EUROPHYSICS LETTERS 15 May 1988

Europhys. Lett., 6 (2), pp. 101-105 (1988)

The Enhanced Storage Capacity in Neural Networks with Low Activity Level.

M. V. TSODYKS (*) and M. V. FEIGEL'MAN (**)

(*) *Institute of Higher Nervous Activity and Neurophysiology USSR Academy of Sciences, Moscow*
(**) *L. D. Landau Institute for Theoretical Physics USSR Academy of Sciences, Moscow*

(received 18 November 1987; accepted in final form 11 March 1988)

PACS. 05.20 – Statistical mechanics.
PACS. 05.90 – Other topics in statistical physics and thermodynamics.
PACS. 87.30 – Biophysics of neurophysiological processes.

Abstract. – The modified Hopfield model defined in terms of «V-variables» ($V = 0; 1$), which is appropriate for storage of correlated patterns, is considered. The learning algorithm is proposed to enhance significantly the storage capacity in comparison with previous estimates. At low levels of neural activity, $p \ll 1$, we obtain $\alpha_c(p) \sim (p|\ln p|)^{-1}$ which resembles Gardner's estimate for the maximum storage capacity.

1. – In the last few years a large number of analytical and numerical results concerning the Hopfield model of a neural network has been obtained. In this model neurons are treated as two-state variables S_i taking the values (± 1), the stored patterns constitute the set $\{\xi_i^\mu\}$ ($\xi_i^\mu = \pm 1$, $1 \leq i \leq N$, $1 \leq \mu \leq L$), where the subscript i numbers different neurons, whereas μ numbers different patterns. The synaptic matrix should be chosen in such a way that the stored patterns would be stationary points of neural dynamics. In the Hopfield scheme [1] it was proposed the following form of this matrix:

$$J_{ij}^{(0)} = \frac{1}{N} \sum_\mu \xi_i^\mu \xi_j^\mu . \tag{1}$$

One of the most important problems arising here is that of the maximum storage capacity. The statistical physics approach to this problem has been developed by Amit *et al.* [2, 3]. It has been shown that the network with the Hopfield learning rule possesses an associative memory if a number of stored patterns does not exceed $\alpha_c(T)N$, where T is the «temperature» which mimicks the intensity of noise in the neural dynamics. $\alpha_c(T)$ is a monotonically decreasing function on the interval $0 \leq T \leq 1$, $\alpha_c(0) \approx 0.14$, $\alpha_c(1) = 0$. This model works successfully if the stored patterns are uncorrelated so that $\text{Prob}(\xi_i^\mu = \pm 1) = 1/2$. Neurophysiological data indicate, however, that the mean level of neural activity is much lower than 1/2. In order to store biased patterns with $\overline{\xi_i^\mu} = a \neq 0$ one should modify the model. Two different but similar approaches to this problem have been suggested in papers [4, 5] both leading to local algorithms with the maximum storage capacity similar to that of the original Hopfield model.

It is also worth mentioning the method of pseudoinverse matrix [6, 7] which allows to store N linear independent patterns without errors, no matter correlated or not. However, all these results are far from the maximal storage capacity which may be achieved by an appropriate choice of the connection matrix. In [8] Gardner has managed to obtain formulae for the maximal capacity without proposing an explicit learning algorithm. In particular, it is possible to store $\alpha_c(a)N$ biased patterns and $\alpha_c(a) \sim [(1-a)\ln(1-a)]^{-1}$ when $a \to 1$ (*i.e.* mean level of activity tends to zero). In this letter we present an explicit learning rule which is in agreement with this prediction. Another explicit model with a similar enhancement in storage capacity was studied in [9].

2. – To begin with, briefly recall the usual scheme of the Hopfield model of storage and retrieval. Each neuron has its presynaptic potential due to the activity of all other neurons

$$h_i = \sum_{j \neq i} J_{ij} \frac{S_j + 1}{2}. \tag{2}$$

The dynamics of the overall network is governed by the equation

$$S_i(t+1) = \operatorname{sign}(h_i(t) - \theta_i), \tag{3}$$

where θ_i is the neural threshold. If we believe that $\sum_{j \neq i} J_{ij} = \theta_i$, then the stationary configurations of eq. (3) are the minima of the Hamiltonian

$$H(S) = -\frac{1}{2} \sum_{ij} J_{ij} S_i S_j. \tag{4}$$

As has been shown in [4, 5] in order to store biased patterns one must modify the Hopfield's form of the matrix (1)

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{L} (\xi_i^\mu - a)(\xi_j^\mu - a), \tag{5}$$

where every component ξ_i^μ in a learnt pattern can be chosen independently with probability

$$p(\xi) = \frac{1}{2}(1+a)\delta(\xi - 1) + \frac{1}{2}\delta(\xi + 1).$$

The main idea of our approach is the following: we take the connection matrix in the same form as in (5) but introduce new variables: $V = (S+1)/2$, $\eta = (\xi + 1)/2$ so that V and η take the values 0, 1. The probability distributions of η and of the matrix J_{ij} become

$$p(\eta) = p\delta(\eta - 1) + (1-p)\delta(\eta), \tag{6}$$

$$J_{ij} = \frac{1}{N} \sum_{\mu} (\eta_i^\mu - p)(\eta_j^\mu - p), \tag{7}$$

and we shall be particularly interested in the case $p \ll 1$. The value of the threshold θ will be chosen so that to optimize the storage capacity. In terms of the previous model this change of variables means adding the threshold $\tilde{\theta}_i = \sum_{j \neq i} J_{ij}$, which correlates with the stored patterns. It turns out that this correlation significantly enhances the performance of the network. To make sure of it, we start with the estimation of the storage capacity; the estimation being based on the requirement of the small «noise-to-signal» ratio in the stored patterns, ensuring their stability. The local field at neuron i in the pattern η^1 is

$$h_i = \sum_{j \neq i} J_{ij} V_j - \theta = \frac{1}{N} \sum_{j \neq i} \sum_{\mu=1}^{L} \tilde{\eta}_i^\mu \tilde{\eta}_j^\mu \eta_j^1 - \theta = p(\eta_i^1 - p) + \frac{1}{N} \sum_{j \neq i} \sum_{\mu > 1} \tilde{\eta}_i^\mu \tilde{\eta}_j^\mu \eta_j^1 - \theta, \tag{8}$$

where $\tilde{\eta}_i^\mu = \eta_i^\mu - p$. The second (noise) term has zero mean and its mean square equals

$$R^2 = \frac{(N-1)(L-1)}{N^2} p^3 \approx \alpha p^3, \qquad \alpha = \frac{L}{N}. \tag{9}$$

The minimal value of the first term in (8) is $-p^2$; we see that if $\theta = 0$, the original patterns are stable under the condition $\alpha \ll p$. Yet, if we take $\theta = \theta_0 p$, $\theta_0 \sim 1$, the stability condition drastically changes

$$\alpha \ll \min\left(\frac{\theta_0^2}{p}, \frac{(1-\theta_0)^2}{p}\right). \tag{10}$$

Now we pass over to the quantitative study of the model. Consider the Hamiltonian

$$H = -\frac{1}{2N} \sum_{ij} J_{ij} V_i V_j + \theta \sum_i V_i, \tag{11}$$

where J_{ij} has the form (7), V_i, $\eta_i^\mu = 0.1$; η_i^μ is an independent random variable with probability (5). The mean-field equations for this problem can be derived by analogy with [3]; so we write them down

$$m = \frac{1}{N} \sum_i \langle\!\langle \tilde{\eta}_i K(\beta H_i) \rangle\!\rangle, \tag{12a}$$

$$q_0 = \frac{1}{N} \sum_i \langle\!\langle K(\beta H_i) \rangle\!\rangle, \tag{12b}$$

$$q = \frac{1}{N} \sum_i \langle\!\langle K^2(\beta H_i) \rangle\!\rangle, \tag{12c}$$

$$r = \frac{q p^2}{[1 - p\beta(q_0 - q)]^2}. \tag{12d}$$

Here

$$K(x) = (1 + \exp[-x])^{-1}, \qquad H_i = \sqrt{\alpha r}\, z - \theta + \frac{\alpha p}{2} \frac{p\beta(q_0 - q)}{1 - p\beta(q_0 - q)} + m\tilde{\eta}_i.$$

The double angular brackets mean the Gaussian average over the variable z with zero mean and unit variance.

The meaning of the new order parameter q_0 is as follows:

$$q_0 = \frac{1}{N} \sum_i \langle V_i \rangle, \tag{13}$$

where the angular brackets denote the temperature average. The parameter $\tilde{m} = m/p$ measures the overlap between the retrieval state and the stored pattern, the «spin glass» parameters q and r have the same meaning as in [3]. At first we study the system (12) at zero temperature limit $\beta \to \infty$. In the limit $p \ll 1$ it takes a rather simple form

$$\tilde{m} = \operatorname{erf}\left(\frac{\theta - \tilde{m}p}{\sqrt{2\alpha r}}\right) - \operatorname{erf}\left(\frac{\theta}{\sqrt{2\alpha r}}\right), \tag{14}$$

$$r/p^2 = p \operatorname{erf}\left(\frac{\theta - \tilde{m}p}{\sqrt{2\alpha r}}\right) + \operatorname{erf}\left(\frac{\theta}{\sqrt{2\alpha r}}\right). \tag{15}$$

When writing down eqs. (14), (15) we approximate r by qp^2 which is valid at $T \to 0$, $p \ll 1$ in the entire region of the parameters α, θ where the nontrivial solution with $\tilde{m} \neq 0$ is existent. The analysis of these equations at $p \ll 1$ leads to the following result (the extended derivation will be given elsewhere):

$$\alpha_c(p) \simeq \frac{\theta_0^2}{2p\,|\ln p|}, \qquad \tilde{m}_c \approx 1 - \frac{\theta_0}{1-\theta_0}\left(\ln^{-1/2}\frac{1}{p}\right) p^{(\theta_0^{-1}-1)^2}, \tag{16}$$

where $\theta_0 = \theta/p$ is not very close to unity: $\varepsilon = 1 - \theta_0 \gg |\ln p|^{-1/2}$. At a further θ_0 increase the value of $\alpha_c(p)$ decreases:

$$\alpha_c(p) \simeq \frac{(1-\theta_0)^2}{p\,|\ln(1-\theta_0)|}. \tag{17}$$

Hence, the maximal storage capacity is achieved at

$$\alpha_c \simeq \frac{1}{2p\,|\ln p|}. \tag{18}$$

Note the remarkable similarity to the general bound obtained by Gardner [8]. We also find the value of the transition temperature $T_c(0)$ at $\alpha = 0$. At $\alpha \to 0$ the only equation needed is (12a) which reads

$$\tilde{m}_0 = K(\tilde{\beta}(\tilde{m} - \theta_0)) - K(-\tilde{\beta}\theta_0), \tag{19}$$

where $\tilde{\beta} = p\beta$, $\theta_0 = \theta/p$, $\tilde{m} = m/p$. From eq. (19) it is clear that the transition is a first-order transition and at $\varepsilon = 1 - \theta_0 \ll 1$ it occurs at the temperature

$$T_c(p) \approx p\,\frac{(1-\theta_0)}{|\ln(1-\theta_0)|}. \tag{20}$$

As is evident from eqs. (16) and (20), the optimal choice of θ_0 depends on the noise level presented in the network. The value of the overlap $\tilde{m}$ at the temperature $T = T_c$ is very close to unity

$$\tilde{m} = 1 - \frac{\varepsilon}{|\ln \varepsilon|}. \tag{21}$$

The stability analysis of the solution and the whole transition line $T_c(\alpha)$ will be given elsewhere.

It should be borne in mind that the «old» theories of the associative memory were formulated in terms of the «V-model» ($V = 0.1$), which seems to be most natural. Then, however, it was replaced by the «S-model» ($S = \pm 1$) without careful analysis of their equivalence. The results of our paper give rise to an amazing conclusion that in some cases such «obvious» simplification may drastically affect the performance of the neural networks.

* * *

One of our authors (MT) is grateful to A. A. FROLOV for helpful discussions of the work.

REFERENCES

[1] HOPFIELD J. J., *Proc. Nat. Acad. Sci. USA*, **79** (1982) 2554.
[2] AMIT D. J., GUTFREUND H. and SOMPOLINSKY H., *Phys. Rev. A*, **32** (1985) 1007.
[3] AMIT D. J., GUTFREUND H. and SOMPOLINSKY H., *Phys. Rev. Lett.*, **55** (1985) 1530; *Ann. Phys. (N.Y.)*, **173** (1987) 30.
[4] AMIT D. J., GUTFREUND H. and SOMPOLINSKY H., *Phys. Rev. A*, **35** (1987) 2283.
[5] FEIGEL'MAN M. V. and IOFFE L. B., *J. Mod. Phys. B*, **1** (1987) 51.
[6] PERSONNAZ L., GUYON I. and DREFUS G., *J. Phys. (Paris) Lett.*, **46** (1985) L-359.
[7] KANTOR I. and SOMPOLINSKY H., *Phys. Rev. A*, **35** (1987) 380.
[8] GARDNER E., Edinburgh preprints 395/87, 396/87.
[9] WILLSHAW D. J., BUNEMAN O. P. and LONGUET-HIGGINS H. C., *Nature (London)*, **222** (1969) 960; WILLSHAW D. J. and LONGUET-HIGGINS H. C., *Machine Intelligence* (Edinburgh University Press, Edinburgh) 1970, Chap. 5, p. 351.