8 Storage Capacity in Associative Networks for (Near) Perfect Recall

We consider a Hopfield network with the standard Hebb-like learning rule and ask how many memories we can imbed in a network of $N$ neurons with the constraint that we will accept at most one bit (one neuron’s output in only one memory state) of error.

\[
\text{Input} = \sum_{j \neq i}^{N} W_{ij} S_j = \frac{1}{N} \sum_{\mu=1}^{p} \sum_{j \neq i}^{N} \zeta_{\mu}^{j} \zeta_{\mu}^{i} S_j
\]  

(8.201)

where $p$ is the number of stored memories, $N$ is the number of neurons and

\[
W_{ij} \equiv \frac{1}{N} \sum_{\mu=1}^{p} \zeta_{\mu}^{i} \zeta_{\mu}^{j}
\]

(8.202)

is the synaptic weight matrix given by the Hebb rule.

Now, check stability of stored state. Make $S_j = \zeta_1^j$, one of the stored memory states, so that

\[
\text{Input} = \frac{1}{N} \sum_{\mu=1}^{p} \sum_{j \neq i}^{N} \zeta_{\mu}^{j} \zeta_{\mu}^{i} \zeta_1^j 
\]

(8.203)

\[
= \frac{1}{N} \zeta_1^i \sum_{j \neq i}^{N} \zeta_1^j + \frac{1}{N} \sum_{\mu \neq 1}^{p} \zeta_{\mu}^{i} \sum_{j \neq i}^{N} \zeta_{\mu}^{j} \zeta_1^j
\]

On average, the second term is zero, so that

\[
\text{Average of Input} = \left( \frac{N - 1}{N} \right) \zeta_1^i \simeq \zeta_1^i
\]

(8.204)

What is the variance? The second term, summed over random vectors with zero mean, consists of the sum of $(p - 1)$ inner products of vectors with $(N - 1)$ terms. Each term is $+1$ or $-1$, i.e., binomially distributed, so that

\[
\text{Variance of Input} = \left( \frac{1}{N} \right)^2 \cdot (p - 1) \cdot (N - 1) \simeq \frac{p}{N}
\]

(8.205)
This results in a fluctuation to the input with a standard deviation, \( \sigma \), of

\[
\sigma = \pm \sqrt{\frac{P}{N}} \tag{8.206}
\]

Noise hurts only if the magnitude exceeds 1. The noise becomes Gaussian for large \( p \) and \( N \), which is the limit of interest, Thus the probability of an error in the recall of all stored states is

\[
P_{\text{error}} = \frac{1}{\sqrt{2\pi} \, \sigma} \left[ \int_{-\infty}^{-1} e^{-x^2/2\sigma^2} \, dx + \int_{1}^{\infty} e^{-x^2/2\sigma^2} \, dx \right] \tag{8.207}
\]

\[
= \frac{\sqrt{2}}{\sqrt{\pi} \, \sigma} \int_{1}^{\infty} e^{-x^2/2\sigma^2} \, dx \tag{8.208}
\]

\[
= \frac{2}{\sqrt{\pi}} \int_{\sqrt{\frac{2p}{N}}}^{\infty} e^{-x^2} \, dx \tag{8.209}
\]

\[
\equiv \text{erfc} \left( \sqrt{\frac{N}{2p}} \right)
\]

where \( \text{erfc}(x) \) is the complementary error function and we again note that the average of the error term is zero. We recall that for \( \frac{N}{2p} \ll 1 \), the complementary error function may be approximated by an asymptotic form, so that

\[
P_{\text{error}} \simeq \frac{1}{\sqrt{\pi} \, \sqrt{2p}} \frac{2p}{N} e^{-N/2p} \tag{8.210}
\]

We have a nice and closed expression in a relevant limit!

Now \( N \cdot \cdot p \) is total number of “bits” in the network. Suppose only less than one bit can be in error. Then

\[
(1 - P_{\text{error}})^{Np} \geq 1 - \frac{1}{Np} \tag{8.211}
\]

But \( Np \) is large and \( P_{\text{error}} \) will be small by construction, so \( 1 - Np \times P_{\text{error}} \geq 1 - \frac{1}{Np} \) and thus

\[
P_{\text{error}} < \frac{1}{(Np)^2} \tag{8.212}
\]

From the expansion of the gaussian error:

\[
\log [P_{\text{error}}] \simeq -\frac{1}{2} \log \pi - \frac{N}{2p} - \log \frac{N}{2p} \tag{8.213}
\]
From the constraint on the desired error:

\[ \log[P_{\text{error}}] < -2 \log(Np) \]  \hspace{1cm} (8.214)

Thus

\[ -\frac{1}{2} \log \pi - \frac{N}{2p} - \log \frac{N}{2p} < -2 \log (Np) \]  \hspace{1cm} (8.215)

We now let \( N \to \infty \) with \( N/p \) constant. To leading order, we have:

\[ \frac{N}{2p} > 2 \log (Np) \]  \hspace{1cm} (8.216)

To go further and solve for \( p \) in terms of \( N \), we assume that

\[ p = \alpha(N)N \]  \hspace{1cm} (8.217)

then

\[ \frac{N}{2\alpha N} > 2 \log (N\alpha N) \]  \hspace{1cm} (8.218)

\[ 1 > 8\alpha \log N \]  \hspace{1cm} (8.219)

and therefore

\[ p < \frac{1}{8} \frac{N}{\log N} \]  \hspace{1cm} (8.220)

Note that \( p \) has a similar scaling for the choice of a fixed error rate.

Thus we see that an associate memory based on a recurrent Hopfield network stores a number of memories that scales more weakly than the number of neurons if one cannot tolerate any errors upon recall. If a finite number of errors can be tolerated, a statistical mechanical deviation shows \( p < 0.14N \).