Derrida's Generalized Random Energy models 2: Gibbs measures and probability cascades.

Anton Bovier[1 2]

Weierstraß–Institut
für Angewandte Analysis und Stochastik
Mohrenstrasse 39, D-10117 Berlin, Germany

Irina Kurkova[3]

Laboratoire de Probabilités et Modèles Aléatoires
Université Paris 6
4, place Jussieu, B.C. 188
75252 Paris, Cedex 5, France

Abstract: This is the second in a series of three papers in which we present a full rigorous analysis of Derrida's Generalized Random Energy Models (GREM). In this paper we still consider only models with finitely many levels. In this context we present two ways to prove the convergence of the Gibbs measures (in a suitable representation). The first is direct and shows convergence to the probability cascades introduced by Ruelle. The second approach is indirect and use the so-called Ghirlanda-Guerra identities, that allow to control the distribution of the Gibbs measures via recursive identities for their moments.

Keywords: Gaussian processes, generalized random energy model, spin glasses, Poisson cascades, probability cascades, Ghirlanda-Guerra identities

AMS Subject Classification: 82B44, 60G70,60K35

[1]e-mail: bovier@wias-berlin.de

[2]Research supported in part by the DFG in the Concentration program "Interacting stochastic systems of high complexity".

[3]e-mail: kourkova@ccr.jussieu.fr

1. Introduction.

In 1980, Derrida proposed two simplified models for mean field spin glasses: the *random energy model* (REM)[D1,D2], and the *generalized random energy model* (GREM)[DG1,DG1,DG2,DG3]. The former consisted of modelling the random energy landscape as simply i.i.d. Gaussian random variables on the set of spin configurations, $\{-1,1\}^N$. This is the second of three papers in which we present a full analysis of these models in the light of recent developments. We refer to the first of these papers [BK1] for more background.

In the present paper we continue the analyis of the subclass of models considered in [BK1]. While in the first paper, we constructed a number of relevant point processes and identified the extremal process as a cluster point process obtained from a Poisson cascade process [Ru], 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 [Ru]. More or less as a corollary, we will then also prove that certain point processes that characterise the mass distribution of the Gibbs measure, converge to the probability cascade processes of Ruelle. We will then show that the so-called Ghirlanda-Guerra identities [GG,AC] hold in these models, and that they suffice to completely characterise the probability cascade processes up to the mean of the two-replica distribution function. Let us note that this fact was first observed in the REM by M. Talagrand [T2,T5] and exploited in the analysis of the Gibbs measures of the p-spin SK models [T3,T4,T5].

This opens an alternative way to the construction of the thermodynamic limit that will bear its fruits fully in the final paper of this series where models with continuous hierarchies will be considered.

1.1. Definition of the models. We will consider Gaussian processes X_σ indexed by the hypercube $\mathcal{S}_N \equiv \{-1,1\}^N$. Let us equip the hypercube $\mathcal{S}_N$ with the natural ultra-metric valuation

$$d_N(\sigma,\sigma') \equiv \frac{1}{N}\left(\min\left(i|\sigma_i \neq \sigma'_i\right) - 1\right) \tag{1.1}$$

(Note that $\exp(-d_N(\sigma,\sigma'))$ is a ultra-metric on $\mathcal{S}_N$; by an abuse of language we will refer often to d_N as 'distance'). We will consider processes whose covariance is given by a (non-decreasing) function of d_N, i.e. we assume that $\mathbb{E}X_\sigma = 0$, and

$$\operatorname{cov}\left(X_\sigma, X_{\sigma'}\right) = \mathbb{E}X_\sigma X_{\sigma'} = A(d_N(\sigma,\sigma')) \tag{1.2}$$

where $A(x)$ is a probability distribution function on $[0,1]$. Note that the mean field spin glasses of the SK-family [SK] can be constructed in an analogous way, except that $d_N(\sigma,\sigma')$ is replaced by the overlap $R_N(\sigma,\sigma') \equiv N^{-1}\sum_{i=1}^N \sigma_i\sigma_i'$.

The standard GREM with *finitely* many hierarchies corresponds to the special case when A is the distribution function of a measure that is supported on a finite number, n, of points $x_1,\dots,x_n \in [0,1]$. In that case we denote the mass of the atoms x_i by a_i, and we set

$$\ln\alpha_i = (x_i - x_{i-1})\ln 2, \quad i = 1,\dots,n \tag{1.3}$$

where $x_0 \equiv 0$. Of course $\sum_{i=1}^n a_i = 1$, and $\prod_{i=1}^n \alpha_i = 2$. The hypercube $\mathcal{S}_N$ can then be considered as a n-fold product $\mathcal{S}_N = \prod_{i=1}^n \mathcal{S}_{N\ln\alpha_i/\ln 2}$. We will 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 represented in the following way. Take $\alpha_1^N + (\alpha_1\alpha_2)^N + \dots + (\alpha_1\alpha_2\cdots\alpha_n)^N$ independent standard Gaussian random variables enumerated as follows: The first α_1^N of them X_{σ_1} are enumerated by α_1^N configurations of spins $\sigma_1 \in \{-1,+1\}^{N\ln\alpha_1/\ln 2}$, the next $(\alpha_1\alpha_2)^N$ of them $X_{\sigma_1\sigma_2}$ — by configurations of spins $\sigma_1\sigma_2 \in \{-1,+1\}^{N\ln(\alpha_1\alpha_2)/\ln 2}$, etc ..., finally the last $(\alpha_1\alpha_2\cdots\alpha_n)^N = 2^N$ of them $X_{\sigma_1\sigma_2,\dots,\sigma_n}$ are enumerated by configurations $\sigma_1\sigma_2\dots\sigma_n \in \{-1,+1\}^N$. Then the process X_σ has the same distribution as

$$X_\sigma \equiv \sqrt{a_1}X_{\sigma_1} + \sqrt{a_2}X_{\sigma_1\sigma_2} + \dots + \sqrt{a_n}X_{\sigma_1\sigma_2\dots\sigma_n}, \quad \text{if } \sigma = \sigma_1\sigma_2\dots\sigma_n, \tag{1.4}$$

The partition function of the GREM at the inverse temperature β is defined as

$$\begin{aligned} Z_{N,\beta} &= 2^{-N}\sum_{\sigma\in\mathcal{S}_N} e^{\beta\sqrt{N}X_\sigma} \\ &= 2^{-N}\sum_{\sigma_1\in\mathcal{S}_{N\ln\alpha_1/\ln 2}} e^{\beta\sqrt{a_1N}X_{\sigma_1}} \sum_{\sigma_2\in\mathcal{S}_{N\ln\alpha_2/\ln 2}} e^{\beta\sqrt{a_2N}X_{\sigma_1\sigma_2}} \dots \sum_{\sigma_n\in\mathcal{S}_{N\ln\alpha_n/\ln 2}} e^{\beta\sqrt{a_nN}X_{\sigma_1\sigma_2\dots\sigma_n}}. \end{aligned} \tag{1.5}$$

The Gibbs measure $\mu_{\beta,N}$ is defined as

$$\mu_{\beta,N}(\sigma) \equiv \frac{e^{\beta\sqrt{N}X_\sigma}}{Z_{\beta,N}} \tag{1.6}$$

To be able to state our main results, we must recall some notation and results from [BK1].

Recall the function $u_{\ln\alpha,N}(x)$, $x\in\mathbb{R}$ given by

$$u_{\ln\alpha,N}(x) = \sqrt{2\ln\alpha N} + \frac{x}{\sqrt{2\ln\alpha N}} - \frac{\ln N + \ln\ln\alpha + \ln 4\pi}{2\sqrt{2\ln\alpha N}}. \tag{1.7}$$

The main result from [BK1] that we will need in the sequal is

Theorem 1.1: *Let $\alpha_i \geq 1$, $0 < a_i < 1$, $i = 1,2,\ldots,n$, $\prod_{i=1}^n \alpha_i = 2$, $\sum_{i=1}^n a_i = 1$. Let $J_1, J_2, \ldots, J_m \in \mathbb{N}$ be the indices such that $0 = J_0 < J_1 < J_2 < \cdots < J_m = n$. Denote by $\bar{a}_l \equiv \sum_{i=J_{l-1}+1}^{J_l} a_i$, $\bar{\alpha}_l \equiv \prod_{i=J_{l-1}+1}^{J_l} \alpha_i$, $l = 1,2,\ldots,m$, and introduce standard Gaussian random variables*

$$\bar{X}^{\sigma_1\ldots\sigma_{J_{l-1}}}_{\sigma_{J_{l-1}+1}\sigma_{J_{l-1}+2}\cdots\sigma_{J_l}} \equiv \Big(\sqrt{a_{J_{l-1}+1}} X_{\sigma_{J_1}\ldots\sigma_{J_{l-1}}\sigma_{J_{l-1}+1}} + \sqrt{a_{J_{l-1}+2}} X_{\sigma_{J_1}\ldots\sigma_{J_{l-1}}\sigma_{J_{l-1}+1}\sigma_{J_{l-1}+2}} + \cdots + \sqrt{a_{J_l}} X_{\sigma_{J_1}\ldots\sigma_{J_{l-1}}\sigma_{J_{l-1}+1}\ldots\sigma_{J_l}}\Big)\Big/\sqrt{\bar{a}_l}. \tag{1.8}$$

Assume that the partition $J_1, J_2, \ldots, J_m$ satisfies the following condition: for all $l = 1,2,\ldots,m$ and all k such that $J_{l-1}+2 \leq k \leq J_l$

$$(a_k + a_{k+1} \cdots + a_{J_l-1} + a_{J_l})/\bar{a}_l \geq \ln(\alpha_k \alpha_{k+1} \cdots \alpha_{J_l-1}\alpha_{J_l})/\ln(\bar{\alpha}_l). \tag{1.9}$$

Then the point process

$$\mathcal{P}_N^{(m)} \equiv \sum_{\sigma_1\ldots\sigma_{J_1}} \delta_{u^{-1}_{\ln\bar{\alpha}_1,N}(\bar{X}_{\sigma_1\ldots\sigma_{J_1}})} \sum_{\sigma_{J_1+1}\ldots\sigma_{J_2}} \delta_{u^{-1}_{\ln\bar{\alpha}_2,N}(\bar{X}^{\sigma_1\cdots\sigma_{J_1}}_{\sigma_{J_1+1}\cdots\sigma_{J_2}})} \cdots \sum_{\sigma_{J_{m-1}+1}\ldots\sigma_{J_m}} \delta_{u^{-1}_{\ln\bar{\alpha}_m,N}(\bar{X}^{\sigma_1\cdots\sigma_{J_{m-1}}}_{\sigma_{J_{m-1}+1}\cdots\sigma_{J_m}})}$$

converges weakly to the process $\mathcal{P}^{(m)}$ on $\mathbb{R}^m$ which is characterised by the following generating functions:

$$\begin{aligned} F_{\Delta_1\times\cdots\times\Delta_k}(z) &\equiv \mathbb{E} z^{\sum_{x_1}\mathbb{1}_{\{x_1\in\Delta_1\}}\cdots\sum_{x_k}\mathbb{1}_{\{x_k\in\Delta_k\}}} \\ &= f_{1,\Delta_1}(f_{2,\Delta_2}(f_{3,\Delta_3}\cdots(f_{k-1,\Delta_{k-1}}(f_{k,\Delta_k}(z)))\cdots)), \qquad |z|<1 \end{aligned} \tag{1.10}$$

where $f_{i,\Delta_i}(z) = e^{K_i(z-1)(e^{-a_i}-e^{-b_i})}$, $\Delta_i = (a_i, b_i]$ with $a_i, b_i \in \mathbb{R}$ or $b_i = \infty$, $i = 1,2,\ldots,k$. with constants $0 < K_1, \ldots, K_m \leq 1$.

Remark: The values of the constants K_l are given in [BK1]. Here we just note that if all $J_l - J_{l-1} - 1$ inequalities in (1.9) for $k = J_{l-1}+2, \ldots, J_l$ are strict, $K_l = 1$.

Remark: The process $\mathcal{P}^{(m)}$ is called a *Poisson cascade* with m levels. It is best understood in terms of the following iterative construction. If $m = 1$, 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 m times. Clearly, the projection of $\mathcal{P}^{(m)}$ on the plane of the first k coordinates is distributed as $\mathcal{P}^{(k)}$, $k = 1, \dots, m-1$.

Remark: Given a_i, α_i, $i = 1, \dots, n$, the partition of indices $J_1, \dots, J_m$ satisfying the condition (1.9) is generally not unique. Therefore this theorem yields a *family* of convergent point processes associated with the GREM.

1.2. Convergence of the partition function.

We now turn to the new results of this paper. The first and technically the key result, will relate the fluctuations of the partition function in the thermodynamic limit to the Poisson cascade processes obtained in the preceding theorem. This is the main step needed to link the GREM to the models constructed by Ruelle [Ru].

For *any* sequence of indices $0 < J_1 < \dots < J_m = n$ the partition function of the GREM can be written as:

$$\begin{aligned} Z_{\beta,N} = & e^{\sum_{j=1}^{m} \left(\beta N \sqrt{2\bar{a}_j \ln \bar{\alpha}_j} - \beta\gamma_j [\ln(N \ln \bar{\alpha}_j) + \ln 4\pi]/2 - N \ln \alpha_j\right)} \\ & \times \sum_{\sigma_1 \dots \sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln \bar{\alpha}_1, N}(\bar{X}_{\sigma_1 \dots \sigma_{J_1}})} \dots \sum_{\sigma_{J_{m-1}+1} \dots \sigma_{J_m}} e^{\beta\gamma_m u^{-1}_{\ln \bar{\alpha}_m, N}(\bar{X}^{\sigma_1 \dots \sigma_{J_{m-1}}}_{\sigma_{J_{m-1}+1} \dots \sigma_{J_m}})} \end{aligned} \tag{1.11}$$

where $\bar{a}_l \equiv \sum_{i=J_{l-1}+1}^{J_l} a_i$, $\bar{\alpha}_l \equiv \prod_{i=J_{l-1}+1}^{J_l} \alpha_i$, $\gamma_l \equiv \sqrt{\bar{a}_l}/\sqrt{2 \ln \bar{\alpha}_l}$, $l = 1, 2, \dots, m$, and the random variables $\bar{X}^{\sigma_1 \dots \sigma_{J_{l-1}}}_{\sigma_{J_{l-1}+1} \dots \sigma_{J_l}}$ are defined in (1.8). Moreover, for any sequence $J_1, \dots, J_m$ satisfying the assumptions of Theorem 1.1, the corresponding point process converges. But, of course, at most one of these sequences can provide through (1.11) the right scaling of the fluctuations of the partition function. As in the analysis of the extremal process in [BK1], it will be the process corresponding to a maximal spacing of the integers J_i satisfying the conditions (1.9) that gives the relevant contribution. Recall that this partition is achieved by setting $J_0 \equiv 0$, and

$$J_l \equiv \min\{J > J_{l-1} : A_{J_{l-1}+1, J} > A_{J+1, k} \ \forall k \geq J+1\}. \tag{1.12}$$

with $A_{j,k} \equiv \sum_{i=j}^{k} a_i / (2 \ln(\prod_{i=j}^{k} \alpha_i))$. The sequence $J_0, J_1, \dots, J_m$ of (1.12) satisfies (1.9) by Proposition 1.4 from [BK1]. However, unlike in the case of the extremal process, it must be cut at some temperature dependent level, $J_{l(\beta)}$. Clearly, $A_{1,J_1} > A_{J_1+1,J_2} > \dots > A_{J_{m-1}+1,J_m}$. Let us put

$$l(\beta) \equiv \max\{l \geq 1 : \beta^2 A_{J_{l-1}+1, J_l} > 1\} \tag{1.13}$$

and $l(\beta) \equiv 0$ if $\beta^2 A_{1,J_1} \leq 1$. According to the notations of (1.11) $A_{J_{l-1}+1,J_l} = \gamma_l^2$ and (1.13) implies that $\beta\gamma_{l(\beta)+1} \leq 1$.

In [CCP] the limit of the free energy was found in terms of (1.12) and (1.13):

Theorem 1.2: [CCP] *With the notation introduced above,*

$$\begin{aligned}&\lim_{N\to\infty} N^{-1} \ln Z_{N,\beta} \equiv -\lim_{N\to\infty} F_{\beta,N} \\ =&\beta\left(\sqrt{2\bar{a}_1 \ln \bar{\alpha}_1} + \cdots + \sqrt{2\bar{a}_{l(\beta)} \ln \bar{\alpha}_{l(\beta)}}\right) - \ln(\bar{\alpha}_1 \cdots \bar{\alpha}_{l(\beta)}) + \sum_{i=J_{l(\beta)}+1}^{n} \beta^2 a_i/2, \quad a.s.\end{aligned} \tag{1.14}$$

and also in L^p, $1 \leq p < \infty$.

The following theorem yields the fluctuations of the partition function and of the free energy.

Theorem 1.3: *Let $\alpha_i \geq 1$, $0 < a_i < 1$, $i = 1, 2, \ldots, n$, $\prod_{i=1}^n \alpha_i = 2$, $\sum_{i=1}^n a_i = 1$. Let $J_1, J_2, \ldots, J_m \in \mathbb{N}$, be the sequence of indices defined by (1.12) and $l(\beta)$ defined by (1.13). Denote by $\bar{a}_l \equiv \sum_{i=J_{l-1}+1}^{J_l} a_i$, $\bar{\alpha}_l \equiv \prod_{i=J_{l-1}+1}^{J_l} \alpha_i$, $\gamma_l = \sqrt{\bar{a}_l}/\sqrt{2\ln\bar{\alpha}_l}$, $l = 1, 2, \ldots, m$.*

Then the properly normalised partition function of the GREM converges in law to the following integral on $\mathbb{R}^m$:

$$\begin{aligned}&e^{\sum_{j=1}^{l(\beta)} \left(-\beta N\sqrt{2\bar{a}_j \ln \bar{\alpha}_j} + \beta\gamma_j[\ln(N \ln \bar{\alpha}_j) + \ln 4\pi]/2 + N \ln \alpha_j\right) - N\sum_{i=J_{l(\beta)+1}}^{n} \beta^2 a_i/2} Z_{\beta,N} \\ &\overset{\mathcal{D}}{\to} C(\beta) \int_{\mathbb{R}^{l(\beta)}} e^{\beta\gamma_1 x_1 + \beta\gamma_2 x_2 + \cdots + \beta\gamma_{l(\beta)} x_{l(\beta)}} \mathcal{P}^{(l(\beta))}(dx_1 \ldots dx_{l(\beta)}).\end{aligned} \tag{1.15}$$

This integral is over the process $\mathcal{P}^{(l(\beta))}$ on $\mathbb{R}^{l(\beta)}$ from Theorem 1.1 with constants K_j from Theorem 1.1. The constant $C(\beta)$ satisfies

$$C(\beta) = 1, \quad \textit{if} \quad \beta\gamma_{l(\beta)+1} < 1, \tag{1.16}$$

and

$$C(\beta) = P\Bigg(\bigcap_{\substack{i:\, J_{l(\beta)}+1 \leq i \leq J_{l(\beta)+1} \\ (a_{J_{l(\beta)}+1} + \cdots + a_i)/\bar{a}_{J_{l(\beta)+1}} = \ln(\alpha_{J_{l(\beta)}+1} \cdots \alpha_i)/\ln\bar{\alpha}_{J_{l(\beta)+1}}}} \left(\sqrt{a_{J_{l(\beta)+1}}} Z_{J_{l(\beta)}+1} + \cdots + \sqrt{a_i} Z_i < 0\right) \Bigg) \tag{1.17}$$

if $\beta\gamma_{l(\beta)+1} = 1$
where $Z_{J_{l(\beta)}+1}, \ldots, Z_{J_{l(\beta)+1}}$ are independent standard Gaussian r.v.

Remark: Note that the event in (1.17) is not empty since the necessary equality of parameters holds at least for $i = J_{l(\beta)+1}$.

Let us give the definition of the integrals $\int_{\mathbb{R}^m} e^{\beta\gamma_1 x_1+\cdots+\beta\gamma_m x_m}\mathcal{P}^{(m)}(dx_1\dots dx_m)$ over the process $\mathcal{P}^{(m)}$ on $\mathbb{R}^m$. The key facts, that were proven in Propositions 3.1, 3.2 of [BK1] are enumerated in

Proposition 1.4: *Assume that the numbers $\gamma_1,\dots,\gamma_m$ are such that $\gamma_1>\gamma_2>\cdots>\gamma_m>0$ and $\beta\gamma_m>1$. Then*

(i) For any $a\in R$ the process $\mathcal{P}^{(m)}$ contains a.s. a finite number of points $(x_1,\dots,x_m)$ such that $\gamma_1x_1+\cdots+\gamma_mx_m>a$.

(ii) The limit

$$\lim_{x\to-\infty}\int_{\substack{(x_1,\dots,x_m)\in\mathbb{R}^m,\\ \exists i,1\le i\le m:\gamma_1x_1+\cdots+\gamma_ix_i>(\gamma_1+\cdots+\gamma_i)x}} e^{\beta\gamma_1x_1+\cdots+\beta\gamma_mx_m}\mathcal{P}^{(m)}(dx_1\dots dx_m)$$

is finite a.s.

For $m=1$ we recall the definition of [BKL]: for all $x\in\mathbb{R}$ $\int_x^\infty e^{\beta\gamma_1x_1}\mathcal{P}^{(1)}(dx)$ being a sum over a finite number of points a.s., $\int_{-\infty}^\infty e^{\beta\gamma_1x_1}\mathcal{P}^{(1)}(dx)\equiv\lim_{x\to-\infty}\int_x^\infty e^{\beta\gamma_1x_1}\mathcal{P}^{(1)}(dx)$ and this limit is finite a.s. If the integral in question is defined up to $m-1$, we may consider

$$\begin{aligned}&\int_{\substack{(x_1,\dots,x_m)\in\mathbb{R}^m,\\ \exists i,1\le i\le m:\gamma_1x_1+\cdots+\gamma_ix_i>(\gamma_1+\cdots+\gamma_i)x}} e^{\beta\gamma_1x_1+\cdots+\beta\gamma_mx_m}\mathcal{P}^{(m)}(dx_1\dots dx_m)\\ &=\sum_{k=1}^m\int_{\substack{(x_1,\dots,x_m)\in\mathbb{R}^m,\\ \forall i=1,\dots,k-1:\gamma_1x_1+\cdots+\gamma_ix_i\le(\gamma_1+\cdots+\gamma_i)x\\ \gamma_1x_1+\cdots+\gamma_kx_k>(\gamma_1+\cdots+\gamma_k)x}} e^{\beta\gamma_1x_1+\cdots+\beta\gamma_mx_m}\mathcal{P}^{(m)}(dx_1\dots dx_m)\end{aligned}\tag{1.18}$$

where each term is a.s. a finite linear combination of already defined integrals. In fact, let us determine the kth term. Given any realisation of $\mathcal{P}^{(m)}$ in $\mathbb{R}^m$, consider its projection on the plane of the first k coordinates distributed as $\mathcal{P}^{(k)}$. By Proposition 1.4 (i) this contains a finite number of points $(x_1,\dots,x_k)$ such that $\gamma_1x_1+\cdots+\gamma_kx_k>(\gamma_1+\cdots+\gamma_k)x$. Whenever the first k coordinates of a point of $\mathcal{P}^{(m)}$ in $\mathbb{R}^m$ are fixed, the remaining $m-k$ coordinates are distributed as $\mathcal{P}^{(m-k)}$ in $\mathbb{R}^{m-k}$. Then the integral of the function $e^{\gamma_{k+1}x_{k+1}+\cdots+\gamma_mx_m}$ over these coordinates is defined by induction and is finite a.s. Thus the kth term of (1.18) is the sum of these integrals taken with the weights $e^{\gamma_1x_1+\cdots+\gamma_kx_k}$ over points $(x_1,\cdots,x_k)$ of $\mathcal{P}^{(k)}$ such that $\forall i=1,\dots,k-1:\gamma_1x_1+\cdots+\gamma_ix_i\le(\gamma_1+\cdots+\gamma_i)x,\ \gamma_1x_1+\cdots+\gamma_kx_k>(\gamma_1+\cdots+\gamma_k)x$. As noticed above, this linear combination has a finite number of terms a.s.

Finally we put

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

which exists a.s. by Proposition 1.4 (ii).

1.3. The Gibbs measures.

We will now turn to an appropriate asymptotic description of the Gibbs measures in the GREM. In the random energy model, we have considered various objects to achieve such a description. The simplest one was Ruelle's process of the Gibbs masses, obtained as the limit of the process $\mathcal{W}_N \equiv \sum_\sigma \delta_{\mu_{\beta,N}(\sigma)}$. Let us point to the fact that together with the information that the replica overlap in the REM can take on only the values 0 and 1, this process describes in a way fully the structure of the Gibbs measure: namely, if one is interested in capturing an arbitrary large fraction $1-p$ of the total mass, then it suffices to consider for some $\epsilon=\epsilon(p)>0$ the atoms of the process $\mathcal{W}$ with mass larger than ϵ, and to place them at random on a set of orthogonal vectors on the infinite-dimensional unit sphere. This yields, in a reasonable sense, an adequate description of the Gibbs mass.

However, in the GREM, this picture is insufficient to describe the Gibbs measure fully. The reason for this is that the overlap distribution may now take on values that are different from 0 and 1. Thus the points carrying the masses described by the process $\mathcal{W}$ are distributed in a more complicated way in space, and must retain more information to capture this spatial picture. Ruelle took this fact into account when defining the "probability cascades" in his version of the GREM. We will describe these objects in the context of the GREM and prove their convergence to Ruelle's cascades.

The overlap distribution.

A key object considered in the physical literature on spin glasses is the distribution of the overlap, i.e. the random probability distribution

$$\tilde{f}_{\beta,N}(q) \equiv \mu_{\beta,N}^{\otimes 2}\left(R_N(\sigma,\sigma')<q\right) \tag{1.19}$$

where $R_N(\sigma,\sigma')\equiv\frac{1}{N}(\sigma,\sigma')$. In the context of the GREM, it is more natural to introduce the distribution of the hierarchical distance, i.e.

$$f_{\beta,N}(q) \equiv \mu_{\beta,N}^{\otimes 2}\left(d_N(\sigma,\sigma')<q\right) \tag{1.20}$$

As we shall see later, these two notions turn out to be asymptotically equivalent in a strong sense. With the notation introduced in Theorem 1.3 according to the partition $J_0,\dots,J_m$ defined in (1.12), let us set

$$q_l \equiv \sum_{n=1}^{l} \frac{\ln \bar{\alpha}_n}{\ln 2} \tag{1.21}$$

and

$$q_{\max}(\beta) \equiv \sum_{n=1}^{l(\beta)} \frac{\ln \bar{\alpha}_n}{\ln 2} \tag{1.22}$$

We will see that the measure $f_{\beta,N}$ converges to a limiting random measure f_β with support on the set $\{0, q_1, \dots, q_{l(\beta)}\}$.

Point processes of masses.

We will introduce a number of point processes that appear to be good candidates for a more detailed description of the Gibbs measure.

Let us introduce the sets

$$B_l(\sigma) \equiv \{\sigma' \in \mathcal{S}_N : d_N(\sigma,\sigma') \geq q_l\}, \quad l = 1, 2, \dots, l(\beta). \tag{1.23}$$

We define point processes $\mathcal{W}_{\beta,N}^{(m)}$ on $(0,1]^m$ given by

$$\mathcal{W}_{\beta,N}^{(m)} \equiv \sum_\sigma \delta_{(\mu_{\beta,N}(B_1(\sigma)),\dots,\mu_{\beta,N}(B_m(\sigma)))} \frac{\mu_{\beta,N}(\sigma)}{\mu_{\beta,N}\left(B_m(\sigma)\right)} \tag{1.24}$$

as well as their projection on the last coordinate,

$$\mathcal{R}_{\beta,N}^{(m)} \equiv \sum_\sigma \delta_{\mu_{\beta,N}(B_m(\sigma))} \frac{\mu_{\beta,N}(\sigma)}{\mu_{\beta,N}\left(B_m(\sigma)\right)} \tag{1.25}$$

It is easy to see that the processes $\mathcal{W}_{\beta,N}^{(m)}$ satisfy

$$\mathcal{W}_{\beta,N}^{(m)}(dw_1,\dots,dw_m) = \int_0^1 W_{\beta,N}^{m+1}(dw_1,\dots,dw_m,dw_{m+1}) \frac{w_{m+1}}{w_m} \tag{1.26}$$

where the integration is of course over the last coordinate w_{m+1}. Note that these processes will in general not all converge, but will do so only when for some σ, $\mu_\beta(B_m(\sigma))$ is strictly positive. From our experience with the Gibbs measure, it is clear that this will be the case precisely when $m \leq l(\beta)$. In fact, we prove the following:

Theorem 1.5: *If $m \leq l(\beta)$, the point process $\mathcal{W}_{\beta,N}^{(m)}$ on $(0,1]^m$ converges weakly to the point process $\mathcal{W}_\beta^{(m)}$ whose atoms $w(i)$ are given in terms of the atoms $(x_1(i),\dots,x_m(i))$ of the point process $\mathcal{P}^{(m)}$ by*

$$
\begin{aligned}
&(w_1(i),\dots,w_m(i))\\
&=\left(\frac{\int \mathcal{P}^{(m)}(dy)\delta(y_1-x_1(i))e^{\beta(\gamma,y)}}{\int \mathcal{P}^{(m)}(dy)e^{\beta(\gamma,y)}},\dots,\frac{\int \mathcal{P}^{(m)}(dy)\delta(y_1-x_1(i))\dots\delta(y_m-x_m(i))e^{\beta(\gamma,y)}}{\int \mathcal{P}^{(m)}(dy)e^{\beta(\gamma,y)}}\right)
\end{aligned}
\tag{1.27}
$$

and the point processes $\mathcal{R}_{\beta,N}^{(m)}$ converge to the point process $\mathcal{R}_\beta^{(m)}$ whose atoms are the last components of the atoms in (1.27).

Of course the most complete object we can reasonably study is the process $\widehat{\mathcal{W}}_\beta \equiv \mathcal{W}_\beta^{l(\beta)}$. Analogously, we will set $\widehat{\mathcal{R}}_\beta \equiv \mathcal{R}_\beta^{l(\beta)}$.

The point processes $\widehat{\mathcal{W}}_\beta$ takes values on vectors whose components form decreasing sequences in $(0,1]$. Moreover, these atoms are naturally clustered in a hierarchical way. These processes were introduced by Ruelle [Ru] and called *probability cascades.*

There is an intimate relation between the distance distributions f_β and these point processes. The first result we shall prove is the following theorem.

Theorem 1.6: *With the notation introduce above, we have that*

(i) The random distribution functions $f_{\beta,N}$ and $\tilde{f}_{\beta,N}$ converge in distribution and in mean to the same random distribution function f_β.

(ii) f_β is a step function with jumps at the values $\{0,q_1,\dots,q_{l(\beta)}\}$. For any $q\in(q_{i-1},q_i]$

$$
f_\beta(q)=\int \widehat{\mathcal{W}}_\beta(dw_1,\dots,dw_{l(\beta)})w_{l(\beta)}(1-w_i),\quad i=1,\dots,l(\beta);
\tag{1.28}
$$

$f_\beta(q)=1$ for $q\in(q_l(\beta),1]$.

We will prove Theorems 1.5 in Section 3 directly as a consequence of Theorem 1.3. We will also prove a more general Theorem 3.1 that imply in particular Theorem 1.6. Its proof is somewhat tedious but in principle straightforward.

1.4. Ghirlanda-Guerra identities.

We now turn to a different approach towards the construction of the processes $\widehat{\mathcal{W}}_\beta$ that will completely avoid the use of the Poisson cascades. In fact, we will see that the convergence of the processes $\widehat{\mathcal{W}}_{\beta,N}$ follows from general principles once we can compute the limiting free energy as a function of the function A describing the covariance of the Gaussian process X_σ. The key role in this approach is played by the so-called *Ghirlanda-Guerra* identities [GG,AC,T2,T5]. The existence of this approach will bear its full fruits in the follow-up paper [BK3] where it will allow us to treat the GREM with continuous hierarchies.

The first result is an explicit expression for the mean of the two-replica distribution function f_β.

Proposition 1.7: *The mean of f_β is given by*

$$\mathbb{E}f_\beta(q) = \begin{cases} \beta^{-1}\sqrt{\frac{2\ln\bar{\alpha}_j}{\bar{a}_j}}, & \text{if} q \in (q_{j-1}, q_j],\ j \leq l(\beta), \\ 1, & \text{if} q > q_{max}(\beta) \end{cases} \tag{1.29}$$

Let us denote by $\mu_{\beta,N}^{\otimes k}$ the distribution of k independent copies of σ under the Gibbs measure. The following proposition corresponds to the Ghirlanda-Guerra relations in our models.

Proposition 1.8: *Assume that the parameters α_i and a_i are such that none of the inequalities (1.9) holds with equality. Then, for any bounded function $h : \mathcal{S}_N^n \to \mathbb{R}$, for all $i = 1, \dots, n$*

$$\lim_{N\uparrow\infty}\Bigg| \mathbb{E}\mu_{\beta,N}^{\otimes n+1}\left(h(\sigma^1,\dots,\sigma^n)\mathbb{1}_{\sigma_1^k\dots\sigma_i^k=\sigma_1^{n+1}\dots\sigma_i^{n+1}}\right) - \frac{1}{n}\mathbb{E}\mu_{\beta,N}^{\otimes n+1}\left(h(\sigma^1,\dots,\sigma^n)\left(\sum_{l\neq k}^n \mathbb{1}_{\sigma_1^l\dots\sigma_i^l=\sigma_1^k\dots\sigma_i^k} + \mathbb{E}\mu_{\beta,N}^{\otimes 2}(\mathbb{1}_{\sigma_1^1\dots\sigma_i^1=\sigma_1^2\dots\sigma_i^2})\right)\right)\Bigg| = 0 \tag{1.30}$$

It is quite remarkable that these relations suffice to determine completely the distribution of the random measures $\widehat{\mathcal{W}}_\beta$ through their generalized moments.

Theorem 1.9: *The random measures $\widehat{\mathcal{W}}_\beta$ are completely determined by the relations (1.30) up to the mean value of the two-replica distance distribution function f_β given in Proposition 1.7.*

Proof: Since it is important to understand this result, we will prove this theorem immediately. Note first that the last term in (1.30) is just

$$\lim_{\mathbb{N}\uparrow\infty}\mathbb{E}\mu_{\beta,N}^{\otimes 2}(\mathbb{1}_{\sigma_1^1\dots\sigma_i^1=\sigma_1^2\dots\sigma_i^2}) = \mathbb{E}\mu_{\beta,N}^{\otimes 2}(\mathbb{1}_{d_N(\sigma,\sigma')\geq\sum_{j=1}^i \ln\alpha_j/\ln 2}) = 1 - \mathbb{E}f_\beta\Big(\sum_{j=1}^i \ln\alpha_j/\ln 2\Big). \tag{1.31}$$

Let now $\underline{\underline{J}} \equiv (\underline{J}_0, \dots, \underline{J}_n)$ denote a nested set of finite subsets of $\mathbb{N}$ such that each $\underline{J}_r = (J_{r,1}, \dots, J_{r,j_r})$ is a collection of disjoint subsets of $\mathbb{N}$, such that for each $J_{r,i}$ there exists $J_{r-1,k}$ such that $J_{r,i} \subset J_{r-1,k}$. Let us fix the notation that $J_{r,i} \equiv \{j_1^{r,i}, \dots, j_{|J_{r,i}|}^{r,i}\}$. Given such a $\underline{\underline{J}}$, define the events

$$\mathcal{A}_{\underline{\underline{J}}} \equiv \prod_{r=1}^{n} \prod_{i=1}^{j_r} \mathbb{1}_{\{\sigma_1^{j_1^{r,i}} \dots \sigma_r^{j_1^{r,i}} = \dots = \sigma_1^{j_{|J_{r,i}|}^{r,i}} \dots \sigma_r^{j_{|J_{r,i}|}^{r,i}}\}} \tag{1.32}$$

Then define, with $J \equiv |\underline{J}_0|$,

$$A_{\underline{\underline{J}}} \equiv \lim_{N \uparrow \infty} \mathbb{E} \mu_{\beta,N}^{\otimes J} \mathcal{A}_{\underline{\underline{J}}} \tag{1.33}$$

Now all we have to realise is that for any such event $\mathcal{A}_{\underline{\underline{J}}}$, as long as $\underline{J}_0$ is of size larger than two, there are l, m, r, $1 \leq r \leq n$, such that for some $\underline{\underline{J}}'$ with $J' = J - 1$, $l \in \underline{\underline{J}}'$ and $m \notin \underline{\underline{J}}'$:

$$\mathcal{A}_{\underline{\underline{J}}} = \mathcal{A}_{\underline{\underline{J}}'} \mathbb{1}_{\sigma_1^l \dots \sigma_r^l = \sigma_1^m \dots \sigma_r^m}. \tag{1.34}$$

Thus using (1.30) with $h = \mathcal{A}_{\underline{\underline{J}}'}$ allows to express $A_{\underline{\underline{J}}}$ in term of $A_{\underline{\underline{J}}'}$ with $\underline{\underline{J}}'$ that have total length strictly less than $\underline{\underline{J}}$, and those of length two. Thus it is obvious that the reduction procedure will yield expressions for any $A_{\underline{\underline{J}}}$ as function of those involving only two replicas. Of course there are only n such sets, depending only on the number of non-empty levels. These of course have an immediate meaning in terms of the two-overlap distribution via (1.31).

This result generalizes the corresponding finding in the REM, where of course the multi-index $\underline{\underline{J}}$ is trivial.

We will now show that the numbers $A_{\underline{\underline{J}}}$ fully determine the processes $\widehat{\mathcal{W}}_\beta$ defined in Theorem 1.5 (in complete analogy to what Talagrand [T2] found in the REM). To see this, note that a random measure $\mathcal{M}_N$ on $\mathbb{R}^d$ converges to $\mathcal{M}$, if for any $k \in \mathbb{N}$, any continuous function $F : \mathbb{R}^k \to \mathbb{R}$, and any collection of continuous, integrable functions $f_1, \dots, f_k$ on $\mathbb{R}^d$,

$$F\left(\int \mathcal{M}_N(dx) f_1(x), \dots, \int \mathcal{M}_N(dx) f_k(x)\right) \to F\left(\int \mathcal{M}(dx) f_1(x), \dots, \int \mathcal{M}(dx) f_k(x)\right) \tag{1.35}$$

(see e.g. [Ka,Re]) By standard approximation arguments, if $\mathcal{M}$ has moments of all orders, it is enough to consider for f_i all monomial functions $f(x) = x_1^{i_1} \dots x_d^{i_d}$, and $F(y) = y_1 \dots y_k$.

Now let $\mathcal{W}_{\beta,N}^{(m)}$ be any process of the form (1.24). Then

$$\begin{aligned}
&\int \mathcal{W}_{\beta,N}^{(m)}(dw) w_1^{i_1} \dots w_l^{i_l} \dots w_m^{i_m} \\
&= \sum_\sigma \Big(\sum_{\sigma'_2,\dots,\sigma'_m} \mu_{\beta,N}(\sigma_1\sigma'_2,\dots,\sigma'_m) \Big)^{i_1} \dots \Big(\sum_{\sigma'_{l+1},\dots,\sigma'_m} \mu_{\beta,N}(\sigma_1 \dots \sigma_l \sigma'_{l+1} \dots \sigma'_m) \Big)^{i_l} \dots \\
&\qquad \dots \Big(\mu_{\beta,N}(\sigma_1 \dots \sigma_m) \Big)^{i_m} \\
&= \mu_{\beta,N}^{\otimes(i_1+\dots+i_m)} \Big(\mathbb{1}_{\{\sigma_1^1 = \dots = \sigma_1^{i_1+\dots+i_m}\}} \dots \\
&\qquad \dots \mathbb{1}_{\{\sigma_1^{i_1+\dots+i_{l-1}+1} \dots \sigma_l^{i_1+\dots+i_{l-1}+1} = \dots = \sigma_1^{i_1+\dots+i_m} \dots \sigma_l^{i_1+\dots+i_m}\}} \\
&\qquad \dots \mathbb{1}_{\{\sigma_1^{i_1+\dots+i_{m-1}+1} \dots \sigma_m^{i_1+\dots+i_{m-1}+1} = \dots = \sigma_1^{i_1+\dots+i_m} \dots \sigma_m^{i_1+\dots+i_m}\}} \Big)
\end{aligned} \tag{1.36}$$

where $i_m \geq 1$ (otherwise, the expression would be infinite). Clearly, taking products of such terms, the resulting objects are precisely expectations of events of the form $\mathcal{A}_{\underline{\underline{J}}}$. This proves Theorem 1.9. $\diamondsuit$

The recursions on the $A_{\underline{\underline{J}}}$ may look forbidding, but if only limited information is sought, they can provide quickly a great amount of information. As a first trivial corollary, they yield a very simple proof of a result of Ruelle [Ru]:

Theorem 1.10: *Any of the one-dimensional processes $\mathcal{R}_\beta^m$ has the same distribution as Ruelle's process with parameter $\mathbb{E} f_\beta(q_m)$, $m = 1, \dots, l(\beta)$.*

Proof: Note that if we are interested only in the one-dimensional processes $\mathcal{R}_\beta^m$, we only need to control moments of the form

$$\int \mathcal{R}_{\beta,N}^{(m)}(dw) w^r = \sum_\sigma \Big(\sum_{\sigma'_{m+1},\dots,\sigma'_n} \mu_{\beta,N}(\sigma_1 \dots \sigma_m \sigma'_{m+1} \dots \sigma'_n) \Big)^r = \mu_{\beta,N}^{\otimes r} \big(\mathbb{1}_{\{\sigma_1^1 \dots \sigma_m^1 = \dots = \sigma_1^r \dots \sigma_m^r\}} \big) \tag{1.37}$$

Thus we have the same type of objects as in the REM, and since the relations (1.30) have the same form for all m, the claimed result follows from the preceding theorem. $\diamondsuit$

To see how simple quantities of interest can be computed very explicitly, consider e.g. in the REM

$$A_n \equiv \lim_{N \uparrow \infty} \mathbb{E} \mu_{\beta,N}^{\otimes n} (\mathbb{1}_{\sigma^1 = \sigma^2 = \dots = \sigma^n}) \tag{1.38}$$

then (1.30) with $h = \mathbb{1}_{\sigma^1=\sigma^2=\dots=\sigma^n}$ provides the recursion

$$\begin{aligned} A_{n+1} &= \frac{n-1}{n}A_n + \frac{1}{n}A_nA_2 = A_n\left(1-\frac{1-A_2}{n}\right) \\ &= \prod_{k=2}^{n}\left(1-\frac{1-A_2}{k}\right)A_2 = \frac{\Gamma(n+A_2)}{\Gamma(n+1)\Gamma(A_2)} \end{aligned} \tag{1.39}$$

But

$$\lim_{N\uparrow\infty}\mu_{\beta,N}^{\otimes 2}(\mathbb{1}_{\sigma^1=\sigma^2=\dots=\sigma^n}) = \int \mathcal{R}_\beta(dw)w^n \tag{1.40}$$

so that (1.36) implies a formula for the mean of the n-th moments of $\mathcal{R}_\beta$,

$$\mathbb{E}\int \mathcal{R}_\beta(dw)w^n = \frac{\Gamma(n+A_2)}{\Gamma(n+1)\Gamma(A_2)} \tag{1.41}$$

where $A_2 = \mathbb{E}\int \mathcal{R}_\beta(dw)w^2$. This result has been obtained by a direct computation by Ruelle ([Ru], Corollary 2.2), but its derivation via the Ghirlanda–Guerra identities is considerably simpler.

In the remainder of this paper we will prove the various results claimed above: In Section 3 we prove Theorem 1.3 and 1.5. Section 4 contains the proof of Theorem 1.6. Finally, in Section 5, we prove Proposition 3.1 and the Ghirlanda-Guerra identities stated in Proposition 1.8

Acknowledgements: A.B. thanks Francis Comets and the Université Paris 7 for their hospitality during the final period of the writing of this paper.

2. Proof Theorems 1.3 and 1.5.

In this section we prove the convergence results for the partition function (Theorem 1.3) and that of the Gibbs measures to Ruelle's processes $\mathcal{W}_\beta^m$ (Theorem 1.5).

Proof of Theorem 1.3. It is based on Lemmas 2.1 and 2.2 below. Lemma 2.1 proves the theorem in the case $l(\beta) = 0$. The proof in general case proceeds by induction on $l(\beta)$. We represent the normalised partition function as the sum of two terms

$$
e^{\sum_{j=1}^{l(\beta)}\left(-\beta N\sqrt{2\bar a_j \ln\bar\alpha_j}+\beta\gamma_j[\ln(N\ln\bar\alpha_j)+\ln 4\pi]/2+N\ln\alpha_j\right)-N\sum_{i=J_{l(\beta)}+1}^{n}\beta^2 a_i/2} Z_{\beta,N}
$$

$$
\begin{aligned}
&=\sum_{\sigma_1\ldots\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})}\ldots\sum_{\sigma_{J_{l(\beta)-1}+1}\ldots\sigma_{J_{l(\beta)}}} e^{\beta\gamma_{l(\beta)}u^{-1}_{\ln\bar\alpha_{l(\beta)},N}(\bar X^{\sigma_1\cdots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\cdots\sigma_{J_{l(\beta)}}})}C(\beta)\\
&+\sum_{\sigma_1\ldots\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})}\ldots\sum_{\sigma_{J_{l(\beta)-1}+1}\ldots\sigma_{J_{l(\beta)}}} e^{\beta\gamma_{l(\beta)}u^{-1}_{\ln\bar\alpha_{l(\beta)},N}(\bar X^{\sigma_1\cdots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\cdots\sigma_{J_{l(\beta)}}})}\\
&\times\Big(e^{-\sum_{i=J_{l(\beta)}+1}^{n}(N\ln\alpha_i+N\beta^2\alpha_i/2)}\sum_{\sigma_{J_{l(\beta)}+1}} e^{\beta\sqrt{a_{J_{l(\beta)}+1}}X_{\sigma_1\cdots\sigma_{J_{l(\beta)}+1}}}\ldots\sum_{\sigma_m}e^{\beta\sqrt{a_m}X_{\sigma_1\ldots\sigma_m}}-C(\beta)\Big)\\
&=I_1^N+I_2^N
\end{aligned}
$$

with the same notation $\bar X^{\sigma_1\ldots\sigma_{J_{l-1}}}_{\sigma_{J_{l-1}+1}\ldots\sigma_{J_l}}$ as in (1.8) and show that I_1^N converges in distribution to the integral (1.15) and that I_2^N converges to zero.

For any $x\in\mathbb{R}$ the difference between I_1^N and the integral I of (1.15) can be written as follows:

$$
\begin{aligned}
&I_1^N-I\\
&=\sum_{k=1}^{l(\beta)}\Big[\sum_{\sigma_1,\ldots\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})}\ldots\sum_{\sigma_{J_{k-1}+1}\ldots\sigma_{J_k}} e^{\beta\gamma_k u^{-1}_{\ln\bar\alpha_k,N}(\bar X^{\sigma_1\cdots\sigma_{J_{k-1}}}_{\sigma_{J_{k-1}+1}\cdots\sigma_{J_k}})}\\
&\qquad\times 1\!\!1_{\{\forall i=1,\ldots,k-1:\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})+\cdots+\gamma_i u^{-1}_{\ln\bar\alpha_i,N}(\bar X^{\sigma_1\cdots\sigma_{J_{i-1}}}_{\sigma_{J_{i-1}+1}\cdots\sigma_{J_i}})\le(\gamma_1+\cdots+\gamma_i)x\}}\\
&\qquad\times 1\!\!1_{\{\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})+\cdots+\gamma_k u^{-1}_{\ln\bar\alpha_k,N}(\bar X^{\sigma_1\cdots\sigma_{J_{k-1}}}_{\sigma_{J_{k-1}+1}\cdots\sigma_{J_k}})>(\gamma_1+\cdots+\gamma_k)x\}}\\
&\quad\times C(\beta)\sum_{\sigma_{J_k+1},\ldots\sigma_{J_{k+1}}} e^{\beta\gamma_{k+1}u^{-1}_{\ln\bar\alpha_1,N}(\bar X^{\sigma_1\cdots\sigma_{J_k}}_{\sigma_{J_k+1}\cdots\sigma_{J_{k+1}}})}\ldots\sum_{\sigma_{J_{l(\beta)-1}+1}\ldots\sigma_{J_{l(\beta)}}} e^{\beta\gamma_{l(\beta)}u^{-1}_{\ln\bar\alpha_{l(\beta)},N}(\bar X^{\sigma_1\cdots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\cdots\sigma_{J_{l(\beta)}}})}\\
&\qquad-C(\beta)\int_{\substack{(x_1,\ldots,x_{l(\beta)})\in\mathbb{R}^{l(\beta)},\\ \forall i=1,\ldots,k-1:\gamma_1x_1+\cdots+\gamma_ix_i\le(\gamma_1+\cdots+\gamma_i)x\\ \gamma_1x_1+\cdots+\gamma_kx_k>(\gamma_1+\cdots+\gamma_k)x}} e^{\beta\gamma_1x_1+\cdots+\beta\gamma_{l(\beta)}x_{l(\beta)}}\mathcal{P}^{(l(\beta))}(dx_1\ldots dx_{l(\beta)})\Big]\\
&\quad+C(\beta)\sum_{\sigma_1,\ldots\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})}\ldots\sum_{\sigma_{J_{l(\beta)-1}+1}\ldots\sigma_{J_{l(\beta)}}} e^{\beta\gamma_{l(\beta)}u^{-1}_{\ln\bar\alpha_{l(\beta)},N}(\bar X^{\sigma_1\cdots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\cdots\sigma_{J_{l(\beta)}}})}\\
&\qquad\times 1\!\!1_{\{\forall i=1,\ldots,k:\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})+\cdots+\gamma_i u^{-1}_{\ln\bar\alpha_i,N}(\bar X^{\sigma_1\cdots\sigma_{J_{i-1}}}_{\sigma_{J_{i-1}+1}\cdots\sigma_{J_i}})\le(\gamma_1+\cdots+\gamma_i)x\}}\\
&\quad-C(\beta)\int_{\substack{(x_1,\ldots,x_{l(\beta)})\in\mathbb{R}^{l(\beta)},\\ \forall i=1,\ldots,k:\gamma_1x_1+\cdots+\gamma_ix_i\le(\gamma_1+\cdots+\gamma_i)x}} e^{\beta\gamma_1x_1+\cdots+\beta\gamma_{l(\beta)}x_{l(\beta)}}\mathcal{P}^{(l(\beta))}(dx_1\ldots dx_{l(\beta)}).
\end{aligned}
\tag{2.1}
$$

Given $\epsilon>0$, one can choose $x<0$ sufficiently large by absolute value such that for all N starting from some $N(x)$, the probability that the last two terms of (2.1) are smaller than

ϵ, is arbitrarily close to 1. Namely, for the last term this is obvious from Proposition 1.4 (ii). For the term before the last it suffices to apply Chebyshev's inequality and to estimate its expectation after. Since $\gamma_1 > \cdots > \gamma_{l(\beta)}$ and $\beta\gamma_{l(\beta)} > 1$, then by Lemma 2.2 for all $N \geq N(x)$ the expectation of this term is smaller than $C(\beta)Ce^{x\sum_{j=1}^{l(\beta)}(\beta\gamma_j - 1)} \downarrow 0$ as $x \downarrow -\infty$. Let us fix $x < 0$ large enough by absolute value and consider the kth term in square brackets, $k = 1, 2, \ldots, l(\beta)$. The inequalities (1.9) being verified for the partition (1.12), we have the point processes convergence stated in Theorem 1.1. If $k = l(\beta)$, then by Proposition 1.4 (i), the integral in the $l(\beta)$th term is over a finite number of points of $\mathcal{P}^{(l(\beta))}$ a.s. Then, indeed, this term converges to zero in distribution. If $1 \leq k < l(\beta)$, then the un-truncated sums over $\sigma_{J_k+1}, \ldots, \sigma_n$ are involved in the kth term. They are distributed as I_N^1 for the GREM of $n - J_k$ levels where the parameter $l(\beta)$ equals $l(\beta) - k$. By induction each of these sums converges in distribution to the integral of $e^{\beta\gamma_{k+1}x_{k+1}+\cdots+\beta\gamma_{l(\beta)}x_{l(\beta)}}$ over the process $\mathcal{P}^{(l(\beta)-k)}$ on $R^{l(\beta)-k}$. Moreover, by Proposition 1.4 (i) the integral in the kth term of (2.1) is the linear combination of a.s. finite number of independent integrals distributed as these. These arguments together imply the convergence to zero of the kth term of (2.1) and complete the analysis of I_N^1.

Let us write a similar representation for I_N^2:

$$
\begin{aligned}
I_2^N
&=\sum_{k=1}^{l(\beta)}\sum_{\sigma_1,\ldots\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})}\cdots\sum_{\sigma_{J_{k-1}+1}\ldots\sigma_{J_k}} e^{\beta\gamma_k u^{-1}_{\ln\bar\alpha_k,N}(\bar X^{\sigma_{J_1}\cdots\sigma_{J_{k-1}}}_{\sigma_{J_{k-1}+1}\cdots\sigma_{J_k}})}\\
&\quad\times 1\!\!\mathrm{I}_{\{\forall i=1,\ldots,k-1:\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})+\cdots+\gamma_i u^{-1}_{\ln\bar\alpha_i,N}(\bar X^{\sigma_{J_1}\cdots\sigma_{J_{i-1}}}_{\sigma_{J_{i-1}+1}\cdots\sigma_{J_i}})\le(\gamma_1+\cdots+\gamma_i)x\}}\\
&\quad\times 1\!\!\mathrm{I}_{\{\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})+\cdots+\gamma_k u^{-1}_{\ln\bar\alpha_k,N}(\bar X^{\sigma_{J_1}\cdots\sigma_{J_{k-1}}}_{\sigma_{J_{k-1}+1}\cdots\sigma_{J_k}})>(\gamma_1+\cdots+\gamma_k)x\}}\\
&\times\sum_{\sigma_{J_k+1},\ldots\sigma_{J_{k+1}}} e^{\beta\gamma_{k+1}u^{-1}_{\ln\bar\alpha_1,N}(\bar X^{\sigma_{J_1}\cdots\sigma_{J_k}}_{\sigma_{J_k+1}\cdots\sigma_{J_{k+1}}})}\cdots\sum_{\sigma_{J_{l(\beta)-1}+1}\ldots\sigma_{J_{l(\beta)}}} e^{\beta\gamma_{l(\beta)}u^{-1}_{\ln\bar\alpha_{l(\beta)},N}(\bar X^{\sigma_{J_1}\cdots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\cdots\sigma_{J_{l(\beta)}}})}\\
&\times\Big(e^{-\sum_{i=J_{l(\beta)}+1}^{n}(N\ln\alpha_i+N\beta^2\alpha_i/2)}\sum_{\sigma_{J_{l(\beta)}+1}} e^{\beta\sqrt{a_{J_{l(\beta)}+1}}X_{\sigma_1\cdots\sigma_{J_{l(\beta)}+1}}}\cdots\sum_{\sigma_m} e^{\beta\sqrt{a_m}X_{\sigma_1\cdots\sigma_m}}-C(\beta)\Big)\\
&+\sum_{\sigma_1,\ldots\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})}\cdots\sum_{\sigma_{J_{l(\beta)-1}+1}\ldots\sigma_{J_{l(\beta)}}} e^{\beta\gamma_{l(\beta)}u^{-1}_{\ln\bar\alpha_{l(\beta)},N}(\bar X^{\sigma_{J_1}\cdots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\cdots\sigma_{J_{l(\beta)}}})}\\
&\quad\times 1\!\!\mathrm{I}_{\{\forall i=1,\ldots,k:\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})+\cdots+\gamma_i u^{-1}_{\ln\bar\alpha_i,N}(\bar X^{\sigma_{J_1}\cdots\sigma_{J_{i-1}}}_{\sigma_{J_{i-1}+1}\cdots\sigma_{J_i}})\le(\gamma_1+\cdots+\gamma_i)x\}}\\
&\times\Big(e^{-\sum_{i=J_{l(\beta)}+1}^{n}(N\ln\alpha_i+N\beta^2\alpha_i/2)}\sum_{\sigma_{J_{l(\beta)}+1}} e^{\beta\sqrt{a_{J_{l(\beta)}+1}}X_{\sigma_1\cdots\sigma_{J_{l(\beta)}+1}}}\cdots\sum_{\sigma_m} e^{\beta\sqrt{a_m}X_{\sigma_1\cdots\sigma_m}}-C(\beta)\Big).
\end{aligned}
\tag{2.2}
$$

Again, by Lemma 2.2, for all $N\ge N(x)$ the expectation of the absolute value of the last term in (2.2) does not exceed $2CC(\beta)e^{x\sum_{j=1}^{l(\beta)}(\beta\gamma_j-1)}\downarrow 0$ as $x\downarrow-\infty$. Then, given $\epsilon>0$, using Chebyshev's inequality one can fix $x<0$ such that for all $N\ge N(x)$ the last term of (2.2) is smaller than ϵ with probability arbitrarily close to 1. Having fixed $x<0$ with sufficiently large absolute value, let us turn to the sum of the first $l(\beta)$ terms. In the kth term of (2.2) the un-truncated sums multiplied by the term in round brackets are distributed as I_2^N in the GREM of $n-J_k$ levels where the parameter $l(\beta)$ equals $l(\beta)-k$. They converge to zero in probability by induction. Furthermore, for any $\delta>0$ one can find large enough numbers $N(\delta,x)$, $R(\delta,x)$ and $X(\delta,x)$ such that for all $N\ge N(\delta,x)$ with probability at least $1-\delta$ the kth term of (2.2) is a linear combination of at most $R(\delta,x)$ such sums with the weights at most $e^{\beta X(\delta,x)}$. The number $R(\delta,x)$ exists by Proposition 1.4 (i) and of the number $X(\delta,x)$ does by Theorem 1.5 of [BK1]. This theorem asserts the point process convergence of extremes $\mathcal{E}_N=\int_{\mathbb{R}^k}\delta_{\delta_1x_1+\cdots+\delta_kx_k}\mathcal{P}_N^k(dx_1,\ldots,dx_k)\to\mathcal{E}=\int_{\mathbb{R}^k}\delta_{\delta_1x_1+\cdots+\delta_kx_k}\mathcal{P}^k(dx_1,\ldots,dx_k)$ where the process $\mathcal{E}$ has at least one point in $[a,\infty)$ with exponentially small probability as $a\to\infty$. Then the kth term in I_2^N converges to zero in probability, as $N\uparrow\infty$, $k=1,\ldots,l(\beta)$. Hence, I_2^N converges to zero in probability. This concludes the proof of the theorem. $\diamondsuit$

Lemma 2.1: *Let in the GREM $l(\beta)=0$. Then*

$$e^{-N\sum_{i=1}^{n}\beta^2 a_i/2} Z_{\beta,N} \stackrel{\mathcal{D}}{\to} C(\beta).$$

The constant $C(\beta)=1$ if $\beta\gamma_1<1$ and

$$C(\beta)=P\Big(\bigcap_{\substack{i:1\leq i\leq J_1\\ (a_1+\cdots+a_i)/\bar{a}_{J_1}=\ln(\alpha_1\cdots\alpha_i)/\ln\bar{\alpha}_{J_1}}} (\sqrt{a_1}Z_1+\cdots+\sqrt{a_i}Z_i<0)\Big) \tag{2.3}$$

if $\beta\gamma_1=1$ where $Z_1,\ldots,Z_{J_1}$ are independent standard Gaussian r.v.

Proof. By Proposition 1.4 of [BK1] $A_{1,j}\leq A_{1,J_1}=\gamma_1^2$ for all $j=1,\ldots,J_1$ and by the definition of J_1 $A_{1,j}<A_{1,J_1}=\gamma_1^2$ for $j>J_1$. Therefore, if $\beta\gamma_1<1$, then $\beta^2A_{1,j}<1$ for all $j=1,2,\ldots,n$; if $\beta\gamma_1=1$, then $\beta^2A_{1,j}\leq 1$ for all $j=1,2,\ldots,n$ and the equality takes place only for those $j\leq J_1$ for which $(a_1+\cdots+a_j)/\bar{a}_{J_1}=\ln(\alpha_1\cdots\alpha_j)/\ln\bar{\alpha}_{J_1}$.

First, we study the case of $\beta\gamma_1<1$, in which we show that $Z_{\beta,N}/\mathbb{E}Z_{\beta,N}\stackrel{\mathcal{D}}{\to}1$ as $N\uparrow\infty$. Let us fix $\epsilon>0$ such that

$$\ln(\alpha_1\cdots\alpha_j)-(\beta^2-(\beta-\epsilon)^2/2)(a_1+\cdots+a_j)>0 \tag{2.4}$$

for all $j=1,\ldots,n$ because of the strict inequality $\beta^2A_{1,j}<1$. We use truncation techniques originally introduced by M. Talagrand. Let us define the truncated partition function

$$\begin{aligned} Z^T_{\beta,N} = 2^{-N}\sum_{\sigma_1} e^{\beta\sqrt{a_1}X_{\sigma_1}}\cdots\sum_{\sigma_n} e^{\beta\sqrt{a_1}X_{\sigma_1\sigma_2\cdots\sigma_n}} \\ \times \mathbb{1}_{\{\forall_{i=1,\ldots,n}\sqrt{a_1}X_{\sigma_1}+\cdots+\sqrt{a_i}X_{\sigma_1\ldots\sigma_i}<(\beta+\epsilon)(a_1+\cdots+a_i)\sqrt{N}\}} \end{aligned} \tag{2.5}$$

and set

$$\bar{Z}^T_{\beta,N}\equiv Z_{\beta,N}-Z^T_{\beta,N} \tag{2.6}$$

We can then write

$$\frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}}=\frac{Z^T_{\beta,N}}{\mathbb{E}Z^T_{\beta,N}}\times\frac{\mathbb{E}Z^T_{\beta,N}}{\mathbb{E}Z_{\beta,N}}+\frac{\bar{Z}^T_{\beta,N}}{\mathbb{E}Z_{\beta,N}}. \tag{2.7}$$

It is easy to compute

$$\begin{aligned} \mathbb{E}Z^T_{\beta,N} &= e^{\beta^2N/2}\mathbb{P}(\forall i=1,\ldots,n:\sqrt{a_1}Z_1+\cdots+\sqrt{a_i}Z_i<\epsilon(a_1+\cdots+a_i)\sqrt{N}) \\ &\sim e^{\beta^2N/2}=\mathbb{E}Z_{\beta,N} \end{aligned} \tag{2.8}$$

where $Z_1,\ldots,Z_n$ are independent standard Gaussian r.v. Thus the second term in (2.7) converges to zero in L_1 and in the first term the factor $\mathbb{E}Z^T_{\beta,N}/\mathbb{E}Z_{\beta,N}$ converges to 1. Finally, for any $\delta > 0$

$$
\begin{aligned}
\mathbb{P}(|Z^T_{\beta,N}/\mathbb{E}Z^T_{\beta,N} - 1| > \delta) &\le \frac{\mathbb{E}(Z^T_{\beta,N} - \mathbb{E}Z^T_{\beta,N})^2}{(\mathbb{E}Z^T_{\beta,N})^2} \\
&\le \frac{\sum_{j=1}^N (\alpha_1\cdots\alpha_j)^{-N}\mathbb{E}e^{2\beta\sqrt{N}(\sqrt{a_1}X_1+\cdots+\sqrt{a_j}X_j)}1\!\!1_{\{\sqrt{a_1}X_1+\cdots+\sqrt{a_j}X_j<(\beta+\epsilon)(a_1+\cdots+a_j)\sqrt{N}\}}}{\delta^2 e^{\beta^2N(a_1+\cdots+a_j)}(1+o(1))} \\
&\le (1+o(1))\delta^{-2}\sum_{j=1}^N(\alpha_1\cdots\alpha_j)^{-N}e^{\beta^2N(a_1+\cdots+a_j)}\mathbb{P}\big(Z<(-\beta+\epsilon)\sqrt{N(a_j+\cdots+a_n)}\big) \\
&\le (1+o(1))\delta^{-2}\sum_{j=1}^N(\alpha_1\cdots\alpha_j)^{-N}e^{\beta^2N(a_1+\cdots+a_j)}e^{-N(\beta-\epsilon)^2(a_j+\cdots+a_n)/2}
\end{aligned}
\tag{2.9}
$$

where $X_1,\dots,X_n$ and Z are independent standard Gaussian r.v. In view of the choice of $\epsilon>0$, (2.9) converges to zero as $N\uparrow\infty$. Then the first factor in the first term of (2.7) $Z^T_{\beta,N}/\mathbb{E}Z^T_{\beta,N}\to 1$ in probability finishing the proof of the lemma in the case $\beta\gamma_1<1$.

Assume now that $\beta\gamma_1=1$. We fix ϵ according to (2.4) for all j such that $\beta^2A_{1,j}<1$. (In particular, the last inequality holds for all $j>J_1$.) For those $j\le J_1$ for which $\beta^2A_{1,j}=1$ we truncate the partition function by $u_{\ln(\alpha_1\cdots\alpha_j),N}(0)\sqrt{a_1+\cdots+a_j}$. Then

$$
\begin{aligned}
Z^T_{\beta,N} =& 2^{-N}\sum_{\sigma_1}e^{\beta\sqrt{a_1}X_{\sigma_1}}\cdots\sum_{\sigma_n}e^{\beta\sqrt{a_1}X_{\sigma_1\sigma_2\cdots\sigma_n}} \\
&\times 1\!\!1_{\{\forall j=1,\dots,n,\ A_{1,j}<A_{1,J_1},\ \sqrt{a_1}X_{\sigma_1}+\cdots+\sqrt{a_j}X_{\sigma_1\cdots\sigma_j}<(\beta+\epsilon)(a_1+\cdots+a_j)\sqrt{N}\}} \\
&\times 1\!\!1_{\{\forall j=1,\dots,J_1,\ A_{1,j}=A_{1,J_1},\ \sqrt{a_1}X_{\sigma_1}+\cdots+\sqrt{a_j}X_{\sigma_1\cdots\sigma_j}<u_{\ln(\alpha_1\cdots\alpha_j),N}(0)\sqrt{a_1+\cdots+a_j}\}}.
\end{aligned}
$$

Let us compute again

$$
\begin{aligned}
&\mathbb{E}Z^T_{\beta,N} \\
&= e^{\beta^2N/2}\mathbb{P}\Big(\forall j=1,\dots,n:\text{if } A_{1,j}<A_{1,J_1},\ \sqrt{a_1}Z_1+\cdots+\sqrt{a_j}Z_j<\epsilon(a_1+\cdots+a_j)\sqrt{N}; \\
&\qquad\qquad \text{and if } A_{1,j}=A_{1,J_1}, \\
&\qquad \sqrt{a_1}Z_1+\cdots+\sqrt{a_j}Z_j<(u_{\ln(\alpha_1\cdots\alpha_j),N}(0)-\beta\sqrt{a_1+\cdots+a_j})\sqrt{a_1+\cdots+a_j}\Big) \\
&\sim C(\beta)e^{\beta^2N/2}=C(\beta)\mathbb{E}Z_{\beta,N}
\end{aligned}
$$

as $u_{\ln(\alpha_1\cdots\alpha_j),N}(0)-\beta\sqrt{a_1+\cdots+a_j}=o(1)$ whenever $\beta^2A_{1,J_1}=\beta^2A_{1,j}=1$. Then the factor $\mathbb{E}Z^T_{\beta,N}/\mathbb{E}Z_{\beta,N}$ converges to $C(\beta)$. The other factor $Z^T_{\beta,N}/\mathbb{E}Z^T_{\beta,N}$ in the first term of (2.7) converges to 1 in probability. In fact, it can be estimated as the sum (2.9) where the terms over j with $A_{1,j} < A_{1,J_1}$ remain the same and the terms over j with $A_{1,j} = A_{1,J_1}$ become:

$$(\alpha_1 \cdots \alpha_j)^{-N} e^{\beta^2 N(a_1+\cdots+a_j)} \mathbb{P}\Big(Z < u_{\ln(\alpha_1\cdots\alpha_j),N}(0) - 2\beta\sqrt{N(a_j+\cdots+a_n)}\Big)$$
$$\leq \frac{(\alpha_1 \cdots \alpha_j)^{-N} e^{\beta^2 N(a_1+\cdots+a_j)} e^{-(u_{\ln(\alpha_1\cdots\alpha_j),N}(0)-2\beta\sqrt{N(a_j+\cdots+a_n)})^2/2}}{\sqrt{2\pi}(2\beta\sqrt{N(a_j+\cdots+a_n)} - u_{\ln(\alpha_1\cdots\alpha_j),N}(0))} = O(1/N), \quad N \uparrow \infty$$

in view of the equality $\beta^2 A_{1,j} = 1$. The second term in (2.7) does not converge to zero in L_1 in this case, as its expectation is about $(1 - C(\beta))$. Nevertheless, we show that it converges to zero in probability. Taking into account (1.11) with $\beta^2 A_{1,j} = 1$, one can estimate it by

$$\frac{\bar Z^T_{\beta,N}}{\mathbb{E}Z_{\beta,N}} \leq \sum_{\substack{j=1,\\ A_{1,j}<A_{1,J_1}}}^{n} 2^{-N} e^{-\beta^2 N/2} \sum_{\sigma_1} e^{\beta\sqrt{a_1}X_{\sigma_1}} \cdots \sum_{\sigma_j} e^{\beta\sqrt{a_1}X_{\sigma_1\sigma_2\cdots\sigma_j}}$$
$$\times \mathbb{1}_{\{\sqrt{a_1}X_{\sigma_1}+\cdots+\sqrt{a_j}X_{\sigma_1\cdots\sigma_j} \geq (\beta+\epsilon)(a_1+\cdots+a_j)\sqrt{N}\}} \sum_{\sigma_{j+1}} e^{\beta\sqrt{a_{j+1}}X_{\sigma_1\cdots\sigma_{j+1}}} \cdots \sum_{\sigma_n} e^{\beta\sqrt{a_n}X_{\sigma_1\sigma_2\cdots\sigma_n}}$$
$$+ \sum_{\substack{j=1,\\ A_{1,j}=A_{1,J_1}}}^{n} e^{-[\ln N + \ln\ln(\alpha_1\cdots\alpha_j)+\ln 4\pi]/2} \sum_{\sigma_1\ldots\sigma_j} e^{u^{-1}_{\ln(\alpha_1\cdots\alpha_j),N}(\bar X_{\sigma_1\cdots\sigma_j})} \mathbb{1}_{\{u^{-1}_{\ln(\alpha_1\cdots\alpha_j),N}(\bar X_{\sigma_1\cdots\sigma_j})>0\}}$$
$$(\alpha_{j+1}\cdots\alpha_n)^{-N} e^{-\beta^2(a_{j+1}+\cdots+a_n)N/2} \sum_{\sigma_{j+1}} e^{\beta\sqrt{a_{j+1}}X_{\sigma_1\cdots\sigma_{j+1}}} \cdots \sum_{\sigma_n} e^{\beta\sqrt{a_n}X_{\sigma_1\sigma_2\cdots\sigma_n}}. \tag{2.10}$$

The first sum in this estimate over j with $A_{1,j} < A_{1,J_1}$ converges to zero in L_1 as the mean of its jth term is of the order $e^{-\epsilon^2(a_1+\cdots+a_j)N/2}$. In the jth term of the last sum we have for all $i = 1, 2, \ldots, j$ $A_{1,i} \leq A_{1,J_1} = A_{1,j}$ by Proposition 1.4 from [BK1]. This means that inequalities ensuring the convergence to the Poisson point process (see (1.6) of Theorem 1.1 in [BK1]) are verified where $a_i/\sum_{i=1}^j a_i$ substitute a_i, $\bar\alpha = \prod_{i=1}^j \alpha_i$, and Theorem 1.1 of [BK1] applies. Therefore, the point process $\sum_{\sigma_1,\ldots\sigma_j} \delta_{u^{-1}_{\ln(\alpha_1\cdots\alpha_j),N}(\bar X_{\sigma_1\cdots\sigma_j})}$ converges weakly to the Poisson point process on $\mathbb{R}$ with the intensity measure $e^{-x}dx$ having almost surely a finite number of points in $[0,\infty)$. Then for any $\epsilon > 0$ one can find $K(\epsilon)$ and $R(\epsilon)$ such that the number of its points in $[0,\infty)$ is smaller than $K(\epsilon)$ and each of them is smaller than $R(\epsilon)$ with probability at least $1-\epsilon$ for all N large enough. The sum over $\sigma_{j+1}, \ldots, \sigma_n$ attached to each point of this process is the properly normalised partition function of the GREM with the right partition (1.12) $J_1 - j, J_2, \ldots, J_m$ and $l(\beta) = 0$. If $J_1 - j = 0$, it converges to a constant by the statement of the lemma just proved in the previous case as $\beta^2 A_{J_1+1,J_2} < \beta^2 A_{1,J_1} = 1$. Otherwise, start induction procedure on the size of the first cell. Then this sum over $\sigma_{j+1}, \ldots, \sigma_n$ with the first cell smaller than J_1 converges to a constant by induction assumption. It follows from these arguments that the whole sum over $\sigma_1, \ldots, \sigma_j$

is bounded in probability: for any $\epsilon > 0$ there exits $M(\epsilon)$ that it is bigger than $M(\epsilon)$ with probability smaller than ϵ for all N sufficiently large. Finally the prefactor $e^{-\ln N/2}$ make it converge in probability to zero. This proves the convergence in probability to zero of the second term in (2.7) and completes the proof of the lemma. $\diamondsuit$.

Lemma 2.2: *Let $0 < a_i < 1$, $\alpha_i > 1$, $\gamma_i = \sqrt{a_i/(2\ln\alpha_i)}$, $i = 1, 2, \dots, m$ be the numbers such that $\gamma_1 > \cdots > \gamma_m$, $\beta\gamma_m > 1$. Let $X_1, \dots, X_m$ be independent standard Gaussian random variables. Then for any $x \in \mathbb{R}$ there exists $N(x)$ such that for all $N \geq N(x)$*

$$\begin{aligned}&\mathbb{E}e^{\beta\gamma_1 u^{-1}_{\ln\alpha_1,N}(X_1)+\cdots+\beta\gamma_m u^{-1}_{\ln\alpha_m,N}(X_m)}\mathbb{1}_{\{\forall i=1,\dots,m:\gamma_1 u^{-1}_{\ln\alpha_1,N}(X_1)+\cdots+\gamma_i u^{-1}_{\ln\alpha_i,N}(X_i)<(\gamma_1+\cdots\gamma_i)x\}}\\ &\leq C(\alpha_1\cdots\alpha_m)^{-N}e^{x\sum_{j=1}^m(\beta\gamma_j-1)}\end{aligned} \tag{2.11}$$

where the constant C does not depend on x.

Proof. The r.h.s. of (2.11) equals

$$\begin{aligned}&e^{\sum_{j=1}^m\left(-\beta N\sqrt{2a_j\ln\alpha_j}+\beta\gamma_j[\ln(N\ln\alpha_j)+\ln 4\pi]/2\right)}\\ &\qquad\times\mathbb{E}e^{\beta\sqrt{a_1}X_1+\cdots+\beta\sqrt{a_m}X_m}1_{\forall i=1,\dots,m:\sqrt{a_1}X_1+\cdots+\sqrt{a_i}X_i<\sqrt{a_1}u_{\ln\alpha_1,N}(x)+\cdots+\sqrt{a_i}u_{\ln\alpha_i,N}(x)}\\ &=e^{\sum_{j=1}^m\left(-\beta N\sqrt{2a_j\ln\alpha_j}+\beta\gamma_j[\ln(N\ln\alpha_j)+\ln 4\pi]/2+N\beta a_j^2/2\right)}\\ &\qquad\times\mathbb{P}\Big(\forall i=1,\dots,m:\sqrt{a_1}X_1+\cdots+\sqrt{a_i}X_i\\ &\qquad\qquad<\sqrt{a_1}(u_{\ln\alpha_1,N}(x)-\beta\sqrt{a_1N})+\cdots+\sqrt{a_i}(u_{\ln\alpha_i,N}(x)-\beta\sqrt{a_iN})\Big).\end{aligned} \tag{2.12}$$

Now Proposition 2.3 below will be applied with

$$c_i(N) = u_{\ln\alpha_i,N}(x) - \beta\sqrt{a_iN} = (\sqrt{2\ln\alpha_i} - \beta\sqrt{a_i})\sqrt{N} + o(1), \quad i = 1, 2, \dots, m$$

where $c_i = \sqrt{2\ln\alpha_i} - \beta\sqrt{a_i} < 0$ as $\beta\gamma_i > \beta\gamma_m > 1$ and $\sqrt{a_{i-1}}c_i - \sqrt{a_i}c_{i-1} > 0$ as $\gamma_{i-1} > \gamma_i$. Thus (2.12) does not exceed

$$\begin{aligned}&(\sqrt{N})^{-m}\tilde{C}e^{\sum_{j=1}^m\left(-\beta N\sqrt{2a_j\ln\alpha_j}+\beta\gamma_j[\ln(N\ln\alpha_j)+\ln 4\pi]/2+N\beta a_j^2/2-(u_{\ln\alpha_j,N}(x)-\beta\sqrt{a_jN})^2/2\right)}\\ &=(\sqrt{N})^{-m}\tilde{C}e^{\sum_{j=1}^m\left(-u^2_{\ln\alpha_j,N}(x)/2+\beta\gamma_j x\right)}\leq C(\alpha_1\cdots\alpha_m)^{-N}e^{x\sum_{j=1}^m(\beta\gamma_j-1)}\end{aligned}$$

for all N large enough. $\diamondsuit$

Proposition 2.3: *Assume that the numbers $0 < a_i < 1$, $c_i < 0$, $i = 1, \dots, m$, are such that $\sqrt{a_{i-1}}c_i - \sqrt{a_i}c_{i-1} > 0$ for all $i = 2, \dots, m$. Let $c_i(N) = c_i\sqrt{N} + o(1)$ as $N \uparrow \infty$ be* *the functions of N with leading constants c_i. Let also $X_1,\dots,X_m$ be independent standard Gaussian random variables. Then for all sufficiently large N*

$$\begin{aligned}&\mathbb{P}\Big(\forall i=1,\dots,m: \sqrt{a_1}X_1+\cdots+\sqrt{a_i}X_i<\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_i}c_i(N)\Big)\\&\le(\sqrt{N})^{-m}\tilde{C}e^{-\sum_{j=1}^{N}c_j^2(N)/2}\end{aligned}\tag{2.13}$$

where the constant $\tilde{C}$ depends only on the constants $c_1,\dots,c_m$.

Proof. It is a rather cumbersome computation where several induction procedures are involved. It is based on the elementary estimate for the density of standard random variables: for any $x>0$

$$\Big(\frac{1}{x}-\frac{1}{x^3}\Big)e^{-x^2/2}\le\int_x^\infty e^{-t^2/2}\le\frac{1}{x}e^{-x^2/2}\tag{2.14}$$

In the case $m=1$ (2.13) follows directly from (2.14). Assume that it is proved to $m-1$ and let us make an induction step to m. Denote by

$$D_m=\{(x_1,\dots,x_m)\in\mathbb{R}^m:\ \forall i=1,\dots,m\ \sqrt{a_1}X_1+\cdots+\sqrt{a_i}X_i<\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_i}c_i(N)\}.$$

For any $\delta>0$ by (2.14) we may write:

$$\begin{aligned}&\mathbb{P}\Big(\forall i=1,\dots,m: \sqrt{a_1}X_1+\cdots+\sqrt{a_i}X_i<\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_i}c_i(N)\Big)\\&\le\frac{\sqrt{a_m}}{\sqrt{(2\pi)^mN}\delta}\int_{D_{m-1}}e^{(\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_m}c_m(N)-\sqrt{a_1}x_1-\cdots-\sqrt{a_{m-1}}x_{m-1})^2/(2a_m)}\\&\qquad\qquad\times e^{-(x_{m-1}^2+\cdots+x_1^2)/2}dx_{m-1}\dots dx_1\\&+\mathbb{P}\Big(\forall i=1,\dots,m-1: \sqrt{a_1}X_1+\cdots+\sqrt{a_i}X_i<\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_i}c_i(N)\\&\qquad\sqrt{a_1}X_1+\cdots+\sqrt{a_{m-1}}X_{m-1}<\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_{m-1}}c_{m-1}(N)+\sqrt{a_m}c_m(N)+\delta\sqrt{N}\Big)\\&\equiv I_{N,m}^1+I_{N,m}^2.\end{aligned}$$

Let us first show that with $\delta>0$ small enough $I_{N,m}^2$ is of exponentially smaller order than the bound (2.13). For this purpose we find k such that

$$\begin{aligned}&(a_{m-1}+\cdots+a_{m-i+1})c_{m-i}\ge\sqrt{a_{m-i}}(\sqrt{a_{m-i+1}}c_{m-i+1}+\cdots+\sqrt{a_m}c_m)\quad\forall i=2,3,\dots,k;\\&(a_{m-1}+\cdots+a_{m-k})c_{m-k-1}<\sqrt{a_{m-k-1}}(\sqrt{a_{m-k}}c_{m-k}+\cdots+\sqrt{a_m}c_m).\end{aligned}\tag{2.15}$$

(If (2.15) fails already for $i=2$ we put $k=1$, if (2.15) holds up to $i=m-1$, we put $k=m-1$.) Then the term $I_{N,m}^2$ is bounded by the probability of the following event for which the estimate (2.13) applies by induction:

$$\begin{aligned}
&\mathbb{P}\Big(\forall i=1,\dots,m-k-1:\sqrt{a_1}X_1+\cdots+\sqrt{a_i}X_i<\sqrt{a_1}(c_1(N)+\delta\sqrt{N}/\sqrt{a_1})+\cdots+\sqrt{a_i}c_i(N)\\
&\quad\sqrt{a_1}X_1+\cdots+\sqrt{a_{m-k-1}}X_{m-k-1}+\sqrt{a_{m-k}+\cdots+a_{m-1}}X_{m-k}\\
&\qquad<\sqrt{a_1}(c_1(N)+\delta\sqrt{N}/\sqrt{a_1})+\cdots+\sqrt{a_{m-1}}c_{m-1}(N)+\sqrt{a_m}c_m(N)\Big)\\
&\le\frac{\tilde{C}}{(\sqrt{N})^{m-k}}e^{-[(c_1(N)+\delta\sqrt{N}/\sqrt{a_1})^2+\cdots+c^2_{m-k-1}(N)]/2}\\
&\qquad\times e^{-[\sqrt{a_{m-k}}c_{m-k}(N)+\cdots+\sqrt{a_m}c_m(N)]^2/[2(a_{m-k}+\cdots+a_{m-1})]}.
\end{aligned}\tag{2.16}$$

Now it suffices to note that

$$(\sqrt{a_{m-k}}c_{m-k}+\cdots+\sqrt{a_m}c_m)^2>(a_{m-k}+\cdots+a_{m-1})(c^2_{m-k}+\cdots+c^2_m).\tag{2.17}$$

This is verified for $k=1$ from $\sqrt{a_{m-1}}c_m>\sqrt{a_m}c_{m-1}$ and for $k>1$ by induction from $k-1$:

$$\begin{aligned}
&(\sqrt{a_{m-k}}c_{m-k}+\cdots+\sqrt{a_m}c_m)^2\\
&>a_{m-k}c^2_{m-k}+(\sqrt{a_{m-k+1}}c_{m-k+1}+\cdots+\sqrt{a_m}c_m)^2\\
&\qquad+2\sqrt{a_{m-k}}c_{m-k}(\sqrt{a_{m-k+1}}c_{m-k+1}+\cdots+\sqrt{a_m}c_m)\\
&>a_{m-k}c^2_{m-k}+(a_{m-k+1}+\cdots+a_{m-1})(c^2_{m-k+1}+\cdots+c^2_m)\\
&\qquad+c^2_{m-k}(a_{m-k+1}+\cdots+a_{m-1})+a_{m-k}(c^2_{m-k+1}+\cdots+c^2_m)\\
&=(a_{m-k}+\cdots+a_{m-1})(c^2_{m-k}+\cdots+c^2_m)
\end{aligned}\tag{2.18}$$

where we used (2.15) with $i=k$ and the assumptions on c_i and a_i of the proposition. Note also that by (2.17) and (2.18) the choice of an appropriate $\delta>0$ in (2.16) depends only on c_i, a_i for $i=k,\dots,m$.

Next, let us concentrate on the analysis of $I^1_{N,m}$. We start another induction procedure to prove that it obeys the bound (2.13) under assumptions of the proposition. It is not difficult to check this for $m=2$ and we skip it. Assume that this holds true up to $m-1$ and consider the case of m. Let us write $I^1_{N,m}$ as

$$\begin{aligned}
I^1_{N,m}&=\frac{\sqrt{a_m}}{\sqrt{(2\pi)^m N}\delta}\int_{D_{m-2}}e^{\frac{[\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_m}c_m(N)-\sqrt{a_1}x_1-\cdots-\sqrt{a_{m-2}}x_{m-2}]^2}{2(a_{m-1}+a_m)}}\\
&\qquad\times e^{-(x^2_{m-2}+\cdots+x^2_1)/2}dx_{m-2}\dots dx_1\\
&\int_{-\infty}^{E_{m-1}}e^{\frac{[(a_{m-1}+a_m)x_{m-1}-\sqrt{a_{m-1}}(\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_m}c_m(N)-\sqrt{a_1}x_1-\cdots-\sqrt{a_{m-2}}x_{m-2})]^2}{2a_m(a_{m-1}+a_m)}}dx_{m-1}
\end{aligned}$$

with the notation

$$E_{m-1} \equiv (\sqrt{a_1}c_1(N) + \cdots + \sqrt{a_{m-1}}c_{m-1}(N) - \sqrt{a_1}x_1 - \cdots - \sqrt{a_{m-2}}x_{m-2})/\sqrt{a_{m-1}}.$$

Whenever $\sqrt{a_1}x_1+\cdots+\sqrt{a_{m-2}}x_{m-2} > \sqrt{a_1}c_1(N)+\cdots+\sqrt{a_{m-2}}c_{m-2}(N)+\sqrt{a_{m-1}}c_{m-1}(N)-a_{m-1}(c_m(N)-\delta'\sqrt{N})/\sqrt{a_m}$ for any $\delta' > 0$, we are allowed to apply to the inner integral the estimate (2.14), otherwise let us just estimate it by $\sqrt{2\pi}$:

$$\begin{aligned} I^1_{N,m} \leq & \frac{\sqrt{a_m a_{m-1}}e^{-c_m^2(N)/2}}{\sqrt{(2\pi)^m N^2(a_{m-1}+a_m)\delta\delta'}} \int_{D_{m-2}} e^{\frac{[\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_{m-1}}c_{m-1}(N)-\sqrt{a_1}x_1-\cdots-\sqrt{a_{m-2}}x_{m-2}]^2}{2a_{m-1}}} \\ & \qquad\qquad \times e^{-(x^2_{m-2}+\cdots+x_1^2)/2}dx_{m-2}\ldots dx_1 \\ & + \frac{\sqrt{a_m}}{\sqrt{(2\pi)^{m-1}}N\delta} \int_{D_{m-3}\cap G^{\delta'}_{m,2}} e^{\frac{[\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_m}c_m(N)-\sqrt{a_1}x_1-\cdots-\sqrt{a_{m-2}}x_{m-2}]^2}{2(a_{m-1}+a_m)}} \\ & \qquad\qquad \times e^{-(x^2_{m-2}+\cdots+x_1^2)/2}dx_{m-2}\ldots dx_1 \\ & \equiv J^1_{N,m} + J^2_{N,m} \end{aligned} \tag{2.19}$$

where

$$\begin{aligned} G^{\delta'}_{m,2} = \Big\{ & (x_1,\ldots,x_m) : \sqrt{a_1}x_1 + \cdots + \sqrt{a_2}x_{m-2} \\ & < \sqrt{a_1}c_1(N) + \cdots + \sqrt{a_{m-2}}c_{m-2}(N) + \sqrt{a_{m-1}}c_{m-1}(N) - a_{m-1}c_m(N)/\sqrt{a_m} + \delta'\sqrt{N}\Big\}. \end{aligned}$$

The term $J^1_{N,m}$ has the form of $I^1_{N,m-1}$ with parameters a_i, c_i,$i=1,\ldots,m-1$. By induction on $I^1_{N,m}$ it obeys the bound claimed in (2.13). In fact, this term brings the main contribution to $I^1_{N,m}$ and to the probability of (2.13). We will show under assumptions of the proposition that with an appropriate choice of $\delta' > 0$ depending only on a_i and c_i, $i=2,\ldots,m$, the contribution of $J^2_{N,m}$ is of exponentially smaller order than the bound of (2.13). Observe that with notations $a'_i = a_i$, $c'_i = c_i$ for $i=1,\ldots,m-3$, $a'_{m-2} = a_{m-2}$, $c'_{m-2} = c_{m-2} + (\sqrt{a_{m-1}}c_{m-1} - a_{m-1}c_m/\sqrt{a_m})/\sqrt{a_{m-2}}$, $a'_{m-1} = a_{m-1} + a_m$, $c'_{m-1} = \sqrt{a_{m-1}+a_m}c_m/\sqrt{a_m}$, $J^2_{N,m}$ takes the form of $I^1_{N,m-1}$ as well. Had the assumptions $\sqrt{a'_i}c'_{i-1} - \sqrt{a'_i}c'_i < 0$ for $i=2,\ldots,m-1$ been verified, then by induction on $I^1_{N,m}$ this term could be estimated by a good bound upper $e^{-({c'}^2_1(N)+\cdots+{c'}^2_{m-1}(N))/2}$ as ${c'}^2_1+\cdots+{c'}^2_{m-1} > c_1^2+\cdots+c_m^2$ and the proof of the proposition would have been finished. Unfortunately the conditions $\sqrt{a'_i}c'_{i-1} - \sqrt{a'_i}c'_i < 0$ are verified for all $i=1,\ldots,m-1$ but not necessarily for $i=m-2$. If for $i=m-2$ this inequality fails, then we are obliged to continue the proof. This leads us to showing by induction subsequently for $k=m-1,\ldots,2$ that under assumptions $\sqrt{a_i}c_{i-1} - \sqrt{a_i}c_i < 0$,

$i = 2, \dots, m$, for all N large enough

$$L_{N,m}^k = \int_{D_{m-k-1} \cap G_{m,k}} e^{\frac{[\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_m}c_m(N)-\sqrt{a_1}x_1-\cdots-\sqrt{a_{m-k}}x_{m-k}]^2}{2(a_{m-k+1}+\cdots+a_{m-1})}} \\ e^{-(x_{m-k}^2+\cdots+x_1^2)/2} dx_{m-k} \dots dx_1 \le e^{-BN(1+o(1))} \tag{2.20}$$

where $G_{m,k} = \{(x_1, \dots, x_m) : \sqrt{a_1}x_1 + \cdots + \sqrt{a_{m-k}}x_{m-k} < \sqrt{a_1}c_1(N) + \cdots + \sqrt{a_{m-k}}c_{m-k}(N) + \cdots + \sqrt{a_{m-1}}c_{m-1}(N) - (a_{m-k+1} + \cdots + a_{m-1})c_m(N)/\sqrt{a_m}\}$ and where the constant $B > c_1^2 + \cdots + c_m^2$. Then the estimate (2.20) in the case $k = 2$ proves that with an appropriate choice of $\delta' > 0$ the contribution of $J_{N,m}^2$ to (2.13) is negligible and concludes the proof of the proposition.

Finally we are left to show (2.20). Let us introduce the notations: $a_i' \equiv a_i$, $c_i' \equiv c_i$ for $i = 1, \dots, m-k-1$, $a_{m-k}' \equiv a_{m-k}$,

$$c_{m-k}' \equiv c_{m-k} + (\sqrt{a_{m-k+1}}c_{m-k+1} + \cdots + a_{m-1}c_{m-1} - (a_{m-k+1} + \cdots + a_{m-1})c_m/\sqrt{a_m})/\sqrt{a_{k-m}},$$

$$a_{m-k+1}' \equiv a_{m-k+1} + \cdots + a_m,$$

$$c_{m-k+1}' \equiv \sqrt{a_{m-k+1} + \cdots + a_m}\, c_m/\sqrt{a_m}.$$

One checks that $\sqrt{a_i'}c_{i-1}' - \sqrt{a_i'}c_i' < 0$ for all $i = 2, \dots, m-k+1$ except may be for $i = m-k$. If this is also true for $i = m-k$ or if $k = m-1$ (as in this case a_{m-k-1}', c_{m-k-1}' are absent) then $L_{N,m}^k$ takes the form of $I_{N,m-k+1}^1$ and by induction on $I_{N,m}^1$ it does not exceed the exponential bound (2.20) with $B = {c'}_1^2 + \cdots + {c'}_{m-k+1}^2$. Then the observation

$$\begin{aligned} &{c'}_1^2 + \cdots + {c'}_{m-k+1}^2 - (c_1^2 + \cdots + c_m^2) \\ &\ge 2c_{m-k} \sum_{i=1}^{k-1} (\sqrt{a_{m-i}}c_{m-i} - a_{m-i}c_m/\sqrt{a_m})/\sqrt{a_{m-k}} + \sum_{i=1}^{k-1} a_{m-i}c_m^2/a_m - \sum_{i=1}^{k-1} c_{m-i}^2 \\ &> c_{m-k} \sum_{i=1}^{k-1} (\sqrt{a_{m-i}}c_{m-i} - a_{m-i}c_m/\sqrt{a_m})/\sqrt{a_{m-k}} + \sum_{i=1}^{k-1} a_{m-i}c_m^2/a_m \\ &\qquad\qquad - \sum_{i=1}^{k-1} \sqrt{a_{m-i}}c_m c_{m-i}/\sqrt{a_m} \\ &= \sum_{i=1}^{k-1} (\sqrt{a_{m-i}}c_{m-i} - a_{m-i}c_m/\sqrt{a_m})(c_{m-k}/\sqrt{a_{m-k}} - c_m/\sqrt{a_m}) > 0 \end{aligned} \tag{2.21}$$

finishes the proof of (2.20) in this case. In particular this proves (2.20) for $k = m-1$.

Suppose now that $k > 1$, (2.20) holds up to $k-1$ and that $\sqrt{a'_{m-k-1}}c'_{m-k} \le \sqrt{a'_{m-k}}c'_{m-k-1}$. Let us find $l \ge k$ such that

$$(a'_{m-k} + \cdots + a'_{m-i})c'_{m-i-1} \ge \sqrt{a'_{m-i-1}}(\sqrt{a'_{m-i}}c'_{m-i} + \cdots + \sqrt{a'_{m-k}}c'_{m-k}) \quad \forall i = 2, 3, \dots, l; \tag{2.22}$$

$$(a'_{m-k} + \cdots + a'_{m-l-1})c'_{m-l-2} < \sqrt{a'_{m-l-2}}(\sqrt{a'_{m-l-1}}c'_{m-l-1} + \cdots + \sqrt{a'_{m-k}}c'_{m-k}). \tag{2.23}$$

(If (2.22) holds for $i = 1, \dots, m-2$, we put $l = m-2$.) Then developing $L^k_{N,m}$ $l-k+2$ times as in (2.19), we obtain

$$\begin{aligned} L^k_{N,m} \le & \sqrt{2\pi} \int_{D_{m-k-2} \cap G^{\delta''}_{m,k-1}} e^{\frac{[\sqrt{a_1}c_1(N) + \cdots + \sqrt{a_m}c_m(N) - \sqrt{a_1}x_1 - \cdots - \sqrt{a_{m-k-1}}x_{m-k-1}]^2}{2(a_{m-k} + \cdots + a_m)}} \\ & \qquad \times e^{-(x^2_{m-k-1} + \cdots + x^2_1)/2} dx_{m-k-1} \dots dx_1 \\ & + \frac{\sqrt{(2\pi)^{l-k+1}(a_{m-k+1} + \cdots + a_m)} e^{-c'^2_{m-k+1}(N)/2}}{\sqrt{N}\delta''} \\ & \times \int_{D'_{m-l-2}} e^{\frac{[\sqrt{a'_1}c_1(N) + \cdots + \sqrt{a'_{m-k}}c'_{m-k}(N) - \sqrt{a'_1}x_1 - \cdots - \sqrt{a'_{m-l-2}}x_{m-l-2}]^2}{2(a'_{m-l-1} + \cdots + a'_{m-k})}} \\ & \qquad \times e^{-(x^2_{m-l-2} + \cdots + x^2_1)/2} dx_{m-l-2} \dots dx_1 \end{aligned} \tag{2.24}$$

where $G^{\delta''}_{m,k-1} = \{(x_1, \dots, x_{m-k-1}) : \sqrt{a_1}x_1 + \cdots + \sqrt{a_{m-k-1}}x_{m-k-1} < \sqrt{a_1}c_1(N) + \cdots + \sqrt{a_{m-k-1}}c_{m-k-1}(N) + \cdots + \sqrt{a_{m-1}}c_{m-1}(N) - (a_{m-k} + \cdots + a_{m-1})c_m(N)/\sqrt{a_m} + \delta''\sqrt{N}\}$ and where D'_{m-l-2} is the domain D_{m-l-2} with a_i, c_i formally replaced by a_i, c'_i (In fact, $D'_{m-l-2} = D_{m-l-2}$ as $a_i = a'_i$, $c_i = c'_i$ for $i = 1, \dots, m-l-2$). The first term of (2.24) with $\delta'' > 0$ small enough obeys the bound (2.20) with a good constant $B > c_1^2 + \cdots + c_m^2$ by induction on $L^k_{N,m}$. The second term has the form of $I^1_{N,m-l+1}$ with parameters satisfying the assumptions of the proposition in view of the choice of l according to (2.23). Then by induction on $I^1_{N,m}$ this term obeys the bound (2.20) with

$$B = c'^2_1 + \cdots + c'^2_{m-l-2} + (\sqrt{a'_{m-l-1}}c'_{m-l-1} + \cdots + \sqrt{a'_{m-k}}c'_{m-k})^2/(a'_{m-l-1} + \cdots + a'_{m-k}) + c'^2_{m-k+1}.$$

It remains to show that $B - c_1^2 - \cdots - c_m^2 > 0$. This can be proved by induction on l (the last one !) starting from (2.21). Then

$$
\begin{aligned}
&(a'_{m-l-1}+\cdots+a'_{m-k})(B-c_1^2-\cdots-c_m^2)\\
&>\sqrt{a_{m-l-1}}c_{m-l-1}\Big(\sqrt{a_{m-l}}c_{m-l}+\cdots+\sqrt{a_{m-1}}c_{m-1}\\
&\qquad\qquad -\frac{a_{m-k+1}+\cdots+a_{m-1}}{\sqrt{a_m}}c_m-\frac{a_{m-l}+\cdots+a_{m-k}}{\sqrt{a_{m-l-1}}}c_{m-l-1}\Big)\\
&+\sum_{j=1}^{l}\sqrt{a_{m-l-1}}c_{m-j}(\sqrt{a_{m-j}}c_{m-l-1}-\sqrt{a_{m-l-1}}c_{m-j})\\
&+\sum_{r=1}^{k}\sqrt{a_{m-l-1}}\frac{a_{m-r}}{a_m}c_m(\sqrt{a_{m-l-1}}c_m-\sqrt{a_m}c_{m-l-1}).
\end{aligned}
\tag{2.25}
$$

The factor at c_{m-l-1} in round brackets in the second line being non-positive by (2.22) for $i=l$, let us estimate $c_{m-l-1}<\sqrt{a_{m-l-1}}c_{m-l}/\sqrt{a_{m-l}}$ and after sum this line with the term for $j=l$ in the sum over j. The resulting factor at c_{m-l} being non-positive by (2.22) for $i=l-1$, we may estimate $c_{m-l}<\sqrt{a_{m-l}}c_{m-l+1}/\sqrt{c_{m-l+1}}$ and sum this term with the one for $j=l-1$. This operation can be repeated for $j=l,\dots,k$. From the other side the factors at c_m in the last line of (2.25) being positive, we may estimate $c_m>\sqrt{a_m}c_{m-r}/\sqrt{a_{m-r}}$ for $r=1,\dots,k-1$ and sum these terms with those in the second line. Thus (2.25) is bigger than

$$
\begin{aligned}
&\sqrt{a_{m-l-1}}c_{m-k}\Big(\sqrt{\frac{a_{m-l-1}a_{m-k-1}}{a_{m-k}}}c_{m-k-1}+\cdots+\sqrt{\frac{a_{m-l-1}a_{m-1}}{a_{m-k}}}c_{m-1}\\
&\qquad\qquad -\frac{a_{m-1}+\cdots+a_{m-k+1}}{\sqrt{a_m}}\sqrt{\frac{a_{m-l-1}}{a_{m-k}}}c_m\Big)\\
&+\sum_{i=1}^{k-1}\sqrt{a_{m-l-1}}c_{m-i}\Big(\sqrt{\frac{a_{m-l-1}a_{m-i}}{a_m}}c_m-\sqrt{a_{m-l-1}}c_{m-i}\Big)\\
&>\sqrt{a_{m-l-1}}c_{m-k}\Big(\sqrt{\frac{a_{m-l-1}a_{m-k-1}}{a_{m-k}}}c_{m-k-1}+\cdots+\sqrt{\frac{a_{m-l-1}a_{m-1}}{a_{m-k}}}c_{m-1}\\
&\qquad\qquad -\frac{a_{m-1}+\cdots+a_{m-k+1}}{\sqrt{a_m}}\sqrt{\frac{a_{m-l-1}}{a_{m-k}}}c_m\Big)\\
&+c_{m-k}\sum_{i=1}^{k-1}\sqrt{a_{m-l-1}}\sqrt{\frac{a_{m-i}}{a_{m-k}}}\Big(\sqrt{\frac{a_{m-l-1}a_{m-i}}{a_m}}c_m-\sqrt{a_{m-l-1}}c_{m-i}\Big)=0.
\end{aligned}
$$

This finishes the proof of (2.20) and completes the proof of the proposition. $\diamondsuit$

Proof of Theorem 1.5: The proof of Theorem 1.5 is not much more than a corollary of Theorem 1.3. We have to show that for any collection of continuous, local functions

$h_1, \dots, h_j$, $j \in \mathbb{N}$, and a continuous function $G : \mathbb{R}^j \to \mathbb{R}$

$$\begin{aligned} &\mathbb{E} G\left(\int \mathcal{W}^m_{\beta,N}(dw) h_1(w), \dots, \int \mathcal{W}^m_{\beta,N}(dw) h_j(w) \right) \\ &\to \mathbb{E} F\left(\int \mathcal{W}^m_{\beta}(dw) h_1(w), \dots, \int \mathcal{W}^m_{\beta}(dw) h_j(w) \right) \end{aligned} \tag{2.26}$$

Now clearly

$$\begin{aligned} &\int \mathcal{W}^k_{\beta,N}(dw) h_i(w) \\ &= \int \mathcal{P}^{(k)}_N(dy) h_i \left(\frac{\int \mathcal{P}^{(k)}_N(dx) e^{\beta(\gamma,x)} \delta(x_1 - y_1)}{\int \mathcal{P}^{(k)}_N(dx) e^{\beta(\gamma,x)}}, \dots, \frac{\int \mathcal{P}^{(k)}_N(dx) e^{\beta(\gamma,x)} \delta(x_1 - y_1) \dots \delta(x_k - y_k)}{\int \mathcal{P}^{(k)}_N(dx) e^{\beta(\gamma,x))}} \right) \end{aligned} \tag{2.27}$$

We will use the fact that convergence in distribution of a sequence $\mathcal{P}^{(k)}_N$ is equivalent to almost sure convergence of a sequence $\mathcal{P}^{(k,*)}_N$ that for each N have the same distribution as $\mathcal{P}^{(k)}_N$ (see e.g. [Ka,RE]). We can of course replace $\mathcal{P}^{(k)}_N$ by $\mathcal{P}^{(k,*)}_N$ in (2.26) without changing anything.

We have already shown that the denominators in the arguments of h_i converge in distribution, i.e. $\int \mathcal{P}^{(k,*)}_N(dx) e^{\beta(\gamma,x)}$ converges a.s. to $\int \mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)}$. The same argument applies to the numerators, i.e. for fixed x the vector

$$\begin{aligned} &\left(\frac{\int \mathcal{P}^{(k,*)}_N(dx) e^{\beta(\gamma,x)} \delta(x_1 - y_1)}{\int \mathcal{P}^{(k,*)}_N(dx) e^{\beta(\gamma,x)}}, \dots, \frac{\mathcal{P}^{(k,*)}_N(dx) e^{\beta(\gamma,x)} \delta(x_1 - y_1) \dots \delta(x_k - y_k)}{\int \mathcal{P}^{(k,*)}_N(dx) e^{\beta(\gamma,x)}} \right) \\ &\to \left(\frac{\int \mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)} \delta(x_1 - y_1)}{\int \mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)}}, \dots, \frac{\mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)} \delta(x_1 - y_1) \dots \delta(x_k - y_k)}{\int \mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)}} \right) \quad \text{a.s.} \end{aligned} \tag{2.28}$$

Moreover, this vector is a continuous function of x, which in turn implies that

$$h_i \left(\frac{\int \mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)} \delta(x_1 - y_1)}{\int \mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)}}, \dots, \frac{\mathcal{P}^{(k*)}(dx) e^{\beta(\gamma,x)} \delta(x_1 - y_1) \dots \delta(x_k - y_k)}{\int \mathcal{P}^{(k*)}(dx) e^{\beta(\gamma,x)}} \right) \tag{2.29}$$

is a continuous function of x, and thus

$$\begin{aligned} &\int \mathcal{P}^{k,*}_N(dx) h_i \left(\frac{\int \mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)} \delta(x_1 - y_1)}{\int \mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)}}, \dots, \frac{\mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)} \delta(x_1 - y_1) \dots \delta(x_k - y_k)}{\int \mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)}} \right) \\ &\to \int \mathcal{P}^{k,*}(x) h_i \left(\frac{\int \mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)} \delta(x_1 - y_1)}{\int \mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)}}, \dots, \frac{\mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)} \delta(x_1 - y_1) \dots \delta(x_k - y_k)}{\int \mathcal{P}^{(k,*)}(dx) e^{\beta(\gamma,x)}} \right) \end{aligned} \tag{2.30}$$

Thus all arguments of G in (2.27) converge (after replacing $\mathcal{P}^{(k)}_N$ by $\mathcal{P}^{(k,*)}_N$) almost surely. Since G is bounded and continuous, this implies convergence of the mean of G, which proves the theorem. $\diamondsuit$

3. The two-replica distribution.

We now turn to the analysis of the overlap distributions. Let us first look at the distribution of the ultrametric valuation, $f_{\beta,N}$, defined in (1.20).

We need some notation. For a given configuration $\sigma_1\sigma_2\dots\sigma_n$ we denote by $\bar\sigma_l = \sigma_{J_{l-1}+1}\cdots\sigma_{J_l}$ its pieces for $l=1,\dots,l(\beta)$ and $\bar\sigma_{l(\beta)+1} = \sigma_{J_{l(\beta)}+1}\dots\sigma_n$ in accordance with the appropriate partition (1.12) and (1.13). Note that the event $\{d_N(\sigma,\sigma')\geq q_i\}$ is equivalent to demanding that $\{\bar\sigma_1\dots\bar\sigma_i = \bar\sigma'_1\dots\bar\sigma'_i\}$. Therefore we have that

$$
\begin{aligned}
1-f_{\beta,N}(q_i) &= \sum_\sigma\sum_{\sigma'}\mu_{\beta,N}(\sigma)\mu_{\beta,N}(\sigma')\mathbb{1}_{\{\bar\sigma_1\dots\bar\sigma_i=\bar\sigma'_1\dots\bar\sigma'_i\}}\\
&=\int\mathcal{R}^i_{\beta,N}(dw_i)w_i^2=\int\mathcal{W}^i_{\beta,N}(dw_1,\dots,dw_i)w_i^2=\int\mathcal{W}^m_{\beta,N}(dw_1,\dots,dw_i)w_iw_m
\end{aligned}
\tag{3.1}
$$

for all $m\geq i$.

The convergence of the point processes $\mathcal{W}^i_{\beta,N}$ for $i\leq l(\beta)$ stated in Theorem 1.5 then immediately implies the convergence in distribution of

$$
f_{\beta,N}(q_i)\overset{\mathcal{D}}{\to}\int\widehat{\mathcal{W}}_\beta(dw_1,\dots,dw_{l(\beta)})w_{l(\beta)}(1-w_i).
\tag{3.2}
$$

for $i=1,\dots,l(\beta)$. To conclude the proof of Theorem 1.6, we must show that f_β is in fact atomic with atoms only at the values q_i, $i=1,\dots,l(\beta)$ and that the limiting overlap distribution function $\tilde f_\beta$ does the same.

For that purpose let us introduce a more complicated looking object. Consider the segments $\Delta_l = [-\ln\bar\alpha_l/\ln 2, \ln\bar\alpha_l/\ln 2]$ for $l=1,2,\dots,l(\beta)$ and
$\Delta_{l(\beta)+1} = [-\ln(\alpha_{J_{l(\beta)}+1}\cdots\alpha_n)/\ln 2, \ln(\alpha_{J_{l(\beta)}+1}\cdots\alpha_n)/\ln 2]$.
Clearly, $(\bar\sigma_l\cdot\bar\sigma'_l)/N\in\Delta_l$, $l=1,\dots,l(\beta)+1$. We introduce a random measure $f^{\otimes l(\beta)+1}_{\beta,N}$ on $\Delta_1\times\Delta_2\times\cdots\times\Delta_{l(\beta)+1}$ induced by the overlaps $\bar\sigma_l\cdot\bar\sigma'_l$ of all the 'levels' of the limiting Poisson cascade as follows: for any intervals $I_l\in\Delta_l$, $l=1,\dots,l(\beta)+1$, set

$$
\begin{aligned}
f^{\otimes l(\beta)+1}_{\beta,N}(I_1\times\cdots\times I_{l(\beta)+1}) &\equiv \frac{\mathbb{E}_{\sigma,\sigma'}\prod_{l=1}^{l(\beta)+1}\mathbb{1}_{(\bar\sigma_l\cdot\bar\sigma'_l)/N\in I_l}e^{\beta\sqrt{N}(X_\sigma+X_{\sigma'})}}{Z^2_{\beta,N}}\\
&=\mu^{\otimes 2}_{\beta,N}\Big(\prod_{l=1}^{l(\beta)+1}\mathbb{1}_{(\bar\sigma_l\cdot\bar\sigma'_l)/N\in I_l}\Big).
\end{aligned}
$$

The following theorem yields then all desired information on the overlap distribution and on the ultrametric distribution functions $\tilde f_\beta$ and f_β. In particular, it implies Theorem 1.6 readily.

Theorem 3.1: *The measure $f^{\otimes l(\beta)+1}_{\beta,N}$ converges in distribution to the following point measure on $\Delta_1 \times \Delta_2 \times \cdots \times \Delta_{l(\beta)+1}$*

$$f^{\otimes l(\beta)+1}_{\beta,N} \stackrel{\mathcal{D}}{\to} Q_0 \delta_{(0,0,\ldots,0)} + \sum_{j=1}^{l(\beta)} Q_j \delta_{(\ln \bar{\alpha}_1/\ln 2,\ldots,\ln \bar{\alpha}_j/\ln 2,0,\ldots,0)} \qquad N \to \infty. \tag{3.3}$$

The random variables $Q_1, \ldots, Q_{l(\beta)}$ are given as

$$Q_l(\beta) \equiv \int \widehat{\mathcal{W}}_\beta(dw_1, \ldots, dw_{l(\beta)}) w_{l(\beta)} (w_j - w_{j+1} \mathbb{1}_{\{j \le l(\beta)-1\}}), \quad j = 1, \ldots, l(\beta). \tag{3.4}$$

Proof. All we really need to establish is the support of the measure $f^{\otimes l(\beta)+1}_{\beta,N}$ in the limit $N \uparrow \infty$. The assertion (3.4) then follow from Theorem 1.5 like (3.2).

The next four assertions would provide us the support of the measure $f^{\otimes l(\beta)+1}_{\beta,N}$. We prove subsequently each of them:

(i) If $l(\beta) = 0$, then for any I such that $0 \notin \mathrm{cl}(I)$

$$f^{\otimes l(\beta)+1}_{\beta,N}(I) \stackrel{\mathcal{D}}{\to} 0 \quad \text{as } N \to \infty. \tag{3.5}$$

(ii) Let $l(\beta) \ge 1$. Then for any $l = 1, 2, \ldots, l(\beta)$ and any $I_l \in \Delta_l$ such that $1 \notin I_l$ and $0 \notin \mathrm{cl}\,(I_l)$

$$f^{\otimes l(\beta)+1}_{\beta,N}(\{1\} \times \cdots \times \{1\} \times I_l \times \Delta_{l+1} \times \cdots \times \Delta_{l(\beta)+1}) \stackrel{\mathcal{D}}{\to} 0. \tag{3.6}$$

(iii) Let $l(\beta) \ge 1$. Let I_{l-1} be such that $1 \notin I_{l-1}$. Then for any $l = 2, 3, \ldots, l(\beta)$, any $I_1, \ldots, I_{l-2}$ and any $I_l, \ldots, I_{l(\beta)+1}$ such that $(0, 0, \ldots, 0) \notin \mathrm{cl}(I_l \times \cdots \times I_{l(\beta)+1})$

$$f^{\otimes l(\beta)+1}_{\beta,N}(I_1 \times \cdots \times I_{l-2} \times I_{l-1} \times I_l \times \cdots \times I_{l(\beta)+1}) \stackrel{\mathcal{D}}{\to} 0. \tag{3.7}$$

(iv) Let $l(\beta) \ge 1$. Then for any $I_1, \ldots, I_{l(\beta)}$ and $I_{l(\beta)+1}$ such that $0 \notin \mathrm{cl}(I_{l(\beta)+1})$

$$f^{\otimes l(\beta)+1}_{\beta,N}(I_1 \times \cdots \times I_{l(\beta)+1}) \stackrel{\mathcal{D}}{\to} 0. \tag{3.8}$$

(cl$I$ denotes the closure if the subset I.) We start with (i) where $\beta^2 A_{1,J_1} \le 1$. Let us fix $\epsilon > 0$ such that

$$\ln(\alpha_1 \ldots \alpha_j) - (\beta^2 - (\beta - \epsilon)^2/2)(a_1 + \cdots + a_j) > 0 \tag{3.9}$$

for all $j = 1, \dots, n$ such that $\beta^2 A_{1,j} < 1$. Introduce the events

$$D_{\sigma_1\dots\sigma_j} = \{\omega : \sqrt{a_1} X_{\sigma_1} + \cdots + \sqrt{a_j} X_{\sigma_1\dots\sigma_j} < (\beta+\epsilon)(a_1 + \cdots + a_j)\sqrt{N}\}$$

if $\beta^2 A_{1,j} < 1$ (which is always true for all $j > J_1$, and if $\beta A_{1,J_1} < 1$, then for all $j \le J_1$ as well) and

$$D_{\sigma_1\dots\sigma_j} = \{\omega : \sqrt{a_1} X_{\sigma_1} + \cdots + \sqrt{a_j} X_{\sigma_1\dots\sigma_j} < u_{\ln(\alpha_1\cdots\alpha_j),N}(0)\sqrt{a_1 + \cdots + a_j}\}$$

if $\beta^2 A_{1,j} = 1$ (that can happen only if $\beta^2 A_{1,J_1} = 1$ and only for $j \le J_1$ such that $A_{1,j} = A_{1,J_1}$). We decompose

$$\begin{aligned} f^{l(\beta)+1}_{\beta,N}(I) =& \Big[2^{-2N} e^{-\beta^2 N} \sum_{\sigma,\sigma':\sigma\cdot\sigma'\in I} e^{\beta\sqrt{N}(X_\sigma + X_{\sigma'})} \mathbb{1}_{\{\cap_{j=1}^n (D_{\sigma_1\dots\sigma_j} \cap D_{\sigma'_1\dots\sigma'_j})\}}\Big] \frac{e^{\beta^2 N}}{(Z_{\beta,N})^2} \\ &+ \Big[2^{-N} e^{-\beta^2 N/2} Z^{-1}_{\beta,N} \sum_{\sigma,\sigma':\sigma\cdot\sigma'\in I} e^{\beta\sqrt{N}(X_\sigma + X_{\sigma'})} \mathbb{1}_{\{\cup_{j=1}^n (\bar{D}_{\sigma_1\dots\sigma_j} \cup \bar{D}_{\sigma'_1\dots\sigma'_j})\}}\Big] \frac{e^{\beta^2 N/2}}{Z_{\beta,N}} \\ \equiv& L^1_N + L^2_N. \end{aligned}$$

The second factors in L^1_N and L^2_N converge in probability to the constants $C^{-2}(\beta)$ and $C^{-1}(\beta)$ respectivelly by Theorem 1.6. The expectation of the first factor in square brackets of L^1_N can be estimated by (2.14) It does not exceed

$$\begin{aligned} &2^{-2N} \sum_{\sigma,\sigma':d_u(\sigma,\sigma')<N\ln\alpha_1/\ln 2} \mathbb{1}_{\{(\sigma\cdot\sigma')/N\in I\}} + \sum_{\substack{j=1 \\ \beta^2 A_{1,j}<1}}^{n} e^{-\ln(\alpha_1\dots\alpha_j)N + (\beta^2 - (\beta-\epsilon)^2/2)(a_1+\cdots+a_j)N} \\ &+ \sum_{\substack{j=1 \\ \beta^2 A_{1,j}=\beta^2 A_{1,J_1}=1}}^{J_1} \frac{(\alpha_1\cdots\alpha_j)^{-N} e^{\beta^2 N(a_1+\cdots+a_j)} e^{-(u_{\ln(\alpha_1\cdots\alpha_j),N}(0) - 2\beta\sqrt{N(a_j+\cdots+a_n)})^2/2}}{\sqrt{2\pi}(2\beta\sqrt{N(a_j+\cdots+a_n)} - u_{\ln(\alpha_1\cdots\alpha_j),N}(0))} \end{aligned} \tag{3.10}$$

The second term in (3.10) goes to zero exponentially fast by the choice of ϵ (3.9); the third term is $O(1/N)$. Finally, it is due to the assumption $0 \not\in l \in I_l$ that the first term converges to zero as well. Namely, for any given $m \in (-1, 1)$, the number of pairs of configurations σ, σ' such that $(\sigma\cdot\sigma')/N = m$ is of the order $2^{2N}(2\pi N(1+m)(1-m))^{-1/2} e^{-N\phi(m)}$ as $N \to \infty$ with $\phi(m) = 1/2((1+m)\ln(1+m) + (1-m)\ln(1-m))$ by Stirling's formula. For any $\epsilon > 0$ one can find $\delta > 0$ such that $\phi(m) > \delta$ for all $|m| > \epsilon$. Then, since $0 \not\in \text{cl}I$, the first term of (3.10) being the sum of these numbers with $|m| > \epsilon$, is exponentially small. It follows that $L^1_N \to 0$ in probability. The first factor in L^2_N is not bigger than $2^{-N} e^{-\beta^2 N/2} \sum_\sigma e^{\beta\sqrt{N} X_\sigma} \mathbb{1}_{\{\cup_{j=1}^n \bar{D}_{\sigma_1\cdots\sigma_j}\}}$ which is exactly $\bar{Z}^T_{\beta,N}/\mathbb{E}Z^T_{\beta,N}$ in terms of the notation of Lemma 2.1. In the proof of that lemma it has already been shown that this term converges to zero in probability (see (2.10) and the discussion following it). Hence, $L_N^2 \to 0$ in probability as well, finishing the proof of the assertion (i).

Let us turn to (ii). We introduce the events

$$E_{\bar\sigma_1\dots\bar\sigma_i} = \{\omega : \gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}}) + \cdots + \gamma_i u^{-1}_{\ln\bar\alpha_i,N}(\bar X^{\sigma_1\dots\sigma_{J_{i-1}}}_{\sigma_{J_{i-1}+1}\cdots\sigma_{J_i}}) \le (\gamma_1+\cdots+\gamma_i)x\}.$$

Then

$$\begin{aligned}
&f^{\otimes l(\beta)+1}_{\beta N}(\{1\}\times\cdots\times\{1\}\times I_l\times\Delta_{l+1}\times\cdots\times\Delta_{l(\beta)+1})\\
&= U_N^{-2}\sum_{k=1}^{l-1}\sum_{\sigma_1\dots\sigma_{J_1}} e^{2\beta\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})}\cdots\sum_{\sigma_{J_{k-1}+1}\dots\sigma_{J_k}} e^{2\beta\gamma_k u^{-1}_{\ln\bar\alpha_k,N}(\bar X^{\sigma_1\cdots\sigma_{J_{k-1}}}_{\sigma_{J_{k-1}+1}\cdots\sigma_{J_k}})}\\
&\qquad\times \mathbb{1}_{\{\cap_{j=1}^{k-1}E^N_{\bar\sigma_1\dots\bar\sigma_j}\cap\bar E^N_{\bar\sigma_1\dots\bar\sigma_k}\}}V_{\bar\sigma_1,\dots,\bar\sigma_k}\\
&+U_N^{-2}\sum_{\sigma_1\dots\sigma_{J_1}} e^{2\beta\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})}\cdots\sum_{\sigma_{J_{l-2}+1}\dots\sigma_{J_{l-1}}} e^{2\beta\gamma_{l-1} u^{-1}_{\ln\bar\alpha_{l-1},N}(\bar X^{\sigma_1\cdots\sigma_{J_{l-2}}}_{\sigma_{J_{l-2}+1}\cdots\sigma_{J_{l-1}}})}\\
&\times\mathbb{1}_{\{\cap_{j=1}^{l-1}E^N_{\bar\sigma_1\dots\bar\sigma_j}\}}\sum_{\substack{\bar\sigma_l,\dots,\bar\sigma_{l(\beta)+1}\\ \bar\sigma'_l,\dots,\bar\sigma'_{l(\beta)+1}:\bar\sigma_l\cdot\bar\sigma'_l\in I_l}} S_{\bar\sigma_1,\dots,\bar\sigma_{l-1}}(\bar\sigma_l,\dots,\bar\sigma_{l(\beta)+1};\bar\sigma'_l,\dots,\bar\sigma'_{l(\beta)+1})\\
&\qquad\times\mathbb{1}_{\{\cup_{k=l}^{l(\beta)}\cup_{p=l}^{l(\beta)}(\bar E_{\bar\sigma_1\dots\bar\sigma_{l-1}\bar\sigma_l\dots\bar\sigma_k}\cap\bar E_{\bar\sigma_1\dots\bar\sigma_{l-1}\bar\sigma'_l\dots\bar\sigma'_p})\}}\\
&+U_N^{-2}\sum_{\sigma_1,\dots\sigma_{J_1}} e^{2\beta\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})}\cdots\sum_{\sigma_{J_{l-2}+1}\dots\sigma_{J_{l-1}}} e^{2\beta\gamma_{l-1} u^{-1}_{\ln\bar\alpha_{l-1},N}(\bar X^{\sigma_1\cdots\sigma_{J_{l-2}}}_{\sigma_{J_{l-2}+1}\cdots\sigma_{J_{l-1}}})}\\
&\times\mathbb{1}_{\{\cap_{j=1}^{l-1}E^N_{\bar\sigma_1\dots\bar\sigma_j}\}}\sum_{\substack{\bar\sigma_l,\dots,\bar\sigma_{l(\beta)+1}\\ \bar\sigma'_l,\dots,\bar\sigma'_{l(\beta)+1}:\bar\sigma_l\cdot\bar\sigma'_l\in I_l}} S_{\bar\sigma_1,\dots,\bar\sigma_{l-1}}(\bar\sigma_l,\dots,\bar\sigma_{l(\beta)+1};\bar\sigma'_l,\dots,\bar\sigma'_{l(\beta)+1})\\
&\qquad\times\mathbb{1}_{\{(\cap_{k=l}^{l(\beta)}E_{\bar\sigma_1\dots\bar\sigma_{l-1}\bar\sigma_l\dots\bar\sigma_k})\cup(\cap_{p=l}^{l(\beta)}E_{\bar\sigma_1\dots\bar\sigma_{l-1}\bar\sigma'_l\dots\bar\sigma'_p})\}}\\
&\equiv U_N^{-2}(M_N^1+M_N^2+M_N^3)
\end{aligned}\tag{3.11}$$

with the notations

$$\begin{aligned}
&S_{\bar\sigma_1,\dots,\bar\sigma_{l-1}}(\bar\sigma_l,\dots,\bar\sigma_{l(\beta)+1};\bar\sigma'_l,\dots,\bar\sigma'_{l(\beta)+1})\\
&\equiv e^{\beta\gamma_l u^{-1}_{\ln\bar\alpha_l,N}(\bar X^{\sigma_1\dots\sigma_{J_{l-1}}}_{\sigma_{J_{l-1}+1}\cdots\sigma_{J_l}})+\beta\gamma_l u^{-1}_{\ln\bar\alpha_l,N}(\bar X^{\sigma_1\dots\sigma_{J_{l-1}}}_{\sigma'_{J_{l-1}+1}\cdots\sigma'_{J_l}})}\\
&\times e^{\beta\gamma_{l(\beta)} u^{-1}_{\ln\bar\alpha_{l(\beta)},N}(\bar X^{\sigma_1\dots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\cdots\sigma_{J_{l(\beta)}}})+\beta\gamma_{l(\beta)} u^{-1}_{\ln\bar\alpha_{l(\beta)},N}(\bar X^{\sigma_1\dots\sigma'_{J_{l(\beta)-1}}}_{\sigma'_{J_{l(\beta)-1}+1}\cdots\sigma'_{J_{l(\beta)}}})}\\
&\times e^{\beta\sqrt N(\sqrt{a_{J_{l(\beta)+1}}}X_{\sigma_1\dots\sigma_{J_{l(\beta)}+1}}+\cdots+\sqrt{a_n}X_{\sigma_1\dots\sigma_n}+\sqrt{a_{J_{l(\beta)+1}}}X_{\sigma_1\dots\sigma'_{J_{l(\beta)}+1}}+\cdots+\sqrt{a_n}X_{\sigma_1\dots\sigma'_n})}\\
&\times e^{-\beta^2N(a_{J_{l(\beta)+1}}+\cdots+a_n)-2N\ln(\alpha_{J_{l(\beta)+1}}\cdots\alpha_n)}
\end{aligned}$$

$$
\begin{aligned}
&V_{\bar{\sigma}_1,\dots,\bar{\sigma}_k} \\
&\equiv \Big[\sum_{\sigma_{J_k+1}\dots\sigma_{J_{l-1}}} e^{2\beta\sqrt{N}(\sqrt{a_{J_k+1}}X_{\sigma_1\dots\sigma_{J_k+1}}+\dots+\sqrt{a_{J_{l-1}}}X_{\sigma_1\dots\sigma_{J_{l-1}}})} \\
&\qquad \sum_{\substack{\sigma_{J_{l-1}+1},\dots,\sigma_n \\ \sigma'_{J_{l-1}+1},\dots,\sigma'_n:\ \bar{\sigma}_l\cdot\bar{\sigma}'_l\in I_l}} e^{\beta\sqrt{N}(\sqrt{a_{J_{l-1}+1}}X_{\sigma_1\dots\sigma_{J_{l-1}+1}}+\dots+\sqrt{a_n}X_{\sigma_1\dots\sigma_n}+\sqrt{a_{J_{l-1}+1}}X_{\sigma_1\dots\sigma'_{J_{l-1}+1}}+\dots+\sqrt{a_n}X_{\sigma_1\dots\sigma'_n})}\Big] \\
&\times \Big[\sum_{\sigma_{J_k+1}\dots\sigma_n} e^{\beta\sqrt{N}(\sqrt{a_{J_k+1}}X_{\sigma_1\dots\sigma_{J_k+1}}+\dots+\sqrt{a_n}X_{\sigma_1\dots\sigma_n})}\Big]^{-2} \\
&\times \Big[\sum_{\sigma_{J_k+1}\dots\sigma_{J_{l(\beta)}}} e^{\beta\gamma_{k+1}u^{-1}_{\ln\bar{\alpha}_{k+1},N}(\bar{X}^{\sigma_1\dots\sigma_{J_k}}_{\sigma_{J_k+1}\dots\sigma_{J_{k+1}}})+\dots+\beta\gamma_{l(\beta)}u^{-1}_{\ln\bar{\alpha}_{l(\beta)},N}(\bar{X}^{\sigma_1\dots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}})} \\
&\quad \sum_{\sigma_{J_{l(\beta)}+1},\dots,\sigma_n} e^{\beta\sqrt{N}(\sqrt{a_{J_{l(\beta)}+1}}X_{\sigma_1\dots\sigma_{J_{l(\beta)}+1}}+\dots+\sqrt{a_n}X_{\sigma_1\dots\sigma_n})-\beta^2N(a_{J_{l(\beta)}+1}+\dots+a_n)/2-N\ln(\alpha_{J_{l(\beta)+1}}\cdots\alpha_n)}\Big]^2
\end{aligned}
$$

and

$$
\begin{aligned}
U_N = &\sum_{\sigma_1\dots\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln\bar{\alpha}_1,N}(\bar{X}_{\sigma_1\dots\sigma_{J_1}})}\dots \sum_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}} e^{\beta\gamma_{l(\beta)}u^{-1}_{\ln\bar{\alpha}_{l(\beta)},N}(\bar{X}^{\sigma_1\dots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}})} \\
&\sum_{\sigma_{J_{l(\beta)}+1},\dots,\sigma_n} e^{\beta\sqrt{N}(\sqrt{a_{J_{l(\beta)+1}}}X_{\sigma_1\dots\sigma_{J_{l(\beta)}+1}}+\dots+\sqrt{a_n}X_{\sigma_1\dots\sigma_n})-\beta^2N(a_{J_{l(\beta)}+1}+\dots+a_n)/2-N\ln(\alpha_{J_{l(\beta)+1}}\cdots\alpha_n)}.
\end{aligned}
$$

The third term in (3.11) $U_N^{-2}M_N^3$ is not bigger than $\tilde{M}_N^3/U_N$ where $\tilde{M}_N^3$ is the sum over the same terms as U_N but truncated by $\mathbb{1}\{\cap_{j=1}^{l(\beta)}E_{\bar{\sigma}_1\dots\bar{\sigma}_j}\}$. By Theorem 1.3 the denominator U_N converges in law to the a.s. finite integral over the process $\mathcal{P}^{(l(\beta))}$. Then for any $\epsilon>0$ one can find a small $\delta(\epsilon)>0$ such that $\mathbb{P}(U_N<\delta(\epsilon))<\epsilon$ for all sufficiently large N. The expectation of the numerator $\tilde{M}_N^3$ is bounded by $e^{x\sum_{j=1}^{l(\beta)}(\beta\gamma_j-1)}$ for all N large enough by Lemma 2.2. Then for any given pair of $\epsilon,\tilde{\epsilon}>0$ one can find $x<0$ sufficiently large by absolute value such that

$$
\begin{aligned}
\mathbb{P}(M_N^3/U_N^2>\tilde{\epsilon}) &\le \mathbb{P}(\tilde{M}_N^3/U_N>\tilde{\epsilon}) \le \mathbb{P}(\tilde{M}_N^3/\delta(\epsilon)>\tilde{\epsilon})+\mathbb{P}(U_N>\delta(\epsilon)) \\
&< \frac{e^{x\sum_{j=1}^{l(\beta)}(\beta\gamma_j-1)}}{\tilde{\epsilon}\delta(\epsilon)}+\epsilon<2\epsilon
\end{aligned}
$$

for all N starting from some number N_0.

Given $\tilde{\epsilon},\epsilon>0$, let us then fix an appropriate $x<0$ and consider the terms M_N^1/U_N^2 and M_N^2/U_N^2 with this choice. To treat M_N^2/U_N^2, it suffices to show that M_N^2 is non-zero with vanishing probability as N tends to infinity. This means that the random variable

$$\eta_N = \sum_{\bar\sigma_1,\dots,\bar\sigma_{l-1}} \mathbb{1}_{\{\cap_{j=1}^{l-1} E^N_{\bar\sigma_1\dots\bar\sigma_j}\}} \sum_{\substack{\bar\sigma_l,\dots,\bar\sigma_{l(\beta)+1} \\ \bar\sigma'_l,\dots,\bar\sigma'_{l(\beta)+1}:\bar\sigma_l\cdot\bar\sigma'_l\in I_l}} \mathbb{1}_{\{\cup_{k=l}^{l(\beta)}\cup_{p=l}^{l(\beta)}(\bar E_{\bar\sigma_1\dots\bar\sigma_{l-1}\bar\sigma_l\dots\bar\sigma_k}\cap \bar E_{\bar\sigma_1\dots\bar\sigma_{l-1}\bar\sigma'_l\dots\bar\sigma'_p})\}}$$

does not equal zero with probability converging to zero. Let us decompose $\eta_N = \sum_{k,p=l}^{l(\beta)} \eta_N^{p,k}$ where

$$\eta_N^{p,k} = \sum_{\bar\sigma_1,\dots,\bar\sigma_k} \mathbb{1}_{\{\cap_{j=1}^{k-1} E_{\bar\sigma_1,\dots,\bar\sigma_j}\cap \bar E_{\bar\sigma_1,\dots,\bar\sigma_k}\}} \sum_{\bar\sigma'_l,\dots,\bar\sigma'_p:\bar\sigma_l\cdot\bar\sigma'_l\in I_l} \mathbb{1}_{\{\cap_{i=1}^{p-1} E_{\bar\sigma_1,\dots,\bar\sigma_{l-1}\bar\sigma'_l\dots\bar\sigma'_i}\cap \bar E_{\bar\sigma_1,\dots,\bar\sigma_{l-1}\bar\sigma'_l,\dots,\bar\sigma'_p}\}}$$

$p,k = l,\dots,l(\beta)$ and estimate for $p\ge k\ge l$

$$\begin{aligned}\eta_N^{p,k} \le \tilde\eta_N^{p,k} = &\sum_{\bar\sigma_1,\dots,\bar\sigma_k} \mathbb{1}_{\{\bar E_{\bar\sigma_1,\dots,\bar\sigma_k}\}} \\ &\sum_{\bar\sigma'_l,\dots,\bar\sigma'_p:\bar\sigma_l\cdot\bar\sigma'_l\in I_l} \mathbb{1}_{\{\gamma_l u^{-1}_{\ln\alpha_l,N}(\bar X^{\sigma_1,\dots,\sigma_{J_{l-1}}}_{\sigma'_{J_{l-1}+1},\dots,\sigma'_{J_l}})+\dots+\gamma_p u^{-1}_{\ln\alpha_p,N}(\bar X^{\sigma_1,\dots,\sigma'_{J_{p-1}}}_{\sigma'_{J_{p-1}+1},\dots,\sigma'_{J_p}})>(\gamma_l+\dots+\gamma_p)x\}}.\end{aligned} \tag{3.12}$$

We will show $\mathbb{P}(\tilde\eta_N^{p,k}\neq 0)\to 0$ as $N\to 0$. First of all, note that for any $\epsilon>0$ there exists $K(\epsilon)$ such that $\mathbb{P}(\sum_{\bar\sigma_1,\dots,\bar\sigma_k}\mathbb{1}_{\{\bar E_{\bar\sigma_1,\dots,\bar\sigma_k}\}} > K(\epsilon)) < \epsilon$ for all N large enough. Namely, by Theorem 1.1 this probability converges to the probability that the process $\mathcal{P}^{(k)}$ in R^k has more than K points in the domain $\gamma_1x_1+\dots+\gamma_kx_k > (\gamma_1+\dots+\gamma_k)x$. This last probability tends to zero as $K\to\infty$ since by Proposition 1.4 (i) $\mathcal{P}^{(k)}$ has a.s. a finite number of points in this domain. Second, observe that the sums of the second level $\sum_{\bar\sigma'_l,\dots,\bar\sigma'_p:\bar\sigma_l\cdot\bar\sigma'_l\in I_l}\mathbb{1}_{\{\cdot\}}$ are *independent* from the indicators in front of them in the first level due to the condition $\bar\sigma_l\cdot\bar\sigma'_l\in I_l$ with $1\notin I_l$. Third, for any given $\bar\sigma_l$ let us compute the number of configurations $\bar\sigma'_l$ such that $\bar\sigma_l\cdot\bar\sigma'_l\in I_l$ and estimate it by Stirling's formula as it was done in (i). This number $\bar\alpha_l^N(1+o(1))\sum_{m\in I_l}(2\pi N(1+m)(1-m))^{-1/2}e^{-N\phi(m)}$ turns out to be exponentially smaller than $\bar\alpha_l^N$ by virtue of the condition $0\not\!\text{cl}\in I_l$. In other words, it is bounded by some $\widehat\alpha_l^N$ where $\widehat\alpha_l<\bar\alpha_l$. These arguments combined together imply

$$\mathbb{P}(\tilde\eta_N^{p,k}\neq 0)\le K(\epsilon)\mathbb{P}(\max_\sigma \widehat X_\sigma > \sqrt{\bar a_l}u_{\ln\bar\alpha_l,N}(x)+\dots+\sqrt{\bar a_p}u_{\ln\bar\alpha_p,N}(x))+\epsilon \tag{3.13}$$

where $\widehat X_\sigma$ are Gaussian random variables with zero mean and variance $\bar a_l+\dots+\bar a_p$ subjected to the GREM of $p-l+1$ levels with parametes $\widehat\alpha_l,\bar\alpha_{l+1},\dots,\bar\alpha_p$ and $\bar a_l,\dots,\bar a_p$. By Theorem 1.5 (iii) from [BK1] $\max_\sigma X_\sigma/\sqrt{N}\to\sqrt{\bar a_l\ln\widehat\alpha_l}+\dots+\sqrt{\bar a_p\ln\bar\alpha_p}$ a.s. which is smaller than the value in (3.13) as $\widehat\alpha_l<\bar\alpha_l$. Then the first term in the r.h.s. of (3.13) converges to zero as $N\to\infty$. This proves $\mathbb{P}(\tilde\eta_N^{p,k}\neq 0)\to 0$ for all $p\ge k\ge l$. The case $k\ge p\ge l$ is symmetric. Then $\mathbb{P}(M_N^2\neq 0)\to 0$ and $M_N^2/U_N^2\to 0$ in probability.

If $l = 1$, then the term M_N^1/U_N^2 in (3.11) is absent and the assertion (ii) is proven. Otherwise we proceed by induction. As in the analysis of the third term, for any given $\epsilon > 0$ we choose $\delta(\epsilon)$ to make the denominator U_N^2 smaller than $\delta(\epsilon)$ with probability smaller than ϵ and after that concentrate on the numerator M_N^1. Assume that (ii) has been proven up to $l-1$ and consider the kth term of M_N^1. For any $\epsilon > 0$ one can find such a large $K(\epsilon)$ that $\mathbb{P}(\sum_{\bar\sigma_1,\ldots,\bar\sigma_k} \mathbb{1}_{\{\bar E^N_{\bar\sigma_1\cdots\bar\sigma_k}\}} > K) < \epsilon$. Once again, by Theorem 1.1, this probability converges to the probability that the process $\mathcal{P}^{(k)}$ in $\mathbb{R}^k$ has more than K points in the domain $\gamma_1 x_1 + \cdots + \gamma_k x_k > (\gamma_1 + \cdots + \gamma_k)x$ and this last probability tends to zero when $K \to \infty$ by Proposition 1.4 (i). Thus, with a large enough probability the kth term is a linear combination of not more than K terms of type $V_{\bar\sigma_1,\ldots,\bar\sigma_k}$. Each of these terms $V_{\bar\sigma_1,\ldots,\bar\sigma_k}$ converges in probability to zero, since the product of the first two factors $[\cdot][\cdot]^{-2}$ of $V_{\bar\sigma_1,\ldots,\bar\sigma_k}$ do so by induction assumption, while the third factor $[\cdot]^2$ converges to the square of an a.s. finite integral over the process $\mathcal{P}^{(l(\beta)-k)}$ by Theorem 1.3. Thus for any $k = 1,\ldots,l-1$, the kth term of M_N^1 converges in probability to zero and the proof of the assertion (ii) is terminated.

The proof of the assertion (iii) goes by induction along the same lines as for (ii) using truncation by $E_{\bar\sigma_1,\ldots,\bar\sigma_k}$, the point process convergence of Theorem 1.1 together with the properties of the process $\mathcal{P}^{(m)}$ of Proposition 1.4, Theorem 1.3 and the estimate of the Lemma 2.2. We omit similar details to (ii) and only emphasize the differences. Observe that to obtain (3.13) it was crucial that the sums of the second level of (3.12) were independent from the indicators in front of them. This property was acheived by the assumption $1 \notin I_l$. In (iii) this independence property is already provided by the assumption $1 \notin I_{l-1}$, since then the random variables $X_{\sigma_1,\ldots,\bar\sigma_{l-1}\bar\sigma_l,\ldots}$ and $X_{\sigma_1,\ldots,\bar\sigma'_{l-1}\bar\sigma'_l\ldots}$ with $(\bar\sigma_{l-1}\cdot\bar\sigma'_{l-1})/N \in I_l$ are independent. Therefore under (iii) the contribution of any I_l such that $0 \notin \mathrm{cl}I_l$ is negligible. Furthermore, if evaluating (3.12), we have $0 \in \mathrm{cl}I_l$, but for some j, $l < j \le p$, $0 \notin \mathrm{cl}I_p$, then anyway $\mathbb{P}(\tilde\eta_N^{p,k} \neq 0) \to 0$: We would get (3.13) where $\widehat{X}_\sigma$ are subjected to the GREM of $p-l+1$ levels with the parameters $\bar\alpha_l,\ldots\widehat{\alpha}_j,\ldots,\bar\alpha_p$, $\widehat{\alpha}_j < \bar\alpha_j$, and then with the maximum smaller than the value in (3.13). Finally, if $\mathrm{cl}I_l,\ldots,\mathrm{cl}I_{l(\beta)}$ all contain zero but $I_{l(\beta)+1}$ does not, then truncating by $E_{\bar\sigma_1,\ldots,\bar\sigma_k}$, applying induction and the same arguments as in (ii) we reduce the problem to evaluating the sum $\sum_{\bar\sigma_{l(\beta)+1},\bar\sigma'_{l(\beta)+1}:(\bar\sigma_{l(\beta)+1}\cdot\bar\sigma'_{l(\beta)+1})/N\in I_{l(\beta)+1}} e^{\beta\sqrt{N(a_{l(\beta)+1}+\cdots+\alpha_n)}(X_{\bar\sigma_{l(\beta)+1}}+X_{\bar\sigma'_{l(\beta)+1}})}$ divided by $e^{-\beta^2 N(a_{l(\beta)+1}+\cdots+\alpha_n)-N\ln(a_{l(\beta)+1}\cdots\alpha_n)}$ where $X_{\bar\sigma_{l(\beta)+1}}$ and $X_{\bar\sigma'_{l(\beta)+1}}$ are independent because of $1 \notin I_{l-1}$. The expectation of this term converges to zero if only $0 \notin \mathrm{cl}I_{l(\beta)+1}$.

It is easy to derive from (i), (ii) and (iii) that the measure $f_{\beta,N}^{\otimes l(\beta)+1}$ in the limit has an atomic structure with atoms $(\{1\} \times \cdots \times \{1\} \times \{0\} \cdots \times \{0\} \times \{0\})$ and it remains to study its asymptotic behaviour on the set $(\{1\} \times \cdots \times \{1\} \times \Delta_{l(\beta)+1})$ and show that in the limit it is concentrated on the unique atom of this set $(\{1\} \times \cdots \times \{1\} \times \{0\})$. This is the subject of (iv). Its proof is carried again by induction as in (ii) and using truncatons by $\mathbb{1}_{\{E_{\bar\sigma_1 \ldots \bar\sigma_k}\}}$. We outline only that the basis of the induction is the statement (i) for $l(\beta) = 0$. All other arguments are the same as in (ii), therefore we omit further details of the proof. $\diamondsuit$

4. Proof of the Ghirlanda-Guerra relations.

In this section we find the mean of the overlap distribution function stated Proposition 1.7 and Ghirlanda-Guerra relations stated in Proposition 1.8. Throughout this section we assume that the model parameters are such that the inequalities (1.9) are all strict.

The following simple observation is the key towards both of these results.

Proposition 4.1: *For any value of β, and any $i = 1, \ldots, n$,*

$$\mathbb{E}\frac{d}{d\sqrt{a}_i}F_{\beta,N} = -\beta^2\sqrt{a_i}\mathbb{E}f_{\beta,N}\Big(\sum_{j=1}^{i}\ln\alpha_j/\ln 2\Big) \tag{4.1}$$

Proof of Proposition 4.1: Obviously,

$$\mathbb{E}\frac{d}{d\sqrt{a_i}}F_{\beta,N} = -N^{-1}\beta\mathbb{E}\frac{\mathbb{E}_\sigma\sqrt{N}X_{\sigma_1\ldots\sigma_i}e^{\beta\sqrt{N}X_\sigma}}{\mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma}} \tag{4.2}$$

Now if X is standard normal variable, and g any function of at most polynomial growth, then

$$\mathbb{E}[Xg(X)] = \mathbb{E}g'(X) \tag{4.3}$$

Using this identity in the right hand side of (4.2) with respect to the average over $X_{\sigma_1\ldots\sigma_i}$, we get immediately that

$$\begin{aligned}\mathbb{E}\frac{\mathbb{E}_\sigma\sqrt{N}X_{\sigma_1\ldots\sigma_i}e^{\beta\sqrt{N}X_\sigma}}{\mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma}} &= N\beta\sqrt{a_i}\mathbb{E}\left(1 - \frac{\mathbb{E}_\sigma\mathbb{E}_{\sigma'}\mathbb{1}_{\sigma_1\ldots\sigma_i=\sigma'_1\ldots\sigma'_2}e^{\beta\sqrt{N}(X_\sigma+X_{\sigma'}}}{(\mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma})^2}\right)\\ &= N\beta\sqrt{a_i}\mathbb{E}\left(1-\mu_{\beta,N}^{\otimes 2}\left(\mathbb{1}_{\sigma_1\ldots\sigma_i=\sigma'_1\ldots\sigma'_i}\right)\right)\end{aligned} \tag{4.4}$$

which is obviously (4.1). $\diamondsuit$

Proof of Proposition 1.7.: Due to the convexity of the function $F_{\beta,N}$ and Proposition 4.1

$$\mathbb{E}f_\beta\Big(\sum_{j=1}^{i}\ln\alpha_j/\ln 2\Big) = -\beta^{-2}(\sqrt{a_i})^{-1}\frac{d}{d\sqrt{a_i}}\lim_{N\to\infty}\mathbb{E}F_{\beta,N}$$

where the last limit is found in Theorem 1.2. Taking the derivative of the r.h.s. of (1.14), the result follows. $\diamondsuit$.

We now turn to the proof Proposition 1.8.

Proof of Proposition 1.8: Let us first note that we need to prove (1.30) only for $i \leq J_{l(\beta)}$, since for large values of i , both sides are trivially equal to zero.

The starting point is the following obvious generalization of (4.4):

Lemma 4.2: *Let $h : \mathcal{S}_N^n \to \mathbb{R}$ be any bounded function of n spins. Then*

$$\begin{aligned} &\frac{1}{\sqrt{N}} \mathbb{E}\mu_{\beta,N}^{\otimes n} \left(X_{\sigma_1^k \dots \sigma_i^k} h(\sigma^1, \dots, \sigma^n) \right) \\ &= \beta\sqrt{a_i} \mathbb{E}\mu_{\beta,N}^{\otimes n+1} \left(h(\sigma^1, \dots, \sigma^n) \left(\sum_{l=1}^n \mathbb{1}_{\sigma_1^k \dots \sigma_i^k = \sigma_1^l \dots \sigma_i^l} - n \mathbb{1}_{\sigma_1^k \dots \sigma_i^k = \sigma_1^{n+1} \dots \sigma_i^{n+1}} \right) \right) \end{aligned} \tag{4.5}$$

Proof: It is completely analogous to the proof of Proposition 4.1 using the integration by parts (4.3). $\diamondsuit$

The strength of Lemma 4.2 comes out when combined with a factorization result that in turn is a consequence of self-averaging.

Lemma 4.3: *Let h be as in the previous lemma. Then under the hypothesis of the theorem,*

$$\lim_{N\uparrow\infty} \frac{1}{\sqrt{N}} \left| \mathbb{E}\mu_{\beta,N}^{\otimes n} \left(X_{\sigma_1^k \dots \sigma_i^k} h(\sigma^1, \dots, \sigma^n) \right) - \mathbb{E}\mu_{\beta,N} \left(X_{\sigma_1^k \dots \sigma_i^k} \right) \mathbb{E}\mu_{\beta,N}^{\otimes n} \left(h(\sigma^1, \dots, \sigma^n) \right) \right| = 0 \tag{4.6}$$

Proof: Let us write

$$\begin{aligned} &\left(\mathbb{E}\mu_{\beta,N}^{\otimes n} \left(X_{\sigma_1^k \dots \sigma_i^k} h(\sigma^1, \dots, \sigma^n) \right) - \mathbb{E}\mu_{\beta,N} (X_{\sigma_1^k \dots \sigma_i^k}) \mathbb{E}\mu_{\beta,N}^{\otimes n} (h(\sigma^1, \dots, \sigma^n)) \right)^2 \\ &= \left(\mathbb{E}\mu_{\beta,N}^{\otimes n} \left[\left(X_{\sigma_1^k \dots \sigma_i^k} - \mathbb{E}\mu_{\beta,N}^{\otimes n} (X_{\sigma_1^k \dots \sigma_i^k}) \right) h(\sigma^1, \dots, \sigma^n) \right] \right)^2 \\ &\leq \mathbb{E}\mu_{\beta,N}^{\otimes n} \left(X_{\sigma_1^k \dots \sigma_i^k} - \mathbb{E}\mu_{\beta,N}^{\otimes n} (X_{\sigma_1^k \dots \sigma_i^k}) \right)^2 \mathbb{E}\mu_{\beta,N}^{\otimes n} \left(h(\sigma^1, \dots, \sigma^n) \right)^2 \end{aligned} \tag{4.7}$$

where the last inequality is the Cauchy–Schwarz inequality applied to the joint expectation with respect to the Gibbs measure and the disorder. Obviously the first factor in the last line is equal to

$$\begin{aligned}
&\mathbb{E}\mu_{\beta,N}\left(X_{\sigma_1^k\dots\sigma_i^k}-\mu_{\beta,N}(X_{\sigma_1^k\dots\sigma_i^k})\right)^2\\
&\quad+\mathbb{E}\Big[\mu_{\beta,N}(X_{\sigma_1^k\dots\sigma_i^k})-\mathbb{E}\mu_{\beta,N}(X_{\sigma_1^k\dots\sigma_i^k})\Big]^2\\
&=-\beta^{-2}\mathbb{E}\frac{d^2}{d\sqrt{a_i}^2}F_{\beta,N}+N\beta^{-2}\mathbb{E}\left(\frac{d}{d\sqrt{a_i}}F_{\beta,N}-\mathbb{E}\frac{d}{d\sqrt{a_i}}F_{\beta,N}\right)^2
\end{aligned}\tag{4.8}$$

We know that by Theorem 1.2 $F_{\beta,N}$ converges as $N\uparrow\infty$ and that the limit is infinitely differentiable as a function of the parameters $a_1,\dots,a_{J(l(\beta))}$, except at the set of values where some of the conditions (1.9) hold with equality; moreover, $-F_{\beta,N}$ is convex in each of the variables $\sqrt{a_i}$. Then standard results of convex analysis (see [Ro], Theorem 25.7) imply that

$$\limsup_{N\uparrow\infty}(-\mathbb{E}\frac{d^2}{d\sqrt{a_i}^2}F_{\beta,N})=-\frac{d^2}{d\sqrt{a_i}^2}\lim_{N\uparrow\infty}\mathbb{E}F_{\beta,N}\tag{4.9}$$

which is finite. Thus, the first term in (4.8) will vanish when divided by N. To see that the coefficient of N of the second term gives a vanishing contribution, we use the general fact that if the variance of family of a convex (or concave) functions tends to zero, then the same is true for its derivative, provided that the second derivative is bounded (see Lemma 8.9 in [BG] or Lemma 4.3 in [T1].) In Theorem 1.3 we have more than established that the variance of $F_{\beta,N}$ tends to zero (this follows of course also from Gaussian concentration of measure estimates [LT]), and hence the result of the Lemma is proven. $\diamondsuit$

Remark: The last argument is the sole reason why we need to exclude the situations when equality holds for some of the equations (1.9). In fact, in this case the model is in a sense at a "critical point" in the space of its parameters.

Finally, we substitute the result of Lemma 4.3 into Lemma 4.2. Using in the resulting expression the fact that by 4.1 and (4.2),

$$\mathbb{E}\mu_{\beta,N}(X_{\sigma_1\dots\sigma_i})/\sqrt{N}=\beta\sqrt{a_i}(1-\mathbb{E}\mu_{\beta,N}^{\otimes 2}(\mathbb{1}_{\sigma_1\dots\sigma_i=\sigma_1'\dots\sigma_i'\}}))\tag{4.10}$$

we obtain the assertion of Proposition 1.8.$\diamondsuit$

References

[AC] M. Aizenmann and P. Contucci, On the stability of the quenched state in mean field spin-glass models, J. Statist. Phys. **92**, 765-783 (1998).

[BG] A. Bovier, V. Gayrard, The Hopfield model as a generalized random mean field model, in Mathematics of spin glasses and neural networks, A. Bovier, P. Picco. Eds., Progress in Probablity, Birkhäuser, Boston, (1997).

[BK1] A. Bovier and I. Kurkova, Derrida's Generalized random Energy models 1. Poisson cascades and the extremal process. preprint (2002).

[BK3] A. Bovier and I. Kurkova, Derrida's Generalized Random Energy models 3. Models with continuous hierarchies. in preparation (2002).

[BKL] A. Bovier, I. Kurkova, and M. Löwe, Fluctuations of the free energy in the REM and the p-spin SK-models, Ann. Probab. **30** (2002).

[CCP] D. Capocaccia, M. Cassandro, and P. Picco, On the existence of Thermodynamics for the Generalised Random Energy Model. J. Stat. Physics **46**, 493–505 (1987).

[D1] B. Derrida, Random energy model: limit of a family of disordered models, Phys. Rev. Letts. **45**, 79-82 (1980).

[D2] B. Derrida, Random energy model: An exactly solvable model of disordered systems, Phys. Rev. B **24**, 2613-2626 (1981)

[D3] B. Derrida, A generalization of the random energy model that includes correlations between the energies, J. Physique. Lett. **46**, 401–407 (1985).

[DG1] B. Derrida and E. Gardner, Solution of the generalized random energy model, J. Phys. C **19**, 2253–2274 (1986).

[DG2] B. Derrida and E. Gardner, Magnetic properties and function $q(x)$ of the generalised random energy model, J. Phys. C **19**, 5783-5798 (1986)

[DG3] E. Gardner and B. Derrida, The probability distribution of the partition function of the random energy model, J. Phys. A**22**, 1975-1981 (1989)

[GG] S. Ghirlanda, F. Guerra. General properties of the overlap probability distributions in disordered spin systems.Towards Parisi ultrametricity, J. Phys. A 31(1998) , 9144-9155.

[Ka] O. Kallenberg, Random measures, 4th edition, Akademie Verlag, Berlin, 1986.

[LT] M. Ledoux and L. Talagrand, Probability in Banach space. Spinger, Berlin-Heidelberg-New York, 1991.

[Re] S.I. Resnick, Extreme values, regular variation , and point processes, Springer, New York, 1987.

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

[SK] D. Sherrington and S. Kirkpatrick, "Solvable model of a spin glass", Phys. Rev. Lett. **35**, 1792-1796 (1972).

[T1] M. Talagrand, Rigorous results for the Hopfield model with many patterns, Probab. Theory Related Fields **110**, 177-276 (1998).

[T2] M. Talagrand, Mean field models for spin glasses: a first course, Lecture Notes of the 2000 École des Probabilités de St. Flour; to appear as Springer Lecture Notes in Mathematics.

[T3] M. Talagrand, Rigorous low temperature results for mean field p-spin interaction models. Probab. Theor. Rel. Fields. **117** 303-360 (2000).

[T4] M. Talagrand,"Self organization in the low-temperature region of a spin glass model", Rev. Math. Phys. **15**, 1-78 (2003).

[T5] M. Talagrand, Spin Glasses: a Challenge to Mathematicians, Ergebnisse der Mathematik und ihrer Grenzgebiete Vol. 46, Springer, Berlin-Heidelberg-New York, 2003.