A short course on mean field spin glasses

Anton Bovier[1] and Irina Kurkova[2]

[1] Weierstrass Institut für Angewandte Analysis und Stochastik, Mohrenstrasse 39, 10117 Berlin, Germany bovier@wias-berlin.de
[2] Laboratoire de Probabilités et Modèles Aléatoires, Université Paris 6, 4, place Jussieu, B.C. 188, 75252 Paris, Cedex 5, France kourkova@ccr.jussieu.fr

1 Preparation: The Curie-Weiss model

The main topic of this lecture series are diordered mean field spin systems. This first section will, however, be devoted to ordered spin systems, and more precisely essentially to the Curie-Weiss model. This will be indispensable to appreciate later the much more complicated Sherrington-Kirkpatrick spin glass.

1.1 Spin systems.

In his Ph.D. thesis in 1924, Ernst Ising [18, 19] attempted to solve a model, proposed by his advisor Lenz, intended to describe the statistical mechanics of an interacting system of magnetic moments. The setup of the model proceeds again from a lattice, $\mathbb{Z}^d$, and a finite subset, $\Lambda \subset \mathbb{Z}^d$. The lattice is supposed to represent the positions of the atoms in a regular crystal. Each atom is endowed with a magnetic moment that is quantized and can take only the two values $+1$ and -1, called the *spin* of the atom. This spin variable at site $x \in \Lambda$ is denoted by σ_x. The spins are supposed to interact via an interaction potential $\phi(x, y)$; in addition, a magnetic field h is present. The energy of a *spin configuration* is then

$$H_\Lambda(\sigma) \equiv -\sum_{x \neq y \in \Lambda} \phi(x, y)\sigma_x\sigma_y - h\sum_{x\in\Lambda}\sigma_x \tag{1}$$

The spin system with Hamiltonian (1) with the particular choice

$$\phi(x, y) = \begin{cases} J, & \text{if } |x-y| = 1 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

is known as the Ising spin system or *Ising model*. This model has played a crucial rôle in the history of statistical mechanics.

The essential game in statistical mechanics is to define, once a Hamiltonian is given, a probability measure, called the *Gibbs measure* , on the space of spin-configuration. This entails some interesting subtleties related to the fact that we would really do this in infinite volume, but I will not enter into these here. For finite volumes, Λ, we can easily define this probability as

$$\mu_{\beta,h,\Lambda}(\sigma) \equiv \frac{\exp\left(-\beta H_\Lambda(\sigma)\right)}{Z_{\beta,h,\Lambda}}, \tag{3}$$

where $Z_{\beta,h.\Lambda}$ is a normalizing factor called the *partition function*,

$$Z_{\beta,h,\Lambda} \equiv \sum_{\sigma\in\mathcal{S}_\Lambda} \exp\left(-\beta H_\Lambda(\sigma)\right). \tag{4}$$

An interesting fact of statistical mechanics is that the behavior of the partition function as a function of the parameters β and h contains an enormous amount of information.

We will call

$$F_{\beta,h,\Lambda} \equiv -\frac{1}{\beta} \ln Z_{\beta,h,\Lambda} \tag{5}$$

the *free energy* of the spin system.

The importance of the Ising model for modern statistical physics can hardly be overestimated. With certain extensions, some of which we will discuss here, it has become a paradigmatic model for systems of large numbers of interacting individual components, and the applications gained from the insight into this model stretch far beyond the original intentions of Lenz and Ising.

1.2 Subadditivity and the existence of the free energy

The main concern of statistical mechanics is to describe systems in the limit when its size tends to infinity. Hence one of the first questions one asks, is whether quantities defined for finite Λ have limits as $\Lambda \uparrow \mathbb{Z}^d$. The free energy (5) is defined in such a way that one can expect this limit to exists. Since these questions will recur, it will be useful to see how such a result can be proven.

It will be useful to note that we can express the Hamiltonian in the equivalent form

$$\widehat{H}_\Lambda(\sigma) = \sum_{x,y\in\Lambda} \phi(x,y)\left(\sigma_x - \sigma_y\right)^2 - h\sum_{x\in\Lambda}\sigma_x \tag{6}$$

which differs from H_Λ only by a constant. Now let $\Lambda = \Lambda_1 \cup \Lambda_2$, where Λ_i are disjoint volumes. Clearly we have that

$$\begin{aligned} Z_{\beta,\Lambda} = &\sum_{\sigma_x, x\in\Lambda_1} \sum_{\tau_y, y\in\Lambda_2} \exp\left(-\beta\left[H_{\Lambda_1}(\sigma) + H_{\Lambda_2}(\tau)\right]\right) \\ &\times \exp\left(-\beta \sum_{x\in\Lambda_1}\sum_{y\in\Lambda_2} \phi(x,y)(\sigma_x - \tau_y)^2\right) \end{aligned} \tag{7}$$

If $\phi(x,y) \geq 0$, this implies that

$$Z_{\beta,\Lambda} \leq Z_{\beta,\Lambda_1} Z_{\beta,\Lambda_2} \tag{8}$$

and therefore

$$-F_{\beta,\Lambda} \leq (-F_{\beta,\Lambda_1}) + (-F_{\beta,\Lambda_2}) \tag{9}$$

The property (8) is called *subadditivity* of the sequence $(-F_{\beta,\Lambda})$. The importance of subadditivity is that it implies convergence, through an elementary analytic fact:

Lemma 1. *Let a_n be a real-valued sequence that satisfies, for any $n, m \in \mathbb{N}$,*

$$a_{n+m} \leq a_n + a_m \tag{10}$$

Then, $\lim_{n\uparrow\infty} n^{-1} a_n$ exists. If, moreover, $n^{-1}a_n$ is uniformly bounded from below, then the limit is finite.

By successive iteration, the lemma has an immediate extension to arrays:

Lemma 2. *Let $a_{n_1,n_2,\ldots,n_d}$, $n_i \in \mathbb{N}$ be a real-valued array that satisfies, for any $n_i, m_i \in \mathbb{N}$,*

$$a_{n_1+m_1,\ldots,n_d+m_d} \leq a_{n_1,\ldots,n_d} + a_{m_1,\ldots m_d} \tag{11}$$

Then, $\lim_{n\uparrow\infty} (n_1 n_2 \ldots, n_d)^{-1} a_{n_1,\ldots,n_d}$ exists.
If $a_n (n_1 n_2 \ldots, n_d)^{-1} a_{n_1,\ldots,n_d} \geq b > -\infty$, then the limit is finite.

Lemma 2 can be used straightforwardly to prove convergence of the free energy over rectangular boxes:

Proposition 1. *If the Gibbs free energy $F_{\beta,\Lambda}$ of a model satisfies the subadditivity property (9), and if $\sup_\sigma H_\Lambda(\sigma)/|\Lambda| \geq C > -\infty$, then, for any sequence Λ_n of rectangles*

$$\lim_{n\uparrow\infty} |\Lambda_n|^{-1} F_{\beta,\Lambda_n} = f_\beta \tag{12}$$

exists and is finite.

Obviously this proposition gives the existence of the free energy for Ising's model, but the range of applications of Proposition 1 is far wider, and virtually covers all lattice spin systems with bounded and absolutely summable interactions. To see this, one needs to realize that strict subadditivity is not really needed, as error terms arising, e.g., from boundary conditions can easily be controlled. Further details can be found in Simon's book [28].

1.3 The Curie–Weiss model

Although the Ising model can be solved exactly in dimensions one (easy) and two (hard), exact solutions in statistical mechanics are rare. To get a quick insight into specific systems, one often introduces exactly solvable *mean field models* . It will be very instructive to study the simplest of these models, the *Curie–Weiss model* in some detail. All we need to do to go from the Ising model to the Curie-Weiss model is to replace the nearest neighbor pair interaction of the Ising model by another extreme choice, namely the assumption that each spin variable interacts with each other spin variable at any site of the lattice with exactly the same strength. Since then the actual structure of the lattice becomes irrelevant, and we simply take $\Lambda = \{1, \dots, N\}$. The strength of the interaction should be chosen of order $1/N$, to avoid the possibility that the Hamiltonian takes on values larger than $O(N)$. Thus, the Hamiltonian of the Curie–Weiss model is

$$H_N(\sigma) = -\frac{1}{N} \sum_{1 \leq i,j \leq N} \sigma_i \sigma_j - h \sum_{i=1}^{N} \sigma_i \tag{13}$$

At this moment it is time to discuss the notion of *macroscopic variables* in some more detail. So far we have seen the magnetization, m, as a thermodynamic variable. It will be reasonable to define another magnetization as a *function* on the configuration space: we will call

$$m_N(\sigma) \equiv N^{-1} \sum_{i=1}^{N} \sigma_i \tag{14}$$

the *empirical magnetization.* Here we divided by N to have a specific magnetization. A function of this type is called a *macroscopic* function, because it depends on all spin variables, and depends on each one of them very little (we will make these notions more rigorous in the next section).

Note that the particular structure of the Curie–Weiss model entails that the Hamiltonian can be written as a function of this single macroscopic function:

$$H_N(\sigma) = -\frac{N}{2} [m_N(\sigma)]^2 - hNm_N(\sigma) \equiv N\Psi_h(m_N(\sigma)) \tag{15}$$

This can be considered as a defining feature of *mean field models.*

Let us now try to compute the free energy of this model. Because of the the interaction term, this problem looks complicated at first. To overcome this difficulty, we do what would appear unusual from our past experience: we go from the ensemble of fixed magnetic field to that of fixed magnetization. That is, we write

$$Z_{\beta,h,N} = \sum_{m \in \mathcal{M}_N} e^{N\beta(\frac{m^2}{2}+mh)} z_{m,N} \tag{16}$$

where $\mathcal{M}_N$ is the set of possible values of the magnetization, i.e.,

$$\mathcal{M}_N \equiv \{m \in \mathbb{R} : \exists \sigma \in \{-1,1\}^N : m_N(\sigma) = m\} \\ = \{-1, -1+2/N, \dots, 1-2/N, 1\} \tag{17}$$

and

$$z_{m,N} \equiv \sum_{\sigma \in \{-1,1\}^N} \mathbb{1}_{m_N(\sigma)=m} \tag{18}$$

is a 'micro-canonical partition function'. Fortunately, the computation of this micro-canonical partition function is easy. In fact, all possible values of m are of the form $m = 1 - 2k/N$, where k runs from 0 to N and counts the number of spins that have the value -1. Thus, the computation of $z_{m,N}$ amounts to the most elementary combinatorial problem, the counting of the number of subsets of size k in the set of the first N integers. Thus,

$$z_{m,N} = \binom{N}{N(1-m)/2} \equiv \frac{N!}{[N(1-m)/2]![N(1+m)/2]!} \tag{19}$$

It is always useful to know the asymptotics of the logarithm of the binomial coefficients which gives, to leading order, for $m \in \mathcal{M}_N$,

$$N^{-1} \ln z_{m,N} \sim \ln 2 - I(m) \tag{20}$$

where

$$I(m) = \frac{1+m}{2}\ln(1+m) + \frac{1-m}{2}\ln(1-m) \tag{21}$$

is called *Cramèr's entropy function* and worth memorizing. Note that by its nature it is a relative entropy.

Some elementary properties of I are useful to know: First, I is symmetric, convex, and takes its unique minimum, 0, at 0. Moreover $I(1) = I(-1) = \ln 2$. Its derivative, $I'(m) = \mathrm{arcth}(m)$, exists in $(-1,1)$. While I is not uniformly Lipschitz continuous on $[-1,1]$, it has the following property:

Lemma 3. *There exists $C < \infty$ such that for any interval $\Delta \subset [-1,1]$ with $|\Delta| < 0.1$, $\max_{x,y\in\Delta} |I(x) - I(y)| \leq C|\Delta| |\ln|\Delta||$.*

We would like to say that $\lim_{N\uparrow\infty} \frac{1}{N} \ln z_{m,N} = \ln 2 + I(m)$. But there is a small problem, due to the fact that the relation (20) does only hold on the N-dependent set $\mathcal{M}_N$. Otherwise, $\ln z_{m,N} = -\infty$. A precise asymptotic statement could be the following:

Lemma 4. *For any $m \in [-1,1]$,*

$$\lim_{\epsilon\downarrow 0} \lim_{N\uparrow\infty} \frac{1}{N} \ln \sum_{m \in \mathcal{M}_M : |m-\tilde{m}|<\epsilon} z_{m,N} = \ln 2 + I(\tilde{m}) \tag{22}$$

Proof. The proof is elementary from properties of $z_{m,N}$ and $I(m)$ mentioned above and is left to the reader.

In probability theory, the following formulation of Lemma 4 is known as *Cramèr's theorem.* It is the simplest so-called *large deviation principle* [12]:

Lemma 5. *Let $A \in \mathcal{B}(\mathbb{R})$ be a Borel-subset of the real line. Define a probability measure p_N by $p_N(A) \equiv 2^{-N} \sum_{m \in \mathcal{M}_N \cap A} z_{m,N}$, and let $I(m)$ be defined in (21) Then*

$$
\begin{aligned}
-\inf_{m \in A} I(m) &\leq \liminf_{N \uparrow \infty} \frac{1}{N} \ln p_N(A) \\
&\leq \limsup_{N \uparrow \infty} \frac{1}{N} \ln p_N(A) \leq -\inf_{m \in \bar{A}} I(m)
\end{aligned}
\tag{23}
$$

Moreover, I is convex, lower-semi-continuous, Lipschitz continuous on $(-1,1)$, bounded on $[-1,1]$, and equal to $+\infty$ on $[-1,1]^c$.

Remark 1. The classical interpretation of the preceding theorem is the following. The spin variables $\sigma_i = \pm 1$ are *independent, identically distributed* binary random variables taking the values ± 1 with equal probability. $m_N(\sigma)$ is the normalized sum of the first N of these random variables. p_N denotes the probability distribution of the random variable m_N, which is inherited from the probability distribution of the family of random variables σ_i. It is well known, by the law of large numbers, that p_N will concentrate on the value $m = 0$, as N tends to ∞. A large deviation principle states in a precise manner how small the probability will be that m_N take on different values. In fact, the probability that m_N will be in a set A, that does not contain 0, will be of the order $\exp(-Nc(A))$, and the value of $c(A)$ is precisely the smallest value that the function $I(m)$ takes on the set A.

The computation of the canonical partition function is now easy:

$$
Z_{\beta,h,N} = \sum_{m \in \mathcal{M}_N} \binom{N}{N(1-m)/2} \exp\left(N\beta\left(\frac{m^2}{2} + hm\right)\right) \tag{24}
$$

and by the preceeding lemma, one finds that:

Lemma 6. *For any temperature, β^{-1}, and magnetic field, h,*

$$
\begin{aligned}
\lim_{N \uparrow \infty} \frac{-1}{\beta N} \ln Z_{\beta,h,N} &= \inf_{m \in [0,1]} \left(-m^2/2 + hm - \beta^{-1}(\ln 2 - I(m)\right) \\
&= f(\beta, h)
\end{aligned}
\tag{25}
$$

Proof. We give the simplest proof, which, however, contains some valuable lessons. We first prove an upper bound for $Z_{\beta,h,N}$:

$$
\begin{aligned}
Z_{\beta,h,N} &\leq N \max_{m \in \mathcal{M}_N} \exp\left(N\beta(\frac{m^2}{2} + hm)\right) \binom{N}{N(1-m)/2} \\
&\leq N \max_{m \in [-1,1]} \exp\left(N\beta(\frac{m^2}{2} + hm) + N(\ln 2 - I(m) - J_N(m))\right)
\end{aligned}
\tag{26}
$$

Hence

$$
\begin{aligned}
&N^{-1}\ln Z_{\beta,h,N} \qquad (27)\\
&\le N^{-1}\ln N + \max_{m\in[-1,1]}\left(\beta\left(\frac{m^2}{2}+hm\right)+\ln 2 - I(m) - J_N(m)\right)\\
&\le \ln 2 + \sup_{m\in[-1,1]}\left(\beta\left(\frac{m^2}{2}+hm\right) - I(m)\right) + N^{-1}O(\ln N)
\end{aligned}
$$

so that

$$
\limsup_{N\uparrow\infty} N^{-1}\ln Z_{\beta,h,N} \le \ \beta \sup_{m\in[-1,1]}\left(\frac{m^2}{2}+hm-\beta^{-1}I(m)\right)+\ln 2 \qquad (28)
$$

This already looks good. Now all we need is a matching lower bound. It can be found simply by using the property that the sum is bigger than its parts:

$$
Z_{\beta,h,N} \ge \max_{m\in\mathcal{M}_N} \exp\left(N\beta\left(\frac{m^2}{2}+hm\right)\right)\binom{N}{N(1-m)/2} \qquad (29)
$$

We see that we will be in business, up to the small problem that we need to pass from the max over $\mathcal{M}_N$ to the max over $[-1,1]$, after inserting the bound for the binomial coefficient in terms of $I(m)$. In fact, we get that

$$
\begin{aligned}
N^{-1}\ln Z_{\beta,h,N} \ge \ln 2 + \beta \max_{m\in\mathcal{M}_N}\left(\frac{m^2}{2}+hm-\beta^{-1}I(m)\right) \qquad (30)\\
- O(\ln N/N)
\end{aligned}
$$

for any N. Now, we can easily check that

$$
\begin{aligned}
\max_{m\in\mathcal{M}_N} \quad &\left|\left(\frac{m^2}{2}+hm-\beta^{-1}I(m)\right)\right. \qquad (31)\\
&- \left.\sup_{m'\in[0,1],|m'-m|\le 2/N}\left(\frac{m^2}{2}+hm-\beta^{-1}I(m)\right)\right| \le C\ln N/N
\end{aligned}
$$

so that

$$
\liminf_{\mathbb{N}\uparrow\infty}\frac{1}{\beta N}\ln Z_{\beta,h,N} \ge \beta^{-1}\ln 2 + \sup_{m\in[-1,1]}\left(\frac{m^2}{2}+hm-\beta^{-1}I(m)\right) \qquad (32)
$$

and the assertion of the lemma follows immediately.

The function $g(\beta,m)\equiv -m^2/2-\beta^{-1}(\ln 2 - I(m))$ is called the *Helmholtz free energy* for zero magnetic field, and

$$
\lim_{\epsilon\downarrow 0}\lim_{N\uparrow\infty}\frac{-1}{\beta N}\ln\sum_{\tilde m:|\tilde m - m|<\epsilon}\widetilde{Z}_{\beta,\tilde m,N} = g(\beta,m) \qquad (33)
$$

where

$$\widetilde{Z}_{\beta,\tilde{m},N} = \sum_{\sigma\in\{-1,1\}^N} e^{\beta H_N(\sigma)} \mathbb{1}_{m_N(\sigma)=m} \tag{34}$$

for $h = 0$. Thermodynamically, the function $f(\beta, h)$ is then called *Gibbs free energy*, and the assertion of the lemma would then be that the Gibbs free energy is the *Legendre transform* of the Helmholtz free energy. The latter is closely related to the rate function of a large deviation principle for the distribution of the magnetization under the Gibbs distribution. Namely, if we define the Gibbs distribution on the space of spin configurations

$$\mu_{\beta,h,N}(\sigma) \equiv \frac{e^{-\beta H_N(\sigma)}}{Z_{\beta,h,N}} \tag{35}$$

and denote by $\tilde{p}_{\beta,h,N}(A) \equiv \mu_{\beta,h,N}\left(\{m_N(\sigma) \in A\}\right)$ the law of m_N under this distribution, then we obtain very easily

Lemma 7. *Let $\tilde{p}_{\beta,h,N}$ be the law of $m_N(\sigma)$ under the Gibbs distribution. Then the family of probability measures $\tilde{p}_{\beta,h,N}$ satisfies a large deviation principle, i.e. for all Borel subsets of $\mathbb{R}$,*

$$\begin{aligned} -\inf_{m\in A}(g(\beta,m) - hm) + f(\beta,h) &\leq \liminf_{N\uparrow\infty} \frac{1}{\beta N} \ln \tilde{p}_{\beta,h,N}(A) \\ &\leq \limsup_{N\uparrow\infty} \frac{1}{\beta N} \ln \tilde{p}_{\beta,h,N}(A) \\ &\leq -\inf_{m\in \bar{A}}(g(\beta,m) - hm) + f(\beta,h) \end{aligned} \tag{36}$$

We see that the thermodynamic interpretation of equilibrium emerges very nicely: the equilibrium value of the magnetization, $m(\beta, h)$, for a given temperature and magnetic field, is the value of m for which the rate function in Lemma 7 vanishes, i.e., which satisfies the equation

$$g(\beta, m(\beta, h)) - hm(\beta, h) = f(\beta, h) \tag{37}$$

By the definition of f (see (25)), this is the case whenever $m(\beta, h)$ realizes the infimum in (25). If $g(\beta, m)$ is strictly convex, this infimum is unique, and, as long as g is convex, it is the set on which $\frac{\partial g(\beta,m)}{\partial m} = h$.

Note that, in our case, $g(\beta, m)$ is not a convex function of m if $\beta > 1$.

In fact, it has two local minima, situated at the values $\pm m^*_\beta$, where m^*_β is defined as the largest solution of the equation

$$m = \tanh \beta m \tag{38}$$

Moreover, the function g is symmetric, and so takes the same value at both minima. As a consequence, the minimizer of the function $g(\beta, m) - mh$, the magnetization as a function of the magnetic field, is not unique at the value

$h = 0$ (and only at this value). For $h > 0$, the minimizer is the positive solution of $m = \tanh(\beta(m+h))$, while for negative h it is the negative solution. Consequently, the magnetization has a jump discontinuity at $h = 0$, where it jumps by $2m^*_\beta$. One says that the Curie–Weiss model exhibits a first order phase transition.

1.4 A different view on the CW model.

We will now have a slightly different look at the Curie-Weiss model. This will be very instructive from the later perspective of the Sherrington-Kirkpatrick model.

To get started, we may want to compute the distribution of the spin variables as such. The perspective here is that of the product topology, so we should consider a fixed finite set of indices which without loss we may take to be $\{1, \ldots, K\}$ and ask for the Gibbs probability that the corresponding spin variables, $\sigma_1, \ldots, \sigma_K$ take specific values, and then take the thermodynamic limit.

To do these computations, it will be useful to make the following choices. The total volume of the system will be denoted $K + N$, where K is fixed and N will later tend to infinity. We will write $\hat\sigma \equiv (\sigma_1, \ldots, \sigma_K)$, and $\check\sigma \equiv (\sigma_{K+1}, \ldots, \sigma_{K+N})$. We set $\sigma = (\hat\sigma, \check\sigma)$. We now re-write the Hamiltonian as

$$
\begin{aligned}
-H_{K+N}(\sigma) = & \frac{1}{2(N+K)} \sum_{i,j \leq K} \sigma_i \sigma_j \\
& + \frac{1}{2(N+K)} \sum_{i,j > K} \sigma_i \sigma_j \\
& + \frac{1}{N+K} \sum_{i=1}^{K} \sigma_i \sum_{j=K+1}^{N+K} \sigma_j .
\end{aligned}
\tag{39}
$$

This can be written as

$$
\begin{aligned}
-H_{K+N}(\sigma) = & \frac{K^2}{2(N+K)} \left(m_K(\hat\sigma)\right)^2 \\
& + \frac{N^2}{2(N+K)} \left(m_N(\check\sigma)\right)^2 \\
& + \frac{N}{N+K} \sum_{i=1}^{K} \sigma_i m_N(\check\sigma).
\end{aligned}
\tag{40}
$$

Now the first term in this sum is of order $1/N$ and can be neglected. Also $N/(N+K) \sim 1 + O(1/N)$. But note that $N^2/(N+K) = N - K(1 - K/N) \sim N - K$. Using these approximations, we see that, up to terms that will vanish in the limit $N \uparrow \infty$,

$$\mu_{\beta,N+K}(\hat{\sigma}) = \frac{\mathbb{E}_{\check{\sigma}} e^{\beta(N-K)(m_N(\check{\sigma}))^2/2} e^{\beta m_N(\check{\sigma})\sum_{i=1}^K \sigma_i}}{\mathbb{E}_{\check{\sigma}} e^{\beta(N-K)(m_N(\check{\sigma}))^2/2} \sum_{\hat{\sigma}} e^{\beta m_N(\check{\sigma})\sum_{i=1}^K \sigma_i}} \\ = \frac{\int \mathbb{Q}_{\beta,N}(dm) e^{-\beta K m^2/2} e^{\beta m \sum_{i=1}^K \sigma_i}}{\int \mathbb{Q}_{\beta,N}(dm) e^{-\beta K m^2/2} \prod_{i=1}^K 2\cosh(\beta m)}. \tag{41}$$

Now we know that $\mathbb{Q}_{\beta,N}$ converges to either a Dirac measure on m^* or the mixture of two Dirac measures on m^* and $-m^*$. Thus it follows from (41) That $\mu_{\beta,N+K}$ converges to a product measure.

But assume that we did not know anything about $\mathbb{Q}_\beta$. Could we find out about it?

First, we write (assuming convergence, otherwise take a subsequence)

$$\mu_\beta(\hat{\sigma}) = \frac{\int \mathbb{Q}_\beta(dm) e^{-\beta K m^2/2} e^{\beta m \sum_{i=1}^K \sigma_i}}{\int \mathbb{Q}_\beta(dm) e^{-\beta K m^2/2} \prod_{i=1}^K 2\cosh(\beta m)} \tag{42}$$

Thus (41) establishes that the Gibbs measure of our model is completely determined by a single probability distribution, $\mathbb{Q}_\beta$, on a scalar random variable. Thus the task of finding the Gibbs measure is reduced to finding this distribution. How could we do this? A natural idea would be to use the Gibbs variational principle that say that the thermodynamic state must minimize the free energy. For this we would just need a representation of the free energy in terms of $\mathbb{Q}_\beta$.

To get there, we write the analog of (41) for the partition function. This yields

$$\frac{Z_{\beta,N+K}}{Z_{\beta,N}} = \int \mathbb{Q}_{\beta,N}(dm) e^{-\beta K m^2/2} \prod_{i=1}^K 2\cosh(\beta m). \tag{43}$$

Now it is not hard to see that the free energy can be obtained as

$$\lim_{N\uparrow\infty} \frac{1}{K\beta} \ln \frac{Z_{\beta,N+K}}{Z_{\beta,N}} = -f_\beta.$$

Thus we get the desired representation of the free energy

$$-f_\beta = \frac{1}{K\beta} \ln \int \mathbb{Q}_{\beta,N}(dm) \exp\left(-\beta K \left(m^2/2 - \beta^{-1} \ln 2\cosh(\beta m)\right)\right). \tag{44}$$

Thus the Gibbs principle states implies that

$$-f_\beta = \sup_{\mathbb{Q}} \frac{1}{K\beta} \ln \int \mathbb{Q}(dm) \exp\left(-\beta K \left(m^2/2 - \beta^{-1} \ln 2\cosh(\beta m)\right)\right), \tag{45}$$

where the supremum is taken over all probability measures on $\mathbb{R}$. It is of course not hard to see that the supremum is realized by any probability measure that has support on the minimizer of the function $m^2/2 - \beta^{-1} \ln\cosh(\beta m)$.

We will see later that a curious analog, with the sup replaced by an inf, of the formula (45) is the key to the solution of the Sherrington-Kirkpatrick model.

Should we not have known about the Gibbs principle, we could instead have observed that (44) can only hold for *all* K, if $\mathbb{Q}_\beta$ is supported on the minimizer of the function $m^2/2 - \beta^{-1} \ln\cosh(\beta m)$.

Remark 2. An other way to reach the same conclusion is to derive the the consistency relation

$$\int \mathbb{Q}_\beta(dm)m = \frac{\int \mathbb{Q}_\beta(dm)e^{-\beta K m^2/2} \left[\cosh(\beta m)\right]^K \tanh(\beta m)}{\int \mathbb{Q}_\beta(dm)e^{-\beta K m^2/2} \left[\cosh(\beta m)\right]^k} \tag{46}$$

for arbitrary K. But then it is clear that this can hold for all K only if $\mathbb{Q}_\beta$ is concentrated on the minimizers of the function $mr/2 - \beta^{-1} \ln\cosh(\beta m)$, which happen also to solve the equation $m^* = \tanh(\beta m^*)$ so that in the end all is consistent.

2 Random mean field models

The naive analog of the Curie–Weiss Hamiltonian with random couplings would be

$$H_N[\omega](\sigma) = -\frac{1}{2N} \sum_{1\leq i,j\leq N} J_{ij}[\omega]\sigma_i\sigma_j \tag{47}$$

for, say, J_{ij} some family of i.i.d. random variables. Thus, we must estimate $\mathbb{P}[\max_\sigma H_N(\sigma) \geq CN]$. But,

$$\begin{aligned} \mathbb{P}[\max_\sigma H_N(\sigma) \geq CN] &\leq \sum_{\sigma\in\mathcal{S}_N} \mathbb{P}[H_N(\sigma) \geq CN] \\ &= \sum_{\sigma\in\mathcal{S}_N} \inf_{t\geq 0} e^{-tCN} \mathbb{E}e^{t\frac{1}{2N}\sum_{i,j\in\Lambda_N\times\Lambda_N} J_{ij}[\omega]\sigma_i\sigma_j} \\ &= \sum_{\sigma\in\mathcal{S}_N} \inf_{t\geq 0} e^{-tCN} \prod_{i,j\in\Lambda_N\times\Lambda_N} \mathbb{E}e^{t\frac{1}{2N}J_{ij}[\omega]\sigma_i\sigma_j} \end{aligned} \tag{48}$$

where we assumed that the exponential moments of J_{ij} exist. A standard estimate then shows that, for some constant c, $\mathbb{E}e^{t\frac{1}{2N}J_{ij}[\omega]\sigma_i\sigma_j} \leq e^{c\frac{t^2}{2N^2}}$, and so

$$\mathbb{P}[\max_\sigma H_N(\sigma) \geq CN] \leq 2^N \inf_{t\geq 0} e^{-tCN} e^{ct^2/2} \leq 2^N e^{-\frac{C^2N^2}{2c}} \tag{49}$$

which tends to zero with N. Thus, our Hamiltonian is never of order N, but at best of order $\sqrt{N}$. The proper Hamiltonian for what is called the *Sherrington–Kirkpatrick model* (or short SK-model), is thus

$$H_N^{SK} \equiv -\frac{1}{\sqrt{2N}} \sum_{i,j\in\Lambda_N\times\Lambda_N} J_{ij}\sigma_i\sigma_j \tag{50}$$

where the random variables $J_{ij} = J_{ji}$ are i.i.d. for $i \leq j$ with mean zero (or at most $J_0 N^{-1/2}$) and variance normalized to one for $i \neq j$ and to two for $i = j$ In its original, and mostly considered, form, the distribution is moreover taken to be Gaussian.

This model was introduced by Sherrington and Kirkpatrick in 1976 [27] as an attempt to furnish a simple, solvable mean-field model for the then newly discovered class of materials called *spin-glasses*.

2.1 Gaussian process

This point of view consists of regarding the Hamiltonian (50) as a *Gaussian random process* indexed by the set $\mathcal{S}_N$, i.e. by the N-dimensional hypercube. Covariance function

$$\begin{aligned} \mathrm{cov}(H_N(\sigma), H_N(\sigma')) &= \frac{1}{2N} \sum_{1\leq i,j,l,k\leq N} \mathbb{E}J_{ij}J_{kl}\sigma_i\sigma_j\sigma'_k\sigma'_l \\ &= \frac{1}{N} \sum_{1\leq i,j\leq N} \sigma_i\sigma'_i\sigma_j\sigma'_j = NR_N(\sigma,\sigma')^2 \end{aligned} \tag{51}$$

where $R_N(\sigma,\sigma') \equiv N^{-1}\sum_{i=1}^N \sigma_i\sigma'_i$ is usually called the *overlap* between the two configurations σ and σ'.

Hamming distance $d_{HAM}(\sigma,\sigma') \equiv \#(i \leq N : \sigma_i \neq \sigma'_i)$, namely $R_N(\sigma,\sigma') = (1-2N^{-1}d_{HAM}(\sigma,\sigma'))$.

More general class:

$$\mathrm{cov}(H_N(\sigma), H_N(\sigma')) = N\xi(R_N(\sigma,\sigma')) \tag{52}$$

normalized such that $\xi(1) = 1$. p-spin SK-models, which are obtained by choosing $\xi(x) = |x|^p$.

$$H_N^{p-SK}(\sigma) = \frac{-1}{\sqrt{N^{p-1}}} \sum_{1\leq i_1,\dots,i_p\leq N} J_{i_1\dots i_p}\sigma_{i_1}\dots\sigma_{i_p} \tag{53}$$

As we will see later, the difficulties in studying the statistical mechanics of these models is closely linked to the understanding of the extremal properties of the corresponding random processes. While Gaussian processes have been heavily analyzed in the mathematical literature (see e.g. [22, 1]), the known results were not enough to recover the heuristic results obtained in the physics literature. This is one reason why this particular field of mean-field spin-glass models has considerable intrinsic interest for mathematics.

2.2 The generalized random energy models

Further classes of models: Use different distances!

Lexicographic distance:

$$d_N(\sigma,\tau) \equiv N^{-1}\left(\min(i : \sigma_i \neq \tau_i) - 1\right) \tag{54}$$

is analogous to the overlap $R_N(\sigma,\tau)$. The corresponding Gaussian processes are then characterized by covariances given by

$$\mathrm{cov}(H_N(\sigma), H_N(\tau)) = NA(d_N(\sigma,\tau)) \tag{55}$$

where A can be chosen to be any non-decreasing function on $[0,1]$, and can be thought of as a probability distribution function. The choice of the lexicographic distance entails some peculiar features. First, this distance is an *ultrametric*, i.e. for any three configurations σ,τ,ρ,

$$d_N(\sigma,\tau) = \min\left(d_N(\sigma,\rho), d_N(\tau,\rho)\right) \tag{56}$$

3 The simples example: the REM

We set

$$H_N(\sigma) = -\sqrt{N} X_\sigma \tag{57}$$

where X_σ, $\sigma \in \mathcal{S}_N$, are 2^N i.i.d. standard normal random variables.

3.1 Ground-state energy and free energy

Lemma 8. *The family of random variables introduced above satisfies*

$$\lim_{N\uparrow\infty} \max_{\sigma\in\mathcal{S}_N} N^{-1/2} X_\sigma = \sqrt{2\ln 2} \tag{58}$$

both almost surely and in mean.

Proof. Since everything is independent,

$$\mathbb{P}\left[\max_{\sigma\in\mathcal{S}_N} X_\sigma \leq u\right] = \left(1 - \frac{1}{\sqrt{2\pi}}\int_u^\infty e^{-x^2/2}dx\right)^{2^N} \tag{59}$$

and we just need to know how to estimate the integral appearing here. This is something we should get used to quickly, as it will occur all over the place. It will always be done using the fact that, for $u > 0$,

$$\frac{1}{u}e^{-u^2/2}\left(1 - 2u^{-2}\right) \leq \int_u^\infty e^{-x^2/2}dx \leq \frac{1}{u}e^{-u^2/2} \tag{60}$$

$$\frac{2^N}{\sqrt{2\pi}}\int_{u_N(x)}^\infty e^{-z^2/2}dx = e^{-x} \tag{61}$$

then (for $x > -\ln N/\ln 2$)

$$u_N(x) = \sqrt{2N\ln 2} + \frac{x}{\sqrt{2N\ln 2}} - \frac{\ln(N\ln 2) + \ln 4\pi}{2\sqrt{2N\ln 2}} + o(1/\sqrt{N}) \tag{62}$$

Thus

$$\mathbb{P}\left[\max_{\sigma\in\mathcal{S}_N} X_\sigma \leq u_N(x)\right] = \left(1 - 2^{-N}e^{-x}\right)^{2^N} \to e^{-e^{-x}} \tag{63}$$

In other terms, the random variable $u_N^{-1}\left(\max_{\sigma\in\mathcal{S}_N} X_\sigma\right)$ converges in distribution to a random variable with double-exponential distribution

Next we turn to the analysis of the partition function.

$$Z_{\beta,N} \equiv 2^{-N}\sum_{\sigma\in\mathcal{S}_N} e^{\beta\sqrt{N}X_\sigma} \tag{64}$$

A first guess would be that a *law of large numbers* might hold, implying that $Z_{\beta,N} \sim \mathbb{E}Z_{\beta,N}$, and hence

$$\lim_{N\uparrow\infty} \Phi_{\beta,N} = \lim_{N\uparrow\infty} \frac{1}{N}\ln \mathbb{E}Z_{\beta,N} = \frac{\beta^2}{2}, \text{a.s.} \tag{65}$$

Holds only for small enough values of β!

Theorem 1. *In the REM,*

$$\lim_{N\uparrow\infty}\mathbb{E}\Phi_{\beta,N}=\begin{cases}\frac{\beta^2}{2}, & \text{for } \beta\le\beta_c\\ \frac{\beta_c^2}{2}+(\beta-\beta_c)\beta_c, & \text{for } \beta\ge\beta_c\end{cases}\tag{66}$$

where $\beta_c=\sqrt{2\ln 2}$.

Proof. We use the method of truncated second moments.

We will first derive an upper bound for $\mathbb{E}\Phi_{\beta,N}$. Note first that by Jensen's inequality, $\mathbb{E}\ln Z\le\ln\mathbb{E}Z$, and thus

$$\mathbb{E}\Phi_{\beta,N}\le\frac{\beta^2}{2}\tag{67}$$

On the other hand we have that

$$\begin{aligned}\mathbb{E}\frac{d}{d\beta}\Phi_{\beta,N}&=N^{-1/2}\mathbb{E}\frac{\mathbb{E}_\sigma X_\sigma e^{\beta\sqrt{N}X_\sigma}}{Z_{\beta,N}}\\&\le N^{-1/2}\mathbb{E}\max_{\sigma\in\mathcal{S}_N}X_\sigma\le\beta\sqrt{2\ln 2}(1+C/N)\end{aligned}\tag{68}$$

for some constant C. Combining (67) and (68), we deduce that

$$\mathbb{E}\Phi_{\beta,N}\le\inf_{\beta_0\ge 0}\begin{cases}\frac{\beta^2}{2}, & \text{for } \beta\le\beta_0\\ \frac{\beta_0^2}{2}+(\beta-\beta_0)\sqrt{2\ln 2}(1+C/N), & \text{for } \beta\ge\beta_0\end{cases}\tag{69}$$

It is easy to see that the infimum is realized (ignore the C/N correction) for $\beta_0=\sqrt{2\ln 2}$. This shows that the right-hand side of (66) is an upper bound.

It remains to show the corresponding lower bound. Note that, since $\frac{d^2}{d\beta^2}\Phi_{\beta,N}\ge 0$, the slope of $\Phi_{\beta,N}$ is non-decreasing, so that the theorem will be proven if we can show that $\Phi_{\beta,N}\to\beta^2/2$ for all $\beta<\sqrt{2\ln 2}$, i.e. that the law of large numbers holds up to this value of β. A natural idea to prove this is to estimate the variance of the partition function One would compute

$$\begin{aligned}\mathbb{E}Z_{\beta,N}^2&=\mathbb{E}_\sigma\mathbb{E}_{\sigma'}\mathbb{E}e^{\beta\sqrt{N}(X_\sigma+X_{\sigma'})}\\&=2^{-2N}\left(\sum_{\sigma\neq\sigma'}e^{N\beta^2}+\sum_\sigma e^{2N\beta^2}\right)\\&=e^{N\beta^2}\left[(1-2^{-N})+2^{-N}e^{N\beta^2}\right]\end{aligned}\tag{70}$$

where all we used is that for $\sigma\neq\sigma'$ X_σ and $X_{\sigma'}$ are independent. The second term in the square brackets is exponentially small if and only if $\beta^2<\ln 2$. For such values of β we have that

$$\begin{aligned}\mathbb{P}\left[\left|\ln \frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}}\right| > \epsilon N\right] &= \mathbb{P}\left[\frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} < e^{-\epsilon N} \text{ or } \frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} > e^{\epsilon N}\right] \\ &\leq \mathbb{P}\left[\left(\frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}} - 1\right)^2 > \left(1-e^{-\epsilon N}\right)^2\right] \\ &\leq \frac{\mathbb{E}Z^2_{\beta,N}/(\mathbb{E}Z_{\beta,N})^2 - 1}{(1-e^{-\epsilon N})^2} \\ &\leq \frac{2^{-N} + 2^{-N}e^{N\beta^2}}{(1-e^{-\epsilon N})^2} \end{aligned} \tag{71}$$

which is more than enough to get (65). But of course this does not correspond to the critical value of β claimed in the proposition!

Instead of the second moment of Z one should compute a truncated version of it, namely, for $c \geq 0$,

$$\widetilde{Z}_{\beta,N}(c) \equiv \mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma} \mathbb{1}_{X_\sigma < c\sqrt{N}} \tag{72}$$

An elementary computation using (60) shows that, if $c > \beta$, then

$$\mathbb{E}\widetilde{Z}_{\beta,N}(c) = e^{\frac{\beta^2 N}{2}}\left(1 - \frac{e^{-N\beta^2/2}}{\sqrt{2\pi N}(c-\beta)}(1+O(1/N)\right) \tag{73}$$

so that such a truncation essentially does not influence the mean partition function. Now compute the mean of the square of the truncated partition function (neglecting irrelevant $O(1/N)$ errors):

$$\mathbb{E}\widetilde{Z}^2_{\beta,N}(c) = (1-2^{-N})[\mathbb{E}\widetilde{Z}_{\beta,N}(c)]^2 + 2^{-N}\mathbb{E}e^{\beta\sqrt{N}2X_\sigma}\mathbb{1}_{X_\sigma < c\sqrt{N}}) \tag{74}$$

where

$$\mathbb{E}\, e^{2\beta\sqrt{N}X_\sigma}\mathbb{1}_{X_\sigma < c\sqrt{N}} = \begin{cases} e^{2\beta^2 N}, & \text{if } 2\beta < c \\ 2^{-N}\frac{e^{2c\beta N - \frac{c^2 N}{2}}}{(2\beta - c)\sqrt{2\pi N}}, & \text{otherwise,} \end{cases} \tag{75}$$

Combined with (73) this implies that, for $c/2 < \beta < c$,

$$\frac{2^{-N}\mathbb{E}\, e^{2\beta\sqrt{N}X_\sigma}\mathbb{1}_{X_\sigma < c\sqrt{N}}}{\left(\mathbb{E}\,\widetilde{Z}_{N,\beta}\right)^2} = \frac{e^{-N(c-\beta)^2 - N(2\ln 2 - c^2)/2}}{(2\beta - c)\sqrt{N}} \tag{76}$$

Therefore, for all $c < \sqrt{2\ln 2}$, and all $\beta < c$,

$$\mathbb{E}\left[\frac{\widetilde{Z}_{\beta,N}(c) - \mathbb{E}\widetilde{Z}_{\beta,N}(c)}{\mathbb{E}\widetilde{Z}_{\beta,N}(c)}\right]^2 \leq e^{-Ng(c,\beta)} \tag{77}$$

with $g(c,\beta) > 0$. Thus Chebyshev's inequality implies that

$$\mathbb{P}\left[|\widetilde{Z}_{\beta,N}(c) - \mathbb{E}\widetilde{Z}_{\beta,N}(c)| > \delta\mathbb{E}\widetilde{Z}_{\beta,N}(c)\right] \leq \delta^{-2}e^{-Ng(c,\beta)} \tag{78}$$

and so, in particular,

$$\lim_{N\uparrow\infty}\frac{1}{N}\mathbb{E}\ln\widetilde{Z}_{\beta,N}(c) = \lim_{N\uparrow\infty}\frac{1}{N}\ln\mathbb{E}\widetilde{Z}_{\beta,N}(c) \tag{79}$$

for all $\beta < c < \sqrt{2\ln 2} = \beta_c$. But this implies that for all $\beta < \beta_c$, we can chose c such that

$$\lim_{N\uparrow\infty}\frac{1}{N}\ln\mathbb{E}Z_{\beta,N} \geq \lim_{N\uparrow\infty}\frac{1}{N}\ln\mathbb{E}\widetilde{Z}_{\beta,N}(c) = \frac{\beta^2}{2} \tag{80}$$

This proves the theorem.

3.2 Fluctuations and limit theorems

Theorem 1. *Let $\mathcal{P}$ denotes the Poisson point process on $\mathbb{R}$ with intensity measure $e^{-x}dx$. Then, in the REM, with $\alpha = \beta/\sqrt{2\ln 2}$, if $\beta > \sqrt{2\ln 2}$,*

$$e^{-N[\beta\sqrt{2\ln 2}-\ln 2]+\frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]}Z_{\beta,N} \stackrel{\mathcal{D}}{\to} \int_{-\infty}^{\infty}e^{\alpha z}\mathcal{P}(dz) \tag{81}$$

and

$$N\left(\Phi_{\beta,N} - \mathbb{E}\Phi_{\beta,N}\right) \stackrel{\mathcal{D}}{\to} \ln\int_{-\infty}^{\infty}e^{\alpha z}\mathcal{P}(dz) - \mathbb{E}\ln\int_{-\infty}^{\infty}e^{\alpha z}\mathcal{P}(dz). \tag{82}$$

Proof. Basically, the idea is very simple. We expect that for β large, the partition function will be dominated by the configurations σ corresponding to the largest values of X_σ. Thus we split $Z_{\beta,N}$ carefully into

$$Z^x_{N,\beta} \equiv \mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma}\mathbb{1}_{\{X_\sigma \leq u_N(x)\}} \tag{83}$$

and $Z_{\beta,N} - Z^x_{\beta,N}$. Let us first consider the last summand. We introduce the random variable

$$\mathcal{W}_N(x) = Z_{\beta,N} - Z^x_{\beta,N} = 2^{-N}\sum_{\sigma\in\mathcal{S}_N}e^{\beta\sqrt{N}X_\sigma}\mathbb{1}_{\{X_\sigma > u_N(x)\}} \tag{84}$$

It is convenient to rewrite this as (we ignore the sub-leading corrections to $u_N(x)$ and only keep the explicit part of (62))

$$\begin{aligned}\mathcal{W}_N(x) &= 2^{-N}\sum_{\sigma\in\mathcal{S}_N}e^{\beta\sqrt{N}u_N(u_N^{-1}(X_\sigma))}\mathbb{1}_{\{u_N^{-1}(X_\sigma)>x\}} \\ &= e^{N(\beta\sqrt{2\ln 2}-\ln 2)-\frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]} && (85)\\ &\quad\times\sum_{\sigma\in\mathcal{S}_N}e^{\alpha u_N^{-1}(X_\sigma)}\mathbb{1}_{\{u_N^{-1}(X_\sigma)>x\}} && (86)\\ &\equiv \frac{1}{C(\beta,N)}\sum_{\sigma\in\mathcal{S}_N}e^{\alpha u_N^{-1}(X_\sigma)}\mathbb{1}_{\{u_N^{-1}(X_\sigma)>x\}} && (87)\end{aligned}$$

where

$$\alpha \equiv \beta/\sqrt{2\ln 2} \tag{88}$$

and $C(b,N)$ is defined through the last identity. The key to most of what follows relies on the famous result on the convergence of the extreme value process to a Poisson point process (for a proof see, e.g., [21]):

Theorem 2. *Let $\mathcal{P}_N$ be point process on $\mathbb{R}$ given by*

$$\mathcal{P}_N \equiv \sum_{\sigma\in\mathcal{S}_N} \delta_{u_N^{-1}(X_\sigma)} \tag{89}$$

Then $\mathcal{P}_N$ converges weakly to a Poisson point process on $\mathbb{R}$ with intensity measure $e^{-x}dx$.

Clearly, the weak convergence of $\mathcal{P}_N$ to $\mathcal{P}$ implies convergence in law of the right-hand side of (85), provided that $e^{\alpha x}$ is integrable on $[x,\infty)$ w.r.t. the Poisson point process with intensity e^{-x}. This is, in fact, never a problem: the Poisson point process has almost surely support on a finite set, and therefore $e^{\alpha x}$ is always a.s. integrable. Note, however, that for $\beta \geq \sqrt{2\ln 2}$ the mean of the integral is infinite, indicating the passage to the low-temperature regime.

Lemma 9. *Let $\mathcal{W}_N(x), \alpha$ be defined as above, and let $\mathcal{P}$ be the Poisson point process with intensity measure $e^{-z}dz$. Then*

$$C(\beta,N)\mathcal{W}_N(x) \xrightarrow{\mathcal{D}} \int_x^\infty e^{\alpha z}\mathcal{P}(dz) \tag{90}$$

Next we show that the contribution of the truncated part of the partition function is negligible compared to this contribution. For this it is enough to compute the mean values

$$\begin{aligned}
\mathbb{E}Z^x_{\beta,N} &\sim e^{N\beta^2/2} \int_{-\infty}^{u_N(x)-\beta\sqrt{N}} \frac{dz}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \\
&\sim e^{N\beta^2/2} \frac{e^{-(u_N(x)-\beta\sqrt{N})^2/2}}{\sqrt{2\pi}(\beta\sqrt{N}-u_N(x))} \\
&\sim \frac{2^{-N}e^{x(\alpha-1)}}{\alpha-1} e^{N(\beta\sqrt{2\ln 2}-\ln 2)-\frac{\alpha}{2}[\ln(N\ln 2)+\ln 4\pi]} \\
&= \frac{e^{x(\alpha-1)}}{\alpha-1}\frac{1}{C(\beta,N)}
\end{aligned} \tag{91}$$

so that

$$C(\beta,N)\mathbb{E}Z^x_{\beta,N} \sim \frac{e^{x(\alpha-1)}}{\alpha-1}$$

which tends to zero as $x \downarrow -\infty$, and so $C(\beta,N)\mathbb{E}Z^x_{\beta,N}$ converges to zero in probability. The assertions of Theorem 1 follow.

3.3 The Gibbs measure

A nice way to do this consists in mapping the hypercube to the interval $(0,1]$ via

$$\mathcal{S}_N \ni \sigma \to r_N(\sigma) \equiv 1 - \sum_{i=1}^{N} (1-\sigma_i) 2^{-i-1} \in (0,1] \tag{92}$$

Define the pure point measure $\tilde{\mu}_{\beta,N}$ on $(0,1]$ by

$$\tilde{\mu}_{\beta,N} \equiv \sum_{\sigma \in \mathcal{S}_N} \delta_{r_N(\sigma)} \mu_{\beta,N}(\sigma) \tag{93}$$

Our results will be expressed in terms of the convergence of these measures. It will be understood in the sequel that the space of measures on $(0,1]$ is equipped with the topology of weak convergence, and all convergence results hold with respect to this topology.

Let us introduce the Poisson point process $\mathcal{R}$ on the strip $(0,1] \times \mathbb{R}$ with intensity measure $\frac{1}{2} dy \times e^{-x} dx$. If (Y_k, X_k) denote the atoms of this process, define a new point process $\mathcal{M}_\alpha$ on $(0,1] \times (0,1]$ whose atoms are (Y_k, w_k), where

$$w_k \equiv \frac{e^{\alpha X_k}}{\int \mathcal{R}(dy,dx) e^{\alpha x}} \tag{94}$$

for $\alpha > 1$. With this notation we have that:

Theorem 3. *If $\beta > \sqrt{2 \ln 2}$, with $\alpha = \beta / \sqrt{2 \ln 2}$, then*

$$\tilde{\mu}_{\beta,N} \xrightarrow{\mathcal{D}} \tilde{\mu}_\beta \equiv \int_{(0,1] \times (0,1]} \mathcal{M}_\alpha(dy,dw) \delta_y w \tag{95}$$

Proof. With $u_N(x)$ defined in (62), we define the point process $\mathcal{R}_N$ on $(0,1] \times \mathbb{R}$ by

$$\mathcal{R}_N \equiv \sum_{\sigma \in \mathcal{S}_N} \delta_{(r_N(\sigma), u_N^{-1}(X_\sigma))} \tag{96}$$

A standard result of extreme value theory (see [21], Theorem 5.7.2) is easily adapted to yield that

$$\mathcal{R}_N \xrightarrow{\mathcal{D}} \mathcal{R}, \text{ as } N \uparrow \infty \tag{97}$$

Note that

$$\mu_{\beta,N}(\sigma) = \frac{e^{\alpha u_N^{-1}(X_\sigma)}}{\sum_\sigma e^{\alpha u_N^{-1}(X_\sigma)}} = \frac{e^{\alpha u_N^{-1}(X_\sigma)}}{\int \mathcal{R}_N(dy,dx) e^{\alpha x}} \tag{98}$$

Since $\int \mathcal{R}_N(dy,dx) e^{\alpha x} < \infty$ a.s., we can define the point process

$$\mathcal{M}_{\alpha,N} \equiv \sum_{\sigma \in \mathcal{S}_N} \delta_{\left(r_N(\sigma), \frac{\exp(\alpha u_N^{-1}(X_\sigma))}{\int \mathcal{R}_N(dy,dx) \exp(\alpha x)}\right)} \tag{99}$$

on $(0,1]\times(0,1]$. Then

$$\tilde{\mu}_{\beta,N} = \int \mathcal{M}_{\alpha,N}(dy,dw)\delta_y w \tag{100}$$

The only non-trivial point in the convergence proof is to show that the contribution to the partition functions in the denominator from atoms with $u_N(X_\sigma) < x$ vanishes as $x \downarrow -\infty$. But this is precisely what we have shown to be the case in the proof of part of Theorem 1. Standard arguments then imply that first $\mathcal{M}_{\alpha,N} \stackrel{\mathcal{D}}{\to} \mathcal{M}_\alpha$, and consequently, (95).

The measure $\tilde{\mu}_\beta$ is in fact closely related to a classical object in probability theory, the α-stable Lévy subordinator. To see this, denote by

$$\mathcal{Z}_\alpha(t) \equiv \int_0^t \int_{-\infty}^{+\infty} e^{\alpha x}\mathcal{R}(dy,dx). \tag{101}$$

Clearly, the probability distribution function associated to the measure $\tilde{\mu}_\beta$ satisfies, for $t\in[0,1]$,

$$\int_0^t \mu_\beta(dx) = \frac{\mathcal{Z}_\alpha(t)}{\mathcal{Z}_\alpha(1)}. \tag{102}$$

Lemma 10. *For any $0<\alpha<1$, the stochastic process $\mathcal{Z}_\alpha(t)$ is the α-stable levy process (subordinator) with Lévy measure $y^{-1/\alpha-1}dy$.*

Proof. There are various ways to prove this result. Note first that the process has independent, identically distributed increments. It is then enough, e.g. to compute the Laplace transform of the one-dimensional distribution, i.e. one shows that

$$\mathbb{E}e^{-\lambda\mathcal{Z}_\alpha(t)} = \exp\left(\int_0^\infty (e^{-\lambda y}-1)y^{-1/\alpha-1}dy\right). \tag{103}$$

This can be done by elementary calculus and is left to the reader.

Let us note that is is not difficult to show that the process

$$\sum_{\sigma\in\mathcal{S}_N} e^{\alpha u_N^{-1}(X_\sigma)}\mathbb{1}_{r_N(\sigma)\le t} \tag{104}$$

converges in the Skorokhod J_1-topology to the α-stable Lévy subordinator. Hence the distribution function of the measures $\tilde{\mu}_{\beta,N}$ to $\tilde{\mu}_\beta$ can be interpreted in the sense of the corresponding convergence of their distribution functions a stochastic processes on Skorokhod space.

3.4 The asymptotic model

In the case $\beta > 1$, we can readily interpret our asymptotic results in terms of an new statistical mechanics model *for the infinite volume limit*. It has the following ingredients:

- State space: $\mathbb{N}$;
- Random Hamiltonian: $x : \mathbb{N} \to \mathbb{R}$, where x_i is the i-th atom of the Poisson process $\mathcal{P}$;
- Temperature: $1/\alpha = \beta_c/\beta$;
- Partition function: $\mathcal{Z}_\alpha = \sum_{i\in\mathbb{N}} e^{\alpha x_i}$;
- Gibbs measure: $\hat{\mu}_\sigma(i) = \mathcal{Z}_\alpha e^{\alpha x_i}$.

Our convergence results so far can be interpreted in terms of this model as follows:

- The partition function of the REM converges, after division by $\exp(\beta\sqrt{N}u_N(0))$, to $\mathcal{Z}_\alpha$;
- If we map the Gibbs measure $\hat{\mu}\alpha$ to the unit interval via

$$\hat{\mu}_\alpha \to \check{\mu}_\alpha = \sum_{i\in\mathbb{N}} \delta_{U_i} \hat{\mu}_\alpha(i), \tag{105}$$

  where U_i, $i \in \mathbb{N}$, is a family of independent random variables that are distributed uniformly on the interval $[0, 1]$. Then $\check{\mu}_\sigma$ has the same distribution as $\tilde{\mu}_\beta$.

This is a reasonably satisfactory picture. What is lacking, however, is a proper reflection of the geometry of the Gibbs measure on the hypercube. Clearly the convergence of the embedded measures on the unit interval is insufficient to capture this.

In the next section we will see how this should be incorporated.

3.5 The replica overlap

If we want to discuss the geometry of Gibbs measures, we first must decide on how to measure distance on the hypercube. The most natural one is the Hamming distance, or its counterpart, the *overlap*, $R_N(\sigma, \sigma')$. Of course we might also want to use the ultrametric, distance d_N, and we will comment on this later.

To describe the geometry of $\mu_{\beta,N}$, we may now ask how much mass one finds in a neighborhood of a point $\sigma \in \mathcal{S}_N$, i.e. we may define

$$\phi_{\beta,N}(\sigma, t) \equiv \mu_{\beta,N} \left(R_N(\sigma, \sigma') > t \right). \tag{106}$$

Clearly this defines a probability distribution on $[-1, 1]$ (as we will see, in reality if will give zero mass to the negative numbers).

Of course these 2^N functions are not very convenient. The first reflex will be to average over the Gibbs measures, i.e. to define

$$\psi_{\beta,N}(t) \equiv \sum_{\sigma\in\mathcal{S}_N} \mu_{\beta,N}(\sigma)\phi_{\beta,N}(\sigma,t) = \mu_{\beta,N}[\omega]\otimes\mu_{\beta,N}[\omega]\left(R_N(\sigma,\sigma')\in dz\right). \tag{107}$$

The following theorem expresses the limit of ψ in the form we would expect, namely in terms of the asymptotic model.

Theorem 4. *For all* $\beta > \sqrt{2\ln 2}$

$$\psi_{\beta,N}(t) \stackrel{\mathcal{D}}{\to} \begin{cases} 0, & \text{if, } t<0;\\ 1-\sum_{i\in\mathbb{N}}\hat\mu_\alpha(i)^2, & \text{if } 0\le t<1,\\ 1, & \text{if } t\ge 1.\end{cases} \tag{108}$$

Proof. The only new thing we have to show is that the function ψ increases only at 0 and at one; that is to say, we have to show that with probability tending to one the overlap $\mathbb{R}_N$ takes on only the values one or zero.

We write for any $\Delta\subset[-1,1]$

$$\psi_{\beta,N}(\Delta)\equiv\mu_{\beta,N}^{\otimes 2}(R_N\in\Delta)\equiv\psi_{\beta,N}(\Delta) = Z_{\beta,N}^{-2}\mathbb{E}_\sigma\mathbb{E}_{\sigma'}\sum_{\substack{t\in\Delta\\ R_N(\sigma,\sigma')=t}} e^{\beta\sqrt{N}(X_\sigma+X_{\sigma'})} \tag{109}$$

We use the truncation introduced in Section 3.2. Note first that, for any interval Δ,

$$\left|\psi_{\beta,N}(\Delta) - Z_{\beta,N}^{-2}\mathbb{E}_\sigma\mathbb{E}_{\sigma'}\sum_{\substack{t\in\Delta\\ R_N(\sigma,\sigma')=t}} \mathbb{1}_{X_\sigma,X_{\sigma'}\ge u_N(x)}e^{\beta\sqrt{N}(X_\sigma+X_{\sigma'})}\right| \le \frac{2Z_{\beta,N}^x}{Z_{\beta,N}} \tag{110}$$

We have already seen in the proof of Theorem (1) that the right-hand side of (110) tends to zero in probability, as first $N\uparrow\infty$ and then $x\downarrow-\infty$. On the other hand, for $t\neq 1$,

$$\begin{aligned}&\mathbb{P}\left[\exists_{\sigma,\sigma,:R_N(\sigma,\sigma')=t}:\, X_\sigma>u_N(x)\wedge X_{\sigma'}>u_N(x)\right] \\ &\le \mathbb{E}_{\sigma\sigma'}\mathbb{1}_{R_N(\sigma,\sigma')=t}\,2^{2N}\mathbb{P}\left[X_\sigma>u_N(x)\right]^2 = \frac{2e^{-I(t)N}e^{-2x}}{\sqrt{2\pi N}\sqrt{1-t^2}}\end{aligned} \tag{111}$$

by the definition of $u_N(x)$ (see (61)). This implies again that any interval $\Delta\subset[-1,1)\cup[-1,0)$ has zero mass. To conclude the proof it is enough to compute $\psi_{\beta,N}(1)$. Clearly

$$\psi_{\beta,N}(1) = \frac{2^{-N}\mathbb{Z}_{2\beta,N}}{Z_{\beta,N}^2} \tag{112}$$

By Theorem 1, one sees easily that

$$\psi_{\beta,N}(1) \xrightarrow{\mathcal{D}} \frac{\int e^{2\alpha z}\mathcal{P}(dz)}{\left(\int e^{\alpha z}\mathcal{P}(dz)\right)^2} \tag{113}$$

It is now very easy to conclude the proof.

The empirical distance distribution.

Rather than just taking the mean of the functions $\phi_{\beta,N}$, we can naturally define their empirical distribution. It is natural to do this biased with their importance in the Gibbs measures. This lead to the object

$$\mathcal{K}_{\beta,N} \equiv \sum_{\sigma\in\mathcal{S}_N} \mu_{\beta,N}(\sigma)\delta_{\phi_{\beta,N}(\sigma,\cdot)}. \tag{114}$$

Here we think of the δ-measure as a measure on probability distribution functions, respectively probability measures on $[-1,1]$, and to $\mathcal{K}_{\beta,N}$ is a random probability measure on the same space. This measure carries a substantial amount of information on the geometry of the Gibbs measure and is in fact the fundamental object to study.

Note that we can define an analogous object in the asymptotic model. We just have to decide on how to measure distance, or overlap, between the points in $\mathbb{N}$. In view of the results above, the natural choice is to say that the overlap between a point and itself is one, and is zero between different points. Then set

$$\mathcal{K}_\alpha = \sum_{i\in\mathbb{N}} \hat{\mu}_\alpha(i)\delta_{\hat{(1-\mu_\alpha(i))}\mathbb{1}\{\cdot\in[0,1)\}+\hat{\mu}_\alpha(i)\mathbb{1}\{\cdot\geq 1\}}. \tag{115}$$

A fairly simple to prove extension of Theorem 4 gives the strongest link between the REM and the asymptotic model.

Theorem 5. *With the standard relation between α and β,*

$$\mathcal{K}_{\beta,N} \to \mathcal{K}_\alpha, \tag{116}$$

where the convergence is in distribution with respect to weak topology of measures on the space to distribution functions equipped with the weak topology.

4 Derrida's Generalized Random Energy models

We will now turn to the investigation of the second class of Gaussian models we have mentioned above, namely Gaussian processes whose covariance is a function of the lexicographic distance on the hypercube (see (55)). B. Derrida introduced these models in the case where A is a step function with finitely many jumps as a natural generalization of the REM and called it the *generalized random energy model* (GREM)[11, 13, 14, 15]. The presentation below is based on results obtained with I. Kurkova [6, 7, 8].

4.1 The GREM and Poisson cascades

A key in the analysis of the REM was the theory of convergence to Poisson processes of the extreme value statistics of (i.i.d.) random variables. In the GREM, analogous results will be needed in the correlated case.

We assume that A is the distribution function of a measure that is supported on a finite number, n, of points $x_1, \dots, x_n \in [0,1]$, as shown in Figure 1. In that case we denote the mass of the atoms x_i by a_i, and we set

$$\ln \alpha_i = (x_i - x_{i-1}) \ln 2, \quad i = 1, \dots, n \tag{117}$$

where $x_0 \equiv 0$. We normalize in such a way that $\sum_{i=1}^n a_i = 1$, and $\prod_{i=1}^n \alpha_i = 2$.

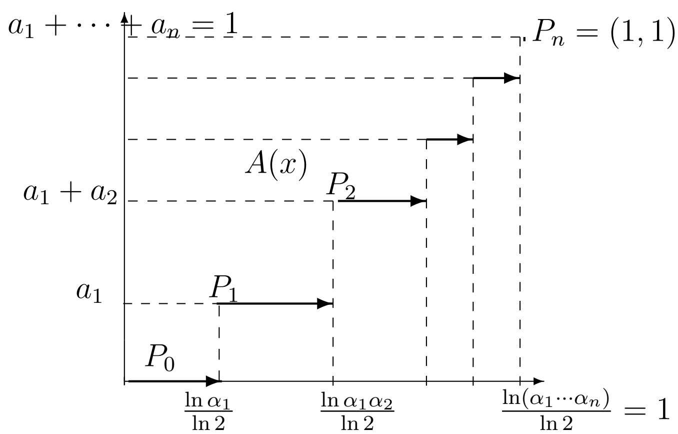

Fig. 1. The function $A(x)$.

It is very useful that there is an explicit representation of the corresponding process X_σ. We write $\sigma = \sigma_1 \sigma_2 \dots \sigma_n$ where $\sigma_i \in \mathcal{S}_{N \ln \alpha_i / \ln 2}$. Usually we will assume that $x_1 > 0$, $x_n = 1$, and all $a_i > 0$.

Then the Gaussian process X_σ can be constructed from independent standard Gaussian random variables $X_{\sigma_1}, X_{\sigma_1 \sigma_2}, \dots, X_{\sigma_1 \dots \sigma_2}$,where $\sigma_i \in \{-1,1\}^{N \ln \alpha_i / \ln 2}$ as

$$X_\sigma \equiv \sqrt{a_1} X_{\sigma_1} + \sqrt{a_2} X_{\sigma_1 \sigma_2} + \cdots + \sqrt{a_n} X_{\sigma_1 \sigma_2 \dots \sigma_n}, \quad \text{if } \sigma = \sigma_1 \sigma_2 \dots \sigma_n. \tag{118}$$

4.2 Poisson cascades and extremal processes

Our first concern is to understand the structure of the extremes of such processes. The key ideas are easiest understood in the case where $n = 2$. Let us consider the set, S_x, of σ_1 for which $X_{\sigma_1} \sim \sqrt{a_1 2N \ln \alpha_1 x}$. We know that the cardinality of this set is rather precisely $\alpha_1^{N(1-x)}$ if $x < 1$. Now all the

$\alpha_1^{N(1-x)}\alpha_2^N = 2^N\alpha_1^{-xN}$ random variables $X_{\sigma_1\sigma_2}$ with $\sigma_1 \in S_x$ are independent, so that we know that their maximum is roughly $\sqrt{2a_2N(\ln 2 - x\ln\alpha_1)}$. Hence, the maximum of the X_σ with $\sigma_1 \in S_x$ is

$$\sqrt{a_1 2N \ln a_1 x} + \sqrt{2a_2N(\ln 2 - x\ln\alpha_1)} \tag{119}$$

Finally, to determine the overall maximum, it suffices to find the value of x that maximizes this quantity, which turns out to be given by $x^* = \frac{a_1\ln 2}{\ln\alpha_1}$, provided the constraint $\frac{a_1\ln 2}{\ln\alpha_1} < 1$ is satisfied. In that case we also find that

$$\sqrt{a_1 2N \ln a_1 x} + \sqrt{2a_2N(\ln 2 - x^*\ln\alpha_1)} = \sqrt{2\ln 2} \tag{120}$$

i.e. the same value as in the REM. On the other hand, if $\frac{a_1\ln 2}{\ln\alpha_1} > 1$, the maximum is realized by selecting the largest values in the first generation, corresponding to $x = 1$, and then for each of them the extremal members of the corresponding second generation. The value of the maximum is then (roughly)

$$\sqrt{a_1 2N \ln a_1} + \sqrt{2a_2N\ln\alpha_2} \leq \sqrt{2\ln 2} \tag{121}$$

where equality holds only in the borderline case $\frac{a_1\ln 2}{\ln\alpha_1} = 1$, which requires more care. The condition $\frac{a_1\ln 2}{\ln\alpha_1} < 1$ has a nice interpretation: it simply means that the function $A(x) < x$, for all $x \in (0,1)$.

In terms of the point processes, the above considerations suggest the following picture (which actually holds true): If $\frac{a_1\ln 2}{\ln\alpha_1} < 1$, the point process

$$\sum_{\sigma\mathcal{S}_N} \delta_{u_N^{-1}(X_\sigma)} \to \mathcal{P} \tag{122}$$

exactly as in the REM, while in the opposite case this process would surely converge to zero. On the other hand, we can construct (in both cases) another point processes,

$$\sum_{\sigma=\sigma_1\sigma_2\in\{-1,+1\}^N} \delta_{\sqrt{a_1}u^{-1}_{\ln\alpha_1,N}(X_{\sigma_1}+\sqrt{a_2}u^{-1}_{\ln\alpha_2,N}(X_{\sigma_1\sigma_2})} \tag{123}$$

where we set

$$u_{\alpha,N}(x) \equiv u_{N\ln\alpha/\ln 2}(x) \tag{124}$$

This point process will converge to a process obtained from a *Poisson cascade*: The process

$$\sum_{\sigma_1\in\{-1,+1\}^{\ln\alpha_1 N}} \delta_{u^{-1}_{\alpha_1,N}(X_{\sigma_1})} \tag{125}$$

converges to a Poisson point process, and, for any σ_1, so do the point processes

$$\sum_{\sigma_2\in\{-1,+1\}^{\ln\alpha_2 N}} \delta_{u^{-1}_{\alpha_2,N}(X_{\sigma_1\sigma_2})} \tag{126}$$

Then the two-dimensional point process

$$\sum_{\sigma=\sigma_1\sigma_2\in\{-1,+1\}^N} \delta_{(u^{-1}_{\alpha_1,N}(X_{\sigma_1}),u^{-1}_{\alpha_2,N}(X_{\sigma_1\sigma_2}))} \tag{127}$$

converges to a *Poisson cascade* in $\mathbb{R}^2$: we place the Poisson process (always with intensity measure $e^{-x}dx$) on $\mathbb{R}$, and then, for each atom, we place an independent PPP on the line orthogonal to the first line that passes through that atom. Adding up the atoms of these processes with the right weight yields the limit of the process defined in (123). Now this second point process does not yield the extremal process, as long as the first one exists, i.e. as long as the process (122) does not converge to zero. Interestingly, when we reach the borderline, the process (122) converges to the PPP with intensity $Ke^{-x}dx$ with $0 < K < 1$, while the cascade process yields points that differ from those of this process only in the sub-leading order.

Having understood the particular case of two levels, it is not difficult to figure out the general situation.

The next result tells us which Poisson point processes we can construct.

Theorem 1. *Let $0 < a_i < 1$, $\alpha_i > 1$, $i = 1, 2, \dots, n$ with $\sum_{i=1}^n a_i = 1$. Set $\bar\alpha \equiv \prod_{i=1}^n \alpha_i$. Then the point process*

$$\sum_{\sigma=\sigma_1\dots\sigma_n\in\{-1,+1\}^{N\ln\bar\alpha/\ln 2}} \delta_{u^{-1}_{\bar\alpha,N}(\sqrt{a_1}X_{\sigma_1}+\sqrt{a_2}X_{\sigma_1\sigma_2}+\dots+\sqrt{a_n}X_{\sigma_1\sigma_2\dots\sigma_n})} \tag{128}$$

converges weakly to the Poisson point process $\mathcal{P}$ on $\mathbb{R}$ with intensity measure $Ke^{-x}dx$, $K \in \mathbb{R}$, if and only if, for all $i = 2, 3, \dots, n$,

$$a_i + a_{i+1} + \dots + a_n \geq \ln(\alpha_i\alpha_{i+1}\cdots\alpha_n)/\ln\bar\alpha \tag{129}$$

Furthermore, if all inequalities in (129) are strict, then the constant $K = 1$. If some of them are equalities, then $0 < K < 1$.

Remark 3. An explicit formula for K can be found in [6].

Remark 4. The conditions (129) can be expressed as $A(x) \leq x$ for all $x \in (0, 1)$.

Theorem 2. *Let $\alpha_i \geq 1$, and set $\bar\alpha \equiv \prod_{i=1}^k \alpha_i$. Let $Y_{\sigma_1}, Y_{\sigma_1\sigma_2}, \dots, Y_{\sigma_1\dots\sigma_k}$ be identically distributed random variables, such that the vectors $(Y_{\sigma_1})_{\sigma_1\in\{-1,1\}^{N\ln\alpha_1/\ln\bar\alpha}}, (Y_{\sigma_1\sigma_2})_{\sigma_2\in\{-1,1\}^{N\ln\alpha_2/\ln\bar\alpha}}, \dots\dots\dots,$ $(Y_{\sigma_1\sigma_2\dots\sigma_k})_{\sigma_k\in\{-1,1\}^{N\ln\alpha_k/\ln\bar\alpha}}$ are independent. Let $v_{N,1}(x), \dots, v_{N,k}(x)$ be functions on $\mathbb{R}$ such that the following point processes*

$$\begin{aligned}
&\sum_{\sigma_1} \delta_{v_{N,1}(Y_{\sigma_1})} \to \mathcal{P}_1\\
&\sum_{\sigma_2} \delta_{v_{N,2}(Y_{\sigma_1\sigma_2})} \to \mathcal{P}_2 \quad \forall\sigma_1\\
&\dots\\
&\sum_{\sigma_k} \delta_{v_{N,k}(Y_{\sigma_1\sigma_2\dots\sigma_k})} \to \mathcal{P}_k \quad \forall\sigma_1\dots\sigma_{k-1}
\end{aligned} \tag{130}$$

converge weakly to Poisson point processes, $\mathcal{P}_1, \dots, \mathcal{P}_k$, on $\mathbb{R}$ with intensity measures $K_1 e^{-x}dx, \dots, K_k e^{-x}dx$, for some constants $K_1, \dots, K_k$. Then the point processes on $\mathbb{R}^k$,

$$\mathcal{P}_N^{(k)} \equiv \sum_{\sigma_1} \delta_{v_{N,1}(Y_{\sigma_1})} \sum_{\sigma_2} \delta_{v_{N,2}(Y_{\sigma_1\sigma_2})} \cdots \sum_{\sigma_k} \delta_{v_{N,k}(Y_{\sigma_1\sigma_2\dots\sigma_k})} \to \mathcal{P}^{(k)} \tag{131}$$

converge weakly to point processes $\mathcal{P}^{(k)}$ on $\mathbb{R}^k$, called Poisson cascades with k levels.

Poisson cascades are best understood in terms of the following iterative construction. If $k = 1$, it is just a Poisson point process on $\mathbb{R}$ with intensity measure $K_1 e^{-x}dx$. To construct $\mathcal{P}^{(2)}$ on $\mathbb{R}^2$, we place the process $\mathcal{P}^{(1)}$ for $k = 1$ on the axis of the first coordinate and through each of its points draw a straight line parallel to the axis of the second coordinate. Then we put on each of these lines independently a Poisson point process with intensity measure $K_2 e^{-x}dx$. These points on $\mathbb{R}^2$ form the process $\mathcal{P}^{(2)}$. This procedure is now simply iterated k times.

Theorems 1 and 2 combined tell us which are the different point processes that may be constructed in the GREM.

Theorem 3. *Let $\alpha_i \geq 1$, $0 < a_i < 1$, such that $\prod_{i=1}^n \alpha_i = 2$, $\sum_{i=1}^n a_i = 1$. Let $J_1, J_2, \dots, J_m \in \mathbb{N}$ be the indices such that $0 = J_0 < J_1 < J_2 < \cdots < J_m = n$. We denote by $\bar{a}_l \equiv \sum_{i=J_{l-1}+1}^{J_l} a_i$, $\bar{\alpha}_l \equiv \prod_{i=J_{l-1}+1}^{J_l} \alpha_i$, $l = 1, 2, \dots, m$, and set*

$$\bar{X}^{\sigma_1\dots\sigma_{J_{l-1}}}_{\sigma_{J_{l-1}+1}\sigma_{J_{l-1}+2}\cdots\sigma_{J_l}} \equiv \frac{1}{\sqrt{\bar{a}_l}} \sum_{i=1}^{J_l - J_{l-1}} \sqrt{a_{J_{l-1}+i}} X_{\sigma_{J_1}\dots\sigma_{J_{l-1}+i}} \tag{132}$$

To a partition $J_1, J_2, \dots, J_m$, we associate the function A_J obtained by joining the sequence of straight line segments going from $(x_{J_i}, A(x_{J_i})$ to $(x_{J_{+1i}}, A(x_[J_{i+1})$, $i = 0, m-1$. A partition is admissible, if $A(x) \leq A_J(x)$, for all $x \in [0,1]$. Then, for any admissible partition, the point process

$$\begin{aligned} \mathcal{P}_N^{(m)} \equiv & \sum_{\sigma_1\dots\sigma_{J_1}} \delta_{u^{-1}_{\bar{\alpha}_1,N}(\bar{X}_{\sigma_1\dots\sigma_{J_1}})} \sum_{\sigma_{J_1+1}\dots\sigma_{J_2}} \delta_{u^{-1}_{\bar{\alpha}_2,N}(\bar{X}^{\sigma_1\dots\sigma_{J_1}}_{\sigma_{J_1+1}\dots\dots\sigma_{J_2}})} \cdots \\ & \cdots \sum_{\sigma_{J_{m-1}+1}\dots\sigma_{J_m}} \delta_{u^{-1}_{\bar{\alpha}_m,N}(\bar{X}^{\sigma_1\dots\sigma_{J_{m-1}}}_{\sigma_{J_{m-1}+1}\dots\sigma_{J_m}})} \end{aligned} \tag{133}$$

converges weakly to the process $\mathcal{P}^{(m)}$ on $\mathbb{R}^m$ defined in Theorem 2 with constants $K_1, \dots, K_m$. If $A_J(x) < A(x)$, for all $x \in (x_{J_i}, x_{J_{i+1}})$, then

$$K_l = 1 \tag{134}$$

Otherwise $0 < K_l < 1$.

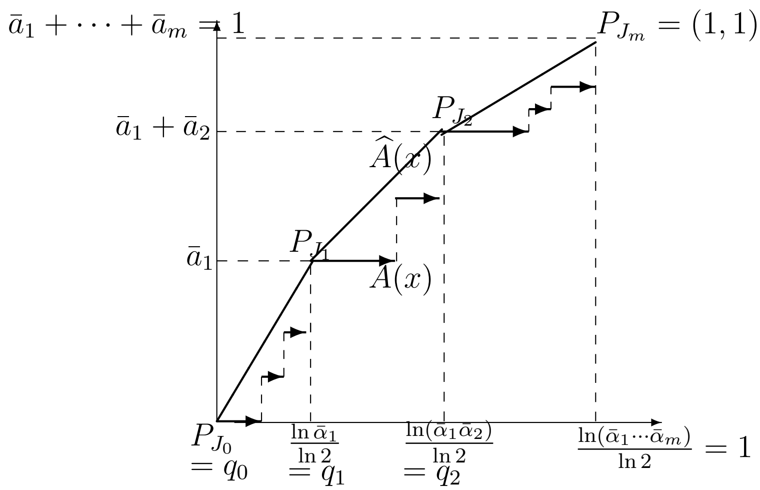

Fig. 2. The concave hull of $A(x)$.

Having constructed all possible point processes, we now find the extremal process by choosing the one that yields the largest values. It is easy to see that this is achieved if as many intermediate hierarchies as possible are grouped together. In terms of the geometrical construction just described, this means that we must choose the partition J in such a way that the function A_J has no convex pieces, i.e. that A_J is the *concave hull*, $\bar{A}$, of the function A (see Fig. 2). (The concave hull, $\bar{A}$, of a function A is the smallest concave function such that $\bar{A}(x) \geq A(x)$, for all x in the domain considered.) Algorithmically, this is achieved by setting $J_0 \equiv 0$, and

$$J_l \equiv \min\{J > J_{l-1} : A_{J_{l-1}+1,J} > A_{J+1,k} \ \forall k \geq J+1\} \tag{135}$$

where $A_{j,k} \equiv \sum_{i=j}^{k} a_i/(2\ln(\prod_{i=j}^{k}\alpha_i))$.

Set $\gamma_l \equiv \sqrt{\bar{a}_l}/\sqrt{2\ln\bar{\alpha}_l}$, $l = 1, 2, \ldots, m$. Clearly, by (135), $\gamma_1 > \gamma_2 > \cdots > \gamma_m$. Define the function $U_{J,N}$ by

$$U_{J,N}(x) \equiv \sum_{l=1}^{m} \left(\sqrt{2N\bar{a}_l \ln\bar{\alpha}_l} - N^{-1/2}\gamma_l(\ln(N(\ln\bar{\alpha}_l)) + \ln 4\pi)/2\right) + N^{-1/2}x \tag{136}$$

and the point process

$$\mathcal{E}_N \equiv \sum_{\sigma\in\{-1,1\}^N} \delta_{U_{J,N}^{-1}(X_\sigma)} \tag{137}$$

Theorem 4. *(i) The point process $\mathcal{E}_N$ converges weakly, as $N \uparrow \infty$, to the point process on $\mathbb{R}$*

$$\mathcal{E} \equiv \int_{\mathbb{R}^m} \mathcal{P}^{(m)}(dx_1, \dots, dx_m)\delta_{\sum_{l=1}^m \gamma_l x_l} \tag{138}$$

where $\mathcal{P}^{(m)}$ is the Poisson cascade introduced in Theorem 3 corresponding to the partition $J_1, \dots, J_m$ given by (135).

(ii)$\mathcal{E}$ exists, since $\gamma_1 > \cdots > \gamma_m$. It is the cluster point process on $\mathbb{R}$ containing an a.s. finite number of points in any interval $[b, \infty)$, $b \in \mathbb{R}$. The probability that there exists at least one point of $\mathcal{E}$ in the interval $[b, \infty)$ is decreasing exponentially, as $b \uparrow \infty$.

The formal proofs of these theorems can be found in [6].

4.3 Convergence of the partition function

We will now turn to the study of the Gibbs measures. Technically, the main step in the proof will be to show that the infinite-volume limit of the properly rescaled partition function can be expressed as a certain functional of Poisson cascade processes, as suggested by Ruelle [25].

For *any* sequence of indices, J_i, such that the function A^J is concave, the partition function can be written as:

$$Z_{\beta,N} = e^{\sum_{j=1}^m \left(\beta N\sqrt{2\bar{a}_j \ln \bar{\alpha}_j} - \beta\gamma_j[\ln(N\ln\bar{\alpha}_j) + \ln 4\pi]/2\right)} \times \tag{139}$$
$$\mathbb{E}_{\sigma_1\dots\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\bar{\alpha}_1,N}(\bar{X}_{\sigma_1\dots\sigma_{J_1}})} \dots \mathbb{E}_{\sigma_{J_{m-1}+1}\dots\sigma_{J_m}} e^{\beta\gamma_m u^{-1}_{\bar{\alpha}_m,N}(\bar{X}^{\sigma_1\dots\sigma_{J_{m-1}}}_{\sigma_{J_{m-1}+1}\dots\sigma_{J_m}})}$$

Clearly, not all of these representations can be useful, i.e the sums in the second line should converge to a finite random variable. For this to happen, from what we learned in the REM, each of the sums should be at 'low temperature', meaning here that $\beta\gamma_\ell > 1$. Moreover, we should expect that there is a relation to the maximum process; in fact, this will follow from the condition that $\gamma_i > \gamma_{i+1}$, for all i that appear. Thus we will have to choose the partition J that yields the extremal process, and we have to cut the representation (140) at some temperature-dependent level, $J_{l(\beta)}$, and treat the remaining hierarchies as *high-temperature* REM's, i.e. replace them by their mean value. The level $l(\beta)$ is determined by

$$l(\beta) \equiv \max\{l \geq 1 : \beta\gamma_l > 1\} \tag{140}$$

and $l(\beta) \equiv 0$ if $\beta\gamma_1 \leq 1$.

From these considerations it is now very simple to compute the asymptotics of partition function. The resulting formula for the free energy was first found in [9]:

Theorem 5. *[9] With the notation introduced above,*

$$\lim_{N\to\infty} \Phi_{\beta,N} = \beta \sum_{i=1}^{l(\beta)} \sqrt{2\bar{a}_i \ln \bar{\alpha}_i} + \sum_{i=J_{l(\beta)}+1}^{n} \beta^2 a_i/2, \text{a.s.} \tag{141}$$

The condition that for $\beta \leq \beta_c$, $l(\beta) = 0$, defines the critical temperature, $\beta_c = 1/\gamma_1$.

The more precise asymptotics of the partition function is as follows.

Theorem 6. *Let $J_1, J_2, \ldots, J_m \in \mathbb{N}$, be the sequence of indices defined by (135) and $l(\beta)$ defined by (140). Then, with the notations introduced above,*

$$e^{-\beta \sum_{j=1}^{l(\beta)} \left(N\sqrt{2\bar{a}_j \ln \bar{\alpha}_j} - \gamma_j [\ln(N \ln \bar{\alpha}_j) + \ln 4\pi]/2\right) - N\beta^2 \sum_{i=J_{l(\beta)}+1}^{n} a_i/2} Z_{\beta,N}$$
$$\xrightarrow{\mathcal{D}} C(\beta) \int_{\mathbb{R}^{l(\beta)}} e^{\beta \sum_{i=1}^{l(\beta)} \gamma_i x_i} \mathcal{P}^{(l(\beta))}(dx_1 \ldots dx_{l(\beta)}) \tag{142}$$

This integral is over the process $\mathcal{P}^{(l(\beta))}$ on $\mathbb{R}^{l(\beta)}$ from Theorem 2 with constants K_j from Theorem 3. The constant $C(\beta)$ satisfies

$$C(\beta) = 1, \text{ if } \beta\gamma_{l(\beta)+1} < 1, \tag{143}$$

and $0 < C(\beta) < 1$, otherwise.

Remark 5. An explicit formula for $C(\beta)$ is given in [6].

The integrals over the Poisson cascades appearing in Theorem 6 are to be understood as

$$\int_{\mathbb{R}^m} e^{\beta\gamma_1 x_1 + \cdots + \beta\gamma_m x_m} \mathcal{P}^{(m)}(dx_1 \ldots dx_m) \tag{144}$$
$$\equiv \lim_{x \downarrow -\infty} \int_{\substack{(x_1,\ldots,x_m) \in \mathbb{R}^m, \\ \exists i, 1 \leq i \leq m: \gamma_1 x_1 + \cdots + \gamma_i x_i > (\gamma_1 + \cdots + \gamma_i)x}} e^{\beta\gamma_1 x_1 + \cdots + \beta\gamma_m x_m} \mathcal{P}^{(m)}(dx_1 \ldots dx_m)$$

The existence of these limits requires the conditions on the γ_i mentioned before, and thus can be seen as responsible for the selection of the partition J and the cut-off level $l(\beta)$. Namely [6]:

Proposition 2. *Assume that $\gamma_1 > \gamma_2 > \ldots > \gamma_m > 0$, and $\beta\gamma_m > 1$. Then*

(i) For any $a \in \mathbf{R}$ the process $\mathcal{P}^{(m)}$ contains a.s. a finite number of points $(x_1, \ldots, x_m)$ such that $\gamma_1 x_1 + \cdots + \gamma_m x_m > a$.
(ii) The limit in (144) exists and is finite a.s.

4.4 The asymptotic model

As in the REM, we are able to reinterpret the convergence of the partition function in terms of an asymptotic statistical mechanics model.

This time, it has the following ingredients:

- State space: $\mathbb{N}^\ell$, that should be though of as an ℓ-level tree with infinite branching number;
- A sequence, $\boldsymbol{\gamma} \equiv (\gamma_1 > \gamma_2 > \cdots > \gamma_\ell)$ of numbers;

- Random Hamiltonian: $\mathcal{H}^{\ell}_{\boldsymbol{\gamma}} : \mathbb{N}^{\ell} \to \mathbb{R}$, where

$$\mathcal{H}^{\ell}_{\boldsymbol{\gamma}} = \sum_{k=1}^{\ell} \gamma_k x_{\boldsymbol{i}_k}, \tag{145}$$

  and $x_{\boldsymbol{i}_k}$ is the $\boldsymbol{i}_k$-th atom of the Poisson process $\mathcal{P}^{(k)}$;
- Temperature: $1/\beta$;
- Partition function: $\mathcal{Z}_\beta = \sum_{\boldsymbol{i}\in\mathbb{N}^\ell} e^{\beta\mathcal{H}(\boldsymbol{i})}$;
- Gibbs measure:

$$\hat{\mu}^{\ell}_{\beta,\boldsymbol{\gamma}}(\boldsymbol{i}) = \mathcal{Z}^{\ell}_{\beta} e^{\beta\mathcal{H}^{\ell}_{\boldsymbol{\gamma}}(\boldsymbol{i})}. \tag{146}$$

Our convergence results so far can be interpreted in terms of this model as follows:

- The partition function of the GREM converges, after multiplication with the correct scaling factor, to $\mathcal{Z}_\beta$;

A new feature compared to the situation in the REM is that that state space $\mathbb{N}^\ell$ of the asymptotic model carries a natural non-trivial distance, namely the hierarchical distance, respectively the corresponding hierarchical overlap

$$d(\boldsymbol{i},\boldsymbol{j}) = \frac{1}{\ell}(\min(k : i_k \neq j_k) - 1). \tag{147}$$

This allows to define in the asymptotic model the analog of the local mass distribution (see (106) as

$$\phi_\beta(\boldsymbol{i},t) \equiv \hat{\mu}_\beta \left(d(\boldsymbol{i},\boldsymbol{j}) > t\right). \tag{148}$$

This allows also to write the empirical distance distribution function for the asymptotic model in in the form

$$\mathcal{K}_\beta \equiv \sum_{\boldsymbol{i}\in\mathbb{N}^\ell} \hat{\mu}_\beta(\boldsymbol{i})\delta_{\phi_\beta(\boldsymbol{i},\cdot)} \tag{149}$$

Clearly we expect K_β to be related to the analogous object in the GREM, i.e. to $\mathcal{K}_{\beta,N}$ defined as in (114). The only additional ingredient is a translation between the overlap on the hypercube and the tree-overlap (148). This is in fact given by the following Lemma:

Lemma 11. *Let*

$$q_\ell \equiv \sum_{n=1}^{\ell} \frac{\ln \bar{\alpha}_n}{\ln 2}, \tag{150}$$

and let $f(q) \equiv \sup\{k : q_k \leq q\}/\ell(\beta)$. *For any* β, *for* $q \leq q_{\max} \equiv q_{\ell(\beta)}$,

$$\lim_{N\uparrow\infty} \mu^{\otimes 2}_{\beta,N} \left(R_N(\sigma,\sigma') \leq q\right) = \hat{\mu}^{\otimes 2}_{\beta} \left(d(\boldsymbol{i},\boldsymbol{j}) \leq f(q)\right). \tag{151}$$

A nontrivial aspect of the lemma above is that the overlap defined in terms of the non-hierarchical R_N is asymptotically given in terms of the distribution of a hierarchical overlap, d. In fact, it would be quite a bit easier to show that

$$\lim_{N\uparrow\infty} \mu_{\beta,N}^{\otimes 2}\left(d_N(\sigma,\sigma') \leq q\right) = \hat{\mu}_{\beta}^{\otimes 2}\left(d(\boldsymbol{i},\boldsymbol{j}) \leq f(q)\right). \tag{152}$$

where d_N is defined in (54). The fact that the two distances are asymptotically the same on the support of the Gibbs measures is remarkable and the simplest instance of the apparent universality of ultrametric structures in spin glasses. Bolthausen and Kistler [5] (see also Jana [20]) have shown that the same occurs in a class of models where the covariance depends on several different hierarchical distances.

The main result on the limiting Gibbs measures can now be formulated as follows:

Theorem 7. *Under the assumptions and with the notation of Lemma (11)*

$$\lim_{\mathbb{N}\uparrow\infty} \mathcal{K}_{\beta,N} = \mathcal{K}_{\beta}^{f}, \tag{153}$$

where $f : [0, q_{\max}] \to [0,1]$ *is defined in Lemma 11 and*

$$\mathcal{K}_{\beta}^{f} = \sum_{\boldsymbol{i}\in\mathbb{N}^{\ell}} \hat{\mu}_{\beta}(\boldsymbol{i})\delta_{\phi_{\beta}(\boldsymbol{i},f(\cdot))}. \tag{154}$$

We see that in the aysmptotic model, we have so far three ingredients: 1) the Poisson cascade; 2) the weights γ_i; 3) the mapping f from that readjusts the tree-distance.

In fact, $\mathcal{K}_{\beta}$ as a probability distribution on distributions of distances contains a lot of *gauge invariance*. In particular, neither the measures $\hat{\mu}_{\beta}$ nor the underlying space $\mathbb{N}^{\ell}$ play a particular rôle. In fact, there is a canonical way to shift all the structure into the ultrametric and to chose as a canonical space the interval $[0,1]$ and as a canonical measure on it the Lebesgue measure. To do this, chose a one-to-one map,

$$\theta : \mathbb{N}^{\ell} \to [0,1] \tag{155}$$

such that for any Borel set, $\mathcal{A} \subset \mathcal{B}([0,1]$,

$$|\mathcal{A}| = \hat{\mu}_{\beta}(\theta^{-1}(\mathcal{A})). \tag{156}$$

Then define the overlap, γ_1, on $[0,1]$, by

$$\gamma_1(x,y) = f^{-1}\left(d\left(\theta^{-1}(x), \theta^{-1}(y)\right)\right). \tag{157}$$

Note that this overlap structure is now random, and in fact contains all the remaining randomness of the system. Then we can write $\mathcal{K}_{\beta}^{f}$ as

$$\mathcal{K}_\beta^f = \int_0^1 dx \delta_{|\{y:\gamma_1(x,y)>\cdot\}|}. \tag{158}$$

This representation allows, in fact, to put all GREMs on a single footing. Namely, one can show that the random overlaps γ_1 can all be obtained by a deterministic time change from the *genealogical distance* of a particular continuous time branching process, the so-called *Neveu CBP*. This observation goes back to an unpublished paper of Neveu [24] and was elaborated on by Bertoin and LeGall [3] and the present authors [8].

5 Gaussian comparison and applications

We now return to the study of the SK type models. We will emphasize here the role of classical comparison results for Gaussian processes. A clever use of them will allow to connect the SK models with the GREMs discussed above. We begin by recalling the basic comparison theorem.

5.1 A theorem of Slepian-Kahane

Lemma 12. *Let X and Y be two independent n-dimensional Gaussian vectors. Let D_1 and D_2 be subsets of $\{1,\dots,n\}\times\{1,\dots,n\}$. Assume that*

$$\begin{aligned} &\mathbb{E}X_iX_j \leq \mathbb{E}Y_iY_j, \quad \text{if} \quad (i,j)\in D_1 \\ &\mathbb{E}X_iX_j \geq \mathbb{E}Y_iY_j, \quad \text{if} \quad (i,j)\in D_2 \\ &\mathbb{E}X_iX_j = \mathbb{E}Y_iY_j, \quad \text{if} \quad (i,j)\not\in D_1\cup D_2 \end{aligned} \tag{159}$$

Let f be a function on $\mathbb{R}^n$, such that its second derivatives satisfy

$$\begin{aligned} &\frac{\partial^2}{\partial x_i \partial x_j} f(x) \geq 0, \quad \text{if} \quad (i,j)\in D_1 \\ &\frac{\partial^2}{\partial x_i \partial x_j} f(x) \leq 0, \quad \text{if} \quad (i,j)\in D_2 \end{aligned} \tag{160}$$

Then

$$\mathbb{E}f(X) \leq \mathbb{E}f(Y) \tag{161}$$

Proof. The first step of the proof consists of writing

$$f(X) - f(Y) = \int_0^1 dt \frac{d}{dt} f(X^t) \tag{162}$$

where we define the interpolating process

$$X^t \equiv \sqrt{t}\,X + \sqrt{1-t}\,Y \tag{163}$$

Next observe that

$$\frac{d}{dt}f(X^t) = \frac{1}{2}\sum_{i=1}^{n}\frac{\partial}{\partial x_i}f(X^t)\left(t^{-1/2}X_i - (1-t)^{-1/2}Y_i\right) \tag{164}$$

Finally, we use the generalization of the standard Gaussian integration by parts formula to the multivariate setting, namely:

Lemma 13. *Let X_i, $i \in \{1,\dots,n\}$ be a multivariate Gaussian process, and let $g : \mathbb{R}^n \to \mathbb{R}$ be a differentiable function of at most polynomial growth. Then*

$$\mathbb{E}g(X)X_i = \sum_{j=1}^{n}\mathbb{E}(X_iX_j)\mathbb{E}\frac{\partial}{\partial x_j}g(X) \tag{165}$$

Applied to the mean of the left-hand side of (164) this yields

$$\mathbb{E}f(X) - \mathbb{E}f(Y) = \frac{1}{2}\sum_{i,j}\int_{0,1} dt\,(\mathbb{E}X_iX_j - \mathbb{E}Y_iY_j)\,\mathbb{E}\frac{\partial^2}{\partial x_j\partial x_i}f(X^t) \tag{166}$$

from which the assertion of the theorem can be read off.

Note that Equation (166) has the flavor of the fundamental theorem of calculus on the space of Gaussian processes.

5.2 The thermodynamic limit through comparison

Theorem 8. *[17] Assume that X_σ is a normalized Gaussian process on $\mathcal{S}_N$ with covariance*

$$\mathbb{E}X_\sigma X_\tau = \xi(R_N(\sigma,\tau)) \tag{167}$$

where $\xi : [-1,1] \to [0,1]$ is convex and even. Then

$$\lim_{N\uparrow\infty}\frac{-1}{\beta N}\mathbb{E}\ln\mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma} \equiv f_\beta \tag{168}$$

exists.

Proof. The proof of this fact is frightfully easy, once you think about using Theorem 12. Choose any $1 < M < N$. Let $\sigma = (\hat\sigma,\check\sigma)$ where $\hat\sigma = (\sigma_1,\sigma_2,\dots,\sigma_M)$, and $\check\sigma = (\sigma_{M+1},\dots,\sigma_N)$. Define independent Gaussian processes $\hat X$ and $\check X$ on $\mathcal{S}_M$ and $\mathcal{S}_{N-M}$, respectively, such that

$$\mathbb{E}\hat X_{\hat\sigma}\hat X_{\hat\tau} = \xi(R_M(\hat\sigma,\hat\tau)) \tag{169}$$

and

$$\mathbb{E}\check X_{\check\sigma}\check X_{\check\tau} = \xi(R_{N-M}(\check\sigma,\check\tau) \tag{170}$$

Set

$$Y_\sigma \equiv \sqrt{\tfrac{M}{N}}\hat{X}_{\hat\sigma} + \sqrt{\tfrac{N-M}{N}}\check{X}_{\check\sigma} \tag{171}$$

Clearly,

$$\begin{aligned} \mathbb{E}Y_\sigma Y_\tau &= \tfrac{M}{N}\xi(R_M(\hat\sigma,\hat\tau)) + \tfrac{N-M}{N}\xi(R_{N-M}(\check\sigma,\check\tau)) \\ &\geq \xi\left(\tfrac{M}{N}R_M(\hat\sigma,\hat\tau) + \tfrac{N-M}{N}(R_{N-M}(\check\sigma,\check\tau))\right) = \xi(R_N(\sigma,\tau)) \end{aligned} \tag{172}$$

Define real-valued functions $F_N(x) \equiv \ln \mathbb{E}_\sigma e^{\beta\sqrt{N}x_\sigma}$ on $\mathbb{R}^{2^N}$. It is straightforward that

$$\mathbb{E}F_N(Y) = \mathbb{E}F_M(X) + \mathbb{E}F_{N-M}(X) \tag{173}$$

A simple computation shows that, for $\sigma \neq \tau$,

$$\frac{\partial^2}{\partial x_\sigma \partial x_\tau} F_N(x) = -\frac{2^{-2N}\beta^2 N e^{\beta\sqrt{N}(x_\sigma + x_\tau)}}{Z^2_{\beta,N}} \leq 0 \tag{174}$$

Thus, Theorem 12 tells us that

$$\mathbb{E}F_N(X) \geq \mathbb{E}F_N(Y) = \mathbb{E}F_M(X) + \mathbb{E}F_{N-M}(X) \tag{175}$$

This implies that the sequence $-\mathbb{E}F_N(X)$ is subadditive, and this in turn implies (see Section 1.2) that the free energy exists , provided it is bounded, which is easy to verify (see e.g. the discussion on the correct normalization in the SK model).

The same ideas can be used for other types of Gaussian processes, e.g. the GREM-type models discussed above [10].

Convergence of the free energy in mean implies readily almost sure convergence. This follows from a general *concentration of measure* principle for functions of Gaussian random variables.

5.3 An extended comparison principle.

As I have mentioned, comparison of the free energy of SK models to simpler models do not immediately seem to work. The idea is to use comparison on a much richer class of processes. Basically, rather than comparing one process to another, we construct an extended process on a product space and use comparison on this richer space. Let us first explain this in an abstract setting. We have a process X on a space $\mathcal{S}$ equipped with a probability measure $\mathbb{E}_\sigma$. We want to compute as usual the average of the logarithm of the partition function $F(X) = \ln \mathbb{E}_\sigma e^{\beta X_\sigma}$. Now consider a second space $\mathcal{T}$ equipped with a probability law $\mathbb{E}_\alpha$. Choose a Gaussian process, Y, independent of X, on this space, and define a further independent process, Z, on the product space $\mathcal{S}\times\mathcal{T}$. Define real valued functions, G, H, on the space of real valued functions on $\mathcal{T}$ and $\mathcal{S}\times\mathcal{T}$, respectively, via $G(y) \equiv \ln \mathbb{E}_\alpha e^{\beta y_\alpha}$ and $H(z) = \ln \mathbb{E}_\sigma \mathbb{E}_\alpha e^{\beta z_{\sigma,\alpha}}$. Note that $H(X+Y) = F(X)+G(Y)$. Assume that the covariances are chosen such that

$$\mathrm{cov}(X_\sigma, X_{\sigma'}) + \mathrm{cov}(Y_\alpha, Y_{\alpha'}) \geq \mathrm{cov}(Z_{\sigma,\alpha}, Z_{\sigma',\alpha'}) \tag{176}$$

Since we know that the second derivatives of H are negative, we get from Theorem 12 that

$$\mathbb{E}F(X) + \mathbb{E}G(Y) = \mathbb{E}H(X+Y) \leq \mathbb{E}H(Z) \tag{177}$$

This is a useful relation if we know how to compute $\mathbb{E}G(Y)$ and $\mathbb{E}H(Z)$. This idea may look a bit crazy at first sight, but we must remember that we have a lot of freedom in choosing the auxiliary spaces and processes to our convenience. Before turning to the issue whether we can find useful computable processes Y and Z, let us see why we could hope to find in this way *sharp* bounds.

5.4 The extended variational principle and thermodynamic equilibrium

To do so, we will show that, in principle, we can represent the free energy in the thermodynamic limit in the form $\mathbb{E}H(Z) - \mathbb{E}G(Y)$. To this end let $\mathcal{S} = \mathcal{S}_M$ and $\mathcal{T} = \mathcal{S}_N$, both equipped with their natural probability measure $\mathbb{E}_\sigma$. We will think of $N \gg M$, and both tending to infinity eventually. We write again $\mathcal{S} \times \mathcal{T} \ni \sigma = (\hat\sigma, \check\sigma)$. Consider the process X_σ on $\mathcal{S}_{N+M}$ with covariance $\xi(R_{N+M}(\sigma, \sigma'))$. We would like to write this as

$$X_\sigma = \hat X_{\hat\sigma} + \check X_{\check\sigma} + Z_\sigma \tag{178}$$

where all three processes are independent. Note that here and in the sequel equalities between random variables are understood to hold in distribution. Moreover, we demand that

$$\mathrm{cov}(\hat X_{\hat\sigma}, \hat X_{\hat\sigma'}) = \xi(\tfrac{M}{N+M} R_M(\hat\sigma, \hat\sigma')) \tag{179}$$

and

$$\mathrm{cov}(\check X_{\check\sigma}, \check X_{\check\sigma'}) = \xi(\tfrac{N}{N+M} R_N(\check\sigma, \check\sigma')) \tag{180}$$

Obviously, this implies that

$$\begin{aligned} \mathrm{cov}(Z_\sigma, Z_{\sigma'}) = {} & \xi\left(\tfrac{M}{N+M} R_M(\hat\sigma, \hat\sigma') + \tfrac{N}{N+M} R_N(\check\sigma, \check\sigma')\right) \\ & - \xi\left(\tfrac{M}{N+M} R_M(\hat\sigma, \hat\sigma')\right) - \xi\left(\tfrac{N}{N+M} R_N(\check\sigma, \check\sigma')\right) \end{aligned} \tag{181}$$

(we will not worry about the existence of such a decomposition; if $\xi(x) = x^p$, we can use the explicit representation in terms of p-spin interactions to construct them). Now we first note that, by super-additivity [2]

$$\lim_{M\uparrow\infty} \frac{1}{\beta M} \liminf_{N\uparrow\infty} \mathbb{E} \log \frac{Z_{\beta,N+M}}{Z_{\beta,N}} = -f_\beta \tag{182}$$

Thus we need a suitable representation for $\frac{Z_{\beta,N+M}}{Z_{\beta,N}}$. But

$$\frac{Z_{\beta,N+M}}{Z_{\beta,N}} = \frac{\mathbb{E}_\sigma e^{\beta\sqrt{N+M}\left(\check{X}_{\check{\sigma}}+Z_\sigma+\hat{X}_{\hat{\sigma}}\right)}}{\mathbb{E}_{\check{\sigma}} e^{\beta\sqrt{N+M}\left(\sqrt{(1-M/(N+M))}X_{\check{\sigma}}\right)}} \tag{183}$$

Now we want to express the random variables in the denominator in the form

$$\sqrt{(1-M/(N+M))}X_{\check{\sigma}} = \check{X}_{\check{\sigma}} + Y_{\check{\sigma}} \tag{184}$$

where Y is independent of $\check{X}$. Comparing covariances, this implies that

$$\begin{aligned}\mathrm{cov}(Y_{\check{\sigma}}, Y_{\check{\sigma}'}) = &(1-M/(N+M))\xi(R_N(\check{\sigma},\check{\sigma}')) \\ &- \xi\left(\tfrac{N}{N+M}R_N(\check{\sigma},\check{\sigma}')\right)\end{aligned} \tag{185}$$

As we will be interested in taking the limit $N\uparrow\infty$ before $M\uparrow\infty$, we may expand in $M/(N+M)$ to see that to leading order in $M/(N+M)$,

$$\begin{aligned}\mathrm{cov}(Y_{\check{\sigma}}, Y_{\check{\sigma}'}) \sim &\tfrac{M}{N+M}R_N(\check{\sigma},\check{\sigma}')\xi'\left(\tfrac{N}{N+M}R_N(\check{\sigma},\check{\sigma}')\right) \\ &- \tfrac{M}{N+M}\xi\left(\tfrac{N}{N+M}R_N(\check{\sigma},\check{\sigma}')\right)\end{aligned} \tag{186}$$

Finally, we note that the random variables $\hat{X}_{\hat{\sigma}}$ are negligible in the limit $N\uparrow\infty$, since their variance is smaller than $\xi(M/(N+M))$ and hence their maximum is bounded by $\sqrt{\xi(M/(N+M))M\ln 2}$, which even after multiplication with $\sqrt{N+M}$ gives no contribution in the limit if ξ tends to zero faster than linearly at zero, which we can safely assume. Thus we see that we can indeed express the free energy as

$$f_\beta = -\lim_{M\uparrow\infty}\liminf_{N\uparrow\infty}\frac{1}{\beta M}\mathbb{E}\ln\frac{\mathbb{E}_{\hat{\sigma}}\widetilde{\mathbb{E}}_{\check{\sigma}}e^{\beta\sqrt{N+M}Z_{\hat{\sigma},\check{\sigma}}}}{\widetilde{\mathbb{E}}_{\check{\sigma}}e^{\beta\sqrt{N+M}Y_{\check{\sigma}}}} \tag{187}$$

where the measure $\widetilde{\mathbb{E}}_{\check{\sigma}}$ can be chosen as a probability measure defined by $\widetilde{\mathbb{E}}_{\check{\sigma}}(\cdot) = \mathbb{E}_{\check{\sigma}}e^{\beta\sqrt{N+M}\check{X}_{\check{\sigma}}}(\cdot)/\check{Z}_{\beta,N,M}$ where $\check{Z}_{\beta,N,M}\equiv\mathbb{E}_{\check{\sigma}}e^{\beta\sqrt{N+M}\check{X}_{\check{\sigma}}}$. Of course this representation is quite pointless, because it is certainly uncomputable, since $\widetilde{\mathbb{E}}$ is effectively the limiting Gibbs measure that we are looking for. However, at this point there occurs a certain miracle: the (asymptotic) covariances of the processes X, Y, Z satisfy

$$\xi(x) + y\xi'(y) - \xi(y) \geq x\xi'(y) \tag{188}$$

for all $x, y\in[-1,1]$, if ξ is convex and even. This comes as a surprise, since we did not do anything to impose such a relation! But it has the remarkable consequence that asymptotically, by virtue of Lemma 12 it implies the bound

$$\mathbb{E}\ln\mathbb{E}_{\hat{\sigma}}e^{\beta\sqrt{M}X_{\hat{\sigma}}} \leq \mathbb{E}\ln\frac{\mathbb{E}_{\hat{\sigma}}\widetilde{\mathbb{E}}_{\check{\sigma}}e^{\beta\sqrt{N+M}Z_{\hat{\sigma},\check{\sigma}}}}{\widetilde{\mathbb{E}}_{\check{\sigma}}e^{\beta\sqrt{N+M}Y_{\check{\sigma}}}} \tag{189}$$

(if the processes are taken to have the asymptotic form of the covariances). Moreover, this bound will hold *even* if we replace the measure $\widetilde{\mathbb{E}}$ by some other probability measure, and even if we replace the overlap R_N on the space $\mathcal{S}_N$ by some other function, e.g. the ultrametric d_N. Seen the other way around, we can conclude that a lower bound of the form (177) can actually be made as good as we want, provided we choose the right measure $\widetilde{\mathbb{E}}$. This observation is due to Aizenman, Sims, and Starr [2]. They call the auxiliary structure made from a space $\mathcal{T}$, a probability measure $\mathbb{E}_\alpha$ on $\mathcal{T}$, a normalized distance q on $\mathcal{T}$, and the corresponding processes, Y and Z, a *random overlap structure*

$$\operatorname{cov}(Y_\alpha, Y_{\alpha'}) = q(\alpha,\alpha')\xi'(q(\alpha,\alpha')) - \xi(q(\alpha,\alpha')) \tag{190}$$

and the process $Z_{\sigma,\alpha}$ on $\mathcal{S}_N \times [0,1]$ with covariance

$$\operatorname{cov}(Z_{\sigma,\alpha}, Z_{\sigma',\alpha'}) \equiv R_N(\sigma,\sigma')\xi'(q(\alpha,\alpha')) \tag{191}$$

With these choices, and naturally X_σ our original process with covariance $\xi(R_N)$, the equation (176) is satisfied, and hence the inequality (177) holds, no matter what choice of q and $\mathbb{E}_\alpha$ we make. Restricting these choices to the random genealogies obtained from Neveu's process by a time change with some probability distribution function m, and $\mathbb{E}_\alpha$ the Lebesgue measure on $[0,1]$, gives the bound we want.

This bound would be quite useless if we could not compute the right-hand side. Fortunately, one can get rather explicit expressions. We need to compute two objects:

$$\mathbb{E}_\alpha \mathbb{E}_\sigma e^{\beta\sqrt{N} Z_{\sigma,\alpha}} \tag{192}$$

and

$$\mathbb{E}_\alpha e^{\beta\sqrt{N} Y_\alpha} \tag{193}$$

In the former we use that Z has the representation

$$Z_{\sigma,\alpha} = N^{-1/2} \sum_{i=1}^N \sigma_i z_{\alpha,i} \tag{194}$$

where the processes $z_{\alpha,i}$ are independent for different i and have covariance

$$\operatorname{cov}(z_{\alpha,i}, z_{\alpha',i}) = \xi'(q(\alpha,\alpha')) \tag{195}$$

Thus at least the σ- average is trivial:

$$\mathbb{E}_\alpha \mathbb{E}_\sigma e^{\beta\sqrt{N} Z_{\sigma,\alpha}} = \mathbb{E}_\alpha \prod_{i=1}^N e^{\ln\cosh(\beta z_{\alpha,i})} \tag{196}$$

Thus we see that, in any case, we obtain bounds that only involve objects that we introduced ourselves and that thus can be manipulated to be computable. In fact, such computations have been done in the context of the Parisi solution [23]. A useful mathematical reference is [4].

This is the form derived in Aizenman, Sims, and Starr [2].

5.5 Parisi auxiliary systems

The key idea of the Parisi solution is to chose as an auxiliary system as the asymptotic model of the GREM.

That is, the space $\mathcal{T}$ in this case is chosen as a n-level infinite tree $\mathbb{N}^n$ equipped with the measure $\hat{\mu}_\beta$ defined in (146).

$\mathcal{T}$ is naturally endowed with its tree overlap, $d(\boldsymbol{i},\boldsymbol{j}) \equiv n^{-1}(\min\{\ell : i_\ell \neq j_\ell\} - 1)$. This distance will play the rôle of the distance q on $\mathcal{T}$. Finally, we define the processes $Y_{\boldsymbol{i}}$ and $Z_{\boldsymbol{i},\sigma}$ with covariances

$$\operatorname{cov}(Y_{\boldsymbol{i}}, Y_{\boldsymbol{j}}) = d(\boldsymbol{i},\boldsymbol{j})\xi'(d(\boldsymbol{i},\boldsymbol{j})) - \xi(d(\boldsymbol{i},\boldsymbol{j})) \equiv h(d(\boldsymbol{i},\boldsymbol{j})) \tag{197}$$

and the process $Z_{\sigma,\boldsymbol{i}}$ on $\mathcal{S}_N \times \mathcal{T}$ with covariance

$$\operatorname{cov}(Z_{\sigma,\boldsymbol{j}}, Z_{\sigma',\boldsymbol{j}}) \equiv R_N(\sigma,\sigma')\xi'(d(\boldsymbol{i},\boldsymbol{j})) \tag{198}$$

It is easy to see that such processes can be constructed as long as h, ξ' are increasing functions. E.g.

$$Y_{\boldsymbol{i}} = \sum_{\ell=1}^{n} \sqrt{h(\ell/n) - h((\ell-1)/n)}\, Y^{(\ell)}_{\boldsymbol{i}_1\dots\boldsymbol{i}_\ell} \tag{199}$$

where $Y^{(\ell)}_{\boldsymbol{i}_1\dots\boldsymbol{i}_\ell}$ are independent standard normal random variables. In this way, we have constructed an explicit random overlap structure, which corresponds indeed to the one generating the Parisi solution.

Note that also the auxiliary structure depends only on the information contained in the empirical distance distribution, $\mathcal{K}_\beta$, associated with the asymptotic model. In fact we could alternatively use $\mathcal{T} = [01]$ equipped with the Lebesgue measure and the random overlap γ_1, as defined in (158). While this is nice conceptually, for actual computations the form above will be more useful, however.

5.6 Computing with Poisson cascades

Lemma 14. *Assume that $\mathcal{P}$ be a Poisson process with intensity measure $e^{-x}dx$. and let $Z_{i,j}$, $i \in \mathbb{N}$, $j \in \mathbb{N}$ and Y be iid standard normal random variables. Let $g_j : \mathbb{R} \to \mathbb{R}$, $j = 1,\dots,M$, be smooth functions, such that, for all $|m| \leq 2$, there exist $C < \infty$, independent of N, such that*

$$\mathbb{E}_Y e^{m g_i(Y)} \equiv e^{L_i(m)} < C \tag{200}$$

Let x_i be the atoms of the Poisson process $\mathcal{P}$ with intensity measure $e^{-x}dx$. Then

$$\mathbb{E}\ln \frac{\sum_{i=1}^\infty e^{\alpha x_i + \sum_{j=1}^M g_j(Z_{i,j})}}{\sum_{i=1}^\infty e^{\alpha x_i}} = \sum_{j=1}^M \alpha L_j(1/\alpha). \tag{201}$$

Proof. Let for simplicity $M = 1$. The numerator on the left in (201) can be written as

$$\int e^{\alpha z}\widetilde{\mathcal{P}}(dz)$$

where $\widetilde{\mathcal{P}}$ is the point process

$$\widetilde{\mathcal{P}} \equiv \sum_j \delta_{z_j+\alpha^{-1}g(Y_j)}$$

This follows from a general fact about Poisson point processes: if $\mathcal{N} \equiv \sum_i \delta_{x_i}$ is a Poisson point process with intensity measure λ on E, and Y_i are iid random variables with distribution ρ, then

$$\widetilde{\mathcal{N}} \equiv \sum_i \delta_{x_i+Y_i}$$

is a Poisson process with intensity measure $\lambda * \rho$ on the set $E + \text{supp}\rho$. This follows from the representation of $\mathcal{N}$ as

$$\mathcal{N} = \sum_{i=1}^{N_\lambda} \delta_{X_i}$$

where N_λ is Poisson with parameter $\int_E \lambda(dx) \equiv |\lambda|$ (if this is finite), and X_i iid random variables with distribution $\lambda/|\lambda|$. Clearly

$$\widetilde{\mathcal{N}} = \sum_i \delta_{x_i+Y_i} = \sum_{i=1}^{N_\lambda} \delta_{X_i+Y_i}$$

is again of the form of a PPP, and the distribution of $X_i + Y_i$ is $\lambda * \rho/|\lambda|$. Since the total intensity of the process is the parameter of N_λ, $|\lambda|$, it follows that the intensity measure of this process is the one we claimed.

Thus, in our case, $\widetilde{\mathcal{P}}$ is a PPP whose intensity measure is the convolution of the measure $e^{-z}dz$ and the distribution of the random variable $\alpha^{-1}g(Y)$. A simple computation shows that this is $\mathbb{E}_Y e^{g(Y)/\alpha}e^{-z}dz$, i.e. a multiple of the original intensity measure!

Finally, one makes the elementary but surprising and remarkable observation that the Poisson point process $\sum_j \delta_{z+\ln \mathbb{E}_Y e^{g(Y)/\alpha}}$ has the same intensity measure, and therefore, $\sum_j e^{\alpha z_j+g(Y_j)}$ has the same law as $\sum_j e^{\alpha z_j}[\mathbb{E}_Y e^{g(Y)/\alpha}]^\alpha$: multiplying each atom with an iid random variable leads to the same process as multiplying each atom by a suitable constant! The assertion of the Lemma follows immediately.

Remark 6. Nobody seems to know who made this discovery. Michael Aizenman told me about it and attributed to David Ruelle, but one cannot find it in his paper. A slightly different proof from the one above can be found in [26].

Let us look first at (193). We can then write

$$\sum_{\boldsymbol{i}} e^{\mathcal{H}^n_{\boldsymbol{\gamma}}(\boldsymbol{i}+\beta\sqrt{M}Y_{\boldsymbol{i}}} = \sum_{\boldsymbol{i}} e^{\mathcal{H}^n_{\boldsymbol{\gamma}}(\boldsymbol{i})+\beta\sqrt{M}Y_{\boldsymbol{i}_{n-1}}+\sqrt{h(x_n)-h(x_{n-1})}Y^{(n)}_{\boldsymbol{i}}}$$
$$= \sum_{\boldsymbol{i}_1\dots\boldsymbol{i}_{n-1}} e^{\sum_{\ell=1}^{n-1}\gamma_\ell x_{\boldsymbol{i}_1\dots\bar{\sigma}_\ell}+\beta\sqrt{M}Y_{\boldsymbol{i}_{n-1}}}$$
$$\times\sum_{\boldsymbol{i}_n} e^{\gamma_n x_{\boldsymbol{i}_{n-1},i_n}+\beta\sqrt{M}\sqrt{h(1)-h(1-1/n)}Y^{(n)}_{\boldsymbol{i}_{n-1},i_n}} \tag{202}$$

Using Lemma 14, the last factor can be replaced by

$$\mathbb{E}_{i_n} e^{\gamma_n x_{\boldsymbol{i}_{n-1},i_n}+\beta\sqrt{M}\sqrt{h(1)-h(1-1/n)}Y^{(n)}_{\boldsymbol{i}_{n-1},i_n}} \tag{203}$$
$$\to \left[\int \frac{dz}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}e^{zm_n\beta\sqrt{M}\sqrt{h(1)-h(1-1/n)}}\right]^{1/m_n}\sum_{i_n} e^{\gamma_n x_{\boldsymbol{i}}} \tag{204}$$
$$= e^{\frac{\beta^2 M}{2}m_n(h(1)-h(1-1/n))}\sum_{i_n} e^{\gamma_n x_{\boldsymbol{i}}} \tag{205}$$

(we use throughout $m_n = 1/\gamma_n$). Note that the last factor is independent of the random variables $x_{i_1,\dots,i_\ell}$ with $\ell < n$. Thus

$$\mathbb{E}\ln\sum_{\boldsymbol{i}} e^{\alpha x_{\boldsymbol{i}}+\beta\sqrt{M}Y_{\boldsymbol{i}}} = \mathbb{E}\ln\sum_{i_1,\dots,i_{n-1}} e^{\sum_{\ell=1}^{n-1}\gamma_\ell x_{i_1,\dots,i_{n-1}}+\beta\sqrt{M}Y_{i_1,\dots,i_{n-1}}}$$
$$+\frac{\beta^2 M}{2}m_n(h(1)-h(1-1/n)) + \mathbb{E}\ln\sum_{i_n} e^{\gamma_n x_{\boldsymbol{i}}} \tag{206}$$

The first term now has the same form as the original one with n replaced by $n-1$, and thus the procedure can obviously iterated. As the final result, we get that a consequence, we get that

$$M^{-1}\mathbb{E}\ln\frac{\sum_{\boldsymbol{i}} e^{x_{\boldsymbol{i}}+\beta\sqrt{M}Y_{\boldsymbol{i}}}}{\sum_{\boldsymbol{i}} e^{x_{\boldsymbol{i}}}}$$
$$= \sum_{\ell=1}^{n}\frac{\beta^2}{2}m_\ell(h(1-\ell/n)-h(1-(\ell-1)/n))$$
$$= \frac{\beta^2}{2}\int_0^1 m(x)x\xi''(x)dx \tag{207}$$

The computation of the expression (192) is now very similar, but gives a more complicated result since the analogs of the expressions (203) cannot be computed explicitly. Thus, after the k-th step, we end up with a new function of the remaining random variables $Y_{i_1\dots i_{n-k}}$. The result can be expressed in the form

$$\frac{1}{M}\mathbb{E}\ln\mathbb{E}_{\boldsymbol{i}}\mathbb{E}_{\sigma}e^{\beta\sqrt{M}Z_{\sigma,\boldsymbol{i}}} = \zeta(0,h,m,\beta) \tag{208}$$

(here h is the magnetic field (which we have so far hidden in the notation) that can be taken as a parameter of the a priori distribution on the σ such that $\mathbb{E}_{\sigma_i}(\cdot) \equiv \frac{1}{2\cosh(\beta h)}\sum_{\sigma_i=\pm 1} e^{\beta h\sigma_i}(\cdot))$ where $\zeta(1,h) = \ln\cosh(\beta h)$, and

$$\zeta(x_{a-1},h) = \frac{1}{m_a}\ln\int\frac{dz}{\sqrt{2\pi}}e^{-z^2/2}e^{m_a\zeta(x_a,h+z\sqrt{\xi'(x_a)-\xi'(x)})} \tag{209}$$

(we put $x_a = a/n$).

In all of the preceeding discussion, the choice of the parameter n and of the numbers $m_i = 1/\gamma_1$ is still free. From

We can now announce Guerra's bound in the following form:

Theorem 9. *[16] Let $\zeta(t,h,m,b)$ be the function defined in terms of the recursion (209). Then*

$$\lim_{N\uparrow\infty} N^{-1}\mathbb{E}\ln Z_{\beta,h,N} \leq \inf_m \zeta(0,h,m,\beta) - \frac{\beta^2}{2}\int_0^1 m(x)x\xi''(x)dx \tag{210}$$

where the infimum is over all probability distribution functions m on the unit interval.

Remark 7. It is also interesting to see that the recursive form of the function ζ above can also be represented in a closed form as the solution of a partial differential equation. Consider the case $\xi(x) = x^2/2$. Then ζ is the solution of the differential equation

$$\frac{\partial}{\partial t}\zeta(t,h) + \frac{1}{2}\left(\frac{\partial^2}{\partial h^2}\zeta(t,h) + m(t)\left(\frac{\partial}{\partial h}\zeta(t,h)\right)^2\right) = 0 \tag{211}$$

with final condition

$$\zeta(1,h) = \ln\cosh(\beta h) \tag{212}$$

If m is a step function, it is easy to see that a solution is obtained by setting, for $x \in [x_{a-1}, x_a)$,

$$\zeta(x,h) = \frac{1}{m_a}\ln\mathbb{E}_z e^{m_a\zeta(x_a,h+z\sqrt{x_a-x})} \tag{213}$$

For general convex ξ, analogous expressions can be obtained through changes of variables [16].

5.7 Talagrand's theorem

In both approaches, it pays to write down the expression of the difference between the free energy and the lower bound, since this takes a very suggestive form.

To do this, we just have to use formula (166) with

$$X^t_{\sigma,\alpha} \equiv \sqrt{t}(X_\sigma + Y_\alpha) + \sqrt{1-t} Z_{\sigma,\alpha} \tag{214}$$

and $f(X^t)$ replaced by $H(X^t) = \ln \mathbb{E}_\sigma \mathbb{E}_\alpha e^{\beta\sqrt{N} Z^t_{\sigma,\alpha}}$. This gives the equality

$$\begin{aligned} H(X+Y) - H(Z) = \frac{1}{2}\mathbb{E}\int_0^1 dt \widetilde{\mu}^{\otimes 2}_{\beta,t,N}(d\sigma, d\alpha)\Big(&\xi(R_N(\sigma,\sigma')) \\ &+ q(\alpha,\alpha')\xi'(q(\alpha,\alpha')) \\ &- \xi(q(\alpha,\alpha')) - R_N(\sigma,\sigma')\xi'(q(\alpha,\alpha'))\Big) \end{aligned} \tag{215}$$

where the measure $\widetilde{\mu}_{\beta,t,N}$ is defined as

$$\widetilde{\mu}_{\beta,t,N}(\cdot) \equiv \frac{\mathbb{E}_\sigma \mathbb{E}_\alpha e^{\beta\sqrt{N} X^t_{\sigma,\alpha}}(\cdot)}{\mathbb{E}_\sigma \mathbb{E}_\alpha e^{\beta\sqrt{N} X^t_{\sigma,\alpha}}} \tag{216}$$

where we interpret the measure $\widetilde{\mu}_{\beta,t,N}$ as a joint distribution on $\mathcal{S}_N \times [0,1]$. Note that for convex and even ξ, the function $\xi(R_N(\sigma,\sigma'))+q(\alpha,\alpha')\xi'(q(\alpha,\alpha'))-\xi(q(\alpha,\alpha'))$ vanishes if and only if $R_N(\sigma,\sigma') = q(\alpha,\alpha')$. Thus for the left hand side of (215) to vanish, the replicated interpolating product measure should (for almost all t), concentrate on configurations where the overlaps in the σ-variables coincide with the genealogical distances of the α-variables. Thus we see that the inequality in Theorem 9 will turn into an equality if it is possible to choose the parameters of the reservoir system in such a way that the the overlap distribution on $\mathcal{S}_N$ aligns with the genealogical distance distribution in the reservoir once the systems are coupled by the interpolation.

This latter fact was proven very recently, and not long after the discovery of Guerra's bound, by M. Talagrand [29].

Theorem 10. *[29] Let $\zeta(t,h,m,b)$ be the function defined in terms of (211) and (212). Then*

$$\lim_{N\uparrow\infty} N^{-1}\mathbb{E}\ln Z_{\beta,h,N} = \inf_m \left(\zeta(0,h,m,.\beta) - \frac{\beta^2}{2}\int_0^1 m(x)x\xi''(x)dx\right) \tag{217}$$

where the infimum is over all probability distribution functions m on the unit interval.

I will not give the complex proof which the interested reader should study in the original paper [29], but I will make some comments on the key ideas. First, Talagrand proves more than the assertion 217. What he actually proves is the following. For any $\epsilon > 0$, there exists a positive integer $n(\epsilon) < \infty$, and a probability distribution function m_n that is a step function with n steps, such that for all $t > \epsilon$,

$$\lim_{N\uparrow\infty} \mathbb{E}\widetilde{\mu}^{\otimes 2}_{\beta,t,N}(d\sigma, d\alpha)\Big(\xi(R_N(\sigma,\sigma')) + q(\alpha,\alpha')\xi'(q(\alpha,\alpha')) \\ -\xi(q(\alpha,\alpha')) - R_N(\sigma,\sigma')\xi'(q(\alpha,\alpha'))\Big) = 0 \tag{218}$$

if the measure $\widetilde{\mu}_{b,t,N}$ corresponds to the genealogical distance obtained from this function m. That is to say, if the coupling parameter t is large enough, the SK model can be aligned to a GREM with any desired number of hierarchies.

References

1. R.J. Adler. On excursion sets, tube formulas and maxima of random fields. *Ann. Appl. Probab.*, 10(1):1–74, 2000.
2. M. Aizenman, R. Sims, and S.L. Starr. An extended variational principle for the SK spin-glass model. *Phys. Rev. B*, 6821(21):4403–4403, 2003.
3. J. Bertoin and J.-F. Le Gall. The Bolthausen-Sznitman coalescent and the genealogy of continuous-state branching processes. *Probab. Theory Related Fields*, 117(2):249–266, 2000.
4. E. Bolthausen and A.-S. Sznitman. On Ruelle's probability cascades and an abstract cavity method. *Comm. Math. Phys.*, 197(2):247–276, 1998.
5. Erwin Bolthausen and Nicola Kistler. On a nonhierarchical version of the generalized random energy model. *Ann. Appl. Probab.*, 16(1):1–14, 2006.
6. A. Bovier and I. Kurkova. Derrida's generalised random energy models. I. Models with finitely many hierarchies. *Ann. Inst. H. Poincaré Probab. Statist.*, 40(4):439–480, 2004.
7. A. Bovier and I. Kurkova. Derrida's generalized random energy models. II. Models with continuous hierarchies. *Ann. Inst. H. Poincaré Probab. Statist.*, 40(4):481–495, 2004.
8. A. Bovier and I. Kurkova. Gibbs measures of Derrida's generalized random energy models and genealogies of Neveu's continuous state branching processe. In *Spin glasses*, volume 1900 of *Lecture Notes in Math.* Springer, Berlin, 2007.
9. D. Capocaccia, M. Cassandro, and P. Picco. On the existence of thermodynamics for the generalized random energy model. *J. Statist. Phys.*, 46(3-4):493–505, 1987.
10. P. Contucci, M. Degli Esposti, C. Giardinà, and S. Graffi. Thermodynamical limit for correlated Gaussian random energy models. *Comm. Math. Phys.*, 236(1):55–63, 2003.
11. B. Derrida. A generalisation of the random energy model that includes correlations between the energies. *J. Phys. Lett.*, 46:401–407, 1985.
12. R.S. Ellis. *Entropy, large deviations, and statistical mechanics*, volume 271 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, New York, 1985.
13. E. Gardner and B. Derrida. Magnetic properties and function $q(x)$ of the generalised random energy model. *J. Phys. C*, 19:5783–5798, 1986.
14. E. Gardner and B. Derrida. Solution of the generalised random energy model. *J. Phys. C*, 19:2253–2274, 1986.
15. E. Gardner and B. Derrida. The probability distribution of the partition function of the random energy model. *J. Phys. A*, 22(12):1975–1981, 1989.

16. F. Guerra. Broken replica symmetry bounds in the mean field spin glass model. *Comm. Math. Phys.*, 233(1):1–12, 2003.
17. F. Guerra and F. L. Toninelli. The thermodynamic limit in mean field spin glass models. *Comm. Math. Phys.*, 230(1):71–79, 2002.
18. E. Ising. Beitrag zur Theorie des Ferro- und Paramagnetismus. Ph.D. thesis, Univ. Hamburg, 1924.
19. E. Ising. *Zeitschrift. f. Physik*, 31:253, 1925.
20. Nabin Kumar Jana. Contributions to random energy models. Thesis, 2008.
21. M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and processes.* Springer Series in Statistics. Springer-Verlag, New York, 1983.
22. M. Ledoux and M. Talagrand. *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin, 1991.
23. M. Mézard, G. Parisi, and M.A. Virasoro. *Spin glass theory and beyond*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific Publishing Co. Inc., Teaneck, NJ, 1987.
24. J. Neveu. A continuous state branching process in relation with the GREM model of spin glass theory. rapport interne 267, Ecole Polytechnique Paris, 1992.
25. D. Ruelle. A mathematical reformulation of Derrida's REM and GREM. *Comm. Math. Phys.*, 108(2):225–239, 1987.
26. A. Ruzmaikina and M. Aizenman. Characterization of invariant measures at the leading edge for competing particle systems. *Ann. Probab.*, 33(1):82–113, 2005.
27. D. Sherrington and S. Kirkpatrick. Solvable model of a spin glas. *Phys. Rev. Letts.*, 35:1792–1796, 1972.
28. B. Simon. *The statistical mechanics of lattice gases. Vol. I.* Princeton Series in Physics. Princeton University Press, Princeton, NJ, 1993.
29. Michel Talagrand. The Parisi formula. *Ann. of Math. (2)*, 163(1):221–263, 2006.

Index

Aizenman, M., 27

binomial coefficient
  asymptotics, 5

concentration of measure, 24
convexity, 9
Cramèr entropy, 5
Curie, P., 4
Curie–Weiss model, 4

distance
  ultrametric, 13

entropy
  Cramèr, 5

ferromagnetism, 2
free energy
  in SK model, 33
  of spin system , 2

Gaussian
  process, 13
Gibbs measure, 2
Guerra's bound, 32

integration by parts
  Gaussian
    multivariate, 23
Ising model, 1
Ising, E., 1
isotherm, 9

Kirkpatrick, S., 12

large deviation
  principle, 6

mean field models, 4
metastate, 20

non-convexity, 9

Parisi solution, 28
partition function, 2
point process
  Poisson, 17
Poisson process
  of extremes, 18
pure state, 20

random overlap structure, 27
random process
  Gaussian, 13

Sherrington, D., 12
Sherrington–Kirkpatrick (SK)-model, 12
Sims, R., 27
spin
  configuration, 1
  variable, 1
spin-glass, 12
Starr, S.L., 27
sub-additivity, 24
subadditivity, 3

Talagrand's theorem, 33
Talagrand, M., 33

Weiss, P., 4