Poisson–Dirichlet statistics for the extremes of a log-correlated Gaussian field

Louis-Pierre Arguin, Olivier Zindy

The Annals of Applied Probability
2014, Vol. 24, No. 4, 1446–1481
DOI: 10.1214/13-AAP952

POISSON–DIRICHLET STATISTICS FOR THE EXTREMES OF A LOG-CORRELATED GAUSSIAN FIELD

By Louis-Pierre Arguin[1] and Olivier Zindy[2]

Université de Montréal and Université Paris 6

We study the statistics of the extremes of a discrete Gaussian field with logarithmic correlations at the level of the Gibbs measure. The model is defined on the periodic interval $[0,1]$, and its correlation structure is nonhierarchical. It is based on a model introduced by Bacry and Muzy [*Comm. Math. Phys.* **236** (2003) 449–475] (see also Barral and Mandelbrot [*Probab. Theory Related Fields* **124** (2002) 409–430]), and is similar to the logarithmic Random Energy Model studied by Carpentier and Le Doussal [*Phys. Rev. E* (3) **63** (2001) 026110] and more recently by Fyodorov and Bouchaud [*J. Phys. A* **41** (2008) 372001]. At low temperature, it is shown that the normalized covariance of two points sampled from the Gibbs measure is either 0 or 1. This is used to prove that the joint distribution of the Gibbs weights converges in a suitable sense to that of a Poisson–Dirichlet variable. In particular, this proves a conjecture of Carpentier and Le Doussal that the statistics of the extremes of the log-correlated field behave as those of i.i.d. Gaussian variables and of branching Brownian motion at the level of the Gibbs measure. The method of proof is robust and is adaptable to other log-correlated Gaussian fields.

1. Introduction. This paper studies the statistics of the extremes of a Gaussian field whose correlations decay logarithmically with the distance. The model is related to the process introduced by Bacry and Muzy [3] (see also Barral and Mandelbrot [4]) and is similar to the *logarithmic random energy model* or *log-REM* studied by Carpentier and Le Doussal [15], and Fyodorov and Bouchaud [24]. Another important log-correlated model is the two-dimensional discrete Gaussian free field.

Received June 2013.
[1]Supported by a NSERC discovery grant and a grant FQRNT *Nouveaux chercheurs*.
[2]Supported in part by the French ANR project MEMEMO2 2010 BLAN 0125.
AMS 2000 subject classifications. Primary 60G15, 60F05; secondary 82B44, 60G70, 82B26.
Key words and phrases. Log-correlated Gaussian fields, Gibbs measure, Poisson–Dirichlet variable, tree approximation, spin glasses.

The statistics of the extremes of log-correlated Gaussian fields are expected to resemble those of i.i.d. Gaussian variables or *random energy model* (REM) and at a finer level, those of branching Brownian motion. In fact, log-correlated fields are conjectured to be the critical case where correlations start to affect the statistics of the extremes. The reader is referred to the works of Carpentier and Le Doussal [15]; Fyodorov and Bouchaud [24]; and Fyodorov, Le Doussal and Rosso [25] for physical motivations of this fact. The analysis for general log-correlated Gaussian field is complicated by the fact that, unlike branching Brownian motion, the correlations do not necessarily exhibit a tree structure.

The approach of this paper is in the spirit of the seminal work of Derrida and Spohn [19] who studied the extremes of branching Brownian motion using the Gibbs measure. The method of proof presented here is robust and applicable to a large class of nonhierarchical log-correlated fields. The model studied here has the advantages of having a graphical representation of the correlations, a continuous scale parameter and no boundary effects (cf. Section 1.1) which make the ideas of the method more transparent. Even though correlations are not tree-like for general log-correlated models, such fields can often be decomposed as a sum of independent fields acting on different scales. The main results of the paper are Theorem 1.4 on the correlations of the extremes and Theorem 1.5 on the statistics of the Gibbs weights. The results show that, in effect, the statistics of the extremes of the log-correlated field are the same as those of branching Brownian motion at the level of the Gibbs measure, as conjectured by Carpentier and Le Doussal [15].

The method of proof is outlined in Section 2. The proof of the first theorem is based on an adaptation of a technique of Bovier and Kurkova [11, 12] originally developed for hierarchical Gaussian fields such as branching Brownian motion. For this purpose, we need to introduce a family of log-correlated Gaussian models where the variance of the fields in the scale-decomposition depends on the scale. The free energy of the perturbed models is computed using ideas of Daviaud [17]. The second theorem on the Poisson–Dirichlet statistics of the Gibbs weights is proved using the first theorem on correlations and general spin glass theory results.

1.1. *A log-correlated Gaussian field.* Following [3], we consider the half-infinite cylinder

$$\mathcal{C}^+ := \{(x,y); x \in [0,1]_\sim, y \in \mathbb{R}_+^*\},$$

where $[0,1]_\sim$ stands for the unit interval where the two endpoints are identified. We write $\|x-x'\| := \min\{|x-x'|, 1-|x-x'|\}$ for the distance on $[0,1]_\sim$.

The following measure is put on $\mathcal{C}^+$:

$$\theta(\mathrm{d}x, \mathrm{d}y) := y^{-2}\,\mathrm{d}x\,\mathrm{d}y.$$

For $\sigma > 0$, *the variance parameter*, there exists a random measure μ on $\mathcal{C}^+$ that satisfies:

(i) for any measurable set A in $\mathcal{B}(\mathcal{C}^+)$, the random variable $\mu(A)$ is a centered Gaussian with variance $\sigma^2\theta(A)$;

(ii) for every sequence of disjoint sets $(A_n)_n$ in $\mathcal{B}(\mathcal{C}^+)$, the Borel σ-algebra associated with $\mathcal{C}^+$, the random variables $(\mu(A_n))_n$ are independent and

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n) \qquad \text{a.s.}$$

Let Ω be the probability space on which μ is defined, and let $\mathbb{P}$ be the law of μ. The space Ω is endowed with the σ-algebras $\mathcal{F}_u$ generated by the random variables $\mu(A)$, for all the sets A at a distance greater than u from the x-axis. The reader is referred to [3] for the existence of the probability space $(\Omega, (\mathcal{F}_u)_u, \mathbb{P})$.

The subsets needed for the definition of the Gaussian field are the cone-like subsets $A_u(x)$ of $\mathcal{C}^+$,

$$A_u(x) := \{(s,y) \in \mathcal{C}^+ : y \geq u, -f(y)/2 \leq s - x \leq f(y)/2\},$$

where $f(y) = y$ for $y \in (0, 1/2)$ and $f(y) = 1/2$ otherwise. See Figure 1 for a depiction of the subsets. Observe that, by construction, if $\|x - x'\| = \ell > u$, then $A_u(x)$ and $A_u(x')$ intersect exactly above the line $y = \ell$.

The Gaussian process $\omega_u = (\omega_u(x), x \in [0,1]_\sim)$ is defined using the random measure μ,

$$\omega_u(x) := \mu(A_u(x)), \qquad x \in [0,1]_\sim. \tag{1.1}$$

By properties (i) and (ii) of μ listed above, the covariance between $\omega_u(x)$ and $\omega_u(x')$ is given by the integral over θ of the intersection of $A_u(x)$ and $A_u(x')$,

$$\mathbb{E}[\omega_u(x)\omega_u(x')] = \int_{A_u(x)\cap A_u(x')} \theta(\mathrm{d}s, \mathrm{d}y). \tag{1.2}$$

The paper focuses on a discrete version of ω_u. Let $N \in \mathbb{N}$, and take $\varepsilon = 1/N$. Define the set

$$\mathcal{X}_N = \mathcal{X}_\varepsilon := \left\{0, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{i}{N}, \ldots, \frac{N-1}{N}\right\}.$$

The notation $\mathcal{X}_N$ and $\mathcal{X}_\varepsilon$ will be used equally depending on the context. For a given N, the log-correlated Gaussian field is the collection of Gaussian centered random variables $\omega_\varepsilon(x)$ for $x \in \mathcal{X}_N$,

$$X = (X_x, x \in \mathcal{X}_N) = (\omega_\varepsilon(x), x \in \mathcal{X}_N). \tag{1.3}$$

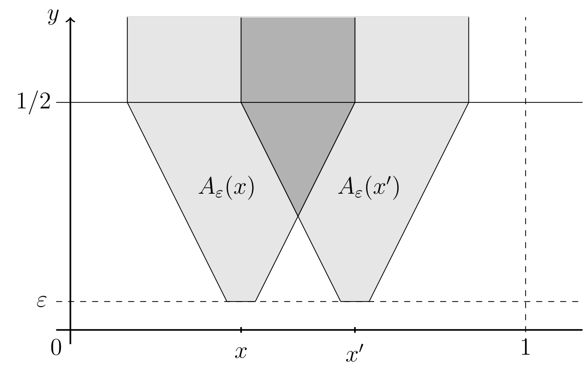

FIG. 1. *The two subsets $A_\varepsilon(x)$ and $A_\varepsilon(x')$ for $\varepsilon = 1/N$. The variance of the variables is given by the integral over $\theta(\mathrm{d}t, \mathrm{d}y) = y^{-2}\,\mathrm{d}t\,\mathrm{d}y$ of the lighter gray area above $\varepsilon = 1/N$, and the covariance by the integral over the intersection of the subsets, the darker gray region.*

A compelling feature of this construction is that a *scale decomposition* for X is easily obtained from property (ii) above. Indeed, it suffices to write the variable X_x as a sum of independent Gaussian fields corresponding to disjoint horizontal strips of $\mathcal{C}^+$. The y-axis then plays the role of the scale.

The covariances of the field are computed from (1.2) by straightforward integration; see also Figure 1.

LEMMA 1.1. *For any $0 < \varepsilon = 1/N < 1/2$,*

$$\mathbb{E}[X_x^2] = \sigma^2(\log N + 1 - \log 2), \qquad x \in \mathcal{X}_N,$$

$$\mathbb{E}[X_x X_{x'}] = \sigma^2(\log(1/\|x - x'\|) - \log 2), \qquad x \neq x' \in \mathcal{X}_N.$$

Similar constructions of log-correlated Gaussian fields using a random measure on cone-like subsets are also possible in two dimensions; see, for example, [30].

1.2. *Main results.* Without loss of generality, the results of this section are stated for the variance parameter $\sigma = 1$. The points where the field is unusually high, *the extremes* or *the high points*, can be studied using a minor adaptation of the arguments of Daviaud for the two-dimensional discrete Gaussian free field [17]. We denote by $|\mathcal{A}|$ the cardinality of a finite set $\mathcal{A}$.

THEOREM 1.2 (Daviaud [17]). *Let*

$$\mathcal{H}_N(\gamma) := \{x \in \mathcal{X}_N : X_x \geq \sqrt{2}\gamma \log N\}$$

be the set of γ-high points. Then for any $0 < \gamma < 1$,

$$\lim_{N\to\infty} \frac{\log|\mathcal{H}_N(\gamma)|}{\log N} = 1 - \gamma^2 \qquad \textit{in probability.}$$

Moreover, for all $\rho > 0$ *there exists a constant* $c = c(\rho) > 0$ *such that*

$$\mathbb{P}(|\mathcal{H}_N(\gamma)| \leq N^{(1-\gamma^2)-\rho}) \leq \exp\{-c(\log N)^2\}$$

for N *large enough.*

The technique of Daviaud is based on a tree approximation introduced by Bolthausen, Deuschel and Giacomin [6] for the discrete two-dimensional Gaussian free field. There, the technique is used to obtain the first order of the maximum. The same argument applies here. Theorem 1.2 and simple Gaussian estimates yield

$$\lim_{N\to\infty} \frac{\max_{x\in\mathcal{X}_N} X_x}{\log N} = \sqrt{2} \qquad \text{a.s.} \tag{1.4}$$

The important feature of Theorem 1.2 and equation (1.4) is that they are identical to the results for N i.i.d. Gaussian variables of variance $\log N$. In other words, the above observables of the high points are not affected by the correlations of the field. The i.i.d. case is called the *random energy model* (REM) in the spin glass literature.

The starting point of the paper is to understand to which extent i.i.d. statistics is a good approximation for more refined observables of the extremes of log-correlated Gaussian fields. To this end, we turn to tools of statistical physics which allow for a good control of the correlations.

First, consider the *partition function* $Z_N(\beta)$ of the model (β stands for the inverse-temperature),

$$Z_N(\beta) := \sum_{x\in\mathcal{X}_N} \exp\{\beta X_x\} \qquad \forall\beta > 0,$$

and the *free energy*

$$f_N(\beta) := \frac{1}{\log N} \log Z_N(\beta) \qquad \forall\beta > 0.$$

Theorem 1.2 is used to compute the free energy of the model.

Corollary 1.3. *Let* $\beta_c := \sqrt{2}$. *Then, for all* $\beta > 0$

$$f(\beta) := \lim_{N\to\infty} f_N(\beta) = \begin{cases} 1 + \dfrac{\beta^2}{2}, & \text{if } \beta < \beta_c, \\ \sqrt{2}\beta, & \text{if } \beta \geq \beta_c, \end{cases} \qquad \text{a.s. and in } L^1.$$

The free energy is the same as for the REM with variance $\log N$. In particular, the model undergoes *freezing* above β_c in the sense that the quantity $f(\beta)/\beta$ is constant.

More importantly, consider the *normalized Gibbs weights* or *Gibbs measure*

$$G_{\beta,N}(x) := \frac{e^{\beta X_x}}{Z_N(\beta)}, \qquad x \in \mathcal{X}_N.$$

By design, the Gibbs measure concentrates on the high points of the Gaussian field. The first main result of the paper is to achieve a control of the correlations at the level of the Gibbs measure. Precisely, with spin glasses in mind, we consider the normalized covariance or *overlap*

$$q(x,y) = q^{(N)}(x,y) := -\frac{\log \|y-x\|}{\log N}, \qquad x,y \in \mathcal{X}_N. \tag{1.5}$$

Clearly, $\|x-y\| = \varepsilon^{q(x,y)}$ and $0 \le q(x,y) \le 1$. Moreover, the overlap $q(x,y)$ is equal to the normalized correlations $\mathbb{E}[X_x X_y]/\mathbb{E}[X_x^2]$ plus a term that goes to zero as N goes to infinity.

A fundamental object, that records the correlations of high points, is the *distribution function of the overlap* sampled from the Gibbs measure. Namely, denote by $G_{\beta,N}^{\times 2}$ the product measure on $\mathcal{X}_N \times \mathcal{X}_N$. Let $(x_1,x_2) \in \mathcal{X}_N^2$ be sampled from $G_{\beta,N}^{\times 2}$. Write for simplicity q_{12} for $q(x_1,x_2)$. The averaged distribution function of the overlap is

$$x_\beta^{(N)}(q) := \mathbb{E}[G_{\beta,N}^{\times 2}\{q_{12} \le q\}], \qquad 0 \le q \le 1. \tag{1.6}$$

The first result is the analogue of results of Derrida and Spohn for the Gibbs measure of branching Brownian motion (see equation (6.19) in [19]), of Chauvin and Rouault on branching random walks [16] and of Bovier and Kurkova on Derrida's *generalized random energy models* (GREM) [11, 18]. It had been conjectured for nonhierarchical log-correlated Gaussian field by Carpentier and Le Doussal; see page 16 in [15].

THEOREM 1.4. *For $\beta > \beta_c$,*

$$\lim_{N\to\infty} x_\beta^{(N)}(q) = \lim_{N\to\infty} \mathbb{E}[G_{\beta,N}^{\times 2}\{q_{12} \le q\}] = \begin{cases} \dfrac{\beta_c}{\beta}, & \text{for } 0 \le q < 1, \\ 1, & \text{for } q = 1. \end{cases}$$

This result is the same as for the REM model [33]. It is therefore consistent with rich statistics of extremes consisting of many high values order one away of each other and whose correlations are either very high or close to 0. This result is in expectation. The typical behavior of the random variable $G_{\beta,N}^{\times 2}\{q_{12} \le q\}$ for q small in terms of β should be exponentially small in β rather than $1/\beta$. To see this, at the heuristic level, it is informative to consider the i.i.d. case where the same phenomenon occurs. Consider N

i.i.d. Gaussian random variables $(X_i)_{1\leq i\leq N}$ of variance $\log N$ ordered in a decreasing way. In this case, $q_{ij}=0$ if $i\neq j$. The following inequality is easily verified:

$$G_{\beta,N}^{\times 2}\{q_{12}=0\}=\frac{\sum_{i\neq j}\mathrm{e}^{\beta X_i}\mathrm{e}^{\beta X_j}}{(\sum_i \mathrm{e}^{\beta X_i})^2}\leq 2\sum_{j\geq 2}\mathrm{e}^{\beta(X_j-X_1)}.$$

In particular, since the gap X_1-X_2 is of order one in the limit and since the density of points at distance x from the maximum is bounded by e^{Cx} for C large enough (see [10] for a precise statement in terms of extremal process), the typical behavior of $G_{\beta,N}^{\times 2}\{q_{12}=0\}$ is expected to be exponentially small in β.

We remark also that for $\beta\leq\beta_c$ the free energy contains all information about the two-overlap distribution. Indeed, since the free energy in Corollary 1.3 is differentiable for every $\beta>0$ including β_c, we have by the convexity of the free energy that the derivative of the limit is the limit of the derivatives. Hence

$$\lim_{N\to\infty} f_N'(\beta)=\lim_{N\to\infty}\beta(1-\mathbb{E}G_{\beta,N}^{\times 2}[q_{12}])=f'(\beta).$$

The first equality is by Gaussian integration by part. It follows that $\lim_N \mathbb{E}[G_{\beta,N}^{\times 2}(q_{12})]=0$ for $\beta\leq\beta_c$. In particular, since the correlations are positive, the overlap of two sampled points is 0 almost surely for every $\beta\leq\beta_c$.

In the case of $\beta>\beta_c$, the first moment of the two-overlap distribution is strictly greater than 0, therefore more information is needed to determine the distribution. One way to proceed would be to obtain enough expectations of functions of q_{12} to determine the distribution. This can be done by adding parameters to the field and consider the appropriate derivative of the free energy of the perturbed model. This is similar in spirit to the p-spin perturbations for the Sherrington–Kirkpatrick model in spin glasses; see, for example, [33]. It turns out that this kind of pertubative approach pioneered by Bovier and Kurkova in [12] for Gaussian fields on trees can be generalized to log-correlated fields. The control of the correlations is achieved by introducing a perturbed version of the model at a specific scale; cf. Section 2.1. In the present case, the proof is more intricate since the structure of correlations of the Gaussian field for finite N is not tree-like or *ultrametric* as in the cases of branching Brownian motion and GREM's. For example, for branching Brownian motion, $q(x,y)$ corresponds to the branching time of the common ancestor of two particles at time t, x and y, divided by t. Because of the branching structure,

(1.7) the inequality $q(x,y)\geq\min\{q(x,z),q(y,z)\}$ is satisfied for all x,y,z.

[The terminology *ultrametric* comes from the fact that the distance induced by the form $q(\cdot,\cdot)$ is ultrametric when (1.7) holds.]

The *Parisi ultrametricity conjecture* in the spin-glass literature states that, even though tree-like correlations might not be present for finite N, ultrametric correlations are recovered in the limit $N\to\infty$ for a large class of Gaussian fields at the level of the Gibbs measure, that is,

$$\lim_{N\to\infty}\mathbb{E}[G_{\beta,N}^{\times 3}\{q_{12}\geq\min\{q_{13},q_{23}\}\}]=1. \tag{1.8}$$

It is not hard to see that Theorem 1.4 implies the ultrametricity conjecture for the Gaussian field considered, since the overlaps can only take value 0 or 1. (In the language of spin glasses, the field is said to admit a *one-step replica symmetry breaking* at low temperature.)

The second main result describes the joint distribution of overlaps sampled from the Gibbs measure. To this end, for $s\geq 2$, we denote the product of Gibbs measure on $\mathcal{X}_N^s$ by $G_{\beta,N}^{\times s}$. We consider the class of continuous functions $F:[0,1]^{s(s-1)/2}\to\mathbb{R}$. We write $\mathbb{E}G_{\beta,N}^{\times s}[F(q_{ll'})]$ for $\mathbb{E}G_{\beta,N}^{\times s}[F(\{q(x_l,x_{l'})\}_{1\leq l<l'\leq s})]$, that is, the averaged expectation of $F(\{q(x_l,x_{l'})\}_{1\leq l<l'\leq s})$ when $(x_1,\ldots,x_s)$ is sampled from $G_{\beta,N}^{\times s}$. We recall the definition of a Poisson–Dirichlet variable. For $0<\alpha<1$, let $\eta=(\eta_i,i\in\mathbb{N})$ be the atoms of a Poisson random measure on $(0,\infty)$ of intensity measure $s^{-\alpha-1}\,\mathrm{d}s$. A *Poisson–Dirichlet variable* ξ of parameter α is a random variable on the space of decreasing weights $\vec{s}=(s_1,s_2,\ldots)$ with $1\geq s_1\geq s_2\geq\cdots\geq 0$ and $\sum_i s_i\leq 1$ which has the same law as

$$\xi\overset{\text{law}}{=}\left(\frac{\eta_i}{\sum_j\eta_j},i\in\mathbb{N}\right)_{\downarrow},$$

where $\downarrow$ stands for the decreasing rearrangement.

Theorem 1.5. *Let $\beta>\beta_c$ and $\xi=(\xi_k,k\in\mathbb{N})$ be a Poisson–Dirichlet variable of parameter β_c/β. Denote by E the expectation with respect to ξ. For any continuous function $F:[0,1]^{s(s-1)/2}\to\mathbb{R}$ of the overlaps of s points,*

$$\lim_{N\to\infty}\mathbb{E}G_{\beta,N}^{\times s}[F(q_{ll'})]=E\left[\sum_{k_1\in\mathbb{N},\ldots,k_s\in\mathbb{N}}\xi_{k_1}\cdots\xi_{k_s}F(\delta_{k_lk_{l'}})\right].$$

It is important to stress that, as in the case of branching Brownian motion and unlike the REM, it is not the collection $(G_{\beta,N}(x),x\in\mathcal{X}_N)_{\downarrow}$ *per se* that converges to a Poisson–Dirichlet variable. Rather, the result suggests that the Poisson–Dirichlet weights are formed by the sum of the Gibbs weights of high points that are arbitrarily close to each other because the continuity of the function F naturally identifies points x,y for which $q(x,y)$ tends to 1 in the limit $N\to\infty$. In the theory of spin glasses, these clusters of high points are often called *pure states*. For more on the connection with spin glasses, the reader is referred to [34] where the pure states are constructed explicitly for mean-field models.

1.3. *Relation to previous results.* Bolthausen and Kistler have studied a family of models called *generalized GREMs* for which the correlations are not ultrametric [8, 9] for finite N. By construction, the overlaps of these models can only take a finite number of values (uniformly in N, the number of variables). They compute the free energies and the Gibbs measure and prove the Parisi ultrametricity conjecture for these. Bovier and Kurkova [11, 12] have obtained the distribution of the Gibbs measure for Gaussian fields, called the CREMs, where the values of the overlaps are not a priori restricted. Their analysis is restricted to models with ultrametric correlations and include the case of branching Brownian motion.

The works of Bolthausen, Deuschel and Zeitouni [7], Bramson and Zeitouni [13] and Ding [20] establish the tightness of the recentered maximum of the two-dimensional discrete Gaussian free field. We expect that their method can be applied to the Gaussian field we consider.

We note that Fang and Zeitouni [23] have studied a branching random walk model where the variance of the motion is time-dependent. This model is related to the simpler GREM model of spin glasses and to the CREM of Bovier and Kurkova. The family of log-correlated Gaussian fields introduced in Section 2.2 is akin to these hierarchical models, where the scale parameter replaces the time parameter.

2. Outline of the proof. The proof is split in three steps, and each can be adapted (with different correlation estimates) to other log-correlated Gaussian fields. The Gaussian field we study has a graphical representation of its correlations as well as no boundary effect which help in illustrating the method.

2.1. *A family of perturbed models.* In this section, we define a family of Gaussian fields for which the variance parameter σ is scale-dependent. It can be seen as the GREM analogue for the nonhierarchical Gaussian field considered here. We restrict ourselves to the case where σ takes two values, which is the one needed for the proof of Theorem 1.4. However, the construction and the results can hold for any finite number of values.

Fix $\varepsilon = 1/N$. We introduce a scale (or time) parameter t by defining for any $t \in [0,1]$,

$$X_x(t) := \omega_{\varepsilon^t}(x), \qquad x \in \mathcal{X}_\varepsilon.$$

Observe that for any fixed x, the process $(X_x(t))_{0\leq t\leq 1}$ has independent increments and is a martingale for the filtration $(\mathcal{F}_{\varepsilon^t}, t \geq 0)$,

$$\mathbb{E}[X_x(t)|\mathcal{F}_{\varepsilon^s}] = X_x(s) \qquad \text{for } t > s.$$

This is a consequence of the defining property (ii) of the random measure μ.

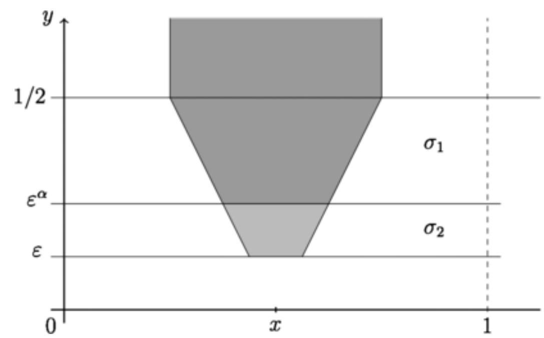

FIG. 2. *The cone associated with the process $Y_x(\cdot)$.*

The parameters of the family of perturbed models are α where $0<\alpha<1$ and $\vec{\sigma}=(\sigma_1,\sigma_2)$ with $\sigma_i>0$, $i=1,2$. For the sake of clarity and to avoid repetitive trivial corrections, it is assumed throughout the paper that N^α and $N^{1-\alpha}$ are integers. The Gaussian field $Y^{(\vec{\sigma},\alpha)}(t)=(Y_x^{(\vec{\sigma},\alpha)}(t), x\in\mathcal{X}_\varepsilon)$ is defined from the field X as follows:

$$Y_x^{(\vec{\sigma},\alpha)}(t)=\begin{cases}\sigma_1 X_x(t), & \text{if } 0<t\le\alpha,\\ \sigma_1 X_x(\alpha)+\sigma_2(X_x(t)-X_x(\alpha)), & \text{if } \alpha<t\le 1.\end{cases} \tag{2.1}$$

The construction is depicted in Figure 2. We write $Y^{(\vec{\sigma},\alpha)}$ for the field $(Y_x^{(\vec{\sigma},\alpha)}(1), x\in\mathcal{X}_\varepsilon)$. The dependence on $\vec{\sigma}$ and α will sometimes be dropped in the notation of Y for simplicity.

Consider the partition function $Z_N^{(\vec{\sigma},\alpha)}(\beta)$ of the perturbed model

$$Z_N^{(\vec{\sigma},\alpha)}(\beta):=\sum_{x\in\mathcal{X}_N}\exp(\beta Y_x), \tag{2.2}$$

and the free energy

$$f_N^{(\vec{\sigma},\alpha)}(\beta):=\frac{1}{\log N}\log Z_N^{(\vec{\sigma},\alpha)}(\beta) \qquad \forall\beta>0.$$

The log number of high points can be computed for the Gaussian field Y using Daviaud's technique recursively. The free energy is then obtained by doing an explicit sum on these high points. This is the object of Sections 3 and 4. The result is better expressed in terms of the free energy of the REM with N i.i.d. Gaussian variables of variance $\sigma^2\log N$,

$$f(\beta;\sigma^2):=\begin{cases}1+\dfrac{\beta^2\sigma^2}{2}, & \text{if } \beta\le\beta_c(\sigma^2):=\dfrac{\sqrt{2}}{\sigma},\\ \sqrt{2}\sigma\beta, & \text{if } \beta\ge\beta_c(\sigma^2).\end{cases}$$

Corollary 1.3 follows from the next result with the choice $\sigma_1=\sigma_2$.

PROPOSITION 2.1. *Let* $V_{12} := \sigma_1^2\alpha + \sigma_2^2(1-\alpha)$. *Then:*

- *Case 1: If* $\sigma_1 \le \sigma_2$,

$$\lim_{N\to\infty} f_N^{(\vec{\sigma},\alpha)}(\beta) = f(\beta; V_{12}).$$

- *Case 2: If* $\sigma_1 \ge \sigma_2$,

$$\lim_{N\to\infty} f_N^{(\vec{\sigma},\alpha)}(\beta) = \alpha f(\beta;\sigma_1^2) + (1-\alpha) f(\beta;\sigma_2^2),$$

where the convergence holds almost surely and in L^1.

The expressions are identical to the free energy of a GREM with two levels. In case 1, it is reduced to a REM. The conditions can be rewritten by defining a piecewise linear function of slopes σ_1^2 and σ_2^2 on the intervals $[0,\alpha]$, $[\alpha,1]$, respectively. In case 1, this function fails to be concave. However, it is easily verified that the effective parameters define the concave hull of the function. The reader is referred to [14] and [11] for more details on the concavity conditions which is very general for the family of GREM models. In case 1 there is one critical value for β, and in case 2 there are two critical values for β corresponding to the respective $\beta_c(\sigma^2)$ of the two effective parameters σ^2. In case 1, the critical β is $\sqrt{2/V_{12}}$, whereas the two critical β's are $\sqrt{2}/\sigma_1$ and $\sqrt{2}/\sigma_2$ in case 2.

2.2. *The Bovier–Kurkova technique.* The proof of Theorem 1.4 relies on determining the overlap distribution of the original model from the free energy of the perturbed ones. This approach has been used by Bovier and Kurkova in the case of the GREM-type models [11, 12].

For $u \in (-1,1)$ and $\alpha \in (0,1)$, consider the field $(Y_x, x \in \mathcal{X}_\varepsilon)$ defined in (2.1) with the choice of parameters $\vec{\sigma} = (1,(1+u))$; see Figure 3. (Recall

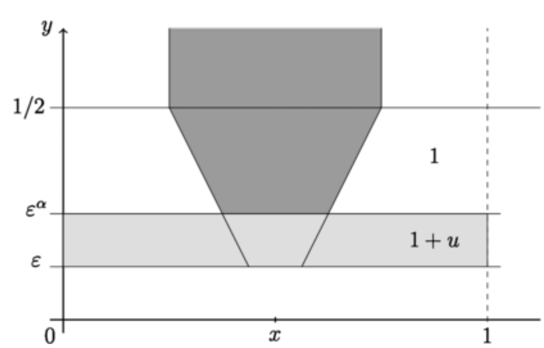

FIG. 3. *The perturbed model where the variance parameter is* $(1+u)$ *on the strip* $[\varepsilon,\varepsilon^\alpha]$ *where* $\varepsilon = 1/N$.

that, for the sake of clarity, it is assumed that N^α and $N^{1-\alpha}$ are integers.) The original Gaussian field (X_x) is recovered at $u=0$. Note that if $u>0$, the parameters correspond to the first case of Proposition 2.1 and if $u<0$, to the second. The field Y can also be represented as follows:

$$Y_x = X_x + u(X_x - X_x(\alpha)), \qquad 1 \le i \le N. \tag{2.3}$$

The proof of the next lemma is a simple integration and is postponed to the Appendix; see Appendix A.2.

Lemma 2.2. *Fix $0<\varepsilon=1/N<1/2$, and $\alpha\in(0,1)$. Let $\tilde{X}_x := X_x - X_x(\alpha)$. Then, for $x\in\mathcal{X}_\varepsilon$*

$$\mathbb{E}[\tilde{X}_x^2] = \mathbb{E}[\tilde{X}_x X_x] = (1-\alpha)\log N, \qquad x \in \mathcal{X}_\varepsilon,$$

and, for $x, x' \in \mathcal{X}_\varepsilon$,

$$\mathbb{E}[\tilde{X}_x X_{x'}] = \begin{cases} (q(x,x')-\alpha)\log N + O_N(1), & \text{if } \alpha < q(x,x') \le 1,\\ 0, & \text{if } 0 \le q(x,x') \le \alpha, \end{cases} \tag{2.4}$$

where $O_N(1)$ is a term uniformly bounded in N, and we recall that $\|x-x'\| = \varepsilon^{q(x,x')}$.

This result and a Gaussian integration by parts yield an important lemma.

Lemma 2.3. *For all $\alpha\in(0,1)$, we have*

$$\beta\int_\alpha^1 x_\beta^{(N)}(s)\,\mathrm{d}s + o_N(1) = \frac{1}{\log N}\mathbb{E}\biggl[\sum_{x\in\mathcal{X}_\varepsilon} G_{\beta,N}(x)(X_x - X_x(\alpha))\biggr],$$

where $o_N(1)$ stands for a term that goes to 0 as N goes to ∞.

Proof. Fix $\varepsilon = 1/N$ and $\alpha\in(0,1)$. Note that $(\tilde{X}_x;(X_{x'}, x'\in\mathcal{X}_\varepsilon))$ is a Gaussian vector of $N+1$ variables. Therefore, Gaussian integration by parts (see Lemma A.3) yields, for all $x\in\mathcal{X}_\varepsilon$,

$$\begin{aligned}\beta^{-1}\mathbb{E}\biggl[\frac{\tilde{X}_x \mathrm{e}^{\beta X_x}}{\sum_{x'\in\mathcal{X}_\varepsilon}\mathrm{e}^{\beta X_{x'}}}\biggr] &= -\sum_{x'\in\mathcal{X}_\varepsilon}\mathbb{E}[\tilde{X}_x X_{x'}]\mathbb{E}\biggl[\frac{\mathrm{e}^{\beta(X_x+X_{x'})}}{(\sum_{z\in\mathcal{X}_\varepsilon}\mathrm{e}^{\beta X_z})^2}\biggr]\\ &\quad + \mathbb{E}[\tilde{X}_x X_x]\mathbb{E}\biggl[\frac{\mathrm{e}^{\beta X_x}}{\sum_{z\in\mathcal{X}_\varepsilon}\mathrm{e}^{\beta X_z}}\biggr].\end{aligned}$$

Lemma 2.2 and elementary manipulations imply

$$(\beta\log N)^{-1}\mathbb{E}\biggl[\sum_{x\in\mathcal{X}_\varepsilon}\tilde{X}_x G_{\beta,N}(x)\biggr]$$

$$
\begin{aligned}
&= \sum_{x,x'\in\mathcal{X}_\varepsilon}\left(\int_\alpha^1 \mathbf{1}_{\{q(x,x')\le s\}}\,\mathrm{d}s\right)\mathbb{E}[G_{\beta,N}(x)G_{\beta,N}(x')] + O\left(\frac{1}{\log N}\right)\\
&= \int_\alpha^1 \mathbb{E}[G_{\beta,N}^{\times 2}\{q_{12}\le s\}]\,\mathrm{d}s + O\left(\frac{1}{\log N}\right),
\end{aligned}
$$

which concludes the proof of the lemma. □

Proof of Theorem 1.4. Fix $\beta > \beta_c = \sqrt{2}$. Write $Z_N^{(u,\alpha)}(\beta)$ for the partition function (2.2) for the choice $\vec{\sigma} = (1,(1+u))$. Direct differentiation and equation (2.3) give

$$
\frac{\mathrm{d}}{\mathrm{d}u}(\mathbb{E}\log Z_N^{(u,\alpha)}(\beta))_{u=0} = \beta\mathbb{E}\left[\sum_{x\in\mathcal{X}_\varepsilon}(X_x - X_x(\alpha))G_{\beta,N}(x)\right],
$$

which, together with Lemma 2.3, yields

$$
\int_\alpha^1 x_\beta^{(N)}(s)\,\mathrm{d}s = \beta^{-2}(\log N)^{-1}\frac{\mathrm{d}}{\mathrm{d}u}(\mathbb{E}\log Z_N^{(u,\alpha)}(\beta))_{u=0} + o_N(1). \tag{2.5}
$$

Observe that $\mathbb{E}f_N^{(u,\alpha)}(\beta) = (\log N)^{-1}\mathbb{E}\log Z_N^{(u,\alpha)}(\beta)$ is a convex function of u. Moreover, by Proposition 2.1, $\mathbb{E}f_N^{(u,\alpha)}(\beta)$ converges. The limit, that we denote $f^{(u,\alpha)}(\beta)$, is also convex in the parameter u. In particular, by a standard result of convexity (see, e.g., Proposition I.3.2 in [32]), at every point of differentiability, the derivative of the limit equals the limit of the derivative

$$
\lim_{N\to\infty}\frac{\mathrm{d}}{\mathrm{d}u}\mathbb{E}f_N^{(u,\alpha)}(\beta) = \frac{\mathrm{d}}{\mathrm{d}u}f^{(u,\alpha)}(\beta) \qquad \forall u \text{ where } u\mapsto f^{(u,\alpha)}(\beta) \text{ is differentiable.} \tag{2.6}
$$

We show $f^{(u,\alpha)}(\beta)$ is differentiable at $u=0$. The derivative can be computed by Proposition 2.1. For u small enough, β is larger than all critical β's. Thus

$$
\frac{\mathrm{d}}{\mathrm{d}u}f^{(u,\alpha)}(\beta) = \begin{cases} \sqrt{2}\beta\dfrac{(1-\alpha)(1+u)}{\sqrt{\alpha+(1-\alpha)(1+u)^2}}, & \text{if } u>0,\\ \sqrt{2}\beta(1-\alpha), & \text{if } u<0. \end{cases} \tag{2.7}
$$

From this, it is easily verified that $f^{(u,\alpha)}(\beta)$ is differentiable at $u=0$ and

$$
\frac{\mathrm{d}}{\mathrm{d}u}(f^{(u,\alpha)}(\beta))_{u=0} = \sqrt{2}\beta(1-\alpha). \tag{2.8}
$$

Equations (2.5), (2.6) and (2.8) together imply

$$
\lim_{N\to\infty}\int_\alpha^1 x_\beta^{(N)}(s)\,\mathrm{d}s = \frac{\sqrt{2}}{\beta}(1-\alpha) \qquad \text{for all } \alpha\in(0,1). \tag{2.9}
$$

Therefore, any weak limit x_β must satisfy $x_\beta(\alpha) \leq \sqrt{2}/\beta$ for any point of continuity $\alpha < 1$, since x_β is nondecreasing. If there exists $0 < \alpha < 1$ such that $x_\beta(\alpha) < \frac{\sqrt{2}}{\beta}$, there would be a contradiction with (2.9), since by right-continuity and monotonicity of x_β we could find $\alpha' > \alpha$ such that

$$\lim_{N\to\infty} \int_\alpha^{\alpha'} x_\beta^{(N)}(s)\,\mathrm{d}s < \frac{\sqrt{2}}{\beta}(\alpha' - \alpha).$$

This proves that any weak limit x_β of $(x_\beta^{(N)}, N \in \mathbb{N})$ is the same and equals $\frac{\sqrt{2}}{\beta}$ on $(0,1)$. The subsequential limits being the same, this proves in particular convergence of the sequence to the desired distribution function. □

2.3. *A spin-glass approach to Poisson–Dirichlet variables.* In this section, the link between Theorems 1.4 and 1.5 is explained. The technique, inspired from the study of spin glasses in particular [2], is general and is of independent interest to prove convergence to Poisson–Dirichlet statistics.

The first step is to find a good space for the convergence of $G_{\beta,N}$. Let C be the compact metric space of $\mathbb{N} \times \mathbb{N}$ covariance matrices with 1 on the diagonal endowed with the product topology on the entries. For a given N, consider the mapping

$$\mathcal{X}_N^{\times\infty} \to C,$$
$$(x_l, l \in \mathbb{N}) \mapsto R^{(N)},$$

where for $l, l' \in \mathbb{N}$

$$R_{l,l'}^{(N)} := \begin{cases} q_{ll'} = q(x_l, x_l'), & \text{if } l \neq l' \\ 1, & \text{if } l = l'. \end{cases}$$

Consider the probability measure $\mathbb{E}G_{\beta,N}^{\times\infty}$ on $\mathcal{X}_N^{\times\infty}$. The push-forward of this probability measure under the above mapping defines a random element of C that we denote $\vec{R}^{(N)}$. Since each point is sampled independently from the same measure, the law of $\vec{R}^{(N)}$ is *weakly exchangeable*, that is, for any permutation π of a finite number of indices,

$$(\vec{R}_{\pi(l)\pi(l')}^{(N)}) \overset{\text{law}}{=} (\vec{R}_{ll'}^{(N)}).$$

The sequence of random matrices $(\vec{R}^{(N)}, N \in \mathbb{N})$ is tight by Prokhorov's theorem since the space C is a compact metric space. Hence, there exists a subsequence $\{\vec{R}^{(N_m)}\}_{m\in\mathbb{N}}$ that converges weakly. Denote the subsequential limit by $\vec{R}$. Observe that $\vec{R}$ is also weakly exchangeable since the mappings on C induced by a finite permutation is continuous. Therefore, by the representation theorem of Dovbysh and Sudakov [21], $\vec{R}$ is constructed like $\vec{R}^{(N)}$ by sampling from a random measure. Precisely, the theorem states that there exists a random probability measure μ_β on a Hilbert space $\mathcal{H}$, with law P

and corresponding expectation E, such that the random matrix $\vec{R}$ has the same law as the Gram matrix of a sequence of vectors $(v_l, l \in \mathbb{N})$ that are sampled under $E\mu_\beta^{\times\infty}$. [In other words, the vectors $(v_l, l \in \mathbb{N})$ are i.i.d. conditionally on μ_β.] The equality in law can be expressed as follows: for any continuous function F on C,

$$\lim_{m\to\infty} \mathbb{E}G_{\beta,N_m}^{\times\infty}[F(q_{ll'})] = E\mu_\beta^{\times\infty}[F(v_l \cdot v_l')]. \tag{2.10}$$

Note that, since $q(x,x') \le 1$, the random measure μ_β is supported on the unit ball. The first consequence of Theorem 1.4 is that for any subsequential limit μ_β,

$$\begin{aligned} E[\mu_\beta^{\times 2}\{v_1 \cdot v_2 \le q\}] &= \lim_{N\to\infty} \mathbb{E}[G_{\beta,N}^{\times 2}\{q_{12} \le q\}] \\ &= \frac{\beta_c}{\beta} 1_{[0,1)}(q) + 1_{\{1\}}(q). \end{aligned} \tag{2.11}$$

The first equality is obtained by bounding $1_{[0,q]}(q_{ll'})$ by continuous functions on $q_{ll'}$ above and below and by applying (2.10). In view of equations (2.10) and (2.11), we see the random measures μ_β as limit points of $(G_{\beta,N})_{N\in\mathbb{N}}$.

The main ingredient to prove Poisson–Dirichlet statistics is a general property of the Gibbs measure $(G_{\beta,N}(x), x \in \mathcal{X}_N)$ of centered Gaussian fields known as the Ghirlanda–Guerra identities. They were introduced in [26] and were proved in a general setting by Panchenko [29].

THEOREM 2.4. *Let μ_β be a subsequential limit of $(G_{\beta,N})_{N\in\mathbb{N}}$ in the sense of (2.10). Then for any $s \in \mathbb{N}$ and any continuous functions F: $[-1,1]^{s(s-1)/2} \to \mathbb{R}$*

$$\begin{aligned} E\mu_\beta^{\times s+1}[v_1 \cdot v_{s+1} F(v_l \cdot v_{l'})] &= \frac{1}{s} E\mu_\beta^{\times 2}[v_1 \cdot v_2] E\mu_\beta^{\times s}[F(v_l \cdot v_{l'})] \\ &\quad + \frac{1}{s}\sum_{k=2}^{s} E\mu_\beta^{\times s}[v_1 \cdot v_k F(v_l \cdot v_{l'})]. \end{aligned} \tag{2.12}$$

PROOF. Recall that we write $G_{\beta,N}^{\times s}$ for the product measure on $\mathcal{X}_N^s$. Also for $(x_1,\dots,x_s) \in \mathcal{X}_N^s$, the overlaps $q(x_l, x_l')$, $1 \le l, l' \le s$, are denoted $q_{ll'}$. In a similar way, we write X_1 for the field X_{x_1} of the first point sampled from $G_{\beta,N}$. It is shown in [29] that, for any β where the free energy $f(\beta)$ is differentiable, the following concentration holds:

$$\lim_{N\to\infty} \frac{1}{\log N} \mathbb{E}G_{\beta,N}[|X_1 - \mathbb{E}G_{\beta,N}(X_1)|] = 0. \tag{2.13}$$

Note that by Corollary 1.3, differentiability holds at all β for the Gaussian field considered. Since the function F is bounded, (2.13) implies

$$\lim_N \frac{1}{\log N}(\mathbb{E}G_{\beta,N}^{\times s}[X_1 F(q_{ll'})] - \mathbb{E}G_{\beta,N}[X_1]\mathbb{E}G_{\beta,N}^{\times s}[F(q_{ll'})]) = 0. \tag{2.14}$$

The two terms can be evaluated by Gaussian integrations by part (see Lemma A.3),

$$\frac{1}{\beta \log N}\mathbb{E}G_{\beta,N}[X_1] = 1 - \mathbb{E}G_{\beta,N}^{\times 2}[q_{12}] + O\left(\frac{1}{\log N}\right) \tag{2.15}$$

and

$$\begin{aligned}&\frac{1}{\beta \log N}\mathbb{E}G_{\beta,N}^{\times s}[X_1 F(q_{ll'})] \\ &\qquad = -s\mathbb{E}G_{\beta,N}^{\times s+1}[q_{1,s+1}F(q_{ll'})] + \sum_{1\le k\le s}\mathbb{E}G_{\beta,N}^{\times s}[q_{1k}F(q_{ll'})] \\ &\qquad\quad + O\left(\frac{1}{\log N}\right).\end{aligned} \tag{2.16}$$

Finally recalling (2.14) and assembling (2.15)–(2.16) yields the Ghirlanda–Guerra identities (see equation (16) in [26]),

$$\begin{aligned}&\mathbb{E}G_{\beta,N}^{\times s+1}[q_{1,s+1}F(q_{ll'})] \\ &\qquad = \frac{1}{s}\mathbb{E}G_{\beta,N}^{\times 2}[q_{12}]\mathbb{E}G_{\beta,N}^{\times s}[F(q_{ll'})] + \frac{1}{s}\sum_{k=2}^{s}\mathbb{E}G_{\beta,N}^{\times s}[q_{1k}F(q_{ll'})] \\ &\qquad\quad + o_N(1).\end{aligned} \tag{2.17}$$

[Note that the term for $k=1$ cancels with the 1 since $q_{11} = 1 + o_N(1)$.] In particular, for any subsequential limit μ_β of $(G_{\beta,N})_N$ in the sense of (2.10), one obtains (2.12) by taking the limit $N\to\infty$ and applying the definition of convergence in the sense of (2.10). □

Equation (2.11) and the Ghirlanda–Guerra identities imply that μ_β is atomic.

Corollary 2.5. *Let μ_β be a subsequential limit of $(G_{\beta,N})_{N\in\mathbb{N}}$ in the sense of (2.10). Then there exist random weights $\xi = (\xi_i; i\in\mathbb{N})_\downarrow$ with $\xi_i \ge 0$, $\sum_{i\in\mathbb{N}}\xi_i = 1$ and orthonormal vectors $(e_i; i\in\mathbb{N}) \subset \mathcal{H}$ such that*

$$\mu_\beta = \sum_{i\in\mathbb{N}}\xi_i\delta_{e_i}, \qquad P\text{-a.s.}$$

Moreover, from (2.11), $E[\sum_{i\in\mathbb{N}}\xi_i^2] = 1 - \frac{\beta_c}{\beta}$.

Proof. Let $(v_l, l\in\mathbb{N})$ be a sequence sampled from $E\mu_\beta^{\times\infty}$. From $(v_l, l\in\mathbb{N})$, we reconstruct μ_β up to isometry. For a fixed l consider the sequence $(v_l\cdot v_{l'}, l'>l)$. This is a sequence of 0's and 1's by (2.11). We first show that, almost surely, for every $l\in\mathbb{N}$, there exists $l'>l$ such that $v_l\cdot v_{l'} = 1$; in particular, since all vectors are in the unit ball, $v_l = v_{l'}$ and $\|v_l\| = 1$. For this, we proceed as in Lemma 1 in [28]. Write $F_s(v_l\cdot v_{l'})=\prod_{l=2}^s(1-v_1\cdot v_l)$. In other words, $F_s(v_l\cdot v_{l'})$ is 1 if $v_1\cdot v_l=0$ for $l=2,\dots,s$, otherwise it is 0. Denote for short $\alpha=1-E\mu_\beta^{\times 2}[v_1\cdot v_2]$. Equation (2.12) implies

$$\begin{aligned}
&E\mu_\beta^{\times s+1}\{v_1\cdot v_l=0, 2\le l\le s+1\}\\
&\quad=E\mu_\beta^{\times s+1}[(1-v_1\cdot v_{s+1})F_s(v_l\cdot v_{l'})]\\
&\quad=\frac{\alpha}{s}E\mu_\beta^{\times s}\{v_1\cdot v_l=0,2\le l\le s\}+\frac{1}{s}\sum_{l=2}^{s}E\mu_\beta^{\times s}\{v_1\cdot v_l=0,2\le l\le s\}\\
&\quad=\frac{s-1+\alpha}{s}E\mu_\beta^{\times s}\{v_1\cdot v_l=0,2\le l\le s\}=\frac{(s-1+\alpha)\cdots(1+\alpha)\alpha}{s!},
\end{aligned}$$

where the last equality is obtained by induction. The last term goes to 0 as $s\to\infty$ since $\alpha<1$, hence

$$E\mu_\beta^{\times\infty}\{v_1\cdot v_l=0,l\ge 2\}=0,$$

from which we deduce that, P-a.s., $\mu_\beta^{\times\infty}\{v_1\cdot v_l=0,l\ge 2\}=0$ and then that, for μ_β-almost all v,

$$\mu_\beta^{\times\infty}\{v\cdot v_l=0,l\ge 2\}=0.$$

Since the vectors v_l are i.i.d. μ_β-sampled, it follows that, P-a.s., for μ_β-almost all v, $\mu_\beta(v\cdot v_1=0)<1$, thus $\mu_\beta(v\cdot v_1=1)>0$ as claimed.

By the reasoning above, a vector that is sampled once in $(v_l,l\in\mathbb{N})$ is sampled infinitely many times $E\mu_\beta^{\times\infty}$-a.s. Moreover, since the vectors are conditionally i.i.d., for $l\in\mathbb{N}$, the following limit exists and must be nonzero:

$$\lim_{n\to\infty}\frac{1}{n}\sum_{j=l+1}^{l+n}1_{v_l}(v_j)>0,\qquad E\mu_\beta^{\times\infty}\text{-a.s.}\tag{2.18}$$

In particular, every sampled vector v_l is an atom a.s. and its weight is measurable with respect to $(v_l,l\in\mathbb{N})$. Moreover, if $v_l\neq v_{l'}$, then $v_l\cdot v_{l'}=0$ $E\mu_\beta^{\times\infty}$-a.s. Therefore the atoms are orthogonal. It remains to consider the different atoms without repetitions and reorder the weights. Let $e_1=v_1$, $e_2=v_{l_2}$ where $l_2=\inf\{l\ge 1: v_l\cdot e_1=0\}$, $e_3=v_{l_3}$ where $l_3=\inf\{l\ge l_2: v_l\cdot e_i=0, i=1,2\}$, and so forth. By construction, $(e_j,j\ge 1)$ are orthonormal vectors. (The collection is not necessarily infinite at this point.) We can assign to each vector e_j its weight $\mu_\beta(\{e_j\})$ by (2.18). The collection can then be ordered in decreasing order to get the result.

The fact that $E[\sum_{i\in\mathbb{N}}\xi_i^2]=1-\frac{\beta_c}{\beta}$ is straightforward from (2.11). $\square$

To finish the proof of Theorem 1.5, it remains to show that the random weights ξ are distributed like a Poisson–Dirichlet variable of parameter $\frac{\beta_c}{\beta}$.

In fact, the parameter is already determined by Corollary 2.5, since for a Poisson–Dirichlet variable ξ' of parameter x, $E[\sum_k (\xi'_k)^2] = 1-x$ holds; see, for example, Corollary 2.2 in [31]. This will also imply that for any converging sequence of $(G_{\beta,N})$ in the sense of (2.10), the limit is the same. In particular, it implies convergence of the whole sequence by compactness.

To prove the Poisson–Dirichlet statistics of the weights ξ, we use the following characterization theorem of the law; see [33], page 22 for details. Define for all $m \in \mathbb{N}$ the joint moments of the weights

$$S(n_1,\dots,n_m) = E \sum_{k_1,\dots,k_m} \xi_{k_1}^{n_1} \cdots \xi_{k_m}^{n_m} \qquad \text{for } n_1,\dots,n_m \geq 1. \tag{2.19}$$

The collection of $S(n_1,\dots,n_m)$, $m \in \mathbb{N}$, determines the law of a random mass-partition, that is, a random variable on ordered sequences $1 \geq r_1 \geq r_2 \geq \cdots \geq 0$ with $\sum_{i\in\mathbb{N}} r_i \leq 1$. If ξ is a Poisson–Dirichlet variable, it is shown in [33], Proposition 1.2.8, that the moments satisfy the recursion relations

$$\begin{aligned} S(n_1+1,\dots,n_m) &= \frac{S(2)}{s} S(n_1,\dots,n_m) + \frac{n_1-1}{s} S(n_1,\dots,n_m) \\ &\quad + \sum_{2\leq l\leq m} \frac{n_l}{s} S(n_1+n_l, n_2,\dots,n_{l-1},n_{l+1},\dots,n_m), \end{aligned} \tag{2.20}$$

where $s = n_1 + \cdots + n_m$. It is not hard to verify that all moments $S(n_1,\dots,n_m)$ (and thus the law of ξ) are determined by recursion from $S(2)$ and the identities (2.20).

It turns out that these identities are satisfied by ξ defined by Theorem 2.4 and Corollary 2.5.

THEOREM 2.6. *Let ξ be a random mass-partition satisfying the assumptions of Corollary 2.5. The moments $S(n_1,\dots,n_m)$ of ξ satisfy (2.20) for any $m \in \mathbb{N}$ and any $n_1,\dots,n_m \in \mathbb{N}$. In particular, ξ has the law of a Poisson–Dirichlet variable of parameter $1-S(2)$.*

PROOF. To deduce (2.20) from (2.12), we follow [33], pages 24–25. The set $\{1,\dots,s\}$ can be decomposed into the disjoint union of sets $I_1,\dots,I_m$ with $|I_j| = n_j$ for all $1 \leq j \leq m$. Consider the functions $(F_j)_{1\leq j\leq m}$ given by $F_j(\delta_{k_l k_{l'}}) := \prod_{k_l,k_{l'}\in I_j} \delta_{k_l k_{l'}}$ and define $F := \prod_{1\leq j\leq m} F_j$. Then elementary manipulations imply (2.20). Note that the second term on the right-hand side of (2.12) yields the last two terms of (2.20). □

3. High points of the perturbed models. In this section, the log-number of high points at a given level is computed for the perturbed models introduced in Section 2. The focus is on the Gaussian field introduced in

Section 2.1, though the technique applies to any perturbed model with a finite number of parameters. The free energies of the models are computed in Section 4.

Let $Y=(Y_x, x\in\mathcal{X}_\varepsilon)$ be the Gaussian field introduced in Section 2.1. Recall the notation and the two choices of parameters in Proposition 2.1:

$$\begin{aligned} &\textit{Case } 1: && \sigma_1\le\sigma_2;\\ &\textit{Case } 2: && \sigma_1\ge\sigma_2. \end{aligned} \tag{3.1}$$

Define also as before $V_{12}:=\sigma_1^2\alpha+\sigma_2^2(1-\alpha)$.

Proposition 3.1.

$$\lim_{N\to\infty}\mathbb{P}\Big(\max_{x\in\mathcal{X}_\varepsilon}Y_x\ge\sqrt{2}\gamma_{\max}\log N\Big)=0,$$

where

$$\gamma_{\max}=\gamma_{\max}(\vec\sigma,\alpha):=\begin{cases}\sqrt{V_{12}}, & \textit{for case 1;}\\ \sigma_1\alpha+\sigma_2(1-\alpha), & \textit{for case 2.}\end{cases}$$

Proposition 3.2. *Let $\mathcal{H}_N^Y(\gamma):=\{x\in\mathcal{X}_\varepsilon:Y_x\ge\sqrt{2}\gamma\log N\}$ be the set of γ-high points. Then, for all $0<\gamma<\gamma_{\max}$,*

$$\lim_{N\to\infty}\frac{\log|\mathcal{H}_N^Y(\gamma)|}{\log N}=\mathcal{E}^{(\vec\sigma,\alpha)}(\gamma)\qquad\textit{in probability,}$$

where in case 1,

$$\mathcal{E}^{(\vec\sigma,\alpha)}(\gamma):=1-\frac{\gamma^2}{V_{12}};$$

and in case 2,

$$\mathcal{E}^{(\vec\sigma,\alpha)}(\gamma):=\begin{cases}1-\dfrac{\gamma^2}{V_{12}}, & \textit{if } \gamma<\dfrac{V_{12}}{\sigma_1},\\[2ex] (1-\alpha)-\dfrac{(\gamma-\sigma_1\alpha)^2}{\sigma_2^2(1-\alpha)}, & \textit{if } \gamma\ge\dfrac{V_{12}}{\sigma_1}.\end{cases}$$

Moreover, for any $\mathcal{E}<\mathcal{E}^{(\vec\sigma,\alpha)}(\gamma)$, there exists c such that

$$\mathbb{P}(|\mathcal{H}_N^Y(\gamma)|\le N^{\mathcal{E}})\le\exp\{-c(\log N)^2\}.$$

3.1. *Proof of Proposition 3.1.* The proof of case 1 is by a union bound,

$$\mathbb{P}\Big(\max_{x\in\mathcal{X}_\varepsilon}Y_x\ge\sqrt{2}\gamma_{\max}\log N\Big)\le N\mathbb{P}(Y_x\ge\sqrt{2}\gamma_{\max}\log N),$$

which goes to zero by a Gaussian estimate; see Lemma A.1. For case 2, we construct a Gaussian field with hierarchical correlations that dominates Y

at the level of the covariances. The result will follow by comparison using Slepian's lemma.

Notice that if $\varepsilon < \|x-x'\| \leq \varepsilon^\alpha$, the corresponding cone-like sets for Y_x and $Y_{x'}$ in $\mathcal{C}^+$ intersect between the lines $y=\varepsilon$ and $y=\varepsilon^\alpha$. Therefore the covariance of the variables satisfies, writing $\ell := \|x-x'\|$,

$$\begin{aligned}\mathbb{E}[Y_x Y_{x'}] &= \sigma_2^2 \int_\ell^{\varepsilon^\alpha} \frac{y-\ell}{y^2}\,\mathrm{d}y + \sigma_1^2\biggl(\int_{\varepsilon^\alpha}^{1/2} \frac{y-\ell}{y^2}\,\mathrm{d}y + \int_{1/2}^{\infty} \frac{1/2-\ell}{y^2}\,\mathrm{d}y\biggr)\\ &\geq \sigma_1^2\biggl(\log\frac{1/2}{\varepsilon^\alpha} - 1\biggr).\end{aligned}$$

By applying the same reasoning when $\varepsilon^\alpha < \|x-x'\| \leq 1/2$, one obtains the following lower bound for the covariance:

$$\mathbb{E}[Y_x Y_{x'}] \geq \begin{cases} 0, & \text{if } \|x-x'\| > \varepsilon^\alpha,\\ \sigma_1^2\biggl(\log\dfrac{1/2}{\varepsilon^\alpha} - 1\biggr), & \text{if } \varepsilon < \|x-x'\| \leq \varepsilon^\alpha.\end{cases} \tag{3.2}$$

Equation (3.2) is used to construct a Gaussian field $\tilde{Y}$. Define the map

$$\begin{aligned}\pi : \mathcal{X}_\varepsilon &\to \mathcal{X}_{\varepsilon^\alpha},\\ x &\mapsto \pi(x),\end{aligned}$$

where $\pi(x)$ is the unique $y \in \mathcal{X}_{\varepsilon^\alpha}$ such that $\|x-y\| \leq \frac{\varepsilon^\alpha}{2}$. (If $\|x-y\| = \frac{\varepsilon^\alpha}{2}$, there are two possibilities for y. We take the right point.) The pre-image of $y \in \mathcal{X}_{\varepsilon^\alpha}$ under π are exactly the points in $\mathcal{X}_\varepsilon$ that are at a distance less than $\frac{\varepsilon^\alpha}{2}$ from y. One can think of $\pi(x)$ as the *ancestor* of x at the scale ε^α.

Consider the following Gaussian variables

$$\begin{aligned} &(g_x^{(1)}, x \in \mathcal{X}_{\varepsilon^\alpha}) \quad \text{i.i.d. Gaussians of variance } \sigma_1^2\alpha\log N - \sigma_1^2\log 2 - \sigma_1^2,\\ &(g_x^{(2)}, x \in \mathcal{X}_{\varepsilon}) \quad \text{i.i.d. Gaussians of variance } \sigma_2^2(1-\alpha)\log N + 2\sigma_1^2.\end{aligned} \tag{3.3}$$

These two families are also assumed independent. Then, the field $\tilde{Y}$ is defined, using the map π above and the Gaussian random variables $g_x^{(i)}$, by

$$\tilde{Y}_x = g_{\pi(x)}^{(1)} + g_x^{(2)}. \tag{3.4}$$

This construction and equation (3.2) directly imply the following comparison lemma.

LEMMA 3.3.

$$\begin{aligned}\mathbb{E}[\tilde{Y}_x^2] &= \mathbb{E}[Y_x^2] \qquad \forall x \in \mathcal{X}_\varepsilon,\\ \mathbb{E}[\tilde{Y}_x\tilde{Y}_y] &\leq \mathbb{E}[Y_xY_y] \qquad \forall\ x \neq y,\ x,y \in \mathcal{X}_\varepsilon.\end{aligned} \tag{3.5}$$

The following corollary is a straightforward consequence of the above lemma and Slepian's lemma; see Corollary 3.12 in [27].

COROLLARY 3.4. *For any* $\lambda > 0$,

$$\mathbb{P}\Big(\max_{x\in\mathcal{X}_\varepsilon} Y_x \geq \lambda\Big) \leq \mathbb{P}\Big(\max_{x\in\mathcal{X}_\varepsilon} \tilde{Y}_x \geq \lambda\Big). \tag{3.6}$$

The Gaussian field $\tilde{Y}$ is almost identical to a GREM model with two levels with parameters $0 < \alpha < 1$ and σ_1, σ_2; see, for example, [11, 18]. In fact the only aspect different from an exact GREM are the terms of order one in the variances of the Gaussian random variables $g_x^{(i)}$'s. However, these do not affect the first order of the maximum. The proof of Proposition 3.1 is concluded by the following standard GREM result. The proof of the lemma is not hard and is omitted for conciseness. The reader is referred to Theorem 1.1 in [11] where a stronger result on the maximum is given and to [10], Lecture 9, for more details on the free energy and on the log-number of high points of a two-level GREM.

LEMMA 3.5. *Let* $\tilde{Y}$ *be the Gaussian field constructed above. Then*

$$\mathbb{P}\Big(\max_{x\in\mathcal{X}_\varepsilon} \tilde{Y}_x \geq \sqrt{2}\gamma_{\max}\log N\Big) \to 0, \qquad N\to\infty,$$

where $\gamma_{\max}$ *is defined in Proposition 3.1.*

3.1.1. *Proof of the upper bound in Proposition 3.2.* The goal is to get an upper bound in probability) for $|\mathcal{H}_N^Y(\gamma)|$ where $\mathcal{H}_N^Y(\gamma) = \{x \in \mathcal{X}_\varepsilon : Y_x \geq \sqrt{2}\gamma\log N\}$.

In case 1, a first moment computation gives the result. Indeed, a Gaussian estimate (see Lemma A.1) gives

$$\mathbb{E}[|\mathcal{H}_N^Y(\gamma)|] = N\mathbb{P}(Y_1 \geq \sqrt{2}\gamma\log N) \leq CN^{\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)},$$

where $\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma) = 1 - \gamma^2/V_{12}$. Therefore, by Markov's inequality, for any $\rho > 0$,

$$\mathbb{P}(|\mathcal{H}_N^Y(\gamma)| \geq N^{\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)+\rho}) \leq CN^{-\rho} \to 0, \qquad N \to 0.$$

In case 2, if $0 < \gamma < V_{12}/\sigma_1 =: \gamma_{\text{crit}}$ the same argument gives the correct bound.

It remains to bound the case $\gamma \geq \gamma_{\text{crit}}$. The argument is essentially an explicit comparison with a 2-level GREM. For the scale α, define

$$\mathcal{H}_{N^\alpha}^Y(\gamma) := \{x \in \mathcal{X}_{\varepsilon^\alpha} : Y_x(\alpha) \geq \sqrt{2}\gamma\log N\}, \qquad \mathcal{E}_1(\gamma) := \alpha - \frac{\gamma^2}{\sigma_1^2\alpha}.$$

A first moment computation yields, for any $0<\gamma_1<\sigma_1\alpha$ and any $\rho>0$,

$$\mathbb{P}(|\mathcal{H}^Y_{N^\alpha}(\gamma_1)|\geq N^{\mathcal{E}_1(\gamma_1)+\rho})\leq CN^{-\rho}\to 0, \qquad N\to 0. \tag{3.7}$$

Similarly, a union bound gives

$$\mathbb{P}\Big(\max_{x\in\mathcal{X}_{\varepsilon^\alpha}} Y_x(\alpha)\geq \sqrt{2}\sigma_1\alpha\log N\Big)\to 0. \tag{3.8}$$

Recall that, for any $x\in\mathcal{X}_\varepsilon$, we denote by $\pi(x)$ the closest point in $\mathcal{X}_{\varepsilon^\alpha}$, hence $\|x-\pi(x)\|\leq\varepsilon^\alpha/2$. We define for all N and $\nu>0$,

$$A_{N,\nu}:=\bigcup_{x\in\mathcal{X}_\varepsilon}\{|Y_x(\alpha)-Y_{\pi(x)}(\alpha)|\geq\nu\log N\}.$$

The parameter ν will be fixed later and will depend on ρ. Using a union bound together with Lemma A.4, we obtain, for all $\nu>0$,

$$\mathbb{P}(A_{N,\nu})\leq CN\mathrm{e}^{-c(\log N)^2}\to 0, \qquad N\to 0. \tag{3.9}$$

We also consider the events giving the log-number of high points at scale α. Precisely, we divide $[0,\sigma_1\alpha]$ in intervals of size $\sigma_1\alpha/M$ where M will be fixed later. Define $\eta_i:=i\sigma_1\alpha/M$, for $0\leq i\leq M$ and

$$I^{(i)}:=[\sqrt{2}\eta_{i-1}\log N;\sqrt{2}\eta_i\log N], \qquad 1\leq i\leq M.$$

By (3.7), the events

$$B_{N,i}:=\{|\mathcal{H}^Y_{N^\alpha}(\eta_{i-1})|\geq N^{\mathcal{E}_1(\eta_{i-1})+\rho/2}\}, \qquad 1\leq i\leq M$$

are such that

$$\mathbb{P}\left(\bigcup_{i=1}^M B_{N,i}\right)\to 0, \qquad N\to 0. \tag{3.10}$$

Therefore, by (3.9) and (3.10), we are reduced to estimate

$$\mathbb{P}\left(\{|\mathcal{H}^Y_N(\gamma)|\geq N^{\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)+\rho}\}\cap A^c_{N,\nu}\cap\bigcap_{i=1}^M B^c_{N,i}\right),$$

which is smaller than

$$\frac{1}{N^{\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)+\rho}}\mathbb{E}\left[|\mathcal{H}^Y_N(\gamma)|;A^c_{N,\nu},\bigcap_{i=1}^M B^c_{N,i}\right]. \tag{3.11}$$

We split the set $\mathcal{H}^Y_N(\gamma)$ into the possible value of the field at scale α

$$\mathcal{H}^{(i)}_N(\gamma):=\{x\in\mathcal{X}_\varepsilon:Y_x\geq\sqrt{2}\gamma\log N;Y_{\pi(x)}(\alpha)\in I^{(i)}\}, \qquad 1\leq i\leq M,$$
$$\mathcal{H}^{(0)}_N(\gamma):=\{x\in\mathcal{X}_\varepsilon:Y_x\geq\sqrt{2}\gamma\log N;Y_{\pi(x)}(\alpha)\leq 0\}.$$

The term in (3.11) can then be bounded above by

$$\frac{1}{N^{\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)+\rho}}\sum_{i=0}^{M}\mathbb{E}[|\mathcal{H}_N^{(i)}(\gamma)|;A_{N,\nu}^c\cap B_{N,i}^c].$$

If $0\le\gamma\le\gamma_{\max}$, note that $\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)$ satisfies $\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)=\max_{0\le\eta\le\sigma_1\alpha}Q(\eta)$ where

$$Q(\eta):=1-\frac{\eta^2}{\sigma_1^2\alpha}-\frac{(\gamma-\eta)^2}{\sigma_2^2(1-\alpha)}.$$

Moreover, if $\gamma_{\mathrm{crit}}\le\gamma\le\gamma_{\max}$, the maximum is attained at $\eta=\sigma_1\alpha$, thus $Q(\eta)\le\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)$ for all $\eta\in[0,\sigma_1\alpha]$. For $1\le i\le M$, one gets

$$\begin{aligned}
&\mathbb{E}[|\mathcal{H}_N^{(i)}(\gamma)|;A_{N,\nu}^c\cap B_{N,i}^c]\\
&\quad=\mathbb{E}\Big[\sum_{x\in\mathcal{X}_\varepsilon}\mathbf{1}_{\{Y_x\ge\sqrt{2}\gamma\log N,Y_{\pi(x)}(\alpha)\in I^{(i)}\}};A_{N,\nu}^c\cap B_{N,i}^c\Big]\\
&\quad\le\mathbb{E}\Big[\sum_{x\in\mathcal{X}_\varepsilon}\mathbf{1}_{\{Y_x-Y_x(\alpha)\ge\sqrt{2}(\gamma-\eta_i-\nu)\log N,Y_{\pi(x)}(\alpha)\ge\sqrt{2}\eta_{i-1}\log N\}};B_{N,i}^c\Big]\\
&\quad\le CN^{\mathcal{E}_1(\eta_{i-1})+\rho/2}N^{1-\alpha}N^{-(\gamma-\eta_i-\nu)^2/(\sigma_2^2(1-\alpha))}\\
&\quad=CN^{\rho/2}N^{1-(\eta_{i-1})^2/(\sigma_1^2\alpha)-(\gamma-\eta_i-\nu)^2/(\sigma_2^2(1-\alpha))},
\end{aligned}$$

where the last inequality follows by the definition of $B_{N,i}$ the independence of the field at different scales and a Gaussian estimate. Since $Q(\eta)\le\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)$ for all $\eta\in[0,\sigma_1\alpha]$, the last term is smaller than $CN^{\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)+3\rho/4}$ by taking ν small enough and M large enough, but fixed. For $i=0$, a similar argument gives also the bound $CN^{\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)+\rho/2}$. Putting this back in (3.11) shows that the term goes to 0 as $N\to\infty$ as desired.

3.1.2. *Proof of the lower bound in Proposition 3.2.* The proof of the lower bound is two-step recursion. Two lemmas are needed. The first is a generalization of the lower bound in Daviaud's theorem; see Theorem 1.2 or [17].

LEMMA 3.6. *Let $0<\alpha'<\alpha''\le1$. Suppose that the parameter σ is constant on the strip $[0,1]_\sim\times[\varepsilon^{\alpha''},\varepsilon^{\alpha'}]$, and that the event*

$$\Xi:=\{\#\{x\in\mathcal{X}_{\varepsilon^{\alpha'}}:Y_x(\alpha')\ge\sqrt{2}\gamma'\log N\}\ge N^{\mathcal{E}'}\}$$

is such that

$$\mathbb{P}(\Xi^c)\le\exp\{-c'(\log N)^2\}$$

for some $\gamma'\ge0$, $\mathcal{E}'>0$ and $c'>0$.

Let

$$\mathcal{E}(\gamma) := \mathcal{E}' + (\alpha'' - \alpha') - \frac{(\gamma - \gamma')^2}{\sigma^2(\alpha'' - \alpha')} > 0.$$

Then, for any γ'' *such that* $\mathcal{E}(\gamma'') > 0$ *and any* $\mathcal{E} < \mathcal{E}(\gamma'')$*, there exists* c *such that*

$$\mathbb{P}(\#\{x \in \mathcal{X}_{\varepsilon^{\alpha''}} : Y_x(\alpha'') \geq \sqrt{2}\gamma'' \log N\} \leq N^{\mathcal{E}}) \leq \exp\{-c(\log N)^2\}.$$

We stress that γ'' may be such that $\mathcal{E}(\gamma'') < \mathcal{E}'$. The second lemma, which follows, serves as the starting point of the recursion and is analogous to Lemma 8 in [6].

LEMMA 3.7. *For any* α_0 *such that* $0 < \alpha_0 < \alpha$*, there exists* $\mathcal{E}_0 = \mathcal{E}_0(\alpha_0) > 0$ *and* $c = c(\alpha_0)$ *such that*

$$\mathbb{P}(\#\{x \in \mathcal{X}_{\varepsilon^{\alpha_0}} : Y_x(\alpha_0) \geq 0\} \leq N^{\mathcal{E}_0}) \leq \exp\{-c(\log N)^2\}.$$

We first conclude the proof of the lower bound in Proposition 3.2 using the two above lemmas.

PROOF OF THE LOWER BOUND OF PROPOSITION 3.2. Let γ such that $0 < \gamma < \gamma_{\max}$. Choose $\mathcal{E}$ such that $\mathcal{E} < \mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)$. It will be shown that for some $c > 0$

$$\mathbb{P}(|\mathcal{H}_N^Y(\gamma)| \leq N^{\varepsilon}) \leq \exp\{-c(\log N)^2\}. \tag{3.12}$$

By Lemma 3.7, for $\alpha_0 < \alpha$ arbitrarily close to 0, there exists $\mathcal{E}_0 = \mathcal{E}_0(\alpha_0) > 0$ and $c_0 = c_0(\alpha_0) > 0$, such that

$$\mathbb{P}(\#\{x \in \mathcal{X}_{\varepsilon^{\alpha_0}} : Y_x(\alpha_0) \geq 0\} \leq N^{\mathcal{E}_0}) \leq \exp\{-c_0(\log N)^2\}. \tag{3.13}$$

Observe that we have $0 \leq \mathcal{E}_0 \leq \alpha_0$. Moreover, let

$$\mathcal{E}_1(\gamma_1) := \mathcal{E}_0 + (\alpha - \alpha_0) - \frac{\gamma_1^2}{\sigma_1^2(\alpha - \alpha_0)}. \tag{3.14}$$

Lemma 3.6 is applied from α_0 to α. For any γ_1 with $\mathcal{E}_1(\gamma_1) > 0$ and any $\mathcal{E}_1 < \mathcal{E}_1(\gamma_1)$, there exists $c_1 > 0$ such that

$$\mathbb{P}(\#\{x \in \mathcal{X}_{\varepsilon^{\alpha}} : Y_x(\alpha) \geq \sqrt{2}\gamma_1 \log N\} \leq N^{\mathcal{E}_1}) \leq \exp\{-c_1(\log N)^2\}.$$

Therefore, Lemma 3.6 can be applied from α to 1 for any γ_1 with $\mathcal{E}_1(\gamma_1) > 0$. Define similarly

$$\mathcal{E}_2(\gamma_1, \gamma_2) := \mathcal{E}_1(\gamma_1) + (1 - \alpha) - \frac{(\gamma_2 - \gamma_1)^2}{\sigma_2^2(1 - \alpha)}. \tag{3.15}$$

Then, for any γ_2 with $\mathcal{E}_2(\gamma_1,\gamma_2)>0$ and $\mathcal{E}_2<\mathcal{E}_2(\gamma_1,\gamma_2)$, there exists $c_2>0$ such that

$$\mathbb{P}(\#\{x\in\mathcal{X}_\varepsilon : Y_x\geq\sqrt{2}\gamma_2\log N\}\leq N^{\mathcal{E}_2})\leq\exp\{-c_2(\log N)^2\}. \tag{3.16}$$

Recalling that $0\leq\mathcal{E}_0\leq\alpha_0$, equation (3.12) follows from (3.16) if it is proved that $\lim_{\alpha_0\to0}\mathcal{E}_2(\gamma_1,\gamma)=\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)$ for an appropriate choice of γ_1 [in particular such that $\mathcal{E}_1(\gamma_1)>0$]. It is easily verified that, for a given γ, the quantity $\mathcal{E}_2(\gamma_1,\gamma)$ is maximized at

$$\gamma_1^*=\gamma\frac{\sigma_1^2(\alpha-\alpha_0)}{V_{12}-\sigma_1^2\alpha_0}.$$

Plugging these back in (3.14) shows that $\mathcal{E}_1(\gamma_1^*)>0$ provided that

$$\gamma<\frac{V_{12}}{\sigma_1}=:\gamma_{\mathrm{crit}},$$

with α_0 small enough (depending on γ). Furthermore, since

$$\mathcal{E}_2(\gamma_1^*,\gamma)=\mathcal{E}_0+(1-\alpha_0)-\frac{\gamma^2}{V_{12}-\sigma_1^2\alpha_0},$$

we obtain $\lim_{\alpha_0\to0}\mathcal{E}_2(\gamma_1^*,\gamma)=\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)$, which completes the proof in the case $0<\gamma<\gamma_{\mathrm{crit}}$.

If $\gamma_{\mathrm{crit}}\leq\gamma<\gamma_{\max}$, the condition $\mathcal{E}_1(\gamma_1^*)>0$ will be violated as α_0 goes to zero. In this case, for $\nu>0$, pick $\gamma_1^{**}=\sigma_1\alpha-\nu$ such that $\mathcal{E}_1(\gamma_1^{**})>0$. The first term in γ_1^{**} corresponds to γ_1^* evaluated at γ_{crit} for $\alpha_0=0$. In particular, $\lim_{\alpha_0\to0,\nu\to0}\mathcal{E}_1(\gamma_1^{**})=0$. From (3.15), this shows that

$$\lim_{\alpha_0\to0,\nu\to0}\mathcal{E}_2(\gamma_1^{**},\gamma)=(1-\alpha)-\frac{(\gamma-\sigma_1\alpha)^2}{\sigma_2^2(1-\alpha)}=\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma).$$

Note that $\mathcal{E}^{(\vec{\sigma},\alpha)}(\gamma)$ is strictly positive if and only if $\gamma<\sigma_1\alpha+\sigma_2(1-\alpha)=\gamma_{\max}$. This concludes the proof of (3.12). $\square$

Proof of Lemma 3.6. Let γ'' such that $\mathcal{E}(\gamma'')>0$ and $\mathcal{E}$ such that $0<\mathcal{E}<\mathcal{E}(\gamma'')$. Pick $\overline{\gamma}>\gamma''$ such that

$$\mathcal{E}(\overline{\gamma})>\mathcal{E}>0. \tag{3.17}$$

Since $\overline{\gamma}>\gamma''$, there exists $\varsigma\in(0,1)$ such that

$$\overline{\gamma}(1-\varsigma)\geq\gamma''. \tag{3.18}$$

For $K\in\mathbb{N}$ (which will be fixed later), we set

$$\eta_\ell:=\alpha'+\frac{\ell-1}{K}(\alpha''-\alpha'),\qquad 1\leq\ell\leq K+1,$$

$$\lambda_\ell:=\gamma'+\frac{\ell-1}{K}(\overline{\gamma}-\gamma')(1-\varsigma),\qquad 1\leq\ell\leq K+1.$$

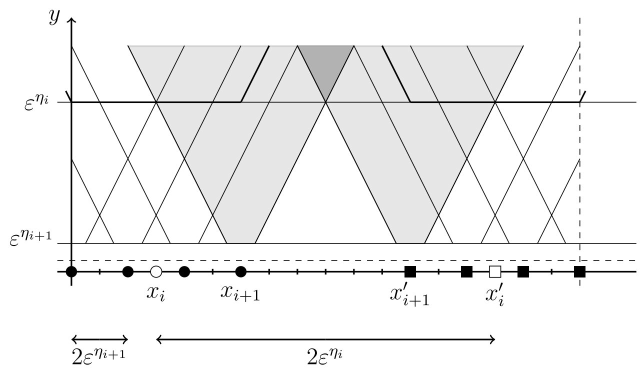

FIG. 4. *Approximation by a tree-like structure. The black circles symbolize the children of the white circle, while the black squares symbolize the children of the white square.*

Observe that the η_ℓ's and the λ_ℓ's satisfy $\eta_1 = \alpha' < \eta_2 < \cdots < \eta_K < \eta_{K+1} = \alpha''$, and $\lambda_1 = \gamma' < \lambda_2 < \cdots < \lambda_K < \lambda_{K+1} = (1-\varsigma)\overline{\gamma} + \varsigma\gamma'$. Consider the sets $\mathcal{A}_\ell$ given by

$$\mathcal{A}_\ell := \{\underline{x}^{(\ell)} = (x_1, \ldots, x_\ell) : x_i \in \mathcal{X}_{2\varepsilon^{\eta_i}}, \forall 1 \le i \le \ell \text{ and } \|x_{i+1} - x_i\| \le \varepsilon^{\eta_i}/2\}$$

for $1 \le \ell \le K+1$. Note that only half of the x_i's in $\mathcal{X}_{\varepsilon^{\eta_i}}$'s are considered. Also, to each x_i we consider the points x_{i+1} in $\mathcal{X}_{2\varepsilon^{\eta_{i+1}}}$ that are close to x_i. By analogy with a branching process, these points can be thought of as the *children* of x_i. The reason for these two choices is that the cones corresponding to the variables $Y_{x_{i+1}}(\eta_{i+1})$ and $Y_{x'_{i+1}}(\eta_{i+1})$ do not intersect below the line $y = \varepsilon^{\eta_i}$ if $x_i \ne x'_i$; see Figure 4.

Now consider, the sets of high points of $\mathcal{A}_\ell$,

$$A_\ell := \{\underline{x}^{(\ell)} \in \mathcal{A}_\ell : Y_{x_i}(\eta_i) \ge \sqrt{2}\lambda_i \log N, \forall 1 \le i \le \ell\}, \qquad 1 \le \ell \le K+1$$

and

$$B_\ell := \{\# A_\ell \ge n_\ell\}, \qquad 1 \le \ell \le K+1,$$

where

$$n_\ell := N^{\mathcal{E}' + (\ell-1)/K((\alpha''-\alpha') - ((\overline{\gamma}-\gamma')^2/(\sigma^2(\alpha''-\alpha'))))}, \qquad 1 \le \ell \le K+1, \tag{3.19}$$

such that $N^{\mathcal{E}'} = n_1$ and $n_{K+1} = N^{\mathcal{E}(\overline{\gamma})}$. Furthermore, with these definitions and the choice of $\overline{\gamma}$ in (3.18) and (3.17), we have for large N

$$\begin{aligned} B_{K+1} &= \{\# A_{K+1} > n_{K+1}\} \\ &\subset \{\#\{x \in \mathcal{X}_{\varepsilon^{\alpha''}} : Y_x(\alpha'') \ge \sqrt{2}((1-\varsigma)\overline{\gamma} + \varsigma\gamma') \log N\} > N^{\mathcal{E}(\overline{\gamma})}\} \\ &\subset \{\#\{x \in \mathcal{X}_{\varepsilon^{\alpha''}} : Y_x(\alpha'') \ge \sqrt{2}\gamma'' \log N\} > N^{\mathcal{E}}\}. \end{aligned}$$

It is thus sufficient to find a bound for $\mathbb{P}(B_{K+1}^c)$ to prove the lemma. For events C_ℓ to be defined in (3.22), we use the elementary bound $\mathbb{P}(B_{K+1}^c) \leq \mathbb{P}(B_{K+1}^c \cap B_K \cap C_K^c) + \mathbb{P}(C_K) + \mathbb{P}(B_K^c)$ which applied recursively gives

$$\mathbb{P}(B_{K+1}^c) \leq \sum_{\ell=2}^{K+1} \left(\mathbb{P}(B_\ell^c \cap B_{\ell-1} \cap C_{\ell-1}^c) + \mathbb{P}(C_{\ell-1})\right) + \mathbb{P}(B_1^c). \tag{3.20}$$

The last term has the correct bound by assumption. It remains to bound the ones appearing in the sum.

On the event B_ℓ, there exist at least n_ℓ high ℓ-branches $\underline{x}^{(\ell)} = (x_1, \dots, x_\ell)$, these are branches that satisfy $Y_{x_i}(\eta_i) \geq \sqrt{2}\lambda_i \log N$ for $1 \leq i \leq \ell$. Select the first n_ℓ such ℓ-branches, and denote them by $\underline{x}_j^{(\ell)} = (x_{j,1}, \dots, x_{j,\ell})$, for all $1 \leq j \leq n_\ell$. Consider the set $\mathcal{A}_{j,\ell}$, the children of $x_{j,\ell}$ at level $\eta_{\ell+1}$: $\mathcal{A}_{j,\ell} := \{x \in \mathcal{X}_{2\varepsilon^{\eta_{\ell+1}}} : \|x - x_{j,\ell}\| \leq \varepsilon^{\eta_\ell}/2\}$. It holds

$$\begin{aligned} B_\ell \cap B_{\ell+1}^c &\subset B_\ell \cap \left\{ \sum_{j=1}^{n_\ell} \sum_{x \in \mathcal{A}_{j,\ell}} \mathbf{1}_{\{Y_x(\eta_{\ell+1}) - Y_{x_{j,\ell}}(\eta_\ell) \geq \sqrt{2}((\overline{\gamma}-\gamma')(1-\varsigma)/K)\log N\}} \leq n_{\ell+1} \right\} \\ &\subset B_\ell \cap \left\{ \sum_{j=1}^{n_\ell} \zeta_j \leq \frac{2n_{\ell+1}}{N^{(\alpha''-\alpha')/K}} \right\}, \end{aligned}$$

where

$$\zeta_j := \frac{1}{|\mathcal{A}_{j,\ell}|} \sum_{x \in \mathcal{A}_{j,\ell}} \mathbf{1}_{\{Y_x(\eta_{\ell+1}) - Y_{x_{j,\ell}}(\eta_\ell) \geq \sqrt{2}((\overline{\gamma}-\gamma')(1-\varsigma)/K)\log N\}}, \tag{3.21}$$

and $|\mathcal{A}_{j,\ell}| = N^{(\alpha''-\alpha')/K}/2$. A crucial point is that $Y_{x_{j,\ell}}(\eta_\ell)$ is not equal to $Y_x(\eta_\ell)$ since $x \neq x_{j,\ell}$ in general. However, it turns out that their value must be very close since the variance of the difference is essentially a constant due to the logarithmic correlations. Precisely, let

$$\begin{aligned} C_\ell := \bigcup_{\underline{x}^{(\ell)} \in \mathcal{A}_\ell} \bigcup_{\substack{x \in \mathcal{X}_{2\varepsilon^{\eta_{\ell+1}}}: \\ \|x - x_\ell\| \leq \varepsilon^{\eta_\ell}/2}} \Big\{ &|Y_{\underline{x}^\ell}(\eta_\ell) - Y_x(\eta_\ell)| \\ &\geq \sqrt{2}\nu \frac{(\overline{\gamma}-\gamma')(1-\varsigma)}{K} \log N \Big\} \end{aligned} \tag{3.22}$$

for $\nu > 0$ which is fixed and will be chosen small later. By Lemma A.4 of the Appendix, $\mathrm{Var}(Y_x(\eta_\ell) - Y_{x'}(\eta_\ell)) \leq \max\{\sigma_1^2, \sigma_2^2\} < \infty$, for every $1 \leq \ell \leq K$, and any $x \in \mathcal{X}_{2\varepsilon^{\eta_\ell}}$, $x' \in \mathcal{X}_{2\varepsilon^{\eta_{\ell+1}}}$ such that $\|x' - x\| \leq \varepsilon^{\eta_\ell}/2$. Therefore, a Gaussian estimate (see Lemma A.1), together with the union-bound give

$$\mathbb{P}(C_\ell) \leq \exp\{-d(\log N)^2\} \tag{3.23}$$

for all $1 \le \ell \le K$ and some $d > 0$.

It remains to bound the first term appearing in the sum of (3.20). On C_ℓ^c, $Y_{x_{j,\ell}}(\eta_\ell)$ can be replaced by $Y_x(\eta_\ell)$ in (3.21), making a small error that depends on ν. Namely, one has $\zeta_j \ge \tilde{\zeta}_j$, where

$$\tilde{\zeta}_j := \frac{1}{|\mathcal{A}_{j,\ell}|} \sum_{x \in \mathcal{A}_{j,\ell}} \mathbf{1}_{\{Y_x(\eta_{\ell+1}) - Y_x(\eta_\ell) \ge \sqrt{2}(1+\nu)((\overline{\gamma}-\gamma')(1-\varsigma)/K)\log N\}}.$$

Note that conditionally on $\mathcal{F}_{\varepsilon^{\eta_\ell}}$, the $\tilde{\zeta}_j$'s are i.i.d. Moreover, since the $\tilde{\zeta}_j$'s are independent of $\mathcal{F}_{\varepsilon^{\eta_\ell}}$, they are also independent of each other. Lemma A.2 of the Appendix guarantees that the sum of the $\tilde{\zeta}_j$ cannot be too low. Observe that

$$\mathbb{E}[\tilde{\zeta}_j] = \mathbb{P}\left(z \ge \sqrt{2}(1+\nu)\frac{(\overline{\gamma}-\gamma')(1-\varsigma)}{K}\log N\right),$$

where z is a centered Gaussian with variance $\sigma^2 \log(\frac{\varepsilon^{\eta_\ell}}{\varepsilon^{\eta_{\ell+1}}}) = \sigma^2 \frac{(\alpha''-\alpha')}{K}\log N$. By a Gaussian estimate, Lemma A.1,

$$\mathbb{E}[\tilde{\zeta}_j] \ge \exp\left\{-\frac{1}{K}\frac{(1+2\nu)^2(\overline{\gamma}-\gamma')^2(1-\varsigma)^2}{\sigma^2(\alpha''-\alpha')}\log N\right\},$$

where $(1+\nu)$ has been replaced by $(1+2\nu)$ to absorb the $1/\sqrt{\log N}$ term in front of the exponential. Consequently, using elementary manipulations,

$$\begin{aligned}
&B_{\ell+1}^c \cap B_\ell \cap C_\ell^c \\
&\quad \subset \left\{\sum_{j=1}^{n_\ell}(\tilde{\zeta}_j - \mathbb{E}[\tilde{\zeta}_j]) \right. \\
&\qquad\qquad \left. \le \frac{2n_{\ell+1}}{N^{(\alpha''-\alpha')/K}} - n_\ell N^{-(1/K)((1+2\nu)^2(\overline{\gamma}-\gamma')^2(1-\varsigma)^2/(\sigma^2(\alpha''-\alpha')))}\right\} \\
&\quad \subset \left\{\left|\sum_{j=1}^{n_\ell}(\tilde{\zeta}_j - \mathbb{E}[\tilde{\zeta}_j])\right| \ge \frac{1}{2}n_\ell N^{-(1/K)((1+2\nu)^2(\overline{\gamma}-\gamma')^2(1-\varsigma)^2/(\sigma^2(\alpha''-\alpha')))}\right\},
\end{aligned}$$

provided

$$\frac{1}{K}\frac{(1+2\nu)^2(\overline{\gamma}-\gamma')^2(1-\varsigma)^2}{\sigma^2(\alpha''-\alpha')} < \frac{1}{K}\frac{(\overline{\gamma}-\gamma')^2}{\sigma^2(\alpha''-\alpha')},$$

that is

$$(1+2\nu)(1-\varsigma) < 1. \tag{3.24}$$

Fix ν small enough such that (3.24) is satisfied. Write for short

$$\mu := \frac{1}{K}\frac{(1+2\nu)^2(\overline{\gamma}-\gamma')^2(1-\varsigma)^2}{\sigma^2(\alpha''-\alpha')}.$$

Then, taking $n = n_\ell$ and $t = n_\ell N^{-\mu}$ in Lemma A.2, we get

$$\mathbb{P}(B_{\ell+1}^c \cap B_\ell \cap C_\ell^c) \le 2 \exp\left\{ \frac{n_\ell^2 N^{-2\mu}}{2n_\ell + (2/3) n_\ell N^{-\mu}} \right\}.$$

By the form of n_ℓ in (3.19), K can be taken large enough so that $n_\ell N^{-2\mu} > N^\delta$ for some $\delta > 0$ and all $\ell = 1, \ldots, K+1$. This completes the proof of the lemma. $\square$

Proof of Lemma 3.7. Take $\alpha' < \alpha_0$ in such a way that $\mathcal{X}_{\varepsilon^{\alpha'}} \subset \mathcal{X}_{\varepsilon^{\alpha_0}}$. Consider the set

$$\Lambda := \{x \in \mathcal{X}_{\varepsilon^{\alpha'}} : Y_x(\alpha') \ge -\sigma_1(\alpha_0 - \alpha') \log N\},$$

and the event

$$A = A_\delta := \{|\Lambda| \ge N^\delta\}, \qquad \delta > 0.$$

The parameters $\mathcal{E}_0$, δ and α' will be chosen later as a function of α_0. By splitting the probability on the event A,

$$\begin{aligned}
&\mathbb{P}(\#\{x \in \mathcal{X}_{\varepsilon^{\alpha_0}} : Y_x(\alpha_0) \ge 0\} \le N^{\mathcal{E}_0}) \\
&\quad \le \mathbb{P}(\#\{x \in \mathcal{X}_{\varepsilon^{\alpha_0}} : Y_x(\alpha_0) \ge 0\} \le N^{\mathcal{E}_0}; A) + \mathbb{P}(A^c) \\
&\quad \le \mathbb{E}[\mathbb{P}(\#\{x \in \Lambda : Y_x(\alpha_0) - Y_x(\alpha') \ge \sigma_1(\alpha_0 - \alpha') \log N\} \le N^{\mathcal{E}_0} | \mathcal{F}_{\varepsilon^{\alpha'}}); A] \\
&\qquad + \mathbb{P}(A^c),
\end{aligned}$$

where the second inequality is obtained by restricting to the set $\Lambda \subset \mathcal{X}_{\varepsilon^{\alpha_0}}$.

First we prove that the definition of A yields a super-exponential decay of the first term for $\mathcal{E}_0$ and δ depending on $\alpha_0 - \alpha'$. The variables $Y_x(\alpha_0) - Y_x(\alpha')$, $x \in \mathcal{X}_{\varepsilon^{\alpha'}}$, are i.i.d. Gaussians of variance $\sigma_1^2(\alpha_0 - \alpha') \log N$. Write for simplicity $(z_i, i = 1, \ldots, N^\delta)$ for i.i.d. Gaussians random variables with variance $\sigma_1^2(\alpha_0 - \alpha') \log N$. A Gaussian estimate (see Lemma A.1) implies

$$\mathbb{P}(z_i \ge \sigma_1(\alpha_0 - \alpha') \log N) \ge \frac{1}{2} \frac{\mathrm{e}^{-(1/2)(\alpha_0 - \alpha') \log N}}{\sqrt{(\alpha_0 - \alpha') \log N}} \ge \mathrm{e}^{-(2/3)(\alpha_0 - \alpha') \log N}.$$

Therefore

$$\begin{aligned}
&\mathbb{E}[\mathbb{P}(\#\{x \in \Lambda : Y_x(\alpha_0) - Y_x(\alpha') \ge \sigma_1(\alpha_0 - \alpha') \log N\} \le N^{\mathcal{E}_0} | \mathcal{F}_{\varepsilon^{\alpha'}}); A] \\
&\quad \le \mathbb{P}\left(\sum_{i=1}^{N^\delta} (\mathbf{1}_{\{z_i \ge \sigma_1(\alpha_0 - \alpha') \log N\}} - \mathbb{P}(z_i \ge \sigma_1(\alpha_0 - \alpha') \log N)) \right. \\
&\qquad\qquad \left. \le N^{\mathcal{E}_0} - N^{\delta - (2/3)(\alpha_0 - \alpha')} \right).
\end{aligned}$$

Lemma A.2 in the Appendix gives a super-exponential decay of the above probability for the choice $\delta > \frac{4}{3}(\alpha_0 - \alpha')$ and $\mathcal{E}_0 - \delta + \frac{2}{3}(\alpha_0 - \alpha') < 0$, for example, $\delta = 2(\alpha_0 - \alpha')$ and $\mathcal{E}_0 = \alpha_0 - \alpha'$.

It remains to show that $\mathbb{P}(A^c)$ has super-exponential decay. We have

$$\mathbb{P}(A^c) \le P\Big(A^c, \max_{x \in \mathcal{X}_{\varepsilon\alpha'}} Y_x(\alpha') \le (\log N)^2\Big) + P\Big(\max_{x \in \mathcal{X}_{\varepsilon\alpha'}} Y_x(\alpha') > (\log N)^2\Big).$$

The second term is easily shown to have the desired decay. We focus on the first. On the event $A^c \cap \{\max_{x \in \mathcal{X}_{\alpha'}} Y_x(\alpha') \le (\log N)^2\}$,

$$\begin{aligned} &\frac{1}{|\mathcal{X}_{\varepsilon\alpha'}|} \sum_{x \in \mathcal{X}_{\varepsilon\alpha'}} \omega_{\alpha'}(x) \\ &\quad = \frac{1}{|\mathcal{X}_{\varepsilon\alpha'}|} \sum_{x \in \Lambda} \omega_{\alpha'}(x) + \frac{1}{|\mathcal{X}_{\varepsilon\alpha'}|} \sum_{x \in \Lambda^c} \omega_{\alpha'}(x) \qquad (3.25) \\ &\quad \le \frac{|\Lambda|}{|\mathcal{X}_{\varepsilon\alpha'}|} (\log N)^2 + \Big(1 - \frac{|\Lambda|}{|\mathcal{X}_{\varepsilon\alpha'}|}\Big)(-\sigma_1(\alpha_0 - \alpha') \log N). \end{aligned}$$

Since $|\mathcal{X}_{\varepsilon\alpha'}| = N^{\alpha'}$, it is easily checked that for $\delta = 2(\alpha_0 - \alpha') < \alpha'$, the above is smaller than $-\frac{2}{3}\sigma_1(\alpha_0 - \alpha') \log N$. Therefore we choose α' such that $\alpha_0 < 3\alpha'/2$. Finally the left-hand side of (3.25) is a Gaussian random variable, whose variance is of order 1. Therefore the probability that it is smaller than $-\frac{2}{3}\sigma_1(\alpha_0 - \alpha') \log N$ is super-exponentially small. This completes the proof of the lemma. $\square$

4. The free energy from the high points: Proof of Proposition 2.1. In this section, we compute the free energy of the perturbed models introduced in Section 2.1. The free energy $f_N^{(\vec{\sigma},\alpha)}(\beta)$ is shown to converge in probability to the claimed expression. The L^1-convergence then follows from the fact that the variables $(f_N^{(\vec{\sigma},\alpha)}(\beta))_{N \ge 1}$ are uniformly integrable. This is a consequence of Borell-TIS inequality. (Another more specific approach used by Capocaccia, Cassandro and Picco [14] for the GREM models could also have been applied here; see Section 3.1 in [14]. Indeed, we clearly have

$$\beta \frac{\max_{x \in \mathcal{X}_N} Y_x}{\log N} \le f_N^{(\vec{\sigma},\alpha)}(\beta) \le 1 + \beta \frac{\max_{x \in \mathcal{X}_N} Y_x}{\log N}.$$

Therefore, uniform integrability follows if it is proved that $\frac{1}{(\log N)^2} \times \mathbb{E}[(\max_{x \in \mathcal{X}_N} Y_x)^2]$ is uniformly bounded. It equals

$$\frac{1}{(\log N)^2} \mathbb{E}\Big[\Big(\max_{x \in \mathcal{X}_N} Y_x - \mathbb{E}\Big[\max_{x \in \mathcal{X}_N} Y_x\Big]\Big)^2\Big] + \frac{1}{(\log N)^2} \mathbb{E}\Big[\max_{x \in \mathcal{X}_N} Y_x\Big]^2.$$

The first term is bounded by the Borell-TIS inequality (see [1], page 50)

$$\mathbb{P}\Big(\Big|\max_{x\in\mathcal{X}_N} Y_x - \mathbb{E}\max_{x\in\mathcal{X}_N} Y_x\Big| > r\Big) \le 2\mathrm{e}^{-r^2/(2V_{12}\log N)} \qquad \forall r>0,$$

which gives

$$\mathbb{E}\bigg[\bigg(\frac{\max_{x\in\mathcal{X}_N} Y_x - \mathbb{E}[\max_{x\in\mathcal{X}_N} Y_x]}{\log N}\bigg)^2\bigg] \le 4\int_0^\infty r\mathrm{e}^{-r^2/(2V_{12})\log N}\,\mathrm{d}r.$$

The right-hand side goes to zero for $N\to\infty$. The term $\frac{1}{\log N}\mathbb{E}[\max_{x\in\mathcal{X}_N} Y_x]$ can be bounded uniformly by comparing with i.i.d. centered Gaussian random variables of variance $V_{12}\log N$ and using Slepian's inequality; see, for example, [1], page 57. Equivalently, one can reason as follows. It is easily checked that the probability that the maximum be negative decreases exponentially with N. Thus to control the second term it suffices to control

$$\frac{1}{\log N}\int_0^\infty \mathbb{P}\Big(\max_{x\in\mathcal{X}_N} Y_x > r\Big)\,\mathrm{d}r.$$

It suffices to split the integral in two intervals: $[0,\sqrt{2V_{12}}\log N)$ and $[\sqrt{2V_{12}}\log N,+\infty)$. The first integral divided by $\log N$ is evidently of order 1. The second integral divided by $\log N$ tends to 0 by a union bound and a Gaussian estimate. The almost-sure convergence is straightforward from the L^1-convergence and the almost-sure self-averaging property of the free energy

$$\lim_{N\to\infty}|f_N^{(\vec{\sigma},\alpha)}(\beta) - \mathbb{E}f_N^{(\vec{\sigma},\alpha)}(\beta)| = 0 \qquad \text{a.s.}$$

This is a standard consequence of concentration of measure (see [33], page 32) since the free energy is a Lipschitz function of i.i.d. Gaussian variables of Lipschitz constant smaller than $\beta/\sqrt{\log N}$. (Note that the Y_x's can be written as a linear combination of i.i.d. standard Gaussians with coefficients chosen to get the correct covariances.)

It remains to prove that the free energy $f_N^{(\vec{\sigma},\alpha)}(\beta)$ converges in probability to the claimed expression in Proposition 2.1. For fixed $\beta>0$ and $\nu>0$, we prove that

$$\lim_{N\to\infty}\mathbb{P}(f_N^{(\vec{\sigma},\alpha)}(\beta) \le f^{(\vec{\sigma},\alpha)}(\beta) - \nu) = 0, \tag{4.1}$$

$$\lim_{N\to\infty}\mathbb{P}(f_N^{(\vec{\sigma},\alpha)}(\beta) \ge f^{(\vec{\sigma},\alpha)}(\beta) + \nu) = 0. \tag{4.2}$$

First, we introduce some notation and give a preliminary result. For simplicity, we will write $\mathcal{E}$ for $\mathcal{E}^{(\vec{\sigma},\alpha)}$ throughout the proof. For any $M\in\mathbb{N}$, consider the partition of $[0,\gamma_{\max}]$ into M intervals $[\gamma_{i-1},\gamma_i[$, where the γ_i's are given by

$$\gamma_i := \frac{i}{M}\gamma_{\max}, \qquad i=0,1,\dots,M.$$

Moreover for any $N \geq 2$, any $M \in \mathbb{N}$ and any $\delta > 0$, define the random variable

$$K_{N,M}(i) := \#\left\{x \in \mathcal{X}_N : \frac{Y_x}{\sqrt{2}\log N} \in [\gamma_{i-1}, \gamma_i[\right\}, \qquad 1 \leq i \leq M,$$

and the events

$$B_{N,M,\delta} := \bigcap_{i=1}^{M} \{N^{\mathcal{E}(\gamma_{i-1})-\delta} - N^{\mathcal{E}(\gamma_i)+\delta} \leq K_{N,M}(i) \leq N^{\mathcal{E}(\gamma_{i-1})+\delta} - N^{\mathcal{E}(\gamma_i)-\delta}\}$$
$$\cap \{\#\{x \in \mathcal{X}_N : Y_x \geq \sqrt{2}\gamma_{\max} \log N\} = 0\}.$$

The next result is a straightforward consequence of Propositions 3.1 and 3.2.

LEMMA 4.1. *For any $M \in \mathbb{N}$ and any $\delta > 0$, we have*

$$\lim_{N\to\infty} \mathbb{P}(B_{N,M,\delta}) = 1.$$

Define the continuous function

$$P_\beta(\gamma) := \mathcal{E}(\gamma) + \sqrt{2}\beta\gamma \qquad \forall \gamma \in [0, \gamma_{\max}].$$

Using the expression of $\mathcal{E}$ in Proposition 3.2 on the different intervals, it is easily checked by differentiation that

$$\max_{\gamma \in [0,\gamma_{\max}]} P_\beta(\gamma) = f^{(\vec{\sigma},\alpha)}(\beta). \tag{4.3}$$

Furthermore, the continuity of $\gamma \mapsto P_\beta(\gamma)$ on $[0, \gamma_{\max}]$ yields

$$\max_{0 \leq i \leq M-1} P_\beta(\gamma_i) \longrightarrow \max_{\gamma \in [0,\gamma_{\max}]} P_\beta(\gamma) = f^{(\vec{\sigma},\alpha)}(\beta), \qquad M \to \infty.$$

Fix $M \in \mathbb{N}$ large enough and $\delta > 0$ small enough, such that

$$\max_{0 \leq i \leq M-1} P_\beta(\gamma_i) \geq f^{(\vec{\sigma},\vec{\alpha})}(\beta) - \frac{\nu}{3}, \tag{4.4}$$

$$\frac{\sqrt{2}\beta}{M} < \frac{\nu}{3}, \tag{4.5}$$

$$\delta < \min\left\{-\frac{1}{2}\max_{1 \leq i \leq M}\{\mathcal{E}(\gamma_i) - \mathcal{E}(\gamma_{i-1})\}, \frac{\nu}{3}, \sqrt{2}\gamma_1\beta\right\}. \tag{4.6}$$

Note that for fixed M, $\max_{1 \leq i \leq M}\{\mathcal{E}(\gamma_i) - \mathcal{E}(\gamma_{i-1})\} < 0$ since $\gamma \mapsto \mathcal{E}(\gamma)$ is a decreasing function on $[0, \gamma_{\max}]$.

Proof of the lower bound (4.1). Observe that the partition function $Z_N^{(\vec{\sigma},\alpha)}(\beta)$ associated with the perturbed model satisfies $Z_N^{(\vec{\sigma},\alpha)}(\beta) \geq \sum_{i=1}^{M} K_{N,M}(i) N^{\sqrt{2}\gamma_{i-1}\beta}$. Therefore on $B_{N,M,\delta}$ we get

$$Z_N^{(\vec{\sigma},\alpha)}(\beta) \geq \sum_{i=1}^{M} (1 - N^{\mathcal{E}(\gamma_i)-\mathcal{E}(\gamma_{i-1})+2\delta}) N^{P_\beta(\gamma_{i-1})-\delta}.$$

This yields on $B_{N,M,\delta}$

$$f_N^{(\vec{\sigma},\alpha)}(\beta) \geq \frac{\log(1 - N^{\max_{1\leq i\leq M}\{\mathcal{E}(\gamma_i)-\mathcal{E}(\gamma_{i-1})\}+2\delta})}{\log N} + \max_{0\leq i\leq M-1} P_\beta(\gamma_i) - \delta.$$

Since for δ in (4.6)

$$\lim_{N\to\infty} (\log N)^{-1} \log(1 - N^{\max_{1\leq i\leq M}\{\mathcal{E}(\gamma_i)-\mathcal{E}(\gamma_{i-1})\}+2\delta}) = 0,$$

the choices of M, δ in (4.4) and (4.6) give that $f_N^{(\vec{\sigma},\alpha)}(\beta) - f^{(\vec{\sigma},\alpha)}(\beta) > -\nu$ on $B_{N,M,\delta}$ for N large enough. Therefore, (4.1) is a consequence of Lemma 4.1.

Proof of the upper bound (4.2). Observe first that the partition function $Z_N^{(\vec{\sigma},\alpha)}(\beta)$ satisfies on $B_{N,M,\delta}$

$$Z_N^{(\vec{\sigma},\alpha)}(\beta) \leq \sum_{i=1}^{M} K_{N,M}(i) N^{\sqrt{2}\gamma_i\beta} + N,$$

the second term coming from the negative values of the field. Since $\mathcal{E}(0) = 1$, on $B_{N,M,\delta}$ and for N large enough, we have using (4.6)

$$K_{N,M}(1) \geq N^{1-\delta} - N^{\mathcal{E}(\gamma_1)+\delta} \geq \tfrac{1}{2} N^{1-\delta},$$

thus $N \leq 2K_{N,M}(1)N^\delta$. Moreover, on $B_{N,M,\delta}$ the random variable $K_{N,M}(i)$ are less than $N^{\mathcal{E}(\gamma_{i-1})+\delta}$ for all $1 \leq i \leq M$. The two last observations imply by the choice of δ

$$Z_N^{(\vec{\sigma},\alpha)}(\beta) \leq \sum_{i=1}^{M} K_{N,M}(i) N^{\sqrt{2}\gamma_i\beta} + 2K_{N,M}(1)N^\delta \leq 3\sum_{i=1}^{M} N^{\mathcal{E}(\gamma_{i-1})+\sqrt{2}\gamma_i\beta+\delta}.$$

Therefore, on the event $B_{N,M,\delta}$, we get

$$f_N^{(\vec{\sigma},\alpha)}(\beta) \leq \frac{\log(3M)}{\log N} + \max_{\gamma\in[0,\gamma_{\max}]} P_\beta(\gamma) + \frac{\sqrt{2}\beta}{M} + \delta.$$

Recalling (4.3) and since $\lim_{N\to\infty}(\log N)^{-1}\log(2M) = 0$, the choices of M and δ in (4.5) and (4.6) imply that $f_N^{(\vec{\sigma},\alpha)}(\beta) - f^{(\vec{\sigma},\alpha)}(\beta) < \nu$ on $B_{N,M,\delta}$ for N large enough. Therefore (4.2) is a consequence of Lemma 4.1.

APPENDIX

A.1. Gaussian estimates, large deviation result and integration by part.

LEMMA A.1 (see, e.g., [22]). *Let X be a standard Gaussian random variable. For any $a > 0$, we have*

$$\frac{(1-2a^{-2})}{\sqrt{2\pi}a} \mathrm{e}^{-a^2/2} \leq \mathbb{P}(X \geq a) \leq \frac{1}{\sqrt{2\pi}a} \mathrm{e}^{-a^2/2}.$$

LEMMA A.2 (see, e.g., [5]). *Let $Z_1,\dots,Z_n$ be i.i.d. real valued random variables satisfying $\mathbb{E}[Z_i]=0$, $\sigma^2=\mathbb{E}[Z_i^2]$ and $\|Z_i\|_\infty\le 1$. Then for any $t>0$,*

$$\mathbb{P}\left(\left|\sum_{i=1}^n Z_i\right|\ge t\right)\le 2\exp\left\{-\frac{t^2}{2n\sigma^2+2t/3}\right\}.$$

LEMMA A.3 (see, e.g., the Appendix of [33]). *Let $(X,Z_1,\dots,Z_d)$ be a centered Gaussian random vector. Then, for any C^1 function $F:\mathbb{R}^d\mapsto\mathbb{R}$, of moderate growth at infinity, we have*

$$\mathbb{E}[XF(Z_1,\dots,Z_d)]=\sum_{i=1}^d\mathbb{E}[XZ_i]\mathbb{E}\left[\frac{\partial F}{\partial z_i}(Z_1,\dots,Z_d)\right].$$

A.2. Proof of Lemma 2.2. Recall that $0<\varepsilon=1/N<1/2$, and $\alpha\in(0,1)$. Also by definition, $\|x'-x\|=\varepsilon^{q(x,x')}$.

It is clear that $\mathbb{E}[\tilde{X}_xX_x]=E[(\tilde{X}_x)^2]$, which is the variance of the centered Gaussian random variable $\mu(A_\varepsilon(x)\setminus A_{\varepsilon^\alpha}(x))$. This variance can be computed and equals

$$\int_\varepsilon^{\varepsilon^\alpha}y^{-1}\,\mathrm{d}y=[\log y]_\varepsilon^{\varepsilon^\alpha}=(1-\alpha)\log N.$$

For the covariance, observe that $\mathbb{E}[\tilde{X}_xX_{x'}]$ is equal to the variance of the random variable $\mu((A_\varepsilon(x)\setminus A_{\varepsilon^\alpha}(x))\cap A_\varepsilon(x'))$. If $\varepsilon<\ell=\|x'-x\|<\varepsilon^\alpha$ [i.e., $\alpha<q(x,x')\le 1$], then the subsets intersect in between the lines $y=\varepsilon$ and $y=\varepsilon^\alpha$, thus

$$\mathbb{E}[\tilde{X}_xX_{x'}]=\int_\ell^{\varepsilon^\alpha}\frac{y-\ell}{y^2}\,\mathrm{d}y=[\log y]_\ell^{\varepsilon^\alpha}+\ell\left[\frac{1}{y}\right]_\ell^{\varepsilon^\alpha}=(q(x,x')-\alpha)\log N+O_N(1).$$

Finally, if $\ell=\|x'-x\|\ge\varepsilon^\alpha$ [i.e., $0\le q(x,x')\le\alpha$], then the set $(A_\varepsilon(x)\setminus A_{\varepsilon^\alpha}(x))\cap A_\varepsilon(x')$ is empty and thus $\mathbb{E}[\tilde{X}_xX_{x'}]=0$.

A.3. A key property of the perturbed models. The following lemma is a key tool to approximate the Gaussian field we consider by a tree. Indeed the difference between the contribution to the Gaussian field at a certain scale for two points that are close can be explicitly computed by integrating parallelograms (see Figure 5 below) and is shown to be small.

LEMMA A.4. *Fix α',α'' as in Lemma 3.6, u such that $\alpha'<u<\alpha''$ and $\delta\in(0,1)$. Then for all $x,x'\in\mathcal{X}_\varepsilon$ such that $\|x-x'\|\le\delta\varepsilon^u$, we have*

$$\mathrm{Var}(Y_x(u)-Y_{x'}(u))\le 2\overline{\sigma}^2\delta,$$

where $\overline{\sigma}$ denotes an upper bound for the σ_i's.

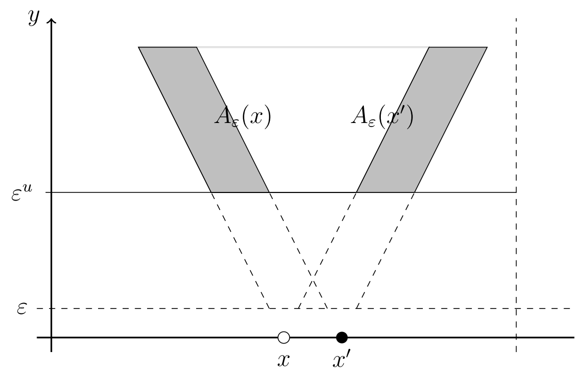

FIG. 5. *The error terms in the tree approximation correspond to the two grey parallelograms in Lemma A.4.*

PROOF. Writing $A := A_{\varepsilon^u}(x)\Delta A_{\varepsilon^u}(x')$, we have

$$\begin{aligned}\operatorname{Var}(Y_x(u) - Y_{x'}(u)) &\le \overline{\sigma}^2 \int_A y^{-2}\,\mathrm{d}s\,\mathrm{d}y = 2\overline{\sigma}^2\|x-x'\| \int_{\varepsilon^u}^{\infty} y^{-2}\,\mathrm{d}y \\ &= 2\overline{\sigma}^2 \frac{\|x-x'\|}{\varepsilon^u} \le 2\overline{\sigma}^2\delta,\end{aligned}$$

which completes the proof of the lemma. □

Acknowledgments. The authors thank Yan Fyodorov, Nicola Kistler, Irina Kurkova and Vincent Vargas for helpful discussions. O. Zindy would like to thank the Courant Institute of Mathematical Science and the Université de Montréal for hospitality and financial support.

REFERENCES

[1] ADLER, R. J. and TAYLOR, J. E. (2007). *Random Fields and Geometry*. Springer, New York. MR2319516

[2] ARGUIN, L.-P. and CHATTERJEE, S. (2013). Random overlap structures: Properties and applications to spin glasses. *Probab. Theory Related Fields* **156** 375–413. MR3055263

[3] BACRY, E. and MUZY, J. F. (2003). Log-infinitely divisible multifractal processes. *Comm. Math. Phys.* **236** 449–475. MR2021198

[4] BARRAL, J. and MANDELBROT, B. B. (2002). Multifractal products of cylindrical pulses. *Probab. Theory Related Fields* **124** 409–430. MR1939653

[5] BENNETT, G. (1962). Probability inequalities for the sum of independent random variables. *J. Amer. Statist. Assoc.* **57** 33–45.

[6] BOLTHAUSEN, E., DEUSCHEL, J.-D. and GIACOMIN, G. (2001). Entropic repulsion and the maximum of the two-dimensional harmonic crystal. *Ann. Probab.* **29** 1670–1692. MR1880237

[7] Bolthausen, E., Deuschel, J. D. and Zeitouni, O. (2011). Recursions and tightness for the maximum of the discrete, two dimensional Gaussian free field. *Electron. Commun. Probab.* **16** 114–119. MR2772390

[8] Bolthausen, E. and Kistler, N. (2006). On a nonhierarchical version of the generalized random energy model. *Ann. Appl. Probab.* **16** 1–14. MR2209333

[9] Bolthausen, E. and Kistler, N. (2009). On a nonhierarchical version of the generalized random energy model. II. Ultrametricity. *Stochastic Process. Appl.* **119** 2357–2386. MR2531095

[10] Bolthausen, E. and Sznitman, A.-S. (2002). *Ten Lectures on Random Media*. *DMV Seminar* **32**. Birkhäuser, Basel. MR1890289

[11] Bovier, A. and Kurkova, I. (2004). Derrida's generalised random energy models. I. Models with finitely many hierarchies. *Ann. Inst. Henri Poincaré Probab. Stat.* **40** 439–480. MR2070334

[12] Bovier, A. and Kurkova, I. (2004). Derrida's generalized random energy models. II. Models with continuous hierarchies. *Ann. Inst. Henri Poincaré Probab. Stat.* **40** 481–495. MR2070335

[13] Bramson, M. and Zeitouni, O. (2012). Tightness of the recentered maximum of the two-dimensional discrete Gaussian free field. *Comm. Pure Appl. Math.* **65** 1–20. MR2846636

[14] Capocaccia, D., Cassandro, M. and Picco, P. (1987). On the existence of thermodynamics for the generalized random energy model. *J. Stat. Phys.* **46** 493–505. MR0883541

[15] Carpentier, D. and Le Doussal, P. (2001). Glass transition for a particle in a random potential, front selection in nonlinear renormalization group, and entropic phenomena in Liouville and Sinh-Gordon models. *Phys. Rev. E (3)* **63** 026110.

[16] Chauvin, B. and Rouault, A. (1997). Boltzmann–Gibbs weights in the branching random walk. In *Classical and Modern Branching Processes (Minneapolis, MN, 1994)* (K. B. Athreya and P. Jagers, eds.). *IMA Vol. Math. Appl.* **84** 41–50. Springer, New York. MR1601693

[17] Daviaud, O. (2006). Extremes of the discrete two-dimensional Gaussian free field. *Ann. Probab.* **34** 962–986. MR2243875

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

[19] Derrida, B. and Spohn, H. (1988). Polymers on disordered trees, spin glasses, and traveling waves. *J. Stat. Phys.* **51** 817–840. MR0971033

[20] Ding, J. (2013). Exponential and double exponential tails for maximum of two-dimensional discrete Gaussian free field. *Probab. Theory Related Fields* **157** 285–299. MR3101848

[21] Dovbysh, L. N. and Sudakov, V. N. (1982). Gram–de Finetti matrices. *J. Soviet. Math.* **24** 3047–3054.

[22] Durrett, R. (2004). *Probability: Theory and Examples*, 3rd ed. Duxbury Press, Belmont, CA.

[23] Fang, M. and Zeitouni, O. (2012). Branching random walks in time inhomogeneous environments. *Electron. J. Probab.* **17** 1–18. MR2968674

[24] Fyodorov, Y. V. and Bouchaud, J.-P. (2008). Freezing and extreme-value statistics in a random energy model with logarithmically correlated potential. *J. Phys. A* **41** 372001, 12. MR2430565

[25] Fyodorov, Y. V., Le Doussal, P. and Rosso, A. (2009). Statistical mechanics of logarithmic REM: Duality, freezing and extreme value statistics of $1/f$ noises generated by Gaussian free fields. *J. Stat. Mech.* **2009** P10005.

[26] Ghirlanda, S. and Guerra, F. (1998). General properties of overlap probability distributions in disordered spin systems. Towards Parisi ultrametricity. *J. Phys. A* **31** 9149–9155. MR1662161

[27] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces: Isoperimetry and Processes*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete (3)* **23**. Springer, Berlin. MR1102015

[28] Panchenko, D. (2010). A connection between the Ghirlanda–Guerra identities and ultrametricity. *Ann. Probab.* **38** 327–347. MR2599202

[29] Panchenko, D. (2010). The Ghirlanda–Guerra identities for mixed p-spin model. *C. R. Math. Acad. Sci. Paris* **348** 189–192. MR2600075

[30] Robert, R. and Vargas, V. (2010). Gaussian multiplicative chaos revisited. *Ann. Probab.* **38** 605–631. MR2642887

[31] Ruelle, D. (1987). A mathematical reformulation of Derrida's REM and GREM. *Comm. Math. Phys.* **108** 225–239. MR0875300

[32] Simon, B. (1993). *The Statistical Mechanics of Lattice Gases. Vol. I*. Princeton Univ. Press, Princeton, NJ. MR1239893

[33] Talagrand, M. (2003). *Spin Glasses: A Challenge for Mathematicians: Cavity and Mean Field Models*. *Ergebnisse der Mathematik und Ihrer Grenzgebiete. 3. Folge.* **46**. Springer, Berlin. MR1993891

[34] Talagrand, M. (2010). Construction of pure states in mean field models for spin glasses. *Probab. Theory Related Fields* **148** 601–643. MR2678900

Département de Mathématiques
et Statistique
Université de Montréal
Montréal, Quebec H3T 1J4
Canada
E-mail: arguinlp@dms.umontreal.ca

Laboratoire de Probabilités
et Modèles Aléatoires
CNRS UMR 7599
Université Paris 6
4 place Jussieu
75252 Paris Cedex 05
France
E-mail: olivier.zindy@upmc.fr