Poisson convergence in the restricted k-partioning problem[1]

Anton Bovier[2]

Weierstraß–Institut
für Angewandte Analysis und Stochastik
Mohrenstrasse 39, D-10117 Berlin, Germany
and
Institut für Mathematik
Technische Universität Berlin
Strasse des 17. Juni 136
12623 Berlin, Germany

Irina Kurkova[3]

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

Abstract: The randomized k-number partitioning problem is the task to distribute N i.i.d. random variables into k groups in such a way that the sums of the variables in each group are as similar as possible. The restricted k-partitioning problem refers to the case where the number of elements in each group is fixed to N/k. In the case $k = 2$ it has been shown that the properly rescaled differences of the two sums in the close to optimal partitions converge to a Poisson point process, as if they were independent random variables. We generalize this result to the case $k > 2$ in the restricted problem and show that the vector of differences between the k sums converges to a $k-1$-dimensional Poisson point process.

Keywords: Number partioning, extreme values, Poisson process, Random Energy Model

AMS Subject Classification: 90C27, 60G70

---

[1]Work supported in part by the DFG research center matheon and the European Science Foundation in the programme RDSES

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

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

1. Introduction.

The number partitioning problem is a classical problem from combinatorial optimization. One considers N numbers $x_1, \ldots, x_N$ and one seeks to partition the set $\{1, \ldots, N\}$ into k disjoint subsets $I_1, \ldots, I_k$, such that the sums $K_\beta \equiv K_\beta(I_1, \ldots, I_k) \equiv \sum_{n \in I_\beta} x_n$ are as similar to each other as possible. This problem can be cast into the language of mean field spin systems [Mer1,Mer2,BFM] by realizing that the set of partitions is equivalent to the set of Potts spin variables $\sigma : \{1, \ldots, N\} \to \{1, \ldots, k\}^N$. We then define the variables

$$K_\beta(\sigma) \equiv \sum_{n=1}^{N} x_n \mathbb{1}_{\sigma_n = \beta}, \quad \beta = 1, \ldots, k. \tag{1.1}$$

One may introduce a "Hamiltonian" as [Mer1,BFM]

$$H_N(\sigma) \equiv \sum_{\beta=1}^{k-1} |K_\beta(\sigma) - K_{\beta+1}(\sigma)| \tag{1.2}$$

and study the minimization problem of this Hamiltonian. In particular, if the numbers x_i are considered as random variables, the problem transforms into the study of a random mean field spin model. For a detailed discussion we refer to the recent paper [BFM].

Mertens [Mer1,Mer2] has argued that the problem is close to the so-called *Random Energy Model (REM)*, i.e. that the random variables $K_\beta(\sigma)$ can effectively be considered as independent random variables for different realizations of σ, at least as far as their extremal properties are concerned. This claim was proven rigorously in a paper by Borgs et al.[BCP] in the case $k = 2$ (see also [BCMP]).

In this paper we extend this result to the case of arbitrary k and under the additional constraint that the cardinalities of the sets I_j are all equal. We formulate this result in the language of multi-dimensional extremal process.

Let $X_1, \ldots, X_N$ be independent uniformly distributed on [0,1] random variables. (We assume that N is always a multiple of k.) Consider the state space of configurations σ of N spins, where each spin takes k possible values $\sigma = (\sigma_1, \ldots, \sigma_N) \in \{1, \ldots, k\}^N$. We will restrict ourselves to configurations such that the number of spins taking each value equals N/k, i.e. $\#\{n : \sigma_n = \beta\} = N/k$ for all $\beta = 1, \ldots, k$. Finally, we must take equivalence classes of these configurations: each class includes $k!$ configurations obtained by a permutation of the values of spins $1, \ldots, k$. We denote by Σ_N the state space of these equivalence classes. Then

$$|\Sigma_N| = \binom{N}{N/k}\binom{N(1-1/k)}{N/k} \cdots \binom{2N/k}{N/k}(k!)^{-1} \sim k^N (2\pi N)^{\frac{1-k}{2}} k^{\frac{k}{2}} (k!)^{-1} \equiv S(k, N). \tag{1.3}$$

Each configuration $\sigma \in \Sigma_N$ corresponds to a partition of $X_1, \ldots, X_N$ into k subsets of N/k random variables, each subset being $\{X_n : \sigma_n = \beta\}$, $\beta = 1, \ldots, k$. Then the vector $\vec{Y}(\sigma) = \{Y^\beta(\sigma)\}_{\beta=1}^{k-1}$ with the coordinates

$$Y^\beta(\sigma) = K_\beta(\sigma) - K_{\beta+1}(\sigma) = \sum_{n=1}^{N} X_n(\mathbb{1}_{\{\sigma_n=\beta\}} - \mathbb{1}_{\{\sigma_n=\beta+1\}}), \quad \beta = 1, \ldots, k-1, \tag{1.4}$$

measures the differences of the sums over the subsets. Our objective is to minimize its norm as most as possible. Our main result is the following theorem.

Theorem 1.1: *Let*

$$V^\beta(\sigma) = k^{\frac{N}{k-1}}(2\pi N)^{-1}k^{\frac{2k-1}{2k-2}}(k!)^{\frac{-1}{k-1}}2\sqrt{6}|Y^\beta(\sigma)|, \quad \beta = 1, \ldots, k-1. \tag{1.5}$$

Then the point process on $\mathbb{R}_+^{k-1}$

$$\sum_{\sigma\in\Sigma_N} \delta_{(V^1(\sigma),\ldots,V^{k-1}(\sigma))}$$

converges weakly to the Poisson point process on $\mathbb{R}_+^{k-1}$ *with intensity measure given by the Lebesgue measure.*

Clearly, from this result we can deduce extremal properties of $H_N(\sigma) = \sum_{\beta=1}^{k-1}|Y^\beta(\sigma)|$ straightforwardly.

Remark: Integer partitioning problem. It is very easy to derive also from our Theorem 1.1 the analogous result for the integer partitioning problem. Let $S_1, \ldots, S_N$ be discrete random variables uniformly distributed on $\{1, 2, \ldots, M(N)\}$ where $M(N) > 1$ is an integer number depending on N. Let us define

$$D^\beta(\sigma) = \sum_{n=1}^{N} S_n(\mathbb{1}_{\{\sigma_n=\beta\}} - \mathbb{1}_{\{\sigma_n=\beta+1\}}).$$

Theorem 1.2: *Assume that* $M(N) \to \infty$ *as* $N \to \infty$ *such that* $\lim_{N\to\infty}(M(N))^{-1}k^{N/(k-1)} = 0$. *Let*

$$W^\beta(\sigma) = M(N)^{-1}k^{\frac{N}{k-1}}(2\pi N)^{-1}k^{\frac{2k-1}{2k-2}}(k!)^{\frac{-1}{k-1}}2\sqrt{6}|D^\beta(\sigma)|, \quad \beta = 1, \ldots, k-1. \tag{1.6}$$

Then the point process on $\mathbb{R}_+^{k-1}$

$$\sum_{\sigma\in\Sigma_N} \delta_{(W^1(\sigma),\ldots,W^{k-1}(\sigma))}$$

converges weakly to the Poisson point process on $\mathbb{R}_+^{k-1}$ *with the intensity measure which is the Lebesgue measure.*

Proof. It follows from Theorem 1.1 by the same coupling argument as in the proof of Theorem 6.4 of [BCP].

The difficulty one is confronted with when proving Theorem 1.1 is that the standard criteria for convergence for extremal processes to Poisson processes that go beyond the i.i.d. case either assume independence, stationarity, and some mixing conditions (see [LLR]), or exchangeability and a very strong form of asymptotic independence of the finite dimensional marginals [Gal,BM]. In the situation at hand, we certainly do not have independence, or stationarity, nor do we have exchangeability. Worse, also the asymptotic factorization of marginals does not hold uniformly in the form required e.g. in [BM].

What saves the day is, however, that the asymptotic factorization conditions hold *on average* on Σ_N, and that one can prove a general criterion for Poisson convergence that requires just that.

Thus the proof of Theorem 1.1 involves two steps. In Section 2 we prove an abstract theorem that gives a criteria for the convergence of an extremal process to a Poisson process, and in Sections 3,4 we show that these are satisfied in the problem at hand.

Unfortunately, and this makes the proof seriously tedious, for certain vectors σ, σ', there appear very strong correlations between $\vec{Y}(\sigma)$ and $\vec{Y}(\sigma')$ that have to be dealt with. Such a problem did already appear in a milder form in the work of Borgs et al [BCP] for $k = 2$, but in the general case $k > 2$ the associated linear algebra problems get much more difficult.

Remark: The unrestricted problem. These linear algebra problems prevented us to complete the study of the unrestricted problem (that is when the sets $I_1, \ldots, I_k$ are not necessarily of size N/k) in the case $k > 2$. In Section 5 we give a conjecture for the result similar to Theorem 1.1 in this case and explain the drawback in the proof that remains to be filled in.

Remark: Dynamical search algorithms. It would be interesting to investigate rigorously the properties of dynamical search algorithms, resp. Glauber dynamics associated to this model. This problem has been studied mainly numerically in a recent paper by Junier and Kurchan [JK]. They argued that the dynamics for long times should be described by an effective trap model, just as in the case of the Random Energy Model. This is clearly going to be the case if the particular updating rules used in [BBG1], [BBG2] for the REM will be employed, namely if the transition probability $p(\sigma, \sigma')$ depends only on the energy of the initial configuration. In the REM this choice could be partly justified by the observation that the deep traps had energies of the order $-N$, while all of their neighbors, typically, would have energies of the order of 1, give or take $\sqrt{\ln N}$. Thus, whatever the choice of the dynamics, the main obstacle to motion will always be the first step away from a deep well.

In the number partitioning problem, the situation is quite different. Let us only consider the case $k = 2$. If σ is one of the very deep wells, then

$$H_N(\sigma) = |\sum_{i=1}^{N} x_i \sigma_i| \approx 2^{-N}\sqrt{N}. \tag{1.7}$$

If σ^j denotes the configuration obtained from σ by inverting one spin, then

$$H_N(\sigma^j) \sim 2|x_j|. \tag{1.8}$$

For a typical sample of x_i's, these values range from $O(1/N)$ to $1 - O(1/N)$. Thus, if we use e.g. the Metropolis updating rule, then the probability of a step from σ to σ^j will be $\sim \exp(-2\beta|x_j|)$. It is by no means clear how high the saddle point between two deep wells will be, and whether they will all be of the same order. This implies that the actual time scale for transition times between deep wells is not obvious, nor it is clear what the trap model describing the long term dynamics would have to be.

Of course, changing the Hamiltonian from $H(\sigma)$ to $\ln H(\sigma)$, as was proposed in [JK], changes the foregoing discussion completely and brings us back to the more REM-like situation.

Acknowledgements: We thank Stephan Mertens for introducing us to the number partitioning problem and for valuable discussions.

2. A general extreme value theorem.

Consider series of M random vectors $\vec{V}_{i,M} = (V^1_{i,M}, \dots, V^p_{i,M}) \in \mathbb{R}^p_+$, $i = 1, \dots, M$.

Notation. We write $\sum_{\alpha(l)}$ when the sum is taken over all possible *ordered* sequences of *different* indices $\{i_1, \dots, i_l\} \subset \{1, \dots, M\}$. We also write $\sum_{\alpha(r_1),\dots,\alpha(r_R)}(\cdot)$ when the sum is taken over all possible *ordered* sequences of disjoint ordered subsets $\alpha(r_1) = (i_1, \dots, i_{r_1})$, $\alpha(r_2) = (i_{r_1+1}, \dots, i_{r_2}), \dots, \alpha(r_R) = (i_{r_1+\dots+r_{R-1}+1}, \dots, i_{r_1+\dots+r_R})$ of $\{1, \dots, M\}$.

Theorem 2.1: *Assume that for all finite $l = 1, 2, \dots$ and all set of constants $c_j^\beta > 0$, $j = 1, \dots, l$, $\beta = 1, \dots, p$ we have*

$$\sum_{\alpha(l)=(i_1,\dots,i_l)} \mathbb{P}\Big(V_{i_j,M}^\beta < c_j^\beta \ \forall j = 1, \dots, l, \beta = 1, \dots, p\Big) \to \prod_{\substack{j=1,\dots,l\\ \beta=1,\dots,p}} c_j^\beta, \quad M \to \infty. \tag{2.1}$$

Then the point process

$$\Pi_M^p = \sum_{i=1}^M \delta_{(V_{i,M}^1,\dots,V_{i,M}^p)} \tag{2.2}$$

on $\mathbb{R}_+^p$ converges weakly as $M \to \infty$ to the Poisson point process $\mathcal{P}^p$ on $\mathbb{R}_+^p$ with the intensity measure which is the Lebesgue measure.

Proof. Denote by $\Pi_M^p(A)$ the number of points of the process Π_M^p in a subset $A \subset \mathbb{R}_+^p$.

The proof of this theorem follows from Kallenberg theorem [Kal] on the week convergence of a point process Π_M^p to the Poisson process Π^p. Applying his theorem in our situation weak convergence holds whenever

(i) For all cubes $A = \prod_{\beta=1}^p [a^\beta, b^\beta)$

$$\mathbb{E}\Pi_M^p(A) \to |A|, \quad M \to \infty. \tag{2.3}$$

(ii) For all finite union $A = \bigcup_{l=1}^L \prod_{\beta=1}^p [a_l^\beta, b_l^\beta)$ of disjoint cubes

$$\mathbb{P}(\Pi_M^p(A) = 0) \to \epsilon^{-|A|}, \quad M \to \infty. \tag{2.4}$$

Our main tool of checking (i) and (ii) is the inclusion-exclusion principle which can be summarized as follows: for any $l = 1, 2, \dots$ and any events $O_1, \dots, O_l$

$$\mathbb{P}\Big(\bigcap_{i=1,\dots,l} O_i\Big) = \sum_{k=0}^{l} \sum_{\substack{\mathcal{A}_k=(i_1,\dots,i_k)\subset\{1,\dots,l\}\\ i_1<i_2<\dots<i_k}} (-1)^k \mathbb{P}\Big(\bigcap_{j=1}^k \bar{O}_{i_j}\Big) \tag{2.5}$$

where $\bar{O}_{i_j}$ are complementary events to O_{i_j}. We use (2.5) to "invert" the inequalities of type $\{V_{i,M}^\beta \geq a^\beta\}$, i.e. to represent their probability as the sum of probabilities of opposite events, that can be estimated by (2.1). The power of the inclusion-exclusion principle comes from the fact that the partial sums of the right-hand side provide upper and lower bounds (Bonferroni inequalities, see [Fe]), i.e. for any $n \leq [l/2]$:

$$\sum_{k=0}^{2n} \sum_{\substack{\mathcal{A}_k=(i_1,\dots,i_k)\\ \subset\{1,\dots,l\}\\ i_1<i_2<\dots<i_k}} (-1)^k \mathbb{P}\Big(\bigcap_{j=1}^k \bar{O}_{i_j}\Big) \geq \mathbb{P}\Big(\bigcap_{i=1,\dots,l} O_i\Big) \geq \sum_{k=0}^{2n+1} \sum_{\substack{\mathcal{A}_k=(i_1,\dots,i_k)\\ \subset\{1,\dots,l\}\\ i_1<i_2<\dots<i_k}} (-1)^k \mathbb{P}\Big(\bigcap_{j=1}^k \bar{O}_{i_j}\Big). \tag{2.6}$$

They imply that it will be enough to compute the limits as $N \uparrow \infty$ of terms for any fixed value of l. Using (2.5), we derive from the assumption of the theorem the following more general statement: Let $A_1, \dots, A_l \in \mathbb{R}_+^p$ be any subsets of volumes $|A_1|, \dots, |A_l|$ that can be represented as unions of disjoint cubes. Then for any $m_1, \dots, m_l$

$$\sum_{\alpha(m_1),\alpha(m_2),\dots,\alpha(m_l)} \mathbb{P}(\vec{V}_{i,M} \in A_j \ \forall i \in \alpha(m_r), \forall r = 1, \dots, l) \to \prod_{r=1}^{l} |A_r|^{m_r}. \tag{2.7}$$

Let us first concentrate on the proof of this statement. We first show it in the case of one subset, $l = 1$, which is a cube $A = \prod_{\beta=1}^{p}[a^\beta, b^\beta)$. Let $m = 1$. We denote by $\sum_{\mathcal{A}\subset\{1,\dots,p\}}$ the sum over all 2^p possible ordered subsets of coordinates : $\mathcal{A}$ denotes the subset of coordinates β such that the inequalities $V_{i,M}^\beta < a^\beta$ are excluded leaving thus $V_{i,M}^\beta < b^\beta$. Then by (2.5) applied to $\bigcap_{\beta=1}^{p}\{V_{i,M}^\beta \geq a^\beta\}$

$$\begin{aligned}\sum_{i=1}^{M}\mathbb{P}(\vec{V}_{i,M} \in A) &= \sum_{i=1}^{M}\mathbb{P}(a^\beta \leq V_{i,M}^\beta < b^\beta, \forall\beta = 1, \dots, p)\\ &= \sum_{i=1}^{M}\sum_{\mathcal{A}\subset\{1,\dots,p\}}(-1)^{|\mathcal{A}|}\mathbb{P}(V_{i,M}^\beta < a^\beta \mathbb{1}_{\beta\notin\mathcal{A}} + b^\beta \mathbb{1}_{\beta\in\mathcal{A}}, \forall\beta = 1, \dots, p)\\ &= \sum_{\mathcal{A}\subset\{1,\dots,p\}}(-1)^{|\mathcal{A}|}\sum_{i=1}^{M}\mathbb{P}(V_{i,M}^\beta < a^\beta \mathbb{1}_{\beta\notin\mathcal{A}} + b^\beta \mathbb{1}_{\beta\in\mathcal{A}}, \forall\beta = 1, \dots, p).\end{aligned} \tag{2.8}$$

The interior sum in (2.8) $\sum_{i=1}^{M}\mathbb{P}(\cdot)$ converges to $\prod_{\beta=1}^{p}(a^\beta \mathbb{1}_{\beta\notin\mathcal{A}} + b^\beta \mathbb{1}_{\beta\in\mathcal{A}})$ by the assumption (2.1). Thus

$$\lim_{M\to\infty}\sum_{i=1}^{M}\mathbb{P}(\vec{V}_{i,M} \in A) = \sum_{\mathcal{A}\subset\{1,\dots,p\}}(-1)^{|\mathcal{A}|}\prod_{\beta=1}^{p}(a^\beta \mathbb{1}_{\beta\notin\mathcal{A}} + b^\beta \mathbb{1}_{\beta\in\mathcal{A}}) = \prod_{\beta=1}^{p}(b^\beta - a^\beta) = |A|. \tag{2.9}$$

Now let $m > 1$. Denote by $\sum_{\mathcal{A}_1,\mathcal{A}_2,\dots,\mathcal{A}_m}$ the sum over all 2^{mp} ordered sequences of all 2^p unordered subsets $\mathcal{A} \subset \{1, \dots, p\}$. Here $\mathcal{A}_j$ is the subset of coordinates corresponding to the jth index in the row $\alpha(m) = (i_1, \dots, i_m)$. Then by (2.5)

$$\begin{aligned}&\sum_{\alpha(m)}\mathbb{P}(\vec{V}_{i,M} \in A \ \forall i \in \alpha(m)) = \sum_{\alpha(m)}\mathbb{P}\left(a^\beta \leq V_{i,M}^\beta < b^\beta \ \forall i \in \alpha(m), \forall\beta = 1, \dots, p\right)\\ &= \sum_{\alpha(m)}\sum_{\mathcal{A}_1,\dots,\mathcal{A}_m}(-1)^{|\mathcal{A}_1|+\cdots|\mathcal{A}_m|}\mathbb{P}\left(V_{i,M}^\beta < a^\beta \mathbb{1}_{\beta\notin\mathcal{A}_j} + b^\beta \mathbb{1}_{\beta\in\mathcal{A}_j} \ \forall i = i_j \in \alpha(m), \forall j = 1, \dots, m, \forall\beta\right)\\ &= \sum_{\mathcal{A}_1,\dots,\mathcal{A}_m}(-1)^{|\mathcal{A}_1|+\cdots|\mathcal{A}_m|}\sum_{\alpha(m)}\mathbb{P}(V_{i,M}^\beta < a^\beta \mathbb{1}_{\beta\notin\mathcal{A}_j} + b^\beta \mathbb{1}_{\beta\in\mathcal{A}_j} \ \forall i = i_j \in \alpha(m), \forall j = 1, \dots, m, \forall\beta).\end{aligned} \tag{2.10}$$

By (2.1) applied to the interior sum of (2.10) $\sum_{\alpha(m)}\mathbb{P}(\cdot)$ we get:

$$\lim_{M\to\infty}\sum_{\alpha(m)}\mathbb{P}(\vec{V}_{i,M}\in A\;\forall i\in\alpha(m))=\sum_{\mathcal{A}_1,\dots,\mathcal{A}_m}(-1)^{|\mathcal{A}_1|+\cdots|\mathcal{A}_m|}\prod_{j=1}^{m}\prod_{\beta=1}^{p}(a^\beta\mathbb{I}_{\beta\notin\mathcal{A}_j}+b^\beta\mathbb{I}_{\beta\in\mathcal{A}_j})=|A|^m.$$

Assume now that $l>1$ and $A_r=\prod_{\beta=1}^{p}[a_r^\beta,b_r^\beta)$, $r=1,\dots,l$. Then

$$\begin{aligned}
&\sum_{\alpha(m_1),\alpha(m_2),\dots,\alpha(m_l)}\mathbb{P}\left(\vec{V}_{i,M}\in A_j\;\forall i\in\alpha(m_r),\forall r=1,\dots,l\right)\\
&=\sum_{\alpha(m_1),\alpha(m_2),\dots,\alpha(m_l)}\;\sum_{\substack{\mathcal{A}_1^1,\dots,\mathcal{A}_{m_1}^1,\\ \dots,\mathcal{A}_1^l,\dots,\mathcal{A}_{m_l}^l}}(-1)^{|\mathcal{A}_1^1|+\cdots+|\mathcal{A}_{m_l}^l|}\mathbb{P}\big(V_{i,M}^\beta<a^\beta\mathbb{I}_{\beta\notin\mathcal{A}_j^r}+b^\beta\mathbb{I}_{\beta\in\mathcal{A}_j^r}\\
&\qquad\qquad\forall i=i_j\in\alpha(m_r),\forall j=1,\dots,m_r,\forall r=1,\dots,l,\forall\beta\big)\\
&=\sum_{\mathcal{A}_1^1,\dots,\mathcal{A}_{m_1}^1}(-1)^{|\mathcal{A}_1^1|+\cdots+|\mathcal{A}_{m_1}^1|}\cdots\sum_{\mathcal{A}_1^l,\dots,\mathcal{A}_{m_l}^l}(-1)^{|\mathcal{A}_1^l|+\cdots+|\mathcal{A}_{m_l}^l|}\sum_{\alpha(m_1),\alpha(m_2),\dots,\alpha(m_l)}\\
&\quad\mathbb{P}\left(V_{i,M}^\beta<a^\beta\mathbb{I}_{\beta\notin\mathcal{A}_j^r}+b^\beta\mathbb{I}_{\beta\in\mathcal{A}_j^r}\;\forall i=i_j\in\alpha(m_r),\forall j=1,\dots,m_r,\forall r=1,\dots,l,\forall\beta\right).
\end{aligned}\tag{2.11}$$

Due to (2.1) applied once more to the interior sum $\sum_{\alpha(m_1),\dots,\alpha(m_l)}\mathbb{P}(\cdot)$, (2.11) converges to

$$\begin{aligned}
&\sum_{\mathcal{A}_1^1,\dots,\mathcal{A}_{m_1}^1}(-1)^{|\mathcal{A}_1^1|+\cdots+|\mathcal{A}_{m_1}^1|}\cdots\sum_{\mathcal{A}_1^l,\dots,\mathcal{A}_{m_l}^l}(-1)^{|\mathcal{A}_1^l|+\cdots+|\mathcal{A}_{m_l}^l|}\prod_{r=1}^{l}\prod_{j=1}^{m_r}\prod_{\beta=1}^{p}(a^\beta\mathbb{I}_{\beta\notin\mathcal{A}_j^r}+b^\beta\mathbb{I}_{\beta\in\mathcal{A}_j^r})\\
&\quad=\sum_{\mathcal{A}_1^1,\dots,\mathcal{A}_{m_1}^1}(-1)^{|\mathcal{A}_1^1|+\cdots+|\mathcal{A}_{m_1}^1|}\cdots\sum_{\mathcal{A}_1^{l-1},\dots,\mathcal{A}_{m_{l-1}}^{l-1}}(-1)^{|\mathcal{A}_1^{l-1}|+\cdots+|\mathcal{A}_{m_{l-1}}^{l-1}|}\\
&\qquad\qquad\prod_{r=1}^{l-1}\prod_{j=1}^{m_r}\prod_{\beta=1}^{p}(a^\beta\mathbb{I}_{\beta\notin\mathcal{A}_j^r}+b^\beta\mathbb{I}_{\beta\in\mathcal{A}_j^r})|A_l|^{m_l}\\
&\quad=|A_1|^{m_1}|A_2|^{m_2}\cdots|A_l|^{m_l}.
\end{aligned}\tag{2.12}$$

Let finally $A_1=\bigcup_{k=1}^{s_1}A_{1,k},\dots,A_l=\bigcup_{k=1}^{s_l}A_{l,k}$ be unions of $s_1,\dots,s_l$ disjoint cubes respectively. Then we may write:

$$\begin{aligned}
&\sum_{\alpha(m_1),\alpha(m_2),\dots,\alpha(m_l)}\mathbb{P}(\vec{V}_{i,M}\in A_j\;\forall i\in\alpha(m_r),\forall r=1,\dots,l)\\
&=\sum_{\substack{m_{1,1},\dots,m_{1,s_1}\geq 0\\ m_{1,1}+\cdots+m_{1,s_1}=m_1}}\cdots\sum_{\substack{m_{l,1},\dots,m_{l,s_l}\geq 0\\ m_{l,1}+\cdots+m_{l,s_l}=m_l}}\;\sum_{\substack{\alpha(m_{1,1}),\dots,\alpha(m_{1,s_1}),\\ \dots,\alpha(m_{l,1}),\dots,\alpha(m_{l,s_l})}}\\
&\quad\mathbb{P}\left(\vec{V}_{i,M}\in A_{r,k}\forall i\in\alpha(m_{r,k})\;\forall r=1,\dots,l,\forall k=1,\dots,s_r\right)
\end{aligned}\tag{2.13}$$

and apply to the interior sum $\sum_{\alpha(m_{1,1}),\dots,,\alpha(m_{l,s_l})}\mathbb{P}(\cdot)$ the statement (2.7) about cubes just proven by (2.12). Then (2.13) converges to

$$\sum_{\substack{m_{1,1},\dots,m_{1,s_1}\geq 0\\ m_{1,1}+\dots+m_{1,s_1}=m_1}}\cdots\sum_{\substack{m_{l,1},\dots,m_{l,s_l}\geq 0\\ m_{l,1}+\dots+m_{l,s_l}=m_l}}\prod_{r=1}^{l}\prod_{k=1}^{s_r}|A_{r,k}|^{m_{r,k}}=\prod_{r=1}^{l}\sum_{\substack{m_{r,1},\dots,m_{r,s_r}\geq 0\\ m_{r,1}+\dots+m_{r,s_r}=m_r}}\prod_{k=1}^{s_r}|A_{r,k}|^{m_{r,k}}$$
$$=\prod_{r=1}^{l}|A_r|^{m_r}. \tag{2.14}$$

This finishes the proof of the statement (2.7).

Now we are ready to turn to the proof of the theorem. The condition (i) has been already shown by (2.9). To verify (ii), let us construct a cube $B=\prod_{\beta=1}^{p}[0,\max_{l=1,\dots,L}b_l^{\beta})$ of volume $|B|$, then clearly $A\subset B$. For any $R>0$ we may write the following decomposition:

$$\begin{aligned}\mathbb{P}(\Pi_M(A)=0)&=\sum_{r=0}^{R}\frac{1}{r!}\sum_{\alpha(r)}\mathbb{P}\left(\vec{V}_{i,M}\in B\setminus A\,\forall i\in\alpha(r),\ \vec{V}_{i,M}\notin B\,\forall i\notin\alpha(r)\right)\\&\quad+\mathbb{P}(\Pi_M(A)=0,\Pi_M(B)>R)\equiv I_1(R,M)+I_2(R,M).\end{aligned}\tag{2.15}$$

Applying the inclusion-exclusion (2.6) principle to $M-r$ events $\{\vec{V}_i\notin B\}$ for $i\notin\alpha(r)$, we may bound $I_1(R,M)$ for all $n\leq[(M-r)/2]$ by

$$\begin{aligned}&\sum_{r=0}^{R}\frac{1}{r!}\sum_{k=0}^{2n}\frac{(-1)^k}{k!}\sum_{\alpha(r),\alpha(k)}\mathbb{P}(\vec{V}_{i,M}\in B\setminus A\ \forall i\in\alpha(r),\ \vec{V}_{i,M}\in B\ \forall i\in\alpha(k))\geq I_1(R,M)\\&\geq\sum_{r=0}^{R}\frac{1}{r!}\sum_{k=0}^{2n+1}\frac{(-1)^k}{k!}\sum_{\alpha(r)\alpha(k)}\mathbb{P}(\vec{V}_{i,M}\in B\setminus A\ \forall i\in\alpha(r),\ \vec{V}_{i,M}\in B\ \forall i\in\alpha(k)).\end{aligned}\tag{2.16}$$

Then for any fixed $n\geq 1$, the statement (2.7) applied to the subsets A/B and B imply:

$$\sum_{r=0}^{R}\frac{|B\setminus A|^r}{r!}\sum_{k=0}^{2n}\frac{(-1)^k|B|^k}{k!}\geq\lim_{M\to\infty}I_1(R,M)\geq\sum_{r=0}^{R}\frac{|B\setminus A|^r}{r!}\sum_{k=0}^{2n+1}\frac{(-1)^k|B|^k}{k!}.\tag{2.17}$$

Since n can be fixed arbitrarily large, it follows that

$$\lim_{M\to\infty}I_1(R,M)=e^{-|B|}\sum_{r=0}^{R}\frac{|B\setminus A|^r}{r!}.\tag{2.18}$$

The statement (2.7) also gives

$$\lim_{M\to\infty}I_2(R,M)\leq\lim_{M\to\infty}\mathbb{P}(\Pi_M^1(B)>R)=\lim_{M\to\infty}\frac{1}{R!}\sum_{\alpha(R)}\mathbb{P}(\vec{V}_{i,M}\in B\ \forall i\in\alpha(R))=\frac{|B|^R}{R!}.\tag{2.19}$$

By choosing R large enough, the limit (2.19) can be done as small as desired and the sum (2.18) can be done as close to the exponent $e^{|B\setminus A|-|B|}$ as wanted. Hence, $\lim_{M\to\infty}\mathbb{P}(\Pi^1_M(A)) = e^{-|A|}$. This concludes the proof of the theorem. $\diamondsuit$

3. Application to number partitioning

We will now prove Theorem 1.1. In fact, the proof will follow directly from Theorem 2.1 and the following proposition:

Proposition 3.1: *Let*

$$S(k,N) = k^N (2\pi N)^{\frac{1-k}{2}} k^{\frac{k}{2}} (k!)^{-1} \tag{3.1}$$

be borrowed from (1.3). We denote by $\sum_{\sigma^1,\dots,\sigma^l\in\Sigma_N}(\cdot)$ *the sum over all possible ordered sequences of different elements of* Σ_N*. Then for any* $l = 1,2,\dots$*, any constants* $c_j^\beta > 0$*,* $j = 1,\dots,l$*,* $\beta = 1,\dots,k-1$ *we have:*

$$\begin{aligned}\sum_{\sigma^1,\dots,\sigma^l\in\Sigma_N} \mathbb{P}\Big(\forall\beta = 1,\dots,k-1, \forall j = 1,\dots,l \frac{|Y^\beta(\sigma^j)|}{\sqrt{2(N/k)\mathrm{var}\, X}} < \frac{c_j^\beta}{S(k,N)^{\frac{1}{k-1}}}\Big)\\ \to \prod_{\substack{j=1,\dots,l\\ \beta=1,\dots,k-1}} (2(2\pi)^{-1/2} c_j^\beta).\end{aligned} \tag{3.2}$$

Informal arguments. Before proceeding with the rigorous proof, let us give intuitive arguments supporting this lemma.

The random variables $\frac{Y^\beta(\sigma^j)}{\sqrt{2(N/k)\mathrm{var}\, X}}$ are the sums of independent identically distributed random variables with the expectations $\mathbb{E}Y^\beta(\sigma^j) = 0$ and the covariance matrix $B_N(\sigma^1,\dots,\sigma^l)$ with the elements

$$b_{i,s}^{\beta,\gamma} = \frac{\mathrm{cov}\,(Y^\beta(\sigma^i), Y^\gamma(\sigma^s))}{2(N/k)\mathrm{var}\, X} = \frac{\sum_{n=1}^N (\mathbb{1}_{\{\sigma_n^i=\beta\}} - \mathbb{1}_{\{\sigma_n^i=\beta+1\}})(\mathbb{1}_{\{\sigma_n^s=\gamma\}} - \mathbb{1}_{\{\sigma_n^s=\gamma+1\}})}{2(N/k)}. \tag{3.3}$$

In particular:

$$b_{i,i,}^{\beta,\beta} = 1,\quad b_{i,i}^{\beta,\beta+1} = -1/2,\quad b_{i,i}^{\beta,\gamma} = 0 \text{ for } \gamma\neq\beta,\beta+1,\ \forall\, i = 1,\dots,k-1. \tag{3.4}$$

Moreover, the property that $b_{i,j}^{\beta,\gamma} = o(1)$ as $N\to\infty$ for all $i\neq j$, β, γ, holds for a number $R(N,l)$ of sets $\sigma^1,\dots,\sigma^l\in\Sigma_N^{\otimes l}$ which is $R(N,l) = |\Sigma_N|^l(1+o(1)) = S(k,N)^l(1+o(1))$ with $o(1)$ exponentially small as $N\to\infty$. For all such sets $\sigma^1,\dots,\sigma^l$, by the Central Limit Theorem, the random variables $\frac{Y^\beta(\sigma^j)}{\sqrt{2(N/k)\mathrm{var}\, X}}$ should behave asymptotically as centered

Gaussian random variables with covariances $b_{i,j}^{\beta,\gamma} = 1_{\{i=j,\beta=\gamma\}} + (-1/2)1_{\{i=j,\gamma=\beta+1\}} + o(1)$. The determinant of this covariance matrix is $1 + o(1)$. Hence, the probability $\mathbb{P}(\cdot)$ defined in (3.2) that these Gaussians belong to the exponentially small segments $[-c_j^\beta S(k,N)^{-1/(k-1)}, c_j^\beta S(k,N)^{-1/(k-1)}]$ is of the order $\prod_{\substack{j=1,\dots,l\\ \beta=1,\dots,k-1}} (2(2\pi)^{-1/2} c_j^\beta S(k,N)^{-1/(k-1)})$. Multiplying this probability by the number of terms $R(N,l)$ we get the result claimed in (3.2).

Let us turn to the remaining tiny part of $\Sigma_N^{\otimes l}$ where $\sigma^l, \dots, \sigma^l$ are such that $b_{i,j}^{\beta,\gamma} \not\to 0$ for some $i \neq j$ as $N \to \infty$. Here two possibilities should be considered differently. The first one is when the covariance matrix $B_N(\sigma^1, \dots, \sigma^l)$ of $\frac{Y^\beta(\sigma^j)}{\sqrt{2(N/k)\text{var}\, X}}$ is non-degenerate. Then invoking again the Central Limit Theorem, the probability $\mathbb{P}(\cdot)$ in this case is of the order

$$(\det B_N(\sigma^1, \dots, \sigma^l))^{-1/2} \prod_{\substack{j=1,\dots,l\\ \beta=1,\dots,k-1}} (2(2\pi)^{-1/2} c_j^\beta S(k,N)^{-1/(k-1)}).$$

But from the definition of $b_{i,j}^{\beta,\gamma}$ $(\det B_N(\sigma^1, \dots, \sigma^l))^{-1/2}$ may grow at most polynomially. Thus the probability $\mathbb{P}(\cdot)$ is about $S(k,N)^{-l}$ up to a polynomial term while the number of sets $\sigma^1, \dots, \sigma^l$ in this part is exponentially smaller than $S(k,N)^l$. Hence, the contribution of all such $\sigma^l, \dots, \sigma^l$ in (3.2) is exponentially small.

The case of $\sigma^1, \dots, \sigma^l$ with $B(\sigma^1, \dots, \sigma^l)$ degenerate is more delicate. Although the number of such $\sigma^1, \dots, \sigma^l$ is exponentially smaller than $S(k,N)^l$, the probability $\mathbb{P}(\cdot)$ is exponentially bigger than $S(k,N)^{-l}$ since the system of $l(k-1)$ random variables $\{Y^\beta(\sigma^i)\}_{\beta=1,\dots,k-1}^{i=1,\dots,l}$ is linearly dependent! First of all, it may happen that there exist $1 \leq i_1 < i_2 < \dots < i_p \leq l$ such that the basis of this system consists of $(k-1)p$ elements $\{Y^\beta(\sigma^{i_j})\}_{\beta=1,\dots,k-1}^{j=1,\dots,p}$. Then the assumption that the elements $\sigma^1, \dots, \sigma^l$ of Σ_N must be different, plays a crucial role: due to it the number of such sets $\sigma^1, \dots, \sigma^l$ in this sum remains small enough compare to the probability $\mathbb{P}(\cdot)$, consequently their total contribution to (3.2) vanishes.

Finally, for some sets $\sigma^1, \dots, \sigma^l$, there is no such $p < l$: for any basis, there exists a number $j \in \{1, \dots, l\}$ such that the random variables $Y^\beta(\sigma^j)$ are included in the basis for some *non-empty* subset of coordinates β and are not included there for the complementary *non-empty* subset of β. This last part is clearly absent in the case $k = 2$. It turns out that its analysis is quite tedious. We manage to complete it only in the case of the constrained problem by evaluating the number of such sets $\sigma^1, \dots, \sigma^l$ where each of spins' values $\{1, \dots, k\}$ figures out exactly N/k times and by showing that the corresponding probabilities $\mathbb{P}(\cdot)$ are negligible compare to this number. The only drawback that remains in the study of the unconstrained problem is precisely the analysis of this part.

Proof of Proposition 3.1. In the course of the proof we will rely on four lemmata that will be stated here but proven separately in Section 4. Let

$$f_N^{\sigma^1,\dots,\sigma^l}(\{t_{\beta,j}\}) = \mathbb{E}\exp\Big(\frac{i}{\sqrt{2(N/k)\mathrm{var}\,X}}\sum_{\substack{j=1,\dots,l,\\ \beta=1,\dots,k-1}} t_{\beta,j}Y^\beta(\sigma^j)\Big) \tag{3.5}$$

be the characteristic function of the random vector $(2(N/k)\mathrm{var}\,X)^{-1/2}\{Y^\beta(\sigma^j)\}_{\substack{j=1,\dots,l,\\ \beta=1,\dots,k-1}}$. Here $\vec{t} = \{t_{\beta,j}\}_{\substack{\beta=1,\dots,k-1,\\ j=1,\dots,l}}$ is the vector with $(k-1)l$ coordinates. Then

$$\mathbb{P}\Big(\forall\beta = 1,\dots,k-1, \forall j = 1,\dots,l \frac{|Y^\beta(\sigma^j)|}{\sqrt{2(N/k)\mathrm{var}\,X}} < \frac{c_j^\beta}{S(k,N)^{\frac{1}{k-1}}}\Big) \tag{3.6}$$

$$= \frac{1}{(2\pi)^{l(k-1)}}\lim_{D\to\infty}\int_{[-D,D]^{l(k-1)}} f_N^{\sigma^1,\dots,\sigma^l}(\{t_{\beta,j}\})\prod_{\substack{j=1,\dots,l,\\ \beta=1,\dots,k-1}}\frac{e^{it_{\beta,j}c_j^\beta S(k,N)^{\frac{-1}{k-1}}} - e^{-it_{\beta,j}c_j^\beta S(k,N)^{\frac{-1}{k-1}}}}{it_{\beta,j}}dt_{\beta,j}.$$

It will be convenient to have in mind the following representation throughout the proof. Any configuration σ gives rise to $k-1$ configurations $\sigma^{(1)},\dots,\sigma^{(k-1)} \in \{-1,0,1\}^N$ such that

$$\sigma_n^{(\beta)} = 1\!\!\mathrm{I}_{\{\sigma_n=\beta\}} - 1\!\!\mathrm{I}_{\{\sigma_n=\beta+1\}}, \quad n = 1,\dots,N. \tag{3.7}$$

We now define the $N\times(k-1)$ matrix $C(\sigma)$ composed of columns $\sigma^{(1)},\dots,\sigma^{(k-1)}$. Then it is composed of types of k rows of length $k-1$: $O_0 = (1,0,\dots,0)$, $O_1 = (-1,1,0\dots,0)$, $O_2 = (0,-1,1,0,\dots,0),\dots, O_{k-2} = (0,\dots,0,-1,1)$, $O_{k-1} = (0,\dots,0,-1)$. They correspond to spin values $1,2,\dots,k$ respectively: if $\sigma_n = \beta$, then the nth row of $C(\sigma)$ is $O_{\beta-1}$.[4] Each of these k rows is repeated N/k times in the construction of $C(\sigma)$. Then

$$Y^\beta(\sigma) = \sum_{n=1}^N X_n\sigma_n^{(\beta)}.$$

Let $C(\sigma^1,\dots,\sigma^l)$ be the $N\times(k-1)l$ matrix composed by the columns $\sigma^{1,(1)},\sigma^{1,(2)},\dots,\sigma^{1,(k-1)},\sigma^{2,(1)},\dots,\sigma^{l,(k-1)}$. Then it is easy to see that the function $f_N^{\sigma^1,\dots,\sigma^l}(\{t_{\beta,j}\})$ is the product of N functions

$$f_N^{\sigma^1,\dots,\sigma^l}(\{t_{\beta,j}\}) = \prod_{n=1}^N \mathbb{E}\exp\Big(\frac{iX_n}{\sqrt{2(N/k)\mathrm{var}\,X}}\{C(\sigma^1,\dots,\sigma^l)\vec{t}\}_n\Big)$$

$$= \prod_{n=1}^N\frac{\exp\Big(\frac{i}{\sqrt{2(N/k)\mathrm{var}\,X}}\{C(\sigma^1,\dots,\sigma^l)\vec{t}\}_n\Big) - 1}{i(\sqrt{2(N/k)\mathrm{var}\,X})^{-1}\{C(\sigma^1,\dots,\sigma^l)\vec{t}\}_n}, \tag{3.8}$$

[4] The case $k=2$ is particular, since here $C(\sigma)$ is the vector with elements ±1; i.e. in this case this reparametrisation just corresponds to passing from values $\{1,2\}$ to $\{-1,+1\}$.

where $\{C(\sigma^1,\dots,\sigma^l)\vec{t}\}_n$ is the nth coordinate of the product of the vector $\vec{t}=\{t_{\beta,j}\}_{\substack{\beta=1,\dots,k-1,\\ j=1,\dots,l}}$ with the matrix $C(\sigma^1,\dots,\sigma^l)$.

We will split the sum of (3.2) into two terms

$$\sum_{\sigma^1,\dots,\sigma^l\in\Sigma_N}\mathbb{P}(\cdot)=\sum_{\substack{\sigma^1,\dots,\sigma^l\in\Sigma_N\\ \operatorname{rank} C(\sigma^1,\dots,\sigma^l)=(k-1)l}}\mathbb{P}(\cdot)+\sum_{\substack{\sigma^1,\dots,\sigma^l\in\Sigma_N\\ \operatorname{rank} C(\sigma^1,\dots,\sigma^l)<(k-1)l}}\mathbb{P}(\cdot) \tag{3.9}$$

and show that the first term converges to the right-hand side of (3.2) while the second term converges to zero.

We start with the second term in (3.9) that we split into two parts

$$\sum_{\substack{\sigma^1,\dots,\sigma^l\in\Sigma_N\\ \operatorname{rank} C(\sigma^1,\dots,\sigma^l)<(k-1)l}}\mathbb{P}(\cdot)=J_N^1+J_N^2. \tag{3.10}$$

In the first part J_N^1 the sum is taken over ordered sets $\sigma^1,\dots,\sigma^l$ of different elements of Σ_N with the following property: the rank r of $C(\sigma^1,\dots,\sigma^l)$ is a multiple of $(k-1)$ and, moreover, there exist configurations $\sigma^{i_1},\dots,\sigma^{i_{r/(k-1)}}$ such that all of $\sigma^{(1),i_1},\sigma^{(2),i_1},\dots,\sigma^{(k-1),i_{r/(k-1)}}$ constitute the basis of the columns of the matrix $C(\sigma^1,\dots,\sigma^l)$, i.e. the rank of $C(\sigma^{i_1},\dots,\sigma^{i_{r/(k-1)}})$ equals r. Consequently, for any $j\in\{1,\dots,l\}\setminus\{i_1,\dots,i_{r/(k-1)}\}$ all of $\sigma^{(1),j},\dots,\sigma^{(k-1),j}$ are linear combinations of the columns of the matrix $C(\sigma^{i_1},\dots,\sigma^{i_{r/(k-1)}})$. In the remaining part, J_N^2, the sum is taken over configurations $\sigma^1,\sigma^2,\dots,\sigma^l$ satisfying the complementary property: for any basis of the columns of $C(\sigma^1,\dots,\sigma^l)$ there exist at least one configuration σ^i such that some of the configurations $\sigma^{(1),i},\dots,\sigma^{(k-1),i}$ are included in this basis and some others are not[5].

The following Lemma 3.2 shows that the sum J_N^1 is taken over sets of different $\sigma^1,\dots,\sigma^l$ such that the matrix of the basis $C(\sigma^{i_1},\dots,\sigma^{i_{r/(k-1)}})$ contains at most $(k^{r/(k-1)}-1)$ different rows.

Lemma 3.2: *Assume that the matrix $C(\sigma^1,\dots,\sigma^l)$ contains all k^l different rows. Assume that a configuration $\tilde\sigma$ is such that each $\tilde\sigma^{(1)},\dots,\tilde\sigma^{(k-1)}$ is a linear combination of the columns of the matrix $C(\sigma^1,\dots,\sigma^l)$. Then the configuration $\tilde\sigma$ is obtained by a permutation of spin values in one of the configurations $\sigma^1,\dots,\sigma^l$, i.e. $\tilde\sigma$ coincides with one of $\sigma^1,\dots,\sigma^l$ as an element of Σ_N.*

[5] In the case $k=2$ the term J_N^2 can obviously not exist. This leads to considerable simplifications.

Remark: In the case $k=2$, Lemma 3.2 has been an important ingredient in the analysis of the Hopfield model. It possibly appeared first in a paper by Koch and Piasko [KP].

In fact, if in J_N^1 the matrix $C(\sigma^{i_1},\dots,\sigma^{i_{r/(k-1)}})$ contained all $k^{r/(k-1)}$ different rows, then by Lemma 3.2 the remaining configurations σ^j with $j\in\{1,\dots,l\}\setminus\{i_1,\dots,i_{r/(k-1)}\}$ would be equal to one of $\sigma^{i_1},\dots,\sigma^{i_{r/(k-1)}}$ as elements of Σ_N, which is impossible since the sum in (3.10) is taken over *different* elements of Σ_N. Thus there can be at most $O((k^{r/(k-1)}-1)^N)$ possibilities to construct $C(\sigma^{i_1},\dots,\sigma^{i_{r/(k-1)}})$ in the sum J_N^1. Furthermore, there is only a N-independent number of possibilities to complete it by linear configurations of its columns up to $C(\sigma^1,\dots,\sigma^l)$. To see this, assume that there are $\nu<k^{r/(k-1)}$ different rows in the matrix $C(\sigma^{i_1},\dots,\sigma^{i_{r/(k-1)}})$ and consider its restriction to these rows which is the $\nu\times r$ matrix $\widetilde{C}(\sigma^{i_1},\dots,\sigma^{i_{r/(k-1)}})$. Then $\widetilde{C}(\sigma^{i_1},\dots,\sigma^{i_{r/(k-1)}})$ has the same rank r as $C(\sigma^{i_1},\dots,\sigma^{i_{r/(k-1)}})$. Now there are not more than $3^{(\nu(l(k-1)-r))}$ ways to complete the matrix $\widetilde{C}$ to a $\nu\times l(k-1)$ matrix with elements $1,-1,0$ such that all added columns of length ν are linear combinations of those of $\widetilde{C}$. But each such choice determines uniquely the coefficients in these linear combinations, and hence the completion of the full $N\times r$ matrix $C(\sigma^{i_1},\dots,\sigma^{i_{r/(k-1)}})$ up to the $N\times l(k-1)$ matrix $C(\sigma^1,\dots,\sigma^l)$ is already fully determined. Thus the number of terms in the sum representing J_N^1 is smaller than

$$\sum_{\nu=r}^{k^{r/(k-1)}-1}\nu^N 3^{(\nu(l(k-1)-r))}=O\big(\big(k^{r/(k-1)}-1\big)^N\big). \tag{3.11}$$

The next proposition gives an a priori estimate for each of these terms.

Lemma 3.3: *There exists a constant $K(k,l)>0$ such that for any different $\sigma^1,\dots,\sigma^l\in\Sigma_N$, any $r=\operatorname{rank}C(\sigma^1,\dots\sigma^l)\le(k-1)l$ and all $N>1$*

$$\mathbb{P}\Big(\forall\beta=1,\dots,k-1,\forall j=1,\dots,l\,\frac{|Y^\beta(\sigma^j)|}{\sqrt{2(N/k)\operatorname{var}X}}<\frac{c_j^\beta}{S(k,N)^{\frac{1}{k-1}}}\Big)\le KS(k,N)^{-r/(k-1)}N^{3r/2}. \tag{3.12}$$

Hence, by Lemma 3.3 each term in J_N^1 is smaller than $KS(k,N)^{-r/(k-1)}N^{3r/2}$ with the leading exponential term $k^{-Nr/(k-1)}$. It follows that $J_N^1=O\big([(k^{r/(k-1)}-1)k^{-r/(k-1)}]^N\big)\to 0$ as $N\to\infty$.

Let us now turn to J_N^2 in (3.10). The next proposition allows to evaluate the number of terms in this sum.

Lemma 3.4: *Let D_N be any $N\times q$ matrix of rank $r\le q$. Assume that for any $N>1$ it is composed only of R different rows taken from a finite set $\mathcal{D}$ of cardinality $R\ge k$.*

Let $Q_N(R,t)$ be the number of configurations σ such that the matrix D_N completed by the columns $\sigma^{(1)},\dots,\sigma^{(k-1)}$ has rank $r+t$ where $1\le t\le k-2$. Then there exists a constant $K(R,t,k)>0$, depending only on R,t,k, such that

$$Q_N(R,t)\le K(R,t,k)\frac{(N(t+1)/k)!}{((N/k)!)^{t+1}}. \tag{3.13}$$

Now, to treat J_N^2, consider $\sigma^1,\dots,\sigma^l$ such that $(k-1)m+t_1+t_2+\cdots+t_s=r$ columns of $C(\sigma^1,\dots,\sigma^l)$ form a basis for the span of all column vectors of this matrix. Then there exist $\sigma^{i_1},\dots,\sigma^{i_m}$ such that all of $\sigma^{(v),i_p}$ are included in the basis for all $v=1,\dots,k-1$, $p=1,\dots,m$, and there exist $\sigma^{j_1},\dots,\sigma^{j_s}$ such that among $\sigma^{(v),j_q}$ $t_q\ge 1$ configurations are included in the basis and other $k-1-t_q\ge 1$ are not, $q=1,\dots,s$. By Lemma 3.4 the number of possibilities to construct such a matrix $C(\sigma^1,\dots,\sigma^l)$ is

$$O\Big(k^{mN}\prod_{q=1}^{s}\frac{(N(t_q+1)/k)!}{((N/k)!)^{t_q+1}}\Big)\sim k^{Nm}\prod_{s=1}^{q}(t_q+1)^{N(t_q+1)/k}$$

up to leading exponential order. The probability in (3.9) is already estimated in Lemma 3.3: it is

$$O(N^{3r/2}S(k,N)^{-r/(k-1)})\sim k^{-Nr/(k-1)}=k^{-N(m(k-1)+t_1+t_2+\cdots+t_s)/(k-1)}.$$

Thus, to conclude that $J_N^2\to 0$ exponentially fast, it suffices to show that for any $k=3,4,\dots$ and any $t=1,2,\dots,k-2$ we have $(t+1)^{(t+1)/k}k^{-t/(k-1)}<1$, which is reduced to the inequality

$$\phi(k,t)=\frac{k-1}{t}\ln(t+1)-\frac{k}{t+1}\ln k<0.$$

It is elementary to check that $\frac{\partial\phi(k,t)}{\partial k}<0$ for all $k\ge t+1$ and $t\ge 1$. Then, given t, it suffices to check this inequality for the smallest value of k which is $k=t+2$, that is that

$$\psi(k)=(k-1)^2\ln(k-1)-k(k-2)\ln k<0.$$

This is easy as $\psi'(k)<0$ for all $k\ge 3$ and $\psi(3)<0$. Hence, $J_N^2\to 0$ as $N\to\infty$. Thus the proof of the convergence to zero of the second term of (3.9) is complete.

We now concentrate on the convergence of the first term of (3.9). Let us fix any $\alpha\in(0,1/2)$ and introduce a subset $\mathcal{R}_{l,N}^{\alpha}\subset\Sigma_N^{\otimes l}$:

$$\mathcal{R}_{l,N}^{\alpha}=\Big\{\sigma^1,\dots,\sigma^l\in\Sigma_N:\forall 1\le i<r\le l,1\le\beta,\gamma,\eta\le k,\beta\ne\gamma$$

$$\Big|\sum_{n=1}^{N}(\mathbb{I}_{\{\sigma_n^i=\beta\}}-\mathbb{I}_{\{\sigma_n^i=\gamma\}})\mathbb{I}_{\{\sigma_n^r=\eta\}}\Big|<N^{\alpha+1/2}\Big\}. \qquad (3.14)$$

This subset can be constructed as follows. Take σ^1 where each of k possible values of spins is present N/k times. Divide each set $\mathcal{A}_\beta\equiv\{i\in\{1,\ldots,N\}:\sigma_i^1=\beta\}$, $\beta\in\{1,\ldots,k\}$, into $N/k+O(N^{\alpha+1/2})$ pieces $\mathcal{A}_{\beta,\gamma}$ of length $N/k^2+O(N^{\alpha+1/2})$. Then the spins of σ^2 have the same value on the subsets of indices which are composed by k such pieces $\mathcal{A}_{\beta,\gamma}$ taken from different $\mathcal{A}_\beta$, $\beta=1,\ldots,k$. Next, divide k^2 subsets $\mathcal{A}_{\beta,\gamma}$ into k pieces $\mathcal{A}_{\beta,\gamma,\delta}$. The spins of σ^3 have the same values on the subsets composed by k^2 such pieces $\mathcal{A}_{\beta,\gamma,\delta}$ of length $N/k^3+O(N^{\alpha+1/2})$ taken from different $\mathcal{A}_{\beta,\gamma}$, etc. It is an easy combinatorial computation to check that with some constant $h>0$

$$|\Sigma_N^{\otimes l}\setminus\mathcal{R}_{l,N}^\alpha|\le k^{Nl}\exp(-hN^{2\alpha}) \qquad (3.15)$$

from where by (1.3)

$$|\mathcal{R}_{l,N}^\alpha|=S(k,N)^l(1+o(1)),\quad N\to\infty. \qquad (3.16)$$

It is also not difficult to see that for any $\sigma^1,\ldots,\sigma^l\in\mathcal{R}_{l,N}^\alpha$ the rank of $C(\sigma^1,\ldots,\sigma^l)$ equals $(k-1)l$. Note that the covariance matrix B_N (see (3.3)) can be expressed as

$$B_N(\sigma^1,\ldots,\sigma^l)=\frac{C^T(\sigma^1,\ldots,\sigma^l)C(\sigma^1,\ldots,\sigma^l)}{2(N/k)\mathrm{var}\,X}. \qquad (3.17)$$

Thus by definition of $\mathcal{R}_{N,l}^\alpha$, its elements satisfy

$$b_{i,j}^{\beta,\gamma}=O(N^{\alpha-1/2})\quad\forall\,\beta,\gamma,i\ne j, \qquad (3.18)$$

uniformly for $\forall\,\sigma^1,\ldots,\sigma^l\in\mathcal{R}_{N,l}^\alpha$. Therefore, for any $\sigma^1,\ldots,\sigma^l\in\mathcal{R}_{l,n}^\alpha$, $\det B_N(\sigma^1,\ldots,\sigma^l)=1+o(1)$ and consequently the rank of $C(\sigma^1,\ldots,\sigma^l)$ equals $(k-1)l$.

By Lemma 3.3 and the estimate (3.16)

$$\sum_{\substack{\sigma^1,\ldots,\sigma^l\not\in\mathcal{R}_{l,N}^\alpha\\ \mathrm{rank}\,C(\sigma^1,\ldots,\sigma^l)=(k-1)l}}\mathbb{P}(\cdot)\le k^{Nl}e^{-hN^{2\alpha}}KN^{3(k-1)l/2}S(k,N)^{-l}\to 0. \qquad (3.19)$$

To complete the study of the first term of (3.9), let us show that

$$\sum_{\sigma^1,\ldots,\sigma^l\in\mathcal{R}_{l,N}^\alpha}\mathbb{P}(\cdot)\to(2\pi)^{-(k-1)l/2}\prod_{\substack{j=1,\ldots,l\\ \beta=1,\ldots,k-1}}(2c_j^\beta) \qquad (3.20)$$

with $\mathbb{P}(\cdot)$ defined by (3.6). Using the representation (3.6), will divide the normalized probability $\mathbb{P}(\cdot)$ of (3.6) into five parts

$$S(k,N)^l\Big(\prod_{\substack{i=1,\dots,k-1,\\ j=1,\dots,l}}(2c_j^\beta)^{-1}\Big)\mathbb{P}(\cdot)=\sum_{i=1}^{5}I_N^i(\sigma^1,\dots,\sigma^l) \tag{3.21}$$

where:

$$I_N^1\equiv(2\pi)^{-l(k-1)}\int\limits_{\|\vec{t}\|<\epsilon N^{1/6}}e^{-\vec{t}B_N(\sigma^1,\dots,\sigma^l)\vec{t}/2}\prod_{\substack{\beta=1,\dots,k-1,\\ j=1,\dots,l}}dt_{\beta,j}, \tag{3.22}$$

$$I_N^2\equiv(2\pi)^{-l(k-1)}\int\limits_{\|\vec{t}\|<\epsilon N^{1/6}}\big(f_N^{\sigma^1,\dots,\sigma^l}(\{t_{\beta,j}\})-e^{-\vec{t}B_N(\sigma^1,\dots,\sigma^l)\vec{t}/2}\big)\prod_{\substack{\beta=1,\dots,k-1,\\ j=1,\dots,l}}dt_{\beta,j}, \tag{3.23}$$

$$I_N^3\equiv(2\pi)^{-l(k-1)}\int\limits_{\epsilon N^{1/6}<\|\vec{t}\|<\delta\sqrt{N}}f_N^{\sigma^1,\dots,\sigma^l}(\{t_{\beta,j}\})\prod_{\substack{\beta=1,\dots,k-1,\\ j=1,\dots,l}}dt_{\beta,j}, \tag{3.24}$$

$$\begin{aligned}I_N^4\equiv&(2\pi)^{-l(k-1)}\int\limits_{\|\vec{t}\|\le\delta\sqrt{N}}f_N^{\sigma^1,\dots,\sigma^l}(\{t_{\beta,j}\})\\ &\times\Big[\prod_{\substack{\beta=1,\dots,k-1\\ j=1,\dots,l}}\frac{e^{it_{\beta,j}c_j^\beta S(k,N)^{-1/(k-1)}}-e^{-it_{\beta,j}c_j^\beta S(k,N)^{-1/(k-1)}}}{2it_{j,\beta}c_j^\beta S(k,N)^{-1/(k-1)}}-1\Big]\prod_{\substack{\beta=1,\dots,k-1,\\ j=1,\dots,l}}dt_{\beta,j}\end{aligned} \tag{3.25}$$

and

$$\begin{aligned}I_N^5\equiv&(2\pi)^{-l(k-1)}\lim_{D\to\infty}\int\limits_{[-D,D]^{l(k-1)}\cap\|\vec{t}\|>\delta\sqrt{N}}f_N^{\sigma^1,\dots,\sigma^l}(\{t_{\beta,j}\})\\ &\times\prod_{\substack{\beta=1,\dots,k-1\\ j=1,\dots,l}}\frac{e^{it_{\beta,j}c_j^\beta S(k,N)^{-1/(k-1)}}-e^{-it_{\beta,j}c_j^\beta S(k,N)^{-1/(k-1)}}}{2it_{\beta,j}c_j^\beta S(k,N)^{-1/(k-1)}}dt_{\beta,j}.\end{aligned} \tag{3.26}$$

for values $\delta,\epsilon>0$ to be chosen appropriately later. We will show that there is a choice such that $I_N^i(\sigma^1,\dots,\sigma^l)\to0$ for $i=2,3,4,5$ and $I_N^1(\sigma^1,\dots,\sigma^l)\to(2\pi)^{-(k-1)l/2}$, uniformly for $\sigma^1,\dots,\sigma^l\in\mathcal{R}_{l,N}^\alpha$ as $N\to\infty$. These facts combined with (3.16) imply the assertion (3.20) and complete the proof of the proposition. The following lemma gives control over some of the terms appearing above.

Lemma 3.5: *There exist constants $C>0$, $\epsilon>0$, $\delta>0$, and $\zeta>0$, such that for all $\sigma^1,\dots,\sigma^l\in\mathcal{R}_{l,N}^\alpha$, the following estimates hold:*

(i)

$$\Big|f_N^{\sigma^1,\dots,\sigma^l}(\{t_{\beta,j}\})-e^{-\vec{t}B_N(\sigma^1,\dots,\sigma^l)\vec{t}/2}\Big|\le\frac{C\|\vec{t}\|^3}{\sqrt{N}}e^{-\vec{t}B_N(\sigma^1,\dots,\sigma^l)\vec{t}/2},\quad \textit{for all } \|\vec{t}\|<\epsilon N^{1/6} \tag{3.27}$$

(ii)

$$\left|f_N^{\sigma^1,\dots,\sigma^l}(\{t_{\beta,j}\})\right| \le e^{-\vec{t}B_N(\sigma^1,\dots,\sigma^l)\vec{t}/2+C|t|^3N^{-1/2}} \quad \textit{for all } \|\vec{t}\| < \delta\sqrt{N}, \tag{3.28}$$

and

(iii)

$$\left|f_N^{\sigma^1,\dots,\sigma^l}(\{t_{\beta,j}\})\right| \le e^{-\zeta\|\vec{t}\|^2} \quad \textit{for all } \|\vec{t}\| < \delta\sqrt{N}. \tag{3.29}$$

We can now estimate the terms I_N^i. First, by a standard estimate on Gaussian integrals,

$$\begin{aligned} I_N^1(\sigma^1,\dots,\sigma^l) &= ((2\pi)^{(k-1)l}\det B_N(\sigma^1,\dots,\sigma^l))^{-1/2} + o(1) \\ &= (2\pi)^{-(k-1)l/2} + o(1), \quad N\to\infty, \end{aligned} \tag{3.30}$$

where $o(1)$ is uniform for $\sigma^1,\dots,\sigma^l \in \mathcal{R}_{l,N}^{\alpha}$ by (3.18) and (3.4). Thus I_N^1 gives the desired main contribution.

The second part $I_N^2(\sigma^1,\dots,\sigma^l) = O(N^{-1/2})$, uniformly for $\sigma^1,\dots,\sigma^l \in \mathcal{R}_{l,N}^{\alpha}$ by the estimates (3.27) and (3.18), (3.4). The third part $I_N^3(\sigma^1,\dots,\sigma^l)$ is exponentially small by (3.29). To treat $I_N^4(\sigma^1,\dots,\sigma^l)$, we note that for any $\epsilon > 0$ one can find N_0 such that for all $N \ge N_0$ and all $\vec{t}$ with $\|\vec{t}\| \le \delta\sqrt{N}$ the quantity in square brackets is smaller than ϵ in absolute value, and apply again (3.29). Finally, we estimate

$$|I_N^5(\sigma^1,\dots\sigma^l)| \le (2\pi)^{-l(k-1)} \int_{\|\vec{t}\|>\delta\sqrt{N}} |f_N^{\sigma^1,\dots,\sigma^l}(\{t_{\beta,j}\})| \prod_{\substack{j=1,\dots,l,\\ \beta=1,\dots,k-1}} dt_{\beta,j}. \tag{3.31}$$

For any $\sigma^1,\dots,\sigma^l \in \mathcal{R}_{l,N}^{\alpha}$ the matrix $C(\sigma^1,\dots,\sigma^l)$ contains all k^l possible different rows and by (3.8) $f_N^{\sigma^1,\dots,\sigma^l}(\{t_{\beta,j}\})$ is the product of k^l different characteristic functions, where each is taken to the power $N/k^l(1+o(1))$. Let us fix from a set of k^l rows of $C(\sigma^1,\dots,\sigma^l)$ any $(k-1)l$ linearly independent and denote by $\bar{C}$ the matrix composed by them. There exists $\eta(\delta) > 0$ such that $\sqrt{\vec{t}\bar{C}^T\bar{C}\vec{t}/(2(1/k)\mathrm{var}X)} \ge \eta$ for all $\vec{t}$ with $\|\vec{t}\| > \delta$. Changing variables $\vec{s} = \vec{t}\bar{C}^T/\sqrt{2(N/k)\mathrm{var}X}$ one gets the bound

$$\begin{aligned} |I_N^5(\sigma^1,\dots,\sigma^5)| \le &(2\pi)^{-l(k-1)}(2(N/k)\mathrm{var}X)^{l(k-1)/2}(\det\bar{C})^{-1} \\ &\times \int_{\|\vec{s}\|>\eta} \prod_{\substack{\beta=1,\dots,k-1,\\ j=1,\dots,l}} \left|\frac{e^{is_{\beta,j}}-1}{is_{\beta,j}}\right|^{Nk^{-l}(1+o(1))} ds_{\beta,j} \\ &\le CN^{l(k-1)/2}(1-h(\eta))^{Nk^{-l}(1+o(1))-2} \int_{\|\vec{s}\|>\eta} \prod_{\substack{\beta=1,\dots,k-1,\\ j=1,\dots,l}} \left|\frac{e^{is_{\beta,j}}-1}{is_{\beta,j}}\right|^2 ds_{\beta,j}, \end{aligned} \tag{3.32}$$

where $h(\eta) > 0$ is chosen such that $|(e^{is} - 1)/s| < 1 - h(\eta)$ for all s with $|s| > \eta/((k-1)l)$ and C is a constant independent of the set $\sigma^1, \dots, \sigma^l$ and N. Thus $I_N^5(\sigma^1, \dots, \sigma^l) \to 0$, uniformly for $\sigma^1, \dots, \sigma^l \in \mathcal{R}_{l,N}^{\alpha}$, and exponentially fast as $N \to \infty$. This concludes the proof of (3.20) and of Proposition 3.1. $\diamondsuit$

4. Proofs of Lemmas 3.2, 3.3, 3.4, 3.5.

Proof of Lemma 3.2. Let first $l = 1$. Without loss of generality we may assume that the first k rows of $C(\sigma^1)$ are different. Then for all $i = 1, \dots, k-1$, the following system of equations has a solution:

$$\begin{aligned} \lambda_1^{(i)} &= \tilde{\sigma}_1^{(i)} \\ -\lambda_1^{(i)} + \lambda_2^{(i)} &= \tilde{\sigma}_2^{(i)} \\ -\lambda_2^{(i)} + \lambda_3^{(i)} &= \tilde{\sigma}_3^{(i)} \\ &\dots \\ -\lambda_{k-2}^{(i)} + \lambda_{k-1}^{(i)} &= \tilde{\sigma}_{k-1}^{(i)} \\ -\lambda_{k-1}^{(i)} &= \tilde{\sigma}_k^{(i)}. \end{aligned} \tag{4.1}$$

Then necessarily $\sum_{n=1}^{k} \tilde{\sigma}_n^{(i)} = 0$ for all $i = 1, 2, \dots, k-1$, since the sum of the left-hand sides of these equations equals 0. But for at least one $j \in \{1, \dots, k\}$ and $i = 1, \dots, k-1$, $\sigma_j^{(i)} \neq 0$, for otherwise $\lambda_s^{(i)} = 0$ for all $s = 1, \dots, k$ $i = 1, \dots, k-1$ and consequently $C(\tilde{\sigma})$ is composed only of zeros, which is impossible. Without loss of generality (by definition of Σ_N we may always permute spin values) we may assume that $\tilde{\sigma}_j^{(1)} \neq 0$.

We will use the following crucial property of the configurations $\tilde{\sigma}^{(1)}, \dots, \tilde{\sigma}^{(k-1)}$:

$$\tilde{\sigma}_n^{(j)} = 1 \Longrightarrow \tilde{\sigma}_n^{(j+1)} = 0, \tilde{\sigma}_n^{(j+2)} = 0, \dots, \tilde{\sigma}_n^{(k-1)} = 0. \tag{4.2}$$

$$\tilde{\sigma}_n^{(j)} = -1 \Longrightarrow \tilde{\sigma}_n^{(j+1)} = 1, \tilde{\sigma}_n^{(j+2)} = 0, \dots, \tilde{\sigma}_n^{(k-1)} = 0. \tag{4.3}$$

It follows that, for a certain number $t_1 \geq 1$ of pairs of indices $n_1^1, n_1^2, \dots, n_{t_1}^1, n_{t_1}^2 \in \{1, \dots, k\}$, we must have that $\tilde{\sigma}_{n_u^1}^{(1)} = 1$ and $\tilde{\sigma}_{n_u^2}^{(1)} = -1$, $u = 1, \dots, t_1$. We say that these $2t_1$ indices are "occupied" from the step $j = 1$ on, since, by (4.2) and (4.3), we know all values $\tilde{\sigma}_{n_u^1}^{(j)} = 0$ for all $j = 2, 3, \dots, k-1$, $\tilde{\sigma}_{n_u^2}^{(2)} = 1$, and $\tilde{\sigma}_{n_u^2}^{(j)} = 0$ for all $j = 3, \dots, k-1$, $u = 1, 2, \dots, t_1$. We say that the other $k - 2t_1$ indices are "free" at step $j = 1$. Then we must attribute to at least t_1 of $k - 2t_1$ spins $\tilde{\sigma}_n^{(2)}$ with "free" indices the value $\tilde{\sigma}_n^{(2)} = -1$ in order to ensure that $\sum_{n=1}^{k} \tilde{\sigma}_n^{(2)} = 0$. We could also attribute to a certain number $t_2 \geq 0$ of pairs of the remaining

$k-3t_1$ spins with "free" indices the values $\tilde{\sigma}_n^{(2)} = \pm 1$. Thus by (4.2), (4.3) for $j=2$ we know the values of $\tilde{\sigma}_n^{(j)}$ for $j=2,3,\ldots,k-1$ for at least $3t_1+2t_2$ indices n. We say that they are "occupied" from $j=2$ on. Among them $\tilde{\sigma}_n^{(3)}=1$ for the number of indices t_1+t_2 and $\tilde{\sigma}_n^{(3)}=0$ for the others $2t_1+t_2$. Then we should assign to the number t_1+t_2 of the remaining $k-3t_1-2t_2$ spins $\tilde{\sigma}_n^{(3)}$ with "free" indices the value $\tilde{\sigma}_n^{(3)}=-1$ to make $\sum_{n=1}^k \sigma_n^{(3)}=0$. We could also attribute to a certain number $t_3 \geq 0$ of pairs of the remaining $k-4t_1-3t_2$ spins the values ± 1. Hence, after the third step, $4t_1+3t_2+2t_3$ indices are "occupied" etc. Finally, after $(j-1)$th step, $jt_1+(j-1)t_2+\ldots+2t_{j-1}$ indices are "occupied", $\tilde{\sigma}_n^{(j)}=1$ for $t_1+\cdots+t_{j-1}$ among these indices, and at the jth step we must put $\tilde{\sigma}_n^{(j)}=-1$ for the same number $t_1+t_2+\ldots+t_{j-1}$ of "free" indices to ensure that $\sum_{n=1}^k \tilde{\sigma}_n^{(j)}=0$. But, if $t_1>1$ or $t_1=1$ but $t_i>0$ for some $2 \leq i \leq k-1$, then, for some $j \leq k-1$, we have

$$k-jt_1-(j-1)t_2-\cdots-2t_{j-1} < t_1+t_2+\cdots+t_{j-1}.$$

(In fact, for $j=k-1$, if, $t_1>1$, then obviously $k-(k-1)t_1<t_1$, and if $t_1=1$ but $t_i>0$ we have $k-(k-1)-2<1$). This means that at the jth step there are not enough "free" indices among the remaining $k-jt_1-(j-1)t_2-\ldots-2t_{j-1}$ ones such that we could assign $\tilde{\sigma}_n^{(j)}=-1$ to ensure $\sum_{n=1}^k \tilde{\sigma}_n^{(j)}=0$. Hence, the only possibility is $t_1=1$ and $t_2=t_3=\cdots=t_{k-1}=0$.

So, at the first step 2 indices get "occupied" and at each step one more index is "occupied". Thus there exists a sequence of k different indices $n_1,n_2,\ldots,n_k \in \{1,\ldots,k\}$ such that $\sigma_{n_i}^{(i)}=1$, $\sigma_{n_{i+1}}^{(i)}=-1$, $\sigma_n^{(i)}=0$ for $n \neq n_i,n_{i+1}$, $i=1,\ldots,k-1$. Solving the system (4.1), we see that $\lambda_{n_i}^{(i)}=\lambda_{n_i+1}^{(i)}=\cdots=\lambda_{n_{i+1}-1}=1$, $\lambda_n^{(i)}=0$ for $n \neq n_i,\ldots,n_{i+1}-1$. Hence, the configuration $\tilde{\sigma}$ is a permutation of the configuration σ^1 such that $\tilde{\sigma}_n=i$, iff $\sigma_{n_i}^1=i$, $i=1,\ldots,k$.

Let us now turn to the case $l>1$. We use induction. Consider k^{l-1} possible columns. We denote linear combinations of them by $\Lambda_\alpha^{(i)}$, $\alpha=1,\ldots,k^{l-1}$. Then for any $i=1,\ldots,k-1$, the following system should have a solution

$$\begin{aligned}
\Lambda_\alpha^{(i)}+\lambda_1^{(i)} &= \tilde{\sigma}_{1,\alpha}^{(i)}\\
\Lambda_\alpha^{(i)}-\lambda_1^{(i)}+\lambda_2^{(i)} &= \tilde{\sigma}_{2,\alpha}^{(i)}\\
\Lambda_\alpha^{(i)}-\lambda_2^{(i)}+\lambda_3^{(i)} &= \tilde{\sigma}_{3,\alpha}^{(i)}\\
\cdots &= \cdots\\
\Lambda_\alpha^{(i)}-\lambda_{k-2}^{(i)}+\lambda_{k-1}^{(i)} &= \tilde{\sigma}_{k-1,\alpha}^{(i)}\\
\Lambda_\alpha^{(i)}-\lambda_{k-1}^{(i)} &= \tilde{\sigma}_{k,\alpha}^{(i)}.
\end{aligned} \tag{4.4}$$

It follows that

$$
\begin{aligned}
2\lambda_1^{(i)} - \lambda_2^{(i)} &= \tilde\sigma_{1,\alpha}^{(i)} - \tilde\sigma_{2,\alpha}^{(i)}\\
-\lambda_1^{(i)} + 2\lambda_2^{(i)} - \lambda_3^{(i)} &= \tilde\sigma_{2,\alpha}^{(i)} - \tilde\sigma_{3,\alpha}^{(i)},\\
\cdots &= \cdots\\
-\lambda_{k-2}^{(i)} + 2\lambda_{k-1}^{(i)} &= \tilde\sigma_{k-1,\alpha}^{(i)} - \tilde\sigma_{k,\alpha}^{(i)}.
\end{aligned}
\tag{4.5}
$$

Given $\tilde\sigma_{1,\alpha}^{(i)} - \tilde\sigma_{2,\alpha}^{(i)}, \dots, \tilde\sigma_{k-1,\alpha}^{(i)} - \tilde\sigma_{k,\alpha}^{(i)}$, this system (4.5) of $k-1$ equations has a unique solution, which does not depend on $\alpha = 1, \dots, k^{l-1}$. Then $\tilde\sigma_{1,\alpha}^{(i)} - \tilde\sigma_{2,\alpha}^{(i)}, \dots, \tilde\sigma_{k-1,\alpha}^{(i)} - \tilde\sigma_{k,\alpha}^{(i)}$ should not depend on α neither. We denote by $\delta_j^{(i)} = \tilde\sigma_{j,\alpha}^{(i)} - \tilde\sigma_{j+1,\alpha}^{(i)}$.

Let us consider two cases. In the first case we assume that, for some $i = 1, \dots, k-1$ and for some $j = 1, \dots, k-1$, $\delta_j^{(i)} \neq 0$. Then it may take values $\pm 1, \pm 2$. Knowing each of these values, we can reconstruct in a unique way $\tilde\sigma_{j,\alpha}^{(i)} = \tilde\sigma_j^{(i)}$ and $\tilde\sigma_{j+1,\alpha}^{(i)} = \tilde\sigma_{j+1}^{(i)}$, which do not depend on α. (If $\delta_j^{(i)} = 1$, then $\tilde\sigma_j^{(i)} = 1$ and $\tilde\sigma_{j+1}^{(i)} = 0$, if $\tilde\delta_j^{(i)} = -1$, then $\tilde\sigma_j^{(i)} = 0$ and $\tilde\sigma_{j+1}^{(i)} = -1$ etc.). Then we can reconstruct the values $\tilde\sigma_{t,\alpha}^{(i)} = \tilde\sigma_j^{(i)} + \sum_{m=t}^{j-1} \delta_m^{(i)}$ for $t = 1, \dots, j-1$, $\tilde\sigma_{t,\alpha}^{(i)} = \tilde\sigma_j^{(i)} - \sum_{m=j}^{t-1} \delta_j^{(i)}$ for $t = j+1, \dots, k$, which consequently do not depend on α. Since the sum of all k^l left-hand sides of equations (4.4) equals zero, it follows that $\sum_\alpha \sum_{j=1}^k \tilde\sigma_{j,\alpha}^{(i)} = 0$. But, since $\tilde\sigma_{j,\alpha}^{(i)} = \tilde\sigma_j^{(i)}$, it follows that $\sum_{j=1}^k \tilde\sigma_j^{(i)} = 0$. Thus, $\Lambda_\alpha^{(i)} = \frac{1}{k}\sum_{j=1}^k \sigma_{j,\alpha}^{(i)} = \frac{1}{k}\sum_{j=1}^k \tilde\sigma_j^{(i)} = 0$ for all α.

The sequence $\tilde\sigma_1^{(i)}, \dots, \tilde\sigma_k^{(i)}$ being not constant and $\sum_{j=1}^k \tilde\sigma_j^{(i)} = 0$, it follows that for some j_1, j_2, $\tilde\sigma_{j_1}^{(i)} = 1$ and $\tilde\sigma_{j_2}^{(i)} = -1$. Using (4.2) and (4.3), we see that $\tilde\sigma_{j_1}^{(i+1)} = 0$ and $\tilde\sigma_{j_2}^{(i+1)} = 1$. Therefore, for some $j = 1, \dots, k-1$ $\delta_j^{(i+1)} \neq 0$, so that we may apply the previous reasoning to the configuration $\tilde\sigma^{(i+1)}$. We get that the values $\tilde\sigma_{j,\alpha}^{(i+1)}$ do not depend on α and that $\Lambda_\alpha^{(i+1)} = 0$, for all α. Applying the analogues of (4.2) and (4.3) backwards, namely

$$
\tilde\sigma_n^{(j)} = -1 \Longrightarrow \tilde\sigma_n^{(j-1)} = 0, \tilde\sigma_n^{(j-2)} = 0, \dots, \tilde\sigma_n^{(1)} = 0, \tag{4.6}
$$

$$
\tilde\sigma_n^{(j)} = 1 \Longrightarrow \tilde\sigma_n^{(j-1)} = -1, \tilde\sigma_n^{(j-2)} = 0, \dots, \tilde\sigma_n^{(1)} = 0, \tag{4.7}
$$

we find that $\tilde\sigma_{j_1}^{(i-1)} = -1$ and $\tilde\sigma_{j_2}^{(i-1)} = 0$. Thus, for some $j = 1, \dots, k-1$, $\delta_j^{(i-1)} \neq 0$ and so we may apply the previous reasoning to the configuration $\tilde\sigma^{(i-1)}$. Hence, $\tilde\sigma_{j,\alpha}^{(i-1)}$ does not depend on α and $\Lambda_\alpha^{(i-1)} = 0$ for all α. Continuing this reasoning subsequently for $\tilde\sigma^{(i+2)}, \dots, \tilde\sigma^{(k)}$ and backwards for $\tilde\sigma^{(i-2)}, \dots, \tilde\sigma^{(1)}$, we derive that none of the values $\tilde\sigma_{j,\alpha}^{(s)}$ depends on α and that $\Lambda_\alpha^{(s)} = 0$ for all α and all $s = 1, \dots, k$. But the system $\Lambda_\alpha^{(s)} = 0$ for all $s = 1, \dots, k-1$ and $\alpha = 1, \dots, k^{l-1}$ has only the trivial solution. Hence the system (4.4) becomes the system (4.1). Invoking the reasoning for $l = 1$, we derive that $\tilde\sigma$ is a permutation of the last configuration σ^l.

Let us now turn to the second case, that is assume that for all i,j $\delta_j^{(i)}=0$. Then the unique solution of (4.5) is $\lambda_1^{(i)}=\cdots=\lambda_{k-1}^{(i)}=0$. Then $\tilde\sigma_{1,\alpha}^{(i)}=\tilde\sigma_{2,\alpha}^{(i)}=\cdots=\tilde\sigma_{k,\alpha}^{(i)}=\tilde\sigma_\alpha^{(i)}$ for all $\alpha=1,\ldots,k^{l-1}$ and all $i=1,\ldots,k-1$. The system (4.4) is reduced to a smaller system $\Lambda_\alpha^{(i)}=\tilde\sigma_\alpha^{(i)}$ corresponding to the matrix $C(\sigma^1,\ldots,\sigma^{l-1})$ with all k^{l-1} different columns. The statement of the lemma holds for it by induction. Thus in this case $\tilde\sigma$ is a permutation of one of $\sigma^1,\ldots,\sigma^{l-1}$. $\diamondsuit$

Proof of Lemma 3.3. Let us remove from the matrix $C(\sigma^1,\ldots,\sigma^l)$ linearly dependent columns and leave only r columns of the basis. They correspond to a certain subset of r configurations $\sigma^{j,(\beta)}$ $j,\beta\in A_r\subset\{1,\ldots,l\}\times\{1,\ldots,k-1\}$, $|A_r|=r$. We denote by $\bar C^r(\sigma^1,\ldots,\sigma^l)$ the $N\times r$ matrix composed by them. Then the probability in the right-hand side of (3.12) is not greater than the probability of the same events for $j,\beta\in A_r$ only. Let $\bar f_N^{\sigma^1,\ldots,\sigma^l}(\{t_{\beta,j}\})$, $j,\beta\in A_r$, be the characteristic function of the vector $(2(N/k)\mathrm{var}\,X)^{-1/2}\{Y^\beta(\sigma^j)\}_{j,\beta\in A_r}$. Then

$$\mathbb{P}\Big(\forall\beta=1,\ldots,k-1,\forall j=1,\ldots,l\frac{|Y^\beta(\sigma^j)|}{\sqrt{2(N/k)\mathrm{var}\,X}}<\frac{c_j^\beta}{S(k,N)^{\frac{1}{k-1}}}\Big)$$
$$\leq\frac{1}{(2\pi)^r}\lim_{D\to\infty}\int\limits_{[-D,D]^r}\bar f_N^{\sigma^1,\ldots,\sigma^l}(\{t_{\beta,j}\})\prod_{j,\beta\in A_r}\frac{e^{it_{\beta,j}c_j^\beta S(k,N)^{\frac{-1}{k-1}}}-e^{-it_{\beta,j}c_j^\beta S(k,N)^{\frac{-1}{k-1}}}}{it_{\beta,j}}dt_{\beta,j}. \tag{4.8}$$

To bound the integrand in (4.8) we use that

$$\left|\frac{e^{it_{\beta,j}c_j^\beta S(k,N)^{\frac{-1}{k-1}}}-e^{-it_{\beta,j}c_j^\beta S(k,N)^{\frac{-1}{k-1}}}}{it_{\beta,j}}\right|\leq\min\Big(2c_j^\beta S(k,N)^{\frac{-1}{k-1}},2(t_{\beta,j})^{-1}\Big). \tag{4.9}$$

Next, let us choose in the matrix $\bar C^r(\sigma^1,\ldots,\sigma^l)$ any r linearly independent rows and construct of them a $r\times r$ matrix $\bar C^{r\times r}$. Then

$$\begin{aligned}|\bar f_N^{\sigma^1,\ldots,\sigma^l}(\{t_{\beta,j}\})|&=\prod_{n=1}^N\left|\mathbb{E}\exp\Big(\frac{iX_n}{\sqrt{2(N/k)\mathrm{var}\,X}}\{\bar C^r(\sigma^1,\ldots,\sigma^l)\vec t\}_n\Big)\right|\\&\leq\prod_{s=1}^r\left|\mathbb{E}\exp\Big(\frac{iX_s}{\sqrt{2(N/k)\mathrm{var}\,X}}\{\bar C^{r\times r}\vec t\}_s\Big)\right|\\&\leq\prod_{s=1}^r\min\Big(1,2\sqrt{2(N/k)\mathrm{var}\,X}(\{\bar C^{r\times r}\vec t\}_s)^{-1}\Big),\end{aligned} \tag{4.10}$$

where $\vec t=\{t_{\beta,j}\}_{j,\beta\in A_r}$. Hence, the absolute value of the integral (4.8) is bounded by the sum of two terms

$$\frac{S(N,k)^{\frac{-r}{k-1}}\prod_{\beta,j\in A_r}(2c_\beta^j)}{(2\pi)^r}\int\limits_{\|\vec t\|<S(k,N)^{\frac{1}{k-1}}}\prod_{s=1}^r\min\Big(1,2\sqrt{2(N/k)\mathrm{var}\,X}(\{\bar C^{r\times r}\vec t\}_s)^{-1}\Big)dt_{\beta,j}$$

$$+\frac{1}{(2\pi)^r}\int\limits_{\|\vec{t}\|>S(k,N)^{\frac{1}{k-1}}}\prod_{\beta,j\in A_r}(2(t_{\beta,j})^{-1})\prod_{s=1}^{r}\min\left(1,2\sqrt{2(N/k)\mathrm{var}\,X}(\{\bar{C}^{r\times r}\vec{t}\}_s)^{-1}\right)dt_{\beta,j}. \tag{4.11}$$

The change of variables $\vec{\eta}=\bar{C}^{r\times r}\vec{t}$ in the first term shows that the integral over $\|\vec{t}\|<S(k,N)^{\frac{1}{k-1}}$ is at most $O(N^{r/2}(\ln S(k,N)^{\frac{1}{k-1}})^r)$, where $\ln S(k,N)^{\frac{1}{k-1}} = O(N)$ as $N\to\infty$. Thus the first term of (4.11) is bounded by $K_1(\bar{C}^{r\times r},k,l)N^{3r/2}S(k,N)^{\frac{-r}{k-1}}$ with some constant $K_1(\bar{C}^{r\times r},k,l)>0$ independent of N. Using the change of variables $\vec{\eta}=S(k,N)^{\frac{-1}{k-1}}\vec{t}$ in the second term of (4.11), one finds that the integral over $\|\vec{t}\|>S(k,N)^{\frac{1}{k-1}}$ is at most $O(S(k,N)^{\frac{-r}{k-1}})$. Thus (4.11) is not greater than $K_2(\bar{C}^{r\times r},k,l)N^{3r/2}S(k,N)^{\frac{-r}{k-1}}$ with some positive constant $K_2(\bar{C}^{r\times r},k,l)$ independent of N.

To conclude the proof, let us recall the fact that there is a finite, i.e. N-independent, number of possibilities to construct the matrix $\bar{C}^{r\times r}$ starting from $C(\sigma^1,\dots,\sigma^l)$ since each of its elements may take only three values $\pm1,0$. Thus there exists less than 3^{r^2} different constants $K_2(\bar{C}^{r\times r},k,l)$ corresponding to different matrices $\bar{C}^{r\times r}$. It remains to take the maximal one over them to get (3.12). $\diamondsuit$

Proof of Lemma 3.4. Throughout the proof we denote by $D_N\cup C(\sigma)$ the matrix D_N completed by the rows $\sigma^{(1)},\dots,\sigma^{(k-1)}$.

Let us denote by $c^1,\dots,c^q$ the system of columns of the matrix D_N. Then we can find the indices $i_1<i_2<\dots<i_{k-t-1}\le k-1$ such that $\sigma^{(i_s)}$ is a linear combination of $c^1,\dots,c^q,\sigma^{(1)},\sigma^{(i_2)},\dots,\sigma^{(i_s-1)}$ for all $s=1,\dots,k-t-1$. Then there exist linear coefficients $a_1(s),\dots,a_{i_s-1}(s)$ such that

$$a_1(s)c^1+\dots+a_q(s)c^q+a_{q+1}(s)\sigma^{(1)}+\dots+a_{i_s-1}(s)\sigma^{(i_s-1)}=\sigma^{(i_s)},\ s=1,\dots,k-t-1. \tag{4.12}$$

(If $r<q$ these coefficients may be not unique, but this is not relevant for the proof.) Since $t\ge1$, without loss of generality (otherwise just make a permutation of spin values $\{1,\dots,k\}$ in σ) we may assume that $i_1>1$.

Initially each of $k-t-1$ systems (4.12) consists of N linear equations. But the number of different rows of D_N being a fixed number R, each of these $k-t-1$ systems (4.12) has only a finite number of *different* equations. Thus, (4.12) are equivalent to $k-t-1$ *finite* (i.e. N-independent) systems of different equations of the form:

$$a_1(s)d_1+\dots+a_q(s)d_q=a_{q+1}(s)\delta_1+\dots+a_{i_s-1}(s)\delta_{i_s-1}+\delta_{i_s}, \tag{4.13}$$

where $d=(d_1,\dots,d_q)$ is one of the R distinct rows of the matrix D_N and $\delta_j=0,1,-1$.

Note that there exist at most $R \times 3^s$ of such equations (4.13) for any $s = 1, \dots, k-t-1$. Consequently, for the given matrix D_N, there exists a *finite* (i.e. N-independent) number of such sets of $k-t-1$ finite systems of distinct equations (4.13). We will denote by $\mathcal{A}$ the set of such sets of $k-t-1$ finite systems of distinct equations (4.13) which do arise from some choice of a spin configuration σ with rank $[D_N \cup C(\sigma)] = r+t$, after the reduction of (4.12) (i.e. after eliminating the same equations among all N in each of $k-t-1$ systems (4.12)). For $\sigma \in \Sigma_N$, we denote by $\alpha(\sigma) \in \mathcal{A}$ the set of $k-t-1$ finite systems of distinct equations (4.13) obtained from (4.12) in this way.

We will prove that for any given element $\alpha \in \mathcal{A}$ we have the estimate:

$$\#\{\sigma : \mathrm{rank}[D_N \cup C(\sigma)] = r+t, \quad \alpha(\sigma) = \alpha\} \leq C \frac{(N(t+1)/k)!}{((N/k)!)^{t+1}} \tag{4.14}$$

where C is a constant that depends only on R, t, k. Since the cardinality of $\mathcal{A}$ is finite and depends only on R, t, and k, this will prove the lemma.

Consider some $\alpha_0 \in \mathcal{A}$. Since by definition of $\mathcal{A}$ there exists σ_0 with the property $\mathrm{rank}[D_N \cup C(\sigma_0)] = r+t$ and $\alpha(\sigma_0) = \alpha_0$, then there exists a solution of all these $k-t-1$ systems of equations α_0. Let $a_i(s)$ be any such solution. For any row $d = (d_1, \dots, d_q)$ of D_N, set

$$\Lambda(s, d) = a_1(s)d_1 + a_2(s)d_2 + \cdots a_q(s)d_q. \tag{4.15}$$

Then to any row d of D_N there corresponds the vector of linear combinations $\Lambda(d) = (\Lambda(1, d), \Lambda(2, d), \dots, \Lambda(k-1-t, d))$. Next, let us divide the set $\mathcal{D}$ of the R different rows of the matrix D_N into m disjoint non-empty subsets $\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_m$ such that two rows $d, \tilde{d}$ are in the same subset, if and only if $\Lambda(d) = \Lambda(\tilde{d})$.

Lemma 4.1: *The partition $\mathcal{D}_i$ defined above satisfies the following properties:*

(i) $m \geq k-t$

(ii) For any pair $d \in \mathcal{D}_i$, $\tilde{d} \in \mathcal{D}_j$, with $i \neq j$, and for any σ, such that $\mathrm{rank}[D_N \cup C(\sigma)] = r+t$ *and $\alpha(\sigma) = \alpha_0$, the rows d and $\tilde{d}$ can not be continued by the same row O of the matrix $C(\sigma)$ in $D_N \cup C(\sigma)$.*

Proof. Let us first show that $\mathcal{D}$ can be divided into three non-empty subsets $\mathcal{D}_0, \mathcal{D}_1, \mathcal{D}_2$, such that $\Lambda(1, d) \neq -1, 0$ for all $d \in \mathcal{D}_0$, $\Lambda(1, d) = -1$ for $d \in \mathcal{D}_1$, $\Lambda(1, d) = 0$ for $d \in \mathcal{D}_2$. First of all, since $\alpha_0 \in \mathcal{A}$, then there exists at least one σ_0 such that $\mathrm{rank}[D_N \cup C(\sigma_0)] = r+t$ and $\alpha(\sigma_0) = \alpha_0$. Let $d^0, \dots, d^{k-1}$ denote k rows (not necessarily different) of D_N that are continued by the rows $O_0, \ldots, O_{k-1}$ of the matrix $C(\sigma_0)$ (recall the definition given in the paragraph following (3.7)) respectively in $D_N \cup C(\sigma_0)$. Now consider a row d^{i_1} that is continued by the row O_{i_1}. The corresponding equation (4.12) with $s = 1$ then reads

$$\Lambda(1, d^{i_1}) = -1.$$

This shows that the set $\mathcal{D}_1 \neq \emptyset$. Similarly, for a row d^j continued by the row O_j with $j > i_1$, the corresponding equation yields

$$\Lambda(1, d^j) = 0.$$

Thus $\mathcal{D}_2 \neq \emptyset$. Finally, consider the rows of the matrix D_N continued by $O_0, \ldots, O_{i_1-1}$. The corresponding i_1 equations (4.12) with $s = 1$ then read :

$$\begin{aligned} \Lambda(d^0, 1) &= -a_{q+1} \\ \Lambda(d^1, 1) &= a_{q+1} - a_{q+2} \\ \Lambda(d^2, 1) &= a_{q+2} - a_{q+3} \\ \cdots &= \cdots \\ \Lambda(d^{i_1-1}, 1) &= a_{q+i_1-1} + 1. \end{aligned} \tag{4.16}$$

The sum of the right-hand sides of these equations equals 1. Thus the left-hand side of at least one equation must be positive. Hence, there exists d^j with $j \in \{1, \ldots, i_1 - 1\}$ such that

$$\Lambda(d^j, 1) \neq -1, 0.$$

Thus also $\mathcal{D}_0 \neq \emptyset$, and so all three sets defined above are non-empty. Moreover, $\mathcal{D}_2$ includes all rows d that are continued by the rows O_j with $j > i_1$ of $C(\sigma_0)$.

Now, let us divide $\mathcal{D}_2$ into two non-empty subsets $\mathcal{D}_{2,1}$, $\mathcal{D}_{2,2}$ according to the value taken by $\Lambda(2, d)$. We define $\mathcal{D}_{2,1} \equiv \{d \in \mathcal{D}_2 : \Lambda(2, d) \neq 0\}$, and $\mathcal{D}_{2,2} \equiv \{d \in \mathcal{D}_2 : \Lambda(2, d) = 0\}$. Note that the row d^{i_2} is an element of $\mathcal{D}_2$ by the observation made above, while using (4.12) with $s = 2$, we get, as before that $\Lambda(2, d^{i_2}) = -1$, and for all $j > i_2$, $\Lambda(2, d^j) = 0$. Thus $\mathcal{D}_{2,1}$ and $\mathcal{D}_{2,2}$ are non-empty. In addition to that, for any row d continued by O_j with $j > i_2$ we have again by (4.12) with $s = 2$ $\Lambda(2, d^j) = 0$. Hence, $\mathcal{D}_{2,1}$ and $\mathcal{D}_{2,2}$ are non-empty, and $\mathcal{D}_{2,2}$ contains all rows d continued by O_j with $j > i_2$ of $C(\sigma_0)$.

Using (4.12) for $s = 3$ we can again split $\mathcal{D}_{2,2}$ into two non-empty subsets $\mathcal{D}_{2,2,1}$ with $\Lambda(d, 3) \neq 0$ and $\mathcal{D}_{2,2,2}$ with $\Lambda(d, 3) = 0$. Furthermore, $\mathcal{D}_{2,2,2}$ contains all rows that are continued by O_j with $j > i_3$ of $C(\sigma_0)$, etc. The same procedure can be repeated up to the step $s = k-1-t-1$. In this way we have subdivided $\mathcal{D}_2$ into $k-1-t-1$ disjoint non-empty subsets. Together with $\mathcal{D}_0$ and $\mathcal{D}_1$, these constitute $k-t$ disjoint subsets $\mathcal{D}_i$. This proves the assertion (i).

Let us now take any σ such that rank $[D_N \cup C(\sigma)] = r+t$ and with $\alpha(\sigma) = \alpha_0$. Assume that d and $\tilde{d}$ are continued by the same row O_j of $C(\sigma)$ in $D_N \cup C(\sigma)$. Since d and $\tilde{d}$ belong to different subsets $\mathcal{D}_i$, for some $u \in \{1, \ldots, k-t-1\}$, $\Lambda(d,u) \neq \Lambda(\tilde{d},u)$. Then, writing (4.12) with $s = u$ along the row d continued by O_j and along the row $\tilde{d}$ continued by O_j we would get either the system

$$\Lambda(d,u) = 0$$
$$\Lambda(\tilde{d},u) = 0$$

if $j > i_u$, or

$$\Lambda(d,u) = -1$$
$$\Lambda(\tilde{d},u) = -1$$

if $j = i_u$, or

$$\Lambda(d,u) - a_{q+j}(u) = 1$$
$$\Lambda(\tilde{d},u) - a_{q+j}(u) = 1$$

if $j = i_u - 1$, or finally

$$\Lambda(d,u) - a_{q+j}(u) + a_{q+j+1}(u) = 0$$
$$\Lambda(\tilde{d},u) - a_{q+j}(u) + a_{q+j+1}(u) = 0,$$

if $j < i_u - 1$. But no one of these four systems has a solution if $\Lambda(d,u) \neq \Lambda(\tilde{d},u)$. This proves (ii). $\diamondsuit$

By (ii) of Lemma 4.1, for any σ such that rank $[D_N \cup C(\sigma)] = r+t$ and $\alpha(\sigma) = \alpha_0$ the set of rows of the matrix D_N is divided into $m \geq k-t$ non-empty disjoint subsets $\mathcal{D}_1, \ldots, \mathcal{D}_m$ and the set of k rows of the matrix $C(\sigma)$ is divided into m non-empty disjoint subsets $\mathcal{C}_1, \ldots, \mathcal{C}_m$ of cardinalities $s_1, \ldots, s_m \geq 1$, respectively, such that the rows in $\mathcal{C}_j$ continue the rows of $\mathcal{D}_j$ only. But s_j rows of the matrix $C(\sigma)$ must be present Ns_j/k times. Thus, first of all, in the matrix D_N, these r_j rows must be present Ns_j/k times as well, for all $j = 1, \ldots, m$. Thus, the number of configurations σ with rank $[D_N \cup C(\sigma)] = r+t$ such that $\alpha(\sigma) = \alpha_0$ does not exceed $\prod_{j=1}^m \binom{Ns_j/k}{N/k}\binom{N(s_j-1)/k}{N/k}\cdots\binom{N/k}{N/k} = ((N/k)!)^{-k}\prod_{j=1}^m (Ns_j/k)!$ which is bounded by $((N/k)!)^{-k}((N(k-m+1)/k)!)((N/k)!)^{m-1}$ for any $s_1, \ldots, s_m \geq 1$ with $s_1 + \cdots + s_m = k$.

By (i) of Lemma 4.1 we have $k-t \leq m \leq k$, so that

$$\frac{((N(k-m+1)/k)!)((N/k)!)^{m-1}}{((N/k)!)^k} = \binom{N(k-m+1)/k}{N/k}\binom{N(k-m)/k}{N/k}\cdots\binom{N/k}{N/k}$$

$$\leq \binom{N(t+1)/k}{N/k}\binom{Nt/k}{N/k}\cdots\binom{N/k}{N/k} = \frac{(N(t+1)/k)!}{((N/k)!)^{(t+1)}}.$$

Hence, for any matrix D_N composed of R different columns

$$\#\{\sigma : \text{rank } [D_N \cup C(\sigma)] = r+t,\ \alpha(\sigma)=\alpha_0\}$$

$$\leq \Big(\sum_{m=k-t}^{k} \sum_{\substack{r_1,\ldots,r_m\geq 1,\\ r_1+\cdots+r_m=R}} \sum_{\substack{s_1,\ldots,s_m\geq 1,\\ s_1+\cdots+s_m=k}} \Big)\binom{N(t+1)/k}{N/k}\binom{Nt/k}{N/k}\cdots\binom{N/k}{N/k} = C\frac{(N(t+1)/k)!}{((N/k)!)^{t+1}}. \tag{4.17}$$

$\diamondsuit$

Proof of Lemma 3.5 The statement (3.29) is an immediate consequence of (3.28) and (3.18), (3.4) if $\delta > 0$ is small enough.

The proof of (3.27) and (3.28) mimics the standard proof of the Berry-Essen inequality. Namely, we use the representation (3.8) of $f_N^{\sigma_1,\ldots,\sigma^l}(\{t_i^j\})$ as a product of N characteristic functions where at most k^l of them are different. Each of them by standard Taylor expansion

$$f_{N,n}^{\sigma^1,\ldots,\sigma^l}(\{t_{\beta,j}\}) = 1 - \frac{\Big(\sum_{\substack{i=1,\ldots,k-1\\ j=1,\ldots,l}}(\mathbb{1}_{\{\sigma_n^j=i\}} - 1_{\{\sigma_n^j=i+1\}})t_{\beta,j}\Big)^2}{4(N/k)\mathrm{var}X}\mathrm{var}X$$

$$-\theta_n\frac{i\Big(\sum_{\substack{i=1,\ldots,k-1\\ j=1,\ldots,l}}(\mathbb{1}_{\{\sigma_n^j=i\}} - 1_{\{\sigma_n^j=i+1\}})t_{\beta,j}\Big)^3}{6((2N/k)\mathrm{var}X)^{3/2}}\mathbb{E}(X-\mathbb{E}X)^3 \equiv 1-r_n \tag{4.18}$$

with $|\theta_n| < 1$. It follows that $|r_n| < C_1\|\vec{t}\|^2N^{-1} + C_2\|\vec{t}\|^3N^{-3/2}$, for some $C_1, C_2 > 0$, all $\sigma^1,\ldots,\sigma^l$, and all n. Then $|r_n| < 1/2$ and $|r_n|^2 < C_3\|\vec{t}\|^3N^{-3/2}$, for some $C_3 > 0$ and all $\vec{t}$ satisfying $\|\vec{t}\| < \delta\sqrt{N}$, with δ enough small. Thus, $\ln f_{N,n}^{\sigma^1,\ldots,\sigma^l}(\{t_{\beta,j}\}) = -r_n + \tilde{\theta}_n r_n^2/2$ (using the expansion $\ln(1+z) = z + \tilde{\theta}z^2/2$ for $\|z\| < 1/2$ with $\|\tilde{\theta}\| < 1$) , with some $|\tilde{\theta}_n| < 1$ for all $\sigma^1,\ldots,\sigma^l$, all n, and all t satisfying $\|\vec{t}\| < \delta\sqrt{N}$. It follows that $f_N^{\sigma^1,\ldots,\sigma^l}(\{t_{\beta,j}\}) = \exp(-\sum_{n=1}^N r_n + \sum_{n=1}^N \tilde{\theta}_n r_n^2/2)$. Here $-\sum_{n=1}^N r_n = -\vec{t}B_N(\sigma^1,\ldots,\sigma^l)\vec{t}/2 + \sum_{n=1}^N p_n$ where $|p_n| \leq C_2\|\vec{t}\|^3N^{-3/2}$. Then

$$f_N^{\sigma^1,\ldots,\sigma^l}(\{t_{\beta,j}\}) = e^{-\vec{t}B_N(\sigma^1,\ldots,\sigma^l)\vec{t}/2}e^{\sum_{n=1}^N(p_n+\tilde{\theta}_n r_n^2/2)}, \tag{4.19}$$

where $|p_n|+|\tilde{\theta}_n r_n^2/2| \leq (C_2+C_3/2)\|\vec{t}\|^3N^{-3/2}$. Hence $|e^{\sum_{n=1}^N(p_n+\tilde{\theta}_n r_n^2/2)}-1| \leq C_4\|\vec{t}\|^3N^{-1/2}$, for all $\vec{t}$ satisfying $\|\vec{t}\| < \epsilon N^{1/6}$ with $\epsilon > 0$ small enough. Moreover, $|\sum_{n=1}^N(p_n+\tilde{\theta}_n r_n^2/2)| \leq C_5\|\vec{t}\|^3N^{-1/2}$, which implies (3.28). This concludes the proof of Lemma 3.5 $\diamondsuit$

5. The unrestricted partioning problem.

In the previous section we considered the state space of spin configurations where the number of spins taking each of k values is exactly N/k. Here we want to discuss what happens if all partitions are permitted. Naturally, we divide again the space of all configurations $\{1,\dots,k\}^N$ into equivalence classes obtained by permutations of spins. Thus our state space $\tilde{\Sigma}_N$ has $k^N(k!)^{-1}$ elements. Let us define the random variables $Y^\beta(\sigma)$ as in the previous section, see (1.4). Then we may state the following conjecture analogous to Theorem 1.1.

Conjecture 5.1: *Let*

$$\tilde{V}^\beta(\sigma) = k^{N/(k-1)}N^{-1/2}k^{1/2}(k!)^{-1/(k-1)}\pi^{-1/2}\sqrt{3}|Y^\beta(\sigma)|, \quad \beta = 1,\dots,k-1. \tag{5.1}$$

Then the point process on $\mathbb{R}_+^{k-1}$

$$\sum_{\sigma\in\tilde{\Sigma}_N} \delta_{(\tilde{V}^1(\sigma),\dots,\tilde{V}^{k-1}(\sigma))}$$

converges to the Poisson point process on $\mathbb{R}_+^{k-1}$ *with the intensity measure which is the Lebesgue measure.*

Using Theorem 2.1, the assertion of the conjecture would be an immediate consequence of the following conjecture, that is the analogue of Proposition 3.1.

Conjecture 5.2: *Denote by* $\sum_{\sigma^1,\dots,\sigma^l\in\Sigma_N}(\cdot)$ *the sum over all possible ordered sequences of different elements of* Σ_N*. Then for any* $l = 1,2,\dots,$ *any constants* $c_j^\beta > 0$, $j = 1,\dots,l$, $\beta = 1,\dots,k-1$ *we have:*

$$\sum_{\sigma^1,\dots,\sigma^l\in\Sigma_N} \mathbb{P}\Big(\forall\beta = 1,\dots,k-1, \forall j = 1,\dots,l \frac{|Y^\beta(\sigma^j)|}{\sqrt{2(N/k)\operatorname{var} X}} < \frac{c_j^\beta}{(k!)^{-1/(k-1)}k^{N/(k-1)}}\Big)$$
$$\to \prod_{\substack{j=1,\dots,l\\ \beta=1,\dots,k-1}} \frac{2c_j^\beta\sqrt{\operatorname{var} X}}{\sqrt{2\pi\mathbb{E}(X^2)}}. \tag{5.2}$$

Remark: One can notice the difference between the right-hand sides of (3.2) and (5.2). In spite of this difference, the proof of this statement proceeds along the same lines as that of Proposition 3.1. The only point that we were not able to complete is that the sum analogous to J_N^2 in (3.10) (recall that it is a sum over sets $\sigma^1,\dots,\sigma^l$ such that the system $\{Y^\beta(\sigma^j)\}_{\substack{j=1,\dots,l,\\ \beta=1,\dots k-1}}$ is linearly dependent and, moreover, for any basis of this system there exists a number $j \in \{1, \ldots, l\}$ such that for some non-empty subset of coordinates $\beta \in \{1, \ldots, k-1\}$ the random variables $Y^{\beta}(\sigma^j)$ are included in this basis and for some non-empty subset of coordinates $\beta \in \{1, \ldots, k-1\}$ they are not included there) converges to 0 as $N \to \infty$. Therefore the whole statement remains a conjecture.

Remark: The case $k = 2$. In the case $k = 2$ the sum J_N^2 is absent. Hence, in this case we can provide an entire proof of (5.2) and therefore prove our conjecture. The result in the case $k = 2$ is not new: it has been already established by Ch. Borgs, J. Chayes and B. Pittel in [BCP], Theorem 2.8. Our Theorem 2.1 gives an alternative proof for it via (5.2).

Finally we sketch the arguments that should lead to (5.2) and explain the differences with (3.2). To start with, similarly to (3.9), we split

$$\sum_{\substack{\sigma^1, \ldots, \sigma^l \in \hat{\Sigma}_N \\ \mathrm{rank}_{C(\sigma^1, \ldots, \sigma^l) = (k-1)l}}} \mathbb{P}(\cdot) + \sum_{\substack{\sigma^1, \ldots, \sigma^l \in \hat{\Sigma}_N \\ \mathrm{rank}_{C(\sigma^1, \ldots, \sigma^l) < (k-1)l}}} \mathbb{P}(\cdot). \tag{5.3}$$

We are able to prove that the first part of (5.3) converges to the left-hand side of (5.2). For that purpose, we introduce again "the main part" of the state space with $\alpha \in (0, 1/2)$:

$$\tilde{\mathcal{R}}_{l,N}^{\alpha} = \Big\{\sigma^1, \ldots, \sigma^l \in \Sigma_N : \forall 1 \leq j \leq l, \forall 1 \leq i < r \leq l, 1 \leq \beta, \gamma, \eta \leq k, \beta \neq \gamma$$

$$\Big|\sum_{n=1}^{N} \mathbb{1}_{\sigma_n^j = \beta} - N/k\Big| < N^{\alpha}\sqrt{N}, \Big|\sum_{n=1}^{N} (\mathbb{1}_{\{\sigma_n^i = \beta\}} - \mathbb{1}_{\{\sigma_n^i = \gamma\}}) \mathbb{1}_{\{\sigma_n^r = \eta\}}\Big| < N^{\alpha}\sqrt{N}\Big\} \tag{5.4}$$

where

$$\|\tilde{\mathcal{R}}_{l,N}^{\alpha}\| \geq k^{Nl}(1 - \exp(-hN^{2\alpha}))(k!)^{-l} \tag{5.5}$$

and split the first term of (5.3) into two terms

$$\sum_{\sigma^1, \ldots, \sigma^l \in \tilde{\mathcal{R}}_{l,N}^{\alpha}} \mathbb{P}(\cdot) + \sum_{\substack{\sigma^1, \ldots, \sigma^l \notin \tilde{\mathcal{R}}_{l,N}^{\alpha} \\ \mathrm{rank}_{C(\sigma^1, \ldots, \sigma^l) = (k-1)l}}} \mathbb{P}(\cdot). \tag{5.6}$$

The second term of (5.6) converges to zero exponentially fast: the number of configurations in it is at most $O(\exp(-hN^{2\alpha})k^{Nl})$ by (5.5), while the probability $\mathbb{P}(\cdot) = O(N^l k^{-Nl})$ by the analogue of Lemma 3.3.

To treat the first term of (5.3), let us stress that an important difference compared to the previous sections is the fact that the variables $Y^{\beta}(\sigma)$ are now not necessarily centered. Namely,

$$\mathbb{E}Y^{\beta}(\sigma) = (\mathbb{E}X) \sum_{n=1}^{N} (\mathbb{1}_{\{\sigma_n = \beta\}} - \mathbb{1}_{\{\sigma_n = \beta+1\}}) = \mathbb{E}X\,[\#\{n : \sigma_n = \beta\} - \#\{n : \sigma_n = \beta + 1\}] \tag{5.7}$$

as it may happen that $\#\{n : \sigma_n = \beta\} \neq \#\{n : \sigma_n = \beta + 1\}$.

Taking this observation into account and proceeding similarly to the analysis of (3.21), we can show that, uniformly for all $\sigma^1, \dots, \sigma^l \in \tilde{\mathcal{R}}^{\alpha}_{l,N}$,

$$\mathbb{P}(\cdot) = \frac{k^{-Nl}(k!)^l \prod_{j,\beta}(2c_j^\beta)}{(2\pi)^{(k-1)l/2}} \exp\Big(- \frac{\mathbb{E}\vec{Y}_j^\beta}{\sqrt{2(N/k)\mathrm{var}X}} \frac{B^{-1}}{2} \frac{\mathbb{E}\vec{Y}_j^\beta}{\sqrt{2(N/k)\mathrm{var}X}}\Big) + o(k^{-Nl}) \quad (5.8)$$

where the matrix B consists of l diagonal blocks $(k-1)\times(k-1)$, each block having 1 on the diagonal, $-1/2$ on the line under the diagonal and 0 everywhere else. Thus the first term of (5.6) by (5.8) and (5.5) equals

$$\sum_{\sigma^1,\dots,\sigma^l \in \tilde{\mathcal{R}}^{\alpha}_{l,N}} \frac{k^{-Nl}(k!)^l \prod_{j,\beta}(2c_j^\beta)}{(2\pi)^{(k-1)l/2}} \exp\Big(- \frac{\mathbb{E}\vec{Y}_j^\beta}{\sqrt{2(N/k)\mathrm{var}X}} \frac{B^{-1}}{2} \frac{\mathbb{E}\vec{Y}_j^\beta}{\sqrt{2(N/k)\mathrm{var}X}}\Big) + o(1)$$
$$= \frac{\prod_{j,\beta}(2c_j^\beta)}{(2\pi)^{(k-1)l/2}} E_{\sigma^1,\dots,\sigma^l} \exp\Big(- \frac{\mathbb{E}\vec{Y}_j^\beta}{\sqrt{2(N/k)\mathrm{var}X}} \frac{B^{-1}}{2} \frac{\mathbb{E}\vec{Y}_j^\beta}{\sqrt{2(N/k)\mathrm{var}X}}\Big) + o(1). \quad (5.9)$$

By the Central Limit Theorem the vector $\sum_{n=1}^N (\mathbb{1}_{\{\sigma_n^j = \beta\}} - \mathbb{1}_{\{\sigma_n^j = \beta+1\}})/\sqrt{2N/k}$ on $\tilde{\Sigma}_N^{\otimes l}$ converges to a Gaussian vector Z_j^β with zero mean and covariance matrix B as $N \to \infty$. Hence, (5.9) converges to

$$\frac{\prod_{j,\beta}(2c_j^\beta)}{(2\pi)^{(k-1)l/2}} \mathbb{E}_Z \exp\Big(- \frac{\mathbb{E}X}{\sqrt{\mathrm{var}\, X}} \vec{Z}_j^\beta \frac{B^{-1}}{2} \vec{Z}_j^\beta \frac{\mathbb{E}X}{\sqrt{\mathrm{var}\, X}}\Big) = \prod_{j,\beta} \frac{2c_j^\beta \sqrt{\mathrm{var}X}}{\sqrt{2\pi}\sqrt{(\mathbb{E}X)^2 + \mathrm{var}\, X}} \quad (5.10)$$

which is the right-hand side of (5.2). This finishes the analysis of the first term of (5.3).

To treat the second term, we split it into two parts J_N^1 and J_N^2 analogously to (3.10). The analysis of J_N^1 is exactly the same as in the proof of Proposition 3.1 and relies on Lemmatas 3.3 and 3.2.

However, the problem with the sum J_N^2 persists. First of all, this sum contains much more terms than in the case of the previous section as it consists essentially of configurations $\sigma^1, \dots, \sigma^l$ where some of the values of spins β among $\{1, \dots, k\}$ figure out more often than others, i.e. $\#\{n : \sigma_n = \beta\} > \#\{n : \sigma_n = \beta + 1\}$. Lemma 3.4 is not valid anymore. Second, for all such configurations σ, the random variables $Y^\beta(\sigma)$ are not centered and consequently the estimate of the probability $\mathbb{P}(\cdot)$ suggested by Lemma 3.3 is too rough. We did not manage to complete the details of this analysis.

References.

[BFM] H. Bauke, S. Franz, and St. Mertens, Number partitioning as random energy model, *Journal of Stat. Mech.: Theor. and Exp.*, P04003 (2004).

[BBG1] G. Ben Arous, A. Bovier, V. Gayrard. Glauber Dynamics for the Random Energy Model 1. Metastable motion on extreme states. *Commun. Math. Phys.* **235**, 379–425 (2003).

[BBG2] G. Ben Arous, A. Bovier, V. Gayrard. Glauber Dynamics for the Random Energy Model 2. Aging below the critical temperature. *Commun. Math. Phys.* **235** (2003).

BCMP C. Borgs, J.T. Chayes, S. Mertens, and B. Pittel. Phase diagram for the constrained integer partitioning problem, *Random Structures Algorithms* **24**, 315–380 (2004).

[BCP] Ch. Borgs, J.T. Chayes, and B. Pittel, Phase transition and finite-size scaling for the integer partitioning problem. Analysis of algorithms (Krynica Morska, 2000). *Random Structures Algorithms* **19**, 247–288 (2001).

[BM] A. Bovier and D. Mason, Extreme value behavior in the Hopfield model, *Ann. Appl. Probab.* **11**, 91–120 (2001).

[Fe] W. Feller, An introduction to probability theory and its applications 1, Wiley, New York, 1957.

[Gal] J. Galambos, On the distribution of the maximum of random variables, *Ann. Math. Statist.* **43**, 516-521 (1972).

[JK] I. Junier and J. Kurchan, Microscopic realizations of the trap model. *J. Phys. A* **37**, 3945–3965 (2003).

[Kal] O. Kallenberg, Random measures, Akademie Verlag, Berlin, 1983.

[KP] H. Koch and J. Piasko, "Some rigorous results on the Hopfield neural network model", J. Stat. Phys. **55**, 903-928 (1989).

[LLR] M.R. Leadbetter, G. Lindgren, and H. Rootzén, Extremes and related properties of random sequences and processes, Springer, Berlin, 1983.

[Mer1] St. Mertens, Random costs in combinatorial optimization, *Phys. Rev. Letts.* **84**, 1347-1350 (2000).

[Mer2] St. Mertens, A physicist's approach to number partitioning. Phase transitions in combinatorial problems (Trieste, 1999), *Theoret. Comput. Sci.* **265**, 79–108 (2001).