Neural Processing Letters **20:** 23–38, 2004.

Further Results on the Asymptotic Memory Capacity of the Generalized Hopfield Network

JIANWEI WU, JINWEN MA★ and QIANSHENG CHENG
Department of Information Science, School of Mathematical Sciences And LMAM, Peking University, Beijing, 100871, China. e-mail: jwna@math.pku.edu.cn

Abstract. This paper presents a further theoretical analysis on the asymptotic memory capacity of the generalized Hopfield network (GHN) under the perceptron learning scheme. It has been proved that the asymptotic memory capacity of the GHN is exactly $2(n-1)$, where n is the number of neurons in the network. That is, the GHN of n neurons can store $2(n-1)$ bipolar sample patterns as its stable states when n is large, which has significantly improved the existing results on the asymptotic memory capacity of the GHN.

Key words. associative memory, asymptotic memory capacity, Hopfield network, pattern recognition, perceptron learning algorithm

1. Introduction

As a typical associative memory model, Hopfield network has been intensively applied to pattern recognition via the sum-of-outer product scheme [1, 2]. However, it had been found by theoretical analysis that the asymptotic memory capacity of Hopfield network of n neurons is only $n/(4\log n)$ and also that the sum-of-outer product scheme cannot be sure to store a set of sample patterns in general [3, 4]. As a matter of fact, these disadvantages seriously restrict the application of Hopfield network to associative memory.

In order to overcome these disadvantages, the generalized Hopfield network (GHN) has been proposed in [5] via using a general zero-diagonal weight matrix instead of the symmetric zero-diagonal weight matrix. Actually, it has been shown in [5] that the GHN with stable states can be stable in the same way as a Hopfield network. Therefore, the GHN can be also applied to associative memory with some learning scheme that makes a set of sample patterns be the stable states of a GHN. Moreover, several such learning schemes have been established on the GHNs for associative memory (e.g. [4, 6–10]). By the theoretical analysis [11], it has been further proved that the asymptotic memory capacity of the GHN of n neurons under the perceptron learning scheme is no less than $(n-1)$, which is much greater than that of Hopfield network under the sum-of-outer product scheme.

In this paper, we have made further theoretical analysis on the asymptotic memory capacity of the GHN and proved that the asymptotic memory capacity of the GHN of n neurons is exactly $2(n-1)$. In the sequel, we introduce our theorem on the

★ Corresponding author.

asymptotic memory capacity of the GHN in Section 2. Section 3 describes several lemmas to prove an important fact that is needed for the proof; the proof is contained in Section 4. Section 5 gives a brief conclusion.

2. The Main Theorem

We begin with a brief description of the GHN model. A GHN is composed of n interconnected neurons defined by $(\mathbf{W}, \theta)$ where $\mathbf{W}$ is an $n \times n$ zero-diagonal matrix with element $w_{i,j}$ denoting the weight on the connection from neuron j to neuron i, and θ is a vector of dimension n with component θ_i denoting the threshold of neuron i. For simplicity, we let $\theta_i = 0$ for $i = 1, 2, \ldots, n$ in this paper.

Every neuron can be in one of two possible states, either 1 or -1. At time t, we let $x_i(t)$ be the state of neuron i and $X(t) = [x_1(t), x_2(t), \ldots, x_n(t)]^{\mathrm{T}}$ be the state of the network. Then, the state of neuron i at time $t+1$ is computed by

$$x_i(t+1) = \mathrm{Sgn}(H_i(t)) = \begin{cases} 1, & \text{if } H_i(t) \geq 0, \\ -1, & \text{otherwise,} \end{cases} \tag{1}$$

where

$$H_i(t) = \sum_{j=1}^{n} w_{i,j} x_j(t).$$

The next state of the network, i.e., $X(t+1)$, can be computed from the current state by performing the evaluation of Equation (l) either at each neuron of the network in the synchronous operation model or at a single neuron at each time in the asynchronous operation mode. However, the stable state $X = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$ of the network in the both operation modes is the same and can be defined by

$$x_i = \mathrm{Sgn}\left(\sum_{j=1}^{n} w_{i,j} x_j \right), \quad \text{for } i = 1, 2, \ldots, n. \tag{2}$$

As a dynamic system, the GHN can have the similar characteristics of content-addressed memory as a Hopfield network, especially in randomly asynchronous mode [5]. When the network starts with an initial state nearby some stable state which constitutes a stored pattern in the memory, it evolves and probably enters the stable state. For associative memory, we have a given sample set $\mathcal{S} = \{X^1, X^2, \ldots, X^m\}$ that consists of m different sample patterns (vectors) in $\{-1, 1\}^n$, where

$$X^j = [x_{j,1}, x_{j,2}, \ldots, x_{j,n}]^{\mathrm{T}} \qquad (j = 1, 2, \ldots, m). \tag{3}$$

Then, the key problem concerning the use of a GHN as an associative memory is how to construct its matrix $\mathbf{W}$ that enables each of $X^1, X^2, \ldots, X^m$ to be a stable state of the network when it is possible. For clarity, we introduce the concept of storablity as follows.

DEFINITION 1. A sample set $\mathcal{S} = \{X^1, X^2, \ldots, X^m\}$ is storable if all m sample patterns $X^1, X^2, \ldots, X^m$ can be the stable states of some GHN $\mathbf{N} = (\mathbf{W}, \mathbf{0})$ where $\mathbf{W}$ is a zero-diagonal real matrix and $\mathbf{0}$ is the zero vector of dimension n.

If $\{X^1, X^2, \ldots, X^m\}$ is storable, the perceptron learning algorithm [12] can be implemented to compute the rows of the desired $\mathbf{W}$ from neuron 1 to neuron n independently, with the threshold value of the perceptron being fixed to be zero. Clearly, $\mathbf{W}$ can be successfully constructed by this perceptron learning scheme if and only if the sample set is storable. Therefore, we can study the asymptotic memory capacity of the GHN under the perceptron learning scheme through the concept of storablity.

Assuming that each sample set $\{X^1, X^2, \ldots, X^m\}$ (which has m different sample vectors without ordinal relation) has the equal probability $1 \Big/ \binom{2^n}{m}$ over the sample set space, we define the probability sequence of storage of the GHN as follows.

$$P(m, n) = P(\{\{X^1, X^2, \ldots, X^m\} \text{ is storable.}\}),$$

where $m, n \in \mathcal{N} = \{1, 2, \ldots\}$. Obviously, $P(m, n)$ decreases with m.

We now introduce the mathematical definition of the asymptotic memory capacity of the GHN based on $P(m, n)$ as follows.

DEFINITION 2. An integer function $C(n)$ is the asymptotic memory capacity of the GHN if it satisfies the following two conditions for any $\varepsilon > 0$:

(i) $\lim_{n\to\infty} P(C(n)(1+\epsilon), n) = 0$,

(ii) $\lim_{n\to\infty} P(C(n)(1-\epsilon), n) = 1$.

With the above preparations, we are ready to introduce our main theorem.

THEOREM 1. *The asymptotic memory capacity* $C(n)$ *of the GHN under the perceptron learning scheme is* $2(n-1)$.

The proof of the main theorem is given in the following sections. This theorem exactly gives the asymptotic memory capacity of the GHN under the peceptron learning scheme. That is, $C(n) = 2(n-1)$. It almost reaches at the upper bound of the asymptotic memory capacity of the GHN obtained in [5]. As to the lower bound $n-1$ of the asymptotic memory capacity of the GHN obtained in [5], this result is significantly improved. Moreover, $C(n)$ is much greater than $n/(4\log n)$ – the asymptotic memory capacity of Hopfield network of n neurons under the sum-of-outer product scheme. Therefore, the GHN has great potentiality for associative memory.

3. Lemmas

In this section, we prove an important fact that will be key to the proof of the main theorem. It is proved step by step with several lemmas described in the following.

We let B^n be the set of all n-dimensional binary vectors, i.e., $B^n = \{0, 1\}^n$, and define

$$A_{n,k} = \{A = (a_{ij})_{n\times n} : a_{ij} \in \{0, 1\}; \mathrm{rank}(A) = k.\}, 1 \le k \le n-1$$
$$\Sigma_n = \{E = \{e_1, e_2, \ldots, e_n\} \subset B^n : e_1, e_2, \ldots, e_n \text{ are linearly independent.}\}$$

That is, Σ_n is the set of all the groups of n linearly independent vectors in B^n, and $A_{n,k}$ is all $n \times n$ binary matrices whose ranks are just $k(1 \le k \le n-1)$.

We assume that each element a_{ij} is an i.i.d. random variable to be 0 or 1 with equiprobability. We let $P(A_{n,k})$ be the probability that its rank is k when we arbitrarily pick up an $n \times n$ binary matrix. Letting A_n be all singular $n \times n$ binary matrices, and adding zero matrix to $A_{n,1}$, we have

$$A_n = |A_{n,1}| + |A_{n,2}| + \cdots + |A_{n,n-1}|,$$

where $|A|$ denotes the number of elements in a set A.

LEMMA 1. *From a set of k m-dimensional binary vectors, we can construct at 2^{2^k} different binary vectors by linear combination.*

See the proof in [13]. □

LEMMA 2. *For a positive integer n, we have*

$$\binom{2^{n+1}}{n+1}(n+1)! = 2^{(n+1)^2} \cdot x_{n+1}, \tag{4}$$

where

$$x_n = \left(1 - \frac{1}{2^n}\right)\left(1 - \frac{2}{2^n}\right)\cdots\left(1 - \frac{n-1}{2^n}\right). \tag{5}$$

Proof. Since $n \ge 1$, we have

$$\frac{\binom{2^{n+1}}{n+1}}{\binom{2^n}{n}} = \frac{2^{2n+1}}{n+1} \cdot \frac{\left(1 - \frac{1}{2^{n+1}}\right)\cdots\left(1 - \frac{n-1}{2^{n+1}}\right)\left(1 - \frac{n}{2^{n+1}}\right)}{\left(1 - \frac{1}{2^n}\right)\cdots\left(1 - \frac{n-2}{2^n}\right)\left(1 - \frac{n-1}{2^n}\right)}.$$

Letting $\alpha(n) = x_{n+1}/x_n$, we get

$$\binom{2^{n+1}}{n+1} = \binom{2^n}{n} \cdot \frac{2^{2n+1}}{n+1}\alpha(n). \tag{6}$$

Multiplying $(n+1)!$ on the both sides of Equation (6), we have

$$\binom{2^{n+1}}{n+1}(n+1)! = \binom{2^n}{n} n!\; 2^{2n+1}\alpha(n). \tag{7}$$

Recursively reducing the number n in the way of Equation (7), we have

$$\begin{aligned}\binom{2^{n+1}}{n+1}(n+1)! &= \binom{2^n}{n} n! 2^{2n+1}\alpha(n) \\ &= \binom{2^{n-1}}{n-1}(n-1)! 2^{2((n-1)+n)+2}\alpha(n-1)\alpha(n) \\ &= \cdots \\ &= \binom{2}{1} 2^{2(1+2+\cdots+(n-1)+n)+n}\alpha(1)\alpha(2)\ldots\alpha(n) \\ &= 2^{(n+1)^2}x_{n+1}.\end{aligned}$$

The proof is completed □

LEMMA 3. *Supposing that u is a random binary vector from B^n and p_k is the probalility that u can be linearly expressed by a group of k random independent vectors (from some n-independent-vector group E in Σ_n), we have that $p_1 \le p_2 \le \cdots \le p_{n-1}$, and*

$$\lim_{n\to\infty} p_{n-1} = 0. \tag{8}$$

Proof. Suppose that C_k is the event that $u(\in B^n)$ is linearly expressed by k linearly independent vectors. When u can be linearly expressed by k vectors, it is certainly linearly expressed by $k+1$ vectors. That is , $C_k \subset C_{k+1}$ and thus $p_k = P(C_k) \le P(C_{k+1}) = p_{k+1}$. Therefore, we have $p_1 \le p_2 \le \ldots \le p_{n-1}$.
As for Eq.(8), we introduce the following notations:

$$\begin{aligned}&\Sigma_k = \{\{e_1, e_2, \cdots, e_k\} \subset B^n : e_1, e_2, \cdots, e_k \text{ are linearly independent.}\},\\ &\Theta_k = \{\{u, e_1, e_2, \cdots, e_k\} : \{e_1, e_2, \cdots, e_k\} \in \Sigma_k, \text{ and } \{u, e_1, e_2, \cdots, e_k\}\\ &\qquad \text{are linearly dependent.}\},\\ &T_k = |\Sigma_k|, \qquad L_k = |\Theta_k|.\end{aligned}$$

According to the definition, we have

$$p_{n-1} = \frac{L_{n-1}}{2^n T_{n-1}} = \frac{L_{n-1}n!}{2^n T_{n-1}n!}. \tag{9}$$

On the one hand, since u can be any binary vector, the set of the total groups of $u, e_1, \cdots, e_{n-1}$ contains Σ_n. That is, $2^n T_{n-1} > T_n = |\Sigma_n|$. Then, $2^n T_{n-1} n! > T_n n! = 2^{n^2} - A_n$. On the other hand, since teh group of $u, e_1, \cdots, e_{n-1} (\in \Theta_{n-1})$ are linearly dependent, $L_{n-1} n! \le A_n$. According to these two equalities, it follows from Eq.(9) that

$$p_{n-1} \le \frac{A_n}{2^{n^2} - A_n} = \frac{\frac{A_n}{2^{n^2}}}{1 - \frac{A_n}{2^{n^2}}}. \tag{10}$$

By Komlos's theorem [13] that $\lim_{n\to\infty} \frac{A_n}{2^{n^2}} = 0$, we finally

$$\lim_{n\to\infty} p_{n-1} \leq \frac{\lim_{n\to\infty} \frac{A_n}{2^{n^2}}}{1 - \lim_{n\to\infty} \frac{A_n}{2^{n^2}}} = 0. \tag{11}$$

The proof is completed □

With above preparation, we now estimate the number of n x n singular binary matrices.

LEMMA 4. *Suppose that M_n is the total number of $n \times n$ binary matrices whose ranks are not larger than $[\log n]$, where $[x]$ is the integer part of a real number x. When n is large enough, we have*

$$M_n < \binom{2^n}{n} n! \left(\frac{n}{2^n}\right) \cdot \frac{\log n}{2^{n(n-[\log n]-\sqrt[3]{n}-1)}} \cdot \frac{1}{x_n}, \tag{12}$$

where the logarithm base is 10 and x_n is given by Equation (5).

Proof. For a k-rank $n \times n$ binary matrix, there are k row vectors which are linearly independent, and can express each of the other $n-k$ row vectors by linear combination. Clearly, the total number of k linearly independent vector groups is at most $\binom{2^n}{k}$. On the other hand, according to Lemma 1, the number of binary vectors which are constructed by the linear combination of these k vectors is at most 2^{2^k}. Moreover, the total number of groups of $n-k$ vectors repeatedly selected from the 2^{2^k} binary vectors is $\binom{2^{2^k}+n-k-1}{n-k}$.

Therefore, the number of $n \times n$ binary matrices whose ranks are no more than k is at most $\binom{2^n}{k}\binom{2^{2^k}+n-k-1}{n-k} n!$

Since

$$2^k \leq 2^{[\log n]} \leq 2^{\log n} = 2^{\log_2 n \frac{\log n}{\log_2 n}},$$

and $\frac{\log n}{\log_2 n} < \frac{1}{3}$ when n is large enough, we have

$$2^k < 2^{\log_2 n \cdot \frac{1}{3}} = \sqrt[3]{n}, \tag{13}$$

when $n > m_1$, where m_1 is a positive integer.

For $1 \leq k \leq [\log n], \lim_{n\to\infty} \frac{n-k-1}{2^{\sqrt[3]{n}}} = 0$. Then, there exists another positive integer m_2 such that $n-k-1 < 2^{\sqrt[3]{n}}$ when $n > m_2$.

when $n > \max(m_1, m_2)$, we have

$$2^{2^k} + n - k - 1 < 2 \cdot 2^{\sqrt[3]{n}} = 2^{\sqrt[3]{n}+1},$$

and thus

$$\frac{\binom{2^n}{k}\binom{2^{2^k}+n-k-1}{n-k}}{\binom{2^n}{n}} \leq \frac{\binom{2^n}{k}\binom{2^{\sqrt[3]{n}+1}}{n-k}}{\binom{2^n}{n}}$$

$$< \frac{n!}{k!2^{n(n-k)}x_n} \cdot \frac{2^{\sqrt[3]{n}+1}!}{(n-k)!(2^{\sqrt[3]{n}+1}-n+k)!}$$

$$= \frac{n(n-1)\cdots(n-k+1)}{k!2^{n(n-k)}x_n} \cdot 2^{\sqrt[3]{n}+1}(2^{\sqrt[3]{n}+1}-1)\cdots(2^{\sqrt[3]{n}+1}-n+k+1)$$

$$< \frac{n^k}{k!2^{n(n-k)}x_n} \cdot 2^{(n-k)(\sqrt[3]{n}+1)}$$

$$= \left(\frac{n}{2^{\sqrt[3]{n}+1}}\right)^k \cdot \frac{1}{k!} \cdot \frac{1}{2^{n(n-k-\sqrt[3]{n}-1)}x_n}$$

$$< \left(\frac{n}{2^{\sqrt[3]{n}}}\right) \cdot \frac{1}{2^{n(n-k-\sqrt[3]{n}-1)}} \cdot \frac{1}{x_n}$$

Therefore, when n is large enough, we get

$$\binom{2^n}{k}\binom{2^{2^k}+n-k-1}{n-k}n! < \binom{2^n}{n}n!\left(\frac{n}{2^{\sqrt[3]{n}}}\right)\frac{1}{2^{n(n-k-\sqrt[3]{n}-1)}} \cdot \frac{1}{x_n}.$$

Summing up the both sides of the above inequality from $k = 1$ to $[\log n]$, we finally have

$$M_n \leq \sum_{k=1}^{[\log n]}\binom{2^n}{k}\binom{2^{2^k}+n-k-1}{n-k}n! < \binom{2^n}{k}n!\left(\frac{n}{2^{\sqrt[3]{n}}}\right)\frac{\log n}{2^{n(n-[\log n]-\sqrt[3]{n}-1)}} \cdot \frac{1}{x_n}.$$

The proof is completed □

LEMMA 5. *When n is large enough and the rank $k \geq [logn] + 1$, we have*

$$P(\hat{A}_{n,k}) \leq \frac{|\Sigma_n|}{\binom{2^n}{n}} 2^{n+1} p_k^{n-1} \binom{2n-2}{n-1}\binom{n}{[\frac{n}{2}]}, \tag{14}$$

where $\hat{A}_{n,k}$ is considered as the event that a random $n \times n$ binary matrix with different column vectors takes the rank of k.

Proof. For each binary matrix in $\hat{A}_{n,k}$, here are k linearly independent column vectors, while the other column vectors can be expressed by the linear combination of these k vectors. We consider that these k vectors are randomly selected from an $E = \{e_1, e_2, \cdots, e_n\} \in \sum_n$. For clarity, we let them be $\{e_{j1}, e_{j2}, \cdots, e_{jk}\} \subset E$ respectively. Clearly, an over estimation of the total number of these $\{e_{j1}, e_{j_2}, \cdots, e_{jk}\}$ *is* $|\sum_n|\binom{n}{k}$.

For convenience of analysis, we introduce the following notations:

$$D = \left\{ u = \sum_{i=1}^{k} l_i e_{j_i} : u \in B^n - \{0\}; \{e_{j1}, e_{j2}, \cdots, e_{jk}\} \subset E \in E_n; l_i \in R. \right\}$$
$$D' = D \cup \{0\}$$

Then, the other $n-k$ column vectors can only be selected from D'. Moreover, a vector in D' an be selected repeatedly. According to the formula of the total probability, we have

$$P(\hat{A}_{n,k}) = \sum_{h=1}^{2^n} P(|D'| = h) P(\hat{A}_{n,k} \| D'| = h).$$

Since $\{e_{j1}, e_{j2}, \ldots, e_{jk}\} \subset D \subset D'$, $P(|D'| = h) = P(|D| = h-1) = 0$ for $1 \le h \le k$.

When $h \ge k+1$, there are $h-1$ different vectors $u_1, u_2, \ldots, u_{h-1}$ in D. According to Lemma 3 and that $u_1, u_2, \ldots, u_{h-1}$ are independently expressed by some $(e_{j1}, e_{j2}, \cdots, e_{jk})$, we have

$$P(|D'| = h) = P(|D| = h-1) = P\left(\bigcap_{i=1}^{h-1} U_i\right) = \prod_{i=1}^{h-1} P(U_i) = p_k^{h-1},$$

where U_i is the event that u_i can be linearly expressed by k vectors in an arbitrary group $E \in \sum_n$.

Furthermore, when D' contains h different vectors, since the column vectors of the matrix subject to $\hat{A}_{n,k}$ should be different, the oher $n-k$ column vectors can not be selected repeatedly. Moreover, $(e_{j1}, e_{j2}, \cdots, e_{jk})$ in D can not be selected as these $n-k$ column vectors. So, we arbitrarily pick up a base from Σ_n, and randomly select k vectors from the base with $|D| = h-1$, the total number of the matrices subject to $\hat{A}_{n,k}$ is overestimated by

$$|\Sigma_n| \binom{n}{k} \binom{h-k}{n-k} n!$$

For the existence of such a matrix, $\binom{h-k}{n-k} \ge 1$ is necessary. That is, there must be at least n vectors in D', that is $h \ge n$. Otherwise, if $|D'| = h < n$, this kind of matrix does not exist and $P(\hat{A}_{n,k} \| D'| = h) = 0$.

Then, we have

$$\begin{aligned} P(\hat{A}_{n,k}) &= \sum_{h=1}^{2^n} P(|D'| = h) P(\hat{A}_{n,k} \| D'| = h) \\ &= \sum_{h=n}^{2^n} p_k^{h-1} \frac{|\Sigma_n| \binom{n}{k} \binom{h-k}{n-k}}{2^{n^2}} n!. \end{aligned} \tag{15}$$

According to Lemma 2, $\binom{2^n}{n} n! = 2^{n^2} x_n$ and $0 < x_n < 1$, we further have

$$P(\hat{A}_{n,k}) \leq \frac{|\Sigma_n|\binom{n}{k}}{\binom{2^n}{n}} p_k^{n-1} \sum_{h=n}^{2^n} p_k^{h-n} \binom{h-k}{n-k}. \tag{16}$$

When n is large enough, since $p_k \leq p_{n-1} \leq \frac{1}{4}$, it can be easily verified that $p_k^{h-n}\binom{h-k}{n-k}$ decreases with h for $h \geq 2n-k$. Therefore, we have

$$\begin{aligned}
\sum_{h=n}^{2^n} p_k^{h-n}\binom{h-k}{n-k} &= \sum_{h=n}^{2n-k-1} p_k^{h-n}\binom{h-k}{n-k} + \sum_{h=2n-k}^{2^n} p_k^{h-n}\binom{h-k}{n-k} \\
&< \sum_{h=n}^{2n-k-1} \binom{h-k}{n-k} + p_k^{h-k}\binom{2n-2k}{n-k}(2^n-2n+k) \\
&< \binom{2n-2k}{n-k+1} + p_k^{n-k}\binom{2n-2k}{n-k}2^n.
\end{aligned}$$

With this result, it follows Eq.(16) that

$$\begin{aligned}
P(\hat{A}_{n,k}) &\leq \frac{|\Sigma_n|\binom{n}{k}}{\binom{2^n}{n}} p_k^{n-1} \sum_{h=n}^{2^n} p_k^{h-n}\binom{h-k}{n-k} \\
&\leq \frac{|\Sigma_n|\binom{n}{k}}{\binom{2^n}{n}} p_k^{n-1}\left[\binom{2n-2k}{n-k+1} + p_k^{n-k}\binom{2n-2k}{n-k}2^n\right] \\
&< \frac{|\Sigma_n|\binom{n}{k}}{\binom{2^n}{n}} p_k^{n-1}\binom{2n-2k}{n-k}2^{n+1}.
\end{aligned}$$

Because $\binom{2n-2k}{n-k}$ decreases with $k(k<n)$, $\binom{n}{k} \leq \binom{n}{[\frac{n}{2}]}$, we finally have for $k > [logn]$

$$P(\hat{A}_{n,k}) \leq \frac{|\Sigma_n|}{\binom{2^n}{n}} 2^{n+1} p_k^{n-1} \binom{2n-2}{n-1}\binom{n}{[\frac{n}{2}]}.$$

The proof is completed □

LEMMA 6. *Suppose that E_n is the event that a random $n \times n$ binary matrix with different row vectors is singular. We have for large n*

$$P(E_n) \leq \frac{1}{2^{n(n-log\ n-\sqrt[3]{n}-1)}} + O((16p_{n-1})^{n-1}), \tag{17}$$

where $O(x)$ is an infinitesimal with the same order of an infinitesimal x.

Proof. By the definition of $\hat{A}_{n,k}$, we have

$$E_n = \bigcup_{k=1}^{n-1} \hat{A}_{n,k}.$$

Since $\bigcup_{k=1}^{[logn]} \hat{A}_{n,k} \subset \bigcup_{k=1}^{[logn]} \hat{A}_{n,k}, p_k \leq p_{n-1}(k \leq n-1)$, according to Lemmas 4&5, we have for large enough n

$$\begin{aligned}
P(E_n) &= P(\bigcup_{k=1}^{n-1} \hat{A}_{n,k}) = P(\bigcup_{k=1}^{[logn]} \hat{A}_{n,k}) + P(\bigcup_{k=[logn]+1}^{n-1} \hat{A}_{n,k}) \\
&\leq P(\bigcup_{k=1}^{[logn]} A_{n,k}) + \sum_{k=[logn]+1}^{n-1} P(\hat{A}_{n,k}) \\
&< \frac{M_n}{2^{n^2}} + \frac{|\sum_n|}{\binom{2^n}{n}} 2^{n+1} \binom{2n-2}{n-1} \binom{n}{\left[\frac{n}{2}\right]} \sum_{k=[logn]+1}^{n-1} p_k^{n-1} \\
&< \frac{M_n}{2^{n^2}} + \frac{|\sum_n|}{\binom{2^n}{n}} 2^{n+1} \binom{2n-2}{n-1} \binom{n}{\left[\frac{n}{2}\right]} p_{n-1}^{n-1}(n-[logn]-2) \\
&< \frac{M_n}{2^{n^2}} + \frac{|\sum_n|}{\binom{2^n}{n}} 2^{n+1} \binom{2n-2}{n-1} \binom{n}{\left[\frac{n}{2}\right]} np_{n-1}^{n-1}.
\end{aligned} \tag{18}$$

By the fact that $\lim_{n\to\infty} \frac{|\sum_n|}{\binom{2^n}{n}} = 1,\ \lim_{n\to\infty} \frac{\binom{2n-2}{n-1}}{\frac{2^{2n-2}}{\sqrt{2n-2}}} = \sqrt{\frac{2}{\pi}}$, and $\lim_{n\to\infty} \binom{n}{\left[\frac{n}{2}\right]} / \frac{2^n}{\sqrt{n}} = \sqrt{\frac{2}{\pi}}$,

we have for large n

$$\begin{aligned}
\lim_{n\to\infty} \frac{\frac{|\sum_n|}{\binom{2^n}{n}} 2^{n+1} \binom{2n-2}{n-1} \binom{n}{\left[\frac{n}{2}\right]} np_{n-1}^{n-1}}{(16p_{n-1})^{n-1}} &= \lim_{n\to\infty} \frac{|\sum_n|}{\binom{2^n}{n}} \cdot \frac{\binom{2n-2}{n-1}}{\frac{2^{2n-2}}{\sqrt{2n-2}}} \cdot \frac{\binom{n}{\left[\frac{n}{2}\right]}}{\frac{2^n}{\sqrt{n}}} \cdot \frac{8}{\sqrt{2-\frac{2}{n}}} \cdot \frac{(2^4 p_{n-1})^{n-1}}{(16p_{n-1})^{n-1}} \\
&= \frac{8\sqrt{2}}{\pi}.
\end{aligned}$$

Since $\lim_{n\to\infty} p_{n-1} = 0$, we then have

$$\frac{|\sum_n|}{\binom{2^n}{n}} 2^{n+1} \binom{2n-2}{n-1} \binom{n}{\left[\frac{n}{2}\right]} np_{n-1}^{n-1} = O((16p_{n-1})^{n-1}). \tag{19}$$

On the other hand, according to Lemma 2, it is clear that for large n

$$
\begin{aligned}
\frac{M_n}{2^{n^2}} &< \frac{\binom{2^n}{n} n!}{2^{n^2}} (\frac{n}{2^{\sqrt[3]{n}}}) \frac{\log n}{2^{n(n-[logn]-\sqrt[3]{n}-1)}} \cdot \frac{1}{x_n} \\
&= (\frac{n}{2^{\sqrt[3]{n}}}) \frac{\log n}{2^{n(n-[logn]-\sqrt[3]{n}-1)}} \\
&< \frac{1}{2^{n(n-[logn]-\sqrt[3]{n}-1)}}.
\end{aligned} \tag{20}
$$

Summing up the results of Eqs. (19)&(20), we have from Eq.(18) that

$$
P(E_n) \leq \frac{1}{2^{n(n-\log n-\sqrt[3]{n}-1)}} + O((16p_{n-1})^{n-1}). \tag{21}
$$

The proof is completed □

The order of $P(E_n) \to 0$ given by Lemma 6 is considerably improved in comparison with the order obtained by Komlos [13]. Actually, this accurate order provides a key to the proof of the main theorem in the next section. Although this result is for $n \times n$ binary matrices, but it holds well for $n \times n$ bipolar matrices since the probability of singular bipolar matrices over all bipolar matrices is just that of singular binary matrices over all binary matrices. Therefore, we will use this result directly for bipolar matrices in the next section.

4. The Proof of the Main Theorem

We begin to give some definitions and results on the perceptron with bipolar input variables. Actually, each neuron in a GHN can be considered as a perceptron with bipolar input variables. Mathematically, a percepron is defined by a weight vector $W = [w_1, w_2, \cdots, w_n]^{\mathrm{T}} \in R^n$ and a threshold value θ such that for an input $X = [x_1, x_2, \cdots, x_n]^{\mathrm{T}} \in R^n$, its output $y = \mathrm{Sgn}(W^{\mathrm{T}}X - \theta)$. Here, we let $\theta = 0$ and $X \in B^n$. For a dichotomy $\{\chi^+, \chi^-\}$ of $\mathcal{S}$, i.e., $\mathcal{S}$ is divided into two subsets χ^+ and χ^-, if there exists a weight vector $W \in R^n$ such that

$$
\begin{aligned}
&W^{\mathrm{T}}X \geq 0, \quad \text{if} \quad X \in \chi^+, \\
&W^{\mathrm{T}}X < 0, \quad \text{if} \quad X \in \chi^-,
\end{aligned}
$$

it is called to be homogeneously linearly separable. In this situation, a perceptron can be implemented to realize such a binary classification by the perceptron learning algorithm.

In the same way, we can define the probability sequence of storage of the perceptron as follows.

$$
H(m,n) = P(\{\{X^1, X^2, \ldots, X^m\} \text{ is homogeneously linearly separable }\}),
$$

where $m, n \in \mathcal{N} = \{1, 2, \ldots\}$. When $\mathcal{S} = \{X^1, X^2, \ldots, X^m\}$ is in general position, that is, each group of $\{X^{i_1}, \ldots, X^{i_k}\} \subset \mathcal{S}$ are linearly independent if $k \leq n$, Cover [14] proved that

$$H(m,n) = C(m,n) = \frac{1}{2^{m-1}} \sum_{k=0}^{n-1} \binom{m-1}{k}.$$

Moreover, Cover [14] further proved that

$$\lim_{n \to \infty} C(2n(1+\varepsilon), n) = \lim_{n \to \infty} \frac{1}{2\pi} \int_{-\infty}^{\frac{-2n\varepsilon}{\sqrt{2n(1+\varepsilon)}}} \mathrm{e}^{\frac{-t^2}{2}} \mathrm{d}t = 0, \tag{22}$$

$$\lim_{n \to \infty} C(2n(1-\varepsilon), n) = \lim_{n \to \infty} \frac{1}{2\pi} \int_{-\infty}^{\frac{-2n\varepsilon}{\sqrt{2n(1+\varepsilon)}}} \mathrm{e}^{\frac{-t^2}{2}} \mathrm{d}t = 1, \tag{23}$$

where $\varepsilon > 0$, which leads to the well-known result that the asymptotic memory capacity of the perceptron with an input $X \in R^n$ is $2n$.

Furthermore, Budinich [15] gave the following inequality:

$$C(m,n) - \frac{2}{2^m \binom{2^n}{m}} \sum_{\pi_m} \sum_{k=1}^{n-1} a_k(\pi_m, n) \leq H(m,n) \leq C(m,n), \tag{24}$$

where $\pi_m = \mathcal{S} = \{X^1, X^2, \cdots, X^m\}$ and $a_k(\pi_m, n)$ is the number of groups of k samples in π_m which are linearly dependent. Clearly, for $k \leq n-1$, $\frac{1}{\binom{m}{k}\binom{2^n}{m}} \sum_{\pi_m} a_k(\pi_m, n)$ is just the probability of the event that k samples out of the m are linearly dependent. It is certainly not larger than that of the event that a random π_m is not in general position.

For $m > n$, we further define

$$\Pi = \{\pi_m \subset B^n : \pi_m \text{ is not in general position.}\}$$
$$\Pi_1 = \{\pi_n \subset B^n : \pi_n \text{ is linear dependent.}\}$$

Since for any $\pi_m \in \Pi$, there exists a $\pi_n \in \Pi_1$ with $\pi_n \subset \pi_m$, we have

$$|\Pi| \leq |\Pi_1| \binom{2^n - n}{m - n}. \tag{25}$$

Suppose that F_m is the event that π_m is not in general position. It follows from Equation (25) that

$$P(F_m) = \frac{|\Pi|}{\binom{2^n}{m}} \leq \frac{|\Pi_1| \binom{2^n - n}{m - n}}{\binom{2^n}{m}} = \frac{\Pi_1}{2^{n^2}} \cdot \frac{2^{n^2} \binom{2^n - n}{m - n}}{\binom{2^n}{m}}.$$

By the following facts

$$\binom{2^n}{m} = \frac{2^{nm}}{m!}x_m, \binom{2^n - n}{m - n} = \frac{2^{n(m-n)}}{(m-n)!} \cdot \frac{x_{m+n}}{x_n},$$

and since $x_l > \frac{1}{\sqrt{2}}$ when l is large enough, we have for large n

$$\frac{2^{n^2}\binom{2^n - n}{m - n}}{\binom{2^n}{m}} = \frac{m!}{(m-n)!} \cdot \frac{x_{m+n}}{x_n x_m} < n!\binom{m}{n} \cdot \frac{1}{x_n x_m}$$
$$< 2(n!)\binom{m}{n}. \tag{26}$$

Thus, we have

$$P(F_m) = \frac{|\Pi|}{\binom{2^n}{m}} \le \frac{|\Pi_1|}{2^{n^2}} \cdot 2(n!)\binom{m}{n}.$$

Since $|\prod_1 |n! \le |E_n|$, where $|E_n|$ denotes the number of the matrices subject to the event E_n, we further have

$$P(F_m) = \frac{|\prod|}{\binom{2^n}{m}} \le \frac{|E_n|}{2^{n^2}} \cdot 2\binom{m}{n} = P(E_n) \cdot 2\binom{m}{n}. \tag{27}$$

Because $\frac{1}{2^m}\sum_{k=1}^{n-1}\binom{m}{k} \le 1$, it follows from Eq.(24) and Eq.(26) that

$$C(m,n) - \frac{|E_n|}{2^{n^2}} \cdot 4\binom{m}{n} \le H(m,n) \le C(m,n). \tag{28}$$

We are now ready to prove the main theorem.

Proof of Theorem 1. For a GHN $N = (\mathbf{W}, \mathbf{0})$, neuron i can be considered as a perceptron with a weight vector $W_i = [w_{i1}, \ldots, w_{i,i-1}, w_{i,i+1} \ldots, w_{i,n}]^{\mathrm{T}} \in R^{n-1}$, i.e., the ith row of $\mathbf{W}$ except the diagonal element $w_{i,i}$ a zero threshold and an input vector $X(i) = [x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n]^{\mathrm{T}}$ based on the input X of the network. Then, a state $X = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$ of the network is stable if $x_i = \mathrm{Sgn}(W_i^{\mathrm{T}} X(i))$ for $i = 1, 2, \ldots, n$.

For a sample set $\mathcal{S} = \{X^1, X^2, \ldots, X^m\}$, it is storable if and only if for each i there exists a weight vector W_i such that $\mathrm{Sgn}(W_i^{\mathrm{T}} X^\mu(i)) = x_{\mu,i}$ for $\mu = 1, 2, \ldots, m$, where $X^\mu(i) = [x_{\mu,1}, \ldots, x_{\mu,i-1}, x_{\mu,i+1}, \ldots, x_{\mu,n}]^{\mathrm{T}}$. For convenience, we let B_i is the event that $\{X^1(i), X^2(i), \ldots, X^m(i)\}$ can be classified according to $x_i^1, x_i^2, \ldots, x_i^m$, respectively, by the perceptron with a weight vector W_i (i.e., neuron i). Then, we have

$$P(m,n) = P\left(\bigcap_{i=1}^{n} B_i\right) \tag{29}$$

For clarity, we let $P_i(m,n) = P(B_i)$ for $i = 1, 2, \cdots, n$. It is clear that $P_i(m,n) = H(m, n-1)$. Thus, $P(B_1) = \cdots = P(B_n)$. According to Cover's result [14] or directly from Equations (17) and (28), we have for $\varepsilon > 0$

$$\lim_{n\to\infty} P_i(2(n-1)(1+\varepsilon), n) = \lim_{n\to\infty} H(2(n-1)(1+\varepsilon), n-1) = 0. \tag{30}$$

Because $P(m,n) = P(\bigcap_{i=1}^n B_i) \le P(B_i)(1 \le i \le n)$, from Equation (29) we further have

$$\lim_{n\to\infty} P(2(n-1)(1+\varepsilon), n) = 0. \tag{31}$$

On the other hand, we have

$$\begin{aligned} P(m,n) &= P(\bigcap\nolimits_{i=1}^n B_i) = 1 - P(\bigcup\nolimits_{i=1}^n \bar{B}_i) \ge 1 - \sum\nolimits_{i=1}^n P(\bar{B}_i) \\ &= 1 - nP(\bar{B}_1) = 1 - n(1 - P_1(m,n)) = 1 - n(1 - H(m, n-1)). \end{aligned} \tag{32}$$

Since

$$H(m, n-1) \ge C(m, n-1) - 4\binom{m}{n-1} P(E_{n-1}),$$

we further have

$$\begin{aligned} & n(1 - H(2(n-1)(1-\varepsilon), n)) \\ & \quad \le n(1 - C(2(n-1)(1-\varepsilon), n-1) + 4\binom{2(n-1)}{n-1}) P(E_{n-1}) \\ & \quad \le n(1 - C(2(n-1)(1-\varepsilon), n-1)) + 4n\binom{2(n-1)}{n-1} P(E_{n-1}). \end{aligned}$$

From Equation (23), it can be easily observed that $1 - C(2(n-1)(1-\epsilon), n-1)$ attenuates to zero exponentially with $\frac{1}{n}$. Certainly, $1 - C(2(n-1)(1-\epsilon), n-1) = O(\frac{1}{n})$. We then get

$$\lim_{n\to\infty} n(1 - C(2(n-1)(1-\epsilon), n-1)) = 0. \tag{33}$$

According to Stirling formula, we have

$$\lim_{n\to\infty} n\binom{2n}{n} \Big/ \sqrt{\frac{n}{\pi}} 2^{2n} = 1.$$

Then, it follows from Lemma 6 that

$$\lim_{n\to\infty} n\binom{2(n-1)}{n-1} P(E_{n-1}) \le K \lim_{n\to\infty} \{\sqrt{\frac{n}{\pi}} 2^{2n} (16p_{n-1})^{n-2}\} = 0. \tag{34}$$

where K is a positive constant.

Based on Equations (33) and (34), we have $\lim_{n\to\infty} n(1 - H(2(n-1)\times (1-\epsilon), n)) = 0$. Therefore, it follows from Equation (32) that

$$\lim_{n\to\infty} P(2(n-1)(1-\epsilon), n) = 1. \tag{35}$$

Summing up the results of Equations (31) and (35), we finally have that the asymptotic memory capacity of the GHN under the perceptron learning scheme is $2(n-1)$.

The proof is completed $\square$

5. Conclusion

We have presented a further analysis of the asymptotic memory capacity of the GHN under the perceptron learning scheme. With a more accurate estimated attenuating order of the probability of the event that a random binary matrix is singular, we have proved that the asymptotic memory capacity of the GHN is exactly $2(n-1)$, where n is the number of neurons in the network. It not only has significantly improved the existing results on the asymptotic memory capacity of the GHN, but also shows that the GHN has great potentiality for associative memory.

Acknowledgement

This is work was supported by the Natural Science Foundation of China(60071004,40035010).

References

[1] Hopfield, J. J.: Neural networks and physical system with emergent collective computational abilities, *Proc. Nat. Acad. Sci, USA*, **79**, (1982), 2554–2558.
[2] Amari, S.: Mathematical foundations of neurocoputing, *Proc. IEEE,* **78**, (1990), 1443–1463.
[3] McMliece, R. E., The capacity of the Hopfield associative memory, *IEEE Trans. Inform. Theory*, IT-33, (1987), 461–483.
[4] Venkatesh, S. S. and Psaltis, D.: Linear and logarithmic capacities in associative memory, *IEEE Trans. Inform. Theory*, **IT-35**, (1989), 558–568.
[5] Ma, J.: The stability of the generalized Hopfield networks in randomly asynchronous mode, *Neural Networks*, **10**(6) (1997), 1109–1116.
[6] Gardner, E.: The space of interactions in neural network models, *J. Phys. A*: *Math. General*, **21**, (1988), 257–270.
[7] Abbott L. F. and Kepler, T. B.: Optimal learning in neural network memories, *J. Phys. A*: *Math. General*, **22**, (1989), L711–L717.
[8] Ma, J.: The asymmetric Hopfield model for associative memory, *Proc. Int. Joint Conf. Neural Networks*(*IJCNN'93*), NAGOYA, 1993, Vol. 3, pp: 2611–2614.
[9] Xu, Z., Hu, G. and Kwong, C.: Asymmetric-type networks: theory and applications, *Neural Networks*, **9**(3) (1996), 483–501.
[10] Ma, J.: The object perceptron learning algorithm on generalised Hopfield net works for associate memory, *Neural Comput. Appl.* **8** (1999), 25–32.
[11] Ma, J.: The asymptotic memory capacity of the generalized Hopfield networks, *Neural Networks*, **12** (1999), 1207–1212.
[12] Rosenblatt, F.: *Principles of Neurodynamics*, Spartan Books: New York, 1962.
[13] Komlos, J.: On The Determinant Of (0,1) Metrices, *Studia Scientiarum Mathemacarum Hungarica*, **2** (1967), 7–21.

[14] Cover, T.: Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, *IEEE Trans. Electron. Comput.*, **EC-14** (1965), 326–334.

[15] Budinich, M.: On linear separability of random subsets of hypercube vertices, *J. Phys. A: Math. General*, **24** (1991), L211–L213.