Content-addressability and learning in neural networks

1988 J. Phys. A: Math. Gen. 21 245

(http://iopscience.iop.org/0305-4470/21/1/029)

J. Phys. A: Math. Gen. **21** (1988) 245-255. Printed in the UK

Content-addressability and learning in neural networks

B M Forrest

Department of Physics, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, UK

Received 1 July 1987

Abstract. The content-addressability of patterns stored in Ising-spin neural network models with symmetric interactions is studied. Numerical results from simulations on the ICL distributed array processor (DAP) involving systems with up to 2048 neurons are presented. Behaviour consistent with finite-size scaling, characteristic of a first-order phase transition, is shown to be exhibited by the basins of attraction of the stored patterns both in the case of the Hopfield model and for systems using a local iterative learning algorithm designed to optimise the basins of attraction. Estimates are obtained for the critical minimum overlaps which an input pattern must have with a stored pattern in order to successfully retrieve it.

1. Introduction

A content-addressable memory is one in which access to any of its stored entries is made without referring to the location of the entry; the entry (fact) may be retrieved by a cue which is a (sufficiently large) fraction of the stored fact itself. Such a memory will then be robust to noisy cues.

Models of McCulloch-Pitts (1943) neural networks have been proposed as distributed memories which exhibit such desirable content-addressable features (Hebb 1949, Little 1974). A lot of interest has centred around fully connected spin-glass models using specific prescriptions for the connection matrix (Hopfield 1982, Amit *et al* 1985a, b, 1987a, b, Mézard *et al* 1986, Gardner 1986, Bruce *et al* 1987) and with models employing iterative error-correcting 'learning' algorithms (Wallace 1986, Bruce *et al* 1986, Diederich and Opper 1987, Gardner 1988, Gardner *et al* 1987a, b, Krauth and Mézard 1987, Poeppel and Krey 1987).

In a network of N neurons, the state of each neuron at time t is modelled as an Ising spin $S_i(t)(i = 1, \ldots, N)$ with dynamics governed by

$$S_i(t+1) = \operatorname{sgn}\{h_i(t)\} \tag{1}$$

where the effective local field $h_i(t)$ on spin i due to the others is mediated through the interaction strengths (synaptic connections) T_{ij}:

$$h_i(t) = \sum_{j=1}^{N} T_{ij} S_j(t) \qquad T_{ii} = 0. \tag{2}$$

The stored entries of the memory are then the configurations (patterns) which are the fixed points of the dynamics (1) or, equivalently, those configurations in which every spin is aligned with its local field:

$$S_i h_i > 0 \qquad i = 1, \ldots, N. \tag{3}$$

0305-4470/88/010245+11$02.50

Such a network behaves as a distributed memory since it is the interaction strengths T_{ij} that determine the stable configurations of (1). So in order to store $p = N\alpha$ nominated patterns $S_i^{(r)}$, $i = 1, \ldots, N$; $r = 1, \ldots, p$, the T_{ij} will have to be prescribed appropriately. However, it is desirable not only to have the $\boldsymbol{S}^{(r)}$ being fixed points of (1), but also to have as large a region of attraction as possible around them. In this way the content-addressability (the maximum number of errors that could be tolerated in an input pattern while remaining in the domain of attraction of the stored pattern) would be maximised.

The Hebbian prescription for the connection strengths

$$T_{ij} = N^{-1} \sum_{r=1}^{p} S_i^{(r)} S_j^{(r)} \qquad T_{ii} = 0 \tag{4}$$

has been shown (Amit *et al* 1985a) to store up to $p \simeq 0.14N$ random uncorrelated patterns. The maximum storage capacity for such patterns is $2N$ patterns for general prescription of the T_{ij} (Cover 1965, Venkatesh 1986a, b) and can be increased if the patterns have correlations (Gardner 1987, 1988). However, no analytical results have as yet been obtained for the content-addressability of fully connected models. So the first question addressed here concerns not only the storage capacity of (4), but also the content-addressability of the stored patterns $\boldsymbol{S}^{(r)}$.

Numerical simulations were carried out on the ICL distributed array processor (DAP)—a 64×64 array of bit-processing elements—for networks of 512, 1024 and 2048 spins. The updating process (1) was executed sequentially in the spin sites $i = 1, \ldots, N$, as it was for all simulations in this paper. (The parallelism in the simulations was achieved in the matrix-vector multiplies (2). The connections were not stored in these simulations, but rather the row of connections corresponding to the spin being updated was recalculated each time as required. More than 3×10^8 conditional adds per second were achieved for $N = 2048$, $p/N = 0.10$, while the number of single-spin updates per second was over 1700 for $N = 512$.)

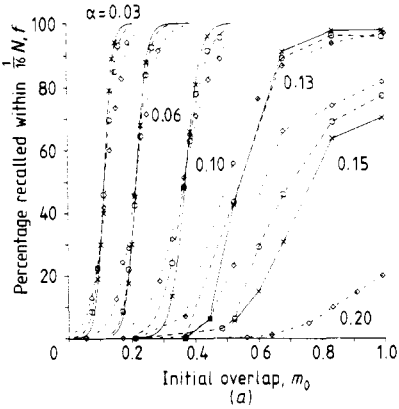
To determine the typical sizes of the basins of attraction (and hence the content-addressability) of the nominated configurations $\boldsymbol{S}^{(r)}$ in (4), states $\boldsymbol{S}^{(r,s)}$, $s = 1, \ldots, n_{m_0}$, were constructed having a given initial overlap m_0 with each pattern:

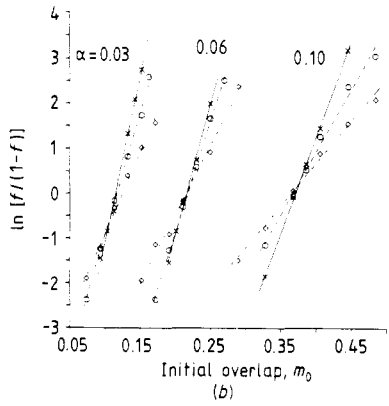
$$m_0 = N^{-1} \sum_{i=1}^{N} S_i^{(r)} S_i^{(r,s)} \tag{5}$$

and iterated to stability under (1) and (2) (where the $\frac{1}{2}(1 + m_0)$ sites at which $S_i^{(r)} = S_i^{(r,s)}$ were random). The nominal pattern $\boldsymbol{S}^{(r)}$ was deemed to have been successfully recalled if the resultant stable state differed in no more than $\frac{1}{16} N$ spin sites. This margin of recall error allows for the possibility of ferromagnetic stable states (i.e. those which are highly correlated with the nominal states) being stored instead of the nominal states themselves. (For $N \to \infty$ and $\alpha \leqslant 0.14$, these ferromagnetic states differ in no more than 1.5% (Amit *et al* 1985a) of spin sites with respect to a nominal state.)

Figure 1 shows the mean fraction $f(m_0)$ (averaged over more than 1000 initial states) of states which were recalled with less than $\frac{1}{16} N$ errors for various initial overlaps m_0 and for $p = N\alpha$ in (4). (Note that error bars, which were typically between 1 and 2%, are suppressed in the interests of clarity.)

The rate at which $f(m_0)$ increases from 0 to near 1 becomes more pronounced for $\alpha \leqslant 0.13$ as the size of the system increases from $N = 512$ through $N = 1024$ to $N = 2048$ spins. This change may approach a discontinuity as $N \to \infty$, displaying a critical





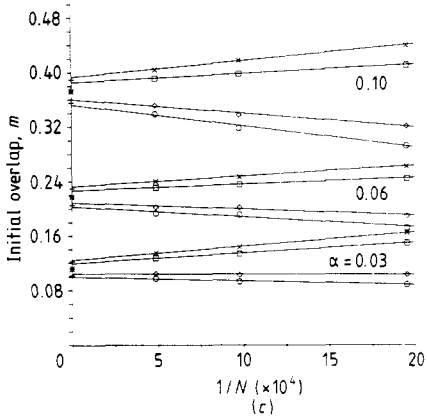


Figure 1. The Hopfield model: (*a*) the fraction, $f(m_0)$, of nominal states that are recalled with less than $\frac{1}{16}N$ errors from initial states having an initial overlap m_0. Best-fit scaling forms (6) are shown for $N = 512$ (◇), 1024 (◯) and 2048 (×) at $\alpha = 0.03$, 0.06 and 0.10. (*b*) The finite-size scaling relation (6) is tested by plotting $\ln[f/(1-f)]$ against m_0; (*c*) The initial overlap, m, required to give a mean recall fraction f is plotted against $1/N$ for four different values of f (◯, $f = 0.2$; ◇, $f = 0.3$; □, $f = 0.7$; ×, $f = 0.8$). Extrapolation to $1/N = 0$ gives $m = m_c$, the critical minimum overlap. *, extrapolation to m_c.

minimum overlap $m_c(\alpha)$ which an initial pattern must have with a stored pattern in order to be in its basin of attraction (and thus ensure its successful retrieval).

This hypothesis is borne out by figure 1(*b*) which demonstrates that the finite-size scaling behaviour

$$f(m_0)/(1-f(m_0)) = C \exp[Na(m_0 - m_c)] \tag{6}$$

is obeyed to a good approximation. Best-fit curves of this form are shown for $\alpha = 0.03$, 0.06 and 0.10 in figure 1(*a*).

Estimates for the critical overlaps $m_c(\alpha)$ are obtained by extrapolating to $N \to \infty$: in figure 1(*c*) the initial overlap $m_0(f)$ required to obtain a given mean recall fraction f is plotted against $1/N$. This analysis then suggests

$$m_c(0.03) = 0.111(10) \qquad m_c(0.06) = 0.218(13) \qquad m_c(0.1) = 0.372(17).$$

(The figures in the brackets on the right-hand side of these results denote the extrapolation errors due to the statistical fluctuations in the simulations for finite N.)

Thus, for example, in a large network storing $0.06N$ patterns, each pattern could be successfully recalled (with less than $\frac{1}{16}N$ errors) from at most 78.2% noise (39.1% spins flipped).

Such scaling behaviour, characteristic of a first-order phase transition, would imply that the basins of attraction of the stored patterns become completely isotropic in the limit of $N \to \infty$.

In contrast to this behaviour, at $\alpha = 0.15$ $f(m_0)$ decreases (figure 1(*a*)) as N increases. This is consistent with an analytical study of this model (Amit *et al* 1985a) which would predict $f(m_0) = 0$ for $0 \leq m_0 \leq 1$ as $N \to \infty$ since this value of α lies beyond the critical value ($\simeq 0.14$) for which no nominal states (or states highly correlated with them) are stored.

Another measure of the content-addressability is described in figure 2: the mean final overlap, m_f, with a nominal configuration which a state acquires from an initial state having overlap m_0. As with the mean recall fraction in figure 1(*a*), m_f is seen to increase more sharply as N increases for $\alpha \leq 0.13$, while it deteriorates for $\alpha = 0.15$. Although, as with f (figure 1(*a*)), m_f stays close to 1 for large enough m_0 at $\alpha \leq 0.13$, m_f only becomes zero for $m_0 \to 0$ since states which are not recalled nevertheless retain a non-zero final overlap with the nominal states.

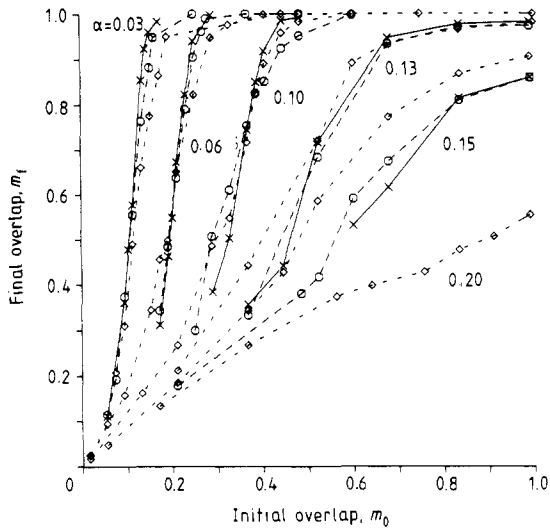


Figure 2. Mean final overlap, m_f, after iteration from states (stored in the Hopfield model) having initial overlap m_0. $N = 512$ (◇), 1024 (◯), 2048 (×).

2. Content-addressability after learning

A 'learning' algorithm for modifying the connection strengths T_{ij} (which is essentially an extension of perceptron learning (Minsky and Papert 1969)) has been shown (Wallace 1986, Bruce *et al* 1986) to be capable of perfectly storing up to at least N patterns on a network of N spins. This algorithm is as follows. An error mask ε_i^r is defined at each site i for each nominal pattern (r):

$$\varepsilon_i^r = \frac{1}{2}\left[1 - \mathrm{sgn}\left(S_i^{(r)} \sum_{j=1}^{N} T_{ij} S_j^{(r)}\right)\right]. \tag{7}$$

In other words, the mask is assigned the value 0 (1) if the spin at site i is (is not) aligned with its local field. The connections T_{ij} are then modified accordingly:

$$T_{ij} \to T_{ij} + N^{-1} \sum_{r=1}^{p} (\varepsilon_i^r + \varepsilon_j^r) S_i^{(r)} S_j^{(r)} \qquad T_{ii} = 0. \tag{8}$$

This form retains the symmetry of the connections and thus ensures the existence of a Lyapunov or energy function (Hopfield 1982)

$$E(\boldsymbol{S}) = -\tfrac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} T_{ij} S_i S_j \tag{9}$$

so that any trajectory under single spin-flip dynamics will terminate at a stable configuration (which minimises E). (An asymmetric rule in (8) could equally well be used. Convergence theorems exist for both the symmetric (Wallace 1986) and asymmetric (Gardner *et al* 1987a) versions.)

Note that this algorithm is local in the sense that any modification to a connection strength only depends on the error mask and states of the spins at the sites at either end of the connection. There is therefore obvious parallelism to be exploited in executing the algorithm.

Starting from the prescription (4) and then using (7) and (8), $N\alpha$ states were stored exactly on a network of $N = 512$ for values of α between 0.13 and 1.0. Figure 3 shows the mean final overlap, m_f, after iteration from a state with overlap m_0. (These results typically involved around 1000 states per value of m_0 and up to 16 000 single-spin updates per second were achieved.) Although all the nominal states are indeed perfectly stored for $\alpha = 0.5$ and $\alpha = 1.0$, they have very poor content-addressability. (In fact, for $\alpha = 1.0$, initial states having merely one spin misaligned out of the 512 results in over 65% of the states failing to be recalled—the non-recalled states end up with a final overlap of around 0.6.) Thus this learning algorithm appears to install no appreciable content-addressability at these higher values of α (0.5 and 1.0), but instead merely creates fixed points of (1) which have negligible regions of attraction.

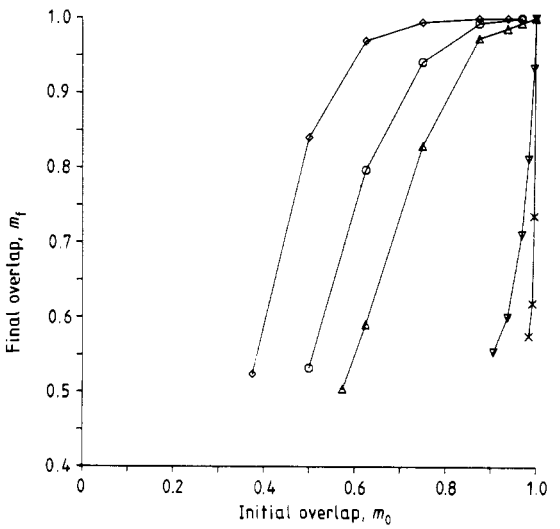


Figure 3. As figure 2, but after use of the learning algorithm (7) and (8). $N = 512$; $\alpha = 0.13$ (◇), 0.15 (◯), 0.20 (△), 0.50 (▽), 1.0 (×).

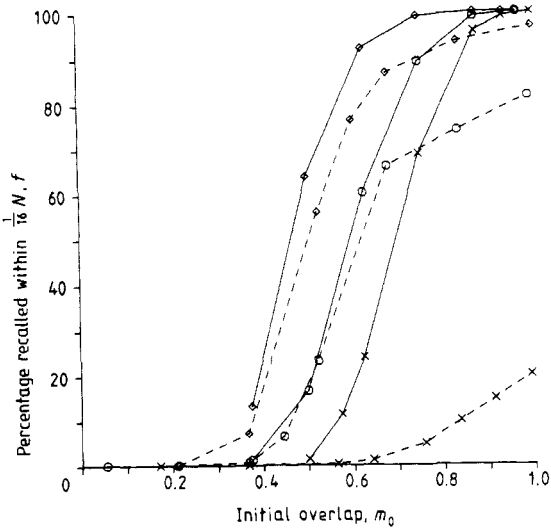


Figure 4. Comparison of the fraction of nominal states recalled with less than $\frac{1}{16}N$ errors from various initial overlaps m_0 before (broken lines) and after (full lines) utilisation of the learning algorithm (7) and (8), starting from the Hopfield-model prescription (4). $N = 512$; $\alpha = 0.13$ (◇), 0.15 (◯), 0.20 (×).

However, content-addressability is substantially improved at $\alpha = 0.15$ and 0.2 (and improved to a lesser extent at $\alpha = 0.13$), as is evident from figure 4, which compares the fraction of states recalled with less than $\frac{1}{16}N$ errors from various initial overlaps m_0 before and after implementation of the learning algorithm. Simulations with larger N would be required to determine whether finite content-addressability would be installed as $N \to \infty$.

3. An improved learning algorithm

The tiny basins of attraction created around the nominal states by the algorithm (7) and (8) for higher values of the storage ratio α are probably due to the fact that when the algorithm terminates, although all the spins $S_i^{(r)}$ in any nominal configuration $\boldsymbol{S}^{(r)}$ are correctly aligned with their local fields (h_i), the majority of them will not be strongly aligned (i.e. $h_i S_i^{(r)}$, although positive, will be small).

Now, if F denotes the set of spins which have been flipped with respect to a nominal state $\boldsymbol{S}^{(r)}$, then an aligned spin $S_i^{(r)}$ will also flip if

$$2S_i^{(r)} \sum_{j \in F} T_{ij} S_j^{(r)} > S_i^{(r)} h_i. \tag{10}$$

Consequently, even if a small number of spins have been flipped, this may be sufficient to misalign many of the local fields and project the state of the system on a trajectory taking it further from the nominal state.

In order to alleviate this problem learning schemes have been proposed (Gardner *et al* 1987a, b, Poeppel and Krey 1987) which involve iterating from 'noisy' versions of nominal configurations. This approach requires sampling a large enough representation of starting configurations for each nominal one.

Other approaches which have been proposed (Diederich and Opper 1987, Gardner 1987, 1988, Krauth and Mézard 1987), and the one adopted here, attack the problem of weak alignment between spin and local field directly by attempting to find solutions for the T_{ij} subject to the more stringent condition

$$S_i^{(r)} h_i > B > 0 \qquad i = 1, \ldots, N;\ r = 1, \ldots, p \tag{11}$$

on the alignment of the spins $S_i^{(r)}$. This was achieved here by using a modification of the error mask (7);

$$\varepsilon_i^r = \frac{1}{2}\left[1 - \operatorname{sgn}\left(S_i^{(r)} \sum_{j=1}^{N} T_{ij} S_j^{(r)} - B\right)\right]. \tag{12}$$

The bound B was scaled to be of the same order of magnitude as $S_i^{(r)} h_i$—if the connections T_{ij} are $O(T)$ then h_i (and hence $S_i^{(r)} h_i$) will be $O(T\sqrt{N})$. The form of B employed here was

$$B = M\langle |T_{ij}| \rangle N^{1/2} \tag{13}$$

where the angular brackets denote the average with respect to all the bonds in the jth row of the T_{ij} matrix, which was performed after all the nominal states had been tested (so that learning was still local during each sweep). The form of B ensures that M is $O(1)$, independent of N and of any rescaling of the T_{ij} (under which the dynamics (1) are invariant).

That the imposition of (11) results in a direct and substantial improvement in the content-addressability of the solution for the connections T_{ij} can be seen from the numerical results presented in figures 5(a) and (b). Both at $\alpha = 0.25$ (figure 5(a)) and $\alpha = 0.5$ (figure 5(b)) the trend from $N = 256$ to $N = 512$ indicates that the ordinary learning algorithm (7)—which corresponds to $B = M = 0$ in (12)—will probably install negligible content-addressability as $N \to \infty$. In contrast, however, the perfect recall fraction, f_p, for a given $M > 0$, exhibits a crossover effect for increasing N similar to the behaviour of the recall fraction for lower values of α in figure 1(a). A best-fit

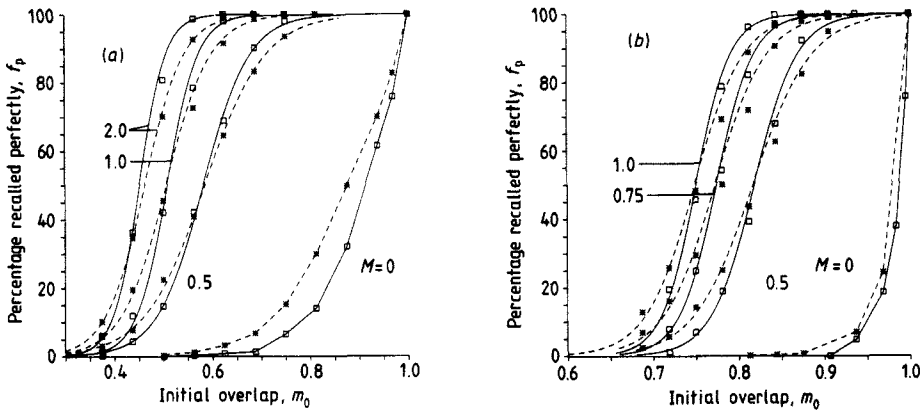


Figure 5. After completion of the improved learning algorithm (12) and (8) at (a) $\alpha = 0.25$ and (b) $\alpha = 0.5$. The fraction, f_p, perfectly recalled from states with initial overlap m_0 is plotted for two different system sizes: $N = 256$ (broken curves) and $N = 512$ (full curves). A best-fit scaling form (6) is shown for the three non-zero values of M.

finite-size scaling form (6) for $f_p(m_0)$ is also plotted in figure 5 and extrapolation to $N \to \infty$ yields the following estimates for the critical minimum overlap $m_c(\alpha, M)$:

$$m_c(\alpha = 0.25, M = 2.0) = 0.44(2) \qquad m_c(\alpha = 0.5, M = 1.0) = 0.75(3).$$

Then, for example, the solution with $M = 1.0$ at $\alpha = 0.5$ on a large network would ensure perfect recall of any of the $\frac{1}{2}N$ nominal patterns from states having no more than 25% noise (12.5% spins flipped).

The rate of decrease of wrongly stored spins is approximately exponential as learning proceeds (figure 6). (One learning cycle is defined as a complete sweep (8) through all the nominal states.) As M increases, this exponential learning rate decreases and will presumably tend to zero as M approaches the optimal value for symmetric connections. These results also demonstrate that the rate of learning appears to be independent of N.

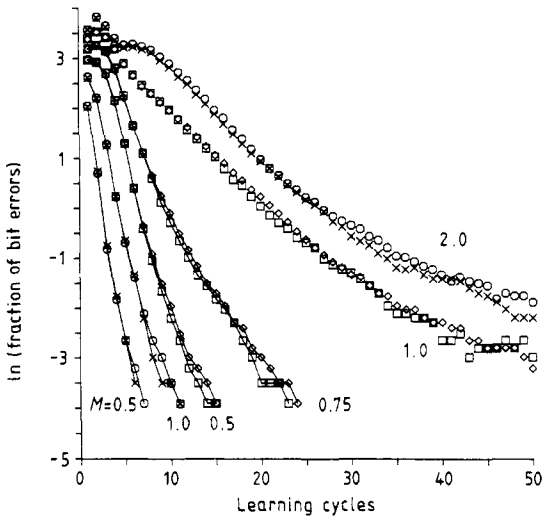


Figure 6. The (natural logarithm of the) fraction of wrongly stored spins as the learning algorithm (12) and (8) proceeds. One learning cycle consists of a complete sweep through all the nominal vectors. $\alpha = 0.25$, $N = 256$ (×), 512 (◯); $\alpha = 0.5$, $N = 256$ (◇), 512 (□).

In figure 7 the number of 'learns' per nominal state is plotted against M; one learn consists of a sweep through one nominal pattern if it has at least one bit (spin) wrongly stored. The number of such learns required grows rapidly as M approaches the largest values used here—$M = 2.0$ at $\alpha = 0.25$ and $M = 1.0$ at $\alpha = 0.5$—suggesting that these values of M are near the largest possible (at least for symmetric T_{ij}).

If connections J_{ij} are chosen which satisfy

$$\sum_{j \neq i} J_{ij}^2 = N \tag{14}$$

then the optimal value of K such that

$$N^{-1/2} S_i^{(r)} \sum_{j \neq i} J_{ij} S_j^{(r)} > K \qquad i = 1, \ldots, N; r = 1, \ldots, p \tag{15}$$

has been shown (Gardner 1987, 1988) to satisfy

$$1/\alpha = (2\pi)^{-1/2} \int_{-K}^{\infty} (t + K)^2 \exp(-\tfrac{1}{2}t^2)\, dt \tag{16}$$

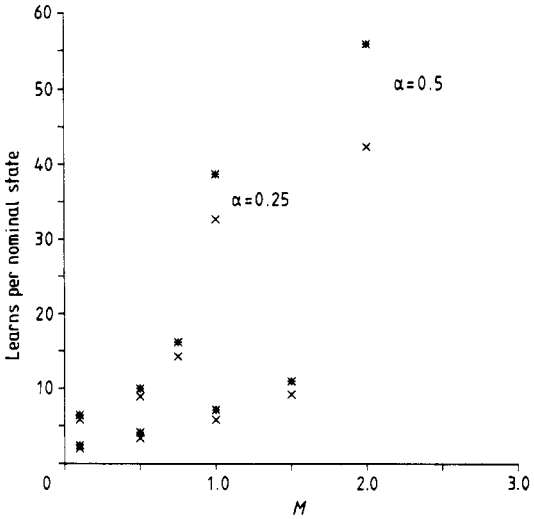


Figure 7. The number of learns required per nominal state; a nominal state contributes one to the number of learns during a learning cycle only if it has at least one spin wrongly stored at that time. $N = 256$ (×), 512 (*).

when, in general, the connections J_{ij} are allowed to assume asymmetric values. Now, if the connections T_{ij} used here are rescaled (under which (1) is invariant) to

$$J_{ij} = T_{ij}\left(N^{-1}\sum_{j=1}^{N} T_{ij}^2\right)^{-1/2} \tag{17}$$

then they satisfy (14) and (11) becomes

$$N^{-1/2}S_i^{(r)}\sum_{j\neq i} J_{ij}S_j^{(r)} > M(2/\pi)^{1/2} \tag{18}$$

as the distribution of T_{ij} remains Gaussian of zero mean (figure 8) after learning—it merely becomes wider for larger M—so that, in (18), M is multiplied by the ratio of

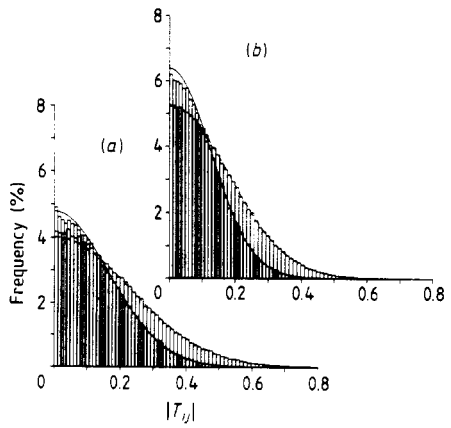


Figure 8. The final distribution of the absolute value of the connections T_{ij} after completion of the learning algorithm (12) and (8). A best-fit Gaussian form is provided. (*a*) $M = 2.0$, $\alpha = 0.25$; (*b*) $M = 1.0$, $\alpha = 0.5$. Gaussian fit: $N = 256$ (- - -), 512 (——).

the mean absolute value of a Gaussian variable (of zero mean) to its width, i.e. $(2/\pi)^{1/2}$. (The width of the final T_{ij} distribution is a factor of $\sqrt{2}$ narrower for $N = 512$ than for $N = 256$. This is consistent with the expectation that the width of the integer distribution NT_{ij} should be proportional to the square root of the number of patterns being learnt, i.e. to $\{N\alpha\}^{1/2}$—the number of modifications to a typical T_{ij} should be proportional to $p = N\alpha$ and these modifications are approximately random.)

Thus the formulation employed in this paper is equivalent to (14) and (15) (for which there exists a convergence theorem (Gardner 1988)) with

$$K = M(2/\pi)^{1/2}. \tag{19}$$

The optimal values of $K(\alpha)$ for $\alpha = 0.25$ and 0.5 are (from (16))

$$K(0.25) \simeq 1.74 \qquad K(0.5) \simeq 1.04.$$

These correspond to

$$M(0.25) \simeq 2.18 \qquad M(0.5) \simeq 1.30$$

which lie above the largest values $M(0.25) = 2.0$ and $M(0.5) = 1.0$ tried here, at which learning becomes rapidly more difficult (figure 7). It may be that these optimal values of M cannot be attained while maintaining the symmetry of the T_{ij}.

4. Conclusions

As is well known by now, the Hopfield model is rather limited in terms of its capacity to store patterns—in the limit of $N \to \infty$ it fails to store more than $0.15N$ patterns. However, the analysis of the numerical studies of this model presented here would indicate that, allowing for some recall errors, it still performs as a reasonable content-addressable memory for storage ratio α up to 0.10. Previous work (Amit 1987) has shown good retrieval for α up to 0.13. Not surprisingly, the content-addressability decreases as more patterns are stored—the growing number of spuriously created states, in addition to the other nominal states themselves, reduce the region of attraction around a nominal state.

Implementation of the 'naive' learning algorithm (7) and (8) has been shown to provide perfect storage up to at least $\alpha = 1.0$, but only achieves appreciable content-addressability for $\alpha \leqslant 0.2$ at $N = 512$. The solution it provides for higher values of α look like being of no use at large N.

In contrast, generalisation of (7)-(12) has been shown to provide finite content-addressability for finite values of M in (13) for α up to at least 0.5. The trend for increasing N suggests that finite content-addressability will be achieved as $N \to \infty$.

The number of bits required for storage of the synapses T_{ij} is larger for this learning algorithm than for the Hopfield model, since a wider distribution of T_{ij} is produced. However, the width of the integer synapses NT_{ij} scales as $\sqrt{N}$ (figure 8), as it does in the Hopfield model, so the number of bits of synaptic information required will be $O(N^2 \ln \sqrt{N})$ in both cases. Therefore, since they both store $O(N^2\alpha)$ bits of information, their relative storage capacities—the number of bits of stored information per bit of synaptic information—will both be of the same order, namely, $O(1/\ln N)$. Hence the improved learning algorithm is superior even in terms of relative capacity since it stores all $N^2\alpha$ bits exactly and provides a larger region of attraction.

It is not clear whether imposing symmetry on the connections will prevent the optimal value of M being attained. However, these results indicate that values of M

not far below the optimal are achievable within this restriction. A feature of this type of algorithm is that only one iteration sweep per nominal state is required for one learning cycle. Algorithms which involve iteration from noisy initial patterns require testing more than one state per nominal state at each learning cycle. The number of possible states differing from a nominal state in a given amount of sites increases exponentially in the number of sites which differ, although the number of states required to train on to achieve finite content-addressability may not necessarily suffer this exponential growth. This latter type of algorithm can provide the opportunity of creating anisotropic basins of attraction, an option not provided by the type used here. Comparison of both types of algorithms would be required to determine which solution provided fastest iteration to stability after learning has been completed, although, unlike the algorithm used here, the 'learning with noise' algorithms can provide for solutions which recover nominal states from noisy versions in one sweep.

It would be interesting to apply this algorithm to much larger systems than $N = 256$ and $N = 512$ to see if the basins of attraction continue to exhibit scaling behaviour for finite M, as the trend for increasing N suggests here.

Acknowledgments

I have benefitted from helpful discussions with E Gardner and useful comments from D Wallace and A Bruce. This work is funded by SERC and GEC Ltd.

References

Amit D J 1987 *Heidelberg Symp. on Glassy Dynamics* ed I Morgenstern and J L Van Hemmen (Berlin: Springer)

Amit D J, Gutfreund H and Sompolinsky H 1985a *Phys. Rev. Lett.* **55** 1530

—— 1985b *Phys. Rev.* A **32** 1007

—— 1987a *Ann. Phys., NY* **173** 30

—— 1987b *Phys. Rev.* A in press

Bruce A D, Canning A, Forrest B, Gardner E and Wallace D J 1986 *Proc. Conf. on Neural Networks for Computing, Snowbird, UT (AIP Conf. Proc. 151)* ed J S Denker (New York: AIP) p 65

Bruce A D, Gardner E and Wallace D J 1987 *J. Phys. A: Math. Gen.* **20** 2909

Cover T M 1965 *IEEE Trans. Electron. Comput.* **EC-14** 326

Diederich S and Opper M 1987 *Phys. Rev. Lett.* **58** 949

Gardner E 1986 *J. Phys. A: Math. Gen.* **19** L1047

—— 1987 *Europhys. Lett.* **4** 481

—— 1988 *J. Phys. A: Math. Gen.* **21** 257

Gardner E, Stroud N and Wallace D J 1987a *Preprint* Edinburgh 87/394

—— 1987b *Preprint* Edinburgh 87/410

Hebb D O 1949 *The Organisation of Behaviour* (New York: Wiley)

Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554

Krauth W and Mézard M 1987 *J. Phys. A: Math. Gen.* **20** L745

Little W A 1974 *Math. Biosci.* **19** 101

McCulloch W S and Pitts W A 1943 *Bull. Math. Biophys.* **5** 115

Mézard M, Nadal J P and Toulouse G 1986 *J. Physique* **47** 1457

Minsky M L and Papert S 1969 *Perceptrons* (Cambridge, MA: MIT Press)

Poeppel G and Krey U 1987 *Preprint* Regensburg

Venkatesh S 1986a *Proc. Conf. on Neural Networks for Computing, Snowbird, UT (AIP Conf. Proc. 151)* ed J S Denker (New York: AIP)

—— 1986b *PhD thesis* California Institute of Technology

Wallace D J 1986 *Lattice Gauge Theory—A Challenge to Large Scale Computing* ed B Bunk and K H Mutter (New York: Plenum)