Introduction to the SK Model

Tomas Dominguez

April 2, 2021

Abstract

The celebrated Parisi formula for the Sherrington-Kirkpatrick model is one of the most important achievements in the field of disordered systems. The ideas involved in its discovery have been applied to a wealth of problems ranging from combinatorial optimization to neural networks and information theory. This talk will consist of two parts. First, I will introduce some of the basic objects and mathematical techniques involved in the study of the Sherrington-Kirkpatrick model. In the second part, I will state the famous Parisi formula, reformulate it in terms of the Ruelle probability cascades and use the Guerra replica symmetry breaking interpolation to prove the upper bound. Time permitting, I will also discuss the Aizenman-Sims-Starr scheme and the main difficulties involved in the proof of the matching lower bound.

Disclaimer
None of the material in these notes is original. All of it appears in some form or another in [4] or [5].

1 Motivation

The field of spin glasses originated in statistical physics with the goal of understanding the unusual magnetic properties of some metal alloys with competing ferromagnetic and anti-ferromagnetic interactions, such as CuMg or AuFe. The Sherrington-Kirkpatrick (SK) model was introduced in 1975 in [10] as a mean field simplification of the Ising version of the Edwards-Anderson model introduced earlier that year in [1]. However, even in the physics literature, it is often motivated as a pure optimization problem called the *Dean's problem*.

Suppose the Dean of a university is tasked with dividing a group of N students indexed by the elements of $\{1,\dots,N\}$ into two dormitories indexed by the elements of $\{-1,+1\}$. An allocation of the N students into the two dormitories can be identified with a *configuration vector*

$$\sigma = (\sigma_1,\dots,\sigma_N) \in \Sigma_N := \{-1,+1\}^N.$$

To make dormitory allocations as informed as possible, the Dean's secretary has provided the Dean with a collection of parameters g_{ij}, called *interaction parameters*, which describe how much student i likes or dislikes student j. A positive parameter means that student i likes student j, while a negative parameter means that student i dislikes student j. If we write $i \sim j$ to indicate that student i and student j are placed in the same dormitory, the Dean should try to maximize the *comfort function*

$$c(\sigma) = \sum_{i\sim j} g_{ij} - \sum_{i\not\sim j} g_{ij} = \sum_{i,j=1}^{N} g_{ij}\sigma_i\sigma_j.$$

It would be interesting to try to understand how this maximum behaves in a typical situation, and one natural way to give this question some meaning is to model the interaction parameters (g_{ij}) as

random variables. The simplest choice is to let the interactions be independent and have the standard Gaussian distribution; however, it can be shown that, in some sense, the distribution of the interaction parameters is not important. This leads us to the *Sherrington-Kirkpatrick model*.

Definition 1.1
The Hamiltonian of the *Sherrington-Kirkpatrick model* is the Gaussian process

$$H_N(\sigma) = \frac{1}{\sqrt{N}} \sum_{i,j=1}^{N} g_{ij}\sigma_i\sigma_j$$

indexed by $\sigma \in \Sigma_N$, where (g_{ij}) is a collection of i.i.d. standard Gaussian random variables. ■

In this definition we have adopted the physics language and replaced the term comfort function by Hamiltonian. Moreover, we have added a normalization factor of $\sqrt{N}$. This has been done to ensure that the maximum of the Hamiltonian is of order N, which will simplify notation. Motivated by the Dean's problem, we would like to understand the behaviour of the maximum

$$\max_{\sigma\in\Sigma_N} H_N(\sigma)$$

as the size of the system N goes to infinity or, as the physicists would say, in the *thermodynamic limit*. This is a very difficult problem, and we will now see how it can be somewhat simplified using standard Gaussian techniques.

2 The Free Energy

The first matter one should address before starting a detailed study of $\max_{\sigma\in\Sigma_N} H_N(\sigma)$ is the scale at which this random quantity should be considered. Using the independence of the standard Gaussian random variables (g_{ij}) it is easy to see that for any two configurations $\sigma^1, \sigma^2 \in \Sigma_N$

$$\mathbb{E}H_N(\sigma^1)H_N(\sigma^2) = \frac{1}{N}\sum_{i,j=1}^{N} \sigma_i^1\sigma_j^1\sigma_k^2\sigma_l^2\mathbb{E}g_{ij}g_{kl} = N\Big(\frac{1}{N}\sum_{i=1}^{N}\sigma_i^1\sigma_i^2\Big)^2 = NR_{1,2}^2, \tag{1}$$

where $R_{1,2} = \frac{1}{N}\sum_{i=1}^{N}\sigma_i^1\sigma_i^2$ is the normalized scalar product between σ^1 and σ^2, also called their *overlap*. Heuristically, this suggests that $\mathbb{E}H_N(\sigma)$ should be of order $\sqrt{N}$ and that the random variables $(H_N(\sigma))_{\sigma\in\Sigma_N}$ should not be too correlated. Since the maximum of 2^N independent Gaussian random variables with variance of order N is of order N on average, this also suggests that $\mathbb{E}\max_{\sigma\in\Sigma_N} H_N(\sigma)$ should be of order N. Remembering that Gaussian random variables, as well as Lipschitz functions of these, concentrate around their expectation, we are led to believe that $\max_{\sigma\in\Sigma_N} H_N(\sigma)$ should be normalized by N and should behave like $\mathbb{E}\max_{\sigma\in\Sigma_N} H_N(\sigma)$. To make this heuristic discussion more rigorous, let us use the Gaussian concentration inequality to show that $\frac{1}{N}\max_{\sigma\in\Sigma_N} H_N(\sigma)$ concentrates about its expectation. We will also show that the expectation of the maximum of 2^N Gaussian random variables with variance of order N is of order no greater than N even if the random variables are not assumed to be independent.

Theorem 2.1 (Gaussian concentration inequality)
If $g = (g_i)_{i\le n}$ is a standard Gaussian vector on $\mathbb{R}^n$ and $F : \mathbb{R}^n \to \mathbb{R}$ is a Lipschitz function, then for any $t \ge 0$

$$\mathbb{P}\{|F(g) - \mathbb{E}F(g)| \ge t\} \le 2\exp\left(-\frac{t^2}{4\|F\|_{\mathrm{Lip}}^2}\right).$$

Proof. For a proof of this standard result see [7, theorem 1.1]. ■

Corollary 2.2 (Concentration of Gaussian maxima)
If $(g_i)_{i\le n}$ are centred Gaussian random variables with $\mathbb{E}g_i^2 \le a$ for $1 \le i \le n$, then for each $t \ge 0$

$$\mathbb{P}\{|\max_{1\le i\le n} g_i - \mathbb{E}\max_{1\le i\le n} g_i| \ge t\} \le 2\exp\Big(-\frac{t^2}{4a}\Big). \tag{2}$$

In particular,

$$\lim_{N\to\infty}\Big|\frac{1}{N}\max_{\sigma\in\Sigma_N} H_N(\sigma) - \frac{1}{N}\mathbb{E}\max_{\sigma\in\Sigma_N} H_N(\sigma)\Big| = 0.$$

Proof. Let $A^T = (a_i)_{i\le n} \in \mathbb{R}^{n\times n}$ be a matrix with $AA^T = (\mathbb{E}g_i g_j)_{1\le i,j\le n}$. If g is a standard Gaussian vector on $\mathbb{R}^n$, then $(g_i)_{i\le n} \overset{d}{=} Ag$. In particular,

$$\max_{1\le i\le n} g_i \overset{d}{=} \max_{1\le i\le n}(Ag)_i = \max_{1\le i\le n} a_i\cdot g = F(g)$$

for the function $F:\mathbb{R}^n\to\mathbb{R}$ defined by $F(g) = \max_{1\le i\le n} a_i\cdot g$. Equation (2) is now immediate from the Gaussian concentration inequality provided we can show that F is Lipschitz with $\|F\|_{\text{Lip}} \le \sqrt{a}$. To do so, fix $g^1, g^2 \in \mathbb{R}^n$ and suppose without loss of generality that $F(g^1) \ge F(g^2)$. If $1 \le j \le n$ is such that $a_j\cdot g^1 = F(g^1)$, then the Cauchy-Schwarz inequality and the fact that $|a_j|^2 = \mathbb{E}g_j^2 \le a$ give

$$|F(g^1) - F(g^2)| = a_j\cdot g^1 - \max_{1\le i\le n} a_i\cdot g^2 \le a_j\cdot(g^1 - g^2) \le \sqrt{a}|g^1 - g^2|$$

as required. Finally, equation (2) implies that for any $\varepsilon > 0$

$$\mathbb{P}\Big\{\Big|\frac{1}{N}\max_{\sigma\in\Sigma_N} H_N(\sigma) - \frac{1}{N}\mathbb{E}\max_{\sigma\in\Sigma_N} H_N(\sigma)\Big| \ge \varepsilon\Big\} \le 2\exp\Big(-\frac{\varepsilon^2}{4}N\Big).$$

Since the right-hand side of this equation is summable in N, the result follows by a simple application of the Borel-Cantelli lemma. ■

Proposition 2.3 (Order of Gaussian maxima)
If $(g(\sigma))$ are centred Gaussian random variables indexed by $\sigma \in \Sigma_N$ with $\mathbb{E}g(\sigma)^2 \le N$, then

$$0 \le \mathbb{E}\max_{\sigma\in\Sigma_N} g(\sigma) \le \sqrt{\log 4}N.$$

Proof. Fix $\beta > 0$, and observe that Jensen's inequality together with the explicit formula for the moment generating function of a Gaussian random variable give

$$\begin{aligned}\mathbb{E}\max_{\sigma\in\Sigma_N} g(\sigma) &= \frac{1}{\beta}\log\exp\mathbb{E}\beta\max_{\sigma} g(\sigma) \le \frac{1}{\beta}\log\mathbb{E}\exp\beta\max_{\sigma} g(\sigma)\\ &\le \frac{1}{\beta}\log\sum_{\sigma\in\Sigma_N}\mathbb{E}\exp\beta g(\sigma) \le \frac{1}{\beta}\log 2^N\exp(\beta^2 N/2)\\ &= \frac{N\log 2}{\beta} + \frac{\beta N}{2}.\end{aligned}$$

Optimizing over $\beta > 0$ to find $\beta^2 = 2\log 2$ yields

$$0 \le \mathbb{E}\max_{\sigma\in\Sigma_N} g(\sigma) \le \sqrt{\log 4}N.$$

This completes the proof. ■

We have now reduced the Dean's problem to understanding the precise asymptotics

$$\lim_{N\to\infty}\frac{1}{N}\mathbb{E}\max_{\sigma\in\Sigma_N}H_N(\sigma). \tag{3}$$

One standard approach to this random optimization problem, which is hinted at by our proof of proposition 2.3, is to think of it as the zero-temperature case of a general family of problems at positive temperature. That is, instead of dealing with (3) directly, we should first try to compute the limit of its 'smooth approximation'

$$\lim_{N\to\infty}\frac{1}{N\beta}\mathbb{E}\log\sum_{\sigma\in\Sigma_N}\exp\beta H_N(\sigma)$$

for every *inverse temperature parameter* $\beta>0$. The validity of this strategy is a consequence of the following elementary bounds.

Proposition 2.4 (Regularizing the maximum)
If $(g(\sigma))$ are random variables indexed by $\sigma\in\Sigma_N$, then for every $\beta>0$

$$\Big|\frac{1}{N}\mathbb{E}\max_{\sigma\in\Sigma_N}g(\sigma)-\frac{1}{N\beta}\mathbb{E}\log\sum_{\sigma\in\Sigma_N}\exp\beta g(\sigma)\Big|\le\frac{\log 2}{\beta}.$$

Proof. It is clear that for any $\beta>0$

$$\frac{1}{N}\mathbb{E}\max_{\sigma\in\Sigma_N}g(\sigma)=\frac{1}{N\beta}\mathbb{E}\log\exp\beta\max_{\sigma\in\Sigma_N}g(\sigma)\le\frac{1}{N\beta}\mathbb{E}\log\sum_{\sigma\in\Sigma_N}\exp\beta g(\sigma)$$

and

$$\frac{1}{N\beta}\mathbb{E}\log\sum_{\sigma\in\Sigma_N}\exp\beta g(\sigma)\le\frac{1}{N\beta}\mathbb{E}\log\sum_{\sigma\in\Sigma_N}\exp\beta\max_{\sigma\in\Sigma_N}g(\sigma)\le\frac{\log(2)}{\beta}+\frac{1}{N}\mathbb{E}\max_{\sigma\in\Sigma_N}g(\sigma).$$

This completes the proof. ■

This result motivates the introduction of the *partition function* and the *free energy* of the Sherrington-Kirkpatrick model.

Definition 2.5
The *partition function* and the *free energy* of the Sherrington-Kirkpatrick model are the random variables

$$Z_N(\beta)=\sum_{\sigma\in\Sigma_N}\exp\beta H_N(\sigma)\text{ and }F_N(\beta)=\frac{1}{N}\mathbb{E}\log Z_N(\beta)$$

respectively. ■

The Dean's problem has now been reduced further to understanding the limit of the free energy

$$\lim_{N\to\infty}F_N(\beta)$$

at every positive temperature $\beta^{-1}>0$. Indeed, if it were possible to prove the existence of $F(\beta)=\lim_{N\to\infty}F_N(\beta)$, then the estimate in proposition 2.4 together with a simple application of Hölder's inequality would allow us to deduce the value of (3).

Proposition 2.6 (From free energy to maximum)
If the limit $F(\beta) = \lim_{N\to\infty} F_N(\beta)$ exists for every $\beta > 0$, then

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\max_{\sigma\in\Sigma_N} H_N(\sigma) = \lim_{\beta\to\infty} \frac{F(\beta)}{\beta}.$$

Proof. By proposition 2.4, for every $\beta > 0$ we have

$$\Big| \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\max_{\sigma\in\Sigma_N} H_N(\sigma) - \frac{F(\beta)}{\beta} \Big| \leq \frac{\log 2}{\beta}. \tag{4}$$

On the other hand, a simple application of Hölder's inequality shows that the function $\beta \mapsto \beta^{-1}(F_N(\beta) - \log(2))$ is increasing for every $N \in \mathbb{N}$. Indeed, if $\beta_1 \leq \beta_2$ then Hölder's inequality for sums yields

$$\begin{aligned}
\beta_1^{-1}(F_N(\beta_1) - \log(2)) &= \frac{1}{N\beta_1}\mathbb{E}\log\frac{1}{2^N}\sum_{\sigma\in\Sigma_N}\exp\beta_1 H_N(\sigma) \\
&\leq \frac{1}{N\beta_1}\mathbb{E}\log\frac{1}{2^N}\Big(\sum_{\sigma\in\Sigma_N}\exp\beta_2 H_N(\sigma)\Big)^{\frac{\beta_1}{\beta_2}} 2^{N(1-\beta_1/\beta_2)} \\
&= \frac{1}{N\beta_2}\mathbb{E}\log\frac{1}{2^N}\sum_{\sigma\in\Sigma_N}\exp\beta_2 H_N(\sigma) \\
&= \beta_2^{-1}(F_N(\beta_2) - \log(2)).
\end{aligned}$$

Letting $N \to \infty$ shows that the function $\beta \mapsto \beta^{-1}(F(\beta) - \log(2))$ is increasing and establishes the existence of $\lim_{\beta\to\infty} \beta^{-1}F(\beta)$. Combining this with equation (4) completes the proof. ■

From here on out, we will forget about the maximum of the Hamiltonian $H_N(\sigma)$ and instead focus on computing the limit of the free energy $F_N(\beta)$. We will often work at a fixed temperature β^{-1} and omit the dependence on the parameter β.

3 The Gibbs Measure

In statistical mechanics, the free energy $F_N(\beta)$ is closely related to another object- the *Gibbs measure*, which is interpreted as the probability of observing the system at temperature β^{-1} in a particular configuration $\sigma \in \Sigma_N$.

Definition 3.1
The *Gibbs measure* of the Sherrington-Kirkpatrick model is the random probability measure on Σ_N defined by

$$G_N(\sigma) = \frac{\exp\beta H_N(\sigma)}{Z_N(\beta)}.$$

The average with respect to the product measure $G_N^{\otimes\infty}$ is denoted by $\langle\cdot\rangle$. ■

Heuristically, the Gibbs measure assigns more weight to the configurations $\sigma \in \Sigma_N$ corresponding to larger values of the Hamiltonian $H_N(\sigma)$, so in the thermodynamic limit the asymptotic Gibbs measure should carry most of its weight on sets where the Hamiltonian takes its largest values. Understanding the Gibbs measure, and in particular the structure of its support, should therefore be useful in computing the limit of the free energy.

Another way to relate the free energy and the Gibbs measure is to notice that the free energy can be thought of as a moment generating function for the distribution of the Hamiltonian under the average Gibbs measure $\mathbb{E}G^{\otimes\infty}$. Indeed, a simple computation shows that

$$\frac{\mathrm{d}}{\mathrm{d}\beta}F_N(\beta) = \frac{1}{N}\mathbb{E}\sum_{\sigma\in\Sigma_N} H_N(\sigma)\frac{\exp H_N(\sigma)}{Z_N(\beta)} = \frac{1}{N}\mathbb{E}\langle H_N(\sigma)\rangle,$$

and similarly

$$\frac{\mathrm{d}^2}{\mathrm{d}\beta^2}F_N(\beta) = \frac{1}{N}\Big(\mathbb{E}\langle H_N(\sigma)^2\rangle - \mathbb{E}\langle H_N(\sigma)\rangle^2\Big).$$

In words, these identities say that the first derivative of the free energy gives the first moment of $H_N(\sigma)$ under the measure $\mathbb{E}G^{\otimes\infty}$, while the second derivative of the free energy corresponds to the second centred moment of $H_N(\sigma)$ under $\mathbb{E}G^{\otimes\infty}$. Of course, higher order derivatives of the free energy give information about higher moments of $H_N(\sigma)$ under $\mathbb{E}G^{\otimes\infty}$. In particular, if we were able to describe the distribution of $H_N(\sigma)$ under $\mathbb{E}G^{\otimes\infty}$, we should be able to gain a lot of information about the free energy $F_N(\beta)$. Since the Hamiltonian $H_N(\sigma)$ depends on the configurations $\sigma\in\Sigma_N$ only through the overlaps $R_{l,l'}$ between configurations $\sigma^l,\sigma^{l'}\in\Sigma_N$, it should actually be enough to understand the distribution under $\mathbb{E}G_N^{\otimes\infty}$ of the infinite random array $(R_{l,l'})_{l,l'\geq 1}$ obtained by sampling i.i.d. *replicas* $\sigma^l\in\Sigma_N$ from the Gibbs measure $G_N^{\otimes\infty}$. To reinforce this idea we will now perform a *cavity computation* that will lead us to the so-called *Aizenman-Sims-Starr representation* and provide a direct link between the free energy and the Gibbs measure.

For $j\geq 0$ let us introduce the quantity $A_j = \mathbb{E}\log Z_{j+1} - \mathbb{E}\log Z_j$ with the convention that $Z_0 = 1$. The free energy may be expressed as the telescopic sum

$$F_N = \frac{1}{N}\mathbb{E}\log Z_N = \frac{1}{N}\sum_{j=0}^{N-1} A_j.$$

Clearly, this representation implies that if the sequence A_N converges, then its limit is also the limit of the free energy F_N. Unfortunately, we will not be able to prove the existence of the limit of A_N in general. Nonetheless, this representation will give us a lot of insight into the form of the limiting free energy. Moreover, it immediately gives us the lower bound

$$\liminf_{N\to\infty} F_N \geq \liminf_{N\to\infty} A_N.$$

The advantage of looking at the quantity A_N over directly studying the free energy F_N is that the former involves the difference between the partition function Z_N of the model with N coordinates and the partition function Z_{N+1} of the model with an additional *cavity coordinate*. To compare these two quantities, let us write $\rho = (\sigma,\varepsilon)\in\Sigma_{N+1}$ for $\sigma\in\Sigma_N$ and $\varepsilon\in\{-1,1\}$. It is clear that

$$H_{N+1}(\rho) = H'_N(\sigma) + \varepsilon z_N(\sigma) + \frac{1}{\sqrt{N+1}}g_{(N+1)(N+1)}, \tag{5}$$

where we have introduced the Gaussian processes

$$H'_N(\sigma) = \frac{1}{\sqrt{N+1}}\sum_{i,j=1}^{N} g_{ij}\sigma_i\sigma_j$$

$$z_N(\sigma) = \frac{1}{\sqrt{N+1}}\sum_{i=1}^{N}(g_{i(N+1)} + g_{(N+1)i})\sigma_i.$$

On the other hand, the Gaussian process $H_N(\sigma)$ can be decomposed as the sum of two independent Gaussian processes

$$H_N(\sigma) \stackrel{d}{=} H'_N(\sigma) + y_N(\sigma), \tag{6}$$

where $(y_N(\sigma))$ is a Gaussian process with covariance structure

$$\mathbb{E}y_N(\sigma^1)y_N(\sigma^2) = \frac{N}{N+1}R_{1,2}^2$$

independent of all other sources of randomness. An explicit example of such a Gaussian process is

$$y_N(\sigma) = \frac{1}{\sqrt{N(N+1)}}\sum_{i,j=1}^{N} g'_{ij}\sigma_i\sigma_j$$

for some i.i.d. standard Gaussian random variables (g'_{ij}) independent of (g_{ij}). Combining the decompositions (5) and (6) yields

$$\begin{aligned} A_N &= \mathbb{E}\log\sum_{\sigma\in\Sigma_N}\sum_{\varepsilon\in\{-1,+1\}}\exp\beta\Big(H'_N(\sigma)+\varepsilon z_N(\sigma)+(N+1)^{-1/2}g_{(N+1)(N+1)}\Big) \\ &\quad -\mathbb{E}\log\sum_{\sigma\in\Sigma_N}\exp\beta\Big(H'_N(\sigma)+y_N(\sigma)\Big) \\ &= \mathbb{E}\log\sum_{\sigma\in\Sigma_N}2\cosh(\beta z_N(\sigma))\exp\beta H'_N(\sigma) - \mathbb{E}\log\sum_{\sigma\in\Sigma_N}\exp(\beta y_N(\sigma))\exp\beta H'_N(\sigma). \end{aligned}$$

If we now denote by $\langle\cdot\rangle'$ the Gibbs average corresponding to the Gibbs measure associated with the Hamiltonian $H'_N(\sigma)$ on Σ_N, this can be written as

$$A_N = \mathbb{E}\log\Big\langle 2\cosh\beta z_N(\sigma)\Big\rangle' - \mathbb{E}\log\Big\langle\exp\beta y_N(\sigma)\Big\rangle'. \tag{7}$$

This equation is known as the *Aizenman-Sims-Starr representation*. Observe that the Gaussian processes $z_N(\sigma)$ and $y_N(\sigma)$ are independent of the randomness of the measure G'_N and their covariance structure

$$\mathbb{E}z_N(\sigma^1)z_N(\sigma^2) = 2\frac{N}{N+1}R_{1,2} \quad\text{and}\quad \mathbb{E}y_N(\sigma^1)y_N(\sigma^2) = \frac{N}{N+1}R_{1,2}^2$$

depends on the spin configurations $\sigma\in\Sigma_N$ only through the overlap array $(R^N_{l,l'})_{l,l'\geq 1}$. If we overlook the fact that the average $\langle\cdot\rangle'$ is with respect to a different Gibbs measure than the one we started with, the Aizenman-Sims-Starr representation gives a very clear connection between the limit of the free energy and the Gibbs measure. Moreover, it suggests that to compute the limit of the free energy (or at least bound it from below) we should study the distributional limit of the overlap array $(R^N_{l,l'})_{l,l'\geq 1}$ of the model on Σ_N under the measure $\mathbb{E}G'^{\otimes\infty}_N$. We will say more about this in section 6 when we will have stated the Parisi formula and discussed its representation in terms of the Ruelle probability cascades. Now, let us show how the Gibbs measure and standard Gaussian techniques can be used to prove the existence of the limit of the free energy.

4 Existence of the Limiting Free Energy

To prove the existence of the limit of the free energy in the Sherrington-Kirkpatrick model, we will use a *Gaussian interpolation* argument based on a simple consequence of the *Gaussian integration by parts formula*.

Theorem 4.1 (Gaussian integration by parts)
If g is a Gaussian vector in $\mathbb{R}^n$ and $\mathbb{E}\left|\frac{\partial F}{\partial x_l}(g)\right| < \infty$ or $\mathbb{E}[g_1 g_l] = 0$ for all $l \leq n$, then

$$\mathbb{E}g_1F(g) = \sum_{1\leq l\leq n}\mathbb{E}g_1g_l\mathbb{E}\frac{\partial F}{\partial x_l}(g).$$

Proof. For a proof of this standard result see [7, lemma 10.2] ■

In the study of spin glasses, the Gaussian integration by parts formula is most typically used to compute averages with respect to the Gibbs measure. To see how this can be done in the greatest possible generality, let us consider two jointly Gaussian vectors $(x(\sigma))$ and $(y(\sigma))$ indexed by some countably infinite set Σ on which is defined a measure G. Consider the random measure G' on Σ defined by the change of density

$$G'(\sigma) = \frac{\exp y(\sigma)}{Z} G(\sigma),$$

where $Z = \sum_{\sigma\in\Sigma} \exp(y(\sigma))G(\sigma)$ is a random normalization factor, and denote by $\langle\cdot\rangle$ the average with respect to the product measure $G'^{\otimes\infty}$.

Corollary 4.2 (Gaussian integration by parts for Gibbs averages)
If the jointly Gaussian processes $(x(\sigma))$ and $(y(\sigma))$ satisfy $\mathbb{E}x(\sigma)^2, \mathbb{E}y(\sigma)^2 \le a$ for all $\sigma \in \Sigma$ and some $a > 0$, then for every bounded and measurable function $\Phi = \Phi(\sigma^1, \ldots, \sigma^n)$

$$\mathbb{E}\langle \Phi x(\sigma^1)\rangle = \mathbb{E}\Big\langle \Phi\Big(\sum_{1\le i\le n} C(\sigma^1, \sigma^i) - nC(\sigma^1, \sigma^{n+1})\Big)\Big\rangle, \tag{8}$$

where $C(\sigma^1, \sigma^2) = \mathbb{E}x(\sigma^1)y(\sigma^2)$.

Proof. By elementary estimates detailed in [4, lemma 1.5], it suffices to prove (8) when G is supported on finitely many points. To do so, let us focus on a single term in the sum

$$\mathbb{E}\langle \Phi x(\sigma^1)\rangle = \sum_{\sigma^1,\ldots,\sigma^n\in\Sigma} \Phi(\sigma^1, \ldots, \sigma^n)\mathbb{E}x(\sigma^1)G'(\sigma^1)\cdots G'(\sigma^n).$$

Introduce the Gaussian vector

$$g = (x(\sigma^1), y(\sigma^1), \ldots, y(\sigma^n), (y(\sigma^{n+1}))_{\sigma^{n+1}\in\Sigma})$$

as well as the bounded function

$$F(g) = G'(\sigma^1)\cdots G'(\sigma^n) = \frac{\exp(y(\sigma^1))\cdots\exp(y(\sigma^n))}{(\sum_{\sigma^{n+1}\in\Sigma}\exp(y(\sigma^{n+1}))G(\sigma^{n+1}))^n} G(\sigma^1)\cdots G(\sigma^n).$$

Observe that for $1 \le i \le n$

$$\frac{\partial F}{\partial y(\sigma^i)} = G'(\sigma^1)\cdots G'(\sigma^n)$$

while for any $\sigma^{n+1} \in \Sigma$

$$\frac{\partial F}{\partial y(\sigma^{n+1})} = -nG'(\sigma^1)\cdots G'(\sigma^{n+1}).$$

Invoking the Gaussian integration by parts formula gives

$$\begin{aligned}\mathbb{E}x(\sigma^1)G'(\sigma^1)\cdots G'(\sigma^n) = &\sum_{1\le i\le n} C(\sigma^1, \sigma^i)\mathbb{E}G'(\sigma^1)\cdots G'(\sigma^n) \\ &-n\sum_{\sigma^{n+1}\in\Sigma} C(\sigma^1, \sigma^{n+1})\mathbb{E}G'(\sigma^1)\cdots G'(\sigma^{n+1}).\end{aligned}$$

Summing over $\sigma^1,\dots,\sigma^n \in \Sigma$ yields

$$\begin{aligned}\mathbb{E}\langle \Phi x(\sigma^1)\rangle &= \mathbb{E}\sum_{\sigma^1,\dots,\sigma^n\in\Sigma} \Phi \sum_{1\le i\le n} C(\sigma^1,\sigma^i)G'(\sigma^1)\cdots G'(\sigma^n)\\ &\quad -\mathbb{E}\sum_{\sigma^1,\dots,\sigma^{n+1}\in\Sigma} n\Phi C(\sigma^1,\sigma^{n+1})G'(\sigma^1)\cdots G'(\sigma^{n+1})\\ &= \mathbb{E}\Big\langle \Phi\Big(\sum_{1\le i\le n} C(\sigma^1,\sigma^i) - nC(\sigma^1,\sigma^{n+1})\Big)\Big\rangle\end{aligned}$$

as required. ■

Before we prove the existence of the limit of the free energy let us recall the classical Fekete lemma. Remember that a sequence $(x_n)_{n=1}^\infty$ is said to be *superadditive* if $x_n + x_m \le x_{n+m}$ for all $n,m \ge 1$.

Lemma 4.3 (Fekete lemma)
If $(x_n)_{n=1}^\infty$ is a superadditive sequence then

$$\lim_{n\to\infty}\frac{x_n}{n} = \sup_{n\ge 1}\frac{x_n}{n}.$$

Proof. It is clear that $\limsup_{n\to\infty}\frac{x_n}{n} \le \sup_{n\ge1}\frac{x_n}{n}$. To prove the reverse inequality fix two integers $n \ge m$ and let $n = mk + r$ be the Euclidean division of n by m. The subadditivity of $(x_n)_{n=1}^\infty$ together with the fact that $0 \le r < m$ reveal that

$$\frac{x_n}{n} \ge \frac{kx_m + x_r}{mk+r} \ge \frac{kx_m + \inf_{0\le r\le m} x_r}{mk+m}.$$

For a fixed $m \in \mathbb{N}$, we have that n tends to infinity if and only if k does. It follows that

$$\liminf_{n\to\infty}\frac{x_n}{n} \ge \liminf_{k\to\infty}\frac{kx_m + \inf_{0\le r\le m} x_r}{mk+m} = \frac{x_m}{m},$$

and taking the supremum over $m \in \mathbb{N}$ completes the proof. ■

Theorem 4.4 (Guerra-Toninelli)
The limit $\lim_{N\to\infty} F_N$ exists.

Proof. By the Fekete lemma it suffices to prove that the sequence $(NF_N)_{N=1}^\infty$ is superadditive. To this end, fix $N,M \in \mathbb{N}$ and consider the three independent Sherrington-Kirkpatrick Hamiltonians $H_N(\rho)$, $H_M(\tau)$ and $H_{N+M}(\sigma)$ on Σ_N, Σ_M and $\Sigma_{N+M} = \Sigma_N \times \Sigma_M$ respectively. Write $\sigma = (\rho,\tau) \in \Sigma_{N+M}$, and for each $t \in [0,1]$ define the interpolating Hamiltonian $H_t(\sigma)$ on Σ_{N+M} by

$$H_t(\sigma) = \sqrt{t}H_{N+M}(\sigma) + \sqrt{1-t}(H_N(\rho) + H_M(\tau)).$$

Associated to this Hamiltonian are the interpolating partition function Z_t, the interpolating Gibbs average $\langle\cdot\rangle_t$ and the interpolating free energy

$$\varphi(t) = \frac{1}{N+M}\mathbb{E}\log Z_t.$$

Observe that $\varphi(0) = \frac{N}{N+M}F_N + \frac{M}{N+M}F_M$ and $\varphi(1) = F_{N+M}$, so superadditivity of the sequence $(NF_N)_{N=1}^\infty$ is equivalent to the fact that $\varphi(0) \le \varphi(1)$. It therefore suffices to prove that $\varphi'(t) \ge 0$ for all $t \in (0,1)$. The Gaussian integration by parts formula in corollary 4.2 yields

$$\varphi'(t) = \frac{1}{N+M}\mathbb{E}\Big\langle \frac{\partial H_t}{\partial t}(\sigma)\Big\rangle_t = \frac{1}{N+M}\mathbb{E}\left\langle C(\sigma^1,\sigma^1) - C(\sigma^1,\sigma^2)\right\rangle_t,$$

where

$$\begin{aligned} C(\sigma^1,\sigma^2) &= \mathbb{E}\frac{\partial H_t}{\partial t}(\sigma^1)H_t(\sigma^2) \\ &= \mathbb{E}\left(\frac{1}{2\sqrt{t}}H_{N+M}(\sigma^1) - \frac{1}{2\sqrt{1-t}}(H_N(\rho^1)+H_N(\tau^1))\right)H_t(\sigma^2) \\ &= \frac{1}{2}\mathbb{E}H_{N+M}(\sigma^1)H_{N+M}(\sigma^2) - \frac{1}{2}\mathbb{E}H_N(\rho^1)H_N(\rho^2) - \frac{1}{2}\mathbb{E}H_M(\tau^1)H_M(\tau^2). \end{aligned}$$

If we write $R_{1,2} = \frac{\sigma^1\cdot\sigma^2}{N+M}$ for the overlap on Σ_{N+M}, $R^-_{1,2} = \frac{\rho^1\cdot\rho^2}{N}$ for the overlap on Σ_N and $R^+_{1,2} = \frac{\tau^1\cdot\tau^2}{M}$ for the overlap on Σ_M, then (1) implies that

$$C(\sigma^1,\sigma^2) = \frac{1}{2}\left[(N+M)R^2_{1,2} - N(R^-_{1,2})^2 - M(R^+_{1,2})^2\right].$$

Since $R_{1,1} = R^-_{1,1} = R^+_{1,1} = 1$ it follows that

$$\varphi'(t) = -\frac{1}{2}\mathbb{E}\Big\langle R^2_{1,2} - \frac{N}{N+M}(R^-_{1,2})^2 - \frac{M}{N+M}(R^+_{1,2})^2\Big\rangle_t.$$

Noting that

$$R_{1,2} = \frac{N}{N+M}\xi(R^1_{1,2}) + \frac{M}{N+M}\xi(R^2_{1,2})$$

and invoking the convexity of the function $x \mapsto x^2$ completes the proof. ■

Although this result proves the existence of the limit $\lim_{N\to\infty} F_N$, it provides absolutely no insight into its value. This is the purpose of the Parisi formula which we now discuss.

5 The Parisi Formula

A formula for the limit of the free energy $F(\beta)$ was proposed by Sherrington and Kirkpatrick in their original paper [10] based on the so-called replica formalism. In this same paper, they observed that their replica symmetric solution exhibits 'unphysical behavior' at low temperature, which means that it can only be correct at high temperature. The correct formula for $F(\beta)$ at all temperatures was famously discovered by Parisi in 1979 in his seminal papers [8] and [9]. In this section we will describe this formula. To alleviate notation we will fix $\beta > 0$ and denote by $\mathscr{D}[0,1]$ the set of discrete probability measures on $[0,1]$. More precisely, we will say that a probability measure $\zeta \in \Pr[0,1]$ is a *discrete functional order parameter*, and write $\zeta \in \mathscr{D}[0,1]$, if there exist $r \geq 1$ and two sequences of parameters

$$0 = \zeta_{-1} < \zeta_0 < \zeta_1 < \ldots < \zeta_{r-1} < \zeta_r = 1 \tag{9}$$

$$0 = q_0 < q_1 < \ldots < q_{r-1} < q_r = 1 \tag{10}$$

with $\zeta(\{q_p\}) = \zeta_p - \zeta_{p-1}$ for $0 \leq p \leq r$. The Parisi formula for the limit of the free energy will be expressed in terms of a variational problem involving the set $\mathscr{D}[0,1]$ as well as the so-called *Parisi functional*. There are many equivalent formulations of the Parisi functional, but we will opt for a recursive definition in terms of the *Parisi sequence*.

Definition 5.1
Given a discrete functional order parameter $\zeta \in \mathscr{D}[0,1]$, the *Parisi sequence* $(X^\zeta_l)_{0\leq l\leq r}$ is the sequence

of random variables defined recursively as follows. Given i.i.d. standard Gaussian random variables $(\eta_p)_{1\le p\le r}$ let

$$X_r^\zeta((\eta_p)_{0\le p\le r}) = \log 2\cosh\Big(\sum_{1\le p\le r}\sqrt{2}\beta(q_p-q_{p-1})^{1/2}\eta_p\Big),$$

and for $0\le l\le r-1$ recursively define $X_l^\zeta((\eta_p)_{1\le p\le l}) = \frac{1}{\zeta_l}\log\mathbb{E}_{\eta_{l+1}}\exp\zeta_l X_{l+1}^\zeta((\eta_p)_{1\le p\le l+1})$. ■

Definition 5.2
The *Parisi functional* is the functional $\mathscr{P}:\mathscr{D}[0,1]\to\mathbb{R}$ defined by

$$\mathscr{P}(\zeta) = X_0^\zeta - \frac{\beta^2}{2}\sum_{0\le p\le r-1}(q_{p+1}^2-q_p^2)\zeta_p = X_0^\zeta - \beta^2\int_0^1 t\zeta(t)\mathrm{d}t,$$

where we have identified ζ with its cumulative distribution function. ■

Theorem 5.3 (The Parisi formula)
The limit of the free energy in the Sherrington-Kirkpatrick model is given by

$$\lim_{N\to\infty}F_N = \inf_\zeta\mathscr{P}(\zeta),$$

where the infimum is taken over all discrete functional order parameters $\zeta\in\mathscr{D}[0,1]$.

Before we discuss the Parisi formula any further, let us mention that the first rigorous proof of this result was given in 2006 by Talagrand in his seminal paper [11]. A more robust proof of the Parisi formula which carried out the scheme described in section 3 was given in [6] by Panchenko following his famous proof of the *Parisi ultrametricity conjecture* in [3]. In this short note we will not even come close to giving a proof of the Parisi formula. Instead, we refer the interested reader to Panchenko's book [4]. We will however, describe a representation of the Parisi functional in terms of the *Ruelle probability cascades* which will be conspicuously similar to equation (7) and use it to prove that the Parisi formula gives an upper bound for the limit of the free energy.

6 The Ruelle Probability Cascades

In this section we will describe a class of random measures on a separable Hilbert space H known as the Ruelle probability cascades. The points and weights of these measures will be indexed by $\mathbb{N}^r$ for some fixed $r\ge 1$. It will be very convenient to think of $\mathbb{N}^r$ as the set of leaves of a rooted tree with vertex set

$$\mathscr{A} = \mathbb{N}^0\cup\mathbb{N}\cup\mathbb{N}^2\cup\ldots\cup\mathbb{N}^r,$$

where $\mathbb{N}^0=\{\emptyset\}$. In this identification $\emptyset$ is the root of the tree and each vertex $\alpha=(n_1,\ldots,n_p)\in\mathbb{N}^p$ for $p\le r-1$ has children

$$\alpha n = (n_1,\ldots,n_p,n)\in\mathbb{N}^{p+1}$$

for all $n\in\mathbb{N}$. Therefore, each vertex $\alpha=(n_1,\ldots,n_p)\in\mathbb{N}^p$ is connected to the root $\emptyset$ by the path

$$\emptyset\to n_1\to(n_1,n_2)\to\cdots\to(n_1,\ldots,n_p)=\alpha$$

whose vertices, omitting the root, we will denote by

$$p(\alpha) = \{n_1,(n_1,n_2),\ldots,(n_1,\ldots,n_p)\}.$$

For each vertex $\alpha\in\mathscr{A}$ we will write $|\alpha|$ for its distance from the root of the tree $\emptyset$, or equivalently for the number of coordinates in α. Moreover, for any two leaves $\alpha,\beta\in\mathbb{N}^r$ we will denote by

$\alpha \wedge \beta = |p(\alpha) \cap p(\beta)|$ the number of common vertices in the paths from the root $\emptyset$ to the leaves α and β. The identification of the index set $\mathbb{N}^r$ with the leaves of the infinitary tree $\mathscr{A}$ is very important: although the points in the support of the random measure we will construct will be indexed by $\alpha \in \mathbb{N}^r$, the construction itself will involve random variables indexed by vertices of the entire tree. A *Ruelle probability cascade* will be determined by two sequences of parameters

$$0 = \zeta_{-1} < \zeta_0 < \zeta_1 < \ldots < \zeta_{r-1} < \zeta_r = 1 \tag{11}$$

$$0 = q_0 < q_1 < \ldots < q_{r-1} < q_r < \infty \tag{12}$$

which we now fix. The parameters (11) will be used to define the weights of the random measure while the sequence (12) will determine its support. These two sequences of parameters can be encoded by a single discrete functional order parameter $\zeta \in \mathscr{D}[0,1]$ defined by

$$\zeta(\{q_p\}) = \zeta_p - \zeta_{p-1} \tag{13}$$

for $0 \leq p \leq r$.

Since it is an easier task, let us begin by describing the support of the Ruelle probability cascade with functional order parameter ζ. For conciseness of notation we will write $\mathrm{RPC}(\zeta)$ for the Ruelle probability cascade with functional order parameter ζ defined by (13).

Definition 6.1
Let (e_α) be an orthonormal sequence in a separable Hilbert space H indexed by $\alpha \in \mathscr{A} \setminus \mathbb{N}^0$. The *support of the Ruelle probability cascade* $\mathrm{RPC}(\zeta)$ is the subset (h_α) of H indexed by $\alpha \in \mathbb{N}^r$and defined by

$$h_\alpha = \sum_{\beta \in p(\alpha)} e_\beta (q_{|\beta|} - q_{|\beta|-1})^{1/2}$$

for each $\alpha \in \mathbb{N}^r$. ■

The dependence of the Ruelle probability cascades on the orthonormal sequence (e_α) and the Hilbert space H will be kept implicit since, in some sense, we will only be interested in this measure up to orthogonal transformations. In fact, our main interest will be in the distribution of the scalar products of points sampled from this measure which of course does not depend on the specific choice of (e_α). A first indication that this distribution should be related to the functional order parameter ζ is the following basic, yet crucial, observation.

Proposition 6.2 (RPC overlaps)
For all $\alpha, \beta \in \mathbb{N}^r$ we have $h_\alpha \cdot h_\beta = q_{\alpha \wedge \beta}$.

Proof. By orthogonality of the sequence (e_α) and the definition of $\alpha \wedge \beta$

$$\begin{aligned} h_\alpha \cdot h_\beta &= \sum_{\varepsilon \in p(\alpha)} \sum_{\delta \in p(\beta)} e_\varepsilon \cdot e_\delta (q_{|\varepsilon|} - q_{|\varepsilon|-1})^{1/2} (q_{|\delta|} - q_{|\delta|-1})^{1/2} \\ &= \sum_{\varepsilon \in p(\alpha) \cap p(\beta)} q_{|\varepsilon|} - q_{|\varepsilon|-1} = q_{\alpha \wedge \beta}. \end{aligned}$$

as required. ■

This result implies that if we place the p'th generation vertices of the tree $\mathscr{A}$ at height q_p and we identify each leaf $\alpha \in \mathbb{N}^r$ with the point h_α in the support of the Ruelle probability cascade $\mathrm{RPC}(\zeta)$,

we can obtain the scalar product $h_\alpha \cdot h_\beta$ by reading off the height at which the paths $p(\alpha)$ and $p(\beta)$ merge. This is best expressed through the following figure.

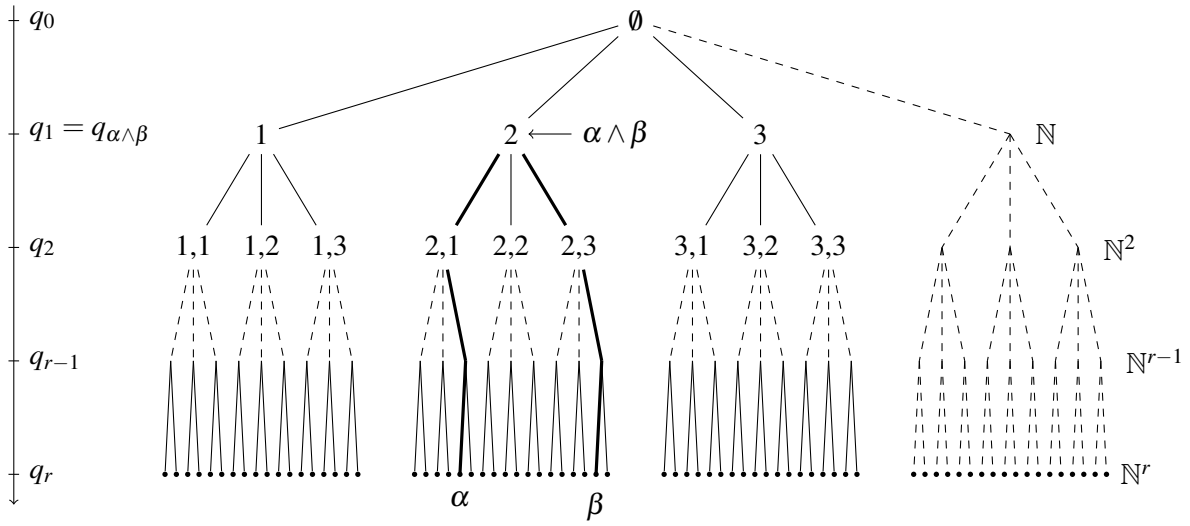

Figure 1: The leaves α and β in this infinite tree index the q_r clusters in the model. Such a tree represents what is called *r-step replica symmetry breaking*.

We now turn our attention to defining the weights of the Ruelle probability cascade RPC(ζ). This requires the notion of a Poisson point process, which we now recall, and many of the properties of these objects which may be found in [7, chapter 7].

Definition 6.3
Let μ be a non-atomic measure admitting the decomposition $\mu = \sum_{n\geq 1}\mu_n$ for some family $(\mu_n)_{n\geq 1}$ of non-atomic and finite measures on a separable metric space S. For each $n \geq 1$ let N_n be a random variable with Poisson distribution $\Pi(\mu_n(S))$ and let $(X_{nl})_{l\geq 1}$ be i.i.d. random variables, also independent of N_n, with distribution $p_n(B) = \frac{\mu_n(B)}{\mu_n(S)}$ for each $B \in \mathscr{S}$. Assume all these random variables are independent for different values of $n \geq 1$. The *Poisson point process on S with mean measure* μ is the set

$$\Pi = \bigcup_{n\geq 1} \Pi_n,$$

where $\Pi_n = \{X_{n1}, \ldots, X_{nN_n}\}$. ■

The weights of the Ruelle probability cascades are defined using the Poisson point process on $(0,\infty)$ with mean measure

$$\mu(\mathrm{d}x) = \eta x^{-1-\eta}\mathrm{d}x \tag{14}$$

for some $\eta \in (0,1)$. Given the importance of the Poisson point process with this mean measure, we will call it the *Poisson-Dirichlet point process* with parameter $\eta \in (0,1)$. Since $\mu((0,\infty)) = \infty$ and $\mu([\varepsilon,\infty)) < \infty$ for every $\varepsilon > 0$, it is immediate from [7, theorem 7.5] that the Poisson-Dirichlet point process Π with parameter $\eta \in (0,1)$ is countably infinite and almost surely admits finitely many points in the interval $[\varepsilon,\infty)$ for each $\varepsilon > 0$. In particular, the decreasing enumeration $(u_n)_{n\geq 1}$ of the points in Π is a well-defined stochastic process. With this in mind, for each $\alpha \in \mathscr{A} \setminus \mathbb{N}^r$ let Π_α be the Poisson-Dirichlet point process with parameter $\zeta_{|\alpha|}$, and let us generate these processes independently for all α. Denote by $(u_{\alpha n})_{n\geq 1}$ the decreasing enumeration of Π_α. In this way, parent vertices $\alpha \in$

$\mathscr{A} \setminus \mathbb{N}^r$ enumerate independent Poisson point processes Π_α and child vertices $\alpha n \in \mathscr{A} \setminus \mathbb{N}^0$ enumerate individual points $u_{\alpha n}$. To each vertex $\alpha \in \mathscr{A} \setminus \mathbb{N}^0$ associate the path dependent quantity

$$w_\alpha := \prod_{\beta \in p(\alpha)} u_\beta.$$

It is not too hard to use induction to prove that $\sum_{\beta \in \mathbb{N}^r} w_\alpha$ is finite almost surely. This allows us to define the weights of the Ruelle probability cascade $\mathrm{RPC}(\zeta)$ by normalizing w_α.

Definition 6.4
The *weights of the Ruelle probability cascade* $\mathrm{RPC}(\zeta)$ are the random weights (ν_α) indexed by $\alpha \in \mathbb{N}^r$ and defined by

$$\nu_\alpha = \frac{w_\alpha}{\sum_{\beta \in \mathbb{N}^r} w_\beta}.$$

for each $\alpha \in \mathbb{N}^r$. The *Ruelle probability cascade* $\mathrm{RPC}(\zeta)$ with functional order parameter ζ is the random measure G on H defined by $G(h_\alpha) = \nu_\alpha$ for each $\alpha \in \mathbb{N}^r$. ■

The Ruelle probability cascades satisfy many fundamental properties which make them indispensable to the study of spin glasses. We will not discuss any of them here, but simply mention a result that links them to the Parisi functional. To state this result as concisely as possible, denote by G the Ruelle probability cascade $\mathrm{RPC}(\zeta)$ and let $\langle \cdot \rangle$ denote the average with respect to $G^{\otimes \infty}$. Let $(\omega_p)_{1 \le p \le r}$ be a sequence of i.i.d. uniform random variables on $[0,1]$ and consider a random variable

$$X_r = X_r(\omega_1, \ldots, \omega_r)$$

determined by the r sources of randomness $(\omega_p)_{1 \le p \le r}$. Suppose that $\mathbb{E} \exp \zeta_{r-1} X_r < \infty$ and define recursively

$$X_l = X_l(\omega_1, \ldots, \omega_l) = \frac{1}{\zeta_l} \log \mathbb{E}_{\omega_{l+1}} \exp \zeta_l X_{l+1} \tag{15}$$

for $0 \le l \le r-1$.

Theorem 6.5 (RPC averages)
If $(\omega_\alpha)_{\alpha \in \mathscr{A} \setminus \mathbb{N}^0}$ is a sequence of i.i.d. uniform random variables on $[0,1]$ and X_0 is defined recursively via (15)*, then*

$$X_0 = \mathbb{E} \log \langle \exp X_r((\omega_\beta)_{\beta \in p(\alpha)}) \rangle = \mathbb{E} \log \sum_{\alpha \in \mathbb{N}^r} \nu_\alpha \exp X_r((\omega_\beta)_{\beta \in p(\alpha)}).$$

Theorem 6.5 is most commonly applied in the following more general formulation. Instead of just considering r sources of randomness $(\omega_p)_{1 \le p \le r}$, use each random variable ω_p to generate N independent copies $\omega_{p,i}$ from the uniform distribution on $[0,1]$, and instead of just considering one random variable X_r, consider N independent random variables

$$X_{r,i} = X_{r,i}(\omega_{1,i}, \ldots, \omega_{r,i}).$$

Assuming that $\mathbb{E} \exp \zeta_{r-1} X_{r,i} < \infty$ for $1 \le i \le N$, define recursively

$$X_{l,i} = X_{l,i}(\omega_{1,i}, \ldots, \omega_{l,i}) = \frac{1}{\zeta_l} \log \mathbb{E}_{\omega_{l+1,i}} \exp \zeta_l X_{l+1,i} \tag{16}$$

for $0 \le l \le r-1$ and $1 \le i \le N$.

Corollary 6.6 (RPC averages generalized)
Let $(\omega_\alpha)_{\alpha\in\mathscr{A}\setminus\mathbb{N}^0}$ be a sequence of i.i.d. uniform random variables on $[0,1]$ and for each $\alpha\in\mathscr{A}\setminus\mathbb{N}^0$ let $(\omega_{\alpha,i})_{1\le i\le N}$ be N independent copies of ω_α. If $(X_{0,i})$ are defined recursively via (16), then

$$\sum_{i=1}^N X_{0,i} = \mathbb{E}\log\Big\langle \exp\sum_{i=1}^N X_{r,i}((\omega_{\beta,i})_{\beta\in p(\alpha)})\Big\rangle = \mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r}\nu_\alpha\exp\sum_{i=1}^N X_{r,i}((\omega_{\beta,i})_{\beta\in p(\alpha)}).$$

Proof. For each $0\le l\le r$ let $Y_l=\sum_{i=1}^N X_{l,i}$. By independence

$$\mathbb{E}\exp\zeta_{r-1}Y_r = \prod_{i=1}^N \mathbb{E}\exp\zeta_{r-1}X_{r,i} < \infty$$

and for $0\le l\le r-1$

$$Y_l = \frac{1}{\zeta_l}\sum_{i=1}^N \log\mathbb{E}_{\omega_{l+1,i}}\exp\zeta_l X_{l+1,i} = \frac{1}{\zeta_l}\log\mathbb{E}_{(\omega_{l+1,1},\dots,\omega_{l+1,N})}\exp\zeta_l Y_{l+1}$$

Since $Y_0=\sum_{i=1}^N X_{0,i}$ the result is now an immediate consequence of theorem 6.5. ∎

Of course, these two results also apply to many i.i.d. random variables $(\omega_p)_{1\le p\le r}$ that are not uniformly distributed on $[0,1]$. For instance, any real-valued random variable can be generated as a measurable function of a uniform random variable by considering its quantile transform. Averaging with respect to this random variable is equivalent to averaging with respect to the uniform randomness that generates it, so corollary 6.6 and theorem 6.5 extend to i.i.d. random variables $(\omega_p)_{1\le p\le r}$ with any distribution on $\mathbb{R}$.

The similarity between the recursive definition of (15) and the Parisi sequence in section 5 suggests that we should be able to express the Parisi functional in terms of the Ruelle probability cascades. To do so, let us introduce two independent Gaussian processes $(Z(h_\alpha))_{\alpha\in\mathbb{N}^r}$ and $(Y(h_\alpha))_{\alpha\in\mathbb{N}^r}$ indexed by the points in the support of the Ruelle probability cascade $\mathrm{RPC}(\zeta)$ with covariance structure

$$\begin{aligned}\mathbb{E}Z(h_{\alpha^1})Z(h_{\alpha^2}) &= 2h_{\alpha^1}\cdot h_{\alpha^2} = 2q_{\alpha^1\wedge\alpha^2}\\ \mathbb{E}Y(h_{\alpha^1})Y(h_{\alpha^2}) &= (h_{\alpha^1}\cdot h_{\alpha^2})^2 = q^2_{\alpha^1\wedge\alpha^2}.\end{aligned}$$

An explicit example of such processes is given by

$$\begin{aligned}Z(h_\alpha) &= \sum_{\gamma\in p(\alpha)}\sqrt{2}(q_{|\gamma|}-q_{|\gamma|-1})^{1/2}\eta_\gamma\\ Y(h_\alpha) &= \sum_{\gamma\in p(\alpha)}(q^2_{|\gamma|}-q^2_{|\gamma|-1})^{1/2}\eta'_\gamma\end{aligned} \tag{17}$$

for independent sequences $(\eta_\alpha)_{\alpha\in\mathscr{A}\setminus\mathbb{N}^0}$ of i.i.d. standard Gaussian random variables. In this notation the first term in the Parisi sequence satisfies

$$X_r^\zeta((\eta_\beta)_{\beta\in p(\alpha)}) = \log 2\cosh\beta Z(h_\alpha) \tag{18}$$

for any $\alpha\in\mathbb{N}^r$. Combining this with theorem 6.5 implies the following representation of the Parisi functional.

Theorem 6.7 (Parisi functional in terms of RPC)
Given a discrete functional order parameter $\zeta \in \mathscr{D}[0,1]$, the Parisi functional at ζ is given by

$$\begin{aligned}\mathscr{P}(\zeta) &= \mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha 2\cosh\beta Z(h_\alpha) - \mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha \exp\beta Y(h_\alpha) \\ &= \mathbb{E}\log\langle 2\cosh\beta Z(\sigma)\rangle - \mathbb{E}\log\langle\exp\beta Y(\sigma)\rangle,\end{aligned} \tag{19}$$

where $\langle\cdot\rangle$ denotes the average with respect to the Ruelle probability cascade $\mathrm{RPC}(\zeta)$.

Proof. To relate the first term in the definition of the Parisi functional to the Ruelle probability cascades recall that each of the standard Gaussian random variables $\eta_1,\dots,\eta_r$ can be expressed as a measurable function of a uniform random variable on $[0,1]$ by means of the quantile transform, and invoke theorem 6.5 and (18) to say that

$$X_0 = \mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha \exp X_r((\eta_\alpha)_{\alpha\in p(\alpha)}) = \mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} 2v_\alpha\cosh\beta Z(h_\alpha).$$

To relate the second term in (19) to the second term in the Parisi functional consider the function

$$Y_r = Y_r(\eta_1,\dots,\eta_r) = \sum_{1\le p\le r}\beta(q_p^2-q_{p-1}^2)^{1/2}\eta_p$$

and define Y_0 iteratively by setting $Y_l = \frac{1}{\zeta_l}\log\mathbb{E}_{\eta_{l+1}}\exp\zeta_l Y_{l+1}$ for $0\le l\le r-1$. It is clear that

$$Y_r((\eta_\beta)_{\beta\in p(\alpha)}) = \beta Y(h_\alpha)$$

for any $\alpha\in\mathbb{N}^r$, so theorem 6.5 reveals that

$$Y_0 = \mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha\exp Y_r((\eta_\alpha)_{\alpha\in p(\alpha)}) = \mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r} v_\alpha\exp\beta Y(h_\alpha).$$

On the other hand, a simple computation leveraging the explicit form of the moment generating function of a standard Gaussian yields

$$\begin{aligned}Y_{r-1} &= \sum_{1\le p\le r-1}\beta(q_p^2-q_{p-1}^2)^{1/2}\eta_p + \frac{1}{\zeta_{r-1}}\log\mathbb{E}_{\eta_r}\exp\zeta_{r-1}\beta(q_r^2-q_{r-1}^2)^{1/2}\eta_r \\ &= \sum_{1\le p\le r-1}\beta(q_p^2-q_{p-1}^2)^{1/2}\eta_p + \frac{\beta^2}{2}\zeta_{r-1}(q_r^2-q_{r-1}^2).\end{aligned}$$

A simple induction on r gives $Y_0 = \frac{\beta^2}{2}\sum_{0\le p\le r-1}\zeta_p(q_{p+1}^2-q_p^2)$ and completes the proof. ∎

Before we use this result to prove the upper bound in the Parisi formula, let us comment on the similarity between (19) and (7). Remember that in section 3 we used a cavity computation to show that $\liminf_{N\to\infty}F_N \ge \liminf_{N\to\infty}A_N$ for the quantity A_N defined by

$$A_N = \mathbb{E}\log\Big\langle 2\cosh\beta z_N(\sigma)\Big\rangle' - \mathbb{E}\log\Big\langle\exp\beta y_N(\sigma)\Big\rangle',$$

where $z_N(\sigma)$ and $y_N(\sigma)$ were independent centred Gaussian processes indexed by $\sigma\in\Sigma_N$ with covariance structure

$$\mathbb{E}z_N(\sigma^1)z_N(\sigma^2) = 2\frac{N}{N+1}R_{1,2} \quad\text{and}\quad \mathbb{E}y_N(\sigma^1)y_N(\sigma^2) = \frac{N}{N+1}R_{1,2}^2.$$

Comparing this to (19) suggests that to prove the lower bound in the Parisi formula we should try to show that the sequence of overlap arrays $(R^N_{l,l'})_{l,l'\ge1}$ under the distribution $\mathbb{E}G_N'^{\otimes\infty}$, or at least an

appropriate subsequence of these, converges weakly to the distribution of an infinite array R generated by some Ruelle probability cascade $\mathrm{RPC}(\zeta)$. Indeed, as the covariance structure of the Gaussian processes $(z_N(\sigma))$ and $(y_N(\sigma))$ depends only on the overlap array and $\frac{N}{N+1} \to 1$, this should imply that

$$\liminf_{N\to\infty} F_N \geq \liminf_{N\to\infty} A_N = \mathscr{P}(\zeta) \geq \inf_{\zeta} \mathscr{P}(\zeta).$$

This is a very loose and simplified description of the strategy adopted by Panchenko in [6] to prove the Parisi formula. As one should expect of a problem that remained open for over 30 years, there are many deep problems one has to tackle when trying to formalize the idea we have just described. Most notably, it is extremely unclear why one would get a Ruelle probability cascade in the limit. Panchenko's approach to this matter involved the famous Ghirlanda-Guerra identities which one can force the Gibbs measure to satisfy by adding to it a small perturbation which does not affect the limit of the free energy. These identities describe how one can generate the off-diagonal entries of the infinite array $(R^N_{l,l'})_{l,l'\geq 1}$ (they only become exact in the limit but we will ignore this subtlety in our vague discussion of them). More precisely, if ζ is the distribution of one overlap $R^N_{1,2}$, they describe how to generate the distribution of $R^N_{1,n+1}$ conditionally on $(R^N_{l,l'})_{l,l\leq n}$. A priori this gives a way to generate the marginal distributions of the overlap array $(R^N_{l,l'})_{l,l'\leq n+1}$, but is insufficient to generate the joint distribution. However, soon after these identities were discovered, it was observed that when combined with the ultrametricity of the support of $(R^N_{l,l'})_{l,l'\leq n+1}$, they are sufficient to deduce the joint distribution of the whole overlap array $(R^N_{l,l'})_{l,l'\geq 1}$. For a very long time, it was believed that the Ghirlanda-Guerra identities and ultrametricity were complimentary properties that were both needed to describe the whole overlap array $(R^N_{l,l'})_{l,l'\geq 1}$ from the distribution of one overlap $R^N_{1,2}$. Panchenko's greatest contribution to the theory of spin glasses was to prove that the Ghirlanda-Guerra identities actually imply ultrametricity. With this great insight, Panchenko was able to use the Dovbysh-Sudakov theorem and the fact that the Ruelle probability cascades satisfy the Ghirlanda-Guerra identities to prove that any subsequential limit of the overlap arrays $(R^N_{l,l'})_{l,l'\geq 1}$ is generated by a Ruelle probability cascade. From this he was able to deduce the Parisi formula with relative ease. All of this is explained in much greater detail and clarity in [4].

7 The Guerra Replica Symmetry Breaking Interpolation

We conclude this note by using the *Guerra replica symmetry breaking interpolation* to show that the Parisi formula gives an upper bound for the limit of the free energy.

Theorem 7.1 (Guerra's RSB bound)
For any functional order parameter $\zeta \in \mathscr{D}[0,1]$ and every $N \in \mathbb{N}$ we have

$$F_N \leq \mathscr{P}(\zeta).$$

Proof. For $1 \leq i \leq N$ let $(Z_i(h_\alpha))$ and $(Y_i(h_\alpha))$ be independent copies of the Gaussian processes defined in (17). For each $0 \leq t \leq 1$ consider the interpolating Hamiltonian on $\Sigma_N \times \mathbb{N}^r$ given by

$$H_{N,t}(\sigma,\alpha) = \sqrt{t}H_N(\sigma) + \sqrt{1-t}\sum_{i=1}^N Z_i(h_\alpha)\sigma_i + \sqrt{t}\sum_{i=1}^N Y_i(h_\alpha).$$

as well as the interpolating free energy

$$\varphi(t) = \frac{1}{N}\mathbb{E}\log\sum_{\sigma,\alpha} \nu_\alpha \exp\beta H_{N,t}(\sigma,\alpha).$$

By theorem 6.5 and corollary 6.6 it is clear that

$$\begin{aligned}\varphi(1) &= \frac{1}{N}\mathbb{E}\log\sum_{\sigma\in\Sigma_N}\exp\beta H_N(\sigma)+\frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r}\nu_\alpha\exp\sum_{i=1}^N\beta Y_i(h_\alpha)\\ &= \frac{1}{N}\mathbb{E}\log\sum_{\sigma\in\Sigma_N}\exp\beta H_N(\sigma)+\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r}\nu_\alpha\exp\beta Y_1(h_\alpha)\end{aligned}$$

To compute $\varphi(0)$ introduce Rademacher random variables $(\varepsilon_i)_{1\le i\le N}$, and for each $1\le m\le N$ write $\mathrm{Av}_{(\varepsilon_1,\dots,\varepsilon_m)}$ to denote the expectation with respect to the randomness of $(\varepsilon_1,\dots,\varepsilon_m)$. Another application of theorem 6.5 and corollary 6.6 gives

$$\begin{aligned}\varphi(0) &= \frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r}\nu_\alpha\sum_{\sigma\in\Sigma_N}\exp\sum_{i=1}^N\beta Z_i(h_\alpha)\sigma_i\\ &= \frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r}\nu_\alpha 2^N\mathrm{Av}_{(\varepsilon_1,\dots,\varepsilon_N)}\prod_{i=1}^N\exp\beta Z_i(h_\alpha)\varepsilon_i\\ &= \frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r}\nu_\alpha\prod_{i=1}^N 2\mathrm{Av}_{\varepsilon_i}\exp\beta Z_i(h_\alpha)\varepsilon_i\\ &= \frac{1}{N}\mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r}\nu_\alpha\prod_{i=1}^N 2\cosh\beta Z_i(h_\alpha)\\ &= \mathbb{E}\log\sum_{\alpha\in\mathbb{N}^r}\nu_\alpha 2\cosh\beta Z_1(h_\alpha).\end{aligned}$$

By theorem 6.7 it is therefore sufficient to prove that $\varphi'(t)\le 0$. To compute the derivative of φ it will be convenient to introduce the Gibbs measure on $\Sigma_N\times\mathbb{N}^r$ defined by

$$\Gamma_t(\sigma,\alpha)=\frac{\nu_\alpha\exp H_{N,t}(\sigma,\alpha)}{Z_t},$$

where $Z_t=\sum_{\sigma,\alpha}\nu_\alpha\exp H_{N,t}(\sigma,\alpha)$ is the partition function of the Hamiltonian $H_{N,t}$. As usual, we will write $\langle\cdot\rangle_t$ for the average with respect to the product measure $\Gamma_t^{\otimes\infty}$. If we write $\rho=(\sigma,\alpha)$ for $\sigma\in\Sigma_N$ and $\alpha\in\mathbb{N}^r$ and denote by $R_{1,2}=\frac{1}{N}\sum_{i=1}^N\sigma_i^1\sigma_i^2$ the overlap between two configurations $\sigma^1,\sigma^2\in\Sigma_N$, the Gaussian integration by parts formula in corollary 4.2 yields

$$\varphi'(t)=\frac{1}{N}\mathbb{E}\Big\langle\frac{\partial H_{N,t}(\rho)}{\partial t}\Big\rangle_t=\frac{1}{N}\mathbb{E}\langle C(\rho^1,\rho^1)-C(\rho^1,\rho^2)\rangle_t,$$

where

$$\begin{aligned}C(\rho^1,\rho^2) &= \mathbb{E}\frac{\partial H_{N,t}(\rho^1)}{\partial t}H_{N,t}(\rho^2)\\ &= \frac{1}{2}\Big(\mathbb{E}H_N(\sigma^1)H_N(\sigma^2)+N\mathbb{E}Y(h_{\alpha^1})Y(h_{\alpha^2})-\sum_{i=1}^N\sigma_i^1\sigma_i^2\mathbb{E}Z_i(h_{\alpha^1})Z_i(h_{\alpha^2})\Big)\\ &= \frac{1}{2}\Big(NR_{1,2}^2+Nq_{\alpha^1\wedge\alpha^2}^2-2q_{\alpha^1\wedge\alpha^2}NR_{1,2}\Big)\\ &= \frac{N}{2}(R_{1,2}-q_{\alpha^1\wedge\alpha^2})^2.\end{aligned}$$

Since $C(\rho^1,\rho^1)=0$ this implies that

$$\varphi'(t)=-\frac{1}{2}\mathbb{E}\Big\langle(R_{1,2}-q_{\alpha^1\wedge\alpha^2})^2\Big\rangle_t\le 0$$

which completes the proof. ■

References

[1] S.F. Edwards and P.W. Anderson. Theory of spin glasses. *Journal of Physics F: Metal Physics*, 5(5):965–974, May 1975.

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

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

[4] D. Panchenko. *The Sherrington-Kirkpatrick Model*. Springer Monographs in Mathematics. Springer New York, 2013.

[5] D. Panchenko. Introduction to the SK model, 2014.

[6] D. Panchenko. The Parisi formula for mixed p-spin models. *The Annals of Probability*, 42(3), May 2014.

[7] D. Panchenko. *Lecture Notes on Probability Theory*. 2019.

[8] G. Parisi. Infinite number of order parameters for spin-glasses. *Phys. Rev. Lett.*, 43:1754–1756, Dec 1979.

[9] G. Parisi. A sequence of approximated solutions to the SK model for spin glasses. *Journal of Physics A: Mathematical and General*, 13(4):L115–L121, Apr 1980.

[10] D. Sherrington and S. Kirkpatrick. Solvable model of a spin-glass. *Phys. Rev. Lett.*, 35:1792–1796, Dec 1975.

[11] M. Talagrand. The Parisi formula. *Annals of Mathematics*, 163(1):221–263, 2006.