leq \mathsf{P}_*(\beta, B) \equiv \inf_{x \in \mathcal{D}} \mathsf{P}[x; \beta, B]. \quad (24)$$

Proof Since β, B are fixed throughout the proof, we will regard $\mathsf{P} : \mathcal{D} \rightarrow \mathbb{R}$ uniquely as a function of $x : [0, 1] \rightarrow [0, 1]$. An important simplifying remark is that P is Lipschitz continuous with respect to the L_1 norm. More precisely, for $x, x' \in \mathcal{D}$, we have

$$|\mathsf{P}[x] - \mathsf{P}[x']| \leq \beta^2 \|x - x'\|_1. \quad (25)$$

We refer to [Gue03] for a proof of this statement.

It is therefore sufficient to prove $\phi_n^{\text{av}}(\beta, B) \leq \mathsf{P}[x]$ for x in a dense subset of \mathcal{D} . The proof proceeds by considering the set $\cup_{K \geq 1} \mathcal{D}_K$, where \mathcal{D}_K is the subset of functions $x : [0, 1] \rightarrow [0, 1]$ non-decreasing and right-continuous which takes at most K distinct values in $[0, 1)$ (plus, eventually, $x(1) = 1$). A function $x \in \mathcal{D}_K$ is parameterized by two vectors $0 \leq q_0 \leq q_1 \leq \dots \leq q_{K-1} \leq q_K \leq q_{K+1} = 1$ and $0 = m_0 \leq m_1 \leq \dots \leq m_K \leq m_{K+1} = 1$, by letting, for $q \in [0, 1)$

$$x(q) = \sum_{i=1}^{K+1} m_i \mathbb{I}\{q \in [q_{i-1}, q_i)\}. \quad (26)$$

(The notation is here slightly different from [Gue03].)

Of particular interest is the case $K = 0$ (replica symmetric, RS) where

$$x(q) = \begin{cases} 0 & \text{if } 0 \leq q < q_0, \\ 1 & \text{if } q_0 \leq q \leq 1. \end{cases} \quad (27)$$

Also important in the analysis of other models is the case $K = 1$ (one-step replica symmetry breaking, 1RSB)

$$x(q) = \begin{cases} 0 & \text{if } 0 \leq q < q_0, \\ m_1 & \text{if } q_0 \leq q < q_1, \\ 1 & \text{if } q_1 \leq q \leq 1. \end{cases} \quad (28)$$

The function $x(q)$ actually achieving the $\inf \mathsf{P}[x]$ has the interpretation of cumulative distribution of the overlap $Q_{\underline{x}, \underline{y}}$ for two configurations drawn from the random measure $\mu_J \times \mu_J$.

The partial differential equation (7) is easily solved for $x \in \mathcal{D}_K$ yielding an explicit expression for $\mathsf{P}[x] = \mathsf{P}_K(m, q)$. We get

$$\mathsf{P}_K(m, q) = \mathbb{E} \log Y_0 - \frac{1}{4} \beta^2 \sum_{i=1}^{K+1} m_i (q_i^2 - q_{i-1}^2), \quad (29)$$

where Y_0 is the random variable constructed as follows. Define, for X_0, \dots, X_{K+1} i.i.d. $\mathbf{N}(0, 1)$,

$$Y_{K+1} = 2 \cosh \left(B + \beta \sum_{\ell=0}^{K+1} \sqrt{q_\ell - q_{\ell-1}} X_\ell \right). \quad (30)$$

where it is understood that $q_{-1} = 0$. Then, recursively, for $\ell = K, K-1, \dots, 0$

$$Y_\ell = \{\mathbb{E}_{\ell+1}(Y_{\ell+1}^{m_{\ell+1}})\}^{1/m_{\ell+1}}. \quad (31)$$

where $\mathbb{E}_{\ell+1}$ denotes expectation with respect to $X_{\ell+1}$. In particular in the replica symmetric case ($K = 0$), we have the following function of the only remaining parameter q_0

$$P_0(q_0) = \mathbb{E} \log 2 \cosh(B + \beta \sqrt{q_0} X) + \frac{1}{4} \beta^2 (1 - q_0)^2. \quad (32)$$

We need now to prove that, for any K , and for any choice of the parameters $\{m_i\}$, $\{q_i\}$, we have $\phi_n^{\text{av}} \leq P_K(m, q)$. This is done by interpolation. Define, for $t \in [0, 1]$,

$$H_t(\underline{x}) = -\beta \sqrt{t} \sum_{i,j=1}^n J_{ij} x_i x_j - B \sum_{i=1}^n x_i - \beta \sqrt{1-t} \sum_{\ell=0}^{K+1} \sqrt{q_\ell - q_{\ell-1}} \sum_{i=1}^n G_i^\ell x_i, \quad (33)$$

with the $\{G_i^\ell\}$ i.i.d. $\mathbf{N}(0, 1)$ random variables. We then construct the partition functions $\{Z_\ell(t)\}_{0 \leq \ell \leq K+1}$ by letting

$$Z_{K+1}(t) \equiv \sum_{\underline{x}} e^{-H_t(\underline{x})}, \quad (34)$$

and then recursively for $\ell \in \{0, \dots, K\}$

$$Z_\ell(t) \equiv \{\mathbb{E}_{\ell+1}(Z_{\ell+1}^{m_{\ell+1}}(t))\}^{1/m_{\ell+1}}, \quad (35)$$

where $\mathbb{E}_{\ell+1}$ denotes expectation with respect to $\{G_i^{\ell+1}\}_{1 \leq i \leq n}$. Finally, we define

$$\Phi(t) \equiv \frac{1}{n} \mathbb{E}_0 \log Z_0(t). \quad (36)$$

Here \mathbb{E}_0 is expectation with respect to both $\{G_i^0\}_{1 \leq i \leq n}$ and $\{J_{ij}\}_{1 \leq i, j \leq n}$.

It is easy to see that $\Phi(1) = \phi_n^{\text{av}}$, since in this case $H_0(\underline{x}) = H(\underline{x})$ is just the SK energy function. Further $\Phi(0) = \mathbb{E} \log Y_0$ (with Y_0 defined as above) since

$$Z_{K+1}(t) \equiv \prod_{i=1}^n \left\{ 2 \cosh \left(B + \beta \sum_{\ell=0}^{K+1} \sqrt{q_\ell - q_{\ell-1}} G_i^\ell \right) \right\}, \quad (37)$$

is the product of n i.i.d. copies of Y_{K+1} . The proof is completed by evaluating the derivative of $\Phi(t)$ with respect to t . This takes the form

$$\Phi'(t) = -\frac{1}{4} \beta^2 \sum_{\ell=1}^{K+1} m_\ell (q_\ell^2 - q_{\ell-1}^2) - \frac{1}{4} \beta^2 \sum_{\ell=0}^K (m_{\ell+1} - m_\ell) \langle (Q(\underline{x}, \underline{y}) - q_\ell)^2 \rangle_\ell, \quad (38)$$

where $\langle \cdot \rangle_\ell$ denotes expectation with respect to an appropriate measure on $(\underline{x}, \underline{y}) \in \{+1, -1\}^n \times \{+1, -1\}^n$. Since the argument in this expectation is a perfect square, it follows that

$$\Phi(1) \leq \Phi(0) - \frac{1}{4} \beta^2 \sum_{\ell=1}^{K+1} m_\ell (q_\ell^2 - q_{\ell-1}^2), \quad (39)$$

which is our claim.

Computing the derivative (38) is not particularly difficult, just a bit laborious. For the sake of simplicity, we will consider the case $K = 0$ (replica symmetric). In that case, applying the definitions we get

$$\Phi(t) = \frac{1}{2}\beta^2(1-t)(1-q_0) + \frac{1}{n}\mathbb{E}\log\left\{\sum_{\underline{x}}e^{-\widehat{H}_t(\underline{x})}\right\}, \quad (40)$$

$$\widehat{H}_t(\underline{x}) = -\beta\sqrt{t}\sum_{i,j=1}^n J_{ij}x_ix_j - B\sum_{i=1}^n x_i - \beta\sqrt{1-t}\sqrt{q_0}\sum_{i=1}^n G_ix_i. \quad (41)$$

We then have

$$\Phi'(t) = -\frac{1}{2}\beta^2(1-q_0) + \frac{\beta}{2n\sqrt{t}}\sum_{i,j=1}^n \mathbb{E}\{J_{ij}\mu_t(x_ix_j)\} - \frac{\beta\sqrt{q_0}}{2n\sqrt{1-t}}\sum_{i=1}^n \mathbb{E}\{G_i\mu_t(x_i)\}, \quad (42)$$

where μ_t is the probability measure $\mu_t(\underline{x}) \propto \exp\{-\widehat{H}_t(\underline{x})\}$, and $\mu_t(x_ix_j)$, $\mu_t(x_i)$ denote the expectations of x_ix_j and x_i . Using Stein's Lemma, we get

$$\begin{aligned} \Phi'(t) &= -\frac{1}{2}\beta^2(1-q_0) + \frac{\beta^2}{4n^2}\sum_{i,j=1}^n \mathbb{E}\{\mu_t(x_i^2x_j^2) - \mu_t(x_ix_j)^2\} - \frac{\beta^2q_0}{2n^2}\sum_{i=1}^n \mathbb{E}\{\mu_t(x_i^2) - \mu_t(x_i)^2\} \\ &= -\frac{1}{4}\beta^2 - \frac{\beta^2}{4n^2}\sum_{i,j=1}^n \mathbb{E}\{\mu_t^{(2)}(x_iy_ix_jy_j)\} + \frac{\beta^2q_0}{2n}\sum_{i=1}^n \mathbb{E}\{\mu_t^{(2)}(x_iy_i)\} \\ &= -\frac{1}{4}\beta^2 - \frac{\beta^2}{4}\mathbb{E}\{\mu_t^{(2)}(Q(\underline{x},\underline{y})^2)\} + \frac{\beta^2}{2}q_0\mathbb{E}\{\mu_t^{(2)}(Q(\underline{x},\underline{y}))\} \end{aligned}$$

where $\mu_t^{(2)} = \mu_t \times \mu_t$ is the product distribution over $\underline{x}, \underline{y}$. By completing the square, we get

$$\Phi'(t) = -\frac{1}{4}\beta^2(1-q_0^2) - \frac{\beta^2}{4}\mathbb{E}\{\mu_t^{(2)}([Q(\underline{x},\underline{y}) - q_0]^2)\},$$

which indeed coincides with Eq. (38) for $K = 0$. □

References

- [GT02] Francesco Guerra and Fabio L. Toninelli. The thermodynamic limit in mean field spin glasses. *Commun. Math. Phys.*, 230:71–79, 2002.
- [Gue03] Francesco Guerra. Broken Replica Symmetry Bounds in the Mean Field Spin Glass Model. *Comm. Math. Phys.*, 233:1–12, 2003.