The Storage Capacity of Generalized Hopfield Models with Semantically Correlated Patterns

Matthias Löwe *

Keywords: Hopfield model, neural networks, storage capacity, Markov chains, large deviations
AMS Subject Classification: 82C32, 82B44, 60K35

Abstract

We analyze the storage capacity of a variant of the Hopfield model with semantically correlated patterns ξ_i^ν (that is the patterns ξ^ν themselves are correlated but consist of independent components ξ_i^ν). We show that there is a class of generalized Hopfield models of neural networks with N neurons that can store $\frac{N}{\gamma \log N}$ or αN spatially correlated patterns (depending on which notion of storage is used), provided that the correlation is generated by a homogeneous Markov chain. The quantities γ and α are independent of the correlations such that these genralized Hopfield models may be regarded as the legitimate representative of the standard Hopfield model in the presence of semantically correlated data.

1 Introduction

The so-called Hopfield model has been intensively studied by physicists, mathematicians and computer scientists since its introduction by Pastur and Figotin [FP77] in 1977 as a solvable model of a spin glass and its reinterpretation by Hopfield [Ho82] as a neural network, or more precisely as an auto-associative memory. In the last two decades the literature on the Hopfield model has become enormously wide and basically splits into the two branches mentioned above: the spin glass aspect of the Hopfield model – a field which is intrinsically related to the statistical mechanics of the model – and the neural networks aspect, which is connected to its (zero temperature) dynamics.

As we do not focus on the spin glass aspect here, we just refer the reader to the excellent survey paper by Bovier and Gayrard [BG98], and the literature cited there, for an overview of the mathematically rigorous results in this area as well as to the fundamental work by Talagrand [T98], which established a breakthrough especially concerning the validity of the so called replica symmetric solution.

Concerning the neural networks facet of the model, among others, the pioneering papers of McEliece, Posner, Rodemich, and Venkatesh [MPRV87] and Newman [N88]

*Universität Bielefeld,Fakultät für Mathematik,Postfach 100131, 33501 Bielefeld, email: loewe@mathematik.uni-bielefeld.de

have to be mentioned as well as those by Bovier and Gayrard [BG92] and Loukianova [Lou94], [Lou97]. An overview over this field can be found in the books by Amit [A89], and Hertz, Krogh, and Palmer [HKP91], as well as in the survey article by Petritis [P96].

It also should be mentioned that besides the mathematically rigorous results there is a variety of physical heuristics about this model, the starting point of which has been the observation that the Hopfield model resembles the Sherrington-Kirkpatrick model of spin glasses made by Amit, Gutfreund and Sompolinsky [AGS87].

In order to describe the essence of our contribution we have to go a little more into detail and introduce some notations. As we will concentrate on the Hopfield model as an associative memory, that is a storage device that is able to reobtain stored information from some input data, we will begin with the information to be stored. This information is encoded in sequences of plus or minus ones of length N - the so called patterns – and we assume that we want to store $M = M(N)$ of such patterns. So altogether this gives an $M \times N$ matrix $((\xi_i^\mu)_{i=1\dots,N})_{\mu=1,\dots,M}$ of ± 1. The central question in this aspect of the Hopfield model is how many of these patterns we can actually store, that is how M scales with N. Up to now this is, of course, not a well posed problem, since we have not defined the Hopfield model yet, nor did we say, what storage means, and finally we have not specified, how the ξ_i^μ are chosen.

Given the matrix of the patterns $(\xi_i^\mu)_{i,\mu}$ the Hopfield model may either be associated with the function (the Hamiltonian of the system)

$$H_N(\sigma) = -\frac{1}{2N} \sum_{i,j=1}^N \sum_{\mu=1}^N \sigma_i \sigma_j \xi_i^\mu \xi_j^\mu \qquad \sigma \in \{-1,+1\}^N \tag{1}$$

on $\{-1,+1\}^N$ or, equivalently with the corresponding Gibbs measure at inverse temperature $\frac{1}{\beta} > 0$ on $\{-1,+1\}^N$

$$\varrho_{N,\beta}(\sigma) := \frac{\exp(-\beta H_N(\sigma))}{\sum_{\sigma'} \exp(-\beta H_N(\sigma'))} \qquad \sigma \in \{-1,+1\}^N. \tag{2}$$

We will henceforth refer to the model described by equations (1) and (2) as the standard Hopfield model. Let us mention that H_N may be rewritten with the help of the overlap vector $(m_N^\mu(\sigma))_{\mu=1,\dots,M}$

$$m_N^\mu(\sigma) := \frac{1}{N} \sum_{j=1}^N \sigma_j \xi_j^\mu \tag{3}$$

in the following convenient way

$$H_N(\sigma) = -\frac{N}{2} ||m_N(\sigma)||^2$$

where $||\cdot||$ denotes the Euclidean norm in $\mathbb{R}^M$.

The idea behind this setup is that, if we assume for a moment orthogonal patterns, that is (ξ^μ) which fulfill

$$m_N^\mu(\xi^\nu) = \delta_{\mu,\nu},$$

the ξ^μ are the exact minima of the energy function H_N. Hence finding the closest local minimum of H_N associated to some input data (for example by a gradient dynamics) may be interpreted as the retrieval of some noised patterns. In general, the assumption of orthogonal patterns is not so far fetched, since e.g. if we choose the ξ_i^μ i.i.d. and unbiased, that is

$$P(\xi_i^\mu = \pm 1) = 1/2 \tag{4}$$

independent of all other ξ_j^ν, by the law of large numbers

$$m_N^\mu(\xi^\nu) \to 0 \quad \text{if } N \to \infty$$

for all $\mu \neq \nu$.

More precisely, a possible retrieval dynamics is defined by the gradient dynamics T_i at each site i that flips the spins at site into the direction of its local field, that is

$$T_i(\sigma) = \operatorname{sgn} \sum_{j=1}^N \sum_{\mu=1}^M \sigma_j \xi_i^\mu \xi_j^\mu. \tag{5}$$

This dynamics will converge to the local minimum which is closest to the starting point. In this concept a pattern is called stored if it is a fixed point of the retrieval dynamics $T = (T_i)_{i=1,\dots N}$ and the storage capacity is understood as the asymptotics of the number of patterns such that all of them are fixed points of T. McEliece et al. [MPRV87] showed that if the patterns are chosen according to (4) the storage capacity of the Hopfield model (in the above concept) is $M = N/(\gamma \log N)$ with $\gamma = 4$ or $\gamma = 6$ depending on whether we concentrate or results with probability converging to one or almost sure results.

Another, more liberal, definition goes back to Newman [N88] and defines storage capacity as the asymptotics of $M(N)$ such that all patterns ξ^μ are surrounded by a sufficiently high energy barrier, such that a retrieval dynamics leads to a state which is close to a "stored pattern", if we start close to it. More precisely, a pattern ξ^μ is called stored in this concept, if there are $\varepsilon > 0$ and $\delta > 0$ such that

$$H_N(\xi^\mu) \leq \inf_{\sigma \in S_{\delta N}(\xi^\mu)} H_N(\sigma) - \varepsilon N. \tag{6}$$

Here the notation $S_r(x)$ means the Hamming sphere of radius r centered in x. Newman [N88] showed that this concept of storage leads to a storage capacity of $M(N) = \alpha_c N$ with $\alpha_c \geq 0.056$ (which later was improved by Loukianova [L94] and Talagrand [T98] to $\alpha_c \geq 0.072$ and $\alpha_c \geq 0.08$, respectively) which is much closer to the results of computer simulations and the non-rigorous replica method [Ho82], [AGS87] that predict a storage capacity of about $M(N) = \alpha_c N$ with $\alpha = 0.14$.

Our motivation for this note originates in a result about unbiased (for biased patterns see e.g. [Lö98b)], but correlated patterns. This situation has been analyzed in [Lö98a] for both cases spatial correlations (that is correlation in i but independence in μ) and semantical correlations (correlation in μ but independence in i) for the special setup that the correlations come from homogeneous Markov chain. Especially the result about semantically correlated patterns obtained there has been quite interesting: While the storage capacity – if defined by the stability under the map T– is *increasing* for an increasing correlation it is *decreasing* (for increasing correlations) when defined by the energy valleys.

A possible explanation for this apparent contradiction is, that the presence of correlations, indeed lowers the energy $H_N(\xi^\mu)$ for each pattern ξ^μ, since the terms

$$\sum_{i,j} \sum_{\nu \neq \mu} \xi_i^\mu \xi_j^\mu \xi_i^\nu \xi_j^\nu$$

tend to be positive, such that the number ξ^μ that are fixed points under T increases. On the other hand the radius of attraction of each pattern shrinks (since the patterns get closer to each other) such that in the limit for large correlations we can store more and more patterns as fixed points but they become less and less attractive (different from what one would call an *associative* memory).

The aim of this note is to present a class of generalized Hopfield models that for semantically correlated patterns work similar to the standard Hopfield model for i.i.d. patterns. The resulting storage capacities will not depend on the strength of the correlation and be of the same order as those of the standard Hopfield model for i.i.d. patterns. Consequently, the radius of attraction δ in the second notion of storage is independent of the correlation (with the possible consequence that two or more patterns lie in the same energy valley).

The two further sections of this paper are organized in the following way: Section 2 contains the the mathematical description of the class of generalized Hopfield models we are going to consider as well as the type of dependencies we allow (Markov chains). Moreover, there we will state our main results. Section 3 will be devoted to the proofs.

Acknowledgment: I'd like to thank Anton Bovier for several discussions on the Hopfield model and in particular on techniques for correlated patterns.

2 The Setup and Statement of the Results

In this section we will describe the class of generalized Hopfield models we are going to consider and the results obtained for them. First of all we will specify the type of patterns we will treat. So throughout the rest of the paper we will assume that the patterns take values ± 1 and that the correlation among them is purely semantical and comes from a homogeneous Markov chain. More precisely, we shall assume that the $(\xi_i^\nu)_{i \in \mathbb{N}, \nu \in \mathbb{N}}$ form a Markov chain with initial distribution

$$P(\xi_i^1 = x_i^1,\ i = 1, \ldots, N) = 2^{-N} \quad \text{for all } x_i^1 \in \{-1, 1\} \text{ and all } i = 1, \ldots, N. \tag{7}$$

and transition probabilities

$$\begin{aligned} & P(\xi_i^\nu = x_i^\nu | \xi_j^\mu = x_j^\mu, j = 1, \ldots, N, \mu = 1, \ldots, \nu - 1) \\ = \ & P(\xi_i^\nu = x_i^\nu | \xi_i^{\nu-1} = x_i^{\nu-1}) = Q(x_i^{\nu-1}, x_i^\nu). \end{aligned} \tag{8}$$

Here Q denotes a symmetric 2×2 matrix with entries

$$Q = \begin{pmatrix} p & 1-p \\ 1-p & p \end{pmatrix}$$

where $0 < p < 1$ and we may assume that $p \neq 1/2$ since this is the case of independent patterns.

To explain our construction of the generalized Hopfield models recall that a basic argument for the functioning of the Hopfield model was that it works well for orthogonal patterns and that i.i.d. patterns are "almost orthogonal". The latter is no more true if the patterns are correlated. Due to this objection against the Hopfield model with correlated patterns other models have been suggested, e.g. the so-called "pseudo-inverse learning rule", which basically proposes to consider a Hamiltonian of the form

$$\tilde{H}_N(\sigma) := -\sum_{i,j=1}^{N} \sigma_i \sigma_j J_{ij}$$

with

$$J_{ij} := \frac{1}{N} \sum_{\mu,\nu=1}^{M(N)} \xi_i^\mu (\mathcal{M}^{-1})_{\mu,\nu} \xi_j^\nu = J_{ji}$$

where $\mathcal{M}^{-1}$ is the inverse of the empirical correlation matrix or overlap matrix $\mathcal{M}$ given by $\mathcal{M}_{\mu,\nu} := m_N^\mu(\xi^\nu)$.
However, the definition of $\mathcal{M}$ shows that $(\mathcal{M}^{-1})_{\mu,\nu}$ and thus also J_{ij} depends on all the spins ξ_i^ν (and not only on those with indices i or j). This property contradicts the so called locality of a neural network, a property which due to physiological constraints assumes that J_{ij} is measurable with respect to $\{\xi_i^\mu, \xi_j^\mu, \mu = 1, \ldots, M\}$. This makes the pseudo-inverse learning rule rather unattractive, even though even in the case of independent patterns its storage capacity is considerably higher than the storage capacity of the standard Hopfield model (up to $M(N) = N$ for i.i.d. patterns).

However, the following can be learned from the pseudo-inverse-rule: if we believe that "almost orthogonality" of the patterns is a basic reason, why the Hopfield model works well as associative memory we should change the geometry of the model for correlated patterns in such a way that the model "regards them as almost orthogonal". The idea for such a "change of geometry" can be obtained from the representation of the Hamiltonian of the standard Hopfield model as a multiple of the Euclidean norm of the overlap vector:

$$H_N(\sigma) = -\frac{N}{2} ||m_N(\sigma)||^2.$$

This leads to the idea to transform the geometry of the underlying space $\mathbb{R}^M$ in such a way that the patterns are again "nearly orthogonal". Thus we are motivated to study a new class of models to be defined now:

We will speak about a measure $\varrho_{\beta,N,G}$ on $\{-1,1\}^N$ as a generalized Hopfield model with geometry G, if $G = (G_{\mu,\nu})_{\mu,\nu}$ is a symmetric, positive semi-definite $M(N) \times$

$M(N)$ matrix and $\varrho_{\beta,N,G}$ is given by

$$\varrho_{\beta,N,G}(\sigma) := \frac{e^{-\beta H_{N,G}(\sigma)}}{Z_{\beta,N,G}}$$

for $\sigma \in \{-1,1\}^N$, where

$$\begin{aligned} H_{N,G}(\sigma) &:= -\frac{N}{2} < m_N(\sigma), G m_N(\sigma) > \\ &= -\frac{1}{2N} \sum_{i,j=1}^{N} \sum_{\mu,\nu=1}^{M(N)} \xi_i^\mu \xi_j^\nu \sigma_i \sigma_j G_{\mu,\nu} \\ &=: -\sum_{i,j=1}^{N} J_{ij} \sigma_i \sigma_j \end{aligned} \tag{9}$$

and $Z_{\beta,N,G}$ again is a normalizing constant to make $\varrho_{\beta,N,G}$ a probability measure:

$$Z_{\beta,N,G} := \sum_{\sigma \in \{-1,1\}^N} e^{-\beta H_{N,G}(\sigma)}.$$

Note that the choice of $G = Id$ (the $M(N) \times M(N)$ identity matrix) reproduces. the standard Hopfield model.

Of course, the properties of these models depend sensitively on the choice of the geometry G. Led by the picture behind the pseudo-inverse rule which is idea that the standard Hopfield model works the better the "more orthogonal" the patterns are, we choose the geometry G in such a way that the correlated patterns are "most orthogonal" in average, that is we choose G as the inverse of the covariance matrix C given by

$$C_{\mu\nu} = E(\xi_1^\nu \xi_1^\mu). \tag{10}$$

We note that our choice of the semantically correlated patterns, that is (7) and (8), implies that

$$C = \begin{pmatrix} 1 & a & \cdots & \cdots & a^{M-1} \\ a & 1 & a & \cdots & a^{M-2} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ a^{M-1} & \cdots & \cdots & a & 1 \end{pmatrix}$$

with $a = 2p - 1$ and thus

$$G = C^{-1} = \frac{1}{1-a^2} \begin{pmatrix} 1 & -a & 0 & \cdots & \cdots & 0 \\ -a & 1+a^2 & -a & 0 & \cdots & 0 \\ 0 & -a & 1+a^2 & -a & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots & 0 & -a & 1+a^2 & -a \\ 0 & \cdots & \cdots & 0 & -a & 1 \end{pmatrix}. \tag{11}$$

Hence

$$H_N(\sigma) := H_{N,G}(\sigma) := \sum_{i,j=1}^{N} \sum_{\mu_1,\mu_2} \sigma_i \sigma_j \xi_i^{\mu_1} \xi_j^{\mu_2} G_{\mu_1,\mu_2} \tag{12}$$

where G_{μ_1,μ_2} is given by (11).

With this setup we will prove the following results concerning the storage capacity of the generalized Hopfield models.

Theorem 2.1 *Assume the random patterns ξ^ν fulfill (7) and (8) and $M(N) = \frac{N}{\gamma \log N}$.*
Then for the generalized Hopfield model (9) with geometry $G = C^{-1}$, where the matrix C is given by (10) and G is given by (11) there exist constants $\gamma_1 > \gamma_2 > \gamma_3 > 0$ independent of p such that the following assertions hold true

1. *If $\gamma > \gamma_1$*
$$P(\liminf_{N\to\infty}(\cap_{\nu=1}^{M(N)} T\xi^\nu = \xi^\nu)) = 1$$
that is the patterns are almost surely stable.

2. *If $\gamma > \gamma_2$*
$$P((\cap_{\nu=1}^{M(N)} T\xi^\nu = \xi^\nu)) = 1 - R_N$$
with $\lim_{N\to\infty} R_N = 0$.

3. *If $\gamma > \gamma_3$ for every fixed ν*
$$P(T\xi^\nu = \xi^\nu)) = 1 - R_N$$
with $\lim_{N\to\infty} R_N = 0$.

Concerning Newman's notion of storage capacity we prove

Theorem 2.2 *Suppose that the random patterns fulfill (7) and (8). There is an $\alpha_c > 0$ not depending on p such that if $M(N) \le \alpha_c N$, then there are $\varepsilon > 0$ and $0 < \delta < 1/2$ such that for the generalized Hopfield model (9) with geometry $G = C^{-1}$ (where again C is given by (10) and G is given by (9)) the following hold true:*

$$P\left(\liminf_{N\to\infty}(\cap_{\nu=1}^{M(N)} \cap_{\sigma \in S_{\delta N}(\xi^\nu)} (H_N(\sigma) \ge H_N(\xi^\nu) + \varepsilon N))\right) = 1$$

where $S_{\delta N}(\xi^\nu)$ is the Hamming sphere of radius δN centered in ξ^ν.

3 Proofs

This section contains the proofs of Theorems 2.1 and 2.2. The idea of the proof is in both cases to apply an exponential Markov-Chebyshev inequality together with an appropriate estimate of the resulting moment generating function. This fairly simple idea in our context becomes a little bit involved due to our different geometry (indeed, even in the standard Hopfield model such a calculation with correlated patterns is not completely trivial, see [Lö98a]).

Proof of Theorem 2.1: Fix $1 \leq \nu \leq M(N)$. Then ξ^ν is stable if and only if $T(\xi_i^\nu) = \xi_i^\nu$ for all $1 \leq i \leq N$. Thus

$$P(\xi^\nu \text{ is not stable }) \leq NP(T\xi_1^\nu = -\xi_1^\nu) = NP(\sum_{j=1}^N \sum_{\mu_1,\mu_2} \xi_1^\nu \xi_j^\nu \xi_1^{\mu_1} \xi_j^{\mu_2} G_{\mu_1,\mu_2} < 0).$$

Hence applying the exponential Chebyshev–Markov inequality we obtain for all $t \geq 0$ and for $1 < \nu < M$

$$\begin{aligned} P(\xi^\nu \text{ is not stable }) &\leq NP(\sum_{j=1}^N \sum_{(\mu_1,\mu_2)\neq(\nu,\nu)} \xi_1^\nu \xi_j^\nu \xi_1^{\mu_1} \xi_j^{\mu_2} G_{\mu_1,\mu_2} \leq -NG_{\nu\nu}) \\ &\leq Ne^{-tN\frac{1+a^2}{1-a^2}} \left(E\exp(-t \sum_{(\mu_1,\mu_2)\neq(\nu,\nu)} \xi_1^\nu \xi_2^\nu \xi_1^{\mu_1} \xi_2^{\mu_2} G_{\mu_1,\mu_2}) \right)^N \end{aligned}$$

and we are led to calculating the integral

$$E\exp(-t \sum_{(\mu_1,\mu_2)\neq(\nu,\nu)} \xi_1^\nu \xi_2^\nu \xi_1^{\mu_1} \xi_2^{\mu_2} G_{\mu_1,\mu_2}).$$

To do so we will center the variables ξ_i^μ conditioned on their past. Thus we put

$$\overline{\xi_i^\mu} = \xi_i^\mu - (2p-1)\xi_i^{\mu-1}. \tag{13}$$

for $i = 1, \ldots, N$. We will now transform the sum

$$\sum_{(\mu_1,\mu_2)\neq(\nu,\nu)} \xi_1^\nu \xi_2^\nu \xi_1^{\mu_1} \xi_2^{\mu_2} G_{\mu_1,\mu_2}$$

into an expression in terms of the centered variables $\overline{\xi_i^\mu}$ with the help of the following lemma

Lemma 3.1 *Let $(X_i)_{i\in\mathbb{N}}$ and be $(Y_i)_{i\in\mathbb{N}}$ homogeneous Markov chains with state space $\{-1,+1\}$ and transition matrix*

$$\Pi = \begin{pmatrix} \pi & 1-\pi \\ 1-\pi & \pi \end{pmatrix}.$$

starting in the state $+1$ at time zero, that is $P(X_0 = +1) = P(Y_0 = +1) = 1$. Let $A = (a_{i,j})$ be a symmetric $(N+1) \times (N+1)$ matrix ($N \in \mathbb{N}$). Define

$$\overline{X}_i = \begin{cases} X_i - (2\pi-1)X_{i-1} & \text{if } i \geq 1 \\ X_0 & \text{otherwise} \end{cases} \tag{14}$$

and

$$\overline{Y}_i = \begin{cases} Y_i - (2\pi - 1)Y_{i-1} & \textit{if } i \geq 1 \\ Y_0 & \textit{otherwise} \end{cases} \tag{15}$$

Then

$$\sum_{i,j=0}^{N} a_{i,j} X_i Y_j = \sum_{i,j=0}^{N} b_{i,j} \overline{X}_i \overline{Y}_j \tag{16}$$

where $b_{i,j} = \sum_{l=i}^{N} \sum_{m=j}^{N} a_{lm}(2\pi - 1)^{l+m-(i+j)}$.

Proof: Note that (14) and (15) imply that

$$X_i = \sum_{k=0}^{i} \gamma^k \overline{X}_{i-k} = \sum_{k=0}^{i} \gamma^{i-k} \overline{X}_k \quad \text{and } Y_i = \sum_{k=0}^{i} \gamma^k \overline{Y}_{i-k} = \sum_{k=0}^{i} \gamma^{i-k} \overline{Y}_k$$

with $\gamma := 2\pi - 1$.
Hence

$$\begin{aligned} \sum_{i,j=0}^{N} a_{i,j} X_i Y_j &= \sum_{i,j=0}^{N} a_{i,j} \Big(\sum_{l=0}^{i} \gamma^{i-l} \overline{X}_l \Big) \Big(\sum_{m=0}^{i} \gamma^{i-m} \overline{Y}_m \Big) \\ = & \sum_{i,j=0}^{N} \Big(\sum_{l=i}^{N} \sum_{m=j}^{N} \gamma^{(l+m)-(i+j)} a_{lm} \overline{X}_i \overline{Y}_j \Big) = \sum_{i,j=0}^{N} b_{i,j} \overline{X}_i \overline{Y}_j. \end{aligned}$$

□

Before we apply Lemma 3.1 we may without loss of generality assume that $\xi_i^\nu = +1$ for all i, since the initial situation is completely symmetric. Therefore

$$\sum_{(\mu_1,\mu_2) \neq (\nu,\nu)} \xi_1^\nu \xi_2^\nu \xi_1^{\mu_1} \xi_2^{\mu_2} G_{\mu_1,\mu_2} = \sum_{(\mu_1,\mu_2) \neq (\nu,\nu)} \xi_1^{\mu_1} \xi_2^{\mu_2} G_{\mu_1,\mu_2}.$$

Due to the structure of the matrix G this can be split

$$\sum_{(\mu_1,\mu_2) \neq (\nu,\nu)} \xi_1^{\mu_1} \xi_2^{\mu_2} G_{\mu_1,\mu_2} = \sum_{\substack{\mu_1,\mu_2 \leq \nu \\ (\mu_1,\mu_2) \neq (\nu,\nu)}} \xi_1^{\mu_1} \xi_2^{\mu_2} G_{\mu_1,\mu_2} + \sum_{\substack{\mu_1,\mu_2 \geq \nu \\ (\mu_1,\mu_2) \neq (\nu,\nu)}} \xi_1^{\mu_1} \xi_2^{\mu_2} G_{\mu_1,\mu_2}.$$

As an immediate application of Lemma 3.1 we obtain

$$\sum_{\substack{\mu_1,\mu_2 \geq \nu \\ (\mu_1,\mu_2) \neq (\nu,\nu)}} \xi_1^{\mu_1} \xi_2^{\mu_2} G_{\mu_1,\mu_2} = \sum_{\substack{\mu_1,\mu_2 \geq \nu \\ (\mu_1,\mu_2) \neq (\nu,\nu)}} \overline{\xi}_1^{\mu_1} \overline{\xi}_2^{\mu_2} H_{\mu_1,\mu_2}$$

with

$$H_{\mu_1,\mu_2} = \sum_{\nu_1=\mu_1}^{M} \sum_{\nu_2=\mu_2}^{M} (2p-1)^{\nu_1+\nu_2-(\mu_1+\mu_2)} G_{\nu_1,\nu_2}.$$

By expressing the coefficients G_{μ_1,μ_2} as well as the centering in terms of $a = 2p - 1$ and by taking into account that $G_{\mu_1,\mu_2} = 0$ if $|\mu_1 - \mu_2| \geq 2$ we obtain

$$\begin{aligned} H_{\mu,\mu} &= \sum_{\nu_1 \geq \mu} a^{2\nu_1 - 2\mu} G_{\nu_1,\nu_1} + 2 \sum_{\substack{\nu_1 = \nu_2 + 1 \\ \nu_2 \geq \mu}} a^{2\nu_2 + 1 - 2\mu} G_{\nu_1,\nu_2} \\ &= \frac{1}{1-a^2} \left[(1+a^2) \sum_{n=0}^{M-\mu-1} a^{2n} + a^{2M-2\mu} - 2a^2 \sum_{n=0}^{M-\mu-1} a^{2n} \right] \\ &= \frac{1}{1-a^2} \left[\frac{1-a^{2M-2\mu}}{1-a^2}(1-a^2) + a^{2M-2\mu} \right] = \frac{1}{1-a^2} \end{aligned}$$

for all $\mu \geq \nu$. For $\mu_1 \neq \mu_2$ we obtain

$$\begin{aligned} H_{\mu_1,\mu_2} := & \left(\frac{1+a^2}{1-a^2} \sum_{n=0}^{M-1-\max \mu_1,\mu_2} a^{2n} + \frac{1}{1-a^2} a^{2(M-\max \mu_1,\mu_2)} \right. \\ & \left. - \frac{1}{1-a^2} \sum_{n=0}^{M-\max \mu_1,\mu_2} a^{2n} - \frac{a^2}{1-a^2} \sum_{n=0}^{M-1-\max \mu_1,\mu_2} a^{2n} \right) a^{|\mu_1-\mu_2|} = 0. \end{aligned}$$

Thus

$$H_{\mu_1,\mu_2} := \begin{cases} \frac{1}{1-a^2} & \text{if } \mu_1 = \mu_2 \\ 0 & \text{if } \mu_1 \neq \mu_2. \end{cases}$$

Analogously,

$$\sum_{\substack{\mu_1,\mu_2 \leq \nu \\ (\mu_1,\mu_2) \neq (\nu,\nu)}} \xi_1^{\mu_1} \xi_2^{\mu_2} G_{\mu_1,\mu_2} = \sum_{\substack{\mu_1,\mu_2 \leq \nu \\ (\mu_1,\mu_2) \neq (\nu,\nu)}} \overline{\xi}_1^{\mu_1} \overline{\xi}_2^{\mu_2} \tilde{H}_{\mu_1,\mu_2}$$

with

$$\tilde{H}_{\mu_1,\mu_2} := \begin{cases} \frac{1}{1-a^2} & \text{if } \mu_1 = \mu_2 \\ 0 & \text{if } \mu_1 \neq \mu_2. \end{cases}$$

Therefore

$$E \exp(-t \sum_{(\mu_1,\mu_2) \neq (\nu,\nu)} \xi_1^\nu \xi_2^\nu \xi_1^{\mu_1} \xi_2^{\mu_2} G_{\mu_1,\mu_2}) = e^{-\frac{t}{1-a^2}} E \exp\Big(-\frac{t}{1-a^2} \sum_{\mu \neq \nu} \overline{\xi}_1^\mu \overline{\xi}_2^\mu \Big)$$

implying that

$$P(\xi^\nu \text{ is not stable }) \leq e^{-tN\frac{2+a^2}{1-a^2}} \left(E \exp\Big(-\frac{t}{1-a^2} \sum_{\mu \neq \nu} \overline{\xi}_1^\mu \overline{\xi}_2^\mu \Big) \right)^N.$$

Since this is true for all $t \geq 0$ we may as well replace t by $\frac{t}{1-a^2}$ and arrive at

$$P(\xi^\nu \text{ is not stable }) \leq N e^{-tN(2+a^2)} \left(E \exp\Big(-t \sum_{\mu \neq \nu} \overline{\xi}_1^\mu \overline{\xi}_2^\mu \Big) \right)^N.$$

The expectation on the right hand side of this expression will be evaluated by iterative integration. To this end observe that

$$E \exp\Big(-t \sum_{\mu \neq \nu} \overline{\xi}_1^\mu \overline{\xi}_2^\mu \Big) = E^{1 \leq \mu \leq M-1} \exp\Big(-t \sum_{\substack{\mu \neq \nu \\ \mu \leq M-1}} \overline{\xi}_1^\mu \overline{\xi}_2^\mu \Big) E^M \exp\Big(-t \overline{\xi}_1^M \overline{\xi}_2^M \Big)$$

where $E^{I'}$ (with $I' \subseteq \{1, \ldots, M\}$) denotes integration with respect to those $\overline{\xi}_1^\mu$ and $\overline{\xi}_2^\mu$ with $\mu in I'$. Now

$$\begin{aligned}
& E^{1\leq\mu\leq M-1} \exp\Big(-t \sum_{\substack{\mu\neq\nu \\ \mu\leq M-1}} \overline{\xi}_1^\mu \overline{\xi}_2^\mu\Big) E^M \exp\Big(-t \overline{\xi}_1^M \overline{\xi}_2^M\Big) \\
= \;& E^{1\leq\mu\leq M-1} \exp\Big(-t \sum_{\substack{\mu\neq\nu \\ \mu\leq M-1}} \overline{\xi}_1^\mu \overline{\xi}_2^\mu\Big) \\
& \quad \times \Big(q \exp(-2t(1-q)\xi_1^{M-1}\xi_2^{M-1}) + (1-q)\exp(2tq\xi_1^{M-1}\xi_2^{M-1})\Big) \\
\leq \;& E^{1\leq\mu\leq M-1} \exp\Big(-t \sum_{\substack{\mu\neq\nu \\ \mu\leq M-1}} \overline{\xi}_1^\mu \overline{\xi}_2^\mu\Big) \cosh(2qt\xi_1^{M-1}\xi_2^{M-1}) \\
\leq \;& E^{1\leq\mu\leq M-1} \exp\Big(-t \sum_{\substack{\mu\neq\nu \\ \mu\leq M-1}} \overline{\xi}_1^\mu \overline{\xi}_2^\mu\Big) \exp(2q^2t^2).
\end{aligned}$$

Here $q = p^2 + (1-p)^2$ is the probability for the Markov chain $Y_\mu := \xi_1^\mu \xi_2^\mu$ to stay in $+1$ when it already is in $+1$ and we have used $|\xi_1^\mu| = |\xi_2^\mu| = 1$ together with the inequalities

$$p\exp(-2(1-p)t) + (1-p)\exp(2pt) \leq \cosh((1+|2p-1|)t)$$

for all $0 < p < 1$ and all $t \in \mathbb{R}$ and $\cosh(x) \leq \exp(x^2/2)$ for all $x \in \mathbb{R}$. Iterating this procedure yields

$$E\exp\Big(-t\sum_{\mu\neq\nu} \overline{\xi}_1^\mu \overline{\xi}_2^\mu\Big) \leq \exp(2q^2t^2(M-1)) \leq \exp(2t^2M).$$

So altogether

$$\begin{aligned}
& P(\xi^\nu \text{ is not stable }) \leq Ne^{-tN(2+a^2)} \exp(2q^2t^2NM) \\
\leq \;& Ne^{-2tN}\exp(2t^2NM).
\end{aligned} \tag{17}$$

Now all the assertions of Theorem 2.1 basically follow from equation (17). First note that the choice of $t = \frac{1}{2M}$ implies that

$$P(\xi^\nu \text{ is not stable }) \leq Ne^{-\frac{N}{2M}}$$

such that if $M = \frac{N}{\gamma \log N}$ with $\gamma > 2 =: \gamma_3$

$$P(\xi^\nu \text{ is not stable }) \to_{N\to\infty} 0.$$

Moreover,

$$P(\exists \nu : \xi^\nu \text{ is not stable }) \leq MP(\xi^1 \text{ is not stable }) \leq NMe^{-\frac{N}{2M}}$$

(again by choosing $t = 1/2M$). Thus $M = \frac{N}{\gamma \log N}$ with $\gamma > 4 =: \gamma_2$ gives

$$P(\exists \nu : \xi^\nu \text{ is not stable }) \to_{N\to\infty} 0$$

while a choice of $\gamma > 6 =: \gamma_1$ even yields

$$P(\liminf_{N\to\infty}(\cap_{\nu=1}^{M(N)} T\xi^\nu = \xi^\nu)) = 1$$

by the Borel-Cantelli Lemma. $\square$

Remark 3.2 *One might observe that we could slightly do better by taking into account the contribution of the terms containing a or q. In such a way we could obtain bounds on the storage capacity that are p-dependent and slightly better than those for the standard Hopfield model with i.i.d. patterns. However, since p and q are bounded this improvement would not be very impressing.*

The proof of Theorem 2.2 is a combination of the ideas yielding the proof of Theorem 2.1 and the ideas introduced by Newman [N88] to analyze the storage capacity of the standard Hopfield model. We will try to be rather brief.

Proof of Theorem 2.2: Using the Hamiltonian $H_N(\sigma)$ of the generalized Hopfield models as given in (12) and putting

$$h_N(\sigma,\delta) := \inf_{\sigma'\in S_\delta(\sigma)} H_N(\sigma').$$

we obtain

$$P\Big(\{\cap_{\nu=1}^{M(N)}(h_N(\xi^\nu,\delta) \geq H_N(\xi^\nu)+\varepsilon N)\}^c\Big) = P\Big(\cup_{\nu=1}^{M(N)} \cup_{J:|J|=\delta N} H_N(\xi^\nu_J) - H_N(\xi^\nu) \leq \varepsilon N\Big)$$

$$\leq \sum_{J:|J|=\delta N} \sum_{\nu=1}^{M(N)} P\Big(H_N(\xi^\nu_J) - H_N(\xi^\nu) \leq \varepsilon N\Big)$$

where ξ^ν_J denotes a configuration differing from ξ^ν exactly in the coordinates indexed by $J \subset \{1,\dots,N\}$ and δN is an integer. For fixed $2 \leq \nu \leq M(N)-1$ and again with $a = 2p-1$ we obtain for all $t \geq 0$

$$\begin{aligned}
& P\Big(H_N(\xi^\nu_J) - H_N(\xi^\nu) \leq \varepsilon N\Big) \\
= \;& P\Big(-\frac{1}{2N}\sum_{\mu_1,\mu_2=1}^{M(N)}\sum_{i,j=1}^{N}(\xi^\nu_{J,i}\xi^\nu_{J,j} - \xi^\nu_i\xi^\nu_j))\xi^{\mu_1}_i\xi^{\mu_2}_j G_{\mu_1,\mu_2} \leq \varepsilon N\Big) \\
= \;& P\Big(\frac{1}{N}\sum_{\mu_1,\mu_2=1}^{M(N)}\sum_{\substack{i\in J, j\notin J\\ i\notin J, j\in J}} \xi^\nu_i\xi^\nu_j\xi^{\mu_1}_i\xi^{\mu_2}_j G_{\mu_1,\mu_2} \leq \varepsilon N\Big) \\
= \;& P\Big(\frac{1}{N}\sum_{(\mu_1,\mu_2)\neq(\nu,\nu)}\sum_{i\in J, j\notin J} \xi^\nu_i\xi^\nu_j\xi^{\mu}_i\xi^{\mu}_j \leq (\varepsilon/2 - \frac{1+a^2}{1-a^2}\delta(1-\delta))N\Big) \\
\leq \;& e^{-t\varepsilon' N} E\exp\Big(-\frac{t}{N}\sum_{(\mu_1,\mu_2)\neq(\nu,\nu)}\sum_{i\in J, j\notin J} \xi^\nu_i\xi^\nu_j\xi^{\mu_1}_i\xi^{\mu_2}_j G_{\mu_1,\mu_2}\Big)
\end{aligned}$$

where we have set $\varepsilon' = -\varepsilon/2 + \frac{1+a^2}{1-a^2}\delta(1-\delta)$.

Now centering the variables ξ_i^μ, that is replacing them by $\overline{\xi_i^\mu} = \xi_i^\mu - (2p-1)\xi_i^{\mu-1}$ yields by the same methods as in the proof of Theorem 2.1 above

$$\begin{aligned} & E\exp\Big(-\frac{t}{N}\sum_{i\in J, j\notin J}\sum_{(\mu_1,\mu_2)\neq(\nu,\nu)} \xi_i^{\mu_1}\xi_j^{\mu_2}G_{\mu_1,\mu_2}\Big) = \exp\Big(-tN\delta(1-\delta)\frac{2}{1-a^2}\Big) \\ \times \quad & E\exp\Big(-\frac{t}{N}\sum_{i\in J, j\notin J}\Big(\sum_{\mu_1,\mu_2>\nu} b_{\mu_1,\mu_2}\overline{\xi_i^{\mu_1}}\,\overline{\xi_j^{\mu_2}} + \sum_{\mu_1,\mu_2<\nu}\tilde b_{\mu_1,\mu_2}\overline{\xi_i^{\mu_1}}\,\overline{\xi_j^{\mu_2}}\Big)\Big), \end{aligned}$$

with

$$b_{\mu_1,\mu_2} = \tilde b_{\mu_1,\mu_2} = \begin{cases} \frac{1}{1-a^2} & \text{if } \mu_1 = \mu_2 \\ 0 & \text{if } \mu_1 \neq \mu_2. \end{cases}$$

More explicitly we obtain

$$\begin{aligned} & E\exp\Big(-\frac{t}{N}\sum_{i\in J, j\notin J}\sum_{(\mu_1,\mu_2)\neq(\nu,\nu)} \xi_i^{\mu_1}\xi_j^{\mu_2}G_{\mu_1,\mu_2}\Big) \\ \leq \quad & \exp\Big(-tN\delta(1-\delta)\frac{2}{1-a^2}\Big)E\exp\Big(-\frac{t}{(1-a^2)N}\sum_{i\in J, j\notin J}\sum_{\mu\neq\nu}\overline{\xi_i^\mu}\,\overline{\xi_j^\mu}\Big). \end{aligned}$$

Next we will replace the $\overline{\xi_i^\mu}$ by Gaussian random variables. To this end again let $E_I^{I'}$ (where $I \subset \{1,\ldots,N\}$ and $I' \subset \{1,\ldots,M\}$) denote the integration with respect to those random variables ξ_i^μ with $i \in I$ and $\mu \in I'$. Especially, if we drop the upper or lower indices we will usually mean the expectation with respect to all the random variables occuring in the argument of the integral.
By Hölder's inequality

$$\begin{aligned} & E\exp\Big(-\frac{t}{(1-a^2)N}\sum_{i\in J, j\notin J}\sum_{\mu\neq\nu}\overline{\xi_i^\mu}\,\overline{\xi_j^\mu}\Big) \leq \\ & \Big(E\exp\Big(-\frac{2t}{(1-a^2)N}\sum_{i\in J, j\notin J}\sum_{\mu=1}^{\nu}\overline{\xi_i^\mu}\,\overline{\xi_j^\mu}\Big)\Big)^{1/2}\Big(E\exp\Big(-\frac{2t}{(1-a^2)N}\sum_{i\in J, j\notin J}\sum_{\mu=\nu+1}^{M}\overline{\xi_i^\mu}\,\overline{\xi_j^\mu}\Big)\Big)^{1/2}. \end{aligned}$$

Now

$$\begin{aligned} & E\exp\Big(-\frac{2t}{(1-a^2)N}\sum_{i\in J, j\notin J}\sum_{\mu=\nu+1}^{M}\overline{\xi_i^\mu}\,\overline{\xi_j^\mu}\Big) \\ = \quad & E^{\nu<\mu\leq M-1}\Big(\exp\Big(-\frac{2t}{(1-a^2)N}\sum_{i\in J, j\notin J}\sum_{\mu=\nu+1}^{M-1}\overline{\xi_i^\mu}\,\overline{\xi_j^\mu}\Big) \\ & E_{J^c}^M E_J^M\Big(\exp\Big(-\frac{2t}{(1-a^2)N}\sum_{i\in J, j\notin J}\overline{\xi_i^M\xi_j^M}\Big)\Big)\Big) \\ = \quad & E^{\nu<\mu\leq M-1}\Big(\exp\Big(-\frac{2t}{(1-a^2)N}\sum_{i\in J, j\notin J}\sum_{\mu=\nu+1}^{M-1}\overline{\xi_i^\mu}\,\overline{\xi_j^\mu}\Big) \\ & E_{J^c}^M\prod_{i\in J}E_i^M\Big(\exp\Big(-\frac{2t}{(1-a^2)N}\overline{\xi_i^M}\sum_{j\notin J}\overline{\xi_j^M}\Big)\Big)\Big) \end{aligned}$$

$$
\begin{aligned}
&\le \; E^{\nu<\mu\le M-1}\Big(\exp\Big(-\frac{2t}{(1-a^2)N}\sum_{i\in J, j\notin J}\sum_{\mu=\nu+1}^{M-1}\overline{\xi_i^\mu}\,\overline{\xi_j^\mu}\Big) \\
&\qquad E^M_{J^c}\prod_{i\in J}\cosh\Big(\xi_i^{M-1}\frac{2(1+|a|)t}{(1-a^2)N}\sum_{j\notin J}\overline{\xi_j^M}\Big)\Big) \\
&\le \; E^{\nu<\mu\le M-1}\Big(\exp\Big(-\frac{2t}{(1-a^2)N}\sum_{i\in J, j\notin J}\sum_{\mu=\nu+1}^{M-1}\overline{\xi_i^\mu}\,\overline{\xi_j^\mu}\Big) \\
&\qquad E^M_{J^c}\prod_{i\in J}\exp\Big(\frac{2(1+|a|)^2t^2}{(1-a^2)^2N^2}(\sum_{j\notin J}\overline{\xi_j^M})^2\Big)\Big) \\
&= \; E^{\nu<\mu\le M-1}\Big(\exp\Big(-\frac{2t}{(1-a^2)N}\sum_{i\in J, j\notin J}\sum_{\mu=\nu+1}^{M-1}\overline{\xi_i^\mu}\,\overline{\xi_j^\mu}\Big) \\
&\qquad E^M_{J^c}\prod_{i\in J}E_{z_i^M}\exp\Big(z_i^M\frac{2(1+|a|)t}{(1-a^2)N}\sum_{j\notin J}\overline{\xi_j^M}\Big)\Big) \\
&= \; E^{\nu<\mu\le M-1}\Big(\exp\Big(-\frac{2t}{(1-a^2)N}\sum_{i\in J, j\notin J}\sum_{\mu=\nu+1}^{M-1}\overline{\xi_i^\mu}\,\overline{\xi_j^\mu}\Big) \\
&\qquad E^M_{J^c}E_{z_J^M}\exp\Big(\frac{2(1+|a|)t}{(1-a^2)N}\sum_{i\in J, j\notin J}z_i^M\overline{\xi_j^M}\Big)\Big),
\end{aligned}
$$

where z_i^M are Gaussian random variables with expectation 0 and identity covariance matrix independent of the ξ_i^μ, $E_{z_i^M}$ denotes the expectation with respect to z_i^M, and finally $E_{z_J^M}$ denotes the expectation with respect to the vector $(z_i^M)_{i\in J}$. Here – among the identities and inequalities already used in the proof of Theorem 2.1 – we have made use of the well known identity

$$
\exp(\frac{1}{2}x^2) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(xy-\frac{1}{2}y^2)dy.
$$

Interchanging the order of integration and using the above technique on every ξ_i^μ we are now able to consecutively replace all the variables ξ_i^μ by Gaussian random variables z_i^μ with expectation zero and identity covariance matrix. This leads to

$$
\begin{aligned}
&E\exp\Big(-\frac{2t}{(1-a^2)N}\sum_{i\in J, j\notin J}\sum_{\mu=\nu+1}^{M}\overline{\xi_i^\mu}\,\overline{\xi_j^\mu}\Big) \\
\le\;& E_z\exp\Big(\frac{2(1+|a|)^2t}{(1-a^2)N}\sum_{i\in J, j\notin J}\sum_{\mu=\nu+1}^{M}z_i^\mu\, z_j^\mu\Big) \\
=\;& E_z\exp\Big(\frac{2(1+|a|)^2t\sqrt{\delta(1-\delta)}}{(1-a^2)}\sum_{\mu=\nu+1}^{M}z^\mu\,\tilde z^\mu\Big) \\
=\;& \Big(E_z\exp\Big(\frac{2(1+|a|)^2t\sqrt{\delta(1-\delta)}}{(1-a^2)}z^1\,\tilde z^1\Big)\Big)^{M-\nu}
\end{aligned}
$$

where (by normalization) the $(z^\mu)_{\mu=\nu+1,\dots,M}$ as well as the $(\tilde z^\mu)_{\mu=\nu+1,\dots,M}$ are now independent Gaussian random variables with expectation 0 and identity covariance matrix, z denotes the vector of the z_i^μ, and E_z is integration with respect to z. Note that the last equality is due to the independence of the Gaussian variables. By Gaussian integration

$$E_z \exp\Big(\frac{2t(1+|a|)^2\sqrt{\delta(1-\delta)}}{(1-a^2)} z^1\, \tilde z^1\Big) = \sqrt{\frac{1}{1-\frac{16t^2\delta(1-\delta)(1+|a|)^4}{(1-a^2)^2}}}$$

if t is small enough (that is, if $\frac{16t^2\delta(1-\delta)(1+|a|)^4}{(1-a^2)^2} < 1$ and moreover, if, for example $\frac{16t^2\delta(1-\delta)(1+|a|)^4}{(1-a^2)^2} < 3/4$

$$\sqrt{\frac{1}{1-\frac{16t^2\delta(1-\delta)(1+|a|)^4}{(1-a^2)^2}}} \le \exp\Big(\frac{16t^2\delta(1-\delta)(1+|a|)^4}{(1-a^2)^2}\Big).$$

Since the term

$$E \exp\Big(-\frac{2t}{(1-a^2)N} \sum_{i\in J, j\notin J} \sum_{\mu=1}^{\nu} \overline{\xi_i^\mu}\, \overline{\xi_j^\mu}\Big)$$

can be treated analogously we arrive at

$$P\Big(H_N(\xi_J^\nu) - H_N(\xi^\nu) \le \varepsilon N\Big) \le \exp\Big(-tN(\delta(1-\delta)\frac{2}{1-a^2} + \varepsilon')\Big) \exp\Big(M \frac{16t^2\delta(1-\delta)(1+|a|)^4}{(1-a^2)^2}\Big)$$

and that means

$$\begin{aligned} &P\Big(\exists\nu\exists J : |J| = \delta N : H_N(\xi_J^\nu) - H_N(\xi^\nu) \le \varepsilon N\Big) \\ \le\ & M\binom{N}{\delta N} \exp\Big(-tN(\delta(1-\delta)\frac{2}{1-a^2} + \varepsilon')\Big) \exp\Big(M \frac{16t^2\delta(1-\delta)(1+|a|)^4}{(1-a^2)^2}\Big) \end{aligned} \tag{18}$$

As $\varepsilon > 0$ as well as $t \ge 0$ were arbitrary (18) simplifies to

$$\begin{aligned} &P\Big(\exists\nu\exists J : |J| = \delta N : H_N(\xi_J^\nu) - H_N(\xi^\nu) \le \varepsilon N\Big) \\ \le\ & M\binom{N}{\delta N} \exp\Big(-tN(\delta(1-\delta)(3+a^2) + \varepsilon/2)\Big) \exp\Big(M(16t^2\delta(1-\delta)(1+|a|)^4)\Big) \end{aligned} \tag{19}$$

(formally by first replacing ε by $\varepsilon/(1-a^2)$ and then $t/(1-a^2)$ by t). To completely get rid of the expression a we replace the right hand side of (19) by the slightly worse bound

$$\begin{aligned} &P\Big(\exists\nu\exists J : |J| = \delta N : H_N(\xi_J^\nu) - H_N(\xi^\nu) \le \varepsilon N\Big) \\ \le\ & M\binom{N}{\delta N} \exp\Big(-tN(3\delta(1-\delta) + \varepsilon/2)\Big) \exp\Big(256t^2 M\delta(1-\delta)\Big). \end{aligned} \tag{20}$$

Obviously, (20) does not depend on a anymore, such that also, if we can find α and δ, such that the right hand side of (20) decreases exponentially they will not depend on a, either. That such α and δ indeed exists, is now fairly standard and can be found in great length, for example, in [N88], [BG92], or [Lö98]. The proof of this fact consists of an application of Stirling's formula to bound

$$\binom{N}{\delta N} \leq \exp(-N(\delta \log \delta + (1-\delta)\log(1-\delta)))$$

together with the choice of the optimal t. For $\alpha \sim \sqrt{\delta(1-\delta)}$ (which is compatible with the above assumptions) the term

$$\exp\Big(-tN(3\delta(1-\delta)+\varepsilon/2)\Big)\exp\Big(256t^2 M\delta(1-\delta)\Big)$$

then is of the order $\exp(-N\sqrt{\delta(1-\delta)})$ which for δ small enough dominates

$$\exp(-N(\delta \log \delta + (1-\delta)\log(1-\delta)))$$

and therefore the probability

$$P\Big(\exists \nu \exists J : |J| = \delta N : H_N(\xi^\nu_J) - H_N(\xi^\nu) \leq \varepsilon N\Big)$$

converges to zero exponentially fast. For the technical details we refer the reader to the references given above. This completes the proof of Theorem 2.2. $\square$

Remark 3.3 *It would be interesting to know whether results in the spirit of the previous sections also hold true, if the correlation of the patterns does not come from a Markov chain (for example if we only require exponentially decreasing correlations or if we try to treat patterns from a higher dimensional Markov random field). On the one hand this seems to be quite plausible since from a theoretical point of view there is no reason why our techniques should be limited to one dimensional Markov chains on the other hand from a technical point of view in all other cases the geometry matrix G is much more difficult to treat and moreover there is no simple way to calculate the moment generating function.*

References

[A89] D.J. Amit; Modelling Brain function; Cambridge University Press, Cambridge (1989) Heidelberg Colloquium on Glassy Dynamics; L. van Hemmen, I. Moregnstern (eds.), Lecture Notes in Physics 275, Springer, Berlin (1987)

[AGS87] D.J. Amit, G. Gutfreund, H. Sompolinsky; Statistical mechanics of neural networks near saturation; Ann. Phys. 173, 30-67 (1987)

[BG92] A. Bovier, V. Gayrard; Rigorous bounds on the storage capacity of the dilute Hopfield model; J. Stat. Phys. 69, 597-627 (1992)

[BG96] A.Bovier, V.Gayrard; An almost sure large deviation principle for the Hopfield model, Ann. Prob. 24, 1444-1475 (1996)

[BG97] A.Bovier, V.Gayrard; The retrieval phase of the Hopfield model, Prob. Th. Rel. Fields 107, 61-98 (1997)

[BG98] A.Bovier, V.Gayrard; Hopfield models as a generalized mean field model, Preprint, to appear in "Mathematics of spin glasses and neural networks", A. Bovier, P.Picco (eds.), "Progress in Probability", Birkhäuser (1998)

[BGP95] A.Bovier, V.Gayrard, P.Picco; Gibbs states for the Hopfield model with extensively many patterns; J. Stat. Phys. 79, 395-414 (1995)

[FP77] L.A. Pastur, A.L. Figotin; Exactly soluble model of a spin-glas; Sov. J. of Low Temprature Phys. 3(6), 378-383 (1977)

[HKP91] J. Hertz, A. Krogh, R. Palmer; Introduction to the Theory of Neural Computation, Addison-Wesley, Redwood City (1991)

[Ho82] J.J. Hopfield; Neural networks and physical systems with emergent collective computational abilities; Proc. nat. Acad. Sci. USA 79, 2554-2558 (1982)

[Lou94] D. Loukianova; Capacité de mémoire dans le modèle de Hopfield; C.R. Acad. Sci. Paris 318, 157-160 (1994)

[Lou97] D. Loukianova; Lower bounds on the restitution error in the Hopfield model; Prob. Th. Rel. Fields 107, 161-176 (1997)

[Lö98a] M. Löwe; On the Strorage Capacity of the Hopfield Model with Weakly Correlated Patterns, Preprint, to appear in: Ann. of Appl. Prob. (1998)

[Lö98b] M. Löwe; On the Strorage Capacity of the Hopfield Model with Biased Patterns, Preprint, to appear in: IEEE Inf. Th. (1998)

[MPRV87] R. McEliece, E. Posner, E. Rodemich. S. Venkatesh; The capacity of the Hopfield associative memory; IEEE Trans. inf. Th. 33, 461-482 (1987)

[N88] C. Newman; Memory capacity in neural networks; Neural Networks 1, 223-238 (1988)

[P96] 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)

[T98] M. Talagrand; Rigorous Results of the Hopfield Model with Many Patterns; Prob. Th. Rel. Fields 110, 177-276 (1998)