SHARP UPPER BOUNDS ON PERFECT RETRIEVAL IN THE HOPFIELD MODEL

Anton Bovier [1]

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

Abstract: We prove a sharp upper bound on the number of patterns that can be stored in the Hopfield model if the stored patterns are required to be fixed points of the gradient dynamic. We also show corresponding bounds on the one-step convergence of the sequential gradient dynamic. The bounds coincide with the known lower bounds and confirm the heuristic expectations. The proof is based on a crucial idea of Loukianova [L] to use the negative association properties of some random variables arising in the analysis.

Keywords: Hopfield model, storage capacity, gradient dynamic, sequential dynamic

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

1. Introduction

We consider the standard Hopfield model [Ho] (see [BP] for recent reviews on mathematical work on this model) with i.i.d. symmetric Bernoulli patterns. The most basic question that can be asked in this model is whether the patterns are fixed points of the gradient dynamics, or, equivalently, whether they are local minima of the Hamiltonian. This question was first raised in a paper by G. Weisbuch and F. Fogelman-Soulié [WF] in 1985 and answered using what is called "signal-to-noise" analysis on a mathematically heuristic level. A more mathematical analysis was later presented by R.J. McEliece et al. [MPRV] in 1986. They computed precisely the probability that all patterns were fixpoints in the limit when the size of the network, N, tends tends to infinity in the case where the number of patterns M scales in a precise way like $\frac{N}{k \log N}\left(1+\frac{\ln t\sqrt{4\pi}+\ln\ln N/2}{\ln N}\right)$ with t a parameter. This probability was seen tends to zero as t goes to infinity, but their proof could not give information if M scaled in a different way. Their result covered also the question of one-step convergence. Beside its limitations, their proof is very complicated and tedious. A simple proof of part of their analysis was later given by S. Martinez [M] and F. Vermet [V]. A good presentation can is given in the review by Petritis [P]). These rigorous results provided *lower bounds* on the number of patterns that could be stored before a patterns failed to be a fixed point, and were, depending on the precise question asked, of the form $M_c \geq \frac{N}{k \ln N}$, with $k = 2, 4, 6$. Similar results have recently been proven for biased and dependent patterns in some cases [Loe]. What was missing on the rigorous level were upper bounds, that is, it was *not* shown that if M exceeded such critical values, then the fixed point property would indeed fail.

The problem of the upper bound on the storage capacity is notorious also when other notions of storage capacity are used (see e.g. Newman [N]). The only rigorous paper that addressed this issue so far is the one by Loukianova [Lou]. She studied the issue whether the presence of local minima of the Hamiltonian can be excluded in some neighborhood of a pattern if $M = \alpha N$. Her results show that indeed there is a function $\delta(\alpha)$ such that for any $\alpha > 0$, there is no minimum in a distance less than $\delta(\alpha) > 0$ of any given patterns, if $M = \alpha N$, with probability tending to one. Unfortunately, the numerical estimates on $\delta(\alpha)$ that come out of the analysis are rather poor, and it was only shown that $\liminf \delta(\alpha) \geq 0.05$. Indeed the analysis was geared to the case of large α, and no attempt was made to verify the results of Weisbuch et al.

In this note I will show that the approach of Loukianova [Lou], with some modifications, can be used to prove that indeed sharp upper and lower bounds on the storage capacity, coinciding with those of [WF], hold true.

While the issue of the storage capacity in this strict sense may be of limited interest by now, signal-to-noise analysis of the kind used in [WF] is still used widely in the analysis of neural networks, with little regard given to the question whether the various assumptions made are applicable in the given situation. In view of this, it may be of interest to understand what is necessary to obtain *rigorous* control of such an analysis.

The gradient dynamic of the Hopfield model is defined by applying at random the maps

$$(T_i(\sigma))_k = \begin{cases} \sigma_k, & \text{if } k \neq i \\ \text{sign}\left(\sum_{j\neq i}\sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu \sigma_j\right), & \text{if k=i} \end{cases} \tag{1.1}$$

Thus a configuration σ is a fixed point of the gradient dynamic if and only if

$$\sigma_i = \text{sign}\left(\sum_{j\neq i}\sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu \sigma_j\right) \tag{1.2}$$

for all i.

Theorem 1: *For any $\gamma > 0$, the following holds:*

(i) If $M(N)$ is such that $\lim_{N\uparrow\infty} \frac{M(N)(2+\gamma)\ln N}{N} \leq 1$, then for any ν

$$\lim_{N\uparrow\infty} I\!P\left[\forall_{i=1}^N \xi_i^\nu = T_i\xi^\nu\right] = 1 \tag{1.3}$$

and

(ii) If $M(N)$ is such that $\lim_{N\uparrow\infty} \frac{M(N)(2-\gamma)\ln N}{N} \geq 1$, then for any ν

$$\lim_{N\uparrow\infty} I\!P\left[\forall_{i=1}^N \xi^\nu = T_i\xi^\nu\right] = 0 \tag{1.4}$$

Remark: As was mentioned before, part (i) has been proven earlier, and the best proof can be found in [P]. There are some essentially obvious extensions that follow from the fact that the proof provides estimates on the probabilities for the events $\xi_i^\nu = \text{sign}\left(\sum_{j\neq i}\sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu \xi_j^\nu\right)$ to *fail* that are of the form $N^{-\gamma/2}$. Therefore, with $\gamma > 2$, one may get either that the probability that *all M* patterns are fixed points tends to zero, or that with probability one, a given pattern is a fixed point for all but a finite number of values N. If $\gamma > 4$ it is even true almost surely that *all M* pattern are fixed points for all but a finite number of values N.

The actual estimates on the probabilities in (ii) we get are such that the Borel-Cantelli lemma also yield the corresponding almost sure statements for any $\gamma > 0$, that is we can repalce (ii) by

(ii') If $M(N)$ is such that $\lim_{N\uparrow\infty} \frac{M(N)(2-\gamma)\ln N}{N} \geq 1$, then for any fixed ν, with probability one, there are only finitely many values N such that $\forall_{i=1}^N T_i\xi^\nu = \xi^\nu$.

Note that on the contrary we cannot show that already for $\gamma > -2$, $I\!P\left[\forall_{\nu=1}^M \forall_{i=1}^N \xi^\nu = T_i\xi^\nu\right]$ tends to zero.

As a further illustration we mention that the method of our proof also yields with almost no modification sharp two sided estimates on the one-step convergence of the *sequential* dynamics. Recall that this dynamics is defined by applying all the maps T_i simultaneously to a configurations, i.e. at each time step we apply the operator T defined through

$$(T\sigma(t))_i \equiv \text{sign}\left(\sum_{j\neq i}\sum_{\mu=1}^{M}\xi_i^\mu\xi_j^\mu\sigma_j\right) \tag{1.5}$$

We want to know under which circumstances T acting on a configuration close to one pattern ξ^μ returns ξ^μ immediately. A worst case analysis yields poor results (but see [Bu] for a careful analysis of this question), so we want to only get results for a typical such initial conditions. To make this notion precise, we introduce a new independent r.v., $\sigma \in \mathcal{S}^\infty$, with the same distribution as the ξ^ν on the probability space $(\Omega, \mathcal{F}, I\!P)$. Then we modify the distribution of this variable by conditioning it to be close to, say, ξ^1. That is we define the new measure $I\!P^\rho[\cdot] \equiv I\!P\left[\cdot | \sum_i \xi_i^1\sigma_i = N\rho\right]$. With this definition we have

Theorem 2: *Let $0 < \rho \leq 1$. For any $\gamma > 0$, the following holds:*

(i) If $M(N)$ is such that $\lim_{N\uparrow\infty}\frac{M(N)(2+\gamma)\ln N}{\rho^2 N} \leq 1$, then

$$\lim_{N\uparrow\infty} I\!P^\rho\left[T\sigma = \xi^1\right] = 1 \tag{1.6}$$

and

(ii) If $M(N)$ is such that $\lim_{N\uparrow\infty}\frac{M(N)(2-\gamma)\ln N}{\rho^2 N} \geq 1$, then

$$\lim_{N\uparrow\infty} I\!P^\rho\left[T\sigma = \xi^1\right] = 0 \tag{1.7}$$

The first part of this theorem is known and due to Komlos and Paturi [KP] (see also [Bu,P,V]). The second part was proven in [MPRV] under special assumptions on the scaling of M, as in the case of Theorem 1.

2. Proofs of the theorems

We choose to give the proof of Theorem 1 in detail first, although it can be seen as a particular case of Theorem 2 with $\rho = 1$. There are just a few modifications necessary to cover the case of general ρ and we indicate those in the proof of Theorem 2.

Proof of Theorem 1: Part (i) is known (see [P]), but we repeat the proof for the convenience and to show how easy it is compared to part (ii). We will in fact prove that

$$\mathbb{P}\left[\exists_{i=1}^N \xi_i^\nu \neq \text{sign}\left(\sum_{j\neq i}\sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu \xi_j^\nu\right)\right] \downarrow 0 \tag{2.1}$$

The whole idea is to observe that

$$\begin{aligned}\text{sign}\left(\sum_{j\neq i}\sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu \xi_j^\nu\right) &= \text{sign}\left((N-1)\xi_i^\nu + \sum_{j\neq i}\sum_{\mu\neq\nu}^M \xi_i^\mu \xi_j^\mu \xi_j^\nu\right)\\ = \xi_i^\nu\,\text{sign}\left((N-1)+\sum_{j\neq i}\sum_{\mu\neq\nu}^M \xi_i^\nu \xi_i^\mu \xi_j^\mu \xi_j^\nu\right)\end{aligned} \tag{2.2}$$

so that the fixed point condition amounts to $\text{sign}\left((N-1)+\sum_{j\neq i}\sum_{\mu\neq\nu}^M \xi_i^\nu \xi_i^\mu \xi_j^\mu \xi_j^\nu\right) = 1$ or $\sum_{j\neq i}\sum_{\mu\neq\nu}^M \xi_i^\nu \xi_i^\mu \xi_j^\mu \xi_j^\nu \leq -(N-1)$. This yields

$$\begin{aligned}\mathbb{P}\left[\exists_{i=1}^N \xi_i^\nu \neq \text{sign}\left(\sum_{j\neq i}\sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu \xi_j^\nu\right)\right] &\leq \sum_{i=1}^N \mathbb{P}\left[\xi_i^\nu \neq \text{sign}\left(\sum_{j\neq i}\sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu \xi_j^\nu\right)\right]\\ \leq N\mathbb{P}\left[\sum_{j=2}^N\sum_{\mu=1}^{M-1} \xi_1^\nu \xi_1^\mu \xi_j^\mu \xi_j^\nu \leq -(N-1)\right] &= N\mathbb{P}\left[\sum_{j=2}^N\sum_{\mu=1}^{M-1} \xi_j^\mu \leq -(N-1)\right]\end{aligned} \tag{2.3}$$

where we used that $\xi_1^\nu\xi_1^\mu\xi_j^\nu\xi_j^\mu$ for $j\neq 1$ and fixed $\mu\neq\nu$ have the same distribution as ξ_j^μ. By a standard Gaussian domination bound one has that

$$\mathbb{P}\left[\sum_{j=2}^N\sum_{\mu=1}^{M-1} \xi_j^\mu \leq -(N-1)\right] \leq \exp\left(-\frac{(N-1)}{2(M-1)}\right) \tag{2.4}$$

from which the claimed result follows immediately.

We turn now to the proof of part (ii). Using our foregoing discussion, here we have to show that

$$\mathbb{P}\left[\forall_{i=1}^N \sum_{j\neq i}\sum_{\mu=1}^{M-1} \xi_i^\mu \xi_j^\mu \geq -(N-1)\right] \downarrow 0 \tag{2.5}$$

The essential difficulty that did not bother us in the previous bound is the lack of independence of the random variables $X_i^\mu \equiv \xi_i^\mu \sum_{j\neq i}\xi_j^\mu$ for different i. The way out of this difficulty was shown by Loukianova. Write $\xi_i^\mu \sum_{j\neq i}\xi_j^\mu = \xi_i^\mu \sum_{j=1}^N \xi_j^\mu - 1$ and define

$$W_\mu \equiv \frac{1}{\sqrt{N}}\sum_{j=1}^N \xi_j^\mu \tag{2.6}$$

We can obviously write

$$
\begin{aligned}
&I\!P\left[\forall_{i=1}^{N}\sum_{j\neq i}\sum_{\mu=1}^{M-1}\xi_i^\mu\xi_j^\mu \geq -(N-1)\right]\\
&=\sum_{\{w_\mu\}_{\mu=1,\dots,M-1}} I\!P\left[\forall_{i=1}^{N}\sum_{\mu=1}^{M-1}[\xi_i^\mu w_\mu - 1/\sqrt{N}] \geq -(N-1)/\sqrt{N}\middle| W_\mu = w_\mu, \forall_\mu\right] I\!P\left[W_\mu = w_\mu, \forall_\mu\right]
\end{aligned}
\tag{2.7}
$$

where the sum is over the (finite) set of values w_μ which the r.v.'s W_μ can attain. The crucial observation made in [Lou] is that under the conditional law $I\!P_w[\cdot] \equiv I\!P[\cdot|W_\mu = w_\mu, \forall_\mu]$, the random variables ξ_i^μ are negatively associated, so that in particular

$$
\begin{aligned}
&I\!P_w\left[\forall_{i=1}^{N}\sum_{\mu=1}^{M-1}[\xi_i^\mu w_\mu - 1/\sqrt{N}] \geq -(N-1)/\sqrt{N}\right] \leq \prod_{i=1}^{N} I\!P_w\left[\frac{1}{\sqrt{N}}\sum_{\mu=1}^{M-1}\xi_i^\mu w_\mu \geq -1+\alpha\right]\\
&\leq \left(I\!P_w\left[\frac{1}{\sqrt{N}}\sum_{\mu=1}^{M-1}\xi_1^\mu w_\mu \geq -1+\alpha\right]\right)^N
\end{aligned}
\tag{2.8}
$$

where here and in the sequal we use the abbreviation $\alpha \equiv \frac{M}{N}$ (for the relevant results on negative associated random variables, and proofs, see [JP]) . Our main task is thus to control the probability $I\!P_w\left[\frac{1}{\sqrt{N}}\sum_{\mu=1}^{M-1}\xi_1^\mu w_\mu \geq -1+\alpha\right]$, and in fact only for a set of w's that have probability tending to one. Recall that the ξ_1^μ are still independent, and, moreover, their distribution under $I\!P_w$ is easily found explicitly. Namely, as given in [Lou],

$$
I\!P_w\left[\xi_1^\mu = \pm 1\right] = \frac{1}{2}\left(1 \pm \frac{w_\mu}{\sqrt{N}}\right) \tag{2.9}
$$

Thus

$$
I\!E_w\left(\frac{1}{\sqrt{N}}\sum_{\mu=1}^{M-1}\xi_1^\mu w_\mu\right) = \frac{1}{N}\sum_\mu w_\mu^2 \tag{2.10}
$$

and

$$
Var_w\left(\frac{1}{\sqrt{N}}\sum_{\mu=1}^{M-1}\xi_1^\mu w_\mu\right) = \frac{1}{N}\sum_\mu w_\mu^2\left(1 - \frac{w_\mu^2}{N}\right) \tag{2.11}
$$

Now both those quantities are close to α with overwhelming probability. Indeed, we have

Lemma 1.1: *Let W_μ be the r.v.'s defined in (2.6). Define the set $\Omega_{\epsilon,c,N} \subset \Omega$ s.t. for all $\omega \in \Omega_{\epsilon,c,N}$*

$$
\frac{1}{M}\sum_\mu W_\mu^2[\omega] \in [1-\epsilon, 1+\epsilon] \tag{2.12}
$$

and

$$
\sup_{\mu=1}^{M-1}|W_\mu[\omega]| \leq \sqrt{2c\ln M} \tag{2.13}
$$

then, for $\epsilon \ll 1$,

$$\mathbb{P}\left[\Omega_{\epsilon,c,N}\right] \geq 1 - 2e^{-\frac{\epsilon^2 M}{8}} - M^{-c+1} \tag{2.14}$$

Proof: This Lemma results from standard and well known estimates on Rademacher sums. ♢

Remark: We note that we can choose without harm $\epsilon = \sqrt{\frac{8c\ln M}{M}}$.

From these observations we should get the idea that for those w that are in the realm of $\Omega_{\epsilon,c,N}$, the random variables $\frac{1}{\sqrt{M}}\sum_\mu \xi_1^\mu w_\mu - 1$ should converge to standard Gaussian, implying that

$$\mathbb{P}_w\left[\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu \geq -1+\alpha\right] \to \int_{-1/\alpha}^{\infty} \frac{dz}{\sqrt{2\pi}} e^{-x^2/2} \tag{2.15}$$

Unfortunately, this observation alone would not give sufficiently sharp estimates. We need some control on the speed of convergence. In [Lou] the speed of convergence was controlled via a Berry-Essén theorem. If we were to follow this approach, we would be able to prove the estimate in (ii only for $\gamma > 1$ which would leave a gap with respect to part (i).

It turns out that for α tending to zero, a better estimate can be obtained using conventional techniques used to obtain large deviation estimates.

Write

$$\mathbb{P}_w\left[\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu \geq -1+\alpha\right] = 1 - \mathbb{P}_w\left[\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu < -1+\alpha\right] \tag{2.16}$$

The second probability concerns already, if α is small, a "rare" event. Thus we may hope to get a good lower bound on it using the method of tilting. That is, for $t \leq 0$ we introduce the measure

$$\mathbb{P}_w^t[\cdot] \equiv \frac{\mathbb{E}_w\left[1_{\{\cdot\}} e^{t\sum_\mu \xi_1^\mu w_\mu}\right]}{\mathbb{E}_w\left[e^{t\sum_\mu \xi_1^\mu w_\mu}\right]} \tag{2.17}$$

Now note that for any $\delta > 0$,

$$\begin{aligned}
&\mathbb{P}_w\left[\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu < -1+\alpha\right] \\
&\geq \mathbb{P}_w\left[-1+\alpha-\delta < \frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu < -1+\alpha\right] \\
&= \mathbb{E}_w^t\left[\mathbb{1}_{\{-1+\alpha-\delta<\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu<-1+\alpha\}} e^{-t\sum_\mu \xi_1^\mu w_\mu}\right] \mathbb{E}_w\left[e^{t\sum_\mu \xi_1^\mu w_\mu}\right] \\
&\geq e^{-t(-1-\delta+\alpha)} \mathbb{E}_w\left[e^{t\sum_\mu \xi_1^\mu w_\mu}\right] \mathbb{P}_w^t\left[-1+\alpha-\delta < \frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu < -1+\alpha\right]
\end{aligned} \tag{2.18}$$

Now choose $t = t^*$ such that

$$I\!E_w^{t^*}\left[\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu\right] = -1-\delta/2+\alpha \tag{2.19}$$

Now a simple calculation shows that

$$I\!E_w\left[e^{\frac{t}{\sqrt{N}}\xi_1^\mu w_\mu}\right] = \cosh(tw_\mu/\sqrt{N}) + \frac{w_\mu}{\sqrt{N}}\sinh(tw_\mu/\sqrt{N}) \tag{2.20}$$

and

$$I\!E_w\left[\frac{1}{\sqrt{N}}\xi_1^\mu w_\mu\right] = \sinh(tw_\mu/\sqrt{N}) + \frac{w_\mu}{\sqrt{N}}\cosh(tw_\mu/\sqrt{N}) \tag{2.21}$$

Hence

$$I\!E_w^t\left[\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu\right] = \frac{1}{\sqrt{N}}\sum_\mu w_\mu\left[\frac{w_\mu/\sqrt{N}+\tanh(tw_\mu/\sqrt{N})}{1+\frac{w_\mu}{\sqrt{N}}\tanh(tw_\mu/\sqrt{N})}\right] \tag{2.22}$$

Using the fact that on $\Omega_{\epsilon,c,N}$ the w_μ are at most like $\sqrt{\ln N}$ and anticipating that t^* will be of order $1/\alpha$. Therefore, if we expand in $tw_\mu/\sqrt{N}$ and keep only the highest terms, we will make errors of order $\ln N/\sqrt{M}$ at most. Thus

$$E_w^t\left[\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu\right] = \frac{1}{N}\sum_\mu w_\mu^2(1+t)(1+O(\ln N/\sqrt{M})) \tag{2.23}$$

and so (anticipating the choice $e \geq \ln N/\sqrt{M}$)

$$t^* = -\frac{1}{\alpha}(1+\delta/2+O(\epsilon)) \tag{2.24}$$

By the same means, we see already that

$$\begin{aligned}
&e^{-t^*(-1-\delta+\alpha)}I\!E_w\left[e^{t^*\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu}\right]\\
&e^{-t^*(-1-\delta+\alpha)}\prod_\mu \cosh(t^*w_\mu/\sqrt{N})\left(1+\frac{w_\mu t^*}{\sqrt{N}}\tanh(t^*w_\mu/\sqrt{N})\right)\\
&\geq e^{-\frac{1}{2\alpha}(1+6\delta+O(\epsilon))}
\end{aligned} \tag{2.25}$$

It remains to consider the last term in (2.18). But here again the variables ξ_1^μ are independent under $I\!P_w^t$ and their sum converges to a Gaussian by the CLT. Since t^* is chosen to guarantee (2.19) and a simple computations shows that the variance

$$\begin{aligned}
&I\!E_w^{t^*}\left(\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu - I\!E_w^{t^*}\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu\right)^2\\
&=\frac{1}{N}\sum_\mu\left(I\!E_w^{t^*}(\xi_1^\mu w_\mu)^2-\left(I\!E_w^{t^*}(\xi_1^\mu w_\mu)\right)^2\right)\\
&=\frac{1}{N}\sum_\mu\left[w_\mu^2\cosh(tw_\mu/\sqrt{N})+\frac{w_\mu}{\sqrt{N}}\sinh(tw_\mu/\sqrt{N})\right.\\
&\left.+w_\mu^2\left[\frac{w_\mu/\sqrt{N}+\tanh(t^*w_\mu/\sqrt{N})}{1+\frac{w_\mu}{\sqrt{N}}\tanh(tw_\mu/\sqrt{N})}\right]^2\right]\\
&=\alpha(1+O(\epsilon))
\end{aligned} \tag{2.26}$$

A often convenient way to exploit this is to use the 2nd order Tchebychev inequality,

$$\mathbb{P}_w^t\left[-1+\alpha-\delta<\frac{1}{\sqrt{N}}\sum_\mu \xi_1^\mu w_\mu<-1+\alpha\right]\geq 1-\frac{4\alpha(1+0(\epsilon))}{\delta^2} \tag{2.27}$$

which allows e.g. to choose $d=4\sqrt{\alpha}$ (which tends to zero) and to have this probability bounded from below by 1/2. Collecting this we see that on $\Omega_{\epsilon,c,N}$ we have indeed

$$\begin{aligned}&\mathbb{P}_w\left[\forall_{i=1}^N \frac{1}{\sqrt{N}}\sum_{\mu=1}^{M-1}\xi_i^\mu w_\mu\geq -1-\alpha\right]\\&\leq\left[1-e^{-\frac{1}{2\alpha}(1+6\delta+O(\epsilon))}/2\right]^N\leq e^{-Ne^{-\frac{1}{2\alpha}(1+6\delta+O(\epsilon))}/2}\end{aligned} \tag{2.28}$$

If $\alpha\geq\frac{\ln N}{2-\gamma}$, with $\gamma>0$, while $\alpha\downarrow 0$, we can choose ϵ and δ such that for sufficiently large N, $Ne^{-\frac{1}{2\alpha}(1+6\delta+O(\epsilon))}>N^{-\gamma/4}$ so that the right hand side of (2.28) tends to zero rapidly. Since $\mathbb{P}[\Omega_{\epsilon,c,N}^c]$ tends to zero also, the proof of Theorem 1 is complete. $\diamondsuit$

Proof of Theorem 2: Let us observe that the condition $\xi_1^1=(T\sigma)_i$ amounts to

$$\sum_{j\neq i}\xi_j^1\sigma_j+\sum_{\mu=2}^M\sum_{j\neq i}\xi_i^1\xi_i^\mu\xi_j^\mu\sigma_j\geq 0 \tag{2.29}$$

But under $\mathbb{P}^\rho$ with probability one we have that $\sum_{j\neq i}\xi_j^1\sigma_j=\rho N-\xi_i^1\sigma_i$. Without any significant further modification of the proof of part (i) of theorem 1 one obtains thus that

$$\mathbb{P}^\rho\left[\exists_i\xi_i^\mu\neq(T\sigma)_i\right]\leq Ne^{-\frac{\rho^2(N-\rho^{-1})^2}{2(M-1)(N-1)}} \tag{2.30}$$

from which the claim in (i) follows.

To prove (ii) we have to deal with the random variables

$$X_i\equiv\sum_{\mu=2}^M\xi_i^1\xi_i^\mu\sum_{j\neq i}\xi_j^\mu\sigma_j=\sum_{\mu=2}^M\left(\xi_i^1\xi_i^\mu\sum_{j=1}^N\xi_j^\mu\sigma_j-\xi_i^1\sigma_i\right) \tag{2.31}$$

It is convenient to introduce the new variables $\hat{\xi}_j^\mu\equiv\xi_j^\mu\sigma_j$ for all $\mu\geq 2$ and for all $j\geq 1$. Then

$$X_i=\xi_i^1\sigma_i\left(\sum_{\mu=2}^M\hat{\xi}_i^\mu\sum_{j=1}^N\hat{\xi}_j^\mu-(M-1)\right) \tag{2.32}$$

Note that the family of random variables $\left\{\sigma_i,\xi_i^1,\hat{\xi}_i^\mu\right\}_{i\geq 1}^{\mu\geq 2}$ has the same distribution as our original family, so that the hats can be dropped in all computations of probabilities. As before we define

$W_\mu = \frac{1}{\sqrt{N}} \sum_{j=1}^N \hat{\xi}_j^\mu$. The trick is now to condition not only on the values of W_μ but also on those of the variables $\xi_i^1 \sigma_i$, that is

$$\begin{aligned} & I\!P^\rho \left[\forall_{i=1}^N \sum_{j \neq i} \sum_{\mu=1}^{M-1} \xi_i^1 \xi_i^\mu \xi_j^\mu \sigma_j \geq -\rho N - \xi_i^1 \sigma_i) \right] = \sum_{x_i = \pm 1} \sum_{\{w_\mu\}_{\mu=1,\dots,M-1}} \\ & I\!P^\rho \left[\forall_{i=1}^N x_i \left(\sum_{\mu=1}^{M-1} \hat{\xi}_i^\mu w_\mu - (M-1)/\sqrt{N} \right) \geq -(\rho N - x_i)/\sqrt{N} \middle| W_\mu = w_\mu, \forall_\mu, \xi_i^1 \sigma_i = x_i, \forall_i \right] \\ & \times I\!P^\rho \left[W_\mu = w_\mu, \forall_\mu \right] I\!P^\rho \left[\xi_i^1 \sigma_i = x_i, \forall_i \right] \end{aligned} \tag{2.33}$$

Due to the different signs of the X_i the estimate (2.8) does now no longer hold, but a sufficiently good estimate is obtained by throwing out all those conditions that have $x_i = -1$, i.e.

$$\begin{aligned} & I\!P^\rho \left[\forall_{i=1}^N x_i \left(\sum_{\mu=1}^{M-1} \hat{\xi}_i^\mu w_\mu - (M-1)/\sqrt{N} \right) \geq -(\rho N - x_i)/\sqrt{N} \middle| W_\mu = w_\mu, \forall_\mu, \xi_i^1 \sigma_i = x_i, \forall_i \right] \\ & \leq \prod_{i: x_i = +1} I\!P^\rho \left[\sum_{\mu=1}^{M-1} \hat{\xi}_i^\mu w_\mu - (M-1)/\sqrt{N} \geq -(\rho N - 1)/\sqrt{N} \middle| W_\mu = w_\mu, \forall_\mu, \xi_i^1 \sigma_i = x_i, \forall_i \right] \\ & \leq \left(I\!P^\rho \left[\sum_{\mu=1}^{M-1} \hat{\xi}_i^\mu w_\mu - (M-1)/\sqrt{N} \geq -(\rho N - 1)/\sqrt{N} \middle| W_\mu = w_\mu, \forall_\mu, \xi_i^1 \sigma_i = x_i, \forall_i \right] \right)^{\rho N} \end{aligned} \tag{2.34}$$

where we used the fact that with probability one, the number of i such that $\sigma_i \xi_i^1$ is ρN. Inserting this bound in (2.33) leaves us with

$$\begin{aligned} & I\!P^\rho \left[\forall_{i=1}^N \sum_{j \neq i} \sum_{\mu=1}^{M-1} \xi_i^1 \xi_i^\mu \xi_j^\mu \sigma_j \geq -\rho N - \xi_i^1 \sigma_i \right] \\ & \leq \sum_{\{w_\mu\}_{\mu=1,\dots,M-1}} \left(I\!P_w \left[\sum_{\mu=1}^{M-1} \xi_i^\mu w_\mu - (M-1)/\sqrt{N} \geq -(\rho N - 1)/\sqrt{N} \right] \right)^{\rho N} \\ & \times I\!P \left[W_\mu = w_\mu, \forall_\mu \right] \end{aligned} \tag{2.35}$$

From here on we may use all the estimates from the proof of Theorem 1. The only difference is the final estimate corresponding to (2.28) becomes $\exp\left(-N\rho e^{-\frac{\rho^2}{2\alpha}(1+6\delta+O(\epsilon))}/2 \right)$ from which the claimed estimate follows.$\diamondsuit$

References

[Bu] D. Burshtein, "Non-direct convergence radius and number of iterations of the Hopfield associative memory", IEEE Trans. Inform. Theory **40**, 838-847 (1994).

[BP] A. Bovier and P. Picco (Eds.), "Mathematical aspects of spin glasses and neural networks", Progress in Probability 41, (Birkhäuser: Boston, Basel, Berlin, 1998).

[Ho] J.J. Hopfield, "Neural networks and physical systems with emergent collective computational abilities", Proc. Natl. Acad. Sci. USA **79**, 2554-2558 (1982).

[KP] J. Komlós and R. Paturi, "Convegence results in an autoassociative memory model", Neural Networks **1**, 239-250 (1988).

[JP] K. Joag-Dev, F. Proschan, "Negative association of random variables, with applications", Ann. Statist. **11**, 286-295 (1983).

[Lou] D. Loukianova, "Lower bounds on the restitution error in the Hopfield model", Prob. Theor. Rel. Fields **107**, 161-176 (1997)

[Loe] M. Löwe, "Storage capacity of the Hopfield model with correlated patterns", to appear in Ann. Appl. Prob. (1998).

[Mar] S. Martinez, "Introduction to neural networks", preprint, Temuco, (1992).

[McE] R.J. McEliece, E.C. Posner, E.R. Rodemich and S.S. Venkatesh, "The capacity of the Hopfield associative memory", IEEE Trans. Inform. Theory **33**, 461-482 (1987).

[WF] G. Weisbuch and F. Fogelman-Soulié, "Scaling laws for the attractor Hopfield networks", J. Physique Letts. **46**, L623-L630 (1985).

[N] Ch.M. Newman, "Memory capacity in neural network models: Rigorous results", Neural Networks **1**, 223-238 (1988).

[P] D. Petritis, "Thermodynamic formalism of neural computing", Dynamics of complex interacting systems (Santiago, 1994), 81–146, Nonlinear Phenom. Complex Systems 2, (Kluwer Acad. Publ.: Dordrecht, 1996).

[V] F. Vermet, "Étude asymptotique d'un réseau neuronal: le modèle de mémoire associative de Hopfield", Thèse, Université de Rennes I (1994).