Extreme Value Behavior in the Hopfield Model

Anton Bovier* and David M. Mason†
Weierstrass Institute University of Delaware

August 25, 1999

Abstract

We study a Hopfield model whose number of patterns M grows to infinity with the system size N, in such a way that $M(N)^2 \log M(N)/N$ tends to zero. In this model the unbiased Gibbs state in volume N can essentially be decomposed into $M(N)$ pairs of disjoint measures. We investigate the distributions of the corresponding weights, and show, in particular, that these weights concentrate for any given N very closely to one of the pairs, with probability tending to one. Our analysis is based upon a new result on the asymptotic distribution of order statistics of certain correlated exchangeable random variables.

AMS Subject Classifications: 82B44, 60G70, 60K35
Keywords: Hopfield model, extreme values, order statistics, metastates, chaotic size-dependence.

1 Introduction and statements of main results.

In recent work, initiated mainly by Newman and Stein [21, 24, 22, 23, 20, 25], it has emerged that in the analysis of disordered systems in statistical mechanics an important aspect is the probabilistic nature of the convergence of finite volume Gibbs states to the infinite volume limit. Most of the previous work in the field has tended to treat a disordered system, for a fixed realization of the disorder, like a particular deterministic system, ignoring the fact that the Gibbs states are actually measure valued random variables. In simple situations (dilute Ising model, random field Ising model, etc.) with only a few infinite volume Gibbs states, this approach was sufficient, since by fixing suitable boundary conditions, deterministic sequences of infinite volume Gibbs states could be constructed that converge almost surely to some infinite volume state. Newman and Stein have pointed out, however, that this naive approach could be inadequate to understand the basic features in systems with a highly

*e-mail: bovier@wias-berlin.de, partially supported by the DFG.
†e-mail: davidm@math.udel.edu, research partially supported by an NSF Grant.

complex phase structure, such as spin glasses. In particular, they argued that a suitable probabilistic description in terms of random measures ("metastates" in their terminology) could be helpful in obtaining some *a priori* information from basic principles, such as symmetries, to classify possible scenarios in different situations. On this basis they argued against the direct applicability of the mean-field picture in the Sherrington-Kirkpatrick model [26] to short-range lattice spin glasses and proposed alternative pictures.

Whenever there is some new conceptual framework, it is always important to have some concrete examples at hand that have been worked out in detail. This has been done in a number of examples, typically taken from mean field models [16, 17, 5, 7], over the last two years. They cover models with finitely many [16, 17] and infinitely many [5, 7] pure states. In the present paper we will consider the case of the standard Hopfield-model with a (not too rapidly) growing number of patterns, that is we will deal with a model with countably many pure states. The construction of the pure states, using symmetry breaking magnetic fields has been achieved some years ago in [1, 2] and many more refined results have been obtained in recent years [5, 3, 6, 4, 12, 13, 28, 29]. However, the question of the convergence of the Gibbs state without a symmetry breaking field has remained unanswered so far. As we will see, this issue is tied to the study of the order statistics of a class of dependent exchangeable random variables whose asymptotic distribution is not covered by known results in extreme value theory. The main technical tool of this paper is a powerful Gaussian distributional approximation result of Zaitsev [30, 31].

We shall begin by briefly describing the model we study (for more details and motivation, see e.g. [3]). Let $\mathcal{S}_N := \{-1,1\}^N$ denote the set of functions $\sigma : \{1,\dots,N\} \to \{-1,1\}$. We call σ a spin configuration and denote by σ_i the value of σ at i. Let $(\Omega, \mathcal{F}, P)$ be an abstract probability space and let ξ_i^μ, $i, \mu \in \mathbb{N}$, denote a family of independent identically distributed random variables on this space. For the purposes of this paper we will assume that the ξ_i^μ are *Rademacher* random variables, namely, $P\{\xi_i^\mu = \pm 1\} = \frac{1}{2}$.

We define random maps $m_N^\mu : \mathcal{S}_N \to [-1,1]$ through

$$m_N^\mu(\sigma) := \frac{1}{N}\sum_{i=1}^N \xi_i^\mu \sigma_i. \tag{1.1}$$

Naturally, these maps 'compare' the configuration σ globally to the *random configuration*

$$\xi^\mu := (\xi_1^\mu, ..., \xi_N^\mu). \tag{1.2}$$

A *Hamiltonian* is now defined as the simple negative function of these variables given by

$$H_N(\sigma) := -\frac{N}{2}\sum_{\mu=1}^{M(N)} (m_N^\mu(\sigma))^2 =: -\frac{N}{2}|m_N(\sigma)|_2^2, \tag{1.3}$$

where $M(N)$ is some, generally increasing, function that will be seen to influence crucially the properties of the model. We let $|\cdot|_2$ denote the Euclidean norm in $\mathbb{R}^M$, and the vector

$m_N(\sigma)$ is always understood to be the $M(N)$-dimensional vector with components $m_N^\mu(\sigma)$. We will always use the abbreviation

$$\alpha := \alpha(N) := \frac{M(N)}{N}. \tag{1.4}$$

Through this Hamiltonian we define in a natural way finite volume *Gibbs measures* on $\mathcal{S}_N$ via

$$d\mu_{N,\beta}(\sigma) := \frac{e^{-\beta H_N(\sigma)}}{Z_{N,\beta}} dP_\sigma, \tag{1.5}$$

where $P_\sigma = (\frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1)^{\otimes N}$ and the probability distribution on $\mathbb{R}^M$ of the overlap parameters given by

$$Q_{N,\beta} := \mu_{N,\beta} \circ m_N^{-1}, \tag{1.6}$$

where the normalizing factor $Z_{N,\beta}$, given by

$$Z_{N,\beta} := 2^{-N} \sum_{\sigma \in \mathcal{S}_N} e^{-\beta H_N(\sigma)} := E_\sigma e^{-\beta H_N(\sigma)} \tag{1.7}$$

is called the *partition function*. We are interested in the large N behavior of these measures. Note that all the objects defined above are random objects. It has been shown first in [1], and later in [3, 28], with more precise estimates, that the measure $Q_{N,\beta}$ is concentrated on the union of $2M$ disjoint balls of radius $\sim \sqrt{\alpha}$. More precisely, set

$$B_\rho(x) := \{y \in \mathbb{R}^M : |x-y|_2 \le \rho\},$$

denote by e_M^μ the $\mu - th$ unit-vector in $\mathbb{R}^M$ and let $m^* := m^*(\beta)$ be the largest solution of the equation $m = \tanh(\beta m)$. In [3] the following result was obtained:

Fact 1.1 *There exist $0 < c_0, C, \gamma_a < \infty$ such that for all $\beta > 1$, $\sqrt{\alpha} < \gamma_a (m^*)^2$, and all ρ satisfying $c_0(\frac{\sqrt{\alpha}}{m^*} \wedge N^{-1/4}) < \rho < m^*/\sqrt{2}$, we have, with probability one, for all but a finite number of indices N,*

$$Q_{N,\beta}\left(\cup_{\mu=1}^M \cup_{s=\pm1} B_\rho(s m^* e_M^\mu)\right) \ge 1 - e^{-C(M \wedge N^{1/2})}. \tag{1.8}$$

Since the balls $B_\rho(s m^* e_M^\mu)$ are disjoint, this result implies that the measure $Q_{\beta,N}$ has the asymptotic decomposition

$$Q_{N,\beta} = \sum_{\mu=1}^M Q_{N,\beta}\left(B_\rho(m^* e_M^\mu)\right)(Q_{N,\beta,\rho}^{+,\mu} + Q_{N,\beta,\rho}^{-,\mu}\right) + O(e^{-C(M \wedge N^{1/2})}), \tag{1.9}$$

where $Q_{N,\beta,\rho}^{s,\mu}$, $s = \pm 1$, denote the conditional measures

$$Q_{N,\beta,\rho}^{s,\mu}(\cdot) = Q_{N,\beta}(\cdot | x \in B_\rho(s m^* e_M^\mu)).$$

What we want to control are the relative weights of these measures, i.e. $Q_{N,\beta}(B_\rho(m^* e^\mu_M))$. In [2, 3] upper bounds on the relative fluctuations of these weights were proven using concentration of measures techniques which show that the relative weights differ by no more than a factor of order $\exp(\sqrt{N})$. However, this method gives no lower bounds on the fluctuations. Thus we must try to get some more explicit control on the form of these weights. This was done, for instance, by Gentz [12, 13] in the course of the proof of a central limit theorem. The following theorem follows easily from the estimates in Section 4.2 of [3] and is also implicit in the proof of Theorem 2.6 of [12], resp. Theorem 2.5 in [13]

Fact 1.2 *With the notation and assumptions of Fact 1.1, for some $C(\beta) > 0$ we have, with probability one, for all but a finite number of indices N, for any $\mu = 1, \dots, M(N)$,*

$$(1.10) \qquad |\log(Z_{N,\beta}Q_{N,\beta}(B_\rho(m^* e^\mu))) - \beta N\phi(m^*) - h(m^*,\beta)\sum_{\nu\neq\mu}(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i^\nu\xi_i^\mu)^2|$$

$$\leq C(\beta)\sqrt{\frac{M^3}{N}},.$$

where

$$\phi(m) := m^2/2 - \beta^{-1}\log\cosh(\beta m)$$

and

$$h(m,\beta) = \beta\frac{m^2}{2[1-\beta(1-m^2)]}.$$

(Note that the condition $M^3/N \to 0$ in the statement of the theorems in [12, 13] is necessary only to assure that the right-hand side in (1.10) vanishes, which we do not require here).

Fact 1.2 tells us that the fluctuations of the weights are governed by the explicitly given random variables (we normalize the variables appearing in (1.10) to have mean zero and variance 1)

$$(1.11) \qquad B_\mu(N,M) := \frac{1}{\sqrt{2M}}\sum_{\nu\neq\mu}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\xi_i^\nu\xi_i^\mu\right)^2 - \frac{M-1}{\sqrt{2M}},$$

provided that their relative fluctuations are large compared to $\frac{M^2}{N}$. We will in fact establish that the spacing of the largest (smallest) of the $B_\mu(N,M)$ is actually on the scale $1/\sqrt{\log M}$, provided $M \to \infty$.

To state our first main result, let us denote the standard normal distribution function by

$$(1.12) \qquad \Phi(u) := \frac{1}{\sqrt{2\pi}}\int_\infty^u e^{-x^2/2}dx$$

and its upper tail by

$$(1.13) \qquad \overline{\Phi}(u) := 1 - \Phi(u).$$

Define for $x \in \mathbb{R}$ and $M \geq 1$

$$u_M(x) := \overline{\Phi}\left(\exp(-x)/M\right). \tag{1.14}$$

It is well known that ([18], page 15)

$$u_M(x) = \widetilde{u}_M(x) + o\left(\frac{1}{\sqrt{\log M}}\right), \tag{1.15}$$

where

$$\widetilde{u}_M(x) := \frac{x}{\sqrt{2\log M}} + (2\log M)^{1/2} - \frac{\log\log M + \log(4\pi)}{2\sqrt{2\log M}}. \tag{1.16}$$

In fact, all the results we state based upon $u_M(x)$, also hold with $u_M(x)$ replaced by $\widetilde{u}_M(x)$.

Define the point process on $\mathbb{R}$ by

$$\Pi_N := \sum_{\mu=1}^{M(N)} \delta_{u_M^{-1}(B_\mu(N,M))}.$$

Theorem 1.1 *Whenever $M(N) \leq N$ satisfies $M(N) \to \infty$, the sequence of point processes $\{\Pi_N\}_{N\geq 1}$ converges weakly with respect to the vague topology to the Poisson point process Π on $\mathbb{R}$ with intensity measure $e^{-x}dx$.*

Set for $x \in \mathbb{R}$

$$m_M(x) = \#\{B_1(N,M), \dots, B_M(N,M) > u_M(x)\}. \tag{1.17}$$

Corollary 1.1 *Whenever $M(N) \leq N$ satisfies $M(N) \to \infty$, we have for all real x and $k \geq 0$*

$$P\{m_{M(N)}(x) = k\} \to \frac{\exp(-kx)}{k!}\exp(-\exp(-x)) \text{ as } N \uparrow \infty. \tag{1.18}$$

Also, as more or less a corollary of Theorem 1.1 we obtain the next result, which asserts that the weights in the decomposition (1.9) are indeed concentrated on a single (random) value of μ with probability tending to one.

Theorem 1.2 *Assume $M(N) \leq N$ satisfies $M(N) \to \infty$ and*

$$\frac{M(N)^2 \log M(N)}{N} \to 0. \tag{1.19}$$

Then with ρ as in Fact 1.1

$$\lim_{N\to\infty} P\left\{\exists \mu : Q_{N,\beta}(B_\rho(m^* e_M^\mu)) \geq \frac{1}{2} - e^{-\frac{\sqrt{M}}{\log M}}\right\} = 1. \tag{1.20}$$

Remark Note that it will not be true, with positive probability, that concentration on a single pair will hold for all N large enough. Rather, occasionally there will be random values of N for which the decomposition (1.9) will give positive weight to several pairs of balls.

Moreover, the estimates used in the proof of Corollary 1.1 together with a law of the iterated logarithm for $B_\mu(N, M(N))$ will allow us to derive (at least for $M(N)$ growing fast enough) that the sequence of indices v_N of the pairs of balls on which the measure $Q_{N,\beta}$ concentrates is transient. This is our next result.

Theorem 1.3 *Assume that $M(N) \leq N$ satisfies (1.19);*

$$M(N) \geq (\log N)^{16+\tau}, \tag{1.21}$$

for some $\tau > 0$;

$$M(2N) \leq 2M(N) \tag{1.22}$$

for all large N; and

$$M(N) - M(N-1) \leq A, \;\; N \geq 2, \tag{1.23}$$

for some $A > 0$. Then for all $\beta > 1$ there is a $d(\beta) > 0$ such that for any fixed $\mu \geq 1$,

$$P\{Q_{N,\beta}(B_\rho(\pm m^* e_M^\mu)) \geq e^{-\beta d(\beta)\sqrt{M 2 \log M}} i.o.\} = 0. \tag{1.24}$$

Remark This result might at first sight look puzzling. Obviously, for any value of N, the probability that the pair of balls with index μ has maximal weight is $1/M(N)$. Thus one might be tempted to believe that the maximum-process is recurrent if the sequence $1/M(N)$ is not summable. But note that the weights for different N are far from independent, which invalidates this argument. Indeed what happens is that the weight of a given ball changes very slowly with N, while the "fresh" patterns that are added as M increases produce almost independent weights which have a good chance to be larger than all previous ones. This explains heuristically the phenomenon described by Theorem 1.3.

Finally we observe that Theorem 1.1 gives a simple corollary on the fluctuations of the free energy, which, as will not come as a surprise, are governed by the Gumbel distribution.

Corollary 1.2 *Under the assumptions of Theorem 1.2, with*

$$a_n = \sqrt{\frac{\log M}{M}}$$

and

$$b_n = \frac{M-1}{\sqrt{\log M}}\sqrt{M} + 2\log M - \frac{\log\log M}{2} - \frac{\log(4\pi)}{2},$$

the sequence of random variables

$$a_n \left(\frac{\log Z_{N,\beta} - N\beta\phi(m^*)}{h(m^*, \beta)} \right) - b_n \to_d Y, \tag{1.25}$$

where Y is a Gumbel random variable with distribution function $G(x) = \exp(-\exp(-x))$, $x \in \mathbb{R}$.

The remainder of the paper is organized as follows. In the next section we provide the analogues of Theorem 1.1 and Corollary 1.1 in an abstract setting for dependent random variables with permutation invariant joint distributions under certain asymptotic assumptions. In Section 3 we apply these results to the random variables $B_\mu(N,M)$. The main task is to show that the appropriate factorization assumptions hold in this case. This is done using some distributional estimates due to Zaitsev [30, 31]. In Section 3.3 we prepare for the proof of Theorem 1.3 by proving a law of the iterated logarithm for the sequence of random variables $B_\mu(N,M)$, as well as an almost sure upper bound on the $\max_\mu B_\mu(N,M)$. In the final Section 4 we show that these results imply Theorems 1.2, 1.3 and Corollary 1.2.

2 Some useful convergence to Poisson process results

We consider the following setting. Let $\{X_i^N\}_{i=1,\dots,N}$ be a family of random variables defined on an abstract probability space such that for any fixed N the distribution of the random variables $X_1^N,\dots,X_N^N$ is invariant under the action of the permutation group acting on the lower indices. Our aim in this section is to establish a number of Poisson convergence results which we need to prove the results stated in the Introduction. Towards this end, consider the following sequence of point processes defined on $\mathbb{R}$

$$\Pi_N := \sum_{i=1}^N \delta_{t_N^{-1}(X_i^N)}, \ \ N \geq 1,$$

where t_N is a sequence of strictly increasing measurable functions from $\mathbb{R}$ onto $\mathbb{R}$.

Theorem 2.1 *Assume that for any integer $k \geq 1$ and any $(x_1,\dots,x_k) \in \mathbb{R}^k$,*

$$N^k P\left\{X_1^N > t_N(x_1),\dots,X_k^N > t_N(x_k)\right\} \to \exp\left(-\sum_{i=1}^k x_i\right), \ \ \text{as } N \to \infty. \tag{2.1}$$

Then the sequence of points processes Π_N converges weakly to the Poisson-point process Π on $\mathbb{R}$ with intensity measure $e^{-x}dx$.

Let $m_N(u)$ denote the number of the variables X_i^N that are greater than u.

Theorem 2.2 *Assume that for all $x \in \mathbb{R}$ and positive integers $k \geq 1$*

$$N^k P\left\{X_1^N > t_N(x),\dots,X_k^N > t_N(x)\right\} \to \exp(-xk), \ \ \text{as } N \to \infty. \tag{2.2}$$

Then for all $x \in \mathbb{R}$ and $k \geq 0$,

$$\lim_{N\to\infty} P\left\{m_N(t_N(x)) = k\right\} = \frac{e^{-xk}}{k!}\exp(-e^{-x}).$$

Remark. This theorem is completely analogous to standard theorems on order statistics in the case of stationary sequences. Assumption 2.2 replaces the usual mixing conditions. For closely related results see [11].

2.1 Proof of Theorems 2.1 and 2.2.

The proof of Theorem 2.1 will follow from Kallenberg's theorem [15] (see also [18]) on the weak convergence of a point process Π_N to the Poisson process Π. Applying his theorem in our situation, weak convergence holds whenever

(i) for all intervals $(c,d] \subset \mathbb{R}$

$$E[\Pi_N((c,d])] \to E[\Pi((c,d])] = e^{-c} - e^{-d}, \text{ as } N \to \infty,$$

and

(ii) for all $B \subset \mathbb{R}$ that are finite unions of disjoint (half-open) intervals,

$$P\{\Pi_N(B) = 0\} \to P\{\Pi(B) = 0\} = \exp\left(-\int_B e^{-x} dx\right), \text{ as } N \to \infty.$$

To verify (i), observe, trivially, that by (2.1), as $N \to \infty$,

$$E[\Pi_N((c,d])] = \sum_{i=1}^{N} P\{t_N(X_i) \in (c,d]\} = NP\{X_1 \in (t_N(c), t_N(d)]\}$$

$$= NP\{X_1 > t_N(c)\} - NP\{X_1 > t_N(d)\} \to e^{-c} - e^{-d}.$$

To prove (ii), consider first the case when B is a single interval, $B = (c,d]$, $c < d$. Clearly, then, for any integer $p \geq 1$ and all $N > p$

$$P\{\Pi_N(B) = 0\} = P\{m_N(c) = m_N(d)\}$$

$$= \sum_{k=0}^{p} P\{m_N(c) = m_N(d) = k\} + P\{m_N(c) = m_N(d) > p\}. \tag{2.3}$$

But using the permutation invariance,

$$P\{m_N(c) = m_N(d) = k\}$$

$$= \binom{N}{k} P\left\{X_1^N > t_N(d), \ldots, X_k^N > t_N(d), X_{k+1}^N \leq t_N(c), \ldots, X_N^N \leq t_N(c)\right\}.$$

The Bonferroni-inequalities (or the inclusion-exclusion principle)[10] provide the following sequence of alternating upper and lower bounds on this probability, namely for any $n \geq 1$,

$$\sum_{l=0}^{2n} (-1)^l \binom{N-k}{l} P\left\{X_1^N > t_N(d), \ldots, X_k^N > t_N(d), X_{k+1}^N > t_N(c), \ldots, X_{k+l}^N > t_N(c)\right\}$$

$$\geq P\left\{X_1^N > t_N(d), \ldots, X_k^N > t_N(d), X_{k+1}^N \leq t_N(c), \ldots, X_N^N \leq t_N(c)\right\} \geq$$

$$\sum_{l=0}^{2n+1} (-1)^l \binom{N-k}{l} P\left\{X_1^N > t_N(d), \ldots, X_k^N > t_N(d), X_{k+1}^N > t_N(c), \ldots, X_{k+l}^N > t_N(c)\right\}$$

Now by (2.1) for each fixed l

$$\binom{N}{k}\binom{N-k}{l} P\left\{X_1^N > t_N(d), \ldots, X_k^N > t_N(d), X_{k+1}^N > t_N(c), \ldots, X_{k+l}^N > t_N(c)\right\}$$

$$= \binom{N}{k}\binom{N-k}{l} e^{-dk-cl} N^{-k-l}(1+o(1)),$$

which as $N \to \infty$ converges to

$$\frac{1}{k!l!} e^{-dk-cl}.$$

Since n can be chosen arbitrarily large we readily argue that for each fixed k

$$\binom{N}{k} P\left\{X_1^N > t_N(d), \ldots, X_k^N > t_N(d), X_{k+1}^N \leq t_N(c), \ldots, X_N^N \leq t_N(c)\right\}$$

$$\to \frac{e^{-dk}}{k!} \exp(-e^{-c}), \text{ as } N \to \infty. \tag{2.4}$$

Furthermore, notice that for each fixed $p \geq 1$

$$P\left\{m_N(c) = m_N(d) > p\right\} \leq P\left\{m_N(d) > p\right\}$$

$$\leq \binom{N}{p} P\left\{X_1^N > t_N(d), \ldots, X_p^N > t_N(d)\right\},$$

which by (2.1) converges to

$$\frac{e^{-pd}}{p!}, \text{ as } N \to \infty. \tag{2.5}$$

Thus we readily conclude from (2.3) (2.4) and (2.5) (letting $p \to \infty$) that

$$\lim_{N\to\infty} P\left\{\Pi_N(B) = 0\right\} = \exp(e^{-d} - e^{-c}) = \exp(-\int_c^d e^{-x} dx).$$

The general case where B is a finite union of disjoint intervals is treated in much the same way and presents, apart from notational complexity, no further difficulties and requires no further conditions. We therefore leave the details to the reader. This completes the proof of Theorem 2.1. Theorem 2.2 has also been proved. □

3 Order statistics for $B_\mu(N,M)$.

It is easy to see that the random variables $B_\mu(N,M)$ defined in (1.11) converge individually and even with respect to the product topology to independent normal variables, provided that $M(N) \to \infty$. However, this is not sufficient to derive the asymptotic distribution of their extremes. One of the main problems is that to study the extreme value behavior one requires control of the convergence in the tails of the distribution, which conventional central limit theorems, and even Berry-Esséen theorems do not provide. The main tool that will give us the required uniform control on the convergence is a Gaussian distributional approximation result that we now describe.

3.1 Gaussian distributional approximation under Bernstein conditions.

For probability measures P and Q on the Borel subsets of $\mathbb{R}^k$, $k \geq 1$, and $\delta > 0$, let

$$\lambda(P,Q,\delta) := \sup\{P(A) - Q(A^\delta),\ Q(A) - P(A^\delta) : A \subset \mathbb{R}^k,\ \text{Borel}\}, \tag{3.1}$$

where A^δ denotes the closed δ–neighborhood of A,

$$A^\delta := \{x \in \mathbb{R}^k : \inf_{y\in A} |x-y|_2 \leq \delta\}$$

with $|\cdot|_2$ as above being the Euclidean norm on $\mathbb{R}^k$. We shall denote (s,t) to be the usual inner product for vectors $s,t \in \mathbb{R}^k$. Further, let $X_1,\dots,X_M$, $M \geq 1$, be independent mean zero random $k-$vectors satisfying for some $\tau > 0$

$$|E(s,X_i)^2(t,X_i)^{m-2}| \leq 2^{-1} m! \tau^{m-2} |t|_2^{m-2} E(s,X_i)^2,\ 1 \leq i \leq M, \tag{3.2}$$

for every $m = 3,4,...$, and for all $s,t \in \mathbb{R}^k$.

Denote the distribution of $X_1+\ldots+X_M$ by P_M and let Q_M be the k–dimensional normal distribution with mean zero and covariance matrix

$$cov(X_1) + \ldots + cov(X_M).$$

The following inequality is contained in Theorem 1.1 of Zaitsev [30] as improved in [31].

Fact 3.1 *For all integers $M \geq 1$ and $\delta \geq 0$*

$$\lambda(P_M, Q_M, \delta) \leq c_{1,k} \exp(-\delta/(c_{2,k}\tau)), \tag{3.3}$$

where $c_{i,k} \leq c_i k^2$ with c_1, c_2 being universal finite positive constants.

3.2 Application to $B_\mu(N,M)$.

We want to use Fact 3.1 for random vectors constructed from a finite collection of the variables $B_\mu(N,M)$. Let us fix $I \subset N$ with cardinality K (and assume that M is so large that $I \subset \{1,\dots,M\}$). Then let us write, for $\mu \in I$,

$$B_\mu(N,M) = \tilde{B}_\mu(N,M) + \Delta_\mu(K,N), \tag{3.4}$$

where

$$\tilde{B}_\mu(N,M) := \frac{1}{\sqrt{2M}} \sum_{\nu \notin I}^{M} \left[\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i^\mu \xi_i^\nu \right)^2 - 1 \right] \tag{3.5}$$

and

$$\Delta_\mu(K,N) := \frac{1}{\sqrt{2M}} \sum_{\nu \in I, \nu \neq \mu} \left[\left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i^\mu \xi_i^\nu \right)^2 - 1 \right]. \tag{3.6}$$

We will denote by $B_I(N,M), \tilde{B}_I(N,M)$, and $\Delta_I(N,M)$, the K-dimensional vectors, whose components are given in (3.4) to (3.6), respectively.

First we shall control the contribution of $\Delta_I(K,N)$. To do this we will need here as well as elsewhere the following special case of Hoeffding's inequality [14] applied to sums of i.i.d. Rademacher random variables: for all $z \geq 0$

$$P\left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \eta_i \geq z \right\} \leq \exp(-z^2/2), \tag{3.7}$$

where $\eta_1, ..., \eta_N$ are i.i.d. Rademacher random variables.

Lemma 3.1

$$P\left\{ |\Delta_I(K,N)|_2 > \delta \right\} \leq 4e^{-1/2} K^2 \exp\left(-\frac{\delta\sqrt{2}M}{2K^{3/2}} \right). \tag{3.8}$$

Proof. Without loss of generality we may assume that $I = \{1,\dots,K\}$. Note that

$$P\left\{ |\Delta_I(K,N)|_2 > \delta \right\} = P\left\{ \sum_{\mu \in I} (\Delta_\mu(K,N))^2 > \delta^2 \right\}$$

$$\leq 2KP\left\{ \frac{1}{\sqrt{2M}} \sum_{k=2}^{K} \left| \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i^k \xi_i^1 \right)^2 - 1 \right| > \delta/\sqrt{K} \right\}$$

$$\leq 2K^2 P\left\{ \left(\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \xi_i^k \xi_i^1 \right)^2 > \frac{\delta\sqrt{2M}}{K^{3/2}} + 1 \right\}$$

$$
(3.9) \qquad \leq 4K^2 P \left\{ \frac{1}{\sqrt{N}} \sum_i \xi_i^k > \sqrt{\frac{\delta\sqrt{2M}}{K^{3/2}} + 1} \right\} \leq 4e^{-1/2} K \exp\left(-\frac{\delta\sqrt{2M}}{2K^{3/2}} \right),
$$

where we use (3.7) to get the last inequality. □

We will see that we can use Lemma 3.1 with $\delta = M^{-1/4}$ to reduce the verification of the hypothesis of Theorem 2.1 to probabilities involving $\widetilde{B}_I(N,M)$ only. We will now show that the random variables $\widetilde{B}_I(N,M)$ are suitable for the application of Fact 3.1. In particular, conditioned on the variables ξ_i^k, $i \in \{1, ..., N\}$, $k \in I$, the summands indexed by $\nu \notin I$, in (3.5) are independent. It remains to establish that they satisfy the Bernstein conditions (3.2).

To simplify the notations we introduce i.i.d. Rademacher random variables η_i and ϵ_i^k, $i \in \{1, \ldots, N\}$ and $k \in \{1, \ldots, K\}$, and the K-dimensional random vectors $X(\epsilon)$ with components

$$
(3.10) \qquad X_k(\epsilon) := (2M)^{-1/2} \left[\left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \epsilon_i^k \eta_i \right)^2 - 1 \right].
$$

We denote by P^ϵ, E^ϵ the conditional law and expectation given the random variables ϵ_i^k. Note that the random vectors $X(\epsilon)$ have the same distribution as the vector summands in (3.5), i.e.

$$
(3.11) \qquad \left(\frac{1}{\sqrt{2M}} \left[\left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \xi_i^1 \xi_i^\nu \right)^2 - 1 \right], \ldots, \frac{1}{\sqrt{2M}} \left[\left(\frac{1}{\sqrt{N}} \sum_{i=1}^N \xi_i^K \xi_i^\nu \right)^2 - 1 \right] \right).
$$

Lemma 3.2 *For any $t, s \in \mathbb{R}^K$ and positive integer $m \geq 2$,*

$$
(3.12) \qquad E^\epsilon (s, X(\epsilon))^2 (t, X(\epsilon))^{m-2} \leq m!(2e)^m \left(\frac{K}{2M} \right)^{m/2} |s|_2^2 |t|_2^{m-2}.
$$

Proof. Obviously for any vector x, $(s,x)^2(t,x)^{m-2} \leq |s|_2^2 |t|_2^{m-2} |x|_2^m$, so that

$$
(3.13) \qquad E^\epsilon (s, X(\epsilon))^2 (t, X(\epsilon))^{m-2} \leq |s|_2^2 |t|_2^{m-2} E^\epsilon |X(\epsilon)|_2^m.
$$

Let us define

$$
V_N := \frac{1}{\sqrt{2M}} \left(\frac{1}{N} \sum_{i=1}^N \eta_i \right)^2.
$$

Observe that under P^ϵ, each of the K components of $X_k(\epsilon)$ has the same marginal distribution as $V_N - 1/(2M)^{1/2}$. Therefore, using Jensen's inequality, we see that for $m \geq 2$

$$
(3.14) \qquad E^\epsilon |X(\epsilon)|_2^m \leq K^{m/2} E^\epsilon |X_1(\epsilon)|^m \leq K^{m/2} 2^{m/2-1} (EV_N^m + (2M)^{-m/2}).
$$

Now by Khintchine's inequality (see Theorem 1 on page 254 of [8]) and Stirling's formula, we have for any positive integer $m \geq 2$

$$EV_N^m \leq (2M)^{-m/2} m^m \leq (2M)^{-m/2} m! e^m. \tag{3.15}$$

Notice also that by a trivial computation

$$EV_N^2 = (2M)^{-1}(3 - 2/N). \tag{3.16}$$

Combining these estimates gives (3.12). □

Next we need a lower bound for $E^\epsilon(s, X(\epsilon))^2$.

Lemma 3.3 *Define for integers $K \geq 1$ and $N \geq 1$, the event*

$$C_{K,N} := \left\{ \sup_{1 \leq k \neq k' \leq K} \left(\frac{1}{N} \sum_{i=1}^{N} \epsilon_i^k \epsilon_i^{k'} \right)^2 \leq \frac{1}{\sqrt{N}} \right\}, \tag{3.17}$$

then

$$P\{C_{K,N}\} \geq 1 - 2K^2 e^{-\frac{\sqrt{N}}{2}} \tag{3.18}$$

and conditioned on the event $C_{K,N}$, for all $s \in \mathbb{R}^K$,

$$E^\epsilon(s, X(\epsilon))^2 \geq \frac{1}{M} |s|_2^2 (1 - K/\sqrt{N}). \tag{3.19}$$

Proof. By a simple computation

$$2ME^\epsilon(s, X(\epsilon))^2 = 2\sum_k s_k^2 + 2\sum_{k \neq k'} s_k s_{k'} \left(\frac{1}{N} \sum_{i=1}^{N} \epsilon_i^k \epsilon_i^{k'} \right)^2. \tag{3.20}$$

But on $C_{K,N}$ we have

$$\left| \sum_{k \neq k'} s_k s_{k'} \left(\frac{1}{N} \sum_{i=1}^{N} \epsilon_i^k \epsilon_i^{k'} \right)^2 \right| \leq \frac{1}{\sqrt{N}} \sum_{k,k'} |s_k||s_k'| \leq \frac{1}{\sqrt{N}} K |s|_2^2, \tag{3.21}$$

from which we get (3.19).

To prove (3.18), just note that

$$P\left\{ \left(\frac{1}{N} \sum_{i=1}^{N} \epsilon_i^k \epsilon_i^{k'} \right)^2 > 1/\sqrt{N} \right\} = 2P\left\{ \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \eta_i > N^{1/4} \right\} \leq 2e^{-\sqrt{N}/2}, \tag{3.22}$$

where we used (3.7) for the last step, from which (3.18) follows easily. □

Putting everything together, from (3.12) and (3.19) we get:

Proposition 3.1 *Whenever $\frac{K}{\sqrt{N}} \leq \frac{1}{2}$, conditioned on the event $C_{K,N}$, the random variables $X(\epsilon)$ satisfy the Bernstein conditions, i.e. for all $m \geq 3$*

$$|E^{\epsilon}(s, X(\epsilon))^2 (t, X(\epsilon))^{m-2}| \leq m! \left(\frac{2e^2 K}{M}\right)^{\frac{m-2}{2}} |t|_2^{m-2} 4e^2 K E^{\epsilon}(s, X(\epsilon))^2$$

$$\leq \frac{m!}{2} \tau^{m-2} |t|_2^{m-2} E^{\epsilon}(s, X(\epsilon))^2, \tag{3.23}$$

with

$$\tau = \sqrt{\frac{128 e^6 K^3}{M}}. \tag{3.24}$$

Let G^{ϵ} denote the Gaussian probability distribution on $\mathbb{R}^K$, with mean zero and covariance matrix

$$cov(Z_I)_{k,k'} := \frac{M-K}{M} \left[\frac{1}{N} \sum_{i=1}^{N} \epsilon_i^k \epsilon_i^{k'}\right]^2. \tag{3.25}$$

Combining Proposition 3.1 with Fact 3.1, and computing the conditional covariance matrix of $X(\epsilon)$, we get by setting $I = \{\mu_1, \ldots, \mu_K\}$ and identifying $\epsilon_i^k = \xi_i^{\mu_k}$ the following corollary.

Corollary 3.1 *Whenever $\frac{K}{\sqrt{N}} \leq \frac{1}{2}$, on the event $C_{K,N}$, for the Gaussian probability distribution G^{ϵ} on $\mathbb{R}^k$ as above, Borel set $A \subset \mathbb{R}^K$, and $\delta \geq 0$,*

$$G^{\epsilon}(A^{\delta}) + c_1 K^2 \exp(-\frac{\delta\sqrt{M}}{c_2 K^2}) \geq P^{\epsilon}\left\{\tilde{B}_I(N, M) \in A\right\} \tag{3.26}$$

and

$$P^{\epsilon}\left\{\tilde{B}_I(N, M) \in A^{\delta}\right\} \geq G^{\epsilon}(A) - c_1 K^2 \exp(-\frac{\delta\sqrt{M}}{c_2 K^2}), \tag{3.27}$$

where c_1, c_2 are finite constants.

Proof. Under the conditional distributution P^{ϵ}, the random variable $\tilde{B}_I(N, M)$ has the same distribution as

$$\sum_{i=1}^{M-K} X^{(i)}(\epsilon),$$

where $X^{(1)}(\epsilon), \ldots, X^{(M-K)}(\epsilon)$ are i.i.d. random $K-$vectors with the same distribution as $X(\epsilon)$. Thus Proposition 3.1 allows us to apply Fact 3.1 to construct $M-K$ independent Gaussian $K-$vectors W_l, $l \in \{1, \ldots, M\} \setminus I$ with mean zero and covariance $cov(W_l)$ equal to the covariance of $X(\epsilon)$ under the law P^{ϵ}. A simple computation shows that the matrix elements of this covariance matrix are given by

$$\frac{1}{M}\left[\frac{1}{N}\sum_{i=1}^{N} \epsilon_i^k \epsilon_i^{k'}\right]^2$$

Now by setting

$$Z_I := \sum_{l \in \{1,\ldots,M\}-I} W_l$$

and using Fact 3.1 with the Bernstein conditions from Proposition 3.1, we readily obtain (3.26) and (3.27). $\square$

We want to apply this result to Borel sets $A(\overrightarrow{u})$ of the form

$$A(\overrightarrow{u}) := \{x \in \mathbb{R}^K : x_i > u_i, \text{ for } x = 1,..,K\},$$

where $\overrightarrow{u} := (u_1, \ldots, u_K)$. Notice that $A(\overrightarrow{u})^\delta \subset A(\overrightarrow{u} - \delta)$ and $A(\overrightarrow{u} + \delta) \subset A(\overrightarrow{u})^\delta$. Hence we get from (3.26) and (3.27) that

$$G^\epsilon(A(\overrightarrow{u} - \delta)) + c_1 K^2 \exp\left(-\frac{\delta\sqrt{M}}{c_2 K^2}\right) \geq P^\epsilon\left\{\tilde{B}_I(N,M) \in A(\overrightarrow{u})\right\}$$

$$\geq G^\epsilon(A(\overrightarrow{u} + \delta)) - c_1 K^2 \exp(-\frac{\delta\sqrt{M}}{c_2 K^2}), \tag{3.28}$$

where $\overrightarrow{u} + a := (u_1 + a, \ldots, u_K + a)$ for any $a \in R$.

It will be convenient to approximate the correlated Gaussian $K-$vector Z_I by an uncorrelated Gaussian $K-$vector Y_I. In fact, for any $0 \leq \gamma < 1$ such that $\gamma^2 I + (cov(Z_I) - I)$ is positive definite, we can write

$$Z_I =_{\mathcal{D}} Y_I + \Delta Z_I \tag{3.29}$$

where Y_I and ΔZ_I are independent Gaussian $K-$vectors with covariances

$$cov(Y_I) = (1-\gamma^2)I, \quad cov(\Delta Z_I) = \gamma^2 I + (cov(Z_I) - I). \tag{3.30}$$

Since on $C_{K,N}$,

$$||cov(Z_I) - I|| \leq \frac{K}{\sqrt{N}} + \frac{K}{M}, \tag{3.31}$$

we may choose $\gamma^2 := \frac{2K}{\sqrt{N}} + \frac{K}{M}$.

We recall the tail bound for a standard normal random variable Z : for all $z \geq 0$,

$$P\{|Z| \geq z\} \leq 2\exp(-z^2/2) \tag{3.32}$$

and the elementary inequalities

$$P\{X_i + Y_i > u \text{ for all } i \in I\}$$

$$\leq P\{X_i \geq u - \delta \text{ for all } i \in I\} + \sum_{i \in I} P\{|Y_i| \geq \delta\} \tag{3.33}$$

and

$$P\{X_i + Y_i > u \text{ for all } i \in I\}$$

$$\geq P\{X_i \geq u + \delta \text{ for all } i \in I\} - \sum_{i \in I} P\{|Y_i| \geq \delta\}. \tag{3.34}$$

Thus using (3.29), (3.32), (3.33), (3.34) we easily get that for any $\overrightarrow{u}$ and $\delta \geq 0$,

$$G_0\left(A((\overrightarrow{u} - \delta)/\sqrt{1-\gamma^2})\right) + 2K\exp(-\frac{\delta^2}{2\gamma^2}) \geq G^\epsilon(A(\overrightarrow{u}))$$

$$\geq G_0\left(A((\overrightarrow{u} + \delta)/\sqrt{1-\gamma^2})\right) - 2K\exp(-\frac{\delta^2}{2\gamma^2}), \tag{3.35}$$

where G_0 denotes the $K-$dimensional standard normal distribution.

Combining these bounds with (3.18), (3.26) and (3.27), we have of course that

$$P\left\{\tilde{B}_I(N,M) \in A(\overrightarrow{u})\right\} \leq G_0\left(A(\overrightarrow{u} - \delta)/\sqrt{1-\gamma^2})\right)$$

$$+2K\exp(-\frac{\delta^2}{2\gamma^2}) + c_1K^2\exp(-\frac{\delta\sqrt{M}}{c_2K^2}) + 2K^2\exp(-\sqrt{N}/2) \tag{3.36}$$

and

$$P\left\{\tilde{B}_I(N,M) \in A(\overrightarrow{u})^\delta\right\} \geq G_0\left(A(\overrightarrow{u} + \delta)/\sqrt{1-\gamma^2})\right)$$

$$-2K\exp(-\frac{\delta^2}{2\gamma^2}) - c_1K^2\exp(-\frac{\delta\sqrt{M}}{c_2K^2}) - 2K^2\exp(-\sqrt{N}/2), \tag{3.37}$$

where $\gamma^2 := \frac{2K}{\sqrt{N}} + \frac{K}{M}$. Furthermore, we obtain from (3.9), (3.33) and (3.34)

$$P\left\{B_I(N,M) \in A(\overrightarrow{u})\right\} \leq P\left\{\tilde{B}_I(N,M) \in A(\overrightarrow{u} - \delta)\right\} + 4e^{-1/2}K^2e^{-\frac{\delta\sqrt{2M}}{2K^{3/2}}} \tag{3.38}$$

and

$$P\left\{B_I(N,M) \in A(\overrightarrow{u})\right\} \geq P\left\{\tilde{B}_I(N,M) \in A(\overrightarrow{u} + \delta)\right\} - 4e^{-1/2}K^2e^{-\frac{\delta\sqrt{2M}}{2K^{3/2}}}. \tag{3.39}$$

Now write

$$\begin{aligned} p_{N,M}(\gamma^2,\delta) \quad &= \quad 2K\exp(-\frac{\delta^2}{2\gamma^2}) + c_1K^2\exp(-\frac{\delta\sqrt{M}}{c_2K^2}) + \\ &\quad\quad 2K^2\exp(-\sqrt{N}/2) + 4e^{-1/2}K^2e^{-\frac{\delta\sqrt{2M}}{2K^{3/2}}}. \end{aligned} \tag{3.40}$$

Collecting the estimates (3.36), (3.37), (3.38) and (3.39), (3.40), we get the following proposition.

Proposition 3.2 *For all integers $1 \le K, M \le N$, satisfying $K/\sqrt{N} \le 1/2$, $\overrightarrow{u} \in \mathbb{R}^K$ and $\delta > 0$*

$$P\{B_I(N,M) \in A(\overrightarrow{u})\} \le G_0\left(A(\overrightarrow{u} - 2\delta)/\sqrt{1-\gamma^2}\right) + p_{N,M}(\gamma^2, \delta) \tag{3.41}$$

and

$$P\{B_I(N,M) \in A(\overrightarrow{u})\} \ge G_0\left(A(\overrightarrow{u} + 2\delta)/\sqrt{1-\gamma^2}\right) - p_{N,M}(\gamma^2, \delta), \tag{3.42}$$

where $\gamma^2 := \frac{2K}{\sqrt{N}} + \frac{K}{M}$.

Of course we have

$$G_0(A(\overrightarrow{u})) = (1 - \Phi(u_1))...(1 - \Phi(u_K)). \tag{3.43}$$

The following elementary lemma allows us to finally do away with the different arguments in the upper and lower bounds $\overrightarrow{u} \pm 2\delta$ in (3.41) and (3.42).

Lemma 3.4 *Let Z be a standard normal variable. There exists a finite positive constant c such that for all $\gamma > 0$, $\delta > 0$ and $u > 0$ satisfying $\sqrt{1-\gamma^2} \ge 1/2$,*

$$\left| P\left\{\sqrt{1-\gamma^2} Z > u + \delta\right\} - P\{Z > u\}\right| \le c(\delta + u\gamma^2)e^{-u^2/2} \tag{3.44}$$

and whenever $u - \delta > 0$ and $u\delta \le 1$

$$\left| P\left\{\sqrt{1-\gamma^2} Z > u - \delta\right\} - P\{Z > u\}\right| \le c(\delta + u\gamma^2)e^{-u^2/2}. \tag{3.45}$$

Proof. We have

$$\begin{aligned} &\left| P\left\{\sqrt{1-\gamma^2} Z > u + \delta\right\} - P\{Z > u\}\right| \\ &= \left| P\left\{Z > u + (u(1-\sqrt{1-\gamma^2}) + \delta)/\sqrt{1-\gamma^2}\right\} - P\{Z > u\}\right| \end{aligned} \tag{3.46}$$

Now since $\sqrt{1-\gamma^2} \ge 1/2$ and $1 - \sqrt{1-\gamma^2} \le \gamma^2$,

$$(u(1-\sqrt{1-\gamma^2}) + \delta)/\sqrt{1-\gamma^2} \le 2u\gamma^2 + 2\delta. \tag{3.47}$$

Thus (3.46) is

$$\le \frac{1}{\sqrt{2\pi}} \int_u^{u+2u\gamma^2+2\delta} e^{-x^2/2} dx \le \sqrt{\frac{2}{\pi}} (\delta + u\gamma^2) e^{-u^2/2}. \tag{3.48}$$

Similarly we can argue that

$$\begin{aligned} |P\left\{\sqrt{1-\gamma^2} Z > u - \delta\right\} - P\{Z > u\}| &\le \frac{1}{\sqrt{2\pi}} \int_{u-\delta}^{u+2u\gamma^2} e^{-x^2/2} dx \\ &\le (\delta + 2u\gamma^2)\frac{1}{\sqrt{2\pi}} e^{-(u-\delta)^2/2} \end{aligned}$$

$$\leq \sqrt{\frac{2}{\pi}}(\delta + u\gamma^2)e^{-(u-\delta)^2/2}. \tag{3.49}$$

Next observe that by $u\delta \leq 1$ we have

$$-\frac{(u-\delta)^2}{2} = -\frac{u^2}{2} + u\delta - \frac{\delta^2}{2} < -\frac{u^2}{2} + 1. \tag{3.50}$$

Therefore

$$\sqrt{\frac{2}{\pi}}(\delta + u\gamma^2)e^{-(u-\delta)^2/2} < e\sqrt{\frac{2}{\pi}}(\delta + u\gamma^2)e^{-u^2/2}. \tag{3.51}$$

Setting $c = e\sqrt{\frac{2}{\pi}}$ completes the proof of the lemma. □

Recalling that the random variables Y_k are mean zero Gaussian random variables with variance $1-\gamma^2$, we have under the conditions in Lemma 3.4,

$$\left|P\left\{Y_k > u+\delta\right\} - \overline{\Phi}(u)\right| \leq c(\delta + u\gamma^2)e^{-u^2/2} \tag{3.52}$$

and

$$\left|P\left\{Y_k > u-\delta\right\} - \overline{\Phi}(u)\right| \leq c(\delta + u\gamma^2)e^{-u^2/2}, \tag{3.53}$$

where $\overline{\Phi}(u)$ is as in (1.13).

Recall from (1.14) the definition of $u_M(x)$. By (1.15) and (1.16), we get for some $C > 0$,

$$u_M(x) = O(\sqrt{\log M}), \text{ for } |x| \leq C\sqrt{\log M}. \tag{3.54}$$

Lemma 3.4 allows us to conclude that with $u_M(\overrightarrow{x}) := (u_M(x_1), ..., u_M(x_K))$,

$$G_0\left(A(u_M(\overrightarrow{x}) \pm \delta)/\sqrt{1-\gamma^2}\right)$$

$$= \prod_{i=1}^{K}\left[\frac{e^{-x_i}}{M} + O\left(\left\{\delta + \left[\frac{2K}{\sqrt{N}} + \frac{K}{M}\right]u_M(x_i)\right\}e^{-u_M^2(x_i)/2}\right)\right]. \tag{3.55}$$

Now it is well-known that for $u > 0$,

$$\frac{e^{-u^2/2}}{u\sqrt{2\pi}}\left(1-u^{-2}\right) \leq 1-\Phi(u) \leq \frac{e^{-u^2/2}}{u\sqrt{2\pi}}, \tag{3.56}$$

and thus for all large M

$$e^{-u_M^2(x_i)/2} \leq 2u_M(x_i)e^{-x_i}/M. \tag{3.57}$$

Inserting this bound into (3.55), taking into account the estimate in (3.54), and choosing $\delta = M^{-1/4}$, we get that for some $D > 0$,

$$\left|M^K P^{\epsilon}\left\{\tilde{B}_I(N,M) \in A(u_M(\overrightarrow{x}))\right\} - \prod_{i=1}^{K} e^{-x_i}\right| \leq M^K p_{N,M}(\overrightarrow{u}, \delta)$$

$$+\sum_{i=1}^{K} e^{-x_i}\left(D\left[M^{-1/4}+\sqrt{\log M}\left(\frac{1}{\sqrt{N}}+\frac{K}{M}\right)\right]\sqrt{\log M}\right), \tag{3.58}$$

which after a little analysis is easily shown to converge to zero as $M(N) \to \infty$, where we use the inequality

$$\left|\prod_{i=1}^{K} a_i - \prod_{i=1}^{K} b_i\right| \leq \sum_{i=1}^{K} |a_i - b_i|$$

holding for all $0 \leq a_i, b_i \leq 1$.

Clearly this shows that the hypotheses of Theorem 2.1 are satisfied for any $\vec{x} \in \mathbb{R}^K$. Thus Theorem 1.1 follows immediately. □

Theorem 1.1 permits us derive the asymptotic distribution of the gap between the largest and second largest order statistic of the $B_\mu(N, M)$. Let

$$\widehat{B}_1(N, M) \geq \widehat{B}_2(N, M) \geq \ldots \geq \tag{3.59}$$

denote the order statistics of the variables $B_\mu(N, M)$.

Proposition 3.3 *Under the hypotheses of Theorem 1.1, for any $\delta \geq 0$,*

$$\lim_{N\to\infty} P\left\{\widehat{B}_1(N, M) - \widehat{B}_2(N, M) \leq u_M(\delta)\right\} \to 1 - e^{-\delta}. \tag{3.60}$$

Proof. This is a corollary of Theorem 1.1. Namely, the weak convergence of the point process implies, in particular, that for any $x, y \in \mathbb{R}$,

$$\begin{aligned}&\lim_{N\to\infty} P\left\{\widehat{B}_1(N, M) \leq u_M(x), \widehat{B}_2(N, M) \leq u_M(y)\right\}\\ &= P\left\{\Pi((x, \infty)) = 0, \Pi((y, x]) \leq 1\right\},\end{aligned} \tag{3.61}$$

and a simple computation shows that for any $x \geq y$,

$$P\left\{\Pi((x, \infty)) = 0, \Pi((y, x]) \leq 1\right\} = e^{-e^{-y}}(e^{-y} - e^{-x} + 1). \tag{3.62}$$

In particular, the joint distribution of $u_M^{-1}\widehat{B}_1(N, M)$ and $u_M^{-1}\ \widehat{B}_2(N, M)$ converges to that of a random 2-vector with joint density

$$p(x, y) = e^{-e^{-y}} e^{-x-y}$$

and therefore

$$\lim_{N\to\infty} P\left\{\widehat{B}_1(N, M) - \widehat{B}_2(N, M) > u_M(\delta)\right\} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{x-\delta} dy e^{-e^{-y}} e^{-x-y} = e^{-\delta},$$

which proves the proposition. □

3.3 Some almost sure behavior of $B_\mu(N, M(N))$

We shall show that for each fixed μ, the sequence of random variables $B_\mu(N, M(N))$ satisfies a law of the iterated logarithm (LIL), more precisely,

Proposition 3.4 *Assume $M(N) \leq N$ is monotone increasing satisfying*

$$(\log N)^{2+\tau} \leq M(N), \tag{3.63}$$

for some $\tau > 0$ and all large N, and (1.22),and (1.23) hold. Then for any fixed index μ

$$\limsup_{N\to\infty} \pm \frac{B_\mu(N, M)}{\sqrt{2\log\log N^2 M(N)}} = 1, \ a.s. \tag{3.64}$$

Proof. The proof is based upon a martingale version of the Kolmogorov LIL due to Stout [27] (see also [9]). It states that if $\{(X_i, \mathcal{F}_i)\}_{i\geq 0}$ is a martingale difference sequence satisfying
(i)

$$s_n^2 := \sum_{i=1}^{n} E[X_i^2 | \mathcal{F}_{i-1}] \to \infty \text{ a.s.},$$

and
(ii)

$$|X_n| \leq \delta_n s_n / \sqrt{\log\log s_n^2} \text{ a.s.},$$

for $\delta_n > 0$, with $\delta_n \to 0$, as $n \to \infty$, then

$$\limsup_{n\to\infty} \pm \frac{\sum_{i=1}^n X_i}{\sqrt{s_n^2 \log\log s_n^2}} = 1, \ \text{ a.s.} \tag{3.65}$$

We will apply this result to the following sequence of random variables, which we will soon prove to be a martingale. Define (for a fixed nonincreasing function $M(N)$)

$$S_N := \sum_{\nu=2}^{M(N)} \left[\left(\sum_{i=1}^{N} \epsilon_i^\nu \right)^2 - N \right] \tag{3.66}$$

where ϵ_i^ν are i.i.d. Rademacher r.v.'s. (Set $S_0 = 0$ and $S_N = 0$ if $M(N) < 2$.) Clearly

$$\{S_N\}_{N\geq 1} =_{\mathcal{D}} \left\{ \sqrt{2M(N)N^2} B_\mu(N, M(N)) \right\}_{N\geq 1}. \tag{3.67}$$

We will first show that $\{S_N\}_{N\geq 0}$ is a martingale with respect to the filtration $\{\mathcal{F}_N\}_{N\geq 0}$ where $\mathcal{F}_N$, $N \geq 1$, denotes the sigma algebra generated by the random variables

$$\{\epsilon_i^\nu : 1 \leq i \leq N, 1 \leq \nu \leq M(N)\}$$

and $\mathcal{F}_0 = \{\varnothing, \Omega\}$. A straightforward computation shows that

$$S_{N+1} - S_N$$

$$= \sum_{M(N) < \nu \le M(N+1)} \left[\left(\sum_{i=1}^{N} \epsilon_i^\nu \right)^2 - N \right] + 2 \sum_{\nu=2}^{M(N)} \epsilon_{N+1}^\nu \sum_{i=1}^{N} \epsilon_i^\nu =: I_{N+1}. \tag{3.68}$$

(Empty sums are defined to be 0.) From this one readily checks that $E[I_{N+1} | \mathcal{F}_N] = 0$, implying that S_N is a martingale.

Next

$$s_N^2 := \sum_{n=1}^{N} E[I_n^2 | \mathcal{F}_{n-1}]$$

$$= \sum_{n=1}^{N} \left[2n^2(1 - 1/n)(M(n) - M(n-1)) + 4 \sum_{\nu=2}^{M(n-1)} \left(\sum_{i=1}^{n-1} \epsilon_i^\nu \right)^2 \right], \tag{3.69}$$

from which one sees immediately that condition (i) holds. To show that condition (ii) is also satisfied, we will first show that as $N \to \infty$,

$$\frac{s_N^2}{2M(N)N^2} \to 1, \text{ a.s.} \tag{3.70}$$

Now it is obvious that $Es_N^2 = ES_N^2 = 2M(N)N^2$. Thus (3.70) will follow, if we can show that as $N \to \infty$,

$$\frac{s_N^2 - Es_N}{2M(N)N^2} \to 0, \text{ a.s.} \tag{3.71}$$

This is the content of the next lemma.

Lemma 3.5 *Let $\{M(N)\}_{N \ge 1}$ be a nondecreasing positive sequence satisfying (1.22) and (3.63) for some $\gamma > 0$. Then (3.71) holds.*

Proof. Write

$$\frac{Es_N^2 - s_N^2}{M(N)N^2} = \frac{4 \sum_{n=2}^{N} \sum_{\nu=2}^{M(n-1)} [(\sum_{i=1}^{n-1} \epsilon_i^\nu)^2 - (n-1)]}{M(N)N^2}$$

$$= \frac{4 \sum_{n=1}^{N-1} S_n}{M(N)N^2} \tag{3.72}$$

We claim that with probability 1

$$\frac{|S_N|}{NM(N)} \to 0, \text{ as } N \to \infty. \tag{3.73}$$

Set $N_k = 2^k$, for $k = 1, 2, \ldots$, and choose any $\delta > 0$. Now $M(N)$ nondecreasing and assumption (1.22)

$$P\left\{\max_{N_{k-1}<N\le N_k} |S_N|/(NM(N)) > 4\delta\right\} \le P\left\{\max_{N_{k-1}<N\le N_k} |S_N| > \delta N_k M(N_k)\right\}, \tag{3.74}$$

which by Doob's inequality and (3.63) is

$$\le \frac{ES^2_{N_k}}{\delta^2 N_k^2 M^2(N_k)} \le \frac{2}{\delta^2 M(N_k)} \le \frac{2}{\delta^2 k^{2+\tau}(\log 2)^{2+\tau}}. \tag{3.75}$$

Since

$$\sum_{k=1}^{\infty} \frac{2}{\delta^2 k^{2+\gamma}(\log 2)^{2+\gamma}} < \infty, \tag{3.76}$$

we conclude (3.73) by the Borel-Cantelli lemma and the arbitrary choice of $\delta > 0$.

Now set

$$Y_N = \frac{S_N}{NM(N)},\ a_N = 4NM(N) \text{ and } A_N = N^2 M(N). \tag{3.77}$$

We see that expression (3.72) has the form

$$A_N^{-1} \sum_{n=1}^{N-1} a_n Y_n, \tag{3.78}$$

where

$$0 \le A_N^{-1} \sum_{n=1}^{N-1} a_n \le 2. \tag{3.79}$$

Since by (3.73), the Y_N converge almost surely to zero, and for each fixed $N_0 \ge 1$,

$$A_N^{-1} \sum_{n=1}^{N_0} a_n Y_n \to 0, \text{ as } N \to \infty, \tag{3.80}$$

it is easy now to conclude (3.71). □

Clearly now, condition (ii) will be verified if we can show that for any $\varepsilon > 0$, almost surely, for all large enough N,

$$|I_N| \le \varepsilon\sqrt{M(N)N^2}/\sqrt{\log\log(M(N)N^2)} =: \varepsilon L_N.$$

By assumption (1.23), it sufficient to prove that almost surely, as $N \to \infty$,

$$\left|\left(\sum_{i=1}^{N} \epsilon_i^{\nu}\right)^2 - N\right| / L_N \to 0 \tag{3.81}$$

and

$$\left|\sum_{\nu=2}^{M} \epsilon_{n+1}^{\nu}\left(\sum_{i=1}^{N} \epsilon_{i}^{\nu}\right)\right| / L_{N} \rightarrow 0. \tag{3.82}$$

Clearly we can apply inequality (3.7) to show that for some $c_1 > 0$ and $c_2 > 0$,

$$P\left\{\left|\left(\sum_{i=1}^{N} \epsilon_{i}^{\nu}\right)^{2}-N\right| \geq \delta L_{N}\right\} \leq c_{2} e^{-\delta c_{1} \sqrt{M / \log \log \left(M N^{2}\right)}}. \tag{3.83}$$

Further, since

$$\sum_{\nu=2}^{M} \epsilon_{n+1}^{\nu}\left(\sum_{i=1}^{N} \epsilon_{i}^{\nu}\right)=_{\mathcal{D}} \sum_{i=1}^{N(M-1)} \eta_{i},$$

we can also apply inequality (3.7) to get for some $c_3 > 0$,

$$P\left\{\left|\sum_{\nu=2}^{M} \epsilon_{n+1}^{\nu}\left(\sum_{i=1}^{n} \epsilon_{i}^{\nu}\right)\right| \geq \delta L_{N}\right\} \leq 2 e^{-\delta^{2} c_{3} N / \log \log \left(M N^{2}\right)}. \tag{3.84}$$

Since we are assuming that $M(N) > (\log M)^{2+\gamma}$, clearly now, using these bounds, we can find $\delta = \delta_n \downarrow 0$ such that both probabilities are summable in N, which implies that condition (ii) holds. Therefore we get that with probability 1

$$\limsup_{N \rightarrow \infty} \pm \frac{S_{N}}{\sqrt{4 M(N) N^{2} \log \log N^{2} M(N)}}=1.$$

The proposition now follows from (3.67). $\square$

Now Corollary 1.1 certainly suggests that $\max_\mu B_\mu(N, M(N)) > \sqrt{\log M}$, for all large N, almost surely. We can, however, only prove the following, somewhat weaker result.

Proposition 3.5 *Assume that for some $\tau > 0$, $M(N) \leq N$ satisfies (3.63). Then there exists a $\rho > 0$ such that*

$$P\left\{\max_{1 \leq \mu \leq M} B_{\mu}(N, M)<\sqrt{\rho \log M}, i.o.\right\}=0. \tag{3.85}$$

Proof. By the Borelli-Cantelli lemma, it suffices to show that

$$\sum_{N=1}^{\infty} P\left\{\max_{1 \leq \mu \leq M} B_{\mu}(N, M)<\sqrt{\rho \log M}\right\}<\infty. \tag{3.86}$$

Now, for any function $K(N) \leq (M(N) \wedge \sqrt{N})/8$, we have

$$P\left\{\max_{1 \leq \mu \leq M} B_{\mu}(N, M)<\sqrt{\rho \log M}\right\} \leq P\left\{\max_{1 \leq \mu \leq K} B_{\mu}(N, M)<\sqrt{\rho \log M}\right\}. \tag{3.87}$$

Let $Z_1, ..., Z_K$ be i.i.d. standard normal random variables. Arguing just as in the proof of Proposition 3.2, we can show for any $0 < \delta < 1/4$, with $\gamma^2 = 2K/\sqrt{N} + K/M$,

$$P\left\{\max_{1\le\mu\le K} B_\mu(N,M) < \sqrt{\rho\log M}\right\}$$

$$\le P\left\{\max_{1\le\mu\le K} Z_\mu \le \frac{\sqrt{\rho\log M}+2\delta}{\sqrt{1-\gamma^2}}\right\} + p_{N,M}(\gamma^2,\delta)$$

$$= \left(P\left\{Z \le \frac{\sqrt{\rho\log M}+2\delta}{\sqrt{1-\gamma^2}}\right\}\right)^{K(N)} + p_{N,M}(\gamma^2,\delta). \tag{3.88}$$

Notice that for all large enough M, using $1-\gamma^2 \ge 1/2$,

$$P\left\{Z \le \frac{\sqrt{\rho\log M}+2\delta}{\sqrt{1-\gamma^2}}\right\} \le P\left\{Z \le \sqrt{4\rho\log M}\right\}.$$

Next using the simple inequality holding for all large enough z

$$P\{Z > z\} \ge (2\pi z)^{-1}\exp(-z^2/2),$$

we obtain

$$\left(P\left\{Z \le \sqrt{4\rho\log M}\right\}\right)^{K(N)} \le \left(1 - \frac{\exp(-2\rho\log M)}{2\pi\sqrt{4\rho\log M}}\right)^{K(N)},$$

which for all large M is

$$\le (1-\exp(-4\rho\log M))^{K(N)} \le \exp(-K(N)M^{-4\rho}).$$

Putting everything together we get that for all large M

$$P\left\{\max_{1\le\mu\le K} B_\mu(N,M) < \sqrt{\rho\log M}\right\}$$

$$\le \exp(-K(N)M^{-4\rho}) + p_{N,M}(\gamma^2,\delta). \tag{3.89}$$

Choosing $0 < 4\rho < 1/16$ and letting $K(N) = M(N)^{1/16+4\rho}$, we see after some analysis that the right hand side of (3.89) is for all large M

$$\le 2\exp(-M(N)^{1/16}).$$

Since our assumption on $M(N)$ implies that

$$\sum_{N=1}^{\infty} \exp(-M(N)^{1/16}) < \infty,$$

we have shown (3.86) and thus (3.85). $\square$

4 Applications to the Hopfield model

In this last section we apply the results obtained for the random variables $B_\mu(N,M)$ to prove, with the help of Facts 1.1 and 1.2, Theorems 1.2, 1.3 and Corollary 1.2.

4.1 Proof of Theorem 1.2

Let us denote $\mu^* := \mu_N^*$ to be any index for which

$$B_{\mu_N^*}(N,M) = \widehat{B}_1(N,M).$$

Fact 1.2 implies that, with probability one, for all N large enough, uniformly in $1 \le \mu \le M(N)$,

$$\begin{aligned}&\log Q_{N,\beta}\left(B_\rho(m^* e_M^\mu)\right) - \log Q_{N,\beta}\left(B_\rho(m^* e_M^{\mu^*})\right)\\ &= c(\beta)\sqrt{M}\left[B_\mu(N,M) - B_{\mu^*}(N,M)\right] + O\left(\sqrt{M^3/N}\right),\end{aligned}$$

where $h(\beta) = h(m^*,\beta)$. But by (1.8) and (1.9), we get that

$$\sum_{\mu=1}^{M} Q_{N,\beta}\left(B_\rho(m^* e_M^\mu)\right) = \frac{1}{2} + O\left(e^{-C(M\wedge N^{1/2})}\right),$$

which implies

$$\begin{aligned}Q_{N,\beta}\left(B_\rho(m^* e_M^{\mu^*})\right) &= \frac{\frac{1}{2} + O\left(e^{-C(M\wedge N^{1/2})}\right)}{1 + \sum_{\mu\neq\mu^*} \frac{Q_{N,\beta}(B_\rho(m^* e_M^\mu))}{Q_{N,\beta}(B_\rho(m^* e_M^{\mu^*}))}}\\ &\ge \frac{\frac{1}{2}}{1 + Me^{-h(\beta)\sqrt{M}[\widehat{B}_1(N,M) - \widehat{B}_2(N,M)] + O\left(\sqrt{M^3/N}\right)}} + O\left(e^{-C(M\wedge N^{1/2})}\right).\end{aligned}$$

Now if

$$u_N^{-1}(\widehat{B}_1(N,M)) - u_N^{-1}(\widehat{B}_2(N,M)) > \delta,$$

then

$$\widehat{B}_1(N,M) - \widehat{B}_2(N,M) > \frac{\delta - o(1)}{\sqrt{2\log M}}.$$

Therefore, by Proposition 3.3, the probability that

$$Q_{N,\beta}\left(B_\rho(m^* e_M^{\mu^*})\right) \ge \frac{\frac{1}{2}}{1 + Me^{-\delta\sqrt{\frac{M}{2\log M}}h(\beta) + O\left(\sqrt{M^3/N}\right)}} + O\left(e^{-C(M\wedge N^{1/2})}\right)$$

is greater than or equal to $e^{-\delta}$, as $N \to \infty$. Further, by the assumption, $M^2 \log M \ll N$, it follows that for any $\delta > 0$,

$$\liminf_{N\to\infty} P\left\{Q_{N,\beta}\left(B_\rho(m^* e_M^{\mu^*})\right) \geq \frac{1}{2} - \frac{M}{2} e^{-\delta\sqrt{\frac{M}{2\log M}}h(\beta)/2}\right\} \geq e^{-\delta}.$$

Now, since for any arbitrary $\delta > 0$,

$$\frac{M}{2} e^{-\delta\sqrt{\frac{M}{2\log M}}h(\beta)/2} \leq e^{-\frac{\sqrt{M}}{\log M}},$$

for all sufficiently large M, this, in turn, implies that

$$\liminf_{N\to\infty} P\left\{Q_{N,\beta}\left(B_\rho(m^* e_M^{\mu^*})\right) \geq \frac{1}{2} - e^{-\frac{\sqrt{M}}{\log M}}\right\} \geq e^{-\delta},$$

for all $\delta > 0$, which yields (1.20). $\square$

4.2 Proof of Theorem 1.3

As above, letting $\mu^* := \mu_N^*$ to be any index for which $B_{\mu_N^*}(N, M) = \widehat{B}_1(N, M)$, we have, almost surely for all large enough N, for any $\mu \geq 1$ fixed,

$$\log Q_{N,\beta}\left(B_\rho(m^* e_M^{\mu})\right) - \log Q_{N,\beta}\left(B_\rho(m^* e_M^{\mu^*})\right)$$

$$= h(\beta)\sqrt{M}[B_\mu(N, M) - B_{\mu^*}(N, M)] + O\left(\sqrt{\frac{M^3}{N}}\right).$$

Now by Propositions 3.4 and 3.5, almost surely, the inequality for any $\varepsilon > 0$

$$B_\mu(N, M) - B_{\mu^*}(N, M) \leq \sqrt{(2+\varepsilon)\log\log(N^2 M)} - \sqrt{\rho \log M}$$

is violated only for finitely many values of N. But since

$$\log N \ll M \text{ and } M^2/\log M \ll N,$$

we have for all large N the bound

$$\log Q_{N,\beta}\left(B_\rho(m^* e_M^{\mu})\right) - \log Q_{N,\beta}\left(B_\rho(m^* e_M^{\mu^*})\right) \leq -\frac{\sqrt{\rho}h(\beta)}{2}\sqrt{M\log M}.$$

Exponentiating gives

$$Q_{N,\beta}\left(B_\rho(m^* e_M^{\mu})\right) e^{\frac{\rho h(\beta)}{2}\sqrt{M\log M}} \leq Q_{N,\beta}\left(B_\rho(m^* e_M^{\mu^*})\right) \leq 1,$$

which finishes the proof of Theorem 1.3. $\square$

4.3 Proof of Corollary 1.2

Finally we prove Corollary 1.2. By (1.9) we have that

$$Z_{N,\beta} = \sum_{\mu=1}^{M} 2Z_{N,\beta}Q_{N,\beta}\left(B_\rho(m^* e^\mu_M)\right) + Z_{N,\beta}O(e^{-C(M\wedge N^{1/2})}). \tag{4.1}$$

Bounding the sum over μ by its maximal term from below and M times its maximal term from above, and using the monotonicity of the logarithm, this implies

$$\begin{aligned}\frac{1}{\sqrt{M}}\log Z_{N,\beta} \le \frac{1}{\sqrt{M}}\max_{1\le\mu\le M}\log\left(Z_{N,\beta}Q_{N,\beta}\left(B_\rho(m^* e^\mu_M)\right)\right) + \frac{\log M}{\sqrt{M}} \\ +\frac{1}{\sqrt{M}}\log\left(1+\frac{O(e^{-C(M\wedge N^{1/2})})}{2M\max_{1\le\mu\le M}Q_{N,\beta}\left(B_\rho(m^* e^\mu_M)\right)}\right)\end{aligned} \tag{4.2}$$

and

$$\begin{aligned}\frac{1}{\sqrt{M}}\log Z_{N,\beta} \ge \frac{1}{\sqrt{M}}\max_{1\le\mu\le M}\log\left(Z_{N,\beta}Q_{N,\beta}\left(B_\rho(m^* e^\mu_M)\right)\right) \\ +\frac{1}{\sqrt{M}}\log\left(1+\frac{O(e^{-C(M\wedge N^{1/2})})}{2\max_{1\le\mu\le M}Q_{N,\beta}\left(B_\rho(m^* e^\mu_M)\right)}\right).\end{aligned} \tag{4.3}$$

On the other hand, (1.9) also implies that

$$2\max_{1\le\mu\le M}Q_{N,\beta}\left(B_\rho(m^* e^\mu_M)\right) \ge \frac{1}{M}\left[1-O(e^{-C(M\wedge N^{1/2})})\right], \tag{4.4}$$

so that in fact

$$\begin{aligned}\frac{1}{\sqrt{M}}\log Z_{N,\beta} = \\ \max_{1\le\mu\le M}\frac{1}{\sqrt{M}}\log\left(Z_{N,\beta}Q_{N,\beta}\left(B_\rho(m^* e^\mu_M)\right)\right) + O\left(\frac{\log 2M}{\sqrt{M}} \vee \sqrt{M}e^{-C(M\wedge N^{1/2})}\right).\end{aligned} \tag{4.5}$$

By Fact 1.2 and definition (1.11),

$$\begin{aligned}\log\left(Z_{N,\beta}Q_{N,\beta}\left(B_\rho(m^* e^\mu_M)\right)\right) \\ = \beta N\phi(m^*) + h(m^*,\beta)\left[M-1+\sqrt{2M}B_\mu(N,M)\right] + O\left(\sqrt{\frac{M^3}{N}}\right).\end{aligned} \tag{4.6}$$

Combining this with (4.5) gives

$$\frac{1}{\sqrt{M}}\log Z_{N,\beta}$$

$$= \left[\beta \frac{N}{\sqrt{M}} \phi(m^*) + h(m^*, \beta) \frac{M-1}{\sqrt{M}}\right] + h(m^*, \beta)\sqrt{2} \max_{1\le\mu\le M} B_\mu(N, M)$$

$$+O\left(\sqrt{\frac{M^2}{N}} \vee \frac{\log M}{\sqrt{M}} \vee \sqrt{M} e^{-C(M\wedge N^{1/2})}\right). \tag{4.7}$$

Next by (1.19), we have

$$a_n \left(\frac{\log Z_{N,\beta} - N\beta\Phi(m^*)}{h(m^*, \beta)}\right) - b_n =$$

$$\widetilde{u}_M^{-1} \max_{1\le\mu\le M} B_\mu(N, M) + O\left(\sqrt{\frac{M^2 \log M}{N}} \vee \frac{(\log M)^{3/2}}{\sqrt{M}} \vee \sqrt{M \log M} e^{-C(M\wedge N^{1/2})}\right)$$

$$= \widetilde{u}_M^{-1} \max_{1\le\mu\le M} B_\mu(N, M) + o(1).$$

Now (1.18) of Corollary 1.1 with $k = 0$ and $u_M(x)$ replaced by $\widetilde{u}_M(x)$ implies for all x

$$\lim_{N\to\infty} P\left\{\widetilde{u}_M^{-1} \max_{1\le\mu\le M} B_\mu(N, M) \le x\right\} = e^{-e^{-x}}. \tag{4.8}$$

(Refer to the comment following (1.16).) This proves Corollary 1.2. □

References

[1] Bovier, A., Gayrard, V., and Picco, P. (1994). Gibbs states of the Hopfield model in the regime of perfect memory. *Probab. Th. Rel. Fields* **100** 329-363.

[2] Bovier, A., Gayrard, V., and Picco, P. (1995). Gibbs states of the Hopfield model with extensively many patterns. *J. Stat. Phys.* **79** 395-414.

[3] Bovier, A. and Gayrard, V. (1997a) The retrieval phase of the Hopfield model: A rigorous analysis of the overlap distribution. *Probab.Theor. Rel. Fields* **107** 61-98.

[4] Bovier, A. and Gayrard, V. (1997b). An almost sure central limit theorem for the Hopfield model. *Markov Proc. Rel. Fields* **3** 151-173.

[5] Bovier, A. and Gayrard, V. (1998a). Hopfield models as generalized random mean field models, in *Mathematical Aspects of Spin Glasses and Neural Networks*, A. Bovier, P. Picco (eds.), Progress in Probability 41, Birkhäuser, Boston-Basel-Berlin.

[6] Bovier, A. and Gayrard, V. (1998b). Metastates in the Hopfield model in the replica symmetric regime. *Math. Phys. Anal. Geom.* **1** 107-144.

[7] Bovier, A., van Enter, A. C. D. and Niederhauser, B. (1999). Stochastic symmetry breaking in a Gaussian Hopfield model. *J. Stat. Phys.* **95** 181-213.

[8] Chow, P. L. and Teicher, H. (1988). *Probability Theory: Independence, Interchangeability, Martingales*, 2nd. ed. Springer Verlag, New York.

[9] Einmahl, U. and Mason, D. M.(1990). Some results on the almost sure behavior of martingales, *Colloquia Math. Soc. J. Bolyai* **57**, 185-195.

[10] Feller, W. (1957). *An Introduction to Probability Theory and Its Applications, Volume 1,* 2nd. ed. Wiley, New York.

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

[12] Gentz, B. (1996). *A Central Limit Theorem for the Overlap in the Hopfield Model*, Ph. D. Thesis, University of Zürich.

[13] Gentz, B. (1998). On the central limit theorem for the overlap in the Hopfield model, in *Mathematical Aspects of Spin Glasses and Neural Networks*, A. Bovier, P. Picco (eds.), Progress in Probability 41, Birkhäuser, Boston-Basel-Berlin.

[14] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Amer. Statist. Assoc.* **58** 13-30.

[15] Kallenberg, O. (1983). *Random Measures*, Akademie Verlag, Berlin.

[16] Külske, C. (1997). Metastates in disordered mean field models: Random field and Hopfield models. *J. Stat. Phys.* **88** 1257-1293.

[17] Külske, C. (1998). Limiting behavior of random Gibbs measures: Metastates in some disordered mean field models, in *Mathematical Aspects of Spin Glasses and Neural Networks*, A. Bovier, P. Picco (eds.), Progress in Probability 41, Birkhäuser, Boston-Basel-Berlin.

[18] Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, Berlin-Heidelberg-New York.

[19] Newman, C. M. (1997). *Topics in Disordered Systems*, Birkhäuser, Boston-Basel-Berlin.

[20] Newman, C. M. and Stein, D. L. (1992). Multiple states and thermodynamic limits in short ranged Ising spin glass models, *Phys. Rev. B* **46** 973-982.

[21] Newman, C. M. and Stein, D. L. (1996a). Non-mean-field behavior of realistic spin glasses. *Phys. Rev. Lett.* **76** 515-518.

[22] Newman, C. M. and Stein, D. L. (1996b). Spatial inhomogeneity and thermodynamic chaos, *Phys. Rev. Lett.* **76** 4821-4824.

[23] Newman, C. M. and Stein, D. L. (1997). Metastate approach to thermodynamic chaos. *Phys. Rev. E* **55** 5194-5211.

[24] Newman, C. M. and Stein, D. L. (1998a). Thermodynamic chaos and the structure of short-range spin glasses, in *Mathematical Aspects of Spin Glasses and Neural Networks*, A. Bovier and P. Picco (eds.), Progress in Probability 41, Birkhäuser, Boston-Basel-Berlin.

[25] Newman, C. M. and Stein, D. L. (1998b). Simplicity of state and overlap structure in finite-volume realistic spin glasses. *Phys. Rev. E* **57** 1356-1369.

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

[27] Stout, W. F. (1970). A martingale analogue of Kolmogorov's law of the iterated logarithm. *Z. Wahrsch. verw. Gebiete* **15** 279-290.

[28] Talagrand, M. (1998). Rigorous results on the Hopfield model with many patterns. *Probab. Theor. Rel. Fields* **110** 177-276.

[29] Talagrand, M. (1998). Exponential inequalities and convergence of moments in the replica-symmetric regime of the Hopfield model, preprint.

[30] Zaitsev, A. Yu. (1987a). On the Gaussian approximation of convolutions under multidimensional analogues of S. N. Bernstein's inequality conditions. *Probab. Th. Rel. Fields* **74** 534-566.

[31] Zaitsev, A. Yu. (1987b). Estimates of the Lévy-Prokhorov distance in the multivariate central limit theorem for random variables with finite exponential moments. *Theory Probab. Appl.* **31** 203-220.

Anton Bovier
Weierstrass–Institut für
Angewandte Analysis und Stochastik
Mohrenstrasse 39
D-10117 Berlin, Germany
bovier@wias-berlin.de

David M. Mason
Department of Mathematical Sciences
University of Delaware
Newark, DE 19716, USA
davidm@freedom.math.udel.edu