ULTRAMETRICITY IN MEAN-FIELD SPIN GLASSES
[after Dmitry Panchenko]

by Erwin BOLTHAUSEN

1. MEAN-FIELD SPIN GLASSES

Spin glasses are Gibbs measures with random interactions. The most natural one is probably the Edwards-Anderson model, which is an Ising type model with nearest neighbor random interactions. On a finite box $V_N \stackrel{\text{def}}{=} \{1, \ldots, N\}^d$ with outer boundary ∂V_N (consisting in the points in $\mathbb{Z}^d \backslash V_N$ which have a neighbor in V_N) one defines the Hamiltonian with boundary condition $\eta \in \{-1, 1\}^{\partial V_N}$ by

$$-H_{V_N}(\sigma|\eta) \stackrel{\text{def}}{=} \sum_{i,j \in V_N, \ i \sim j} J_{ij}\sigma_i\sigma_j + \sum_{i \in V_N, j \in \partial V_N, i \sim j} J_{ij}\sigma_i\eta_j,$$

for $\sigma \in \{-1, 1\}^{V_N}$, where $i \sim j$ means that the points are neighbors on the lattice. One always takes $J_{ij} = J_{ji}$, so that the sum is over the undirected bonds of the graph V_N. In the Ising model, one would have $J_{ij} = J = \text{const} > 0$. The Hamiltonian defines a Gibbs measure on $\Sigma_{V_N} \stackrel{\text{def}}{=} \{-1, 1\}^{V_N}$ with boundary condition η, and with inverse temperature parameter $\beta > 0$, by

$$\mu_{V_N, \eta}(\sigma) \stackrel{\text{def}}{=} \frac{1}{Z_{N,\eta,\beta}} \exp\left[-\beta H_{V_N}(\sigma|\eta)\right].$$

The normalizing constant $Z_{N,\eta,\beta}$, the so-called partition function (as it is a function of β) is

$$Z_{N,\eta,\beta} \stackrel{\text{def}}{=} \sum_{\sigma \in \Sigma_{V_N}} \exp\left[-\beta H_{V_N}(\sigma|\eta)\right].$$

A natural question is about the possible limits with $N \to \infty$ and a sequence $\{\eta_N\}$ of boundary conditions. In particular, one is interested to know whether the set of measures on $\{-1, 1\}^{\mathbb{Z}^d}$ which can be obtained as limits contains a unique element, and if not how many extreme points in this set are. (The set is convex.)

For the Ising model, this is a well studied question, still with many open questions particularly for $d \geq 3$. The Edwards-Anderson model takes the J_{ij} as i.i.d. random variables, for instance centered Gaussian ones, or ± 1 coin tossings. The Hamiltonian itself is then a random variable, and the Gibbs measure is a random measure: If the J's

are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then the partition function is a function of $\omega \in \Omega$: $Z_{N,\eta,\beta,\omega}$, and the Gibbs measure also depends on the realizations ω. Typically, one is interested in so-called quenched properties, i.e. properties which hold for almost all ω (under $\mathbb{P}$). Unfortunately, for the Edwards-Anderson model, the most basic questions are mathematically completely open.

Mean-field models offer a possibility to investigate some of the basic questions, like the existence of phase transitions, in a mathematically much simpler way than for short range models like the Ising model. In mean-field models, the interactions are not local or short range, but a spin variable σ_i interacts with the others in a more global way, for instance through their means. However, the notion of an infinite Gibbs measure typically makes no longer sense, but one can still define the limiting free energy, mean magnetization, critical exponents, etc. Of course, in many respects, mean-field models are too simple to give even qualitatively the same answers as in short range models, for instance about critical exponents.

Given the mathematical difficulty in understanding short range spin glasses, it is natural to investigate mean-field type spin glasses, in the hope that they are tractable. This was the motivation of Sherrington and Kirkpatrick [23] to propose their now famous model. In the end, it turned out that a mathematically rigorous understanding is possible, in contrast to the present situation for short range models, but it took a long way, and the results are still far from covering all aspects. Here is the SK-model:

One starts with a countable number of standard Gaussian random variables, defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, indexed as a matrix $(g_{ij})_{1 \leq i < j}$. If we specify the dependence on ω, we write $g_{ij}(\omega)$. Of course, one can take $\Omega = \mathbb{R}^{\mathbb{N}}$, but probabilists usually don't like to fix that. Then for any $N \in \mathbb{N}$, one defines the random Hamiltonian

$$(1) \qquad H_{N,\omega}(\sigma) \stackrel{\text{def}}{=} -\frac{1}{\sqrt{N}} \sum_{1 \leq i < j \leq N} g_{ij}(\omega) \sigma_i \sigma_j,$$

where $\sigma = (\sigma_1, \ldots, \sigma_N) \in \Sigma_N \stackrel{\text{def}}{=} \{-1, 1\}^N$. The minus sign is of course of no relevance, and is put only to please the physicists. A slight generalization is to allow the presence of an external field in the form

$$(2) \qquad H_{N,\omega}(\sigma) \stackrel{\text{def}}{=} -\frac{1}{\sqrt{N}} \sum_{1 \leq i < j \leq N} g_{ij}(\omega) \sigma_i \sigma_j - h \sum_{i=1}^{N} \sigma_i,$$

with $h \in \mathbb{R}$ a parameter. The external field adds some non-trivial complications. We will however stick to the Hamiltonian (1). As in the usual mean-field models, there is no geometric structure of the index set $\{1, \ldots, N\}$.

From standard mean-field models one would expect a factor $\frac{1}{N}$ instead of $\frac{1}{\sqrt{N}}$. However, a moment's reflection shows that the square root is the appropriate scaling: The main issue is that, given a site i, the influence of the other sites on σ_i is of order 1. To achieve this in the random situation, one has to have the factor $\frac{1}{\sqrt{N}}$ in front.

One then defines in the usual way the partition function which now depends on the Gaussian variables g_{ij}, i.e. on ω:

$$(3) \qquad Z_{N,\beta,\omega} \overset{\text{def}}{=} \sum_{\sigma} \exp\left[-\beta H_{N,\omega}(\sigma)\right],$$

and the random Gibbs measure

$$(4) \qquad G_{N,\beta,\omega}(\sigma) \overset{\text{def}}{=} \frac{1}{Z_{N,\beta,\omega}} \exp\left[-\beta H_{N,\omega}(\sigma)\right].$$

There are two sources of probability, namely for fixed ω the probability measure $G_{N,\beta,\omega}$ on Σ_N, and then the randomness of this law itself, as it depends in a non-trivial way on ω. One usually calls this a random probability distribution.

REMARK 1. — *For the sake of simplicity, we restrict the discussion here completely to the original SK-model. Everything works for so-called p-spin models and mixtures of p-spin models. The p-spin models have the Hamiltonian*

$$H_N^{(p)}(\sigma) \overset{\text{def}}{=} \frac{1}{N^{(p-1)/2}} \sum_{1 \leq i_1 < i_2 < \cdots < i_p \leq N} g_{i_1,i_2,\dots,i_p} \sigma_{i_1} \sigma_{i_2} \cdots \sigma_{i_p},$$

where the g's as before are i.i.d. standard Gaussians.

We can now formulate the ultrametricity conjecture, although its importance will stay quite mysterious for the moment (and maybe still till the end of these notes).

A metric d on a space S is called an **ultrametric**, if it satisfies the stronger triangle inequality:

$$d(x,y) \leq \max\left(d(x,z), d(y,z)\right), \ \forall x, y, z \in S.$$

This is essentially equivalent with the metric space having a tree structure, S being the set of leaves, and d being the (weighted) graph distance. Σ_N is a metric space under the Hamming distance, counting the number of sites on which two elements differ. This is evidently not an ultrametric. However, the ultrametricity conjecture states that it is approximately so under the above Gibbs measure, as $N \to \infty$.

The Hamming distance is not quite appropriate due to the inherent symmetry of the Gibbs measure under the reflection $\sigma \to -\sigma$. A better distance is the L^2-distance between $H_N(\sigma)$ and $H_N(\sigma')$. It can be expressed through the **overlaps**

$$(5) \qquad R_N(\sigma, \sigma') \overset{\text{def}}{=} \frac{1}{N} \sum_{i=1}^{N} \sigma_i \sigma_i',$$

which will play a very important rôle. The importance is coming from the fact that the covariances of the Hamiltonians are given in terms of the overlaps

$$(6) \qquad \mathbb{E} H_N(\sigma) H_N(\sigma') = \frac{N}{2} R_N(\sigma, \sigma')^2 - \frac{1}{2}.$$

Then

$$\|H_N(\sigma) - H_N(\sigma')\|_{L^2}^2 = N\left(1 - R_N^2(\sigma, \sigma')\right),$$

and by norming with $\sqrt{N}$ we take as the distance between σ, σ':

$$d(\sigma, \sigma') \stackrel{\text{def}}{=} \sqrt{1 - R_N^2(\sigma, \sigma')}.$$

Evidently, it is not quite a metric as $d(\sigma, -\sigma) = 0$.

Let's now consider a triple $(\sigma, \sigma', \sigma'') \in \Sigma_N^3$. For fixed ω we may consider the product measure $G_{N,\beta,\omega}^{\otimes 3}$ on Σ_N^3 which means that we choose the three elements $\sigma, \sigma', \sigma''$ independently. Such independent copies are called "replicas" in the literature. However, one has to remember that this independence is for fixed ω (which in physics jargon is the so-called "quenched" measure). $\sigma, \sigma', \sigma''$ are far from independent, if one considers the joint law. The ultrametricity conjecture states

CONJECTURE 2. — *Given any $\varepsilon > 0$, one has for any $\beta \geq 0$ (and $h \in \mathbb{R}$ in case of the presence of an external field)*

$$(7) \qquad \lim_{N \to \infty} \mathbb{E} G_{N,\beta,\cdot}^{\otimes 3} \left(\frac{d(\sigma, \sigma')}{N} \geq \max\left(\frac{d(\sigma, \sigma'')}{N}, \frac{d(\sigma', \sigma'')}{N} \right) + \varepsilon \right) = 0.$$

Expressed in terms of the overlaps, the statement reads that for all $\varepsilon > 0$

$$(8) \qquad \lim_{N \to \infty} \mathbb{E} G_{N,\beta,\cdot}^{\otimes 3} \left(\frac{|R_N(\sigma, \sigma')|}{N} \leq \min\left(\frac{|R_N(\sigma, \sigma'')|}{N}, \frac{|R_N(\sigma', \sigma'')|}{N} \right) - \varepsilon \right) = 0.$$

In the above form, this is an unproved statement. The results which has been proved is somewhat weaker, and will be discussed in more details later.

The relevance of ultrametricity is not immediately evident. In physics literature, the ultrametricity however played a crucial rôle in the non-rigorous derivations for instance of the Parisi formula for the free energy.

There is a rather puzzling point connected with the overlaps. The SK-Gibbs weights are completely symmetric under the transformation $\sigma \to -\sigma$. Therefore, the distribution of $R_N(\sigma, \sigma')$ under $G_{N,\beta,\omega}^{\otimes 2}$ is symmetric on $\mathbb{R}$. Later, we will however work with Gibbs measures under slightly perturbed Hamiltonian, for which for large N, Talagrand's positivity property holds: $R_N \geq 0$ with large probability (≈ 1 in the $N \to \infty$ limit).

There is one aspect one has to be aware of, to see that the statement of Conjecture 2 is highly non-trivial. Take first the trivial case where $\beta = 0$, so that the σ_i are just coming from independent coin tossing (biased if $h \neq 0$). Then by the law of large numbers, $R(\sigma, \sigma')$ is for large N close to a fixed constant (0 if $h = 0$) and therefore also $d(\sigma, \sigma')$ is close to a constant. Then of course, (7) is trivially true. However, in the SK-case, the overlaps $R(\sigma, \sigma')$ stay random also in the $N \to \infty$ limit, under independent replicas, provided β is large enough, i.e. in the so-called spin glass phase. If they stay random, then there is a priori no reason why (7) should hold.

A basic quantity of interest is the so-called free energy:

$$f(\beta) \stackrel{\text{def}}{=} \lim_{N \to \infty} \frac{1}{N} \log Z_{N,\beta}.$$

In the case of the presence of an external field, one would have $f(\beta, h)$. The existence is far from evident, and in principle, the limit, if it exists, could be a random variable. However, it is not too difficult to prove the self-averaging of the free energy, meaning that $\frac{1}{N} \log Z_{N,\beta}$ is for large N close to its $\mathbb{E}$-expectation. This follows from standard Gaussian isoperimetry inequalities, see e.g. [17]. Therefore, the question is if

$$\lim_{N\to\infty} \frac{1}{N} \mathbb{E} \log Z_{N,\beta}$$

exists. For small β, this question had first been addressed in two papers [2], [13] where it is proved among other things, that for β small one has

$$(9) \qquad f(\beta) = \lim_{N\to\infty} \frac{1}{N} \mathbb{E} \log Z_{N,\beta} = \frac{\beta^2}{4} + \log 2$$

which is $\lim_{N\to\infty} N^{-1} \log \mathbb{E} Z_{N,\beta}$. This value cannot be correct for large β, as it would imply that the entropy is negative, as has already been observed by Sherrington and Kirkpatrick. Therefore, the result already proves that there is a phase transition in the system. The existence of the limit for all β was first proved by Guerra and Toninelli [16] by a simple but very clever argument.

In the original paper by Sherrington and Kirkpatrick [23], (9) was predicted based on non-rigorous replica computation. This computation, and an ansatz which has become known under the name "replica symmetry" predicted the following formula for the free energy

$$(10) \qquad f^{\mathrm{RS}}(\beta) = \inf_{q\in[0,1]} \left\{ \frac{\beta^2(1-q)}{4} + \int \log \cosh(\beta\sqrt{q}z) \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz \right\} + \log 2.$$

It is easy to see that $f^{\mathrm{RS}}(\beta) = \beta^2/4 + \log 2$ for $\beta \leq 1$ with a phase transition at $\beta = 1$ after which $f^{\mathrm{RS}}(\beta) < \beta^2/4 + \log 2$. However, it was already noted by Sherrington and Kirkpatrick that for large β, one has $f(\beta) \neq f^{\mathrm{RS}}(\beta)$, as equality would imply that the entropy of the system is negative. In fact, one now knows that

$$f(\beta) < f^{\mathrm{RS}}(\beta) < \beta^2/4$$

for large β. The correct $f(\beta)$ is given by a much more complicated variational formula, the famous Parisi formula, see [21], first rigorously proved in two papers by Guerra [15] and Talagrand [24], and recently, in a way connecting it directly to the ultrametricity problem by Panchenko [20]. We will give the Parisi formula later (30), (31). For more background of the physics theory, see [18].

We finish this short introduction by describing a much simpler model, the so-called **random energy model**, REM for short. Although not directly relevant for the SK-model, it has played a considerable rôle for the understanding of the latter.

The SK-model is completely described through the field $\{H_N(\sigma)\}_{\sigma\in\Sigma_N}$ of random variables. As the g_{ij} are i.i.d. standard Gaussians, the field is a centered Gaussian field (in the $h = 0$ case), and so it is described by its covariances given above in (6). Given the difficulty of the SK-model, Derrida [10] investigated the situation where the Hamiltonian consists of i.i.d. Gaussians (centered). In order to catch some of the

features of the SK-model, one should take them with variances of order N, and then one can as well take them equal to N: $\{H_N(\sigma)\}_{\sigma \in \Sigma_N}$ is a field of independent centered Gaussians with variance N. The structure of Σ_N as a discrete hypercube is no longer of relevance. This model is easy to analyze, but it is interesting that it has a phase transition.

$$f(\beta) \overset{\text{def}}{=} \lim_{N \to \infty} \frac{1}{N} \log Z_N(\beta) = \lim_{N \to \infty} \frac{1}{N} \log \sum_\sigma \mathrm{e}^{-\beta H_N(\sigma)}$$

$$= \lim_{N \to \infty} \frac{1}{N} \mathbb{E} \log Z_N(\beta)$$

is given by

$$f(\beta) = \begin{cases} \log 2 + \frac{\beta^2}{2} & \text{for } \beta \leq \sqrt{2 \log 2} \\ \sqrt{2 \log 2}\, \beta & \text{for } \beta \geq \sqrt{2 \log 2} \end{cases}.$$

Therefore, there is a critical point $\beta_{\mathrm{cr}} \overset{\text{def}}{=} \sqrt{1 \log 2}$ at which f has no continuous second derivative. (The first derivative is still continuous.) The proof of this formula is easy. For $t > 0$

$$\mathbb{P}(H_N(\sigma) \geq tN) \approx \exp\left[-\frac{t^2 N}{2}\right],$$

up to a irrelevant polynomial factor (in N). Therefore, by the law of large numbers, one has for $t < \sqrt{2 \log 2}$

$$\#\{\sigma : H_N(\sigma) \geq tN\} \approx \exp\left[\left(\log 2 - \frac{t^2}{2}\right) N\right]$$

with probability ≈ 1, and for $t > \sqrt{2 \log 2}$, $\#\{\sigma : H_N(\sigma) \geq tN\} = \emptyset$, with large probability. From that one easily gets

$$\sum_\sigma \exp[\beta H_N(\sigma)] \approx \exp\left[N \log 2 + N \sup_{t \leq \sqrt{2 \log 2}}\left(\beta t - \frac{t^2}{2}\right)\right]$$

$$= \exp[N f(\beta)],$$

with $f(\beta)$ as given above. The approximations are up to subexponential factors which are negligible in the limit.

Although the REM is too simple to shed much light on the behavior of more complicated spin glasses, there are some aspects which are similar, as we will explain later. The REM, and a generalization of it, the so-called generalized random energy model GREM introduced in [11], have played a very important rôle in the recent mathematical development of mean-field spin-glass theory.

2. INFINITE OVERLAP STRUCTURES

For mean-field models, there is usually no concept of an infinite Gibbs measure, and therefore, strictly speaking, also no mathematically sound definition of "pure states".

The fact that in [18] and other texts in the physics literature, the notion of "pure states" plays an important rôle, has led to misunderstandings and controversies. However, for SK-type models, there is a way out, surprising at first, which was developed by a number of authors, Aizenman, Arguin, Ghirlanda, Guerra, and others, and finally most successfully by Panchenko. The basic observation is that essential all relevant information is encoded in the law of the overlaps of replicas under the $\mathbb{P}$-average of the infinite product of the Gibbs measure. To be precise: Consider an arbitrary random probability measure $G_{N,\omega}$ on $\Sigma_N = \{-1, 1\}^N$, where we assume, in future without specially mentioning, that the map $\omega \to G_{N,\omega}(\sigma)$ is measurable for any $\sigma \in \Sigma_N$. The infinite product $G_{N,\omega}^{\otimes \mathbb{N}}$ is a probability measure on $\Sigma_N^{\otimes \mathbb{N}}$, depending still (measurably) on ω. Then we define the averaged measure by

$$\mathrm{AV}(d\sigma) \stackrel{\mathrm{def}}{=} \int G_{N,\omega}^{\otimes \mathbb{N}}(d\sigma)\,\mathbb{P}(d\omega)$$

on $\Sigma_N^{\otimes \mathbb{N}}$, the latter equipped with the infinite product σ-field.

Next, we consider overlaps from "replicas": Given $\left\{\sigma^\ell\right\}_{\ell \in \mathbb{N}} \in \Sigma_N^{\otimes \mathbb{N}}$, $\sigma^\ell = \left(\sigma_1^\ell, \ldots, \sigma_N^\ell\right)$, we define the infinite matrix R^N of overlaps with components

$$R_{\ell,\ell'}^N \stackrel{\mathrm{def}}{=} \frac{1}{N} \sum_{i=1}^N \sigma_i^\ell \sigma_i^{\ell'}.$$

The matrix elements are in $[-1, 1]$, and the diagonal elements are 1. Evidently, the matrix is symmetric and positive semidefinite. We write ν_N for the distribution of R^N under AV. The set $\mathbb{M}$ of infinite, symmetric, positive semi-definite matrices with entries $\in [-1, 1]$ is clearly a compact space under the topology of componentwise convergence. The set of probability measures $\mathcal{P}(\mathbb{M})$ on this space is therefore a compact space, too, under weak convergence. The sequence $\{\nu_N\}$ has therefore convergent subsequences, and one can consider the set of possible limits of subsequences, as $N \to \infty$.

This concept is very ingenious: In classical (short range) systems, one would try to consider limits $\lim_{N \to \infty} G_N$. This does not make sense for mean field models, as the interactions between different sides vanish in the $N \to \infty$ limit. The above concept of limits $\lim_{N \to \infty} \nu_N$, maybe along subsequences, makes however perfectly sense, and there is a beautiful structure theorem for the possible limits. It turns out that all possible limits are generated in an abstract way similarly as the finite N laws. The limits can in fact be described through a randomization, as will be explained now. This randomization acts as a kind of an "infinite Gibbs measure", but it is not constructed from the original Gibbs measure, but only through the distribution of the overlaps, and an abstract representation theorem.

The distribution ν_N has an important symmetry property: If $\tau : \mathbb{N} \to \mathbb{N}$ is a permutation of finitely many elements, let $\hat{\tau} : \mathbb{M} \to \mathbb{M}$ be the mapping which exchanges the indices of the matrices accordingly. Evidently $\nu_N \hat{\tau}^{-1} = \nu_N$ where $\nu_N \hat{\tau}^{-1}$ denotes the induced measure under the mapping $\hat{\tau}$. Let $\mathcal{P}^{\mathrm{inv}}(\mathbb{M})$ be the set of probability measures on $\mathbb{M}$ which have this invariance property. $\mathcal{P}^{\mathrm{inv}}(\mathbb{M})$ is a closed subset of $\mathcal{P}(\mathbb{M})$. We

formulate now the key abstract representation property for elements of $\mathcal{P}^{\mathrm{inv}}(\mathbb{M})$. For that, let H be one of the standard (real) infinite dimensional separable Hilbert spaces with Borel-σ-field $\mathcal{H}$.

THEOREM 3 (Dovbysh-Sudakov). — *Any $\nu \in \mathcal{P}^{\mathrm{inv}}(\mathbb{M})$ can be represented in the following way. There exist a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a Markov kernel G from $(\Omega, \mathcal{F})$ to $(H \times \mathbb{R}^+, \mathcal{H} \otimes \mathcal{B}_{\mathbb{R}^+})$ such that ν is the law of $(H \times \mathbb{R}^+)^{\mathbb{N}} \ni ((h_\ell, a_\ell))_{\ell \in \mathbb{N}} \rightarrow (\langle h_\ell, h_{\ell'}\rangle (1 - \delta_{\ell, \ell'}) + a_\ell \delta_{\ell, \ell'})_{\ell, \ell'} \in \mathbb{M}$ under $\int \mathbb{P}(d\omega) G^{\otimes \mathbb{N}}(\omega, \cdot)$. ($\langle \cdot, \cdot \rangle$ is the inner product in H.)*

A short explanation about the formalities: A Markov kernel from a measurable space $(\Omega, \mathcal{F})$ to a second one $(\Omega', \mathcal{F}')$ is a mapping $G : \Omega \times \mathcal{F}' \rightarrow [0, 1]$ such that for any ω, the mapping $A \rightarrow G(\omega, A)$ is a probability measure on $\mathcal{F}'$, and for any $A \in \mathcal{F}'$, the mapping $\omega \rightarrow G(\omega, A)$ is measurable. For fixed ω, $G^{\otimes \mathbb{N}}(\omega, \cdot)$ denotes the infinite product measure on $(\Omega'^{\mathbb{N}}, \mathcal{F}'^{\otimes \mathbb{N}})$, and $\int \mathbb{P}(d\omega) G^{\otimes \mathbb{N}}(\omega, \cdot)$ is simply the averaged one.

Essentially, the Dovbysh-Sudakov theorem says that any measure in $\mathcal{P}^{\mathrm{inv}}(\mathbb{M})$ is the law of the matrix of inner products under the averaged infinite product of a random probability on H. There is a slight modification on the diagonal, as this applies only to the off-diagonal part of the random matrix, and the diagonal elements are produced by a separate random mechanism, as described above.

For the case of the overlap distribution ν_N, the representation is already given by the Gibbs distribution, as we can regard Σ_N as a subset of a Hilbert space. The theorem therefore constructs a kind of a substitute for the limit of the Gibbs measures.

In the application of the theorem to the SK-model it will turn out that any possible limit ν of ν_N will have an additional property, namely that it satisfies the Ghirlanda-Guerra identities. This will imply that G will have a very special structure. This structure is the content of the next section.

3. RUELLE'S PROBABILITY CASCADES

3.1. The Poisson-Dirichlet point process

Let $\lambda \in (0, 1)$ be a parameter. We consider a point process on the positive real line $\mathbb{R}^+ = (0, \infty)$ with intensity measure $\mu_\lambda(dx) = \lambda x^{-1-\lambda} dx$. A Poisson point process is a random point configuration on the base space, here $\mathbb{R}^+$. Intuitively, an infinitesimal interval $[x, x + dx]$ contains a point with probability $\lambda x^{-1-\lambda} dx$, and no point with probability $1 - \lambda x^{-1-\lambda} dx$, and the events in different infinitesimal intervals are independent. From this it follows that in every finite interval I, the number of points in this interval is Poisson distributed with parameter $\int_I \lambda x^{-1-\lambda} dx$. We write PPP (μ) for a Poisson point process with intensity measure μ. As $\mu_\lambda([a, \infty)) < \infty$, $a > 0$, there are only finitely many points in the interval $[a, \infty)$. On the other hand, as $\mu_\lambda(\mathbb{R}^+) = \infty$, there are almost surely infinitely many points on $\mathbb{R}^+$. As our point process has a largest

point, we can order the random points downwards leading to a decreasing sequence of real-valued random variables $\xi_1 > \xi_2 > \cdots > 0$ which represent the random points of the point process. From the fact that $\int_0^1 x\mu_\lambda(dx) < \infty$, it follows that $\sum_i \xi_i 1_{\xi_i \leq 1}$ has finite expectation and is therefore finite. So, it follows that $\sum_i \xi_i < \infty$ almost surely as the number of points above 1 is finite. We put $\bar{\xi}_i \stackrel{\text{def}}{=} \xi_i / \sum_j \xi_j$, leading to a point process in $(0,1)$ whose points sum up to 1, i.e. to a random probability distribution on the integers, with decreasing weights. This point process is usually called the Poisson-Dirichlet point process with parameter λ. We denote it by $\mathrm{PD}(\lambda)$. It is easy to see that a $\mathrm{PD}(\lambda)$ is not a Poisson point process. We have fixed the ξ_i to be decreasing, and therefore the $\bar{\xi}_i$, too.

We will write expectations under $\{\bar{\xi}_i\}$ by $\langle \cdot \rangle$, whereas the probability measure which governs the point process is denoted by $\mathbb{P}$, and expectations by $\mathbb{E}$. The point processes have a number of crucial invariance properties. For instance, it is not difficult to prove the following result, which follows from the special form of the intensity measure of the original Poisson point process.

PROPOSITION 4. — *Assume* $\{\xi_i\}$ *is a* $\mathrm{PPP}\left(\lambda x^{-\lambda-1}dx\right)$ *with* $\lambda \in (0,1)$, *and assume that* $\{X_i\}$ *are i.i.d. positive random variables, independent also of the point process, satisfying* $\mathbb{E}\left(X_i^\lambda\right) < \infty$. *Then* $\left\{\xi_i X_i / \sum_j \xi_j X_j\right\}$ *is a Poisson-Dirichlet point process with parameter* λ.

If the ξ_i are ordered downwards, then of course the $\xi_i X_i$ are no longer ordered. Ordering the new points therefore defines a random permutation of the natural numbers.

For the setting described above, there is a very simple notion of overlaps. If we fix the point process $\bar{\xi} = \{\bar{\xi}_i\}$, ordered downwards, which as remarked above, is a (random) probability distribution on $\mathbb{N}$, we consider independent copies $\sigma^1, \sigma^2, \ldots$ of $\mathbb{N}$-valued random variables, distributed according to $\bar{\xi}$. This amounts just to consider the infinite product measure $\bar{\xi}^{\otimes \mathbb{N}}$ on $\mathbb{N}^{\mathbb{N}}$. The σ^i are called "replicas". Given these replicas, we define their overlaps in a rather trivial way:

$$(11) \qquad R_{ij} \stackrel{\text{def}}{=} \begin{cases} 1 & \text{if } \sigma^i = \sigma^j \\ 0 & \text{if } \sigma^i \neq \sigma^j. \end{cases}$$

Evidently (R_{ij}) is a random element in $\mathcal{M}$. The R_{ij} satisfy simple restrictions. In fact, one just divides $\mathbb{N}$ according to the equivalence relation $i \sim j \iff \sigma^i = \sigma^j$, which defines a partition of $\mathbb{N}$, and then the overlap matrix is one inside the same class. Let's consider this partition restricted to $\{1, \ldots, N\}$, $N \in \mathbb{N}$. For a given partition of this set into k disjoint subset $A_1, \ldots, A_k$ with $n_i = |A_i|$, $\sum_{i=1}^k n_i = N$, the $\bar{\xi}^{\otimes N}$-probability for this partition, or the corresponding overlap matrix, is given by

$$\sum_{i_1,\ldots,i_k}^* \bar{\xi}_{i_1}^{n_1} \bar{\xi}_{i_2}^{n_2} \cdot \ldots \cdot \bar{\xi}_{i_k}^{n_k},$$

where $\sum^*$ means that the summation indices are all different. By applying fairly standard tools from point process theory, one can easily evaluate

$$\Phi\left(n_1,\ldots,n_k\right) \overset{\text{def}}{=} \mathbb{E} \sum_{i_1,\ldots,i_k}^* \bar{\xi}_{i_1}^{n_1} \bar{\xi}_{i_2}^{n_2} \cdot \ldots \cdot \bar{\xi}_{i_k}^{n_k}.$$

PROPOSITION 5. — *Let $n_1,\ldots,n_k \in \mathbb{N}$, and $N \overset{\text{def}}{=} \sum_{i=1}^k n_i$. If $\{\bar{\xi}_i\}$ is a $\mathrm{PD}(\lambda)$ then*

$$(12) \qquad \Phi\left(n_1,\ldots,n_k\right) = \frac{(k-1)!}{(N-1)!}\lambda^{k-1}\prod_{i=1}^k g\left(n_i,\lambda\right)$$

where

$$g\left(n,\lambda\right) \overset{\text{def}}{=} (n-1-\lambda)(k-2-\lambda)\cdots(1-\lambda)$$

if $n \geq 2$, and $g\left(1,\lambda\right) \overset{\text{def}}{=} 1$.

We again define AV_λ to be the measure $\mathrm{AV}_\lambda \overset{\text{def}}{=} \int \bar{\xi}^{\otimes\mathbb{N}} d\mathbb{P}$ on $\mathbb{N}^\mathbb{N}$. The above functions Φ define the distribution of the overlap matrix under this measure uniquely. A special case is $k = 1$, $n_1 = 2$ which gives

$$(13) \qquad \mathrm{AV}_\lambda\left(R_{1,2} = 1\right) = \mathbb{E}\sum_i \bar{\xi}_i^2 = 1 - \lambda.$$

The proposition has as a corollary the **Ghirlanda-Guerra identities** as we will now prove.

The identities express the conditional distribution of $R_{1,n+1}$ under AV_λ conditioned on $R^{(n)} \overset{\text{def}}{=} (R_{ij})_{i,j\leq n}$. As remarked above, $R^{(n)}$ is the same as the partition of $\{1,\ldots,n\}$ through the equivalence relation given by equality of the replicas. $\{R_{1,n+1} = 1\}$ is the event that the "newcomer" $\sigma^{(n+1)}$ belongs to the same class (via equality) as the first one. We compute the conditional probability of this event given the partition of $\{1,\ldots,n\}$. Let n_1 be the number of elements in the class of σ^1 among the first n, or to put it simpler, 1 plus the number of replicas among $\sigma^2,\ldots,\sigma^n$ which are equal to σ^1, and let $n_2,\ldots,n_k$ be the other numbers. (They will cancel out in the computation.) Then

$$\mathrm{AV}_\lambda\left(R_{1,n+1} = 1|R^{(n)}\right) = \frac{\Phi\left(n_1+1,\ldots,n_k\right)}{\Phi\left(n_1,\ldots,n_k\right)}$$

$$(14) \qquad\qquad = \frac{g\left(n_1+1,\lambda\right)}{ng\left(n_1,\lambda\right)} = \frac{n_1-\lambda}{n}$$

$$= \frac{n_1-1}{n} + \frac{1-\lambda}{n} = \frac{n_1-1}{n} + \frac{1}{n}\mathrm{AV}_\lambda\left(R_{1,2} = 1\right).$$

This can be interpreted that conditionally on $R^{(n)}$, $R_{1,n+1}$ picks with probability $1-1/n$ one of the places $j \in \{2,\ldots,n\}$ and then $R_{1,j}$, and with probability $1/n$, it chooses $R_{1,n+1}$ independently with the distribution of $R_{1,2}$, i.e.

$$(15) \qquad \mathcal{L}_{\mathrm{AV}_\lambda}\left(R_{1,n+1}|R^{(n)}\right) = \frac{1}{n}\sum_{j=2}^n \delta_{R_{1,j}} + \frac{1}{n}\mathcal{L}_{\mathrm{AV}_\lambda}\left(R_{1,2}\right)$$

(15) is called the **Ghirlanda-Guerra identity** for the Poisson-Dirichlet distribution.

(13) and (14) determine the probability distribution of (R_{ij}) under AV_λ uniquely.

The proof can be found in [20]: One derives (12) recursively from the two identities.

It's not difficult to place the above simple setting into the framework of Section 2: One chooses in the Hilbert space H an orthonormal sequence $\{h_i\}$ and identifies it with $i \in \mathbb{N}$. $\{\bar{\xi}_i\}$ defines a random probability measure on H by giving h_i weight $\bar{\xi}_i$. This plays the rôle of G of the last section. We therefore have a (rather trivial) example of a structure as described in the last section, with the additional property that the Ghirlanda-Guerra identities are satisfied. This is too simple for our purpose, and we generalize it in the next subsection.

It may be instructive to see in which way this point process appears as a limit of the REM, i.e. when the $H_N(\sigma)$ are independent Gaussians with variance N, $1 \le \sigma \le 2^N$. It is a standard and simple fact from extreme value theory that for some sequence of real numbers $a_N \to \infty$, the point process

$$\sum_\sigma \delta_{-H_N(\sigma)-a_N} \to^w \mathrm{PPP}(\mu)$$

with $\mu(dx) = \sqrt{2\log 2}\exp\left[-\sqrt{2\log 2}x\right]dx$. $\to^w$ here means convergence in distribution, i.e. convergence of the laws. The random point measure on the left hand side is a random element in the space of Radon measures on $\mathbb{R}$ equipped with the vague topology. This means that as $a_N \to \infty$, the bulk of the mass of the measure disappears to $-\infty$. The sequence $\{a_N\}$ can easily be given explicitly, but it is of no importance for us.

It is then an equally standard fact from point process theory, that

$$\sum_\sigma \delta_{\exp[-\beta H_N(\sigma)-\beta a_N]} \to^w \mathrm{PPP}(\mu')$$

where μ' is the measure on $\mathbb{R}^+$ given by $\mu'(dx) = \lambda x^{-\lambda-1}dx$, with $\lambda = \sqrt{2\log 2}/\beta$.

Remark now that the Gibbs measure of the REM, which is obtained by normalizing the $\exp[-\beta H_N(\sigma)]$ can as well be obtained by normalizing $\exp[-\beta H_N(\sigma) - \beta a_N]$. From that, it is plausible, that if $\beta > \sqrt{2\log 2}$, i.e. $\lambda < 1$, one has

$$(16) \qquad \sum_\sigma \delta_{G_{N,\beta}(\sigma)} \to^w \mathrm{PD}(\lambda).$$

There is a slight problem to justify that, since the normalization operation is not continuous, but it can easily be done.

In a sense, the program is now to achieve something like that for the SK-model, with the outcome being the more complicated objects of the next section. However, the reader is warned that a statement like (16) is for the SK-model presently totally beyond reach, and can in fact not be correct in this strong form. For much simpler models, like branching random walks, Gaussian free fields, and other so-called log-correlated models, such results have been obtained recently, or are under intensive research, see

for instance [5], [1], [9] and others, but presently, nobody dares to touch the SK-model from this viewpoint.

3.2. The cascades

We move to a more complicated model which gives a more complicated notion of overlaps, and which satisfies the same Ghirlanda-Guerra property as was described in the last section, but with the crucial difference, that overlaps are now not just 0-1-valued.

The model was invented by David Ruelle [22] who gave a reinterpretation of Derrida's generalized random energy model. Ruelle's probability cascades are indeed the limits of Derrida's GREMs, a fact which Ruelle evidently considered to be evident, but which was mathematically proved only later by Bovier and Kurkova [8]. This point is however irrelevant for the discussion here, and we just describe the Ruelle version of the GREM.

Let $K \in \mathbb{N}$ be fixed. This counts the number of levels the cascades has.

The cascades are defined by piling several of the point processes, with increasing parameters. Fix $K \in \mathbb{N}$ and parameters $0 < \lambda_1 < \lambda_2 < \cdots < \lambda_K < 1$. Choose first a PPP $\left(\lambda_1 t^{-\lambda_1 - 1} dt\right)$ described by $\left\{\xi_{i_1}^{(1)}\right\}_{i_1 \in \mathbb{N}}$, where it is assumed that the variables are ordered downwards. Then for any i_1 choose independently PPP $\left(\lambda_2 t^{-\lambda_2 - 1} dt\right)$'s, i.e. the countable number of point processes $\left\{\xi_{i_1,i_2}^{(2)}\right\}_{i_2 \in \mathbb{N}}$. Then for any pair (i_1, i_2) one chooses independently PPP $\left(\lambda_2 t^{-\lambda_2 - 1} dt\right)$'s $\left\{\xi_{i_1,i_2,i_3}^{(3)}\right\}_{i_3 \in \mathbb{N}}$ etc., up to level K. We can multiply all these variables, obtaining

$$X_{\mathbf{i}} \stackrel{\text{def}}{=} \xi_{i_1}^{(1)} \xi_{i_1,i_2}^{(2)} \cdot \cdots \cdot \xi_{i_1,i_2,\ldots,i_K}^{(K)}, \ \mathbf{i} = (i_1, \ldots, i_K).$$

It is not difficult to see that $\sum_{\mathbf{i}} X_{\mathbf{i}} < \infty$, and therefore, we can again normalize

(17)
$$\bar{X}_{\mathbf{i}} \stackrel{\text{def}}{=} \frac{X_{\mathbf{i}}}{\sum_i X_{\mathbf{i}}}.$$

If we order these random points in $(0, 1)$ downwards, one obtains a random probability distribution on $\mathbb{N}$. This involves a random bijection $\psi : \mathbb{N} \to \mathbb{N}^K$, giving $\pi_i \stackrel{\text{def}}{=} \bar{X}_{\psi(i)}$ with $\pi_1 > \pi_2 > \cdots$. A quite astonishing observation which is not difficult to prove, and which is a consequence of invariance properties like the one stated in Proposition 4, is that, as a point process, $\{\pi_i\}$ is simply a PD (λ_K). Therefore, nothing new seems to have come up by this complicated construction. However, there is now inherently a non-trivial overlap structure of the points. We define $\tau : \mathbb{N} \times \mathbb{N} \to \{0, \ldots, K\}$ by setting $\tau(i, j) = k$ if the first k components of $\psi(i)$ agree with the ones of $\psi(j)$ but not the next. Evidently, $\tau(i, j) = K$ holds only if $i = j$. The relation $\tau(i, j) \geq k$ defines a random equivalence relation on $\mathbb{N}$, i.e. a random partition $\mathcal{P}_k$ of $\mathbb{N}$, with $\mathcal{P}_K$ being the trivial one consisting of the set of one point sets, and $\mathcal{P}_0$ the trivial one with just one class $\mathbb{N}$. All the other ones are non-trivial, and evidently, $\mathcal{P}_{k+1}$ is a finer partition than $\mathcal{P}_k$. Actually, viewed backwards as a sequence of random partitions $\mathcal{P}_K, \mathcal{P}_{K-1}, \ldots, \mathcal{P}_1, \mathcal{P}_0$, it has a nice Markovian structure, as first proved in [7]. A most

astonishing fact is that as random objects the point processes $\{\pi_i\}$ and the sequence $\mathcal{P} = (\mathcal{P}_K, \mathcal{P}_{K-1}, \ldots, \mathcal{P}_1, \mathcal{P}_0)$ are stochastically independent. Remark that the original sequences of point processes fix ψ and therefore also the sequences of partitions, or equivalently the function τ. One should keep in mind a trivial point: If one "forgets" from where the point process $\{\pi_i\}$ is coming, one cannot define the overlaps. They are only defined by keeping the information about the cascades.

As in the last section, we define the random probability distribution $\pi^{\otimes \mathbb{N}}$ on $\mathbb{N}^{\mathbb{N}}$ with projections (the "replicas") $\sigma^i : \mathbb{N}^{\mathbb{N}} \to \mathbb{N}$. For two replicas σ^i, σ^j we define the overlap through

$$R_{ij} \overset{\text{def}}{=} \tau\left(\sigma^i, \sigma^j\right)$$

which defines a random (under $\mathbb{P}$) matrix, taking now values in $0, \ldots, K$.

Usually, in order to keep the relations with the SK-overlaps, one wants to have the overlaps to take values in $[0, 1]$ which can be achieved with a fixed monotone function $q : \{0, \ldots, K\} \to [0, 1]$, typically with

(18) $$q_0 = 0 < q_1 < \cdots < q_K \leq 1,$$

but this is not important for the moment. Remark that randomness enters here again in two ways: First, the point processes and therefore π and τ are chosen random, and given these, we choose the σ^i randomly under π. We again denote the averaged law by AV_λ, this time with the parameter $\lambda = (\lambda_1, \ldots, \lambda_K)$.

One can easily compute the distribution of the overlaps under AV_λ:

$$\text{AV}_\lambda\left(R_{1,2} = k\right) = \lambda_{k+1} - \lambda_k, \ \ k = 0, \ldots, K,$$

with the convention that $\lambda_0 \overset{\text{def}}{=} 0$ and $\lambda_{K+1} \overset{\text{def}}{=} 1$. If one uses the "reweighting" of the overlaps by the q_k in (18), one has

$$\text{AV}_\lambda\left(R_{1,2} = q_k\right) = \lambda_{k+1} - \lambda_k, \ \ k = 0, \ldots, K.$$

This distribution will become very important later, and is called the **Parisi measure** ζ on $\mathbb{R}^+$ which for the moment is discrete:

(19) $$\zeta\left(q_k\right) = \lambda_{k+1} - \lambda_k, \ \ k = 0, \ldots, K.$$

The issue with the random bijection ψ ordering the probabilities may be slightly puzzling: There are two ways to describe the model: Either we stay with the original numbering of the Gibbs weights by $\bar{X}_{\mathbf{i}}$, and then the overlap of $\mathbf{i}$ and $\mathbf{j} \in \mathbb{N}^K$ is deterministically determined, or we take as index the natural numbers and the Gibbs weights π_i, but then, the overlap is itself random through the random bijection ψ.

Remarkably, the matrix R under AV_λ satisfies the Ghirlanda-Guerra identity:

Given $n \in \mathbb{N}$, $n \geq 2$, one has for $k = 1, \ldots, K$

(20) $$\text{AV}_\lambda\left(R_{1,n+1} = k \mid (R_{ij})_{i,j \leq n}\right) = \frac{1}{n} \sum_{j=2}^{n} 1_{R_{1,j}=k} + \frac{1}{n}\text{AV}_\lambda\left(R_{1,2} = k\right).$$

The proof is not too difficult. For the Ruelle probability cascades, essentially everything can be computed explicitly.

In order to see the relation with the SK-model and Section 2it is convenient to view the above random measure as a measure on a countable subset of a separable Hilbert space H. This is always possible: Given a function q as (18), one can define a mapping

$$\rho : \mathbb{N}^K \to H$$

such that for $\mathbf{i} = (i_1, \ldots, i_K)$, $\mathbf{j} = (j_1, \ldots, j_K)$, one has $\|\rho(\mathbf{i})\|_H^2 = q_K$ and $\langle \rho(\mathbf{i}), \rho(\mathbf{j}) \rangle_H = q_{\tau(\mathbf{i},\mathbf{j})}$. Together with the random Gibbs weights from the Ruelle cascade, one then has a Markov kernel G from the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which the point processes are defined to the Hilbert space. The replicas are then simply independent random choices σ^i under $G(\omega, \cdot)$ and the overlaps R_{ij} are defined via the inner products $\langle \sigma^i, \sigma^j \rangle$. Under AV_λ, the distribution of the overlap is invariant under (finite) permutations of the indices. Furthermore, by the very construction from the Ruelle cascades, one has the **ultrametricity property**

$$\langle \sigma^1, \sigma^3 \rangle \geq \min\left(\langle \sigma^1, \sigma^2 \rangle, \langle \sigma^2, \sigma^3 \rangle\right),$$

so the ultrametricity property (8) is trivially satisfied in this case.

The big breakthrough by Panchenko is the deep result that the Ghirlanda-Guerra identity essentially characterizes the overlap structure. The main step is a proof that the identities imply ultrametricity, and ultrametricity together with Ghirlanda-Guerra implies that the distribution of the overlap matrix is coming from a Ruelle cascade, at least if the Parisi measure is finitely supported. This is the content of the next section.

We finish this section by describing the continuous time Markov process introduced in [7], which, in principle, avoids the necessity to work with finitely many levels.

The Markov process $\{\Gamma_t\}_{t \geq 0}$ takes values in the compact space E of partitionings of $\mathbb{N}$. Transitions are only allowed to coarser partitionings, i.e. for $s \leq t$, we have $\Gamma_s \prec \Gamma_t$, where for two partitionings Λ, Λ' of $\mathbb{N}$ into pairwise disjoint sets, $\Lambda \prec \Lambda'$ means that any set $A \in \Lambda$ is a subset of a set in Λ'. The process is described in terms of its traces on the finite subsets $I_N \stackrel{\text{def}}{=} \{1, \ldots, N\} \subset \mathbb{N}$. The finite set of partitionings of I_N is denoted by E_N. Of course, any partitioning Λ of $\mathbb{N}$ induces a partitioning of I_N which is denoted by $\pi_N(\Lambda)$. We define the process $\{\Gamma_t^N\}_{t \geq 0}$ for every N which in the end will satisfy $\Gamma_t^N = \pi_N(\Gamma_t)$. $\{\Gamma_t^N\}_{t \geq 0}$ is a time homogeneous Markov process on the finite set E_N. Of course, in general, there is no reason that a Markov process on E induces a Markov process on E_N, but in our special case, it is true. A Markov process on a finite space is described by its transition probabilities R_t^N, which is the matrix $\left(R_t^N(\Lambda, \Lambda')\right)_{\Lambda, \Lambda' \in E_N}$ given as

$$R_t^N = \exp\left[t A^N\right],$$

where A^N is the infinitesimal generator, i.e. a matrix with non-negative off-diagonal elements, and row sums 0. In our case, the description is very easy: If Λ has $n \geq 2$

classes and Λ' is obtained from Λ by clumping exactly $k \geq 2$ classes of Λ into one, then

$$A^N (\Lambda, \Lambda') = \left[(n-1) \binom{n-2}{k-2} \right]^{-1}.$$

All other matrix elements $A^N (\Lambda, \Lambda')$ with $\Lambda \neq \Lambda'$ are 0. Furthermore,

$$A^N (\Lambda, \Lambda) = - \sum_{\Lambda' : \Lambda' \neq \Lambda} A^N (\Lambda, \Lambda').$$

Let $\{\Gamma_t^N\}_{t \geq 0}$ be the Markov process on E_N which starts with the trivial partitioning of I_N into one-point subsets. There is a very simple description of this process which is based on the fact that also the total number of classes forms a Markov process. If at a time $t \geq 0$, Γ_t^N has $n \geq 2$ sets, then it stays there for an exponential time with expectation $(n-1)^{-1}$ after which it jumps to a partitioning by clumping a random number $\kappa \geq 2$ of sets with the probability $\frac{n}{n-1} \frac{1}{k(k-1)}$ that $\kappa = k$. Conditional that k sets are clumped, the choice which ones is done uniformly over all $\binom{n}{k}$ possibilities. Of course, if the process at time t has only one class left, then there is nothing to be clumped anymore, and the process stays there for ever.

These processes are compatible for different N: If $N' > N$, then the process $\{\Gamma_t^N\}_{t \geq 0}$ is the trace of $\{\Gamma_t^{N'}\}_{t \geq 0}$ on E_N. This implies that one can define the process $\{\Gamma_t\}_{t \geq 0}$ on E, and in fact, one can prove that it is a strong Feller process. The process has the property that at all times $t > 0$, Γ_t has infinitely many classes, and each of the classes is an infinite subset of $\mathbb{N}$.

Then the joint law of the (non-trivial) partitionings $\mathcal{P}_{K-1}, \ldots, \mathcal{P}_1$ coming from the Ruelle cascades is the law of the sequence $(\Gamma_{u_1}, \Gamma_{u_2}, \ldots, \Gamma_{u_{K-1}})$ with

$$\mathrm{e}^{-u_i} = \frac{\lambda_{K-i}}{\lambda_{K-i+1}}.$$

This means that the overlap structure is obtained by observing the Markov process $\{\Gamma_t\}$ at discrete time points.

In this formulation, there is no necessity to stick to finitely many levels, and at first sight, the approach using this Markov process appears more convenient. However, as the invariance properties like Proposition 4 are easier to use in the point process setting, most of the arguments in the literature, up to now, use the "finitely many level setting", and let $K \to \infty$ in the end.

4. THE GHIRLANDA-GUERRA IDENTITIES IMPLY ULTRAMETRICITY

We step back to the situation of Theorem 3 of a random probability distribution G on the separable (real) Hilbert space H, i.e. a Markov kernel from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to the unit ball of H. The infinite product $G^{\otimes \mathbb{N}}$ defines a Markov kernel from Ω to $H^{\mathbb{N}}$. Elements in $H^{\mathbb{N}}$ are denoted by $\sigma = (\sigma^1, \sigma^2, \dots)$. We also write $\sigma^{(n)}$ for $(\sigma^1, \dots, \sigma^n)$. The overlap matrix is defined by the inner products

$$R_{ij} = \left\langle \sigma^i, \sigma^j \right\rangle,$$

and we write $R^{(n)} \overset{\text{def}}{=} (R_{ij})_{i,j \leq n}$.

Consider then AV to be the measure $\int G(\omega, \cdot)^{\otimes \mathbb{N}} \mathbb{P}(d\omega)$ on $H^{\mathbb{N}}$. From the very definition of AV, one sees that the distribution of (R_{ij}) is the same as that of $(R_{\pi(i),\pi(j)})$ for any permutation π of finitely many elements. There will occasionally be the necessity to integrate a measurable function ϕ defined on $\Omega \times H^{\mathbb{N}}$:

$$\int \left(\int \phi(\omega, \sigma) G^{\otimes \mathbb{N}}(\omega, d\sigma) \right) \mathbb{P}(d\omega).$$

By an abuse of notation, we will also simply write $\int \phi \, d\text{AV}$ for this expression, and for a measurable subset $A \subset \Omega \times H^{\mathbb{N}}$ we write $\text{AV}(A)$. The reader should keep in mind that if an event or a function depends explicitly on ω, the AV-integrals have to be understood in this sense.

DEFINITION 6. — *We say that the pair $(\mathbb{P}, G)$ satisfies the Ghirlanda–Guerra identity, if for any $N \geq 2$, the conditioned law of $R_{1,N+1}$ given $R^{(N)} = (R_{ij})_{i,j \leq N}$ satisfies*

$$(21) \qquad \text{AV}\left(R_{1,N+1} \in A \,\middle|\, R^{(N)} \right) = \frac{1}{N} \sum_{j=2}^{N} 1_A(R_{1,j}) + \frac{1}{N} \text{AV}(R_{12} \in A).$$

THEOREM 7 (Panchenko). — *Assume (21). Then*

$$(22) \qquad \text{AV}\left(R_{1,2} \geq \min(R_{1,3}, R_{2,3}) \right) = 1.$$

THEOREM 8 (Panchenko). — *Assume (21). Then the distribution of R under* AV *is completely characterized by the Parisi measure*

$$\zeta \overset{\text{def}}{=} \text{AV}(R_{1,2} \in \cdot).$$

In the case where ζ is supported by a finite set, the AV*-law of R is that coming from the Ruelle cascades with this Parisi measure.*

REMARK 9. — *That ultrametricity, Ghirlanda-Guerra and a finite support imply that the distribution of R is that of a Ruelle cascade is not difficult, and was known before. The important step was the proof of the ultrametricity.*

PROPOSITION 10. — *If $(\mathbb{P}, G)$ satisfies the Ghirlanda-Guerra identities, then*

a) *(Talagrand's positivity result)*

$$\mathrm{AV}\,(R_{1,2} \geq 0) = 1.$$

b) *There exists a constant $q^* \leq 1$ such that*

$$\mathrm{AV}\,\left(\|\sigma_i\|_H^2 = q^*\right) = 1.$$

We will not prove this. See [20], Theorem 2.15 and Theorem 2.16.

b) says that the self-overlaps are all constant. The matrix $R^{(3)}$ therefore takes values in the set of symmetric positive semidefinite matrices with q^* on the diagonal. Denote by $\mathbf{S}^{(n)}$ the compact support of the distribution of $R^{(n)}$ under AV. The ultrametricity claim (22) is equivalent with the statement that if

$$\begin{pmatrix} q^* & a & b \\ a & q^* & c \\ b & c & q^* \end{pmatrix}$$

is in $\mathbf{S}^{(3)}$, then none of a, b, c is strictly smaller than the other two. Without loss of generality, we may assume that $a \leq b \leq c$. So, by a slight abuse of notation, we write $\mathbf{S}^{(3)}$ for the set of triples (a, b, c), $a \leq b \leq c$, such that the above matrix is in the support. Ultrametricity is violated if we find an $(a, b, c) \in \mathbf{S}^{(3)}$ with $a < b$. One easily sees that one needs only to consider $c < q^*$. If $(a, b, q^*) \in \mathbf{S}^{(3)}$ then it's evident that $a = b$. Therefore, the difficult task remains to prove

(23) $$(a, b, c) \in \mathbf{S}^{(3)}, \ a \leq b \leq c < q^* \Longrightarrow a = b.$$

Panchenko's proof of this crucial statement is very ingenious. It relies on a theorem stating an invariance property of AV under a class of transformations if the Ghirlanda-Guerra identities are satisfied. This invariance property implies a replication property of the support. We state here Panchenko's invariance property in a simplified version which is sufficient for the purpose:

Let $f : \mathbb{R} \to \mathbb{R}$ be a bounded measurable function, and Φ be a function of $R^{(n)}$. Let also

(24) $$F_f(\sigma) = \int \exp\left[f\left(\langle \sigma, \sigma' \rangle\right)\right] G(d\sigma').$$

One should pay attention to the fact that F also depends explicitly on $\omega \in \Omega$ through the Gibbs measure G. In order to stress this, we occasionally write $F(\omega, \sigma)$.

PROPOSITION 11. —

(25) $$\mathrm{AV}\,(\Phi) = \mathrm{AV}\left(\Phi \frac{\exp\left[\sum_{i=1}^{n-1} f(R_{i,n}) + \mathrm{AV}\,(f(R_{1,2}))\right]}{F(\sigma^n)^n}\right).$$

Proof. — We replace f by tf: Write $\varphi(t)$ for the right hand side of the above expression, and $\varphi^{(k)}(t)$ for the k-th derivative of φ at t. As $\varphi(0) = \mathrm{AV}\,(\Phi)$ it would suffice to prove that the first derivative is 0 for $t < 1$. It however seems to be difficult to prove this directly. The problem is that $\varphi(t)$ for $t > 0$ is not a function of $R^{(n)}$ because of the

denominator, and therefore, there seems to be no way to apply the Ghirlanda-Guerra identities directly to evaluate $\varphi'(t)$ for $t > 0$. It however turns out that $\varphi^{(k)}(0)$ can be expressed as an AV expectation of a function of $R^{(n+k)}$, and by an application of the identity, one gets $\varphi^{(k)}(0) = 0$ for all k. Together with some simple bounds for $\sup_{t \leq 1} |\varphi^{(k)}(t)|$, one gets $\varphi(1) = \varphi(0)$. The computation of the derivatives is not difficult, but a bit messy. Here is just the first one. Let

$$\psi\left(R^{(n)}\right) \stackrel{\text{def}}{=} \sum\nolimits_{i=1}^{n-1} f\left(R_{i,n}\right) + \text{AV}\left(f\left(R_{1,2}\right)\right),$$

$$F_t\left(\sigma\right) \stackrel{\text{def}}{=} \int \exp\left[tf\left(\langle \sigma, \sigma' \rangle\right)\right] G\left(d\sigma'\right).$$

$$
\begin{aligned}
\varphi'(t) &= \text{AV}\left(\Phi \frac{d}{dt} \frac{\exp\left[t\psi\left(R^{(n)}\right)\right]}{\left(F_t\left(\sigma^n\right)\right)^n}\right) \\
&= \text{AV}\left(\Phi\psi\left(R^{(n)}\right) \frac{\exp\left[t\psi\left(R^{(n)}\right)\right]}{\left(F_t\left(\sigma^n\right)\right)^n}\right) \\
&\quad - n\text{AV}\left(\Phi \frac{\exp\left[t\psi\left(R^{(n)}\right)\right]}{\left(F_t\left(\sigma^n\right)\right)^{n+1}} \int f\left(\langle \sigma^n, \sigma' \rangle\right) \exp\left[tf\left(\langle \sigma_n, \sigma' \rangle\right)\right] G\left(d\sigma'\right)\right).
\end{aligned}
$$

At $t = 0$, this gives by Ghirlanda-Guerra identity

$$
\begin{aligned}
&\text{AV}\left(\Phi\psi\left(R^{(n)}\right)\right) - n\text{AV}\left(\Phi f\left(R_{n,n+1}\right)\right) \\
&= \text{AV}\left(\Phi\psi\left(R^{(n)}\right)\right) - \sum_{j=1}^{n-1} \text{AV}\left(\Phi f\left(R_{j,n}\right)\right) - \text{AV}f\left(R_{1,2}\right)\text{AV}\left(\Phi\right) \\
&= 0.
\end{aligned}
$$

We skip the computation for the higher order derivatives (see [20] Theorem 2.18). □

Consider now a further bounded function $w : \mathbb{R} \to \mathbb{R}$ and define

$$S_w\left(\omega, \sigma\right) \stackrel{\text{def}}{=} \int w\left(\langle \sigma, \sigma' \rangle\right) G\left(\omega, d\sigma'\right)$$

$$T_w\left(\omega, \sigma\right) \stackrel{\text{def}}{=} \frac{\int w\left(\langle \sigma, \sigma' \rangle\right) \exp\left[f\left(\langle \sigma, \sigma' \rangle\right)\right] G\left(\omega, d\sigma'\right)}{F\left(\omega, \sigma\right)}.$$

Then, for any bounded measurable function $\psi : \mathbb{R} \to \mathbb{R}$

COROLLARY 12. — *For Φ as above and any bounded measurable function W, one has*

$$
\begin{aligned}
&\text{AV}\left(\Phi\left(R^{(n)}\right)\psi\left(S\left(\sigma^n\right)\right)\right) \\
&= \text{AV}\left(\Phi\left(R^{(n)}\right)\psi\left(T\left(\sigma^n\right)\right) \frac{\exp\left[\sum_{i=1}^{n-1} f\left(R_{i,n}\right) + \text{AV}\left(f\left(R_{1,2}\right)\right)\right]}{F\left(\sigma^n\right)^n}\right).
\end{aligned}
$$

Proof. — It suffices to prove the statement for polynomials ψ. So, we take $\psi(x) = x^k$.

$$\Phi\left(R^{(n)}\right) T\left(\sigma^n\right)^k = \frac{\Phi\left(R^{(n)}\right)}{F\left(\sigma^n\right)^k} \int \exp\left[\sum_{j=1}^{k} f\left(R_{n,n+j}\right)\right]$$

$$\times \prod_{j=1}^{k} w\left(R_{n,n+j}\right) G\left(\cdot, d\sigma^{n+j}\right).$$

Therefore, if we put

$$\Phi'\left(R^{(n+k)}\right) \overset{\text{def}}{=} \Phi\left(R^{(n)}\right) \prod_{j=1}^{k} w\left(R_{n,n+j}\right),$$

the right hand side of the claimed equation is

$$\mathrm{AV}\left(\Phi'\left(R^{(n+k)}\right) \frac{\exp\left[\sum_{i=1}^{n-1} f\left(R_{i,n}\right) + \sum_{i=n+1}^{n+k} f\left(R_{i,n}\right) + \mathrm{AV}\left(f\left(R_{1,2}\right)\right)\right]}{F\left(\sigma^n\right)^{n+k}}\right)$$

which, by the previous proposition, equals

$$\mathrm{AV}\left(\Phi\left(R^{(n)}\right) \prod_{j=1}^{k} w\left(R_{n,n+j}\right)\right) = \mathrm{AV}\left(\Phi\left(R^{(n)}\right) S\left(\sigma^n\right)^k\right).$$

$\square$

PROPOSITION 13. — *Assume that $(a, b, c) \in \mathbf{S}^{(3)}$, $c < q^*$, and that the Ghirlanda-Guerra identities (21) are satisfied. Then for every m, there exists a $3m \times 3m$ matrix $(r_{ij})_{i,j \leq 3m} \in \mathbf{S}^{(3m)}$ which satisfies*

$$r_{ij} \leq c, \ \forall i \neq j$$

$$r_{ij} = a, \ \text{for } 1 \leq i \leq m, \ m+1 \leq j \leq 2m,$$

$$r_{ij} = b, \ \text{for } 1 \leq i \leq m, \ 2m+1 \leq j \leq 3m,$$

$$r_{ij} = c, \ \text{for } m+1 \leq i \leq 2m, \ 2m+1 \leq j \leq 3m.$$

Proof that Proposition 13 implies (23). — Assume $(a, b, c) \in \mathbf{S}^{(3)}$, $a \leq b \leq c < q^*$. If $\sigma^i \in H$ satisfy $\langle \sigma^i, \sigma^j \rangle = r_{ij}$ and the r_{ij} satisfy the above properties, then

$$\langle \bar{\sigma}^1, \bar{\sigma}^2 \rangle = a, \ \langle \bar{\sigma}^1, \bar{\sigma}^3 \rangle = b, \ \langle \bar{\sigma}^2, \bar{\sigma}^3 \rangle = c,$$

where

$$\bar{\sigma}^i \overset{\text{def}}{=} \frac{1}{m} \sum_{j=m(i-1)+1}^{mi} \sigma^j,$$

and

$$\left\|\bar{\sigma}^1\right\|_H^2 = \frac{q^*}{m} + \frac{1}{m^2} \sum_{i \neq j,\ 1 \leq i,j \leq m} \langle \sigma^i, \sigma^j \rangle \leq \frac{q^* - c}{m} + c,$$

and the same estimate for $\left\|\bar{\sigma}^2\right\|_H^2$, $\left\|\bar{\sigma}^3\right\|_H^2$. Therefore,

$$\left\|\bar{\sigma}^2 - \bar{\sigma}^3\right\|_H^2 \leq \frac{2\,(q^* - c)}{m}.$$

On the other hand

$$\begin{aligned}
b - a &= \langle \bar{\sigma}^1, \bar{\sigma}^3 - \bar{\sigma}^2 \rangle \leq \left\|\bar{\sigma}^1\right\|_H \left\|\bar{\sigma}^2 - \bar{\sigma}^3\right\|_H \\
&\leq \frac{2\,(q^* - c)}{m}.
\end{aligned}$$

This implies $a \geq b$, i.e. $a = b$. $\qquad\square$

Proof of Proposition 13. — The replication property follows by induction of the following statement:

CLAIM 14. — *Let $A \in \mathbf{S}^{(n)}$ satisfy $a_n^* \overset{\text{def}}{=} \max(a_{1,n}, \dots, a_{n-1,n}) < q^*$. Then there exists an extension $A' \in \mathbf{S}^{(n+1)}$ of A such that $a_{i,n+1} = a_{i,n}$ for $i \leq n - 1$, and $a_{n,n+1} \leq a_n^*$. (Extension here means that A is the $n \times n$-matrix obtained from removing the last column and row of A'.)*

To prove the claim, we will prove that, for all $\varepsilon > 0$, one has

(26) $\quad \mathrm{AV}\left(R^{(n)} \in U_\varepsilon(A),\ |R_{i,n+1} - a_{i,n}| \leq \varepsilon,\ i \leq n - 1,\ R_{n,n+1} < a_n^* + \varepsilon\right) > 0,$

where $U_\varepsilon(A)$ denotes the componentwise ε-neighborhood of A, which evidently proves the claim. The argument is best done indirectly, by assuming that for some $\varepsilon > 0$

(27) $\quad \mathrm{AV}\left(R^{(n)} \in U_\varepsilon(A),\ |R_{i,n+1} - a_{i,n}| \leq \varepsilon,\ i \leq n - 1,\ R_{n,n+1} < a_n^* + \varepsilon\right) = 0$

Define

$$\Sigma\left(\sigma^{(n-1)}\right) \overset{\text{def}}{=} \left\{\sigma \in H : \left|\langle \sigma, \sigma^i \rangle - a_{i,n}\right| \leq \varepsilon,\ i \leq n - 1\right\},$$

and let A' be the matrix obtained from A by erasing the last row and the last column. A reformulation of (27) gives

$$\mathrm{AV}\left(R^{(n-1)} \in U_\varepsilon(A'),\ \sigma^n, \sigma^{n+1} \in \Sigma\left(\sigma^{(n-1)}\right),\ \langle \sigma^n, \sigma^{n+1} \rangle < a_n^* + \varepsilon\right) = 0.$$

We use Corollary 12 with $w(x) = 1_{x \geq a_n^* + \varepsilon}$, and $f = tw$ with $t \geq 0$ which will be chosen later. Then

$$S_w(\omega, \sigma^n) \overset{\text{def}}{=} G\left(\omega, \{\sigma : \langle \sigma, \sigma^n \rangle \geq a_n^* + \varepsilon\}\right).$$

It is easy to see that $A \in \mathbf{S}^{(n)}$ implies that given $\varepsilon > 0$, there exist $\delta > 0$, $0 < p < 1/2$, such that

(28) $\quad \mathrm{AV}\left(R^{(n)} \in U_\varepsilon(A),\ p \leq S(\sigma^n) \leq 1 - p\right) \geq \delta.$

F according to (24) is:

$$F(\omega, \sigma^n) = S(\omega, \sigma^n)\left(e^t - 1\right) + 1 \geq 1.$$

By Corollary 12, applied to $\psi(x) = 1_{p \leq x \leq 1-p}$, we get

$$\mathrm{AV}\left(\frac{\mathrm{e}^{\sum_{i=1}^{n-1} f(R_{i,n}) + \mathrm{AV}(f(R_{1,2}))}}{F(\sigma^n)^n} ; R^{(n)} \in U_\varepsilon(A), \ p \leq T(\sigma^n) \leq 1-p\right) \geq \delta.$$

Evidently, $f(R_{i,n}) = 0$ on $R^{(n)} \in U_\varepsilon(A)$, and

$$\mathrm{AV}(f(R_{1,2})) = t\mathrm{AV}(R_{1,2} \geq a_n^* + \varepsilon) = t\gamma,$$

where $\gamma < 1$. Therefore, as $F(\sigma^n) \geq 1$, we get

(29) $$\mathrm{AV}\left(R^{(n)} \in U_\varepsilon(A), \ T(\sigma^n) \leq 1-p\right) \geq \delta\mathrm{e}^{-\gamma t}.$$

An elementary computation gives that $T(\omega, \sigma^n) \leq 1-p$ implies

$$S(\omega, \sigma^n) \leq \frac{1-p}{p}\mathrm{e}^{-t}.$$

Consider now

$$\Lambda\left(\omega, \sigma^{(n-1)}\right) \overset{\text{def}}{=} \left\{\sigma^n \in H : \sigma^n \in \Sigma\left(\sigma^{(n-1)}\right), \ T(\omega, \sigma^n) \leq 1-p\right\},$$

and

$$\Pi \overset{\text{def}}{=} \left\{\left(\omega, \sigma^{(n-1)}\right) \in \Omega \times H^{n-1} : G\left(\omega, \Lambda\left(\omega, \sigma^{(n-1)}\right)\right) > 0\right\}.$$

Then

$$\mathrm{AV}\left(R^{(n)} \in U_\varepsilon(A), \ T(\sigma^n) \leq 1-p\right)$$
$$= \int \mathbb{P}(d\omega) \int_{U_\varepsilon(A')} G^{\otimes(n-1)}\left(\omega, d\sigma^{(n-1)}\right) G\left(\omega, \Lambda\left(\omega, \sigma^{(n-1)}\right)\right).$$

For any $\left(\omega, \sigma^{(n-1)}\right) \in \Pi$, we can choose a $\sigma' \in \Lambda\left(\omega, \sigma^{(n-1)}\right)$ for which one has $S(\omega, \sigma') \leq (1-p)\mathrm{e}^{-t}$. However, fixing this σ', almost all $\sigma^n \in \Sigma\left(\sigma^{(n-1)}\right)$ satisfy $\langle\sigma', \sigma^n\rangle \geq a_n^* + \varepsilon$ by (27), and therefore, for any $\left(\omega, \sigma^{(n-1)}\right) \in \Pi$, one has

$$G\left(\omega, \Lambda\left(\omega, \sigma^{(n-1)}\right)\right) \leq G\left(\omega, \Sigma\left(\sigma^{(n-1)}\right)\right) \leq S(\sigma') \leq \frac{1-p}{p}\mathrm{e}^{-t}.$$

Therefore

$$\mathrm{AV}\left(R^{(n)} \in U_\varepsilon(A), \ T(\sigma^n) \leq 1-p\right)$$
$$\leq \int \left(\mathbb{P} \otimes G^{\otimes(n-1)}\right)\left(d\left(\omega, \sigma^{(n-1)}\right)\right) 1_\Pi\left(\left(\omega, \sigma^{(n-1)}\right)\right) G\left(\omega, \Lambda\left(\omega, \sigma^{(n-1)}\right)\right)$$
$$\leq \frac{1-p}{p}\mathrm{e}^{-t}.$$

As the left hand side is $\geq \delta\mathrm{e}^{-\gamma t}$, by (29), it follows that

$$\delta \leq \frac{1-p}{p}\mathrm{e}^{\gamma t}\mathrm{e}^{-t},$$

which leads to a contradiction when choosing t large enough, as $\gamma < 1$. $\qquad\square$

5. ULTRAMETRICITY IMPLIES THE PARISI FORMULA

The first proof of the Parisi formula was given by Guerra [15] and Talagrand [24]. Guerra first proved the upper bound for $f(\beta)$, also including the case of an external field. A bit later, Talagrand then proved the lower bound. Guerra's proof of the upper bound relies on a very clever interpolation argument. It's easiest to formulate his argument in terms of the Random Overlap Structures introduced in [6]. We will not discuss it here, but just mention that Guerra proved that the finite N free energy can be bounded from above by the corresponding Parisi expression for any N. Guerra's proof depends on very special monotonicity properties of the SK-model. These monotonicity properties are also satisfied by p-spin models, but there are many other mean-field spin glasses where they are not satisfied, for instance the perceptron, the K-SAT, and even the bipartite SK-model. To this day, there is no proof of the upper bound without using such monotonicity properties. Panchenko's new proof of the lower bound does not replace Guerra's argument.

Although Guerra's proof, particularly if framed in the setup of the Aizenman-Sims-Starr concepts, strongly suggests that there must be ultrametricity behind the Parisi formula, Talagrand's technically very complicated proof of the lower bound bypasses ultrametricity altogether, and actually doesn't give a clue how to prove it.

Panchenko's proof of ultrametricity sketched in the last section gives a proof of ultrametricity for the SK- and related models, not quite in the form given in (7), but for a model with a slightly perturbed Hamiltonian. The reason is that the Ghirlanda-Guerra identities can be proved only for the SK-model with a perturbed Hamiltonian. The perturbation is however so small that it does not affect the free energy. Together with Guerra's upper bound for the free energy, this then gives a new proof of the lower bound, proving the Parisi formula. We give a rough outline of the chain of arguments. Unfortunately, it is still technically somewhat involved, and we cannot give all details.

First, we have to present the Parisi formula:

Consider two chains of parameters

$$0 < \lambda_1 < \cdots < \lambda_K < 1,$$

$$0 = q_0 < q_1 < \cdots < q_K = 1,$$

where $K \in \mathbb{N}$. As remarked before, this can be collected into the Parisi measure

$$\zeta \overset{\text{def}}{=} \sum_{k=0}^{K} (\lambda_{k+1} - \lambda_k)\, \delta_{q_k},$$

where $\lambda_0 = 0$ and $\lambda_{K+1} = 1$.

Consider also K independent standard Gaussian random variables $Z_1, \ldots, Z_K$. Then set

$$X_K = \log \cosh\left(\beta \sum_{k=1}^{K} Z_k \sqrt{q_k - q_{k-1}}\right) + \log 2.$$

Next define recursively downwards for $m = 0, \ldots, K - 1$

$$X_m = \frac{1}{\lambda_{m+1}} \log E_{Z_{m+1}} \exp\left[\lambda_{m+1} X_{m+1}\right],$$

where $E_{Z_{m+1}}$ refers to taking the expectation with respect to Z_{m+1}. X_0 is then a real number. (There is a slight complication in the case, the SK-model has an external field: In this case, X_0 will still be random.) The Parisi functional is given by

(30)
$$\mathcal{P}\left(\zeta\right) \stackrel{\text{def}}{=} X_0 - \frac{\beta^2}{4} \sum_{k=1}^{K} \lambda_k \left(q_k^2 - q_{k-1}^2\right).$$

The Parisi formula for the SK-model (with $h = 0$) is:

THEOREM 15 (Parisi, Guerra, Talagrand). —

(31)
$$f\left(\beta\right) = \inf_{\zeta} \mathcal{P}\left(\zeta\right).$$

The replica symmetric expression (10) is obtained by restricting the infimum to Parisi measures with $K = 2$, but it is believed (but mathematically not proved) that for $\beta > 1$, the infimum is *not* attained for a finite K.

The next observation is that the Parisi functional is closely connected with the Ruelle cascades described in Section 3.2.

Consider the cascades with parameters λ_i as above and independent standard Gaussian $Z_{\mathbf{i}^{(k)}}$ for $1 \leq k \leq K$, $\mathbf{i} \in \mathbb{N}^K$, where $\mathbf{i}^{(k)} \stackrel{\text{def}}{=} (i_1, \ldots, i_k)$. For an increasing function ψ, define

$$g_\psi\left(\mathbf{i}\right) \stackrel{\text{def}}{=} \sum_{k=1}^{K} Z_{\mathbf{i}^{(k)}} \left(\psi\left(q_k\right) - \psi\left(q_{k-1}\right)\right).$$

LEMMA 16. —

$$\mathcal{P}\left(\zeta\right) = \log 2 + \mathbb{E} \log \sum_{\mathbf{i}} \bar{X}_{\mathbf{i}} \cosh g_{\psi_1}\left(\mathbf{i}\right) - \mathbb{E} \log \sum_{\mathbf{i}} \bar{X}_{\mathbf{i}} \cosh g_{\psi_2}\left(\mathbf{i}\right),$$

where

$$\psi_1\left(x\right) \stackrel{\text{def}}{=} \beta^2 x, \quad \psi_2\left(x\right) \stackrel{\text{def}}{=} \frac{\beta^2 x^2}{2},$$

and where the "Gibbs weights" $\bar{X}_{\mathbf{i}}$ are given by (17).

The proof of the lemma is not too complicated, but needs an application of invariance properties, like 4. To get the connection of this expression of $\mathcal{P}\left(\zeta\right)$ with the SK-model, we have to discuss shortly the so-called cavity method in the version found by Aizenman, Sims and Starr [6]). From this computation, one will see that the free energy of the

SK-model is related to expressions which resemble the expression given in the above lemma. The idea is to relate the N system with the $(N+1)$-system.

$$Z_{N+1} = \sum_{\sigma \in \{-1,1\}^{N+1}} \exp\left[\frac{\beta}{\sqrt{N+1}} \sum_{i<j\leq N+1} g_{ij}\sigma_i\sigma_j\right]$$

$$= \sum_{\sigma \in \{-1,1\}^{N+1}} \exp\left[\frac{\beta}{\sqrt{N+1}} \sum_{i<j\leq N} g_{ij}\sigma_i\sigma_j + \frac{\beta}{\sqrt{N+1}}\sigma_{N+1}\sum_{i=1}^{N} g_{i,N+1}\sigma_i\right]$$

$$= \sum_{\sigma \in \{-1,1\}^{N}} \exp\left[\frac{\beta}{\sqrt{N+1}} \sum_{i<j\leq N} g_{ij}\sigma_i\sigma_j\right] 2\cosh\left(\frac{\beta}{\sqrt{N+1}}\sum_{i=1}^{N} g_{i,N+1}\sigma_i\right)$$

$$= 2\left\langle\cosh\left(\beta z_N(\sigma)\right)\right\rangle',$$

where

$$(32) \qquad z_N(\sigma) \stackrel{\text{def}}{=} \frac{1}{\sqrt{N+1}}\sum_{i=1}^{N} g_{i,N+1}\sigma_i,$$

and $\langle\cdot\rangle'$ refers to taking the Gibbs expectation on the N system, but with the Hamiltonian

$$H_N' \stackrel{\text{def}}{=} -\frac{1}{\sqrt{N+1}} \sum_{1\leq i<j\leq N} g_{ij}\sigma_i\sigma_j.$$

We can also write Z_N in terms of this Hamiltonian, as

$$\frac{1}{\sqrt{N}} \sum_{1\leq i<j\leq N} g_{ij}\sigma_i\sigma_j \stackrel{\mathcal{L}}{=} \frac{1}{\sqrt{N+1}} \sum_{1\leq i<j\leq N} g_{ij}\sigma_i\sigma_j + y_N(\sigma)$$

with

$$(33) \qquad y_N(\sigma) \stackrel{\text{def}}{=} \frac{1}{\sqrt{N(N+1)}} \sum_{1\leq i<j\leq N} g_{ij}'\sigma_i\sigma_j$$

with some new independent standard Gaussians g_{ij}', independent also of the g_{ij}. $\stackrel{\mathcal{L}}{=}$ means equality in law. Therefore if we set $F_N \stackrel{\text{def}}{=} N^{-1}\mathbb{E}\log Z_N$, we get

$$\mathbb{E}\log Z_N = \mathbb{E}\log\langle y_N(\sigma)\rangle + \mathbb{E}\log Z_{N,\beta}',$$

$$\mathbb{E}\log Z_{N+1} = \mathbb{E}\log\langle 2z_N(\sigma)\rangle + \mathbb{E}\log Z_{N,\beta}',$$

and therefore

$$(34) \qquad A_N = \mathbb{E}\log Z_{N+1} - \mathbb{E}\log Z_N$$

$$= \mathbb{E}\log\left\langle 2\cosh\left(\beta z_N(\sigma)\right)\right\rangle' - \mathbb{E}\log\left\langle\exp\left[\beta y_N(\sigma)\right]\right\rangle',$$

$$F_N = \frac{1}{N}\sum_{j=0}^{N-1} A_j,$$

and

(35) $$f(\beta) = \lim_{N \to \infty} F_N \geq \liminf_{N \to \infty} A_N.$$

This is the basis for getting a lower bound by relating (34) to the Parisi expression (30). The existence of the first limit was proved first in [16]. Of course, if one could prove the existence of $\lim_{N \to \infty} A_n$, and identify it with the Parisi expression, one would have finished the proof of the Parisi formula. This seems not to be possible. However, the upper bound for f is known by the Guerra's interpolation technique.

The task which remains is to prove that the right-hand side of (35) is bounded from below by the Parisi expression. This is based on the fact, which was known since long, see [14], [3], that a slightly perturbed SK-model satisfies the Ghirlanda-Guerra identities. This perturbation is technically a bit awkward, and we cannot explain it here in details. Essentially one considers a Hamiltonian

$$H_N^{\mathrm{pert}}(\sigma) = H_N(\sigma) + \Pi_N(\sigma)$$

where

$$\Pi_N(\sigma) \overset{\mathrm{def}}{=} \sum_{p=1}^{\infty} 2^{-p} x_p N^{-p/2} \sum_{j_1,\ldots,j_p=1}^{N} g'_{i_1,\ldots,i_p} \sigma_{i_1} \cdots \sigma_{i_p},$$

where the $g'_{i_1,\ldots,i_p}$ are standard Gaussians, independent of the Gaussians in the main part H_N, and where still the parameters $x_p \in [1, 2]$ have to be chosen carefully in order to make the following arguments work. The main point is that $\mathrm{var}(\Pi_N(\sigma))$ is of order 1, due to the scaling by $N^{-p/2}$. It is not difficult to see that this implies that the free energy is not affected in the $N \to \infty$ limit by the perturbation. The perturbation, when done carefully, actually even with an averaging over the additional parameters x_p, implies that the distribution of the overlaps satisfies approximately the Ghirlanda-Guerra identities, with an error which goes to 0 as $N \to \infty$, and also, as proved by Talagrand, that the overlaps are positive, with large probability. The Ghirlanda-Guerra identities in this framework were first proved in [14], [3], and variants were proved by many other authors. See also the book by Talagrand [25].

It then follows that any weak limit along a subsequence of the distribution of the overlaps

$$R_{i,j}^N \overset{\mathrm{def}}{=} \frac{1}{N} \sum_{k=1}^{N} \sigma_k^i \sigma_k^j$$

under $\mathrm{AV}^{\mathrm{pert}} \overset{\mathrm{def}}{=} \int G_{N,\beta}^{\mathrm{pert} \otimes \mathbb{N}} d\mathbb{P}$, where $\mathbb{P}$ governs the laws of the Gaussian interaction variables, has the exchangeability property, which, by the Dovbysh-Sudakov theorem, implies the representation as the law of the inner products in a Hilbert space H, under the law given by a random measure G^{DS} on H, with an abstract probability space behind, which we denoted by $(\Omega, \mathcal{F}, \mathbb{P}^{\mathrm{DS}})$. The fact that R^N under $\mathrm{AV}^{\mathrm{pert}}$ satisfies Ghirlanda-Guerra implies that $(G^{\mathrm{DS}}, \mathbb{P}^{\mathrm{DS}})$ satisfies Ghirlanda-Guerra exactly, and therefore, by the Theorems 7 and 8, is ultrametric, and determined by the Parisi measure $\nu(dx) = \mathrm{AV}^{\mathrm{DS}}(R_{1,2} \in dx)$. In case ν has finite support, the distribution is that

coming from the Ruelle cascade with this Parisi measure. As we don't know anything about convergence along the original sequence, there can be arbitrarily many of such Dovbysh-Sudakov pairs, and Parisi measures.

It's now tempting to apply this to the evaluation of A_N for large N. We first choose a subsequence, such that $A_{N_k} \to^{k \to \infty} \liminf_{N \to \infty} A_N$. Then we choose a further subsequence of $\{A_{N_k}\}$ such that the law of R^N along this subsequence converges to the law of the inner products under $\mathrm{AV}^{\mathrm{DS}}$. We would like to conclude that $\liminf_{N \to \infty} A_N$ can be expressed through $\mathrm{AV}^{\mathrm{DS}}$. This is not quite obvious as the A_N cannot directly be expressed through the overlaps. It requires an additional approximation argument. The basic observation is that the covariances of $y_N(\sigma)$, and $z_N(\sigma)$ given in (33) and (32) are expressed in terms of the overlaps of σ. This implies that A_N can be approximated by functions of finite restrictions of the overlap matrix. In case the Parisi measure ν has finite support, one can conclude in this way, using Lemma 16 that

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

In the case of a general Parisi measure, appearing through the limits along the subsequences defined above, an additional approximation is needed, but one concludes that

$$\liminf_{N \to \infty} A_N \geq \inf_\xi \mathcal{P}(\xi).$$

Together with Guerra's upper bound, this finally proves the Parisi formula.

REFERENCES

[1] Aidekon, E., Berestycki, J., Brunet, E., and Shi, Z.: *Branching Brownian motion seen from its tip.* Prob. Th. Rel. Fields **157** (2013), 405-451.

[2] Aizenman, M., Lebowitz, J., and Ruelle, D.: *Some rigorous results on the Sherrington-Kirkpatrick model.* Comm. Math. Phys. **112** (1987), 3-20.

[3] Aizenman, M., and Contucci, P.: *On the stability of the quenched state in mean-field spin glasses.* J. Stat. Phys. **92** (1998), 765-783.

[4] Arguin, L.-P., and Aizenman, M.: *On the structure of quasi-stationary competing particle systems.* Ann. Prob. **37** (2009), 1080-1113.

[5] Arguin, L.-P., Bovier, A., and Kistler, N.: *The extremal process of branching Brownian motion.* Prob. Th. Rel. Fields **157** (2013), 535-574.

[6] Aizenman, M., Sims, R., and Starr, S.L.: *An extended variational principle for the SK spin-glass model.* Phys. Rev. B **68** (2003), 214-403.

[7] Bolthausen, E., and Sznitman, A.-S.: *On Ruelle's probability cascades and an abstract cavity method.* Comm. Math. Phys. **197** (1998), 247-276.

[8] Bovier, A., and Kurkova, I.: *Derrida's generalized random energy models.* Ann. Inst. H. Poincaré Probab. Stat. **40** (2004), 439-480.

[9] Bramson, M., and Zeitouni, O.: *Tightness of the recentered maximum of the two–dimensional discrete Gaussian Free Field.* Comm. Pure Appl. Math. **65** (2011), 1–20.

[10] Derrida, B.: *Random energy model: an exactly solvable model of disordered systems.* Phys. Rev. B **24** (1981), 2613-2626.

[11] Derrida, B.: *A generalization of the random energy model that includes correlations between the energies.* J. Phys. Lett. **46** (1985), 401-407.

[12] Dovbysh, L.N., Sudakov, V.N.: *Gram-de Finetti matrices.* Zap. Nauchn. Sem. Leningrad. Otdel. Mat. Inst. Steklov **119** (1982), 77-86.

[13] Fröhlich, J., and Zegarlinski, B.: *Some comments on the Sherrington-Kirkpatrick model of spin glasses.* Comm. Math. Phys. **112** (1987), 553-566.

[14] Ghirlanda, S., and Guerra, F.: *General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity.* J. Phys. A **31** (1998), 9149-9155.

[15] Guerra, F.: *Broken replica symmetry bound in the mean field spin glass model.* Comm. Math. Phys. **233** (2003), 1-12.

[16] Guerra, F., and Toninelli, F.L.: *The thermodynamic limit in mean field spin glass models.* Comm. Math. Phys. **230** (2002), 71-79.

[17] Ledoux, M.: *Isoperimetry and Gaussian Analysis.* Springer LN in Math. **1648** (1996).

[18] Mézard, M., Parisi, G., and Virasoro, M.: *Spin glass theory and beyond.* World Scientific, Singapore, 1987.

[19] Panchenko, D.: *The Parisi ultrametric conjecture.* Annals of Math. **177** (2013), 383-393.

[20] Panchenko, D.: *The Sherrington-Kirkpatrick model.* Springer, New York, 2013.

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

[22] Ruelle, D.: *A mathematical reformulation of Derrida's REM and GREM.* Comm. Math. Phys. **108** (1987), 225-239.

[23] Sherrington, D., and Kirkpatrick, S.: *Solvable model of a spin glass.* Phys. Rev. Lett. **35** (1972), 1992-1996.

[24] Talagrand, M.: *The Parisi formula.* Annals of Math. **163** (2006), 221-263.

[25] Talagrand, M.: *Mean-field models for spin glasses.* Springer, New York, 2011.

Erwin BOLTHAUSEN

Institut für Mathematik
Universität Zürich
Winterthurerstrasse 190
CH-8057 Zürich (Suisse)
E-mail : eb@math.uzh.ch