

LETTER TO THE EDITOR

A memory which forgets

Giorgio Parisi

Dipartimento di Fisica, II Università di Roma 'Tor Vergata', Via Orazio Raimondo, Roma 00173, Italy and INFN, sezione di Roma, Italy

Received 7 March 1986

Abstract. The model of Hopfield for a neural network with associative memory is modified by the introduction of a maximum value for the synaptic strength; in this way old patterns are automatically forgotten and the memory recalls only the most recent ones. If the parameters are correctly chosen, the memory never goes into the state of total confusion characteristic of the Hopfield model.

In recent years the mechanism for which neural networks behave as associative memories (more precisely as content-addressable memories) has been extensively studied. It seems that considerable progress in this field was made by the introduction of very stylised models which are far from realistic; the advantage of these models is the possibility of performing simple computer simulations and of doing analytic studies; in this way we hope to clarify the basic issues of the theory of networks with associative memory.

A very interesting model has been proposed by Hopfield [1]: each neuron may stay in two states, firing or quiescent (the i th neuron is represented by a spin variable σ_i which may take the values ± 1); the synaptic strength is assumed to be symmetric, i.e. the influence ($J_{i,k}$) of the i th neuron on the k th neuron is the same when i and k are exchanged ($J_{i,k} = J_{k,i}$), and the input patterns are stored using the generalised Hebb rule for modifying the synaptic strength. An 'energy' function $E[\sigma]$ can be associated with each configuration $\{\sigma\}$ of the network and the time evolution of the neural network is such that the asymptotic stable states at large times are the minima of $E[\sigma]$ with respect to $\{\sigma\}$.

For simplicity let us say that the network remembers a given input pattern $\{\sigma\}$ if the asymptotic state is $\{\sigma\}$ or very near to $\{\sigma\}$ when the initial state of the network is equal to $\{\sigma\}$ (different and more restrictive definitions can be used); in other words $E[\sigma]$ must have a minimum near each of the input patterns which are remembered.

The Hopfield model is also very interesting because it has many points in common with spin glasses and a very sophisticated and rich theory has been recently constructed for spin glasses [2].

Under the strong assumption that the input patterns are uncorrelated, both numerical simulation [1, 3] and analytic computations [4] show that the storage capacity of such a network is proportional to N . If the number M of input patterns becomes larger than a critical value M_c ($M_c \propto 0.14N$) the network goes into a state of total confusion and a negligible amount of patterns are remembered; in contrast, if M is smaller than M_c , practically all input patterns are remembered.

Any memory which has been well designed should not go into a state of total confusion when overloaded: the most welcome reaction should be that old inputs are forgotten in order to leave room for new inputs. It has recently been suggested that the state of total confusion may be avoided if the variation in the synaptic strength induced by the storing of the M th pattern increases with M as $\exp(\alpha M)$ [5]: the synaptic strengths essentially depend only on the last input patterns and the old patterns are automatically forgotten; by choosing an appropriate value of α the state of total confusion is avoided and the memory still keeps the capability of storing a number of patterns proportional to N .

This mechanism is very promising; however it is certainly interesting to explore other possibilities. In this letter I would like to suggest another mechanism which allows old patterns to be forgotten. The idea is very simple: while in the original Hopfield model the synaptic strength could take arbitrarily large (positive or negative) values, here I propose (less unrealistically) that the synaptic strength is bounded, i.e.

$$|J_{i,k}| < A. \quad (1)$$

The generalised Hebb rule still holds with the exception that if a modification of the synaptic strength $J_{i,k}$ should violate the bound (1), this modification is not operative. In this way, when we store a given pattern, the non-linear constraint (1) is such that the information about old patterns gradually deteriorates and is finally lost at the end. By choosing the value of A with care we can avoid the state of total confusion and still keep a storing capability proportional to N .

Let us examine the details of the model before showing the computer simulations.

The energy function is (as usual)

$$E[\sigma] = -\frac{1}{2} \sum_{i \neq k} J_{i,k} \sigma_i \sigma_k. \quad (2)$$

The new synaptic strengths ($J_{i,k}^{\text{new}}$) after having stored a pattern $\{p\}$ (the p_i also may have only the values ± 1) are given by

$$J_{i,k}^{\text{new}} = f(J_{i,k}^{\text{old}} + C p_i p_k) \quad (3)$$

where $J_{i,k}^{\text{old}}$ is the old value of the synaptic strength, C is a normalisation constant (which for convention we take equal to $N^{-1/2}$) and $f(x)$ is a function which characterises the model. If

$$f(x) = x \quad (4)$$

(3) is the generalised Hebb rule. The model proposed in this letter corresponds to the choice

$$\begin{aligned} f(x) &= -A && \text{for } x < -A \\ f(x) &= x && \text{for } -A < x < A \\ f(x) &= A && \text{for } A < x. \end{aligned} \quad (5)$$

It is clear that any other non-linear function $f(x)$ with saturation would suffice. Equation (5) is retained for its simplicity.

We know that in the Hopfield model ($A \rightarrow \infty$) the state of total confusion is reached after $M = \alpha N$ independent patterns have been stored, with $\alpha \approx 0.14$; the distribution of a given synaptic strength $J_{i,k}$ is a Gaussian with variance $\alpha^{1/2}$. In order to avoid the state of total confusion the value of A should be such that the distribution of the J is seriously modified for $\alpha = 0.14$, so A cannot be much larger than $(0.14)^{1/2} \approx 0.4$.

In order to test the validity of this model I have performed numerical simulations at different values of N ($N = 100, 200$ and 400) and A (for a total amount of about 100 h of CPU on a VAX 750). A large number of independent patterns (greater than N) have been stored and I have investigated the number of patterns which are remembered by the memory (by definition the pattern is remembered if no more than 2% of the neurons in the final configuration differ from the stored pattern; other reasonable definitions lead to quite similar results).

For A greater than a critical value ($A \approx 0.7$) practically no patterns are remembered while for smaller values of A the last patterns are well memorised. In figure 1 we plot some estimates of the storage capability of the memory (i.e. the average number of stored patterns which are remembered) as a function of A : this quantity has a maximum at an intermediate value of A (≈ 0.35). The existence of a maximum is not surprising: in the limit $A \rightarrow 0$ only the information concerning the very last patterns is not deteriorated and for large values of A we stay in the state of total confusion. It is also apparent that the storage capacity is proportional to N at fixed A : the maximum (at $A \approx 0.35$) is about $0.05N$.

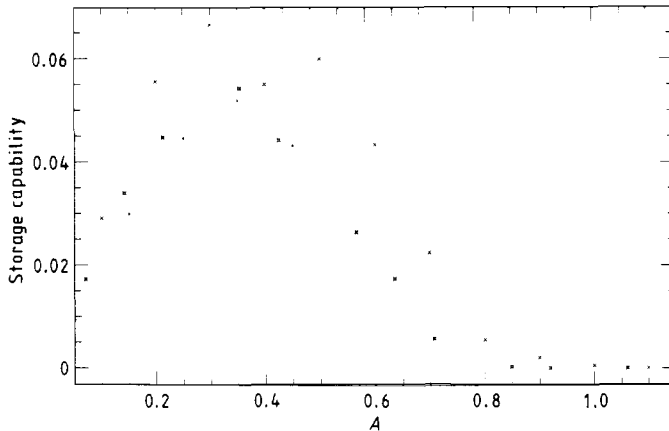


Figure 1. The storage capability as a function of A for $N = 100$ (\times), $N = 200$ (\times) and $N = 400$ (\cdot).

In figure 2 I show the probability of remembering the $(M - k)$ th pattern after M patterns have been stored for $A \approx 0.35$ as a function of $x = k/N$. It is evident that only the most recent states are remembered with high probability and that such a probability decreases with x ; it seems also that this retrieval probability has a finite limit when $N \rightarrow \infty$ at fixed x . For N in the range 200–400, only if x is smaller than $\approx 0.04N$ are we confident that the input patterns have been memorised (i.e. the retrieval probability is higher than 90%). In the range of N studied in this letter, about 70% of the total storage capacity is in this safe region ($x < 0.04N$): more detailed investigations are needed to establish if in the limit $N \rightarrow \infty$ all the storage capacity moves into the safe region and the curve plotted in figure 2 becomes a theta function (i.e. 1 for $x < x_c$ and 0 for $x > x_c$).

The conclusion is that the introduction of bounds on the synaptic strengths and consequently of non-linearities in the generalised Hebb rule may modify the memory model of Hopfield in such a way that the state of total confusion is avoided and the

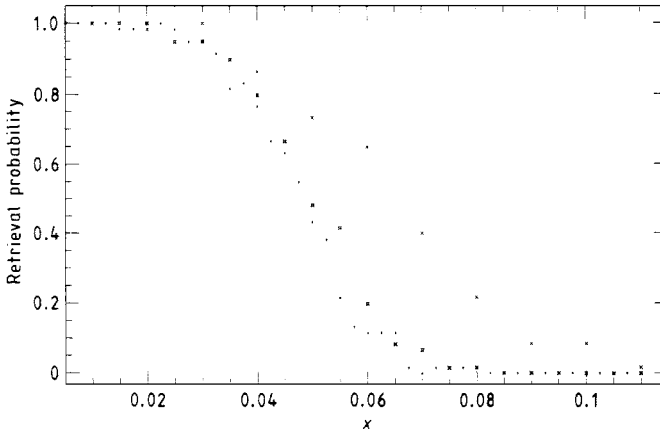


Figure 2. The retrieval probability of a given pattern as a function of $x = k/N$ (for $N = 100$ (\times), $N = 200$ (π) and $N = 400$ (\cdot) at $A \approx 0.35$) after k patterns have been subsequently stored in the network (i.e. the retrieval probability of the j th pattern after $j+k$ patterns are stored).

memory remembers only the last inputs (with very high probability). The safe storage capability of such a neural network is obviously smaller than the original model and it is about $0.04N$.

It is a pleasure for me to thank M Virasoro and M Mezard for many long discussions on memory organisation and neural networks. I am grateful to G Toulouse for having communicated the results of [5] prior to publication and for a discussion on the model presented in this letter.

References

- [1] Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554
- [2] Parisi G 1984 *Mathematical Physics VII* ed W E Brittin, K E Gustafson and W Wyss (Amsterdam: North-Holland) p 337
Mezard M, Parisi G and Virasoro M 1986 *The Spin Glass Theory and Beyond* (Singapore: World Scientific)
- [3] Kinzel W Z 1985 *Z. Phys. B* **60** 205
- [4] Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. Lett.* **55** 1930
- [5] Changeaux J P, Dehare S, Nadal J P and Toulouse G 1986 *Europhys. Lett.* in press