Generalization in the Hopfield Model
Leonid B. Litinskii
High Pressure Physics Institute
of Russian Academy of Sciences
Russia, 142092 Troitsk Moscow region,
e-mail: u10740@dialup.podolsk.ru

In the Hopfield model the ability of the network to generalization is studied in the case of the network trained by one input image (the standard).

Basic Model

The maximization problem with the symmetrical connection matrix is considered:
\[
\begin{cases}
F(\vec{\sigma}) = \sum_{i,j} J_{ij} \sigma_i \sigma_j \to \max, \quad \sigma_i = \{\pm 1\} \forall i \\
J_{ij} = J_{ji}, \quad J_{ii} \text{ does not matter.}
\end{cases}
\] (1)

The configuration vector \(\vec{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n)\) providing the solution of the problem is called the ground state.
The Hebbian-like representation exists for every \(J\),
\[
J = S^T \cdot S,
\]
where \(S\) is a real \((p \times n)\)-matrix and \(p = \text{rank } J\). (2)

We would like to investigate the special case of the \((p \times n)\)-matrix \(S\):
\[
S = \begin{pmatrix}
1 - x & 1 & \ldots & 1 & 1 & \ldots & 1 \\
1 & 1 - x & \ldots & 1 & 1 & \ldots & 1 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
1 & 1 & \ldots & 1 - x & 1 & \ldots & 1
\end{pmatrix}, \quad \begin{cases}
x \text{ is real} \\
p + q = n.
\end{cases}
\] (3)

The rows of the matrix \(S\) are the generalized memorized patterns.
The meaningful interpretation of the problem:

the network had to be learned by \( p \)-time showing of the standard \( \vec{\varepsilon}(n) \),

\[
\vec{\varepsilon}(n) = (1, 1, \ldots, 1) \in \mathbb{R}^n,
\]

but an error crept into the learning process and the network was learned by \( p \) distorted copies \( \vec{s}^{(l)} \) of the standard:

\[
\vec{s}^{(l)} = (1, \ldots, 1, 1 - x, 1, \ldots, 1), \quad l = 1, 2, \ldots, p. \tag{4}
\]

The real number \( x \) is called the distortion parameter.

The problem under investigation, Eqs.(1) -(3), is very close to the problem of generalization in the case of one embedded pattern.

**Main Results**

1°. The local maxima of the functional \( F(\vec{\sigma}) \) necessarily have the form

\[
\vec{\sigma}^* = (\sigma_1, \sigma_2, \ldots, \sigma_p, 1, \ldots, 1), \tag{5}
\]

and

\[
F(\vec{\sigma}^*) \propto x^2 - 2x(q + p \cos w) \cos w + (q + p \cos w)^2, \tag{6}
\]

where \( w \) is the angle between vectors \( \vec{\sigma}' \) and \( \vec{\varepsilon}(p) = (1, 1, \ldots, 1) \in \mathbb{R}^p : 

\[
\cos w = \frac{\sum_{i=1}^{p} \sigma_i}{p} = \frac{(\vec{\sigma}', \vec{\varepsilon}(p))}{\| \vec{\sigma}' \| \cdot \| \vec{\varepsilon}(p) \|}. \tag{7}
\]

Then, the vectors \( \vec{\sigma}^* \) (5) with the \( p \)-dimensional parts \( \vec{\sigma}' \) equidistant from \( \vec{\varepsilon}(p) \), provide the same value of the functional \( F \).

2°. Evidently,

\[
\cos w \equiv \cos w_k = 1 - \frac{2k}{p}, \quad k = 0, 1, \ldots, p,
\]
and the vectors $\vec{\sigma}^\ast$ (5) are grouped into $p + 1$ classes $\Sigma_k$ on which the functional $F(\vec{\sigma}^\ast)$ is constant:

$$ \Sigma_k = \{ \vec{\sigma}^\ast \mid \text{exactly } k \text{ coordinates of } \vec{\sigma}^\ast \text{ are equal to ”–1”} \}. $$

The number of the vectors $\vec{\sigma}^\ast$ in the class $\Sigma_k$ is equal $\binom{p}{k}$.

3°. To find the ground state dependence on $x$, it is necessary to analyze the family of the straight lines

$$ L_k(x) = (q + p \cos w_k)^2 - 2x(q + p \cos w_k) \cos w_k, \ k = 0, 1, \ldots, p. \quad (8) $$

In the region where $L_k(x)$ majorizes all the other straight lines, the ground state belongs to the class $\Sigma_k$ and it is $\binom{p}{k}$ times degenerated.

4°. Theorem.

*When* $x$ *increases from* $-\infty$ *to* $\infty$, *the ground state in consecutive order belongs to the classes* $\Sigma_0, \Sigma_1, \ldots, \Sigma_{k_{\max}}$.

The $k$th rebuilding of the ground state ($\Sigma_{k-1} \rightarrow \Sigma_k$) occurs at the point $x_k$ of the intersection of the straight lines $L_{k-1}(x)$ and $L_k(x)$:

$$ x_k = p \frac{n - (2k - 1)}{n + p - 2(2k - 1)}, \ k = 1, 2, \ldots, k_{\max}, \quad (9) $$

where

$$ k_{\max} = \min \left( p, \left\lceil \frac{n + p + 2}{4} \right\rceil \right). $$

The functional has no other local maxima.

The Theorem relates the quality of the learning of the network with the value of the distortion $x$ during the learning stage and with the length $p$ of the learning sequence. It is reasonable, that the error of the network increases with the increase of the distortion $x$: when $x \in (x_k, x_{k+1})$ the class $\Sigma_k$ (”the truth” understood by the network) differs from the standard $\vec{\varepsilon}(n)$ by $k$ coordinates (others interpretations see below).
In the Fig.1 is given the typical behavior of the straight lines $L_k(x)$. The rebuildings of the ground state occurs at the points $x_k$ of the intersection of the straight lines $L_{k-1}$ and $L_k$. Inside the interval $(x_k, x_{k+1})$ the ground state belongs to the class $\Sigma_k$. When $x$ increases:

a). all the rebuildings of the ground state occur: $k_{max} = p$;  
b). only $k_{max} = \left\lceil \frac{n+p+2}{4} \right\rceil < p$ rebuildings of the ground state occur.

**Generalizations of Basic Model**

1°. When the standard $\vec{\varepsilon}(n)$ is changed by an *arbitrary* configuration vector $\vec{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_n)$, $\alpha_i = \{\pm 1\}$, the Theorem remains valid, but the vectors $\vec{\sigma}^*$ (5) have the form

$$\vec{\sigma}^* = (\alpha_1 \sigma_1, \alpha_2 \sigma_2, \ldots, \alpha_p \sigma_p, \alpha_{p+1}, \ldots, \alpha_n).$$

2°. When we rotate the memorized patterns (4) as a whole, all their first $p$ coordinates are distorted.

Suppose the rotation matrix $\mathbf{U} = (u_{ij})$ transforms the first $p$ coordinates
of $n$-dimensional vectors only:

$$
\begin{align*}
\vec{u} &= \mathbf{U} \cdot \vec{\varepsilon}(p) = (u_1, u_2, \ldots, u_p), \\
\vec{u}_l &= \sum_{i=1}^{p} u_{li}, \quad l = 1, 2, \ldots, p; \quad \| \vec{u} \|_2^2 = p.
\end{align*}
$$

Then the memorized patterns take the form:

$$
\vec{s}^{(l)} = (u_1 - xu_{1l}, u_2 - xu_{2l}, \ldots, u_p - xu_{pl}, 1, \ldots, 1), \quad l = 1, 2, \ldots, p.
$$

It is easy to see, that if the standard $\vec{\varepsilon}(n)$ does not change after the rotation ($u_l \equiv 1$) all the results of the "Basic Model" remain unchanged.

More interesting is the case when the standard $\vec{\varepsilon}(n)$ shifts after the rotation ($u_l \neq 1$). Again the only important configuration vectors are $\vec{\sigma}^*$ (5) and the functional $F(\vec{\sigma}^*)$ is given by the same expression (6). But now $w$ is the angle between vectors $\vec{\sigma}'$ and $\vec{u}$:

$$
\cos w = \frac{\sum_{i=1}^{p} \sigma_i \cdot u_i}{p} = \frac{(\vec{\sigma}', \vec{u})}{\| \vec{\sigma}' \| \cdot \| \vec{u} \|}.
$$

The vectors $\vec{\sigma}^*$ are grouped in the classes $\Sigma_k^{(U)}$ on which $F(\vec{\sigma}^*)$ is constant:

$$
\Sigma_k^{(U)} = \{ \vec{\sigma}^* | \text{ with } p\text{-dimensional parts } \vec{\sigma}' \text{ equidistant from } \vec{u} \}.
$$

The number of the different classes $\Sigma_k^{(U)}$ is given by the number $t$ of different values of $\cos w$ (11):

$$
\cos w_0 > \cos w_1 > \ldots > \cos w_t; \quad \cos w_k = -\cos w_{t-k}, \quad \forall k \leq t.
$$

Then we have the following generalization of the Theorem:

when $x$ increases from $-\infty$ to $\infty$, the ground state in consecutive order belongs to the classes

$$
\Sigma_0^{(U)}, \Sigma_1^{(U)}, \ldots, \Sigma_{k_{\text{max}}}^{(U)}.
$$

The $k$th rebuilding of the ground state ($\Sigma_{k-1}^{(U)} \to \Sigma_k^{(U)}$) occurs at the point $x_k$ of the intersection of the straight lines $L_{k-1}(x)$ and $L_k(x)$:

$$
x_k = \frac{p}{2} \left[ 1 + \frac{q}{q + p(\cos w_{k-1} + \cos w_k)} \right], \quad k = 1, 2, \ldots, k_{\text{max}}.
$$
If $x_1 > \frac{3}{4}p$, all the rebuildings take place ($k_{max} = t$). If $x_1 < \frac{3}{4}p$, the rebuildings stop when the denominator in Eq.(12) becomes negative.

**Note.** The compositions of the classes $\Sigma_k^{(U)}$ are determined by the values of $\{u_i\}_1^p$ only. But the choice of $\{u_i\}$ is completely in our hands! Then we can create the Hopfield type network with a preassigned set of the fixed points.

3°. The memorized patterns can be obtained from $\vec{\varepsilon}(n)$ by the identical synchronous distortions of its $m$ coordinates. Suppose $n = p \times m + q$ and the matrix $S$ is

$$S = \begin{pmatrix} 1-x & \ldots & 1-x & 1 & \ldots & 1 & \ldots & 1 & \ldots & 1 \end{pmatrix}$$

$$\vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots \; \vdots$$

$$1 & \ldots & 1 & \ldots & 1 & \ldots & 1 - x & \ldots & 1 - x & 1 \ldots 1 \end{pmatrix}$$

The "suspicious" configuration vectors are the piecewise constant vectors

$$\vec{\sigma}^* = (\sigma_1, \ldots, \sigma_1, \sigma_2, \ldots, \sigma_2, \ldots, \sigma_p, \ldots, \sigma_p, 1, \ldots, 1); \quad (13)$$

the values of the functional are

$$F(\vec{\sigma}^*) \propto x^2 - 2x \cos w \left( \frac{q}{m} + p \cos w \right) + \left( \frac{q}{m} + p \cos w \right)^2,$$

where, as in Eq.(7), $w$ is the angle between $\vec{\sigma}' = (\sigma_1, \sigma_2, \ldots, \sigma_p)$ and $\vec{\varepsilon}(p)$.

Again the vectors (13) are grouped into classes $\Sigma_k^{(m)}$, whose structure is similar to the structure of the classes $\Sigma_k$. Then we have the generalization of the Theorem: the value of the parameter $x$, which corresponds to the $k$th rebuilding of the ground state ($\Sigma_{k-1}^{(m)} \rightarrow \Sigma_k^{(m)}$), is

$$x_k = p \frac{n}{m} - (2k - 1)$$

and

$$k_{max} = \min \left( p, \left[ \frac{n}{m} + p + 2 \right] \right).$$
1°. \( x_1 = p \frac{n-1}{n+p-2} \geq \frac{p}{2}. \)

In fact \( x_1 \) is the boundary of the distortions up to which the network reproduces the standard from its distorted copies correctly. The boundary depends on the length \( p \) of the learning sequence. Of course, the network is learned correctly, if the value of the distortions does not exceed \( \frac{p}{2} \).

2°. \( x_1 \) is monotonically increasing function of \( p \) and \( n \).

*Let \( n \) and \( x \) be fixed.* Merely due to an increase of \( p \) the boundary \( x_1 \) can be forced to exceed \( x \) (if \( x \) is not too large). As a result \( x \) turns out to be on the left of a new position of \( x_1 \), i.e. in the region where the only fixed point is the standard \( \vec{\epsilon}(n) \). In other words, only by an increase of the length \( p \) of the learning sequence we can force the network to understand correctly ”the truth” it is tried to be learned. It is in agreement with the practical experience: the greater the length of the learning sequence, the better the signal can be read through noise.

*Let \( p \) and \( x \) be fixed.* As above, merely due to an increase of the number \( n \) the value of \( x_1 \) can be forced to exceed \( x \). This result is reasonable too: if \( p \) is fixed, the greater is \( n \), the smaller is the relative weight \( \frac{p}{n} \) of the distorted coordinates. Naturally, the less is the relative distortion, the better must be the result of the learning.
Here we change the notation for the standard $\vec{e}(n)$ and introduce another standard $\vec{e}^\prime(n)$:

\begin{align*}
\vec{e}(+) &= (1, 1, \ldots, 1, 1, \ldots, 1) = \vec{e}(n) \\
\vec{e}(\prime) &= (-1, -1, \ldots, -1, 1, \ldots, 1).
\end{align*}

In their not coincident parts the standards $\vec{e}(+)$ and $\vec{e}(\prime)$ are opposed with each other, i.e. they are two opposite ”statements”. Any of the network fixed points $\vec{\sigma}^\prime$ (5) is an intermediate statement between $\vec{e}(+)$ and $\vec{e}(\prime)$, which is drawn towards either one edge of the scale, or the other. And the network ”feels” this.

Indeed, when the distortion $x$ is not very large ($x < p$), the number $k$ of the ground state does not exceed $\frac{p}{2}$, and the ground state more resembles $\vec{e}(+)$ than $\vec{e}(\prime)$. In other words, the memorized patterns are interpreted by the network as the distorted copies of the standard $\vec{e}(+)$. But if during the learning stage the distortion exceeds $p$ ($x > p$), the number of ground state exceeds $\frac{p}{2}$ and the ground state resembles $\vec{e}(\prime)$. Now the network interprets the memorized patterns as the distorted copies of another standard $\vec{e}(\prime)$.

This is in agreement with the practical experience: we interpret deviations in the image of a standard as permissible only up to some boundary. If only this boundary is exceeded, the patterns are interpreted as the distortions of quite different standard. For the network of the considered type this boundary is $p$.

One extra argument to support this interpretation: from Eq.(9) it is easy to see that when $p = const$ and $n \to \infty$ all $x_k$ stick to one point

\[ x_k \equiv p, \quad k = 1, 2, 3, \ldots, p; \]

then for $x < p$ the ground state belongs to the class $\Sigma_0 = \vec{e}(+)$, whereas for $x > p$ the ground state belongs to the class $\Sigma_p = \vec{e}(\prime)$. 

\[ 3^\circ \quad x_{\frac{p+1}{2}} = p. \] (Without the loss of generality we assume that $p$ is odd.)
4°. \( k_{max} = \begin{cases} p, & \text{when } \frac{p-1}{n-1} < \frac{1}{3} \\ \left[\frac{n+p+2}{4}\right], & \text{when } \frac{p-1}{n-1} > \frac{1}{3} \end{cases} \)

So, \( p \) is the boundary for the permissible distortions \( x \). The question is, what do the memorized patterns with large distortions \( x \) mean? We treat the increase of \( x \) above \( p \) as the more and more negation of the standard \( \vec{\varepsilon}(+) \). As if the network is learned by the memorized patterns, which deny the standard \( \vec{\varepsilon}(+) \). In other words, the network is relearned by presentation of negative examples.

There is big and clear to everybody difference between the relearning with the help of negative examples and the learning of the opposite truth. The relearning is characterized by some specific difficulties: (1) the better the incorrect truth has been understood, the more difficult (and sometimes even impossible) to correct it; (2) it is comparatively easy to correct the result slightly, but it is much more difficult to revise it in the main, etc. We think, that the dependence of \( k_{max} \) on \( p \) is the reflection of just these problems.

When the number \( p \) of the parameters which have to be corrected is not very great (\( \frac{p-1}{n-1} < \frac{1}{3} \)), the network can be relearned by simple presentation of negative examples. In this case \( k_{max} = p \) and, when ”the denial” of the standard \( \vec{\varepsilon}(+) \) is rather strong (\( x > x_p \)), as ”a new” truth the network understands the opposite standard \( \vec{\varepsilon}(-) \). But if the number of the corrected parameters is great (\( \frac{p-1}{n-1} > \frac{1}{3} \)), to relearn the network it is not sufficient to present the negative examples. In this case \( \left[\frac{p+1}{2}\right] < k_{max} < p \) and whatever large \( x \) is (\( x_{k_{max}} < x < \infty \)), as a new truth the network understands not the opposite standard \( \vec{\varepsilon}(-) \), but one of the statements intermediate between \( \vec{\varepsilon}(+) \) and \( \vec{\varepsilon}(-) \). Though the understood truth is drawn towards \( \vec{\varepsilon}(-) \), since \( k_{max} > \frac{p}{2} \).

Of course, our interpretation is open for discussion. But it seems that in real life there are a lot of examples, which confirm our conception.
New Results (in preparation)

The generalization of the Basic Model to the cases:

1°. The functional $F(\vec{\sigma})$ in the problem (1)-(3) has the form

$$F(\vec{\sigma}) = \sum_{i,j=1}^{n} J_{ij}\sigma_i\sigma_j + h \sum_{i=1}^{n} \sigma_i.$$ 

In physics such a linear term describes the magnetic field.

In the Fig.2 the straight lines $h_k(x)$ divide the plane $(x, h)$ into the regions where the ground state belongs to the different classes $\Sigma_k,

$$h_{k-1}(x) = 2p(n - 2k + 1)\left(\frac{x}{x_k} - 1\right), \quad k = 1, 2, \ldots$$

2°. The distortions $x_l$ are different for all the memorized patterns (4):

$$\vec{s}^{(l)} = (1, \ldots, 1, 1 - x_l, 1, \ldots, 1), \quad l = 1, 2, \ldots, p.$$

Suppose

$$x_1 \geq x_2 \geq \ldots, \geq x_p > 0.$$
It can be shown that, firstly, only one of the configuration vectors
\[ \tilde{\sigma}^*(k) = (-1, -1, \ldots, -1, 1, \ldots, 1), \quad k = 0, 1, \ldots, k_{max}, \]
can be the ground state. Here
\[ k_{max} \begin{cases} 
= p, & \text{when } p < \frac{n}{2} \\
\leq \left\lfloor \frac{n}{2} \right\rfloor, & \text{when } p > \frac{n}{2}.
\end{cases} \]
And secondly, in order to the vector \( \tilde{\sigma}^*(k) \) be a ground state, the fulfillment of the inequalities
\[ x_{k+1} - \frac{\sum_{i=1}^{k} x_i - p \sum_{j=k+2}^{p} x_j}{n - 2k - 1} \leq p \leq x_k - \frac{\sum_{i=1}^{k-1} x_i - p \sum_{j=k+1}^{p} x_j}{n - 2k + 1} \]
is necessary and sufficient conditions.

When \( p \) is fixed and \( n \to \infty \) these inequalities are much more simpler:
\[ x_{k+1} \leq p \leq x_k. \]

Note, that for sufficiently large \( n \) and \( p > x_1 \), the standard \( \tilde{\epsilon}(n) = \tilde{\sigma}^*(0) \) is the ground state. It seems, that for the Hebb connection matrix the last result clarifies the meaning of the well-known Latin saying "Repetitio est mater studiorum" – showing the same pattern many times (inevitably each time with a distortion), we seek to make the number of the presentations \( p \) greater than the maximal distortion.

1 Litinskii L.B. 1997 Energy functional and fixed points of neural network. cond-mat/9706280. Also in: M.Marinaro and R.Tagliaferri (Eds.) 1998 ”Neural Nets. WIRN VIETRI-97. Proceedings...” (Berlin: Springer-Verlag) p.338.
2 Fontanari J.F. 1990 J. Phys. France v.51 p.2421.
3 Krebs P.R., Theuman W.K. 1993 J. Phys. A26 p.3983.
4 Litinskii L.B. 1999 Theor. and Math. Phys. v.118 p.133. cond-mat/9906197.
5 Litinskii L.B. 1994 Theor. and Math. Phys. v.101 p.1492.