Available online at www.sciencedirect.com
SCIENCE DIRECT®

Ann. I. H. Poincaré – PR 40 (2004) 439–480

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES

www.elsevier.com/locate/anihpb

Derrida's Generalised Random Energy models 1: models with finitely many hierarchies

Anton Bovier [a,b,1], Irina Kurkova [c]

[a] *Weierstraß Institut für Angewandte Analysis und Stochastik, Mohrenstrasse 39, 10117 Berlin, Germany*
[b] *Institut für Mathematik, Technische Universität Berlin, Strasse des 17. Juni 136, 10623 Berlin, Germany*
[c] *Laboratoire de probabilités et modèles aléatoires, université Paris 6, 4, place Jussieu, B.C. 188, 75252 Paris, Cedex 5, France*

Received 6 June 2002; received in revised form 2 September 2003; accepted 2 February 2004

Available online 12 May 2004

Abstract

This is the first of a series of three papers in which we present a full rigorous analysis of a class of spin glass models introduces by Derrida under the name of Generalised Random Energy Models (GREM). They are based on Gaussian random processes on the hypercube $\{-1,1\}^N$ with a hierarchical correlation structure. In this first paper we analyse the models with a finite number of hierarchies. In particular, we identify the thermodynamic limit of the Gibbs measures with Ruelle's probability cascades.
© 2004 Elsevier SAS. All rights reserved.

Résumé

Cet article est le premier d'une série de trois articles où nous présentons une analyse entièrement rigoureuse de la classe des modèles de verres de spin introduite par Derrida sous le nom de Generalised Random Energy Models (GREM). Ces modèles sont basés sur des processus gaussiens sur l'hypercube $\{-1,1\}^N$ ayant une structure de corrélations hierarchique. Dans ce papier nous analysons les modèles ayant un nombre fini de hierarchies. En particulier, nous identifions la mesure de Gibbs dans la limite thermodynamique avec les cascades de probabilités de Ruelle.
© 2004 Elsevier SAS. All rights reserved.

MSC: 82B44; 60G70; 60K35

Keywords: Gaussian processes; Extreme values; Order statistics; Generalised random energy model; Spin glasses; Poisson cascades; Ghirlanda–Guerra identities

Mots-clés : Processus Gaussiens ; Valeurs extrêmes ; Statistiques d'ordre ; Generalised Random Energy Model ; Verres de spin ; Cascades de Poisson ; Identités de Ghirlanda–Guerra

E-mail addresses: bovier@wias-berlin.de (A. Bovier), kourkova@ccr.jussieu.fr (I. Kurkova).
[1] Research supported in part by the DFG in the Concentration program "Interacting stochastic systems of high complexity".

0246-0203/$ – see front matter © 2004 Elsevier SAS. All rights reserved.
doi:10.1016/j.anihpb.2003.09.002

1. Introduction

In 1980, Derrida proposed two simplified models for mean field spin glasses: the *random energy model* (REM) [9,10], and the *generalised random energy model* (GREM) [11–14]. 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 model can be seen formerly as the limit of the so-called p-spin SK-models [32], when p tends to infinity [9]. In spite of its simplicity, this model has proven to be a rather instructive toy model, and has received considerable attention in the mathematical community [2,6,14,16,17,20,22,23,27,30,34,36]. Of course, in many respects this model is mathematically almost trivial, and physically quite unrealistic, as all the dependence structure that is present in more realistic models like the SK model, is absent. The GREM was introduced in view of *keeping* dependence, while simplifying it to a *hierarchical* structure to yield a mathematically more tractable model. In fact, the GREMs can be seen as a class of models that is obtained by equipping the hypercube $\{-1,1\}^N$ with a tree structure and an associated ultra-metric distance, and then considering standardised Gaussian random fields on the hypercube whose correlation function depends only on this distance. We will call these models *"Derrida's models"* in contrast to the *"Sherrington/Kirkpatrick* (*SK*) *models"* [31] where the covariance depends on the Hamming distance, respectively the overlap $R_N(\sigma,\sigma') = N^{-1}\sum_{i=1}^N \sigma_i\sigma_j$. Let us mention in passing that the study of such processes may represent an independent interest in fields other than physics. E.g., in mathematical finance, such processes may represent a very reasonable model for the distribution of assets of a financial portfolio.

In this paper and two companion papers [4,5] we will undertake a comprehensive study of this class of models. In the present paper we consider only the case when the covariance depends on the ultra-metric distance via a step function with finitely many steps. This corresponds to the classical GREM of Derrida. We will see in [4] how these results can be extended to the general case.

Surprisingly, even the GREM is mathematically not fully understood. Capocaccia et al. [8] derived the exact formula for the mean free energy, making rigorous and extending results of Derrida and Gardner [12]. Galvez et al. [20] construct a certain point process for the energies the case of the two-level tree. Most of the intuitive insight into the model stems from a paper by Ruelle [30] who in fact formulates a new model based on Poisson point processes that are suggested to represent the asymptotics of the fluctuations of the partition function and the Gibbs measures of Derrida's models in the thermodynamic limit. An unpublished note of Neveu [26] contains a rough outline of how these connection can be established, but even in the REM a full proof was published only recently [6,34].

An important ingredient in the analysis of the REM is the theory of convergence to Poisson processes of the extreme value statistics of (i.i.d.) random variables that is, of course, very well known (see e.g. [24,28]). In the GREM, analogous results will be needed in the correlated case. Here, standard results on extremal processes are much rarer. There are criteria under which the extremal processes have the same limits as in the independent case. We would like to point to a result of Bramson [7] that is related to the marginal situation. Here we will be interested in situations when this is not the case. The first results of this paper will be general Poisson convergence results for the extremes of hierarchically correlated Gaussian random processes, where the limits will be described as cascades of Poisson processes. In a second step, we will prove the convergence of Gibbs measures to random measures described in terms of such cascades. These results will be obtained by rather explicit and elementary computations. In a third step we will show that a different approach based on so-called Ghirlanda–Guerra identities [19,1] allows to recover many of these results. This was observed in the REM by Talagrand [34]. Before stating our results, we begin with a precise definitions of our models.

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

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

(note that $1 - d_N(\sigma, \sigma')$ is a ultra-metric distance on $\mathcal{S}_N$). 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

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

where $A(x)$ is a probability distribution function on $[0, 1]$.

The proper 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 $a_i > 0$, although at some later stage we will have to look also at what happens when $x_1 = 0$ and $a_n = 0$.

Then the Gaussian process X_σ can be constructed from $\alpha_1^N + (\alpha_1 \alpha_2)^N + \dots + (\alpha_1 \alpha_2 \cdots \alpha_n)^N$ independent standard Gaussian random variables $X_{\sigma_1}, X_{\sigma_1 \sigma_2}, \dots, X_{\sigma_1 \dots \sigma_2}$, where $\sigma_i \in \{-1, 1\}^{N \ln \alpha_i / \ln 2}$ as

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

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

$$Z_{N,\beta} = 2^{-N} \sum_{\sigma \in \mathcal{S}_N} e^{\beta \sqrt{N} X_\sigma} \tag{1.5}$$

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

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

1.2. Convergence of associated point processes

We define the function (see e.g. [24])

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

Note that if X is a standard normal r.v., $\mathbb{P}(X > u_{\ln \alpha, N}(x)) \sim \alpha^{-N} e^{-x}$, as $N \uparrow \infty$.

Theorem 1.1. *Let $n \in \mathbb{N}$, $n \geqslant 1$, $0 < a_i < 1$, $\alpha_i > 1$, $i = 1, 2, \dots, n$. Set $\bar{\alpha} \equiv \prod_{i=1}^n \alpha_i$ and assume that $\sum_{i=1}^n a_i = 1$. The point process*

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

converges weakly to the Poisson point process $\mathcal{P}$ on $\mathbb{R}$ with intensity measure $K e^{-x} dx$, $K \in \mathbb{R}$ if and only if the following inequalities hold true:

$$a_i + a_{i+1} + \dots + a_n \geqslant \ln(\alpha_i \alpha_{i+1} \cdots \alpha_n) / \ln \bar{\alpha} \quad \textit{for all } i = 2, 3, \dots, n. \tag{1.8}$$

Furthermore, if all inequalities in (1.8) *are strict, then the constant $K = 1$. If some of them are equalities, then K is given as*

$$K = \mathbb{P}\Bigg(\bigcap_{\substack{k:\ a_k+\cdots+a_n \\ =\ln(\alpha_k\cdots\alpha_n)/\ln\tilde{\alpha}}} \Bigg(Z_{k-1}\sqrt{\frac{a_{k-1}}{(a_{k-1}+\cdots+a_n)(a_k+\cdots+a_n)}} + Z_{k-2}\sqrt{\frac{a_{k-2}}{(a_{k-2}+\cdots+a_n)(a_{k-1}+\cdots+a_n)}} + \cdots + Z_1\sqrt{\frac{a_1}{(a_1+\cdots+a_n)(a_2+\cdots+a_n)}} < 0 \Bigg)\Bigg),$$

where $Z_1, \ldots, Z_{n-1}$ are $n-1$ independent standard Gaussian random variables.

Remark. Theorem 1.1 gives a sharp criterion under which the correlations do not influence the properties of the extremal process. They are *strictly weaker* than the sufficient conditions one obtains e.g. from Slepian's Gaussian comparison lemma (see e.g. [24], Theorem 4.2.1).

Theorem 1.2. *Let $\alpha_i \geqslant 1$, $i = 1, 2, \ldots, k$, $\tilde{\alpha} \equiv \prod_{i=1}^k \alpha_i$. Let $Y_{\sigma_1}, Y_{\sigma_1\sigma_2}, \ldots, Y_{\sigma_1\ldots\sigma_k}$ be $\alpha_1^N + \cdots + (\alpha_1\cdots\alpha_k)^N$ identically distributed random variables enumerated as in the definition of the GREM. Assume that $1 + \alpha_1^N + \cdots + (\alpha_1\cdots\alpha_{k-1})^N$ vectors $(Y_{\sigma_1})_{\sigma_1\in\{-1,1\}^{N\ln\alpha_1/\ln\tilde{\alpha}}}$, $(Y_{\sigma_1\sigma_2})_{\sigma_2\in\{-1,1\}^{N\ln\alpha_2/\ln\tilde{\alpha}}}$ $\forall\sigma_1 \in \{-1,1\}^{N\ln\alpha_1/\ln\tilde{\alpha}}, \ldots,$ $(Y_{\sigma_1\sigma_2\ldots\sigma_k})_{\sigma_k\in\{-1,1\}^{N\ln\alpha_k/\ln\tilde{\alpha}}}$ $\forall\sigma_1\ldots\sigma_{k-1} \in \{-1,1\}^{N\ln(\alpha_1\cdots\alpha_{k-1})/\ln\tilde{\alpha}}$ are independent. Let also $v_{N,1}(x), \ldots, v_{N,k}(x)$ be functions on $\mathbb{R}$ such that the following point processes*

$$\begin{aligned}
&\sum_{\sigma_1} \delta_{v_{N,1}(Y_{\sigma_1})} \to \mathcal{P}_1 \\
&\sum_{\sigma_2} \delta_{v_{N,2}(Y_{\sigma_1\sigma_2})} \to \mathcal{P}_2 \quad \forall\sigma_1 \\
&\cdots \\
&\sum_{\sigma_k} \delta_{v_{N,k}(Y_{\sigma_1\sigma_2\ldots\sigma_k})} \to \mathcal{P}_k \quad \forall\sigma_1\ldots\sigma_{k-1}
\end{aligned} \tag{1.9}$$

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

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

converge weakly to point processes $\mathcal{P}^{(k)}$ on $\mathbb{R}^k$, which is characterised by the following generating functions:

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

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 $b_i = \infty$, $i = 1, 2, \ldots, k$.

Moreover, the following independence properties of the counting random variables of the process $\mathcal{P}^{(k)}$, $\sum_{x_1} \mathbb{1}_{\{x_1\in\Delta_1^j\}}\cdots\sum_{x_k} \mathbb{1}_{\{x_k\in\Delta_k^j\}}$, corresponding to the intervals $\Delta_1^j \times\cdots\times \Delta_k^j$, $\Delta_i^j = [a_i^j, b_i^j)$, $j = 1, 2, \ldots, r$, $r > 1$, $i = 1, \ldots, k$, hold true:

(i) *If the first components of these intervals are disjoint, i.e. $a_1^1 \leqslant b_1^1 \leqslant a_1^2 \leqslant b_1^2 \leqslant \cdots \leqslant a_1^r \leqslant b_1^r$, then these r.v. are independent.*

(ii) *If the first $l-1$ components of these intervals coincide and the lth components are disjoint, i.e. $\Delta_i^1 = \cdots = \Delta_i^r$ for $i = 1, \ldots, l-1$ and $a_l^1 \leqslant b_l^1 \leqslant a_l^2 \leqslant b_l^2 \leqslant \cdots \leqslant a_l^k \leqslant b_l^r$, then these r.v. are conditionally independent under condition that $\sum_{x_1} \mathbb{1}_{\{x_1\in\Delta_1\}}\cdots\sum_{x_{l-1}} \mathbb{1}_{\{x_{l-1}\in\Delta_{l-1}\}}$ is fixed.*

Remark. This theorem is a generalisation of Theorem 3 in [20]: we do not specify the distribution of the r.v. $Y_{\sigma_1\ldots\sigma_i}$ in the assumptions, and do not impose their independence (but only of their vectors).

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

Theorems 1.1 and 1.2 combined give a first important result that establish which different point processes may be constructed in the GREM.

Theorem 1.3. *Let $\alpha_i \geqslant 1$, $0 < a_i < 1$, $i = 1, 2, \dots, n$, $\prod_{i=1}^n \alpha_i = 2$, $\sum_{i=1}^n a_i = 1$. Let $J_1, J_2, \dots, J_m \in \mathbb{N}$ be the indices such that $0 = J_0 < J_1 < J_2 < \cdots < J_m = n$. We denote by $\bar{a}_l \equiv \sum_{i=J_{l-1}+1}^{J_l} a_i$, $\bar{\alpha}_l \equiv \prod_{i=J_{l-1}+1}^{J_l} \alpha_i$, $l = 1, 2, \dots, m$, and introduce standard Gaussian random variables*

$$\bar{X}^{\sigma_1\dots\sigma_{J_{l-1}}}_{\sigma_{J_{l-1}+1}\sigma_{J_{l-1}+2}\cdots\sigma_{J_l}} \equiv \big(\sqrt{a_{J_{l-1}+1}}X_{\sigma_{J_1}\dots\sigma_{J_{l-1}}\sigma_{J_{l-1}+1}} + \sqrt{a_{J_{l-1}+2}}X_{\sigma_{J_1}\dots\sigma_{J_{l-1}}\sigma_{J_{l-1}+1}\sigma_{J_{l-1}+2}} + \cdots + \sqrt{a_{J_l}}X_{\sigma_{J_1}\dots\sigma_{J_{l-1}}\sigma_{J_{l-1}+1}\dots\sigma_{J_l}}\big)/\sqrt{\bar{a}_l}. \tag{1.11}$$

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

$$(a_k + a_{k+1} + \cdots + a_{J_l-1} + a_{J_l})/\bar{a}_l \geqslant \ln(\alpha_k\alpha_{k+1}\cdots\alpha_{J_l-1}\alpha_{J_l})/\ln(\bar{\alpha}_l). \tag{1.12}$$

Then the point process

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

converges weakly to the process $\mathcal{P}^{(m)}$ on $\mathbb{R}^m$ defined in Theorem 1.2 with constants $K_1, \dots, K_m$. The constant

$$K_l = 1 \tag{1.13}$$

if all $J_l - J_{l-1} - 1$ inequalities in (1.12) for $k = J_{l-1} + 2, \dots, J_l$ are strict. Otherwise

$$K_l = \mathbb{P}\Bigg(\bigcap_{\substack{k:\ J_{l-1}+1\leqslant k\leqslant J_l,\\ (a_k+\cdots+a_{J_l})/\bar{a}_{J_l}=\ln(\alpha_k\cdots\alpha_{J_l})/\ln\bar{\alpha}_l}} \Bigg(Z_{k-1}\sqrt{\frac{a_{k-1}}{(a_{k-1}+\cdots+a_{J_l})(a_k+\cdots+a_{J_l})}} + Z_{k-2}\sqrt{\frac{a_{k-2}}{(a_{k-2}+\cdots+a_{J_l})(a_{k-1}+\cdots+a_{J_l})}} + \cdots + Z_{J_{l-1}+1}\sqrt{\frac{a_{J_{l-1}+1}}{(a_{J_{l-1}+1}+\cdots+a_{J_l})(a_{J_{l-1}+2}+\cdots+a_{J_l})}} < 0\Bigg)\Bigg), \tag{1.14}$$

where $Z_{J_{l-1}+1}, \dots, Z_{J_l}$ are independent standard Gaussian r.v.

Remark. Given a_i, α_i, $i = 1, \dots, n$, the partition of indices $J_1, \dots, J_m$ satisfying the condition (1.12) is generally not unique. Therefore this corollary yields a *family* of convergent point processes associated with the GREM. In the sequel we will frequently drop the upper index and write $\mathcal{P} \equiv \mathcal{P}^{(m)}$, in particular when the value of m can be read off from the arguments.

1.3. The extremal process

We can now formulate a result on the extreme order statistics of the random variables X_σ. It is clear that the extremal process should be constructed from one of the cascade processes that were constructed above. It remains to find out which one. The answer is simple: it is the one that provides the largest values. To do so, one has simply to try to group as many intermediate hierarchies together to a single point process. In other words, among the possible choices for the integers $J_1, \dots, J_m$ allowed in Theorem 1.3, we must choose those that have maximal spacing. Algorithmically, this is achieved by setting $J_0 \equiv 0$, and

$$J_l \equiv \min\{J > J_{l-1}\colon\ A_{J_{l-1}+1,J} > A_{J+1,k}\ \forall k \geqslant J+1\} \tag{1.15}$$

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

Of course it will be essential that

Proposition 1.4. *The sequence $J_1, \dots, J_m$ defined by* (1.15) *verifies the conditions* (1.12) *for all k such that $J_{l-1}+2 \leqslant k \leqslant J_l$ and all $l = 1, 2, \dots, m$.*

Set $\bar{a}_l \equiv \sum_{i=J_{l-1}+1}^{J_l} a_i$, $\bar{\alpha}_l \equiv \prod_{i=J_{l-1}+1}^{J_l} \alpha_i$, $\gamma_l \equiv \sqrt{\bar{a}_l}/\sqrt{2\ln\bar{\alpha}_l}$, $l = 1, 2, \dots, m$. Clearly by (1.15) $\gamma_1 > \gamma_2 > \dots > \gamma_m$. Define the function $U_{J,N}$ by

$$U_{J,N}(x) \equiv \sum_{l=1}^m \left(\sqrt{2N\bar{a}_l \ln\bar{\alpha}_l} - N^{-1/2}\gamma_l\left(\ln\left(N\left(\ln\bar{\alpha}_l\right)\right) + \ln 4\pi\right)/2\right) + N^{-1/2}x \tag{1.16}$$

and the point process

$$\mathcal{E}_N \equiv \sum_{\sigma\in\{-1,1\}^N} \delta_{U_{J,N}^{-1}(X_\sigma)}. \tag{1.17}$$

Theorem 1.5. (i) *The point process $\mathcal{E}_N$ converges weakly, as $N \uparrow \infty$, to the point process on $\mathbb{R}$*

$$\mathcal{E} \equiv \int_{\mathbb{R}^m} \mathcal{P}^{(m)}(dx_1, \dots, dx_m)\delta_{\sum_{l=1}^m \gamma_l x_l}, \tag{1.18}$$

where $\mathcal{P}^{(m)}$ is the Poisson cascade introduced in Theorem 1.3 *corresponding to the partition $J_1, \dots, J_m$ given by* (1.15).

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

(iii) *Furthermore, we have* $\max_\sigma(X_\sigma/\sqrt{N}) \to \sqrt{\bar{a}_1 2\ln\bar{\alpha}_1} + \dots + \sqrt{\bar{a}_m 2\ln\bar{\alpha}_m}$ *a.s. and also* $\mathbb{E}(\max_\sigma X_\sigma/\sqrt{N}) \to \sqrt{2\bar{a}_1\ln\bar{\alpha}_1} + \dots + \sqrt{2\bar{a}_m\ln\bar{\alpha}_m}$.

Remark. Note that if some of the inequalities in the condition (1.12) hold with equality, there are several point processes that to leading order give the same contribution to the extremes. The degeneracy is lifted in favour of the process with the longest increments in the J_i through the $\ln N$ term in the functions $u_{\ln\alpha,N}$ defined in (1.7). If one considers what happens here as a function of the parameters, one sees that the process with fewer levels 'dies out', and the process with extra levels takes over. While the values of the extremes are in leading order the same at the coexistence point, as the inequality gets strictly violated, these values now begin to drop substantially.

1.4. Convergence of the partition function

We will now turn to the study of the Gibbs measures. Technically, the main step in the proof will be to show that the infinite volume limit of the properly rescaled partition function can be expressed as a certain functional of Poisson cascade processes, as suggested by Ruelle [30].

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

$$Z_{\beta,N} = e^{\sum_{j=1}^m (\beta N\sqrt{2\bar{a}_j \ln\bar{\alpha}_j} - \beta\gamma_j[\ln(N\ln\bar{\alpha}_j)+\ln 4\pi]/2 - N\ln\alpha_j)} \times \sum_{\sigma_1\ldots\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln\bar{\alpha}_1,N}(\bar{X}_{\sigma_1\ldots\sigma_{J_1}})} \ldots \sum_{\sigma_{J_{m-1}+1}\ldots\sigma_{J_m}} e^{\beta\gamma_m u^{-1}_{\ln\bar{\alpha}_m,N}(\bar{X}^{\sigma_1\ldots\sigma_{J_{m-1}}}_{\sigma_{J_{m-1}+1}\ldots\sigma_{J_m}})}, \tag{1.19}$$

where $\bar{a}_l \equiv \sum_{i=J_{l-1}+1}^{J_l} a_i$, $\bar{\alpha}_l \equiv \prod_{i=J_{l-1}+1}^{J_l} \alpha_i$, $\gamma_l \equiv \sqrt{\bar{a}_l}/\sqrt{2\ln\bar{\alpha}_l}$, $l = 1, 2, \ldots, m$, and the random variables $\bar{X}^{\sigma_1\ldots\sigma_{J_{l-1}}}_{\sigma_{J_{l-1}+1}\ldots\sigma_{J_l}}$ are defined in (1.11). Moreover, for any sequence $J_1, \ldots, J_m$ satisfying the assumptions of Theorem 1.3, the corresponding point process converges. But, of course, at most one of these sequences can provide through (1.19) the right scaling of the fluctuations of the partition function. As in the analysis of the extremal process, it will be the process corresponding to a maximal spacing of the integers J_i given in (1.15). However, unlike in the case of the extremal process, it must be cut at some temperature dependent level, $J_{l(\beta)}$. Clearly, $A_{1,J_1} > A_{J_1+1,J_2} > \cdots > A_{J_{m-1}+1,J_m}$. Let us put

$$l(\beta) \equiv \max\{l \geqslant 1\colon\ \beta^2 A_{J_{l-1}+1,J_l} > 1\} \tag{1.20}$$

and $l(\beta) \equiv 0$ if $\beta^2 A_{1,J_1} \leqslant 1$. According to the notations of (1.19) $A_{J_{l-1}+1,J_l} = \gamma_l^2$ and (1.20) implies that $\beta\gamma_{l(\beta)+1} \leqslant 1$.

In [8] the limit of the free energy was found in terms of (1.15) and (1.20):

Theorem 1.6 [8]. *With the notation introduced above,*

$$\lim_{N\to\infty} N^{-1}\ln Z_{N,\beta} \equiv -\lim_{N\to\infty} F_{\beta,N} = \beta\big(\sqrt{2\bar{a}_1\ln\bar{\alpha}_1} + \cdots + \sqrt{2\bar{a}_{l(\beta)}\ln\bar{\alpha}_{l(\beta)}}\big) - \ln(\bar{\alpha}_1\cdots\bar{\alpha}_{l(\beta)}) + \sum_{i=J_{l(\beta)}+1}^{n} \beta^2 a_i/2, \quad a.s. \tag{1.21}$$

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

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

Theorem 1.7. *Let $\alpha_i \geqslant 1$, $0 < a_i < 1$, $i = 1, 2, \ldots, n$, $\prod_{i=1}^n \alpha_i = 2$, $\sum_{i=1}^n a_i = 1$. Let $J_1, J_2, \ldots, J_m \in \mathbb{N}$, be the sequence of indices defined by* (1.15) *and $l(\beta)$ defined by* (1.20). *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 $\mathbb{R}^m$:

$$e^{\sum_{j=1}^{l(\beta)}(-\beta N\sqrt{2\bar{a}_j\ln\bar{\alpha}_j} + \beta\gamma_j[\ln(N\ln\bar{\alpha}_j)+\ln 4\pi]/2 + N\ln\alpha_j) - N\sum_{i=J_{l(\beta)}+1}^n \beta^2 a_i/2} Z_{\beta,N} \xrightarrow{\mathcal{D}} C(\beta) \int_{\mathbb{R}^{l(\beta)}} e^{\beta\gamma_1 x_1 + \beta\gamma_2 x_2 + \cdots + \beta\gamma_{l(\beta)} x_{l(\beta)}} \mathcal{P}^{(l(\beta))}(dx_1 \ldots dx_{l(\beta)}). \tag{1.22}$$

This integral is over the process $\mathcal{P}^{(l(\beta))}$ on $\mathbb{R}^{l(\beta)}$ from Theorem 1.3 *with constants K_j from Theorem* 1.3. *The constant $C(\beta)$ satisfies*

$$C(\beta) = 1, \quad \text{if } \beta\gamma_{l(\beta)+1} < 1, \tag{1.23}$$

and

$$C(\beta) = P\Bigg(\bigcap_{\substack{i:\, J_{l(\beta)}+1 \leqslant i \leqslant J_{l(\beta)+1} \\ (a_{J_{l(\beta)}+1}+\cdots+a_i)/\tilde{a}_{J_{l(\beta)+1}} \\ =\ln(\alpha_{J_{l(\beta)}+1}\cdots\alpha_i)/\ln\tilde{\alpha}_{J_{l(\beta)+1}}}} \Big(\sqrt{a_{J_{l(\beta)}+1}} Z_{J_{l(\beta)}+1} + \cdots + \sqrt{a_i} Z_i < 0\Big)\Bigg) \tag{1.24}$$

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 (1.24) is not empty since the necessary equality of parameters holds at least for $i = J_{l(\beta)+1}$.

The integrals over the Poisson cascades appearing in Theorem 1.5 are to be understood as

$$\int_{\mathbb{R}^m} e^{\beta\gamma_1 x_1 + \cdots + \beta\gamma_m x_m} \mathcal{P}^{(m)}(dx_1 \ldots dx_m)$$
$$\equiv \lim_{x\to-\infty} \int_{\substack{(x_1,\ldots,x_m)\in\mathbb{R}^m, \\ \exists i, 1\leqslant i\leqslant m:\ \gamma_1 x_1+\cdots+\gamma_i x_i > (\gamma_1+\cdots+\gamma_i)x}} e^{\beta\gamma_1 x_1 + \cdots + \beta\gamma_m x_m} \mathcal{P}^{(m)}(dx_1 \ldots dx_m). \tag{1.25}$$

The existence of this limit under the hypotheses of the theorem are ensured by

Proposition 1.8. *Assume that the numbers* $\gamma_1, \ldots, \gamma_m$ *are such that* $\gamma_1 > \gamma_2 > \cdots > \gamma_m > 0$ *and* $\beta\gamma_m > 1$. *Then*

(i) *For any* $a \in \mathbf{R}$ *the process* $\mathcal{P}^{(m)}$ *contains a.s. a finite number of points* $(x_1, \ldots, x_m)$ *such that* $\gamma_1 x_1 + \cdots + \gamma_m x_m > a$.
(ii) *The limit in* (1.25) *exists and is finite a.s.*

1.5. The Gibbs measures

We will now turn to the asymptotic description of the Gibbs measures in the GREM. In the REM one considers Ruelle's process of the Gibbs masses, obtained as the limit of the process $\mathcal{W}_N \equiv \sum_\sigma \delta_{\mu_{\beta,N}(\sigma)}$. Together with the information that the replica overlap in the REM can take on only the values 0 and 1 in the limit, this process describes 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 $\varepsilon = \varepsilon(p) > 0$ 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.

In the GREM, this picture is insufficient since 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. 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.

1.5.1. 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

$$\tilde{f}_{\beta,N}(q) \equiv \mu_{\beta,N}^{\otimes 2}\big(R_N(\sigma,\sigma') \leqslant q\big) \tag{1.26}$$

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.

$$f_{\beta,N}(q) \equiv \mu_{\beta,N}^{\otimes 2}\big(d_N(\sigma,\sigma') \leqslant q\big). \tag{1.27}$$

An interesting and rather important result will be the fact that these two notions coincide in the thermodynamics limit, implying that the Gibbs measures concentrate on sets where between any two points the two distances coincide.

With the notation introduced in Theorem 1.7 according to the partition $J_0,\dots,J_m$ defined in (1.15), let us set

$$q_l \equiv \sum_{n=1}^{l} \frac{\ln \bar{\alpha}_n}{\ln 2} \tag{1.28}$$

and

$$q_{\max}(\beta) \equiv \sum_{n=1}^{l(\beta)} \frac{\ln \bar{\alpha}_n}{\ln 2}. \tag{1.29}$$

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

1.5.2. 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

$$B_l(\sigma) \equiv \big\{\sigma' \in \mathcal{S}_N\colon\ d_N(\sigma,\sigma') \geqslant q_l\big\}, \quad l = 1,2,\dots,l(\beta). \tag{1.30}$$

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

$$\mathcal{W}_{\beta,N}^{(m)} \equiv \sum_{\sigma} \delta_{(\mu_{\beta,N}(B_1(\sigma)),\dots,\mu_{\beta,N}(B_m(\sigma)))} \frac{\mu_{\beta,N}(\sigma)}{\mu_{\beta,N}(B_m(\sigma))} \tag{1.31}$$

as well as their projection on the last coordinate,

$$\mathcal{R}_{\beta,N}^{(m)} \equiv \sum_{\sigma} \delta_{\mu_{\beta,N}(B_m(\sigma))} \frac{\mu_{\beta,N}(\sigma)}{\mu_{\beta,N}(B_m(\sigma))}. \tag{1.32}$$

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

$$\mathcal{W}_{\beta,N}^{(m)}(dw_1,\dots,dw_m) = \int_0^1 W_{\beta,N}^{m+1}(dw_1,\dots,dw_m,dw_{m+1}) \frac{w_{m+1}}{w_m} \tag{1.33}$$

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 $m \leqslant l(\beta)$. In fact, we prove the following:

Theorem 1.9. *If $m \leqslant l(\beta)$, 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*

$$\begin{aligned}&\big(w_1(i),\dots,w_m(i)\big)\\ &\quad= \left(\frac{\int \mathcal{P}^{(m)}(dy)\delta(y_1 - x_1(i))e^{\beta(\gamma,y)}}{\int \mathcal{P}^{(m)}(dy)e^{\beta(\gamma,y)}},\dots,\frac{\int \mathcal{P}^{(m)}(dy)\delta(y_1 - x_1(i))\dots\delta(y_m - x_m(i))e^{\beta(\gamma,y)}}{\int \mathcal{P}^{(m)}(dy)e^{\beta(\gamma,y)}}\right)\end{aligned} \tag{1.34}$$

and the point processes $\mathcal{R}^{(m)}_{\beta,N}$ *converge to the point process* $\mathcal{R}^{(m)}_{\beta}$ *whose atoms are the last components of the atoms in* (1.34).

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

The point processes $\widehat{\mathcal{W}}_\beta$ take 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 [30] 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 1.10. *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* $q \in [q_{i-1}, q_i)$

$$f_\beta(q) = \int \widehat{\mathcal{W}}_\beta(dw_1, \dots, dw_{l(\beta)}) w_{l(\beta)}(1-w_i), \quad i = 1, \dots, l(\beta); \tag{1.35}$$

$f_\beta(q) = 1$ *for* $q \geqslant q_l(\beta)$.

1.6. 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 [19,1,33,35]. The existence of this approach will bear its full fruits in the follow-up paper [4] where it will allow us to treat the GREM with continuous hierarchies.

First we get an explicit expression for the mean of f_β.

Proposition 1.11. *The mean of* f_β *is given by*

$$\mathbb{E} f_\beta(q) = \begin{cases} \beta^{-1}\sqrt{\frac{2\ln \bar{\alpha}_j}{\bar{a}_j}}, & \text{if } q \in [q_{j-1}, q_j),\ j \leqslant l(\beta), \\ 1, & \text{if } q \geqslant q_{\max}(\beta). \end{cases} \tag{1.36}$$

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

Proposition 1.12. *Assume that the parameters* α_i *and* a_i *are such that none of the inequalities* (1.12) *holds with equality. Then, for any bounded function* $h : \mathcal{S}^n_N \to \mathbb{R}$, *for all* $i = 1, \dots, n$

$$\begin{aligned} \lim_{N\uparrow\infty} \Bigg| & \mathbb{E}\mu^{\otimes n+1}_{\beta,N}\big(h(\sigma^1, \dots, \sigma^n)\mathbb{1}_{\sigma^k_1 \dots \sigma^k_i = \sigma^{n+1}_1 \dots \sigma^{n+1}_i}\big) \\ & - \frac{1}{n}\mathbb{E}\mu^{\otimes n+1}_{\beta,N}\left(h(\sigma^1, \dots, \sigma^n)\left(\sum_{l\neq k}^{n} \mathbb{1}_{\sigma^l_1 \dots \sigma^l_i = \sigma^k_1 \dots \sigma^k_i} + \mathbb{E}\mu^{\otimes 2}_{\beta,N}(\mathbb{1}_{\sigma^1_1 \dots \sigma^1_i = \sigma^2_1 \dots \sigma^2_i})\right)\right)\Bigg| = 0. \end{aligned} \tag{1.37}$$

It is quite remarkable that these relations suffice to determine completely the distribution of the random measures $\widehat{\mathcal{W}}_\beta$ through their generalised moments. In the REM, this fact was first pointed out by Talagrand [34].

Theorem 1.13. *The random measures $\widehat{\mathcal{W}}_\beta$ are completely determined by the relations* (1.37) *up to the mean value of the two-replica distance distribution function f_β given in Proposition* 1.11.

Proof. Since it is important to understand this result, we will prove this theorem immediately. Note first that the last term in (1.37) is just

$$\lim_{N\uparrow\infty} \mathbb{E}\mu_{\beta,N}^{\otimes 2}(\mathbb{1}_{\sigma_1^1\ldots\sigma_i^1=\sigma_1^2\ldots\sigma_i^2}) = \mathbb{E}\mu_{\beta,N}^{\otimes 2}(\mathbb{1}_{d_N(\sigma,\sigma')\geqslant\sum_{j=1}^i \ln\alpha_j/\ln 2}) = 1-\mathbb{E}f_\beta\left(\sum_{j=1}^i \ln\alpha_j/\ln 2\right). \tag{1.38}$$

Let now $\underline{\underline{J}} \equiv (\underline{J}_0,\ldots,\underline{J}_n)$ denote a nested set of finite subsets of $\mathbb{N}$ such that each $\underline{J}_r = (J_{r,1},\ldots,J_{r,j_r})$ 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 $J_{r,i}\equiv\{j_1^{r,i},\ldots,j_{|J_{r,i}|}^{r,i}\}$. Given such a $\underline{\underline{J}}$, define the events

$$\mathcal{A}_{\underline{\underline{J}}} \equiv \prod_{r=1}^n\prod_{i=1}^{j_r} \mathbb{1}_{\{\sigma_1^{j_1^{r,i}}\ldots\sigma_r^{j_1^{r,i}}=\cdots=\sigma_1^{j_{|J_{r,i}|}^{r,i}}\ldots\sigma_r^{j_{|J_{r,i}|}^{r,i}}\}}. \tag{1.39}$$

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

$$A_{\underline{\underline{J}}} \equiv \lim_{N\uparrow\infty} \mathbb{E}\mu_{\beta,N}^{\otimes J}\mathcal{A}_{\underline{\underline{J}}}. \tag{1.40}$$

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, $1\leqslant r\leqslant n$, such that for some $\underline{\underline{J}}'$ with $J'=J-1$, $l\in\underline{\underline{J}}'$ and $m\notin\underline{\underline{J}}'$:

$$\mathcal{A}_{\underline{\underline{J}}} = \mathcal{A}_{\underline{\underline{J}}'}\mathbb{1}_{\sigma_1^l\ldots\sigma_r^l=\sigma_1^m\ldots\sigma_r^m}. \tag{1.41}$$

Thus using (1.37) 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 (1.38).

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

We will now show that expectations of the numbers $A_{\underline{\underline{J}}}$ fully determine the processes $\widehat{\mathcal{W}}_\beta$ defined in Theorem 1.9 (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 $f_1,\ldots,f_k$ on $\mathbb{R}^d$,

$$F\left(\int \mathcal{M}_N(dx)f_1(x),\ldots,\int \mathcal{M}_N(dx)f_k(x)\right)\to F\left(\int \mathcal{M}(dx)f_1(x),\ldots,\int \mathcal{M}(dx)f_k(x)\right) \tag{1.42}$$

(see e.g. [21,28]) By standard approximation arguments, if $\mathcal{M}$ has moments of all orders, it is enough to consider for f_i all monomial functions $f(x)=x_1^{i_1}\ldots x_d^{i_d}$, and $F(y)=y_1\ldots y_k$. Now let $\mathcal{W}_{\beta,N}^{(m)}$ be any process of the form (1.31). Then

$$\begin{aligned}\int \mathcal{W}_{\beta,N}^{(m)}(dw)w_1^{i_1}\ldots w_l^{i_l}\ldots w_m^{i_m}\\ = \sum_\sigma\Bigg(\sum_{\sigma_2',\ldots,\sigma_m'}\mu_{\beta,N}(\sigma_1\sigma_2',\ldots,\sigma_m')\Bigg)^{i_1}\ldots\Bigg(\sum_{\sigma_{l+1}',\ldots,\sigma_m'}\mu_{\beta,N}(\sigma_1\ldots\sigma_l\sigma_{l+1}'\ldots\sigma_m')\Bigg)^{i_l}\cdots\end{aligned}$$

$$\begin{aligned}
&\times \big(\mu_{\beta,N}(\sigma_1\dots\sigma_m)\big)^{i_m}\\
&= \mu_{\beta,N}^{\otimes(i_1+\dots+i_m)}\big(\mathbb{1}_{\{\sigma_1^1=\dots=\sigma_1^{i_1+\dots+i_m}\}}\cdots\mathbb{1}_{\{\sigma_1^{i_1+\dots+i_{l-1}+1}\dots\sigma_l^{i_1+\dots+i_{l-1}+1}=\dots=\sigma_1^{i_1+\dots+i_m}\dots\sigma_l^{i_1+\dots+i_m}\}}\cdots\\
&\quad\times \mathbb{1}_{\{\sigma_1^{i_1+\dots+i_{m-1}+1}\dots\sigma_m^{i_1+\dots+i_{m-1}+1}=\dots=\sigma_1^{i_1+\dots+i_m}\dots\sigma_m^{i_1+\dots+i_m}\}}\big),
\end{aligned}\tag{1.43}$$

where $i_m \geqslant 1$ (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 1.13. □

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 [30]:

Theorem 1.14. *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

$$\begin{aligned}
\int \mathcal{R}_{\beta,N}^{(m)}(dw)w^r &= \sum_\sigma\Bigg(\sum_{\sigma'_{m+1},\dots,\sigma'_n}\mu_{\beta,N}(\sigma_1\dots\sigma_m\sigma'_{m+1}\dots\sigma'_n)\Bigg)^r\\
&= \mu_{\beta,N}^{\otimes r}\big(\mathbb{1}_{\{\sigma_1^1\dots\sigma_m^1=\dots=\sigma_1^r\dots\sigma_m^r\}}\big).
\end{aligned}\tag{1.44}$$

Thus we have the same type of objects as in the REM, and since the relations (1.37) have the same form for all m, the claimed result follows from the preceding theorem. □

In the remainder of this paper we will present the proofs of the theorems stated above.

2. Proofs of point process convergence results

We use the relation $A(N) \sim B(N)$ to indicate that $\lim_{N\uparrow\infty}\frac{A(N)}{B(N)} = 1$. The technical key to Theorem 1.1 is the following lemma:

Lemma 2.1. *Let $n \in \mathbb{N}$, $n \geqslant 1$, $0 < a_i < 1$, $\alpha_i > 1$, $i = 1,2,\dots,n$, and $\bar\alpha \equiv \prod_{i=1}^n \alpha_i$. Assume that $\sum_{i=1}^n a_i = 1$. If all inequalities* (1.8) *hold true, then for all $x \in \mathbb{R}$*

$$\mathbb{P}\big(\forall\sigma\in\{-1,+1\}^{\ln\bar\alpha/\ln 2}\colon \sqrt{a_1}X_{\sigma_1} + \sqrt{a_2}X_{\sigma_1\sigma_2} + \dots + \sqrt{a_n}X_{\sigma_1\sigma_2\dots\sigma_n} < u_{\ln\bar\alpha,N}(x)\big) \sim e^{-Ke^{-x}},\tag{2.1}$$

as $N\uparrow\infty$, where K is defined in Theorem 1.1. *If at least one of inequalities* (1.8) *fails, then*

$$\mathbb{P}\big(\forall\sigma\in\{-1,+1\}^{\ln\bar\alpha/\ln 2}\colon \sqrt{a_1}X_{\sigma_1} + \sqrt{a_2}X_{\sigma_1\sigma_2} + \dots + \sqrt{a_n}X_{\sigma_1\sigma_2\dots\sigma_n} < u_{\ln\bar\alpha,N}(x)\big) \sim 1.\tag{2.2}$$

Proof. If $n = 1$, then all $\bar\alpha^N$ random variables in the event of (2.1) are independent. In this case the statement of the lemma is a classical result of extreme values theory. Let now $n \geqslant 2$. We define

$$\begin{aligned}
P(i,y) \equiv \mathbb{P}\Bigg(&\exists\sigma_{i+1}\in\{-1,+1\}^{\ln\alpha_i/\ln 2},\dots,\sigma_n\in\{-1,+1\}^{\ln\alpha_n/\ln 2}\colon\\
&\frac{\sqrt{a_i}X + \sqrt{a_{i+1}}X_{\sigma_{i+1}} + \sqrt{a_{i+2}}X_{\sigma_{i+1}\sigma_{i+2}} + \dots + \sqrt{a_n}X_{\sigma_{i+1}\dots\sigma_n}}{\sqrt{a_i+\dots+a_n}} > y\Bigg),
\end{aligned}\tag{2.3}$$

where $X, X_{\sigma_{i+1}}, \dots, X_{\sigma_{i+1}\dots\sigma_n}$ are all i.i.d. standard normal r.v.'s. We will estimate these probabilities by induction downwards from $i = n-1$, using the identity

$$P(i,y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(1 - \left[1 - P\left(i+1, \frac{\sqrt{a_i + \dots + a_n}\,y - \sqrt{a_i}\,t}{\sqrt{a_{i+1} + \dots + a_n}}\right)\right]^{\alpha_{i+1}^N}\right) e^{-t^2/2}\,dt. \tag{2.4}$$

First we prove the assertion of the lemma under assumption that all inequalities in (1.8) are strict. In this case we will prove inductively for $i = n-1, n-2, \dots, 1$ that

$$P(i,y) \sim \frac{(\alpha_{i+1}\cdots\alpha_n)^N e^{-y^2/2}}{\sqrt{2\pi}\,y}, \tag{2.5}$$

as $N \uparrow \infty$, uniformly for $y \in [\sqrt{2(\ln\bar{\alpha} - \delta_i)N(a_i + \dots + a_n)}, \infty)$ with some $\delta_i > 0$.

Note that $u_{\ln\bar{\alpha},N}(x) \in [\sqrt{2(\ln\bar{\alpha} - \delta_1)N}, \infty)$ for any $x \in \mathbb{R}$, and thus

$$P\big(1, u_{\ln\bar{\alpha},N}(x)\big) \sim (\alpha_2\cdots\alpha_n)^N e^{-u^2_{\ln\bar{\alpha},N}(x)/2}/\sqrt{2\pi}\,u_{\ln\bar{\alpha},N}(x) \sim \alpha_1^{-N} e^{-x}.$$

Since the probability in the l.h.s. of (2.1) equals $(1 - P(1, u_{\ln\bar{\alpha},N}(x)))^{\alpha_1^N} \sim e^{\alpha_1^N \ln(1-\alpha_1^{-N}e^{-x})} \sim e^{-e^{-x}}$, (2.5) indeed proves the assertion of the lemma in this case.

Let us now turn to the proof of (2.5). Note that $P(i,y) \leqslant (\alpha_{i+1}\cdots\alpha_N)^N P(X > y)$, where X is a standard Gaussian random variable, and so the upper bound for (2.5) is immediate by Proposition A.1 of the appendix. We will establish the lower bound for (2.5) by induction. Let first $i = n-1$. The identity (2.4) reads

$$P(n-1,y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(1 - \left[1 - \mathbb{P}\left(X > \frac{\sqrt{a_{n-1} + a_n}\,y - \sqrt{a_{n-1}}\,t}{\sqrt{a_n}}\right)\right]^{\alpha_n^N}\right) e^{-t^2/2}\,dt. \tag{2.6}$$

Let us fix $\delta'_{n-1} > 0$ such that $\ln\alpha_n - (\ln\bar{\alpha} - \delta'_{n-1})a_n < 0$, which is possible only due to the strict inequality in (1.8) for $i = n$. For all y, t satisfying

$$\frac{\sqrt{a_{n-1} + a_n}\,y - \sqrt{a_{n-1}}\,t}{\sqrt{a_n}} \geqslant \sqrt{2(\ln\bar{\alpha} - \delta'_{n-1})Na_n} \tag{2.7}$$

by Proposition A.1 we have

$$\alpha_n^N \mathbb{P}\left(X > \frac{\sqrt{a_{n-1} + a_n}\,y - \sqrt{a_{n-1}}\,t}{\sqrt{a_n}}\right) \leqslant \alpha_n^N e^{(\ln\bar{\alpha} - \delta'_{n-1})Na_n} \to 0, \quad N \uparrow \infty.$$

Then, using the elementary inequalities

$$-x - x^2 \leqslant \ln(1-x) \leqslant -x, \quad 1 + x \leqslant e^x \leqslant 1 + x + x^2 \quad \text{for } |x| < 1/2 \tag{2.8}$$

and again Proposition A.1, it is easy to deduce the asymptotics of the integrand in (2.6):

$$\begin{aligned} &1 - \left[1 - \mathbb{P}\left(X > \frac{\sqrt{a_{n-1} + a_n}\,y - \sqrt{a_{n-1}}\,t}{\sqrt{a_n}}\right)\right]^{\alpha_n^N} \\ &\quad \sim \frac{\alpha_n^N \sqrt{a_n}}{\sqrt{2\pi}(\sqrt{a_{n-1} + a_n}\,y - \sqrt{a_{n-1}}\,t)} e^{-(\sqrt{a_{n-1}+a_n}\,y - \sqrt{a_{n-1}}\,t)^2/(2a_n)}, \quad \text{as } N \uparrow \infty, \end{aligned}$$

uniformly for all y, t satisfying (2.7). Observe that condition (2.7) is equivalent to $t \leqslant t_{cr}(y) \equiv (\sqrt{a_{n-1} + a_n}\,y - \sqrt{2(\ln\bar{\alpha} - \delta'_{n-1})Na_n})/\sqrt{a_{n-1}}$. Thus in view of (2.6), $P(n-1,y)$ obeys the following asymptotic lower bound

$$P(n-1,y) \geqslant \frac{\alpha_n^N}{2\pi} \int_{-\infty}^{t_{cr}(y)} \frac{\sqrt{a_n}}{\sqrt{2\pi}(\sqrt{a_{n-1} + a_n}\,y - \sqrt{a_{n-1}}\,t)} e^{-(\sqrt{a_{n-1}+a_n}\,y - \sqrt{a_{n-1}}\,t)^2/(2a_n) - t^2/2}\,dt,$$

as $N \uparrow \infty$, uniformly in $y \in \mathbb{R}$. After the change of variables $s = (\sqrt{a_{n-1}+a_n}t - \sqrt{a_{n-1}}y)/\sqrt{a_n}$ the last integral becomes

$$\frac{\alpha_n^N e^{-y^2/2}}{\sqrt{2\pi}\,y}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{s_{cr}(y)}\frac{e^{-s^2/2}}{1-\frac{\sqrt{a_{n-1}}s}{\sqrt{a_n}y}}\,ds,$$

where

$$s_{cr}(y) = \bigl(y - \sqrt{2(\ln\bar{\alpha} - \delta'_{n-1})N(a_{n-1}+a_n)}\bigr)\sqrt{a_n}/\sqrt{a_{n-1}}.$$

It yields the necessary asymptotic lower bound for (2.5) as $N \uparrow \infty$ uniform for $y \in [\sqrt{2(\ln\bar{\alpha} - \delta_{n-1})N(a_{n-1}+a_n)}, \infty)$ if we set $0 < \delta_{n-1} < \delta'_{n-1}$.

Assuming that (2.1) is proved for $i = n-1, n-2, \dots, l+1$, we will prove it for $i = l$. Let us fix $\delta'_l > 0$ such that $\delta'_l \leqslant \delta_{l+1}$ and $\ln(\alpha_{l+1}\cdots\alpha_n) - (\ln\bar{\alpha} - \delta'_l)(a_{l+1}+\cdots+a_n) < 0$ due to the strict inequality in (1.8) for $i = l+1$. Then for all y and t verifying the condition

$$\frac{\sqrt{a_l+\cdots+a_n}y - \sqrt{a_l}t}{\sqrt{a_{l+1}+\cdots+a_n}} \geqslant \sqrt{2(\ln\bar{\alpha} - \delta'_l)N(a_{l+1}+\cdots+a_n)} \tag{2.9}$$

we have by the induction hypothesis (2.1) for $i = l+1$

$$\begin{aligned}\alpha_{l+1}^N P\left(l+1, \frac{\sqrt{a_l+\cdots+a_n}y - \sqrt{a_l}t}{\sqrt{a_{l+1}+\cdots+a_n}}\right) &\leqslant (\alpha_{l+1}\cdots\alpha_n)^N e^{-(\sqrt{a_l+\cdots+a_n}y - \sqrt{a_l}t)^2/2(a_{l+1}+\cdots+a_n)}\\ &\leqslant (\alpha_{l+1}\cdots\alpha_n)^N e^{-(\ln\bar{\alpha}-\delta'_l)N(a_{l+1}+\cdots+a_n)},\end{aligned}$$

which tends to zero, as $N \uparrow \infty$. Furthermore, again by the induction hypothesis for $i = l+1$ and elementary exponential inequalities (2.8)

$$\begin{aligned}&1 - \left[1 - P\left(l+1, \frac{\sqrt{a_l+\cdots+a_n}y - \sqrt{a_l}t}{\sqrt{a_{l+1}+\cdots+a_n}}\right)\right]^{\alpha_{l+1}^N}\\ &\quad\sim \frac{(\alpha_{l+1}\cdots\alpha_n)^N\sqrt{a_{l+1}+\cdots+a_n}}{\sqrt{2\pi}(\sqrt{a_l+\cdots+a_n}y - \sqrt{a_l}t)}e^{-(\sqrt{a_l+\cdots+a_n}y - \sqrt{a_l}t)^2/2(a_{l+1}+\cdots+a_n)}\end{aligned}$$

uniformly for all y, t satisfying (2.9), that is for

$$t \leqslant t_{cr(y)} \equiv \bigl(\sqrt{a_l+\cdots+a_n}y - \sqrt{2(\ln\bar{\alpha} - \delta'_l)N}(a_{l+1}+\cdots+a_n)\bigr)/\sqrt{a_l}.$$

In view of this estimate the identity (2.4) for $i = l$ gives the following asymptotic lower bound for $P(l, y)$ as $N \uparrow \infty$ uniform for $y \in \mathbb{R}$:

$$\begin{aligned}P(l,y) &\geqslant \frac{(\alpha_{l+1}\cdots\alpha_n)^N\sqrt{a_{l+1}+\cdots+a_n}}{2\pi}\int_{-\infty}^{t_{cr}(y)}\frac{e^{-(\sqrt{a_l+\cdots+a_n}y - \sqrt{a_l}t)^2/2(a_{l+1}+\cdots+a_n) - t^2/2}}{\sqrt{a_l+\cdots+a_n}y - \sqrt{a_l}t}\,dt\\ &= \frac{(\alpha_{l+1}\cdots\alpha_n)^N e^{-y^2/2}}{\sqrt{2\pi}\,y}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{s_{cr}(y)}\frac{e^{-s^2/2}}{1-\frac{\sqrt{a_l}s}{\sqrt{a_{l+1}+\cdots+a_n}y}}\,ds,\end{aligned}$$

where

$$s_{cr}(y) \equiv \bigl(y - \sqrt{(\ln\bar{\alpha} - \delta'_l)N(a_l+\cdots+a_n)}\bigr)\sqrt{a_{l+1}+\cdots+a_n}/\sqrt{a_l}. \tag{2.10}$$

This bound has the asymptotics stated in (2.5), uniformly in $y \in [\sqrt{(\ln\bar{\alpha} - \delta_l)N(a_l + \cdots + a_n)}, \infty)$ if only $\delta_l > 0$ is chosen to be $\delta_l < \delta_l'$.

The analysis is more delicate whenever some of the inequalities (1.8) degenerate into equalities. Assume that $k = \max\{i > 1, a_i + \cdots + a_n = \ln(\alpha_i \cdots \alpha_n)/\ln\bar{\alpha}\}$. Then for $i = n-1, n-2, \ldots, k$ (2.5) remains valid in the respective domains of y. For $i = k-1$ we will prove the following two assertions $(\mathbf{a}_{k-1})$ and $(\mathbf{b}_{k-1})$.

$(\mathbf{a}_{k-1})$ For any $C \in \mathbb{R}$ we have

$$P(k-1, y) \sim \frac{(\alpha_k \cdots \alpha_n)^N e^{-y^2/2}}{\sqrt{2\pi}\, y} \times \mathbb{P}\left(Z_{k-1} > \left[y - \sqrt{2(\ln\bar{\alpha})(a_{k-1} + \cdots + a_n)N}\right]\sqrt{\frac{a_k + \cdots + a_n}{a_{k-1}}}\right), \quad \text{as } N \uparrow \infty, \tag{2.11}$$

uniformly in $y \in [\sqrt{2\ln\bar{\alpha}(a_{k-1} + \cdots + a_n)N} + C, \infty)$.

$(\mathbf{b}_{k-1})$ For any $\varepsilon > 0$, there exist a negative constant $\tilde{C}_{k-1}(\varepsilon) \in \mathbb{R}$ of large enough absolute value, and $N(\varepsilon) \in \mathbb{N}$, such that, for all $N \geqslant N(\varepsilon)$ and all $y \in (0, \sqrt{2\ln\bar{\alpha}(a_{k-1} + \cdots + a_n)N} + \tilde{C}_{k-1}(\varepsilon)]$

$$P(k-1, y) \leqslant \frac{\varepsilon(\alpha_k \cdots \alpha_n)^N e^{-y^2/2}}{\sqrt{2\pi}\, y}. \tag{2.12}$$

The proof is based again on (2.4) for $i = k-1$. First of all, let us remark that the asymptotics of $P(k, y)$ is determined by (2.5) uniformly in the interval of $y \in [u_{\ln(\alpha_k \cdots \alpha_n), N}(x), \infty)$ for any $x \in \mathbb{R}$. This is ensured by the equality (1.1) for $i = k$, since consequently

$$u_{\ln(\alpha_k \ldots \alpha_n), N}(x) = \sqrt{2(\ln\bar{\alpha})N(a_k + \cdots + a_n)} + \mathrm{o}(1), \quad N \uparrow \infty. \tag{2.13}$$

Therefore for any $x \in \mathbb{R}$

$$\begin{aligned}
&\alpha_k^N P\left(k, \frac{\sqrt{a_{k-1} + \cdots + a_n}\, y - \sqrt{a_{k-1}}\, t}{\sqrt{a_k + \cdots + a_n}}\right) \\
&\quad \sim \frac{(\alpha_k \cdots \alpha_n)^N \sqrt{a_k + \cdots + a_n}}{\sqrt{2\pi}(\sqrt{a_{k-1} + \cdots + a_n}\, y - \sqrt{a_{k-1}}\, t)} 7 e^{-(\sqrt{a_{k-1} + \cdots + a_n}\, y - \sqrt{a_{k-1}}\, t)^2/2(a_k + \cdots + a_n)} \\
&\quad \leqslant \frac{(\alpha_k \cdots \alpha_n)^N e^{-u^2_{\ln(\alpha_k \cdots \alpha_n), N}(x)/2}}{\sqrt{2\pi}\, u_{\ln(\alpha_k \cdots \alpha_n), N}(x)} \sim e^{-x}, \quad N \uparrow \infty,
\end{aligned}$$

uniformly for y, t such that

$$\frac{\sqrt{a_{k-1} + \cdots + a_n}\, y - \sqrt{a_{k-1}}\, t}{\sqrt{a_k + \cdots + a_n}} \geqslant u_{\ln(\alpha_k \cdots \alpha_n), N}(x). \tag{2.14}$$

Then, given an arbitrary $\varepsilon > 0$, using the inequalities (2.8), one can choose a sufficiently large $x_0 = x_0(\varepsilon) \in \mathbb{R}$ and $N(\varepsilon, x_0)$ to bound the integrand in (2.4) for $i = k-1$ as follows:

$$\begin{aligned}
&(1-\varepsilon)\frac{(\alpha_k \cdots \alpha_n)^N \sqrt{a_k + \cdots + a_n}}{\sqrt{2\pi}(\sqrt{a_{k-1} + \cdots + a_n}\, y - \sqrt{a_{k-1}}\, t)} e^{-(\sqrt{a_{k-1} + \cdots + a_n}\, y - \sqrt{a_{k-1}}\, t)^2/2(a_k + \cdots + a_n)} \\
&\quad \leqslant 1 - \left[1 - P\left(k, \frac{\sqrt{a_{k-1} + \cdots + a_n}\, y - \sqrt{a_{k-1}}\, t}{\sqrt{a_k + \cdots + a_n}}\right)\right]^{\alpha_k^N} \\
&\quad \leqslant (1+\varepsilon)\frac{(\alpha_k \cdots \alpha_n)^N \sqrt{a_k + \cdots + a_n}}{\sqrt{2\pi}(\sqrt{a_{k-1} + \cdots + a_n}\, y - \sqrt{a_{k-1}}\, t)} e^{-(\sqrt{a_{k-1} + \cdots + a_n}\, y - \sqrt{a_{k-1}}\, t)^2/2(a_k + \cdots + a_n)}
\end{aligned} \tag{2.15}$$

for all $N \geqslant N(\varepsilon, x_0)$ and all y, t under condition (2.14), that is for t up to $t_{cr}(y) \equiv (y\sqrt{a_{k-1}+\cdots+a_n} - u_{\ln(\alpha_k\cdots\alpha_n),N}(x_0)\sqrt{a_k+\cdots+a_n})/\sqrt{a_{k-1}}$. After making this choice, let us decompose the integral (2.4) for $i = k-1$ into two terms I_1 and I_2: in the first one the integration is from $-\infty$ to $t_{cr}(y)$ and in the second from $t_{cr}(y)$ to $+\infty$. In view of (2.15), for all $N \geqslant N(\varepsilon, x_0)$, I_1 is bounded from below and above by

$$\frac{1}{2\pi}\int_{-\infty}^{t_{cr}(y)} \frac{(\alpha_k\cdots\alpha_n)^N\sqrt{a_k+\cdots+a_n}}{(\sqrt{a_{k-1}+\cdots+a_n}y-\sqrt{a_{k-1}}t)} e^{-(\sqrt{a_{k-1}+\cdots+a_n}y-\sqrt{a_{k-1}}t)^2/2(a_k+\cdots+a_n)-t^2/2}\,dt$$
$$= \frac{(\alpha_k\cdots\alpha_n)^N e^{-y^2/2}}{2\pi y}\int_{-\infty}^{(y\sqrt{a_k+\cdots+a_n}-u_{\ln(\alpha_k\cdots\alpha_n),N}(x_0)\sqrt{a_{k-1}+\cdots+a_n})/\sqrt{a_{k-1}}} \frac{e^{-s^2/2}}{1-\frac{\sqrt{a_{k-1}}s}{\sqrt{a_k+\cdots+a_n}y}}\,ds$$

times the factors $(1-\varepsilon)$ and $(1+\varepsilon)$ respectively. Combined with the observation (2.13), this integral yields the asymptotics stated in $(\mathbf{a}_{k-1})$. Moreover for any $\varepsilon > 0$ one can find $\tilde{C}_{k-1}(\varepsilon)$ and $N(\varepsilon)$ such that for all $y \in (0, \sqrt{2(\ln\bar{\alpha})N(a_{k-1}+\cdots+a_n)} + \tilde{C}_{k-1}(\varepsilon)]$ and $N \geqslant N(\varepsilon)$ I_1 obeys the bound of $(\mathbf{b}_{k-1})$.

The term I_2 brings no significant contribution to the asymptotics of $(\mathbf{a}_{k-1})$ as it verifies the bound (2.12) uniformly on the whole interval $y \in (0, \infty)$. To show this, we simply bound its integrand from above by $e^{-t^2/2}$ and consider separately two intervals of y:

$$\Delta_1 = \left(0, \sqrt{\frac{2(a_k+\cdots+a_n)\ln(\alpha_k\cdots\alpha_n)+\tilde{\delta}_{k-1}}{a_{k-1}+\cdots+a_n}N}\right] \quad \text{and}$$
$$\Delta_2 = \left(\sqrt{\frac{2(a_k+\cdots+a_n)\ln(\alpha_k\cdots\alpha_n)+\tilde{\delta}_{k-1}}{a_{k-1}+\cdots+a_n}N}, \infty\right),$$

where $\tilde{\delta}_{k-1}$ is any fixed number $0 < \tilde{\delta}_{k-1} < a_{k-1}$. The estimate of I_2 in Δ_1 is trivial: for any $\varepsilon > 0$ one can find $N(\varepsilon)$ such that for all $N \geqslant N(\varepsilon)$ and all $y \in \Delta_1$

$$I_2 \leqslant \frac{\varepsilon\ln(\alpha_k\cdots\alpha_n)^N e^{-N((a_k+\cdots+a_n)\ln(\alpha_k\cdots\alpha_n)+\tilde{\delta}_{k-1})/(a_{k-1}+\cdots+a_n)}}{\sqrt{2\pi\frac{2(a_k+\cdots+a_n)\ln(\alpha_k\cdots\alpha_n)+\tilde{\delta}_{k-1}}{a_{k-1}+\cdots+a_n}N}} \leqslant \frac{\varepsilon\ln(\alpha_k\cdots\alpha_n)^N e^{-y^2/2}}{\sqrt{2\pi}\,y}. \tag{2.16}$$

Since $\tilde{\delta}_{k-1} > 0$, then for all $y \in \Delta_2$ $\tau_{cr}(y) > 0$. Here Proposition A.1 applies:

$$I_2 \leqslant \frac{1}{\sqrt{2\pi}t_{cr}(y)} e^{-t_{cr}(y)^2/2} \leqslant \frac{(\alpha_k\cdots\alpha_n)^N C(\tilde{\delta}_{k-1})}{\sqrt{2\pi}\,y} e^{-y^2/2} e^{f(y)/2}, \tag{2.17}$$

where

$$f(y) = \big(-y^2(a_k+\cdots+a_n) - u^2_{\ln(\alpha_k\cdots\alpha_n),N}(x_0)(a_k+\cdots+a_n)$$
$$+ 2u_{\ln(\alpha_k\cdots\alpha_n),N}(x_0)y\sqrt{a_{k-1}+\cdots+a_n}\sqrt{a_k+\cdots+a_n} - 2a_{k-1}N\ln(\alpha_k\cdots\alpha_n)\big)/a_{k-1}.$$

The function $f(y)$ reaches its maximum at

$$y = u_{\ln(\alpha_k\cdots\alpha_n),N}(x_0)\sqrt{a_{k-1}+\cdots+a_n}/\sqrt{a_k+\cdots+a_n}$$

where it equals

$$u^2_{\ln(\alpha_k\cdots\alpha_n),N}(x_0) - 2N\ln(\alpha_k\cdots\alpha_n) = -\ln N - \ln\ln(\alpha_k\cdots\alpha_n) - \ln 4\pi + 2x_0 + o(1).$$

Hence, $e^{f(y)/2}$ being of the order at most $N^{-1/2}$, I_2 does not contribute to the asymptotics of $(\mathbf{a}_{k-1})$ and satisfies the bound of $(\mathbf{b}_{k-1})$ for all $y \in \Delta_2$ as well when N is bigger some $N(\varepsilon)$. This concludes the proof of $(\mathbf{a}_{k-1})$ and $(\mathbf{b}_{k-1})$.

To continue the proof of the lemma, let us assume that the inequalities (1.8) are strict for $i = k-1, k-2, \dots, l-1$ and $(\mathbf{a}_i^1)$, $(\mathbf{b}_i)$ (see below) hold true for $i = k-2, \dots, l$ respectively. Under these assumptions we show that

- if the inequality (1.8) is also strict for $i = l$, then $(\mathbf{a}_{l-1}^1)$ and $(\mathbf{b}_{l-1})$ remain true;
- if (1.8) turns into the equality for $i = l$, then $(\mathbf{a}_{l-1}^2)$ and $(\mathbf{b}_{l-1})$ take place.

The assertion $(\mathbf{b}_i)$ is the same as $(\mathbf{b}_{k-1})$ above with $k-1$ replaced by i. The assertions $(\mathbf{a}_i^1)$, $(\mathbf{a}_i^2)$ are the following:

$(\mathbf{a}_i^1)$ For any $C \in \mathbb{R}$ we have

$$P(i, y) \sim \frac{(\alpha_{i+1} \cdots \alpha_n)^N e^{-y^2/2}}{\sqrt{2\pi}\, y} \mathbb{P}\big(A_i^k(y)\big)$$

uniformly in the interval of $y \in [\sqrt{2 \ln \bar{\alpha}(a_i + \cdots + a_n)N} + C, \infty)$.

$(\mathbf{a}_i^2)$ For any $C \in \mathbb{R}$ we have

$$P(i, y) \sim \frac{(\alpha_{i+1} \cdots \alpha_n)^N e^{-y^2/2}}{\sqrt{2\pi}\, y} \times \mathbb{P}\left(A_i^k(y) \cap \left[Z_i\big(y - \sqrt{2 \ln \bar{\alpha}(a_i + \cdots + a_n)N}\big)\sqrt{\frac{a_{i+1} + \cdots + a_n}{a_i}}\right]\right)$$

uniformly in the interval of $y \in [\sqrt{2 \ln \bar{\alpha}(a_i + \cdots + a_n)N} + C, \infty)$. The events $A_i^k(y)$ are defined as:

$$\begin{aligned} A_i^k(y) \equiv \Bigg(Z_{k-1} > \Bigg[& y\sqrt{\frac{a_{k-1} + \cdots + a_n}{a_i + \cdots + a_n}} - Z_i \sqrt{\frac{a_i(a_{k-1} + \cdots + a_n)}{(a_i + \cdots + a_n)(a_{i+1} + \cdots + a_n)}} \\ & - Z_{i+1}\sqrt{\frac{a_{i+1}(a_{k-1} + \cdots + a_n)}{(a_{i+1} + \cdots + a_n)(a_{i+2} + \cdots + a_n)}} - \cdots - Z_{k-3}\sqrt{\frac{a_{k-3}(a_{k-1} + \cdots + a_n)}{(a_{k-3} + \cdots + a_n)(a_{k-2} + \cdots + a_n)}} \\ & - Z_{k-2}\sqrt{\frac{a_{k-2}}{a_{k-2} + \cdots + a_n}} - \sqrt{2(\ln \bar{\alpha})(a_{k-1} + \cdots + a_n)N} \Bigg] \sqrt{\frac{a_k + \cdots + a_n}{a_{k-1}}} \Bigg). \end{aligned} \tag{2.18}$$

By virtue of $(\mathbf{a}_l^1)$, $(\mathbf{b}_l)$, and the inequalities (2.8), for any given $\varepsilon > 0$ there exist a sufficiently large $x_0 = x_0(\varepsilon) \in \mathbb{R}$, and $N(x_0, \varepsilon)$, such that, for all $N \geqslant N(x_0, \varepsilon)$, all $y \in \mathbb{R}$, and all $t \leqslant t_{cr}(y) \equiv (y\sqrt{a_{l-1} + \cdots + a_n} - u_{\ln(\alpha_l \cdots \alpha_n), N}(x_0)\sqrt{a_l + \cdots + a_n})/\sqrt{a_{l-1}}$, the integrand in (1.8) for $i = l-1$ can be estimated from above and below by $\alpha_l^N P(l, \cdot) e^{-t^2/2}(1 \pm \varepsilon)$ times. Having x_0 fixed, we again split the integral (1.8) into two terms, I_1 and I_2, where the integration is from $-\infty$ to $t_{cr}(y)$ and from $t_{cr}(y)$ to $+\infty$, respectively. By the same arguments as in (2.17), the term I_2 does not contribute to $(\mathbf{a}_{l-1}^1)$ or $(\mathbf{a}_{l-1}^2)$, and satisfies the bound of condition $(\mathbf{b}_{l-1})$.

To treat the term I_1 we distinguish the two cases when the inequality (1.8) is strict for $i = l$, and when equality holds, respectively.

In the second case, the assumed equality entails

$$u_{\ln(\alpha_l \cdots \alpha_n), N}(x_0) = \sqrt{2 \ln \bar{\alpha}(a_l + \cdots + a_n)N} + \mathrm{o}(1). \tag{2.19}$$

This means that the probability $P(l, \cdot)$ entering the integrand of I_1 has the asymptotics claimed in $(\mathbf{a}_l^1)$ uniformly in the interval of integration $t \leqslant t_{cr}(y)$ and all y. Then for any $\varepsilon > 0$, one can find $N(x_0, \varepsilon)$ such that, for all $N \geqslant N(x_0, \varepsilon)$, the term I_1 is bounded from above and below by

$$\frac{(\alpha_l\cdots\alpha_n)^N\sqrt{a_l+\cdots+a_n}}{2\pi}\int_{-\infty}^{t_{cr}(y)}\frac{e^{-(\sqrt{a_{l-1}+\cdots+a_n}y-\sqrt{a_{l-1}}t)^2/2(a_l+\cdots+a_n)-t^2/2}}{\sqrt{a_{l-1}+\cdots+a_n}y-\sqrt{a_{l-1}}t}$$
$$\times\mathbb{P}\left(A_l^k\left(\frac{y\sqrt{a_{l-1}+\cdots+a_n}-t\sqrt{a_{l-1}}}{\sqrt{a_l+\cdots+a_n}}\right)\right)dt\,(1\pm\varepsilon)^2 \tag{2.20}$$
$$=\frac{(\alpha_l\cdots\alpha_n)^N e^{-y^2/2}}{2\pi y}\int_{-\infty}^{s_{cr}(y)}\frac{e^{-s^2/2}}{1-\frac{\sqrt{a_{l-1}}s}{\sqrt{a_l+\cdots+a_n}y}}$$
$$\times\mathbb{P}\left(A_l^k\left(y\sqrt{\frac{a_l+\cdots+a_n}{a_{l-1}+\cdots+a_n}}-s\sqrt{\frac{a_{l-1}}{a_{l-1}+\cdots+a_n}}\right)\right)ds\,(1\pm\varepsilon)^2 \tag{2.21}$$
$$=\frac{(\alpha_l\cdots\alpha_n)^N e^{-y^2/2}}{2\pi y}\int_{-\infty}^{s_{cr}(y)}\frac{e^{-s^2/2}}{1-\frac{\sqrt{a_{l-1}}s}{\sqrt{a_l+\cdots+a_n}y}}\mathbb{P}\big(A_{l-1}^k(y,s)\big)\,ds\,(1\pm\varepsilon)^2,$$

where $s_{cr}(y)=(y\sqrt{a_l+\cdots+a_n}-u_{\ln(\alpha_l\cdots\alpha_n),N}(x_0)\sqrt{a_{l-1}+\cdots+a_n})/\sqrt{a_{l-1}}$ and $A_{l-1}^k(y,s)$ is the event $A_{l-1}^k(y)$ with Z_{l-1} replaced by the variable of integration s. Finally, we substitute (2.19) into the upper limit of integration $s_{cr(y)}$ and the assertions $(\mathbf{a}_{l-1}^2)$ and $(\mathbf{b}_{l-1})$ follow.

Let us now turn to the case of the inequality (1.8) strict for $i=l$. Given $\varepsilon>0$, one can choose a negative constant $\tilde{C}_l(\varepsilon)$ of sufficiently large modulus such that $(\mathbf{a}_l^1)$ is satisfied and that

$$P(Z<\tilde{C}_l(\varepsilon)\sqrt{a_{l-1}+\cdots+a_n}/\sqrt{a_{l-1}})<\varepsilon,\quad \text{where } Z\in\mathcal{N}(0,1). \tag{2.22}$$

Next, let us decompose $I_1\equiv I_1'+I_2''$, s.t. in I_1' the integration is from $-\infty$ to $\tilde{t}_{cr}(y)$ and in I_2 – from $\tilde{t}_{cr}(y)$ to $t_{cr}(y)$, where

$$\tilde{t}_{cr}(y)\equiv\big(y\sqrt{a_{l-1}+\cdots+a_n}-\sqrt{a_l+\cdots+a_n}\big[\sqrt{2(\ln\bar{\alpha})N(a_l+\cdots+a_n)}+\tilde{C}_l(\varepsilon)\big]\big)/\sqrt{a_{l-1}}.$$

The choice of $\tilde{C}_l(\varepsilon)$ enables us to apply to the integrand of I_2'' the estimate $(\mathbf{b}_l)$. It is easy to see, changing variables as in (2.20), that I_2'' is bounded by $\varepsilon(\alpha_l\cdots\alpha_n)^N e^{-y^2/2}/\sqrt{2\pi}\,y$, for all N large enough, and all y in the intervals of $(\mathbf{a}_{l-1}^1)$ and $(\mathbf{b}_{l-1})$.

To consider I_1', we are entitled to estimate its integrand by $(\mathbf{a}_l^1)$ uniformly in the interval of integration. Then for all N large enough this term is bounded from above and below by the integral (2.20) with $t_{cr}(y)$ replaced by $\tilde{t}_{cr}(y)$. Consequently, it equals (2.21) with $s_{cr}(y)$ replaced by

$$\tilde{s}_{cr}(y)=\big(y\sqrt{a_l+\cdots+a_n}-\big[\sqrt{2(\ln\bar{\alpha})N(a_l+\cdots+a_n)}+\tilde{C}_l(\varepsilon)\big]\sqrt{a_{l-1}+\cdots+a_n}\big)/\sqrt{a_{l-1}}.$$

Combined with (2.22), this integral yields the asymptotics $(\mathbf{a}_{l-1}^1)$ and the bound $(\mathbf{b}_{l-1})$.

Following this recurrence procedure up to $i=1$, we obtain for a particular $y=u_{\ln\bar{\alpha},N}(x)$

$$P\big(1,u_{\ln\bar{\alpha},N}(x)\big)\sim K(\alpha_2\cdots\alpha_n)^N e^{-u^2_{\ln\bar{\alpha},N}(x)/2}/\sqrt{2\pi}u_{\ln\bar{\alpha},N}(x)\sim\alpha_1^{-N}Ke^{-x},$$

where K is as defined in Theorem 1.1. Thus the probability in the l.h.s. of (2.1) equals

$$\big(1-P\big(1,u_{\ln\bar{\alpha},N}(x)\big)\big)^{\alpha_1^N}\sim e^{\alpha_1^N\ln(1-K\alpha_1^{-N}e^{-x})}\sim e^{-Ke^{-x}}. \tag{2.23}$$

It remains to establish the assertion (2.2) if for some k, $2\leqslant k\leqslant n$, the inequality (1.8) is violated. For this purpose we prove by induction for $i=k-1,k-2,\ldots,1$ that for any $\varepsilon>0$ and $\bar{\delta}_i>\bar{\delta}_{i+1}>0$ there exists a number $N(\varepsilon,\bar{\delta}_i)$ such that

$$\mathbb{P}(i,y)\leqslant\frac{\varepsilon\ln(\alpha_k\cdots\alpha_n)^N e^{-y^2/2}}{\sqrt{2\pi}\,y} \tag{2.24}$$

for all $N \geqslant N(\varepsilon, \bar{\delta}_i)$ and all $y \in (0, \sqrt{\frac{2(a_i+\cdots+a_n)\ln(\alpha_k\cdots+\alpha_n)-2\bar{\delta}_i}{a_k+\cdots+a_n}N}]$. The arguments of the proof are very similar to those above. Then starting from $\bar{\delta}_{k-1}$ such that $(\mathbf{a}_k + \cdots + a_n)\ln\bar{\alpha} < \ln(\alpha_k \cdots \alpha_n) - \bar{\delta}_{k-1}$ and ending with $\bar{\delta}_1$ sufficiently close to $\bar{\delta}_{k-1}$, we make $y = u_{\ln\bar{\alpha},N}(x)$ belong to the interval of y where (2.24) with $i = 1$ holds true. Thus, $P(1, u_{\ln\bar{\alpha},N}(x)) = \mathrm{o}(\alpha_1^{-N} e^{-x})$ and the assertion (2.2) follows as in (2.23). This completes the proof of the lemma. □

Proof of Theorem 1.1. The necessity of (1.8) is immediate from the assertion (2.2) of Lemma 2.1. The sufficiency follows from Kallenberg's theorem [21] (see also [24]) on the weak convergence of a point process $\mathcal{P}_N$ to the Poisson process $\mathcal{P}$. Applying his theorem in our situation, weak convergence holds whenever

(i) For all intervals $(a, b] \in \mathbb{R}$

$$\mathbb{E}\sum_{\sigma} \mathbb{1}_{\{u^{-1}_{\ln\bar{\alpha},N}(X_\sigma)\in(a,b]\}} \to Ke^{-a} - Ke^{-b}, \quad \text{as } N \uparrow \infty.$$

(ii) For all finite unions of half-open disjoint intervals $\Delta = \Delta_1 \cup \cdots \cup \Delta_k$, $\Delta_i = (a_i, b_i]$, $a_1 < b_1 < a_2 < b_2 < \cdots < a_k < b_k$

$$\mathbb{P}\big\{\forall\sigma\colon u^{-1}_{\ln\bar{\alpha},N}(X_\sigma) \notin \Delta\big\} \to \sum_{i=1}^{k} e^{-K(e^{-a_i}-e^{-b_i})}.$$

The proof of assertion (i) is obvious, and the proof of assertion (ii) is completely analogous to that of Lemma 2.1. We leave the details to the reader. □

Proof of Theorem 1.2. We proceed by induction. In the case $k = 1$, $\mathcal{P}^{(1)} = \mathcal{P}_1$ is the Poisson point process on $\mathbb{R}$ and the statement of the theorem is already contained in the assumptions.

Assume that (1.10) is proved for $k - 1$. Taking into account the identical distribution of the r.v. $Y_{\sigma_1\cdots\sigma_i}$, $i = 1, 2, \ldots, k$, and the independence imposed on their vectors we may write:

$$\begin{aligned}
F^N_{\Delta_1\times\cdots\times\Delta_k}(z) &\equiv \mathbb{E}z^{\sum_{\sigma_1}\mathbb{1}_{\{v_{N,1}(Y_{\sigma_1})\in\Delta_1\}}\cdots\sum_{\sigma_k}\mathbb{1}_{\{v_{N,k}(Y_{\sigma_1\ldots\sigma_k})\in\Delta_k\}}} \\
&= \mathbb{E}\left(\mathbb{E}\left(z^{\sum_{\sigma_k}\mathbb{1}_{\{v_{N,k}(Y_{\sigma_1\ldots\sigma_k})\in\Delta_n}} \;\Big|\; \sum_{\sigma_1}\mathbb{1}_{\{v_{N,1}(Y_{\sigma_1})\in\Delta_1\}}\cdots\sum_{\sigma_k}\mathbb{1}_{\{v_{N,k-1}(Y_{\sigma_1\ldots\sigma_{k-1}})\in\Delta_{k-1}\}}\right)\right) \\
&= \mathbb{E}\big(\mathbb{E}z^{\sum_{\sigma_k}\mathbb{1}_{\{v_{N,k}(Y_{\sigma_1\ldots\sigma_k})\in\Delta_k\}}}\big)^{\sum_{\sigma_1}\mathbb{1}_{\{v_{N,1}(Y_{\sigma_1})\in\Delta_1\}}\cdots\sum_{\sigma_{k-1}}\mathbb{1}_{\{v_{N,k-1}(Y_{\sigma_1\ldots\sigma_{k-1}})\in\Delta_{k-1}\}}}.
\end{aligned}$$

Clearly, $\mathbb{E}z^{\sum_{\sigma_k}\mathbb{1}_{\{v_{N,k}(Y_{\sigma_1\ldots\sigma_k})\in\Delta_k\}}} \to f_k(z)$ as $N \uparrow \infty$. It will be convenient to introduce

$$\zeta_N(z) = \mathbb{E}z^{\sum_{\sigma_k}\mathbb{1}_{\{v_{N,k}(Y_{\sigma_1\ldots\sigma_k})\in\Delta_k\}}}/f_k(z) - 1$$

(which tends to zero as $N \uparrow \infty$) and

$$\eta_{N,k-1} \equiv \sum_{\sigma_1}\mathbb{1}_{\{v_{N,1}(Y_{\sigma_1})\in\Delta_1\}}\cdots\sum_{\sigma_{k-1}}\mathbb{1}_{\{v_{N,k-1}(Y_{\sigma_1\ldots\sigma_{k-1}})\in\Delta_{k-1}\}}.$$

With this notations we have

$$F^N_{\Delta_1\times\cdots\times\Delta_k}(z) = \mathbb{E}\big(f_k(z)\big)^{\eta_{N,k-1}} + \mathbb{E}\big[f_k(z)^{\eta_{N,k-1}}\big(\big(1+\zeta_N(z)\big)^{\eta_{N,k-1}} - 1\big)\big] \equiv I_1^N + I_2^N. \tag{2.25}$$

The term I_2^N in this representation converges to zero as $N \uparrow \infty$. In fact, due to the convergence (1.9) and $a_1 > -\infty$, for any $\varepsilon > 0$ one can find a sufficiently large $C(\varepsilon) > 0$ and $N(\varepsilon)$ such that

$$P\big(\eta_{N,k-1} \geqslant C(\varepsilon)\big) \leqslant P\left(\sum_{\sigma_1}\mathbb{1}_{v_{N,1}(Y_{\sigma_1})\in\Delta_1} \geqslant C(\varepsilon)\right) < \varepsilon$$

for all $N \geqslant N(\varepsilon)$. Then

$$|I_2^N| \leqslant 2\varepsilon + \mathbb{E}\big(\big|f_k(z)\big|^{\eta_{N,k-1}}\big|\big(1+\zeta_N(z)\big)^{\eta_{N,k-1}} - 1\big|\mathbb{1}_{\eta_{N,k-1}<C(\varepsilon)}\big)$$

for all $N \geqslant N(\varepsilon)$. The second term in (2.25) tends to zero, as $N \uparrow \infty$, since $\zeta_N(Z) \to 0$. It remains to analyse I_1^N. But this is the generating function of the model with $k-1$ levels at the point $f_k(z)$ and $|f_k(z)| < 1$. Thus by the induction hypothesis

$$\lim_{N\uparrow\infty} F^N_{\Delta_1\times\cdots\times\Delta_k}(z) = \lim_{N\uparrow\infty} I_1^N = f_1\big(f_2\big(\cdots\big(f_k(z)\big)\cdots\big)\big)$$

and (1.10) is proved.

Let us establish the independence property (i). We take $r = 2$ for shortness, the case of $r > 2$ is analogous. Again in view of the identical distribution of the r.v. $Y_{\sigma_1\cdots\sigma_i}$, and of the independence of their vectors, we get:

$$\begin{aligned}
F_{\Delta_1^1\times\cdots\times\Delta_k^1\cup\Delta_1^2\times\cdots\times\Delta_k^2}(z) &= \lim_{N\uparrow\infty} F^N_{\Delta_1^1\times\cdots\times\Delta_k^1\cup\Delta_1^2\times\cdots\times\Delta_k^2}(z)\\
&= \lim_{N\uparrow\infty} \mathbb{E}\Big[\big(\mathbb{E}z^{\sum_{\sigma_2}\mathbb{1}_{\{v_{N,2}(Y_{\sigma_2})\in\Delta_2^1\}}\cdots\sum_{\sigma_n}\mathbb{1}_{\{v_{N,k}(Y_{\sigma_1\ldots\sigma_k})\in\Delta_k^1\}}}\big)^{\sum_{\sigma_1}\mathbb{1}_{\{v_{N,1}(Y_{\sigma_1})\in\Delta_1^1\}}}\\
&\quad\times\big(\mathbb{E}z^{\sum_{\sigma_2}\mathbb{1}_{\{v_{N,2}(Y_{\sigma_2})\in\Delta_2^2\}}\cdots\sum_{\sigma_k}\mathbb{1}_{\{v_{N,k}(Y_{\sigma_1\ldots\sigma_k})\in\Delta_k^2\}}}\big)^{\sum_{\sigma_1}\mathbb{1}_{\{v_{N,1}(Y_{\sigma_1})\in\Delta_1^2\}}}\Big]\\
&= \lim_{N\uparrow\infty} \mathbb{E}\Big[\big(f_{2,\Delta_2^1}\big(f_{3,\Delta_3^1}\big(\cdots\big(f_{k,\Delta_k^1}(z)\big)\cdots\big)\big)\big)^{\sum_{\sigma_1}\mathbb{1}_{\{v_{N,1}(Y_{\sigma_1})\in\Delta_1^1\}}}\\
&\quad\times\big(f_{2,\Delta_2^2}\big(f_{3,\Delta_3^2}\big(\cdots\big(f_{k,\Delta_k^2}(z)\big)\cdots\big)\big)\big)^{\sum_{\sigma_1}\mathbb{1}_{\{v_{N,1}(Y_{\sigma_1})\in\Delta_1^2\}}}\Big],
\end{aligned} \tag{2.26}$$

where in the last line we applied (1.10) with $n-1$ levels together with the arguments similar to (2.25). Recall that the counting variables $\sum_{x_1}\mathbb{1}_{x_1\in\Delta_1}$ of the Poisson point process $\mathcal{P}_1$ corresponding to non-intersecting segments Δ_1^1 and Δ_1^2 are independent. Therefore (2.26) equals the product of the generating functions $f_{1,\Delta_1^1}(\ldots(f_{n,\Delta_n^1}(z))\ldots)$ and $f_{1,\Delta_1^2}(\ldots(f_{k,\Delta_k^2}(z))\ldots)$ finishing the proof of the property (i). The proof of (ii) is analogous. □

Proof of Proposition 1.4. Suppose that for some l and some k, $J_{l-1}+2 \leqslant k \leqslant J_l$, the inequality (1.12) is not true. Then $A_{J_{l-1}+1,k-1} > A_{k,J_l}$. In view of the definition of J_l either for some $j_0 > J_l$ $A_{J_{l-1}+1,k-1} \leqslant A_{k,j_0}$ or there exists p, $k \leqslant p < J_l$, such that $A_{J_{l-1}+1,k-1} \leqslant A_{k,p}$ and for all $j > p$ $A_{J_{l-1}+1,k-1} > A_{k,j}$. Under the former condition $A_{J_{l-1}+1,k-1} < A_{J_l+1,j_0}$ and $A_{k,J_l} < A_{J_l+1,j_0}$, whence $A_{J_{l-1}+1,J_l} < A_{J_l+1,j_0}$ contradicting the definition of J_l. Under the latter condition $A_{k,p} > A_{p+1,j}$, $A_{J_{l-1}+1,k-1} > A_{p+1,j}$, and thus $A_{J_{l-1}+1,p} > A_{p+1,j}$ for all $j > p$. This implies $J_l \leqslant p$ contradicting $p < J_l$. □

Proof of Theorem 1.5 and Proposition 1.8. The proof of Theorem 1.5 requires the two statements of Proposition 1.8 which we therefore prove at the same time.

The proof proceeds by induction. For $m = 1$ (i) and (ii) are immediate from Theorem 1.3. To see that the two assertions of Proposition 1.8 hold, note that $\mathcal{P}^{(1)}$ has a finite number of points in $[x,\infty)$ a.s. for any $x \in \mathbb{R}$. Then the integral $\int_x^\infty e^{c_1x_1}\mathcal{P}^{(1)}(dx)$ is finite a.s. while $\mathbb{E}\int_{-\infty}^x e^{c_1x_1}\mathcal{P}^{(1)}(dx) = e^{(c_1-1)x}/(c_1-1) \to 0$ as $x \to -\infty$.

Assume that (i), (ii) of both Theorem 1.5 and Proposition 1.8 are true up to $m-1$. First we show the existence of $\mathcal{E}$ for m. We may consider $\mathcal{E}$ as the cluster process (see [15], Chapter 8) with the process N_c of the centres of clusters $\gamma_1x_1 + \cdots + \gamma_{m-1}x_{m-1}$ at the projection points of $\mathcal{P}^{(m)}$ on the plane of the first $m-1$ coordinates and the independent Poisson point processes of the intensity $e^{-x}dx$ corresponding to each cluster. Then by Proposition 8.2.II in [15] it suffices to show that for any bounded $A \in \mathbb{R}$ $\int_R \mathbb{P}(A \mid \gamma_1x_1 + \cdots + \gamma_{m-1}x_{m-1})N_c(dx)$ is finite a.s. where $\mathbb{P}(A \mid \gamma_1x_1 + \cdots + \gamma_{m-1}x_{m-1})$ is the probability that A contains no point of the cluster with the centre $\gamma_1x_1 + \cdots + \gamma_{m-1}x_{m-1}$. It suffices to prove this for $A = [a,\infty)$. A simple calculation gives that

$$\mathbb{P}(A \mid \gamma_1 x_1 + \cdots + \gamma_{m-1} x_{m-1}) = 1 - \exp\{-e^{-(a-\gamma_1 x_1 - \cdots - \gamma_{m-1} x_{m-1})/\gamma_m}\} \leqslant e^{(-a+\gamma_1 x_1 + \cdots + \gamma_{m-1} x_{m-1})/\gamma_m} \tag{2.27}$$

and

$$\int_R e^{(\gamma_1 x_1 + \cdots + \gamma_{m-1} x_{m-1})/\gamma_m} N_c(dx) = \int_{R^{m-1}} e^{(\gamma_1 x_1 + \cdots + \gamma_{m-1} x_{m-1})/\gamma_m} \mathcal{P}^{(m-1)}(dx_1, \ldots, dx_{m-1}) < \infty \quad \text{a.s.} \tag{2.28}$$

by (i) of Proposition 1.8 for $m-1$, since $\gamma_1/\gamma_m > \cdots > \gamma_{m-1}/\gamma_m > 1$. Then $\mathcal{E}$ exists and has an a.s. finite number of points in any bounded subset of $\mathcal{R}$. Furthermore, let $y = a/(\gamma_1 + \cdots + \gamma_m)$. Then

$$\begin{aligned} P_j \equiv \mathbb{P}\big(\exists (x_1, \ldots, x_m) \in \mathcal{P}^{(m)}\colon\ & \forall i = 1, 2, \ldots, j-1 \gamma_1 x_1 + \cdots + \gamma_i x_i \leqslant (\gamma_1 + \cdots + \gamma_i) y, \\ & \gamma_1 x_1 + \cdots + \gamma_j x_j > (\gamma_1 + \cdots + \gamma_j) y\big) \\ \leqslant \int_{\substack{(x_1,\ldots,x_k)\in\mathbb{R}^k:\\ \forall i=1,2,\ldots,j-1\ \gamma_1 x_1+\cdots+\gamma_i x_i \leqslant (\gamma_1+\cdots+\gamma_i)y,\\ \gamma_1 x_1+\cdots+\gamma_j x_j > (\gamma_1+\cdots+\gamma_j)y}} & e^{-x_1-\cdots-x_j}\, dx_1 \ldots dx_j = e^{-jy} \end{aligned} \tag{2.29}$$

leading to the exponential bound

$$\mathbb{P}\big(\exists x \in \mathcal{E}\colon\ x > a\big) \leqslant \sum_{j=1}^{m} P_j \leqslant \sum_{j=1}^{m} e^{-ja/(\gamma_1+\cdots+\gamma_j)}.$$

It follows by Borel–Cantelli lemma that $\mathcal{E}$ has an a.s. finite number of points in any interval $[a, \infty)$ and (ii) of Theorem 1.5 is proven for m.

Now we turn to (ii) of Proposition 1.8 We must show that for any $a \in \mathbf{R}$

$$\mathbb{P}\big(\exists \text{ infinitely many } (x_1, \ldots, x_m) \in \mathcal{P}^{(m)}\colon\ \gamma_1 x_1 + \cdots + x_m \gamma_m > a\big) = 0. \tag{2.30}$$

For any $\varepsilon > 0$ one can find a large $A(\varepsilon, a)$ such that the probability (2.30) is not bigger than

$$\begin{aligned} \mathbb{P}\big(\exists \text{ inf. many } & (x_1, \ldots, x_m) \in \mathcal{P}^{(m)} \gamma_1 x_1 + \cdots + x_m \gamma_m > a \\ & \forall i = 1, \ldots, m-1,\ |\gamma_1 x_1 + \cdots + \gamma_i x_i| < (\gamma_1 + \cdots + \gamma_i) A(\varepsilon, a)\big) + \varepsilon. \end{aligned} \tag{2.31}$$

This follows from the exponential bound analogous to (2.29) and the property (ii) applied to the processes $\mathcal{E}$ constructed from $\mathcal{P}^{(1)}, \ldots, \mathcal{P}^{(m-1)}$. The probability (2.31) does not exceed

$$\begin{aligned} \sum_{j=1}^{m-1} \mathbb{P}\big(\exists \text{ inf. many } & (x_{m-j+1}, \ldots, x_m) \in \mathcal{P}^{(j)}\colon \\ & \gamma_{m-j+1} x_{m-j+1} + \cdots + x_m \gamma_m > a - (\gamma_1 + \cdots + \gamma_j) A(a, \varepsilon)\big) + \varepsilon, \end{aligned}$$

where the terms in the sum equal zero by induction assumption. Since $\varepsilon > 0$ is arbitrary, (2.30) follows.

To show (i) of Proposition 1.8 we consider

$$\int_{\substack{(x_1,\ldots,x_m)\in\mathbb{R}^m,\\ \exists i, 1\leqslant i\leqslant m\colon c_1 x_1+\cdots+c_i x_i > (c_1+\cdots+c_i)y}} e^{c_1 x_1 + \cdots + c_m x_m} \mathcal{P}^{(m)}(dx_1 \ldots dx_m)$$

$$= \sum_{k=1}^{m} \int\limits_{\substack{(x_1,\dots,x_m)\in\mathbb{R}^m,\\ \forall i=1,\dots,k-1:\ c_1x_1+\cdots+c_ix_i\leqslant(c_1+\cdots+c_i)y\\ c_1x_1+\cdots+c_kx_k>(c_1+\cdots+c_k)y}} e^{c_1x_1+\cdots+c_mx_m}\mathcal{P}^{(m)}(dx_1\dots dx_m), \tag{2.32}$$

where each term is determined by induction. In fact, for the kth term, given any realisation of $\mathcal{P}^{(m)}$ in $\mathbb{R}^m$, take its projection on the plane of the first k coordinates. By (2.30) for $k<m$ it contains a finite number of points $(x_1,\dots,x_k)$ in the domain of $\mathbb{R}^k$ where $c_1x_1+\cdots+c_kx_k>(c_1+\cdots+c_k)y$. 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 $e^{c_{k+1}x_{k+1}+\cdots+c_mx_m}$ over these coordinates is defined by induction and is finite a.s. Thus the kth term of (2.32) is the sum of an a.s. finite number of terms and each of them is a.s. finite. Finally we put

$$\int\limits_{\mathbb{R}^m} e^{c_1x_1+\cdots+c_mx_m}\mathcal{P}^{(m)}(dx_1\dots dx_m) \equiv \lim_{x\to-\infty} \int\limits_{\substack{(x_1,\dots,x_m)\in\mathbb{R}^m,\\ \exists i,1\leqslant i\leqslant m:\ c_1x_1+\cdots+c_ix_i>(c_1+\cdots+c_i)y}} e^{c_1x_1+\cdots+c_mc_m}\mathcal{P}^{(m)}(dx_1\dots dx_m).$$

This limit is finite a.s. as it verifies Cauchy's criterion. Denote by

$$I(y) = \int\limits_{\substack{(x_1,\dots,x_m)\in\mathbb{R}^m,\\ \exists i,1\leqslant i\leqslant m:\ c_1x_1+\cdots+c_ix_i>(c_1+\cdots+c_i)y}} e^{c_1x_1+\cdots+c_mx_m}\mathcal{P}^{(m)}(dx_1\dots dx_m).$$

Then for any $\varepsilon>0$ and any $y,y'\in\mathbb{R}$, $y'<y$ under assumptions $c_1>\cdots>c_m>1$

$$\begin{aligned} P\big(\big|I(y)-I(y')\big|>\varepsilon\big) &\leqslant \varepsilon^{-2}\mathbb{E}\big(I(y')-I(y)\big)\\ &\leqslant \varepsilon^{-2} \int\limits_{\substack{(x_1,\dots,x_m)\in\mathbb{R}^m:\\ \forall i=1,2,\dots,m\ c_1x_1+\cdots+c_ix_i\leqslant(c_1+\cdots+c_i)x}} e^{(c_1-1)x_1+\cdots+(c_m-1)x_m}\,dx_1\dots dx_m\\ &= \varepsilon^{-2}e^{(c_1-1)y+\cdots+(c_m-1)y}\to 0, \quad y\to-\infty \end{aligned}$$

and Borel–Cantelli lemma yields the result.

To prove (i) of Theorem 1.5, note that

$$U_{J,N}^{-1}(X_\sigma) = \gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1,\dots,\sigma_{J_1}}) + \gamma_2 u^{-1}_{\ln\bar\alpha_2,N}\big(\bar X^{\sigma_1,\dots,\sigma_{J_1}}_{\sigma_{J_1+1}+\cdots+\sigma_{J_2}}\big) + \cdots + \gamma_1 u^{-1}_{\ln\bar\alpha_1,N}\big(\bar X^{\sigma_1,\dots,\sigma_{J_{m-1}}}_{\sigma_{J_{m-1}+1}+\cdots+\sigma_{J_m}}\big).$$

The convergence (i) is equivalent to the convergence of the following probability generating functionals (p.g.fl.s): for any continuous function $h(x)$, $0<h(x)\leqslant 1$, with $1-h$ vanishing outside a bounded set $\mathbb{E}\prod_{x\in\mathbb{E}_N}h(x)\to\mathbb{E}\prod_{x\in E}h(x)$, see Proposition 9.1.VII in [15]. Let us write

$$\begin{aligned} \mathbb{E}\prod_{x\in\mathbb{E}_N}h(x) - \mathbb{E}\prod_{x\in E}h(x) &= \mathbb{E}\prod_{\substack{x_1^N,\dots,x_{m-1}^N\\ \exists x_m^N:\ \gamma_1x_1^N+\cdots\gamma_mx_m^N\in A}} \mathbb{E}\Bigg[\Bigg(\prod_{x_m^N} h(\gamma_1x_1^N+\cdots+\gamma_{m-1}x_{m-1}^N+\gamma_mx_m^N)\\ &\quad -\prod_{x_m} h(\gamma_1x_1^N+\cdots++\gamma_{m-1}x_{m-1}^N+\gamma_mx_m)\Bigg)\Bigg|(x_1^N,\dots,x_{m-1}^N)\Bigg]\\ &\quad +\mathbb{E}\prod_{x_1^N,\dots,x_{m-1}^N}\prod_{x_m} h(\gamma_1x_1^N+\cdots+\gamma_{m-1}x_{m-1}^N+\gamma_mx_m)\\ &\quad -\mathbb{E}\prod_{x_1,\dots,x_{m-1}}\prod_{x_m} h(\gamma_1x_1+\cdots+\gamma_{m-1}x_{m-1}+\gamma_mx_m), \end{aligned} \tag{2.33}$$

where we denote by $(x_1^N,\dots,x_m^N)$, $(x_1,\dots,x_m)$ the points of the processes $\mathcal{P}_N^{(m)}$ and $\mathcal{P}^{(m)}$ in $\mathbb{R}^m$ respectively and by A the support of $1-h$. For any fixed $x_1^N,\dots,x_{m-1}^N$ the term in square brackets converges to zero by Theorem 1.3 and p.g.fl.s. convergence criterion applied to $\mathcal{P}_N^{(1)}$ and $\mathcal{P}^{(1)}$ in $\mathbb{R}$. The statement (i) for $m-1$ and the fact that $\mathcal{E}$ has an a.s. finite number of points in any bounded set imply that for any $\varepsilon>0$ one can find $K(\varepsilon)$ such that the number of terms $\mathbb{E}[\cdot]$ in the product over $\{x_1^N,\dots,x_{m-1}^N\colon \exists x_m^N, \gamma_1 x_1^N+\dots+\gamma_m x_m^N\in A\}$ is smaller than $K(\varepsilon)$ with probability at least $1-\varepsilon$ for all N large enough. Taking into account also that $0<h(x)<1$, we derive that the first term in (2.33) converges to zero. The second term converges to zero by the p.g.fl.s. convergence criterion applied to $\mathcal{E}_N$ and $\mathcal{E}$ constructed from the first $m-1$ coordinates on the function

$$\tilde h(x)=\prod_{y\in\operatorname{supp}\mathcal{P}^{(1)}} h(x+\gamma_m y).$$

Finally we turn to (iii) of Theorem 1.5. First of all, consider the case of Theorem 1.1, i.e. $m=1$. By Lemma 2.1 we have $\mathbb{P}(\max(X_\sigma/\sqrt{N})<u_{\ln\bar\alpha,N}(x)/\sqrt{N})\to e^{-e^{-x}}$ as $N\to\infty$ where $e^{-e^{-x}}\to 1$ if $x\to+\infty$ and $e^{-e^{-x}}\to 0$ if $x\to-\infty$. It follows that $\max(X_\sigma/\sqrt{N})\to\sqrt{2\ln\bar\alpha}$ in probability. Let $m>1$. Take *any* partition J satisfying (1.12) and $\bar a_1,\dots,\bar a_m,\bar\alpha_1,\dots,\bar\alpha_m$ subjected to it. Then from the tree's structure of X_σ in the GREM (1.4) and the previous result for $m=1$ applied to each cell of the partition J $\liminf_{N\to\infty}\max(X_\sigma/\sqrt{N})\geqslant\sqrt{2\bar a_1\ln\bar\alpha_1}+\dots+\sqrt{2\bar a_m\ln\bar\alpha_m}$ in probability. The maximal value $\max_J\sqrt{2\bar a_1\ln\bar\alpha_1}+\dots+\sqrt{2\bar a_m\ln\bar\alpha_m}$ is attained on the partition (1.15). And it gives the true value of the maximum of X_σ: by (i) and (ii) of Theorem 1.5 for some $h>0$ $\mathbb{P}(\max(X_\sigma/\sqrt{N})>(\sqrt{\bar a_1}u_{\ln\bar\alpha_1,N}(x)+\dots+\sqrt{\bar a_m}u_{\ln\bar\alpha_m,N}(x))/\sqrt{N})\to\mathbb{P}(\exists y\in\mathcal{E}\colon\ y>x)\leqslant e^{-hx}$. Thus $\max(X_\sigma/\sqrt{N})\to\sqrt{2\bar a_1\ln\bar\alpha_1}+\dots+\sqrt{2\bar a_m\ln\bar\alpha_m}$ in probability. To deduce from this the convergence of the expectation, note that the function $\max(X_\sigma)$ is the Lipschitz function of $\alpha_1^N+\dots+(\alpha_1\cdots\alpha_n)^N$ independent Gaussian r.v. $X_{\sigma_1},\dots,X_{\sigma_1\dots\sigma_n}$ with the Lipschitz constant equal to 1 and apply the concentration of measure inequality (A.2) to it with $\varepsilon=\delta\sqrt{N}$. Once the convergence of the expectation proven, the inequality (A.2) applied for the second time shows that $\max(X_\sigma/\sqrt{N})$ converges in probability exponentially fast. Then Borel–Cantelli lemma proves the a.s. convergence. □

3. Proof Theorems 1.7 and 1.9

Proof of Theorem 1.7. It is based on Lemmas 3.1 and 3.2 below. Lemma 3.1 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

$$\begin{aligned}
&e^{\sum_{j=1}^{l(\beta)}(-\beta N\sqrt{2\bar a_j\ln\bar\alpha_j}+\beta\gamma_j[\ln(N\ln\bar\alpha_j)+\ln 4\pi]/2+N\ln\alpha_j-N\sum_{i=J_{l(\beta)}+1}^n\beta^2a_i/2}Z_{\beta,N}\\
&\quad=\sum_{\sigma_1\dots\sigma_{J_1}}e^{\beta\gamma_1u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\dots\sigma_{J_1}})}\dots\sum_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}}e^{\beta\gamma_{l(\beta)}u^{-1}_{\ln\bar\alpha_{l(\beta)},N}(\bar X^{\sigma_1\dots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}})}C(\beta)\\
&\qquad+\sum_{\sigma_1\dots\sigma_{J_1}}e^{\beta\gamma_1u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\dots\sigma_{J_1}})}\dots\sum_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}}e^{\beta\gamma_{l(\beta)}u^{-1}_{\ln\bar\alpha_{l(\beta)},N}(\bar X^{\sigma_1\dots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}})}\\
&\qquad\times\left(e^{-\sum_{i=J_{l(\beta)}+1}^n(N\ln\alpha_i+N\beta^2\alpha_i/2)}\sum_{\sigma_{J_{l(\beta)}+1}}e^{\beta\sqrt{a_{J_{l(\beta)}+1}}X_{\sigma_1\dots\sigma_{J_{l(\beta)}+1}}}\dots\sum_{\sigma_m}e^{\beta\sqrt{a_m}X_{\sigma_1\dots\sigma_m}}-C(\beta)\right)\\
&\quad=I_1^N+I_2^N
\end{aligned}$$

with the same notation $\bar X^{\sigma_1\dots\sigma_{J_{l-1}}}_{\sigma_{J_{l-1}+1}\dots\sigma_{J_l}}$ as in (1.11) and show that I_1^N converges in distribution to the integral (1.22) 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 (1.22) can be written as follows:

$$
\begin{aligned}
I_1^N - I = \sum_{k=1}^{l(\beta)} \Bigg[& \sum_{\sigma_1,\dots,\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln\tilde\alpha_1,N}(\bar X_{\sigma_1\dots\sigma_{J_1}})} \dots \sum_{\sigma_{J_{k-1}+1}\dots\sigma_{J_k}} e^{\beta\gamma_k u^{-1}_{\ln\tilde\alpha_k,N}(\bar X^{\sigma_1\dots\sigma_{J_{k-1}}}_{\sigma_{J_{k-1}+1}\dots\sigma_{J_k}})} \\
& \times \mathbb{1}_{\{\forall i=1,\dots,k-1:\ \gamma_1 u^{-1}_{\ln\tilde\alpha_1,N}(\bar X_{\sigma_1\dots\sigma_{J_1}})+\dots+\gamma_i u^{-1}_{\ln\tilde\alpha_i,N}(\bar X^{\sigma_1\dots\sigma_{J_{i-1}}}_{\sigma_{J_{i-1}+1}\dots\sigma_{J_i}})\leqslant(\gamma_1+\dots+\gamma_i)x\}} \\
& \times \mathbb{1}_{\{\gamma_1 u^{-1}_{\ln\tilde\alpha_1,N}(\bar X_{\sigma_1\dots\sigma_{J_1}})+\dots+\gamma_k u^{-1}_{\ln\tilde\alpha_k,N}(\bar X^{\sigma_1\dots\sigma_{J_{k-1}}}_{\sigma_{J_{k-1}+1}\dots\sigma_{J_k}})>(\gamma_1+\dots+\gamma_k)x\}} \\
& \times C(\beta) \sum_{\sigma_{J_k+1},\dots,\sigma_{J_{k+1}}} e^{\beta\gamma_{k+1} u^{-1}_{\ln\tilde\alpha_1,N}(\bar X^{\sigma_1\dots\sigma_{J_k}}_{\sigma_{J_k+1}\dots\sigma_{J_{k+1}}})} \dots \sum_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}} e^{\beta\gamma_{l(\beta)} u^{-1}_{\ln\tilde\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)}}})} \\
& - C(\beta) \int\limits_{\substack{(x_1,\dots,x_{l(\beta)})\in\mathbb{R}^{l(\beta)},\\ \forall i=1,\dots,k-1:\ \gamma_1 x_1+\dots+\gamma_i x_i\leqslant(\gamma_1+\dots+\gamma_i)x\\ \gamma_1 x_1+\dots+\gamma_k x_k>(\gamma_1+\dots+\gamma_k)x}} e^{\beta\gamma_1 x_1+\dots+\beta\gamma_{l(\beta)}x_{l(\beta)}} \mathcal{P}^{(l(\beta))}(dx_1\dots dx_{l(\beta)}) \Bigg] \\
& + C(\beta) \sum_{\sigma_1,\dots,\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln\tilde\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\tilde\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)}}})} \\
& \times \mathbb{1}_{\{\forall i=1,\dots,k:\ \gamma_1 u^{-1}_{\ln\tilde\alpha_1,N}(\bar X_{\sigma_1\dots\sigma_{J_1}})+\dots+\gamma_i u^{-1}_{\ln\tilde\alpha_i,N}(\bar X^{\sigma_1\dots\sigma_{J_{i-1}}}_{\sigma_{J_{i-1}+1}\dots\sigma_{J_i}})\leqslant(\gamma_1+\dots+\gamma_i)x\}} \\
& - C(\beta) \int\limits_{\substack{(x_1,\dots,x_{l(\beta)})\in\mathbb{R}^{l(\beta)},\\ \forall i=1,\dots,k:\ \gamma_1 x_1+\dots+\gamma_i x_i\leqslant(\gamma_1+\dots+\gamma_i)x}} e^{\beta\gamma_1 x_1+\dots+\beta\gamma_{l(\beta)}x_{l(\beta)}} \mathcal{P}^{(l(\beta))}(dx_1\dots dx_{l(\beta)}). \qquad (3.1)
\end{aligned}
$$

Given $\varepsilon > 0$, 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 (3.1) are smaller than ε, is arbitrarily close to 1. Namely, for the last term this is obvious from Proposition 1.8(ii). For the term before the last it suffices to apply Chebyshev's inequality and to estimate its expectation after. Since $\gamma_1 > \dots > \gamma_{l(\beta)}$ and $\beta\gamma_{l(\beta)} > 1$, then by Lemma 3.2 for all $N \geqslant N(x)$ 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, $k = 1, 2, \dots, l(\beta)$. The inequalities (1.12) being verified for the partition (1.15), we have the point processes convergence stated in Theorem 1.3. If $k = l(\beta)$, then by Proposition 1.8(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 $1 \leqslant k < l(\beta)$, then the un-truncated sums over $\sigma_{J_k+1}, \dots, \sigma_n$ 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 $e^{\beta\gamma_{k+1}x_{k+1}+\dots+\beta\gamma_{l(\beta)}x_{l(\beta)}}$ over the process $\mathcal{P}^{(l(\beta)-k)}$ on $R^{l(\beta)-k}$. Moreover, by Proposition 1.8(i) the integral in the kth term of (3.1) 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 (3.1) and complete the analysis of I_N^1.

Let us write a similar representation for I_N^2:

$$
\begin{aligned}
I_2^N = \sum_{k=1}^{l(\beta)} & \sum_{\sigma_1,\dots,\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln\tilde\alpha_1,N}(\bar X_{\sigma_1\dots\sigma_{J_1}})} \dots \sum_{\sigma_{J_{k-1}+1}\dots\sigma_{J_k}} e^{\beta\gamma_k u^{-1}_{\ln\tilde\alpha_k,N}(\bar X^{\sigma_{J_1}\dots\sigma_{J_{k-1}}}_{\sigma_{J_{k-1}+1}\dots\sigma_{J_k}})} \\
& \times \mathbb{1}_{\{\forall i=1,\dots,k-1:\ \gamma_1 u^{-1}_{\ln\tilde\alpha_1,N}(\bar X_{\sigma_1\dots\sigma_{J_1}})+\dots+\gamma_i u^{-1}_{\ln\tilde\alpha_i,N}(\bar X^{\sigma_{J_1}\dots\sigma_{J_{i-1}}}_{\sigma_{J_{i-1}+1}\dots\sigma_{J_i}})\leqslant(\gamma_1+\dots+\gamma_i)x\}} \\
& \times \mathbb{1}_{\{\gamma_1 u^{-1}_{\ln\tilde\alpha_1,N}(\bar X_{\sigma_1\dots\sigma_{J_1}})+\dots+\gamma_k u^{-1}_{\ln\tilde\alpha_k,N}(\bar X^{\sigma_{J_1}\dots\sigma_{J_{k-1}}}_{\sigma_{J_{k-1}+1}\dots\sigma_{J_k}})>(\gamma_1+\dots+\gamma_k)x\}}
\end{aligned}
$$

$$\times \sum_{\sigma_{J_k+1},\dots,\sigma_{J_{k+1}}} e^{\beta\gamma_{k+1}u^{-1}_{\ln\bar\alpha_1,N}(\bar X^{\sigma_{J_1}\dots\sigma_{J_k}}_{\sigma_{J_k+1}\dots\sigma_{J_{k+1}}})} \dots \sum_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}} e^{\beta\gamma_{l(\beta)}u^{-1}_{\ln\bar\alpha_{l(\beta)},N}(\bar X^{\sigma_{J_1}\dots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}})}$$

$$\times\left(e^{-\sum_{i=J_{l(\beta)}+1}^{n}(N\ln\alpha_i+N\beta^2\alpha_i/2)}\sum_{\sigma_{J_{l(\beta)}+1}} e^{\beta\sqrt{a_{J_{l(\beta)}+1}}X_{\sigma_1\dots\sigma_{J_{l(\beta)}+1}}}\dots\sum_{\sigma_m} e^{\beta\sqrt{a_m}X_{\sigma_1\dots\sigma_m}} - C(\beta)\right)$$

$$+\sum_{\sigma_1,\dots,\sigma_{J_1}} e^{\beta\gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\dots\sigma_{J_1}})}\dots\sum_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}} e^{\beta\gamma_{l(\beta)}u^{-1}_{\ln\bar\alpha_{l(\beta)},N}(\bar X^{\sigma_{J_1}\dots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}})}$$

$$\times \mathbb{1}_{\{\forall i=1,\dots,k:\ \gamma_1 u^{-1}_{\ln\bar\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})+\cdots+\gamma_i u^{-1}_{\ln\bar\alpha_i,N}(\bar X^{\sigma_{J_1}\dots\sigma_{J_{i-1}}}_{\sigma_{J_{i-1}+1}\dots\sigma_{J_i}})\leqslant(\gamma_1+\cdots+\gamma_i)x\}}$$

$$\times\left(e^{-\sum_{i=J_{l(\beta)}+1}^{n}(N\ln\alpha_i+N\beta^2\alpha_i/2)}\sum_{\sigma_{J_{l(\beta)}+1}} e^{\beta\sqrt{a_{J_{l(\beta)}+1}}X_{\sigma_1\dots\sigma_{J_{l(\beta)}+1}}}\dots\sum_{\sigma_m} e^{\beta\sqrt{a_m}X_{\sigma_1\dots\sigma_m}} - C(\beta)\right). \tag{3.2}$$

Again, by Lemma 3.2, for all $N \geqslant N(x)$ the expectation of the absolute value of the last term in (3.2) does not exceed $2CC(\beta)e^{x\sum_{j=1}^{l(\beta)}(\beta\gamma_j-1)} \downarrow 0$ as $x \downarrow -\infty$. Then, given $\varepsilon > 0$, using Chebyshev's inequality one can fix $x < 0$ such that for all $N \geqslant N(x)$ the last term of (3.2) 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 (3.2) 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 S $N \geqslant N(\delta, x)$ with probability at least $1 - \delta$ the kth term of (3.2) 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. This theorem asserts the point process convergence of extremes $\mathcal{E}_N = \int_{\mathbb{R}^k} \delta_{\delta_1 x_1+\cdots+\delta_k x_k}\mathcal{P}_N^k(dx_1,\dots,dx_k) \to \mathcal{E} = \int_{\mathbb{R}^k}\delta_{\delta_1 x_1+\cdots+\delta_k x_k}\mathcal{P}^k(dx_1,\dots,dx_k)$ 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 $N \uparrow \infty$, $k = 1,\dots,l(\beta)$. Hence, I_2^N converges to zero in probability. This concludes the proof of the theorem. □

Lemma 3.1. *Let in the GREM $l(\beta) = 0$. Then*

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

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

$$C(\beta) = P\left(\bigcap_{\substack{i:\ 1\leqslant i\leqslant J_1\\ (a_1+\cdots+a_i)/\bar a_{J_1}=\ln(\alpha_1\cdots\alpha_i)/\ln\bar\alpha_{J_1}}}(\sqrt{a_1}Z_1+\cdots+\sqrt{a_i}Z_i<0)\right) \tag{3.3}$$

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

Proof. By Proposition 1.4, $A_{1,j} \leqslant A_{1,J_1} = \gamma_1^2$ for all $j = 1,\dots,J_1$ 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 $\beta^2 A_{1,j} < 1$ for all $j = 1,2,\dots,n$; if $\beta\gamma_1 = 1$, then $\beta^2 A_{1,j} \leqslant 1$ for all $j = 1,2,\dots,n$ and the equality takes place only for those $j \leqslant J_1$ 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 $Z_{\beta,N}/\mathbb{E}Z_{\beta,N} \xrightarrow{\mathcal{D}} 1$ as $N \uparrow \infty$. Let us fix $\varepsilon > 0$ such that

$$\ln(\alpha_1\cdots\alpha_j) - \left(\beta^2 - (\beta-\varepsilon)^2/2\right)(a_1+\cdots+a_j) > 0 \tag{3.4}$$

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

$$Z^T_{\beta,N}=2^{-N}\sum_{\sigma_1}e^{\beta\sqrt{a_1}X_{\sigma_1}}\dots\sum_{\sigma_n}e^{\beta\sqrt{a_1}X_{\sigma_1\sigma_2\dots\sigma_n}}\mathbb{1}_{\{\forall_{i=1,\dots,n}\sqrt{a_1}X_{\sigma_1}+\dots+\sqrt{a_i}X_{\sigma_1\dots\sigma_i}<(\beta+\varepsilon)(a_1+\dots+a_i)\sqrt{N}\}} \tag{3.5}$$

and set

$$\bar{Z}^T_{\beta,N}\equiv Z_{\beta,N}-Z^T_{\beta,N}. \tag{3.6}$$

We can then write

$$\frac{Z_{\beta,N}}{\mathbb{E}Z_{\beta,N}}=\frac{Z^T_{\beta,N}}{\mathbb{E}Z^T_{\beta,N}}\times\frac{\mathbb{E}Z^T_{\beta,N}}{\mathbb{E}Z_{\beta,N}}+\frac{\bar{Z}^T_{\beta,N}}{\mathbb{E}Z_{\beta,N}}. \tag{3.7}$$

It is easy to compute

$$\begin{aligned}\mathbb{E}Z^T_{\beta,N}&=e^{\beta^2N/2}\mathbb{P}\big(\forall i=1,\dots,n\colon\ \sqrt{a_1}Z_1+\dots+\sqrt{a_i}Z_i<\varepsilon(a_1+\dots+a_i)\sqrt{N}\big)\\&\sim e^{\beta^2N/2}=\mathbb{E}Z_{\beta,N}\end{aligned} \tag{3.8}$$

where $Z_1,\dots,Z_n$ are independent standard Gaussian r.v. Thus the second term in (3.7) 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 $\delta>0$

$$\begin{aligned}&\mathbb{P}\big(|Z^T_{\beta,N}/\mathbb{E}Z^T_{\beta,N}-1|>\delta\big)\\&\quad\leqslant\frac{\mathbb{E}(Z^T_{\beta,N}-\mathbb{E}Z^T_{\beta,N})^2}{(\mathbb{E}Z^T_{\beta,N})^2}\\&\quad\leqslant\frac{\sum_{j=1}^N(\alpha_1\cdots\alpha_j)^{-N}\mathbb{E}e^{2\beta\sqrt{N}(\sqrt{a_1}X_1+\dots+\sqrt{a_j}X_j)}\mathbb{1}_{\{\sqrt{a_1}X_1+\dots+\sqrt{a_j}X_j<(\beta+\varepsilon)(a_1+\dots+a_j)\sqrt{N}\}}}{\delta^2e^{\beta^2N(a_1+\dots+a_j)}(1+\mathrm{o}(1))}\\&\quad\leqslant\big(1+\mathrm{o}(1)\big)\delta^{-2}\sum_{j=1}^N(\alpha_1\cdots\alpha_j)^{-N}e^{\beta^2N(a_1+\dots+a_j)}\mathbb{P}\Big(Z<(-\beta+\varepsilon)\sqrt{N(a_j+\dots+a_n)}\Big)\\&\quad\leqslant\big(1+\mathrm{o}(1)\big)\delta^{-2}\sum_{j=1}^N(\alpha_1\cdots\alpha_j)^{-N}e^{\beta^2N(a_1+\dots+a_j)}e^{-N(\beta-\varepsilon)^2(a_j+\dots+a_n)/2},\end{aligned} \tag{3.9}$$

where $X_1,\dots,X_n$ and Z are independent standard Gaussian r.v. In view of the choice of $\varepsilon>0$, (3.9) converges to zero as $N\uparrow\infty$. Then the first factor in the first term of (3.7) $Z^T_{\beta,N}/\mathbb{E}Z^T_{\beta,N}\to1$ 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 (3.4) for all j such that $\beta^2A_{1,j}<1$. (In particular, the last inequality holds for all $j>J_1$.) For those $j\leqslant J_1$ for which $\beta^2A_{1,j}=1$ we truncate the partition function by $u_{\ln(\alpha_1\cdots\alpha_j),N}(0)\sqrt{a_1+\dots+a_j}$. Then

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

Let us compute again

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

as $u_{\ln(\alpha_1\cdots\alpha_j),N}(0)-\beta\sqrt{a_1+\dots+a_j}=\mathrm{o}(1)$ 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 (3.7) converges to 1 in probability. In fact, it can be estimated as the sum (3.9) 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:

$$\begin{aligned}&(\alpha_1\cdots\alpha_j)^{-N}e^{\beta^2N(a_1+\dots+a_j)}\mathbb{P}\big(Z<u_{\ln(\alpha_1\cdots\alpha_j),N}(0)-2\beta\sqrt{N(a_j+\dots+a_n)}\big)\\ &\quad\leqslant\frac{(\alpha_1\cdots\alpha_j)^{-N}e^{\beta^2N(a_1+\dots+a_j)}e^{-(u_{\ln(\alpha_1\cdots\alpha_j),N}(0)-2\beta\sqrt{N(a_j+\dots+a_n)})^2/2}}{\sqrt{2\pi}(2\beta\sqrt{N(a_j+\dots+a_n)}-u_{\ln(\alpha_1\cdots\alpha_j),N}(0))}=\mathrm{O}(1/N),\quad N\uparrow\infty,\end{aligned}$$

in view of the equality $\beta^2A_{1,j}=1$. The second term in (3.7) 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 (1.19) with $\beta^2A_{1,j}=1$, one can estimate it by

$$\begin{aligned}\frac{\bar Z^T_{\beta,N}}{\mathbb{E}Z_{\beta,N}}\leqslant&\sum_{\substack{j=1,\\ A_{1,j}<A_{1,J_1}}}^{n}2^{-N}e^{-\beta^2N/2}\sum_{\sigma_1}e^{\beta\sqrt{a_1}X_{\sigma_1}}\dots\sum_{\sigma_j}e^{\beta\sqrt{a_1}X_{\sigma_1\sigma_2\dots\sigma_j}}\\ &\times\mathbb{1}_{\{\sqrt{a_1}X_{\sigma_1}+\dots+\sqrt{a_j}X_{\sigma_1\dots\sigma_j}\geqslant(\beta+\varepsilon)(a_1+\dots+a_j)\sqrt{N}\}}\sum_{\sigma_{j+1}}e^{\beta\sqrt{a_{j+1}}X_{\sigma_1\dots\sigma_{j+1}}}\dots\sum_{\sigma_n}e^{\beta\sqrt{a_n}X_{\sigma_1\sigma_2\dots\sigma_n}}\\ &+\sum_{\substack{j=1,\\ A_{1,j}=A_{1,J_1}}}^{n}e^{-[\ln N+\ln\ln(\alpha_1\cdots\alpha_j)+\ln4\pi]/2}\sum_{\sigma_1\dots\sigma_j}e^{u^{-1}_{\ln(\alpha_1\cdots\alpha_j),N}(\bar X_{\sigma_1\dots\sigma_j})}\mathbb{1}_{\{u^{-1}_{\ln(\alpha_1\cdots\alpha_j),N}(\bar X_{\sigma_1\cdots\sigma_j})>0\}}\\ &\times(\alpha_{j+1}\cdots\alpha_n)^{-N}e^{-\beta^2(a_{j+1}+\dots+a_n)N/2}\sum_{\sigma_{j+1}}e^{\beta\sqrt{a_{j+1}}X_{\sigma_1\dots\sigma_{j+1}}}\dots\sum_{\sigma_n}e^{\beta\sqrt{a_n}X_{\sigma_1\sigma_2\dots\sigma_n}}.\end{aligned}\tag{3.10}$$

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 $e^{-\varepsilon^2(a_1+\dots+a_j)N/2}$. In the jth term of the last sum we have for all $i=1,2,\dots,j$ $A_{1,i}\leqslant A_{1,J_1}=A_{1,j}$ by Proposition 1.4. This means that inequalities ensuring the convergence to the Poisson point process (see (1.8) of Theorem 1.1) 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 applies. Therefore, the point process $\sum_{\sigma_1,\dots,\sigma_j}\delta_{u^{-1}_{\ln(\alpha_1\cdots\alpha_j),N}(\bar X_{\sigma_1\dots\sigma_j})}$ 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 $\varepsilon>0$ one can find $K(\varepsilon)$ and $R(\varepsilon)$ such that the number of its points in $[0,\infty)$ is smaller than $K(\varepsilon)$ and each of them is smaller than $R(\varepsilon)$ with probability at least $1-\varepsilon$ for all N large enough. The sum over $\sigma_{j+1},\dots,\sigma_n$ attached to each point of this process is the properly normalised partition function of the GREM with the right partition (1.15) $J_1-j,J_2,\dots,J_m$ 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 $\beta^2A_{J_1+1,J_2}<\beta^2A_{1,J_1}=1$. Otherwise, start induction procedure on the size of the first cell. Then this sum over $\sigma_{j+1},\dots,\sigma_n$ 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 $\sigma_1,\dots,\sigma_j$ is bounded in probability: for any $\varepsilon>0$ there exits $M(\varepsilon)$ that it is bigger than $M(\varepsilon)$ 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 (3.7) and completes the proof of the lemma. □

Lemma 3.2. *Let $0 < a_i < 1$, $a_i > 1$, $\gamma_i = \sqrt{a_i/(2\ln\alpha_i)}$, $i = 1, 2, \dots, m$, 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 $N \geqslant N(x)$*

$$\begin{aligned}
&\mathbb{E}e^{\beta\gamma_1 u^{-1}_{\ln\alpha_1,N}(X_1)+\cdots+\beta\gamma_m u^{-1}_{\ln\alpha_m,N}(X_m)}\mathbb{1}_{\{\forall i=1,\dots,m:\ \gamma_1 u^{-1}_{\ln\alpha_1,N}(X_1)+\cdots+\gamma_i u^{-1}_{\ln\alpha_i,N}(X_i)<(\gamma_1+\cdots+\gamma_i)x\}}\\
&\quad\leqslant C(\alpha_1\cdots\alpha_m)^{-N}e^{x\sum_{j=1}^m(\beta\gamma_j-1)}
\end{aligned}\tag{3.11}$$

where the constant C does not depend on x.

Proof. The l.h.s. of (3.11) equals

$$\begin{aligned}
&e^{\sum_{j=1}^m(-\beta N\sqrt{2a_j\ln\alpha_j}+\beta\gamma_j[\ln(N\ln\alpha_j)+\ln 4\pi]/2)}\\
&\quad\times\mathbb{E}e^{\beta\sqrt{a_1}X_1+\cdots+\beta\sqrt{a_m}X_m}\mathbb{1}_{\forall i=1,\dots,m:\ \sqrt{a_1}X_1+\cdots+\sqrt{a_i}X_i<\sqrt{a_1}u_{\ln\alpha_1,N}(x)+\cdots+\sqrt{a_i}u_{\ln\alpha_i,N}(x)}\\
&\quad= e^{\sum_{j=1}^m(-\beta N\sqrt{2a_j\ln\alpha_j}+\beta\gamma_j[\ln(N\ln\alpha_j)+\ln 4\pi]/2+N\beta a_j^2/2)}\mathbb{P}\big(\forall i=1,\dots,m:\ \sqrt{a_1}X_1+\cdots+\sqrt{a_i}X_i\\
&\qquad<\sqrt{a_1}\big(u_{\ln\alpha_1,N}(x)-\beta\sqrt{a_1N}\big)+\cdots+\sqrt{a_i}\big(u_{\ln\alpha_i,N}(x)-\beta\sqrt{a_iN}\big)\big).
\end{aligned}\tag{3.12}$$

Proposition 3.3 below will be applied with

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

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 (3.12) does not exceed

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

for all N large enough. □

Proposition 3.3. *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 $c_i(N) = c_i\sqrt{N} + \mathrm{o}(1)$ 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*

$$\begin{aligned}
&\mathbb{P}\big(\forall i=1,\dots,m:\ \sqrt{a_1}X_1+\cdots+\sqrt{a_i}X_i<\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_i}c_i(N)\big)\\
&\quad\leqslant(\sqrt{N})^{-m}\tilde{C}e^{-\sum_{j=1}^N c_j^2(N)/2}
\end{aligned}\tag{3.13}$$

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

Proof. This is a rather cumbersome computation involving several induction procedures. It is based on a standard estimate for the density of standard random variables (Proposition A.1) that is given in Appendix A.

In the case $m = 1$ (3.13) follows directly from Proposition A.1. Assume that it is proved to $m - 1$ and let us make an induction step to m. Denote by

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

For any $\delta > 0$ by Proposition A.1 we may write:

$$
\begin{aligned}
&\mathbb{P}\big(\forall i=1,\dots,m\colon \sqrt{a_1}X_1+\dots+\sqrt{a_i}X_i<\sqrt{a_1}c_1(N)+\dots+\sqrt{a_i}c_i(N)\big)\\
&\quad\leqslant \frac{\sqrt{a_m}}{\sqrt{(2\pi)^m N}\delta}\int\limits_{D_{m-1}} e^{(\sqrt{a_1}c_1(N)+\dots+\sqrt{a_m}c_m(N)-\sqrt{a_1}x_1-\dots-\sqrt{a_{m-1}}x_{m-1})^2/(2a_m)}\\
&\qquad\times e^{-(x_{m-1}^2+\dots+x_1^2)/2}dx_{m-1}\dots dx_1\\
&\qquad+\mathbb{P}\big(\forall i=1,\dots,m-1\colon \sqrt{a_1}X_1+\dots+\sqrt{a_i}X_i<\sqrt{a_1}c_1(N)+\dots+\sqrt{a_i}c_i(N)\\
&\qquad\qquad \sqrt{a_1}X_1+\dots+\sqrt{a_{m-1}}X_{m-1}<\sqrt{a_1}c_1(N)+\dots+\sqrt{a_{m-1}}c_{m-1}(N)+\sqrt{a_m}c_m(N)+\delta\sqrt{N}\big)\\
&\quad\equiv I^1_{N,m}+I^2_{N,m}.
\end{aligned}
$$

Let us first show that with $\delta>0$ small enough $I^2_{N,m}$ is of exponentially smaller order than the bound (3.13). For this purpose we find k such that

$$
\begin{aligned}
&(a_{m-1}+\dots+a_{m-i+1})c_{m-i}\geqslant\sqrt{a_{m-i}}\big(\sqrt{a_{m-i+1}}c_{m-i+1}+\dots+\sqrt{a_m}c_m\big)\quad \forall i=2,3,\dots,k;\\
&(a_{m-1}+\dots+a_{m-k})c_{m-k-1}<\sqrt{a_{m-k-1}}\big(\sqrt{a_{m-k}}c_{m-k}+\dots+\sqrt{a_m}c_m\big).
\end{aligned}\tag{3.14}
$$

(If (3.14) fails already for $i=2$ we put $k=1$, if (3.14) holds up to $i=m-1$, we put $k=m-1$.) Then the term $I^2_{N,m}$ is bounded by the probability of the following event for which the estimate (3.13) applies by induction:

$$
\begin{aligned}
&\mathbb{P}\big(\forall i=1,\dots,m-k-1\colon \sqrt{a_1}X_1+\dots+\sqrt{a_i}X_i<\sqrt{a_1}\big(c_1(N)+\delta\sqrt{N}/\sqrt{a_1}\big)+\dots+\sqrt{a_i}c_i(N)\\
&\quad\sqrt{a_1}X_1+\dots+\sqrt{a_{m-k-1}}X_{m-k-1}+\sqrt{a_{m-k}+\dots+a_{m-1}}X_{m-k}\\
&\quad<\sqrt{a_1}\big(c_1(N)+\delta\sqrt{N}/\sqrt{a_1}\big)+\dots+\sqrt{a_{m-1}}c_{m-1}(N)+\sqrt{a_m}c_m(N)\big)\\
&\quad\leqslant\frac{\tilde C}{(\sqrt{N})^{m-k}}e^{-[(c_1(N)+\delta\sqrt{N}/\sqrt{a_1})^2+\dots+c^2_{m-k-1}(N)]/2}\\
&\qquad\times e^{-[\sqrt{a_{m-k}}c_{m-k}(N)+\dots+\sqrt{a_m}c_m(N)]^2/[2(a_{m-k}+\dots+a_{m-1})]}.
\end{aligned}\tag{3.15}
$$

Now it suffices to note that

$$
\big(\sqrt{a_{m-k}}c_{m-k}+\dots+\sqrt{a_m}c_m\big)^2>(a_{m-k}+\dots+a_{m-1})\big(c^2_{m-k}+\dots+c^2_m\big).\tag{3.16}
$$

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

$$
\begin{aligned}
&\big(\sqrt{a_{m-k}}c_{m-k}+\dots+\sqrt{a_m}c_m\big)^2\\
&\quad>a_{m-k}c^2_{m-k}+\big(\sqrt{a_{m-k+1}}c_{m-k+1}+\dots+\sqrt{a_m}c_m\big)^2\\
&\qquad+2\sqrt{a_{m-k}}c_{m-k}\big(\sqrt{a_{m-k+1}}c_{m-k+1}+\dots+\sqrt{a_m}c_m\big)\\
&\quad>a_{m-k}c^2_{m-k}+(a_{m-k+1}+\dots+a_{m-1})(c^2_{m-k+1}+\dots+c^2_m)\\
&\qquad+c^2_{m-k}(a_{m-k+1}+\dots+a_{m-1})+a_{m-k}(c^2_{m-k+1}+\dots+c^2_m)\\
&\quad=(a_{m-k}+\dots+a_{m-1})(c^2_{m-k}+\dots+c^2_m),
\end{aligned}\tag{3.17}
$$

where we used (3.14) with $i=k$ and the assumptions on c_i and a_i of the proposition. Note also that by (3.16) and (3.17) the choice of an appropriate $\delta>0$ in (3.15) depends only on c_i, a_i for $i=k,\dots,m$.

Next, we concentrate on the analysis of $I^1_{N,m}$. We start another induction to prove that it obeys the bound (3.13) under the 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 $m-1$. We want to prove it for m. We write $I^1_{N,m}$ as

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

with the notation

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

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 Proposition A.1, otherwise let us just estimate it by $\sqrt{2\pi}$:

$$\begin{aligned}
I^1_{N,m} &\leqslant \frac{\sqrt{a_m a_{m-1}}e^{-c_m^2(N)/2}}{\sqrt{(2\pi)^m N^2(a_{m-1}+a_m)}\delta\delta'} \int\limits_{D_{m-2}} e^{\frac{[\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_{m-1}}c_{m-1}(N)-\sqrt{a_1}x_1-\cdots-\sqrt{a_{m-2}}x_{m-2}]^2}{2a_{m-1}}} \\
&\quad \times e^{-(x^2_{m-2}+\cdots+x^2_1)/2} dx_{m-2}\dots dx_1 \\
&\quad + \frac{\sqrt{a_m}}{\sqrt{(2\pi)^{m-1}N}\delta} \int\limits_{D_{m-3}\cap G^{\delta'}_{m,2}} e^{\frac{[\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_m}c_m(N)-\sqrt{a_1}x_1-\cdots-\sqrt{a_{m-2}}x_{m-2}]^2}{2(a_{m-1}+a_m)}} \\
&\quad \times e^{-(x^2_{m-2}+\cdots+x^2_1)/2} dx_{m-2}\dots dx_1 \\
&\equiv J^1_{N,m} + J^2_{N,m},
\end{aligned} \tag{3.18}$$

where

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

The term $J^1_{N,m}$ has the form of $I^1_{N,m-1}$ with parameters a_i, c_i, $i=1,\dots,m-1$. By induction it obeys the bound claimed in (3.13). In fact, this term gives the main contribution to $I^1_{N,m}$ and to the probability of (3.13). We will show that, under assumptions of the proposition, with an appropriate choice of $\delta'>0$ depending only on a_i and c_i, $i=2,\dots,m$, the contribution of $J^2_{N,m}$ is of exponentially smaller order than the bound on (3.13). Observe that, with notations $a'_i=a_i$, $c'_i=c_i$ for $i=1,\dots,m-3$, $a'_{m-2}=a_{m-2}$, $c'_{m-2}=c_{m-2}+(\sqrt{a_{m-1}}c_{m-1}-a_{m-1}c_m/\sqrt{a_m})/\sqrt{a_{m-2}}$, $a'_{m-1}=a_{m-1}+a_m$, $c'_{m-1}=\sqrt{a_{m-1}+a_m}c_m/\sqrt{a_m}$, $J^2_{N,m}$ 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 $i=2,\dots,m-1$ been verified, then by induction on $I^1_{N,m}$ this term could be estimated by $e^{-(c'^2_1(N)+\cdots+c'^2_{m-1}(N))/2}$ as $c'^2_1+\cdots+c'^2_{m-1} > c_1^2+\cdots+c_m^2$ and the proof of the proposition would be finished. Unfortunately the conditions $\sqrt{a'_i}c'_{i-1}-\sqrt{a'_i}c'_i<0$ are verified for all $i=1,\dots,m-1$, 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 $k=m-1,\dots,2$ that under assumptions $\sqrt{a_i}c_{i-1}-\sqrt{a_i}c_i<0$, $i=2,\dots,m$, for all N large enough

$$\begin{aligned}
L^k_{N,m} &= \int\limits_{D_{m-k-1}\cap G_{m,k}} e^{\frac{[\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_m}c_m(N)-\sqrt{a_1}x_1-\cdots-\sqrt{a_{m-k}}x_{m-k}]^2}{2(a_{m-k+1}+\cdots+a_{m-1})}} e^{-(x^2_{m-k}+\cdots+x^2_1)/2} dx_{m-k}\dots dx_1 \\
&\leqslant e^{-BN(1+o(1))},
\end{aligned} \tag{3.19}$$

where $G_{m,k} = \{(x_1,\dots,x_m)\colon \sqrt{a_1}x_1 + \cdots + \sqrt{a_{m-k}}x_{m-k} < \sqrt{a_1}c_1(N) + \cdots + \sqrt{a_{m-k}}c_{m-k}(N) + \cdots + \sqrt{a_{m-1}}c_{m-1}(N) - (a_{m-k+1} + \cdots + a_{m-1})c_m(N)/\sqrt{a_m}\}$ and where the constant $B > c_1^2 + \cdots + c_m^2$. Then the estimate (3.19) in the case $k = 2$ proves that, with an appropriate choice of $\delta' > 0$, the contribution of $J^2_{N,m}$ to (3.13) is negligible, which implies the assertion of the proposition.

It remains to prove (3.19). Let us introduce the notations: $a'_i \equiv a_i$, $c'_i \equiv c_i$ for $i = 1,\dots,m-k-1$, $a'_{m-k} \equiv a_{m-k}$,

$$\begin{aligned}
&c'_{m-k} \equiv c_{m-k} + \bigl(\sqrt{a_{m-k+1}}c_{m-k+1} + \cdots + a_{m-1}c_{m-1} - (a_{m-k+1} + \cdots + a_{m-1})c_m/\sqrt{a_m}\bigr)/\sqrt{a_{k-m}},\\
&a'_{m-k+1} \equiv a_{m-k+1} + \cdots + a_m,\\
&c'_{m-k+1} \equiv \sqrt{a_{m-k+1} + \cdots + a_m}\,c_m/\sqrt{a_m}.
\end{aligned}$$

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 $L^k_{N,m}$ takes the form of $I^1_{N,m-k+1}$ and by induction on $I^1_{N,m}$ it does not exceed the exponential bound (3.19) with $B = c'^2_1 + \cdots + c'^2_{m-k+1}$. Then the observation

$$\begin{aligned}
&c'^2_1 + \cdots + c'^2_{m-k+1} - (c_1^2 + \cdots + c_m^2)\\
&\quad\geqslant 2c_{m-k}\sum_{i=1}^{k-1}(\sqrt{a_{m-i}}c_{m-i} - a_{m-i}c_m/\sqrt{a_m})/\sqrt{a_{m-k}} + \sum_{i=1}^{k-1}a_{m-i}c_m^2/a_m - \sum_{i=1}^{k-1}c_{m-i}^2\\
&\quad> c_{m-k}\sum_{i=1}^{k-1}(\sqrt{a_{m-i}}c_{m-i} - a_{m-i}c_m/\sqrt{a_m})/\sqrt{a_{m-k}} + \sum_{i=1}^{k-1}a_{m-i}c_m^2/a_m - \sum_{i=1}^{k-1}\sqrt{a_{m-i}}c_mc_{m-i}/\sqrt{a_m}\\
&\quad= \sum_{i=1}^{k-1}(\sqrt{a_{m-i}}c_{m-i} - a_{m-i}c_m/\sqrt{a_m})(c_{m-k}/\sqrt{a_{m-k}} - c_m/\sqrt{a_m}) > 0
\end{aligned} \tag{3.20}$$

finishes the proof of (3.19) in this case. In particular this proves (3.19) for $k = m-1$.

Suppose now that $k > 1$, (3.19) holds up to $k-1$ and that $\sqrt{a'_{m-k-1}}c'_{m-k} \leqslant \sqrt{a'_{m-k}}c'_{m-k-1}$. Let us find $l \geqslant k$ such that

$$(a'_{m-k} + \cdots + a'_{m-i})c'_{m-i-1} \geqslant \sqrt{a'_{m-i-1}}\bigl(\sqrt{a'_{m-i}}c'_{m-i} + \cdots + \sqrt{a'_{m-k}}c'_{m-k}\bigr) \quad \forall i = 2,3,\dots,l; \tag{3.21}$$

$$(a'_{m-k} + \cdots + a'_{m-l-1})c'_{m-l-2} < \sqrt{a'_{m-l-2}}\bigl(\sqrt{a'_{m-l-1}}c'_{m-l-1} + \cdots + \sqrt{a'_{m-k}}c'_{m-k}\bigr). \tag{3.22}$$

(If (3.21) holds for $i = 1,\dots,m-2$, we put $l = m-2$.) Then developing $L^k_{N,m}$ $l-k+2$ times as in (3.18), we obtain

$$\begin{aligned}
L^k_{N,m} \leqslant{}& \sqrt{2\pi}\int\limits_{D_{m-k-2}\cap G^{\delta''}_{m,k-1}} e^{\frac{[\sqrt{a_1}c_1(N)+\cdots+\sqrt{a_m}c_m(N)-\sqrt{a_1}x_1-\cdots-\sqrt{a_{m-k-1}}x_{m-k-1}]^2}{2(a_{m-k}+\cdots+a_m)}}\\
&\times e^{-(x^2_{m-k-1}+\cdots+x_1^2)/2}\,dx_{m-k-1}\dots dx_1\\
&+ \frac{\sqrt{(2\pi)^{l-k+1}(a_{m-k+1}+\cdots+a_m)}\,e^{-c'^2_{m-k+1}(N)/2}}{\sqrt{N}\delta''}\\
&\times \int\limits_{D'_{m-l-2}} e^{\frac{[\sqrt{a'_1}c_1(N)+\cdots+\sqrt{a'_{m-k}}c'_{m-k}(N)-\sqrt{a'_1}x_1-\cdots-\sqrt{a'_{m-l-2}}x_{m-l-2}]^2}{2(a'_{m-l-1}+\cdots+a'_{m-k})}}\\
&\times e^{-(x^2_{m-l-2}+\cdots+x_1^2)/2}\,dx_{m-l-2}\dots dx_1,
\end{aligned} \tag{3.23}$$

where

$$G^{\delta''}_{m,k-1} = \big\{(x_1,\dots,x_{m-k-1}):\ \sqrt{a_1}x_1+\cdots+\sqrt{a_{m-k-1}}x_{m-k-1} < \sqrt{a_1}c_1(N)+\cdots+\sqrt{a_{m-k-1}}c_{m-k-1}(N)+\cdots+\sqrt{a_{m-1}}c_{m-1}(N) - (a_{m-k}+\cdots+a_{m-1})c_m(N)/\sqrt{a_m}+\delta''\sqrt{N}\big\}$$

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 (3.23) with $\delta''>0$ small enough obeys the bound (3.19) 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 (3.22). Then by induction on $I^1_{N,m}$ this term obeys the bound (3.19) with

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

It remains to show that $B-c_1^2-\cdots-c_m^2>0$. This is done by induction on l starting from (3.20).

$$\begin{aligned}
&(a'_{m-l-1}+\cdots+a'_{m-k})(B-c_1^2-\cdots-c_m^2)\\
&\quad>\sqrt{a_{m-l-1}}c_{m-l-1}\bigg(\sqrt{a_{m-l}}c_{m-l}+\cdots+\sqrt{a_{m-1}}c_{m-1}\\
&\qquad-\frac{a_{m-k+1}+\cdots+a_{m-1}}{\sqrt{a_m}}c_m-\frac{a_{m-l}+\cdots+a_{m-k}}{\sqrt{a_{m-l-1}}}c_{m-l-1}\bigg)\\
&\qquad+\sum_{j=1}^{l}\sqrt{a_{m-l-1}}c_{m-j}\big(\sqrt{a_{m-j}}c_{m-l-1}-\sqrt{a_{m-l-1}}c_{m-j}\big)\\
&\qquad+\sum_{r=1}^{k}\sqrt{a_{m-l-1}}\frac{a_{m-r}}{a_m}c_m\big(\sqrt{a_{m-l-1}}c_m-\sqrt{a_m}c_{m-l-1}\big).
\end{aligned}\tag{3.24}$$

Since the factor at c_{m-l-1} in round brackets in the second line is non-positive by (3.21) for $i=l$, let us estimate $c_{m-l-1}<\sqrt{a_{m-l-1}}c_{m-l}/\sqrt{a_{m-l}}$ and, sum this line with the term for $j=l$ in the sum over j. Since the resulting factor at c_{m-l} is non-positive by (3.21) 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 $j=l,\dots,k$. On the other hand, since the factors at c_m in the last line of (3.24) is 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 (3.24) is bigger than

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

This finishes the proof of (3.19) and completes the proof of the proposition. □

Proof of Theorem 1.9. The proof of Theorem 1.9 is almost immediate from Theorem 1.7. 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 $G : \mathbb{R}^j \to \mathbb{R}$

$$\mathbb{E}G\Big(\int \mathcal{W}^m_{\beta,N}(dw)h_1(w), \dots, \int \mathcal{W}^m_{\beta,N}(dw)h_j(w)\Big) \\ \to \mathbb{E}F\Big(\int \mathcal{W}^m_{\beta}(dw)h_1(w), \dots, \int \mathcal{W}^m_{\beta}(dw)h_j(w)\Big). \tag{3.25}$$

Now clearly

$$\int \mathcal{W}^k_{\beta,N}(dw)h_i(w) = \int \mathcal{P}^{(k)}_N(dy)h_i\Big(\frac{\int \mathcal{P}^{(k)}_N(dx)e^{\beta(\gamma,x)}\delta(x_1-y_1)}{\int \mathcal{P}^{(k)}_N(dx)e^{\beta(\gamma,x)}}, \dots, \\ \frac{\int \mathcal{P}^{(k)}_N(dx)e^{\beta(\gamma,x)}\delta(x_1-y_1)\dots\delta(x_k-y_k)}{\int \mathcal{P}^{(k)}_N(dx)e^{\beta(\gamma,x)}}\Big). \tag{3.26}$$

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. [21,28]). We can of course replace $\mathcal{P}^{(k)}_N$ by $\mathcal{P}^{(k,*)}_N$ in (3.25) 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

$$\Big(\frac{\int \mathcal{P}^{(k,*)}_N(dx)e^{\beta(\gamma,x)}\delta(x_1-y_1)}{\int \mathcal{P}^{(k,*)}_N(dx)e^{\beta(\gamma,x)}}, \dots, \frac{\mathcal{P}^{(k,*)}_N(dx)e^{\beta(\gamma,x)}\delta(x_1-y_1)\dots\delta(x_k-y_k)}{\int \mathcal{P}^{(k,*)}_N(dx)e^{\beta(\gamma,x)}}\Big) \\ \to \Big(\frac{\int \mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}\delta(x_1-y_1)}{\int \mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}}, \dots, \frac{\mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}\delta(x_1-y_1)\dots\delta(x_k-y_k)}{\int \mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}}\Big) \quad \text{a.s.} \tag{3.27}$$

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

$$h_i\Big(\frac{\int \mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}\delta(x_1-y_1)}{\int \mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}}, \dots, \frac{\mathcal{P}^{(k*)}(dx)e^{\beta(\gamma,x)}\delta(x_1-y_1)\dots\delta(x_k-y_k)}{\int \mathcal{P}^{(k*)}(dx)e^{\beta(\gamma,x)}}\Big) \tag{3.28}$$

is a continuous function of x, and thus

$$\int \mathcal{P}^{k,*}_N(dx)h_i\Big(\frac{\int \mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}\delta(x_1-y_1)}{\int \mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}}, \dots, \frac{\mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}\delta(x_1-y_1)\dots\delta(x_k-y_k)}{\int \mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}}\Big) \\ \to \int \mathcal{P}^{k,*}(x)h_i\Big(\frac{\int \mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}\delta(x_1-y_1)}{\int \mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}}, \dots, \\ \frac{\mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}\delta(x_1-y_1)\dots\delta(x_k-y_k)}{\int \mathcal{P}^{(k,*)}(dx)e^{\beta(\gamma,x)}}\Big). \tag{3.29}$$

Thus all arguments of G in (3.26) converge (after replacing $\mathcal{P}^{(k)}_N$ by $\mathcal{P}^{(k,*)}_N$) almost surely. Since G is bounded and continuous, this implies convergence of G in mean, proving the theorem. □

4. 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 (1.27).

We need some notation. For a given configuration $\sigma_1\sigma_2\dots\sigma_n$ we denote by $\bar\sigma_l = \sigma_{J_{l-1}+1}\dots\sigma_{J_l}$ 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 (1.15) and (1.20). Note that the event $\{d_N(\sigma,\sigma') > q_i\}$ is equivalent to demanding that $\{\bar\sigma_1\dots\bar\sigma_i = \bar\sigma'_1\dots\bar\sigma'_i\}$. Therefore we have that

$$\begin{aligned}1 - f_{\beta,N}(q_i) &= \sum_\sigma\sum_{\sigma'}\mu_{\beta,N}(\sigma)\mu_{\beta,N}(\sigma')\mathbb{1}_{\{\bar\sigma_1\dots\bar\sigma_i=\bar\sigma'_1\dots\bar\sigma'_i\}}\\ &= \int \mathcal{R}^i_{\beta,N}(dw_i)w_i^2 = \int \mathcal{W}^i_{\beta,N}(dw_1,\dots,dw_i)w_i^2 = \int \mathcal{W}^m_{\beta,N}(dw_1,\dots,dw_i)w_iw_m \qquad (4.1)\end{aligned}$$

for all $m \geqslant i$.

The convergence of the point processes $\mathcal{W}^i_{\beta,N}$ for $i \leqslant l(\beta)$ stated in Theorem 1.9 then immediately implies the convergence in distribution of

$$f_{\beta,N}(q_i) \xrightarrow{\mathcal{D}} \int \widehat{\mathcal{W}}_\beta(dw_1,\dots,dw_{l(\beta)})w_{l(\beta)}(1-w_i) \qquad (4.2)$$

for $i = 1,\dots,l(\beta)$. To conclude the proof of Theorem 1.10, 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

$$\Delta_l = [-\ln\bar\alpha_l/\ln 2, \ln\bar\alpha_l/\ln 2] \quad \text{for } l = 1,2,\dots,l(\beta)$$

and

$$\Delta_{l(\beta)+1} = \big[-\ln(\alpha_{J_{l(\beta)}+1}\cdots\alpha_n)/\ln 2, \ln(\alpha_{J_{l(\beta)}+1}\cdots\alpha_n)/\ln 2\big].$$

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 $\Delta_1\times\Delta_2\times\dots\times\Delta_{l(\beta)+1}$ 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

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

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 1.10 readily.

Theorem 4.1. *The measure $f^{\otimes l(\beta)+1}_{\beta,N}$ converges in distribution to the following point measure on $\Delta_1\times\Delta_2\times\dots\times\Delta_{l(\beta)+1}$*

$$f^{\otimes l(\beta)+1}_{\beta,N} \xrightarrow{\mathcal{D}} Q_0\delta_{(0,0,\dots,0)} + \sum_{j=1}^{l(\beta)} Q_j\delta_{(\ln\bar\alpha_1/\ln 2,\dots,\ln\bar\alpha_j/\ln 2,0,\dots,0)} \quad N\to\infty. \qquad (4.3)$$

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

$$Q_l(\beta) \equiv \int \widehat{\mathcal{W}}_\beta(dw_1,\dots,dw_{l(\beta)})w_{l(\beta)}(w_j - w_{j+1}\mathbb{1}_{\{j\leqslant l(\beta)-1\}}), \quad j = 1,\dots,l(\beta). \qquad (4.4)$$

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 (4.4) then follow from Theorem 1.9 like (4.2).

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 $0\notin \mathrm{cl}(I)$

$$f_{\beta,N}^{\otimes l(\beta)+1}(I)\xrightarrow{\mathcal{D}} 0 \quad \text{as } N\to\infty. \tag{4.5}$$

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

$$f_{\beta,N}^{\otimes l(\beta)+1}\big(\{1\}\times\cdots\times\{1\}\times I_l\times\Delta_{l+1}\times\cdots\times\Delta_{l(\beta)+1}\big)\xrightarrow{\mathcal{D}} 0. \tag{4.6}$$

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

$$f_{\beta,N}^{\otimes l(\beta)+1}(I_1\times\cdots\times I_{l-2}\times I_{l-1}\times I_l\times\cdots\times I_{l(\beta)+1})\xrightarrow{\mathcal{D}} 0. \tag{4.7}$$

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

$$f_{\beta,N}^{\otimes l(\beta)+1}(I_1\times\cdots\times I_{l(\beta)+1})\xrightarrow{\mathcal{D}} 0. \tag{4.8}$$

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

$$\ln(\alpha_1\cdots\alpha_j)-\big(\beta^2-(\beta-\varepsilon)^2/2\big)(a_1+\cdots+a_j)>0 \tag{4.9}$$

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

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

if $\beta^2 A_{1,j}<1$ (which holds for all $j>J_1$, and, if $\beta A_{1,J_1}<1$, for all $j\leqslant J_1$ as well), and

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

if $\beta^2 A_{1,j}=1$ (which can happen only if $\beta^2 A_{1,J_1}=1$ and $j\leqslant J_1$ such that $A_{1,j}=A_{1,J_1}$). Set

$$\begin{aligned}
f_{\beta,N}^{l(\beta)+1}(I)&=\left[2^{-2N}e^{-\beta^2 N}\sum_{\sigma,\sigma'\colon \sigma\cdot\sigma'\in I} e^{\beta\sqrt{N}(X_\sigma+X_{\sigma'})}\mathbb{1}_{\{\bigcap_{j=1}^n (D_{\sigma_1\dots\sigma_j}\cap D_{\sigma'_1\dots\sigma'_j})\}}\right]\frac{e^{\beta^2 N}}{(Z_{\beta,N})^2}\\
&\quad+\left[2^{-N}e^{-\beta^2 N/2}Z_{\beta,N}^{-1}\sum_{\sigma,\sigma'\colon \sigma\cdot\sigma'\in I} e^{\beta\sqrt{N}(X_\sigma+X_{\sigma'})}\mathbb{1}_{\{\bigcup_{j=1}^n (\bar D_{\sigma_1\dots\sigma_j}\cup \bar D_{\sigma'_1\dots\sigma'_j})\}}\right]\frac{e^{\beta^2 N/2}}{Z_{\beta,N}}\\
&\equiv L_N^1+L_N^2.
\end{aligned}$$

The second factors in L_N^1 and L_N^2 converge in probability to constants $C^{-2}(\beta)$ and $C^{-1}(\beta)$ by Theorem 1.10. The mean of the first factor in square brackets of L_N^1 is bounded by

$$\begin{aligned}
&2^{-2N}\sum_{\sigma,\sigma'\colon d_u(\sigma,\sigma')<N\ln\alpha_1/\ln 2}\mathbb{1}_{\{(\sigma\cdot\sigma')/N\in I\}}+\sum_{\substack{j=1\\ \beta^2 A_{1,j}<1}}^{n} e^{-\ln(\alpha_1\cdots\alpha_j)N+(\beta^2-(\beta-\varepsilon)^2/2)(a_1+\cdots+a_j)N}\\
&\quad+\sum_{\substack{j=1\\ \beta^2 A_{1,j}=\beta^2 A_{1,J_1}=1}}^{J_1}\frac{(\alpha_1\cdots\alpha_j)^{-N}e^{\beta^2 N(a_1+\cdots+a_j)}e^{-(u_{\ln(\alpha_1\cdots\alpha_j),N}(0)-2\beta\sqrt{N(a_j+\cdots+a_n)})^2/2}}{\sqrt{2\pi}(2\beta\sqrt{N(a_j+\cdots+a_n)}-u_{\ln(\alpha_1\cdots\alpha_j),N}(0))}.
\end{aligned} \tag{4.10}$$

The second term in (4.10) goes to zero exponentially fast; the third term is $\mathrm{O}(1/N)$. The first term converges to zero as well. Namely, for any $m\in(-1,1)$, the number of pairs of configurations σ,σ' such that $(\sigma\cdot\sigma')/N=m$ is of order $2^{2N}(2\pi N(1+m)(1-m))^{-1/2}e^{-N\phi(m)}$, with $\phi(m)=1/2((1+m)\ln(1+m)+(1-m)\ln(1-m))$. For any $\varepsilon>0$ one can find $\delta>0$ such that, for all $|m|>\varepsilon$ $\phi(m)>\delta$. Since $0\notin \mathrm{cl}\, I$, the first term of (4.10) being

the sum of these numbers with $|m| > \varepsilon$, is exponentially small. It follows that $L_N^1 \to 0$ in probability. The first factor in L_N^2 is bounded by $2^{-N} e^{-\beta^2 N/2} \sum_\sigma e^{\beta\sqrt{N} X_\sigma} \mathbb{1}_{\{\cup_{j=1}^n \bar{D}_{\sigma_1 \dots \sigma_j}\}}$, which equals $\bar{Z}_{\beta,N}^T / \mathbb{E} Z_{\beta,N}^T$ with the notation of Lemma 3.1. In the proof of that lemma it was shown that this converges to zero in probability (see (3.10) and the discussion following it). Hence, $L_N^2 \to 0$ in probability as well, and assertion (i).

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

$$E_{\bar{\sigma}_1 \dots \bar{\sigma}_i} = \left\{\omega\colon\ \gamma_1 u_{\ln \bar{\alpha}_1, N}^{-1}(\bar{X}_{\sigma_1 \dots \sigma_{J_1}}) + \dots + \gamma_i u_{\ln \bar{\alpha}_i, N}^{-1}\left(\bar{X}_{\sigma_{J_{i-1}+1} \dots \sigma_{J_i}}^{\sigma_1 \dots \sigma_{J_{i-1}}}\right) \leqslant (\gamma_1 + \dots + \gamma_i) x\right\}.$$

Then

$$\begin{aligned}
&f_{\beta N}^{\otimes l(\beta)+1}\big(\{1\} \times \dots \times \{1\} \times I_l \times \Delta_{l+1} \times \dots \times \Delta_{l(\beta)+1}\big) \\
&\quad = U_N^{-2} \sum_{k=1}^{l-1} \sum_{\sigma_1 \dots \sigma_{J_1}} e^{2\beta\gamma_1 u_{\ln \bar{\alpha}_1, N}^{-1}(\bar{X}_{\sigma_1 \dots \sigma_{J_1}})} \dots \sum_{\sigma_{J_{k-1}+1} \dots \sigma_{J_k}} e^{2\beta\gamma_k u_{\ln \bar{\alpha}_k, N}^{-1}(\bar{X}_{\sigma_{J_{k-1}+1} \dots \sigma_{J_k}}^{\sigma_1 \dots \sigma_{J_{k-1}}})} \\
&\qquad \times \mathbb{1}_{\{\cap_{j=1}^{k-1} E_{\bar{\sigma}_1 \dots \bar{\sigma}_j}^N \cap \bar{E}_{\bar{\sigma}_1 \dots \bar{\sigma}_k}^N\}} V_{\bar{\sigma}_1, \dots, \bar{\sigma}_k} \\
&\qquad + U_N^{-2} \sum_{\sigma_1 \dots \sigma_{J_1}} e^{2\beta\gamma_1 u_{\ln \bar{\alpha}_1, N}^{-1}(\bar{X}_{\sigma_1 \dots \sigma_{J_1}})} \dots \sum_{\sigma_{J_{l-2}+1} \dots \sigma_{J_{l-1}}} e^{2\beta\gamma_{l-1} u_{\ln \bar{\alpha}_{l-1}, N}^{-1}(\bar{X}_{\sigma_{J_{l-2}+1} \dots \sigma_{J_{l-1}}}^{\sigma_1 \dots \sigma_{J_{l-2}}})} \\
&\qquad \times \mathbb{1}_{\{\cap_{j=1}^{l-1} E_{\bar{\sigma}_1 \dots \bar{\sigma}_j}^N\}} \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})\}} \\
&\qquad + U_N^{-2} \sum_{\sigma_1, \dots \sigma_{J_1}} e^{2\beta\gamma_1 u_{\ln \bar{\alpha}_1, N}^{-1}(\bar{X}_{\sigma_1 \dots \sigma_{J_1}})} \dots \sum_{\sigma_{J_{l-2}+1} \dots \sigma_{J_{l-1}}} e^{2\beta\gamma_{l-1} u_{\ln \bar{\alpha}_{l-1}, N}^{-1}(\bar{X}_{\sigma_{J_{l-2}+1} \dots \sigma_{J_{l-1}}}^{\sigma_1 \dots \sigma_{J_{l-2}}})} \\
&\qquad \times \mathbb{1}_{\{\cap_{j=1}^{l-1} E_{\bar{\sigma}_1 \dots \bar{\sigma}_j}^N\}} \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})\}} \\
&\quad \equiv U_N^{-2}(M_N^1 + M_N^2 + M_N^3)
\end{aligned} \tag{4.11}$$

with the notations

$$\begin{aligned}
&S_{\bar{\sigma}_1, \dots, \bar{\sigma}_{l-1}}(\bar{\sigma}_l, \dots, \bar{\sigma}_{l(\beta)+1}; \bar{\sigma}'_l, \dots, \bar{\sigma}'_{l(\beta)+1}) \\
&\quad \equiv e^{\beta\gamma_l u_{\ln \bar{\alpha}_l, N}^{-1}(\bar{X}_{\sigma_{J_{l-1}+1} \dots \sigma_{J_l}}^{\sigma_1 \dots \sigma_{J_{l-1}}}) + \beta\gamma_l u_{\ln \bar{\alpha}_l, N}^{-1}(\bar{X}_{\sigma'_{J_{l-1}+1} \dots \sigma'_{J_l}}^{\sigma_1 \dots \sigma_{J_{l-1}}})} \\
&\qquad \times e^{\beta\gamma_{l(\beta)} u_{\ln \bar{\alpha}_{l(\beta)}, N}^{-1}(\bar{X}_{\sigma_{J_{l(\beta)-1}+1} \dots \sigma_{J_{l(\beta)}}}^{\sigma_1 \dots \sigma_{J_{l(\beta)-1}}}) + \beta\gamma_{l(\beta)} u_{\ln \bar{\alpha}_{l(\beta)}, N}^{-1}(\bar{X}_{\sigma'_{J_{l(\beta)-1}+1} \dots \sigma'_{J_{l(\beta)}}}^{\sigma_1 \dots \sigma'_{J_{l(\beta)-1}}})} \\
&\qquad \times e^{\beta\sqrt{N}(\sqrt{a_{J_{l(\beta)}+1}} X_{\sigma_1 \dots \sigma_{J_{l(\beta)}+1}} + \dots + \sqrt{a_n} X_{\sigma_1 \dots \sigma_n} + \sqrt{a_{J_{l(\beta)+1}}} X_{\sigma_1 \dots \sigma'_{J_{l(\beta)}+1}} + \dots + \sqrt{a_n} X_{\sigma_1 \dots \sigma'_n})} \\
&\qquad \times e^{-\beta^2 N(a_{J_{l(\beta)}+1} + \dots + a_n) - 2N \ln(\alpha_{J_{l(\beta)+1}} \cdots \alpha_n)}
\end{aligned}$$

and

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

and

$$
\begin{aligned}
U_N = & \sum_{\sigma_1\dots\sigma_{J_1}} e^{\beta\gamma_1u^{-1}_{\ln\tilde\alpha_1,N}(\bar X_{\sigma_1\cdots\sigma_{J_1}})}\cdots\sum_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}} e^{\beta\gamma_{l(\beta)}u^{-1}_{\ln\tilde\alpha_{l(\beta)},N}(\bar X^{\sigma_1\dots\sigma_{J_{l(\beta)-1}}}_{\sigma_{J_{l(\beta)-1}+1}\dots\sigma_{J_{l(\beta)}}})} \\
& \times\sum_{\sigma_{J_{l(\beta)}+1},\dots,\sigma_n}\exp\Big\{\beta\sqrt{N}\Big(\sqrt{a_{J_{l(\beta)}+1}}X_{\sigma_1\dots\sigma_{J_{l(\beta)}+1}}+\cdots+\sqrt{a_n}X_{\sigma_1\dots\sigma_n}\Big)-\beta^2N(a_{J_{l(\beta)}+1}+\dots+a_n)/2 \\
& -N\ln(\alpha_{J_{l(\beta)+1}}\cdots\alpha_n)\Big\}.
\end{aligned}
$$

The third term in (4.11) $U_N^{-2}M_N^3$ 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 $\mathbb{1}\{\cap_{j=1}^{l(\beta)}E_{\tilde\sigma_1\dots\tilde\sigma_j}\}$. By Theorem 1.7 the denominator U_N converges in law to the a.s. finite integral over the process $\mathcal{P}^{(l(\beta))}$. Then for any $\varepsilon>0$ one can find a small $\delta(\varepsilon)>0$ such that $\mathbb{P}(U_N<\delta(\varepsilon))<\varepsilon$ 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 $\varepsilon,\tilde\varepsilon>0$ one can find $x<0$ sufficiently large by absolute value such that

$$
\mathbb{P}(M_N^3/U_N^2>\tilde\varepsilon)\leqslant\mathbb{P}(\tilde M_N^3/U_N>\tilde\varepsilon)\leqslant\mathbb{P}\big(\tilde M_N^3/\delta(\varepsilon)>\tilde\varepsilon\big)+\mathbb{P}\big(U_N>\delta(\varepsilon)\big)<\frac{e^{x\sum_{j=1}^{l(\beta)}(\beta\gamma_j-1)}}{\tilde\varepsilon\delta(\varepsilon)}+\varepsilon<2\varepsilon
$$

for all N starting from some number N_0.

Given $\tilde\varepsilon,\varepsilon>0$, 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

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

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

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

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

$$\eta_N^{p,k} \leqslant \tilde{\eta}_N^{p,k} = \sum_{\bar\sigma_1,\dots,\bar\sigma_k} \mathbb{1}_{\{\bar E_{\bar\sigma_1,\dots,\bar\sigma_k}\}} \times \sum_{\bar\sigma_l',\dots,\bar\sigma_p':\bar\sigma_l\cdot\bar\sigma_l'\in I_l} \mathbb{1}_{\{\gamma_l u^{-1}_{\ln\alpha_l,N}(\bar X^{\sigma_1,\dots,\sigma_{J_{l-1}}}_{\sigma'_{J_{l-1}+1},\dots,\sigma'_{J_l}})+\dots+\gamma_p u^{-1}_{\ln\alpha_p,N}(\bar X^{\sigma_1,\dots,\sigma'_{J_{p-1}}}_{\sigma'_{J_{p-1}+1},\dots,\sigma'_{J_p}})>(\gamma_l+\dots+\gamma_p)x\}}. \tag{4.12}$$

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 $\varepsilon > 0$ there exists $K(\varepsilon)$ such that $\mathbb{P}(\sum_{\bar\sigma_1,\dots,\bar\sigma_k} \mathbb{1}_{\{\bar E_{\bar\sigma_1,\dots,\bar\sigma_k}\}} > K(\varepsilon)) < \varepsilon$ for all N large enough. Namely, by Theorem 1.3 this probability converges to the probability that the process $\mathcal{P}^{(k)}$ in $\mathbf{R}^k$ has more than K points in the domain $\gamma_1 x_1 + \dots + \gamma_k x_k > (\gamma_1 + \dots + \gamma_k)x$. This last probability tends to zero as $K \to \infty$ since by Proposition 1.8(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 $\bar\alpha_l^N(1 + \mathrm{o}(1)) \sum_{m\in I_l}(2\pi N(1+m)(1-m))^{-1/2} e^{-N\phi(m)}$ turns out to be exponentially smaller than $\bar\alpha_l^N$ by virtue of the condition $0 \not\in \mathrm{cl}\, I_l$. In other words, it is bounded by some $\hat\alpha_l^N$ where $\hat\alpha_l < \bar\alpha_l$. These arguments combined together imply

$$\mathbb{P}(\tilde\eta_N^{p,k} \neq 0) \leqslant K(\varepsilon)\mathbb{P}\big(\max_\sigma \hat X_\sigma > \sqrt{\bar a_l} u_{\ln\bar\alpha_l,N}(x) + \dots + \sqrt{\bar a_p} u_{\ln\bar\alpha_p,N}(x)\big) + \varepsilon \tag{4.13}$$

where $\hat X_\sigma$ 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 parameters $\hat\alpha_l, \bar\alpha_{l+1}, \dots, \bar\alpha_p$ and $\bar a_l, \dots, \bar a_p$. By Theorem 1.5(iii) $\max_\sigma X_\sigma/\sqrt{N} \to \sqrt{\bar a_l \ln\hat\alpha_l} + \dots + \sqrt{\bar a_p \ln\bar\alpha_p}$ a.s. which is smaller than the value in (4.13) as $\hat\alpha_l < \bar\alpha_l$. Then the first term in the r.h.s. of (4.13) converges to zero as $N \to \infty$. This proves $\mathbb{P}(\tilde\eta_N^{p,k} \neq 0) \to 0$ for all $p \geqslant k \geqslant l$. The case $k \geqslant p \geqslant l$ 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 (4.11) is absent and the assertion (ii) is proven. Otherwise we proceed by induction. As in the analysis of the third term, for any given $\varepsilon > 0$ we choose $\delta(\varepsilon)$ to make the denominator U_N^2 smaller than $\delta(\varepsilon)$ 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 $\varepsilon > 0$ one can find such a large $K(\varepsilon)$ that $\mathbb{P}(\sum_{\bar\sigma_1,\dots,\bar\sigma_k} \mathbb{1}_{\{\bar E^N_{\bar\sigma_1\dots\bar\sigma_k}\}} > K) < \varepsilon$. Once again, by Theorem 1.3, this probability converges to the probability that the process $\mathcal{P}^{(k)}$ in $\mathbb{R}^k$ has more than K points in the domain $\gamma_1 x_1 + \dots + \gamma_k x_k > (\gamma_1 + \dots + \gamma_k)x$ and this last probability tends to zero when $K \to \infty$ by Proposition 1.8(i). Thus, with a large enough probability the kth term is a linear combination of not more than K terms of type $V_{\bar\sigma_1,\dots,\bar\sigma_k}$. Each of these terms $V_{\bar\sigma_1,\dots,\bar\sigma_k}$ converges in probability to zero, since the product of the first two factors $[\cdot][\cdot]^{-2}$ of $V_{\bar\sigma_1,\dots,\bar\sigma_k}$ 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 1.7. Thus for any $k = 1, \dots, l-1$, 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 $E_{\bar\sigma_1,\dots,\bar\sigma_k}$, the point process convergence of Theorem 1.3 together with the properties of the process $\mathcal{P}^{(m)}$ of Proposition 1.8, Theorem 1.7 and the estimate of the Lemma 2.2. We omit similar details to (ii) and only emphasise the differences. Observe that to obtain (4.13) it was crucial that the sums of the second level of (4.12) were independent from the indicators in front of them. This property was achieved 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 $X_{\sigma_1,\dots,\bar\sigma_{l-1}\bar\sigma_l,\dots}$ and $X_{\sigma_1,\dots,\bar\sigma_{l-1}'\bar\sigma_l'\dots}$ 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 (4.12), we have $0 \in \mathrm{cl}\, I_l$, but for some j, $l < j \leqslant p$, $0 \notin \mathrm{cl}\, I_p$, then anyway

$\mathbb{P}(\tilde{\eta}_N^{p,k} \neq 0) \to 0$: We would get (4.13) where $\hat{X}_\sigma$ are subjected to the GREM of $p-l+1$ levels with the parameters $\bar{\alpha}_l, \dots \hat{\alpha}_j, \dots, \bar{\alpha}_p, \hat{\alpha}_j < \bar{\alpha}_j$, and then with the maximum smaller than the value in (4.13). Finally, if cl $I_l, \dots,$ cl $I_{l(\beta)}$ all contain zero but $I_{l(\beta)+1}$ does not, then truncating by $E_{\bar{\sigma}_1,\dots,\bar{\sigma}_k}$, applying induction and the same arguments as in (ii) we reduce the problem to evaluating the sum

$$\sum_{\bar{\sigma}_{l(\beta)+1},\bar{\sigma}'_{l(\beta)+1}:\ (\bar{\sigma}_{l(\beta)+1}\cdot\bar{\sigma}'_{l(\beta)+1})/N\in I_{l(\beta)+1}} e^{\beta\sqrt{N(a_{l(\beta)+1}+\dots+\alpha_n)}(X_{\bar{\sigma}_{l(\beta)+1}}+X_{\bar{\sigma}'_{l(\beta)+1}})}$$

divided by $e^{-\beta^2 N(a_{l(\beta)+1}+\dots+\alpha_n)-N\ln(a_{l(\beta)+1}\cdots\alpha_n)}$ 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 $0 \notin$ 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 $(\{1\}\times\dots\times\{1\}\times\{0\}\times\dots\times\{0\}\times\{0\})$ and it remains to study its asymptotic behaviour on the set $(\{1\}\times\dots\times\{1\}\times\Delta_{l(\beta)+1})$ and show that in the limit it is concentrated on the unique atom of this set $(\{1\}\times\dots\times\{1\}\times\{0\})$. This is the subject of (iv). Its proof is carried again by induction as in (ii) and using truncations by $\mathbb{1}_{\{E_{\bar{\sigma}_1\dots\bar{\sigma}_k}\}}$. 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. □

5. Proof of the Ghirlanda–Guerra relations

In this section we find the mean of the overlap distribution function stated Proposition 1.11 and Ghirlanda–Guerra relations stated in Proposition 1.12. Throughout this section we assume that the model parameters are such that the inequalities (1.12) are all strict.

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

Proposition 5.1. *For any value of β, and any $i=1,\dots,n$,*

$$\mathbb{E}\frac{d}{d\sqrt{a_i}}F_{\beta,N} = -\beta^2\sqrt{a_i}\,\mathbb{E}f_{\beta,N}\left(\sum_{j=1}^{i}\ln\alpha_j/\ln 2\right). \tag{5.1}$$

Proof. Obviously,

$$\mathbb{E}\frac{d}{d\sqrt{a_i}}F_{\beta,N} = -N^{-1}\beta\mathbb{E}\frac{\mathbb{E}_\sigma\sqrt{N}X_{\sigma_1\dots\sigma_i}e^{\beta\sqrt{N}X_\sigma}}{\mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma}}. \tag{5.2}$$

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

$$\mathbb{E}\big[Xg(X)\big] = \mathbb{E}g'(X). \tag{5.3}$$

Using this identity in the right-hand side of (5.2) with respect to the average over $X_{\sigma_1\dots\sigma_i}$, we get immediately that

$$\begin{aligned}\mathbb{E}\frac{\mathbb{E}_\sigma\sqrt{N}X_{\sigma_1\dots\sigma_i}e^{\beta\sqrt{N}X_\sigma}}{\mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma}} &= N\beta\sqrt{a_i}\,\mathbb{E}\left(1-\frac{\mathbb{E}_\sigma\mathbb{E}_{\sigma'}\mathbb{1}_{\sigma_1\dots\sigma_i=\sigma'_1\dots\sigma'_2}e^{\beta\sqrt{N}(X_\sigma+X_{\sigma'})}}{(\mathbb{E}_\sigma e^{\beta\sqrt{N}X_\sigma})^2}\right)\\ &= N\beta\sqrt{a_i}\,\mathbb{E}\big(1-\mu_{\beta,N}^{\otimes 2}(\mathbb{1}_{\sigma_1\dots\sigma_i=\sigma'_1\dots\sigma'_i})\big)\end{aligned} \tag{5.4}$$

which is obviously (5.1). □

Proof of Proposition 1.11. Due to the convexity of the function $F_{\beta,N}$ and Proposition 5.1

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

where the last limit is found in Theorem 1.2. Taking the derivative of the r.h.s. of (1.21), the result follows. □

Proof of Proposition 1.12. Let us first note that we need to prove (1.37) only for $i \leqslant J_{l(\beta)}$, since for large values of i, both sides are trivially equal to zero.

The starting point is the following obvious generalisation of (5.4):

Lemma 5.2. *Let $h : \mathcal{S}_N^n \to \mathbb{R}$ be any bounded function of n spins. Then*

$$\frac{1}{\sqrt{N}}\mathbb{E}\mu_{\beta,N}^{\otimes n}\big(X_{\sigma_1^k\ldots\sigma_i^k}h(\sigma^1,\ldots,\sigma^n)\big)$$
$$= \beta\sqrt{a_i}\,\mathbb{E}\mu_{\beta,N}^{\otimes n+1}\left(h(\sigma^1,\ldots,\sigma^n)\left(\sum_{l=1}^{n}\mathbb{1}_{\sigma_1^k\ldots\sigma_i^k=\sigma_1^l\ldots\sigma_i^l} - n\mathbb{1}_{\sigma_1^k\ldots\sigma_i^k=\sigma_1^{n+1}\ldots\sigma_i^{n+1}}\right)\right). \tag{5.5}$$

Proof. It is completely analogous to the proof of Proposition 5.1 using the integration by parts (5.3). □

The strength of Lemma 5.2 comes out when combined with a factorisation result that in turn is a consequence of self-averaging.

Lemma 5.3. *Let h be as in the previous lemma. Then under the hypothesis of the theorem,*

$$\lim_{N\uparrow\infty}\frac{1}{\sqrt{N}}\big|\mathbb{E}\mu_{\beta,N}^{\otimes n}\big(X_{\sigma_1^k\ldots\sigma_i^k}h(\sigma^1,\ldots,\sigma^n)\big) - \mathbb{E}\mu_{\beta,N}(X_{\sigma_1^k\ldots\sigma_i^k})\mathbb{E}\mu_{\beta,N}^{\otimes n}\big(h(\sigma^1,\ldots,\sigma^n)\big)\big| = 0. \tag{5.6}$$

Proof. Let us write

$$\big(\mathbb{E}\mu_{\beta,N}^{\otimes n}\big(X_{\sigma_1^k\ldots\sigma_i^k}h(\sigma^1,\ldots,\sigma^n)\big) - \mathbb{E}\mu_{\beta,N}(X_{\sigma_1^k\ldots\sigma_i^k})\mathbb{E}\mu_{\beta,N}^{\otimes n}\big(h(\sigma^1,\ldots,\sigma^n)\big)\big)^2$$
$$= \big(\mathbb{E}\mu_{\beta,N}^{\otimes n}\big[\big(X_{\sigma_1^k\ldots\sigma_i^k} - \mathbb{E}\mu_{\beta,N}^{\otimes n}(X_{\sigma_1^k\ldots\sigma_i^k})\big)h(\sigma^1,\ldots,\sigma^n)\big]\big)^2$$
$$\leqslant \mathbb{E}\mu_{\beta,N}^{\otimes n}\big(X_{\sigma_1^k\ldots\sigma_i^k} - \mathbb{E}\mu_{\beta,N}^{\otimes n}(X_{\sigma_1^k\ldots\sigma_i^k})\big)^2\mathbb{E}\mu_{\beta,N}^{\otimes n}\big(h(\sigma^1,\ldots,\sigma^n)\big)^2, \tag{5.7}$$

where the last inequality is the Cauchy–Schwartz 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

$$\mathbb{E}\mu_{\beta,N}\big(X_{\sigma_1^k\ldots\sigma_i^k} - \mu_{\beta,N}(X_{\sigma_1^k\ldots\sigma_i^k})\big)^2 + \mathbb{E}\big[\mu_{\beta,N}(X_{\sigma_1^k\ldots\sigma_i^k}) - \mathbb{E}\mu_{\beta,N}(X_{\sigma_1^k\ldots\sigma_i^k})\big]^2$$
$$= -\beta^{-2}\mathbb{E}\frac{d^2}{d\sqrt{a_i}^2}F_{\beta,N} + N\beta^{-2}\mathbb{E}\left(\frac{d}{d\sqrt{a_i}}F_{\beta,N} - \mathbb{E}\frac{d}{d\sqrt{a_i}}F_{\beta,N}\right)^2. \tag{5.8}$$

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 $a_1,\ldots,a_{J(l(\beta))}$, except at the set of values where some of the conditions (1.12) hold with equality; moreover, $-F_{\beta,N}$ is convex in each of the variables $\sqrt{a_i}$. Then standard results of convex analysis (see [29], Theorem 25.7) imply that

$$\limsup_{N\uparrow\infty}\left(-\mathbb{E}\frac{d^2}{d\sqrt{a_i}^2}F_{\beta,N}\right) = -\frac{d^2}{d\sqrt{a_i}^2}\lim_{N\uparrow\infty}\mathbb{E}F_{\beta,N} \tag{5.9}$$

which is finite. Thus, the first term in (5.8) 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 [3] or Lemma 4.3 in [32]). In Theorem 1.7 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 [25]), and hence the result of the lemma is proven. □

Remark. The last argument is the sole reason why we need to exclude the situations when equality holds for some of Eqs. (1.12). 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 5.3 into Lemma 5.2. Using in the resulting expression the fact that by 5.1 and (5.2),

$$\mathbb{E}\mu_{\beta,N}(X_{\sigma_1\ldots\sigma_i})/\sqrt{N} = \beta\sqrt{a_i}\big(1 - \mathbb{E}\mu_{\beta,N}^{\otimes 2}(\mathbb{1}_{\sigma_1\ldots\sigma_i=\sigma_1'\ldots\sigma_i'\}})\big) \tag{5.10}$$

we obtain the assertion of Proposition 1.12. □

Acknowledgements

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

Appendix A

Proposition A.1. *For any* $x > 0$

$$\left(\frac{1}{x} - \frac{1}{x^3}\right)e^{-x^2/2} \leqslant \int_x^\infty e^{-t^2/2}\,dt \leqslant \frac{1}{x}e^{-x^2/2}. \tag{A.1}$$

See [18], Lemma 2, p. 175.

Proposition A.2. *Let* $f(x_1,\ldots,x_M)$ *be a function of* M *variables with Lipschitz constant* L *and* $X_1,\ldots,X_M$ *independent standard Gaussian r.v.*

$$\mathbb{P}\big(\big|f(X_1,\ldots,X_M) - \mathbb{E}f(X_1,\ldots,X_M)\big| > \varepsilon\big) \leqslant \exp\big(-\varepsilon^2/(2L)\big). \tag{A.2}$$

See e.g. [25], p. 21.

References

[1] M. Aizenmann, P. Contucci, On the stability of the quenched state in mean field spin-glass models, J. Statist. Phys. 92 (1998) 765–783.
[2] A. Bovier, Statistical Mechanics of Disordered Systems, in: MaPhySto Lecture Notes, vol. 10, Aarhus, 2001.
[3] A. Bovier, V. Gayrard, The Hopfield model as a generalised random mean field model, in: A. Bovier, P. Picco (Eds.), Mathematics of Spin Glasses and Neural Networks, in: Progress in Probablity, Birkhäuser, Boston, 1997.
[4] A. Bovier, I. Kurkova, Derrida's generalised random energy models 2. Models with continuous hierarchies, Ann. IHP (2002), submitted for publication.
[5] A. Bovier, I. Kurkova, Gibbs measures of Derrida's generalised random energy models and geneologies of Neveu's continuous state branching process, Probab. Theor. Relat. Fields (2003), submitted for publication.
[6] A. Bovier, I. Kurkova, M. Löwe, Fluctuations of the free energy in the REM and the p-spin SK-models, Ann. Probab. 30 (2002) 605–651.
[7] M.D. Bramson, Maximal displacement of branching Brownian motion, Comm. Pure Appl. Math. 31 (1978) 531–581.
[8] D. Capocaccia, M. Cassandro, P. Picco, On the existence of thermodynamics for the generalised random energy model, J. Statist. Phys. 46 (1987) 493–505.
[9] B. Derrida, Random energy model: limit of a family of disordered models, Phys. Rev. Lett. 45 (1980) 79–82.
[10] B. Derrida, Random energy model: an exactly solvable model of disordered systems, Phys. Rev. B 24 (1981) 2613–2626.
[11] B. Derrida, A generalisation of the random energy model that includes correlations between the energies, J. Phys. Lett. 46 (1985) 401–407.
[12] B. Derrida, E. Gardner, Solution of the generalised random energy model, J. Phys. C 19 (1986) 2253–2274.

[13] B. Derrida, E. Gardner, Magnetic properties and function $q(x)$ of the generalised random energy model, J. Phys. C 19 (1986) 5783–5798.
[14] E. Gardner, B. Derrida, The probability distribution of the partition function of the random energy model, J. Phys. A 22 (1989) 1975–1981.
[15] D.J. Daley, D. Vere-Jones, An Introduction to the Theory of Point Processes, Springer, New York, 1988.
[16] T.C. Dorlas, J.R. Wedagedera, Large deviations and the random energy model, Int. J. Mod. Phys. B 15 (2001) 1–15.
[17] T.H. Eisele, On a third order phase transition, Comm. Math. Phys. 90 (1983) 125–159.
[18] W. Feller, An Introduction to Probability Theory and its Applications, vol. I, Wiley, New York, 1966.
[19] 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.
[20] A. Galvez, S. Martinez, P. Picco, Fluctuations in Derrida's random energy and generalised random energy models, J. Statist. Phys. 54 (1989) 515–529.
[21] O. Kallenberg, Random Measures, fourth ed., Akademie Verlag, Berlin, 1986.
[22] M.F. Kratz, P. Picco, On a representation of Gibbs measure for R.E.M., preprint Samos 151, 2002.
[23] I. Kurkova, Temperature dependence of the Gibbs State in the random energy model, J. Stat. Phys. 111 (1/2) (2003) 35–56.
[24] M.R. Leadbetter, G. Lindgren, H. Rootzén, Extremes and Related Properties of Random Sequences and Processes, Springer, Berlin, 1983.
[25] M. Ledoux, L. Talagrand, Probability in Banach Space, Springer, Berlin, 1991.
[26] J. Neveu, A continuous state branching process in relation with the GREM model of spin glass theory, Rapport interne no 267, École Polytechnique, 1992.
[27] E. Olivieri, P. Picco, On the existence of thermodynamics for the random energy model, Comm. Math. Phys. 96 (1991) 125–144.
[28] S.I. Resnick, Extreme Values, Regular Variation, and Point Processes, Springer, New York, 1987.
[29] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, NJ, 1970.
[30] D. Ruelle, A mathematical reformulation of Derrida's REM and GREM, Comm. Math. Phys. 108 (1987) 225–239.
[31] D. Sherrington, S. Kirkpatrick, Solvable model of a spin glass, Phys. Rev. Lett. 35 (1972) 1792–1796.
[32] M. Talagrand, Concentration of measure and isoperimetric inequalities in product space, Publ. Math. IHES 81 (1995) 73–205.
[33] M. Talagrand, Rigorous results for the Hopfield model with many patterns, Probab. Theory Related Fields 110 (1998) 177–276.
[34] M. Talagrand, Mean Field Models for Spin Glasses: a First Course, Lecture Notes of the 2000 École des Probabilités de St. Flour, Lecture Notes in Mathematics, Springer, submitted for publication.
[35] M. Talagrand, Self organisation in the low-temperature region of a spin glass model, Rev. Math. Phys. 15 (2003) 1–78.
[36] M. Talagrand, Spin Glasses: a Challenge to Mathematicians, in: Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 46, Springer, Berlin, 2003.