Mean-field analysis of hierarchical associative networks with 'magnetisation'

1988 J. Phys. A: Math. Gen. 21 2211

(http://iopscience.iop.org/0305-4470/21/9/033)

J. Phys. A: Math. Gen. **21** (1988) 2211-2224. Printed in the UK

Mean-field analysis of hierarchical associative networks with 'magnetisation'

A Krogh† and J A Hertz‡

† Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen Ø, Denmark
‡ Nordita, Blegdamsvej 17, 2100 Copenhagen Ø, Denmark

Received 21 September 1987, in final form 23 December 1987

Abstract. A Hopfield-type neural network that can store ultrametrically organised patterns with a finite magnetisation, or bias, is studied. Both the patterns and their ancestors are remembered in one network. A homogeneous field (or threshold) is included in the model. In zero field the (replica symmetric) mean-field equations map onto those of the Hopfield model, implying that transition temperatures, storage capacity, etc, are the same. By varying appropriately the field and/or a constant added to all the synaptic strengths, it is possible to climb up and down the hierarchical tree or to focus on a certain level in the hierarchy and thereby increase the capacity slightly.

1. Introduction

The Hopfield model [1] of associative memory in neural networks is today reasonably well understood. A major breakthrough came with the mean-field theory (MFT) of Amit *et al* [2], who managed to calculate various critical quantities, e.g. the load capacity $\alpha_c(T)$, which is the maximum number of stable memories at temperature T divided by the number of spins (or neurons) N.

One of the drawbacks of the Hebb prescription used for the synaptic strengths in the Hopfield model is that only a very limited number of correlated patterns can be memorised before they 'destroy each other'. Therefore several groups have studied other rules that are capable of storing a particular kind of correlated pattern [3-8].

Kanter and Sompolinsky studied a very general rule [7], proposed by Personnaz *et al* [9, 10], that can store any set of linearly independent patterns, but at the cost of being fairly complicated and very unbiological in its general form. By using this approach we have proposed a storage rule for ultrametrically correlated patterns [3] which is a simple generalisation of the Hebb rule. Parga and Virasoro [8] and Feigelman and Ioffe [6] have found similar results from different arguments.

Ultrametric organised patterns are interesting for several reasons. First, one may argue that memory is often intrinsically hierarchical, lending itself naturally to ultrametric encoding. Second, a layered structure of the perception system may lead to ultrametric organisation of memories, as discussed by Parga and Virasoro [8]. Third, the ultrametric structure of the SK spin glass makes it interesting to study other systems with this structure.

In this paper we apply the mean-field theory (Amit's recipe) to a class of ultrametrically organised memories. These patterns can have a finite magnetisation; this is motivated by the fact that in real brains most of the neurons are passive at any time.

0305-4470/88/092211+14$02.50

Very interesting results on magnetised networks have been obtained by Amit *et al* [4] on patterns with the same overlap between any two patterns (hierarchy with one level). This has been generalised by Gutfreund [5] to a hierarchy of patterns in a layered structure of networks. The present paper can be seen as a supplement to those works. We have especially concentrated on the effect of a homogeneous external field on the network, which in a sense corresponds to their notion of 'constrained dynamics'.

Gardner *et al* [11] (see also [12]) have calculated the maximum storage capacity of networks with or without magnetisation (with asymmetric couplings allowed). For zero magnetisation this maximum capacity is $\alpha_c = 2$ (and larger when magnetisation is different from zero) which is substantially higher than the capacity of the Hopfield model ($\alpha_c = 0.14$) and also higher than for models using the pseudo-inverse method ($\alpha_c = 1$) [7]. There are no explicit solutions for the optimal couplings but iterative algorithms exist [11, 12].

This work is in a sense complementary to the approach taken in these rules. The starting point is a relatively simple storage rule derived by the pseudo-inverse method. Our aim is to study the effect of variations of macroscopic parameters, rather than simply to maximise the capacity.

1.1. A storage prescription for ultrametric patterns

The p patterns are organised in an ultrametric structure of l levels [13]. To identify a pattern we use l numbers $\alpha_1 \ldots \alpha_l \equiv \bar{\alpha}_l$, describing the way to go in the hierarchical tree. If two patterns $\xi^{\bar{\alpha}_l}$, $\xi^{\bar{\beta}_l}$ have their nearest common branching point at level m (i.e. $\alpha_1 = \beta_1, \ldots, \alpha_m = \beta_m$ and $\alpha_{m+1} \neq \beta_{m+1}$), the overlap is

$$Q^{\bar{\alpha}_l \bar{\beta}_l} = \frac{1}{N} \sum_{i=1}^{N} \xi_i^{\bar{\alpha}_l} \xi_i^{\bar{\beta}_l} = Q_m. \tag{1}$$

The group averages (or ancestor states) $\xi^{\bar{\alpha}_m}$ ($\bar{\alpha}_m = \alpha_1 \ldots \alpha_m$) are

$$\xi^{\bar{\alpha}_m} = \frac{1}{k_m} \sum_{\alpha_{m+1}} \cdots \sum_{\alpha_l} \xi^{\bar{\alpha}_m \alpha_{m+1} \cdots \alpha_l} \tag{2}$$

where k_m is the number of patterns in the groups at level m ($p = k_0$).

We have recently proposed a storage rule for this type of hierarchy [3]. In the present notation it is

$$J_{ij}^0 = \frac{p}{N\eta_0} \xi_i^0 \xi_j^0 + \frac{1}{N} \sum_{m=1}^{l} \frac{k_m}{\eta_m} \sum_{\bar{\alpha}_m} (\xi_i^{\bar{\alpha}_{m-1}\alpha_m} - \xi_i^{\bar{\alpha}_{m-1}})(\xi_j^{\bar{\alpha}_{m-1}\alpha_m} - \xi_j^{\bar{\alpha}_{m-1}}) \tag{3}$$

where η_m are the eigenvalues of the overlap matrix Q:

$$\eta_m = Q_l + (k_{l-1} - 1)Q_{l-1} + \ldots + (k_m - k_{m+1})Q_m - k_m Q_{m-1} \tag{4}$$

(Q_{-1} is formally put equal to 0).

1.2. Outline of the paper

In the next section our model is presented. The way we build the hierarchical tree is described and the storage rule is generalised to incorporate a field.

Section 3 concerns the mean-field theory (MFT). The replica formalism is employed and saddle-point equations for the order parameters are found. In the end we restrict ourselves to the replica symmetric theory.

Finally, § 4 is devoted to the retrieval states and, in particular, the storage capacity, $\alpha_c(T=0)$, under different circumstances. In addition, the critical temperature is considered.

2. The model

2.1. Generating the ultrametric structure

The ultrametric structure can be generated by an inhomogeneous Markov process [8, 13]. The one described here is similar to one proposed by Parga and Virasoro [8] and Feigelman and Ioffe [6] and has also been used by Gutfreund [5].

First the 'oldest' ancestor state is generated. This determines the probability distribution for its descendants, the second oldest ancestors. They determine the probability distribution for the next generation and so on. When the branching ratio goes to infinity the mean values $\xi^{\bar{\alpha}_m}$ (in equation (3)) become equal to these ancestors. Therefore the ancestors are also called $\xi^{\bar{\alpha}_m}$—hopefully this will not cause confusion.

The process looks as follows:

$$\begin{aligned} &\xi_i^0 = a_0 \qquad \text{for all } i \text{ ('oldest' ancestor)} \\ &P(\xi^{\bar{\alpha}_m \alpha_{m+1}} = \pm a_{m+1} | \xi^{\bar{\alpha}_m}) = \tfrac{1}{2}(1 \pm \xi^{\bar{\alpha}_m}/a_{m+1}) \end{aligned} \tag{5}$$

where

$$0 \leqslant a_0 < a_1 < \ldots < a_l = 1.$$

Here we have chosen positive a, but they could just as well be negative.

It can easily be shown that

$$\langle\langle \xi^{\bar{\alpha}_m} \rangle\rangle = a_0 \tag{6}$$

$$\langle\langle \xi^{\bar{\alpha}_m \alpha_{m+1} \ldots \alpha_l} \xi^{\bar{\alpha}_m \beta_{m+1} \ldots \beta_s} \rangle\rangle = a_m^2 \tag{7}$$

where $\langle\langle\ \rangle\rangle$ is an average over the probability distribution of the patterns (5). Equation (7) tells us that the overlap between two states with nearest common ancestor at level m is

$$Q_m = a_m^2. \tag{8}$$

When the number of patterns is finite and $N \to \infty$, fluctuations in these overlaps will be suppressed. But when $\alpha = p/N$ is finite (as $N \to \infty$) this is not the case, and the fluctuations will add up and act as a noise—exactly the situation in the Hopfield model [2].

2.2. Infinite branching ratio

In the following mean-field analysis it is assumed that the branching ratio at each branching point goes to infinity, i.e. $k_m/k_{m+1} \to \infty$. This can be done by putting

$$k_m = b^{l-m} \tag{9}$$

and let b (the branching number) go to infinity. It is now obvious that the mean value $\xi^{\bar{\alpha}_m}$ tends to the ancestor as $b \to \infty$.

Note that the eigenvalues (4) of the overlap matrix become

$$\begin{aligned} &\lim_{b\to\infty} (\eta_0) = k_0 a_0^2 \qquad (a_0 \neq 0) \\ &\lim_{b\to\infty} (\eta_m) = k_m (a_m^2 - a_{m-1}^2). \end{aligned} \tag{10}$$

In this limit our storage rule (3) is identical to the one of Parga and Virasoro [8], as is shown in [3].

2.3. Storage prescription

Using these results ((6) and (10)) the first term in our prescription for the couplings (3) becomes a constant,

$$J_{ij}^0 = \frac{1}{N} + \frac{1}{N} \sum_{m=1}^{l} \frac{k_m}{\eta_m} \sum_{\tilde{\alpha}_m} \delta_i^{\tilde{\alpha}_m} \delta_j^{\tilde{\alpha}_m} \tag{11}$$

where the δ notation should be clear.

In the following analysis a homogeneous external field h is included. This field is (except for sign) equivalent to a firing threshold of the neurons. Now the local field becomes

$$h_i = \sum_j J_{ij} S_j + h. \tag{12}$$

The original rule ((3) or (11)) was derived with the assumption that $h = 0$ and it is therefore natural to modify it by including the field in its derivation. This is done in the way proposed by Personnaz *et al* [9, 10], which is also the basis for (3). For a memory to be stable the local fields in this state must be parallel to the spins. That is ensured by requiring

$$\sum_j J_{ij} \xi_j^\mu + h = \xi_i^\mu \qquad \mu = 1, \ldots, p \tag{13}$$

where the 'final' patterns $\xi_i^{\tilde{\alpha}_l}$ are now numbered from 1 to p. (In [11, 12] this rather restrictive equality is replaced by an inequality saying that the local field times the spin should be positive.)

Now the two $N \times p$ matrices Ξ and E are introduced, where all elements in E are 1 and $\Xi = (\xi^1, \xi^2, \ldots, \xi^p)$. Equation (13) is now written in the following way:

$$J\Xi + hE = \Xi \tag{14}$$

which is solved for J:

$$J = \Xi\Xi^I - hE\Xi^I \tag{15}$$

where Ξ^I is the pseudo-inverse [10],

$$\Xi^I = (\Xi^T\Xi)^{-1}\Xi^T = (1/N)Q^{-1}\Xi^T. \tag{16}$$

The first term in (15) is the one that gives J_{ij}^0 (3), the second becomes (for the ultrametric case)

$$-\frac{h}{N} EQ^{-1}\Xi^T = -\frac{h}{N\eta_0} E\Xi^T. \tag{17}$$

Finally, the J matrix is

$$J_{ij} = J_{ij}^0 - \frac{h}{N\eta_0} \sum_{\mu=1}^{p} \xi_j^\mu = J_{ij}^0 - \frac{hp}{N\eta_0} \xi_j^0. \tag{18}$$

Unless all ξ_j^0 are equal this matrix is asymmetrical. In the present work all the ξ_j^0 are equal to a_0, when the number of patterns are large, so J_{ij} is symmetric and becomes

$$J_{ij} = J_{ij}^0 - h/a_0. \tag{19}$$

Now we introduce a parameter c in the first term of the original storage prescription (11), so the rule becomes

$$J_{ij} = \frac{c}{N} + \frac{1}{N} \sum_{m=1}^{l} \frac{k_m}{\eta_m} \sum_{\bar{\alpha}_m} \delta_i^{\bar{\alpha}_m} \delta_j^{\bar{\alpha}_m}. \tag{20}$$

The first term describes a contribution to the bond strengths which is the same for all pairs (i, j) of spins. This contribution is ferromagnetic or antiferromagnetic depending on the sign of c. In our previous rule (3), derived for zero external field, this interaction was ferromagnetic ($c = 1$), but it need not be for sufficiently large field: our new rule (19) corresponds to $c = 1 - h/a_0$. ($c = 0$ corresponds to the case of unconstrained dynamics of Amit *et al* [4] and $c = -h/a_0$ to their constrained dynamics.)

Although the rule we have derived in this section prescribes a particular value of $c\,(1 - h/a_0)$ for a given external field, in subsequent sections we will vary c and h independently and see how the capacity is affected.

It would be natural to vary the other coefficients in the storage prescription around the value, k_m/η_m, given by our rule (as Feigelman and Ioffe do [6]), but for the sake of simplicity we only vary c and the field h.

2.4. Stable ancestor states

In the limit of large branching ratio the average states are identical to the ancestor states, as discussed in § 2.2, so $|\xi_i^{\bar{\alpha}_m}| = a_m$. Set

$$\sigma_i^{\bar{\alpha}_m} = \frac{1}{a_m} \xi_i^{\bar{\alpha}_m} = \frac{k_m}{a_m p} \sum_{\bar{\alpha}_l} \xi_i^{\bar{\alpha}_l} \tag{21}$$

then $\lim_{b \to \infty} (|\sigma_i^{\bar{\alpha}_m}|) = 1$.

Linear combinations of the 'final' patterns (those at level l) which have elements $+1$ or -1, like (21), are stable states on the same footing as $\xi^{\bar{\alpha}_l}$. This is a result of the construction of J_{ij} (3), as discussed by Kanter and Sompolinsky [7] and noticed also by Parga and Virasoro [8].

This means that the ancestor patterns represented by (21) are also remembered in this model. We will see that by tuning the parameters c and h it is possible to focus on ancestors at different levels of the hierarchy.

3. Mean-field theory

3.1. Replica calculations

This analysis follows closely Amit *et al* [2].

The Hamiltonian becomes

$$\begin{aligned} H &= -\tfrac{1}{2} \sum_{i \neq j} J_{ij} S_i S_j + H_{\text{ext}} \\ &= H_k - \frac{1}{2N} \sum_{m=1}^{l} \frac{k_m}{\eta_m} \sum_{\bar{\alpha}_m} \left(\sum_i \delta_i^{\bar{\alpha}_m} S_i \right)^2 - \frac{c}{2N} \left(\sum_i S_i \right)^2 + H_{\text{ext}} \end{aligned} \tag{22}$$

where H_k is constant (in the limit of large N) and H_{ext} represents external fields. An external field is put on some patterns (or ancestors) $\xi^{\bar{\alpha}_m}$, $\bar{\alpha}_m \in A$, in the following way:

$$H_{\text{ext}} = -h \sum_i S_i - \sum_A h^{\bar{\alpha}_m} \sum_i \delta_i^{\bar{\alpha}_m} S_i \tag{23}$$

(these fields will look less strange when we introduce the order parameters).

The replica method [2] is used to calculate the free energy. The moments of the partition function Z are

$$\langle\langle Z^n\rangle\rangle = \left\langle\!\!\left\langle \operatorname{tr}\exp\left(-\beta\sum_{\rho=1}^{n} H^\rho\right)\right\rangle\!\!\right\rangle$$

$$= \exp(-\beta H_k)\left\langle\!\!\left\langle \operatorname{tr}\exp(-\beta H_{\text{ext}})\exp\left\{\sum_\rho\left[\frac{\beta}{2N}\sum_{m=1}^{l}\frac{k_m}{\eta_m}\sum_{\bar\alpha_m}\left(\sum_i \delta_i^{\bar\alpha_m} S_i^\rho\right)^2\right.\right.\right.$$

$$\left.\left.\left.+\frac{\beta c}{2N}\left(\sum_i S_i^\rho\right)^2\right]\right\}\right\rangle\!\!\right\rangle. \tag{24}$$

With the help of a Gaussian integral identity this can be transformed to

$$\langle\langle Z^n\rangle\rangle = \exp(-\beta H_k)\left\langle\!\!\left\langle \operatorname{tr}\int\left[\prod_{\rho m\bar\alpha_m}\left(\frac{\beta N\eta_m}{2\pi k_m}\right)^{1/2} \mathrm{d}x_\rho^{\bar\alpha_m}\right]\left[\prod_\rho\left(\frac{\beta N}{2\pi}\right)^{1/2}\mathrm{d}y_\rho\right]\right.\right.$$

$$\times\exp\left[-\beta N\sum_\rho\left(\tfrac{1}{2}\sum_{m,\bar\alpha_m}\frac{\eta_m}{k_m}(x_\rho^{\bar\alpha_m})^2+\tfrac{1}{2}y_\rho^2-{\sum}'' x_\rho^{\bar\alpha_m}\frac{1}{N}\sum_i\delta_i^{\bar\alpha_m}S_i^\rho\right)\right]$$

$$\left.\left.\times\exp\left[-\beta N\sum_\rho\left(-{\sum}'(x_\rho^{\bar\alpha_m}+h^{\bar\alpha_m})\frac{1}{N}\sum_i\delta_i^{\bar\alpha_m}S_i^\rho-(\sqrt{c}\,y_\rho+h)\frac{1}{N}\sum_i S_i^\rho\right)\right]\right\rangle\!\!\right\rangle. \tag{25}$$

The prime means that the sum is over all condensed patterns, $\bar\alpha_m \in A$, and a double prime means a sum over the rest.

3.2. Order parameters

When $N\to\infty$ the saddle-point approximation can be used on the integral of (25). The saddle-point equations give

$$x_\rho^{\bar\alpha_m} = \frac{k_m}{\eta_m}\frac{\partial}{\partial h^{\bar\alpha_m}}\langle\langle \ln Z^n\rangle\rangle = \frac{k_m}{\eta_m}\left\langle\!\!\left\langle \frac{1}{N}\sum_i \delta_i^{\bar\alpha_m}\langle S_i^\rho\rangle\right\rangle\!\!\right\rangle \tag{26}$$

$$y_\rho = \frac{\partial}{\partial h}\langle\langle \ln Z^n\rangle\rangle = \sqrt{c}\left\langle\!\!\left\langle \frac{1}{N}\sum_i\langle S_i^\rho\rangle\right\rangle\!\!\right\rangle. \tag{27}$$

So the y order parameter is just the magnetisation per spin times $\sqrt{c}$.

The other order parameter is a generalised overlap with some nice properties. Assume the system condenses into one state $\xi^{\bar\beta_s}$, i.e. $\langle S_i\rangle = x\xi_i^{\bar\beta_s}$, then it is found from (26) and (10) that

$$x^{\bar\alpha_m} = \frac{x}{a_m^2 - a_{m-1}^2}(\langle\langle \xi^{\bar\alpha_m}\xi^{\bar\beta_s}\rangle\rangle - \langle\langle \xi^{\bar\alpha_{m-1}}\xi^{\bar\beta_s}\rangle\rangle). \tag{28}$$

It then follows from (7) that

$$x^{\bar\alpha_m} = \begin{cases} x & \text{if } m\leqslant s \text{ and } \alpha_1,\ldots,\alpha_m = \beta_1,\ldots,\beta_m \\ 0 & \text{otherwise.}\end{cases} \tag{29}$$

Putting this in words: if $\xi^{\bar\alpha_m}$ is an ancestor of (or equal to) $\xi^{\bar\beta_s}$ then $x^{\bar\alpha_m}$ is equal to x, otherwise it is equal to zero.

The system condenses into some states $\bar{\alpha}_m \in A$, $x^{\bar{\alpha}_m} \neq 0$. Then the average $\langle\!\langle\ \rangle\!\rangle$ in (25) over the rest can be done. This averaging is done in the appendix, and the result is identical to the one obtained in the Hopfield model [2]. Therefore one can introduce the familiar order parameters $q_{\rho\sigma}$ and $r_{\rho\sigma}$, use the saddle-point approximation and immediately write down the free energy per spin:

$$f = f_{\text{const}} + \frac{1}{2n} \sum_{\rho m \bar{\alpha}_m} \frac{\eta_m}{k_m} (x_\rho^{\bar{\alpha}_m})^2 + \frac{1}{2n} \sum_\rho (y_\rho)^2 + \frac{\alpha}{2\beta n} \sum_{\rho=1}^{n} \ln \lambda_\rho$$
$$+ \frac{\alpha\beta}{2n} \sum_{\rho \neq \sigma} r_{\rho\sigma} q_{\rho\sigma} - \frac{1}{n\beta} \left\langle\!\!\left\langle \ln \operatorname{tr}_{S^\rho} \exp\left(\tfrac{1}{2}\alpha\beta^2 \sum_{\rho \neq \sigma} r_{\rho\sigma} S^\rho S^\sigma \right.\right.\right.$$
$$\left.\left.\left. + \beta \sum_{\rho m \bar{\alpha}_m} x_\rho^{\bar{\alpha}_m} \delta^{\bar{\alpha}_m} S^\rho + \sum_\rho (\sqrt{c}\, y_\rho + h) S^\rho \right) \right\rangle\!\!\right\rangle \tag{30}$$

where we have now put $h^{\bar{\alpha}_m}$ equal to zero as they are not needed any more. The sums in (30) are only over the condensed patterns and λ_ρ are the eigenvalues of the matrix $\delta_{\rho\sigma} - \beta q_{\rho\sigma}$, where

$$q_{\rho\sigma} = \left\langle\!\!\left\langle \frac{1}{N} \sum_i \langle S_i^\rho \rangle \langle S_i^\sigma \rangle \right\rangle\!\!\right\rangle \tag{31}$$

and

$$r_{\rho\sigma} = \alpha^{-1} \left\langle\!\!\left\langle \sum_{m, \bar{\alpha}_m}'' (a_m^2 - a_{m-1}^2) x_\rho^{\bar{\alpha}_m} x_\sigma^{\bar{\alpha}_m} \right\rangle\!\!\right\rangle. \tag{32}$$

3.3. Replica symmetric theory

The replica symmetry assumption is

$$x_\rho^{\bar{\alpha}_m} = x^{\bar{\alpha}_m} \qquad y_\rho = y \tag{33}$$

$$r_{\rho\sigma} = r \qquad q_{\rho\sigma} = q \qquad (\rho \neq \sigma). \tag{34}$$

Now letting n go to zero in (30) gives an expression for the free energy (again the calculations are identical to the Hopfield case):

$$f = f_{\text{const}} + \tfrac{1}{2} \sum_{m, \bar{\alpha}_m} \frac{\eta_m}{k_m} (x^{\bar{\alpha}_m})^2 + \tfrac{1}{2} y^2 + \frac{\alpha}{2\beta} \left(\ln(1 - \beta + \beta q) - \frac{\beta q}{1 - \beta + \beta q} \right) + \tfrac{1}{2} \alpha\beta r (1 - q)$$
$$- \frac{1}{\beta} \int \frac{dz}{\sqrt{2\pi}} \exp(-\tfrac{1}{2} z^2) \left\langle\!\!\left\langle \ln 2 \cosh \beta \left(\sqrt{\alpha r}\, z + \sum_{m, \bar{\alpha}_m} x^{\bar{\alpha}_m} \delta^{\bar{\alpha}_m} + \sqrt{c}\, y + h \right) \right\rangle\!\!\right\rangle. \tag{35}$$

The variation of f with respect to the order parameters gives the saddle-point equations (the Gaussian average is included in $\langle\!\langle\ \rangle\!\rangle$):

$$\frac{\eta_m}{k_m} x^{\bar{\alpha}_m} = \left\langle\!\!\left\langle \delta^{\bar{\alpha}_m} \tanh \beta \left(\sqrt{\alpha r}\, z + \sum_{m, \bar{\alpha}_m} x^{\bar{\alpha}_m} \delta^{\bar{\alpha}_m} + cM + h \right) \right\rangle\!\!\right\rangle \tag{36}$$

$$M = \frac{y}{\sqrt{c}} = \left\langle\!\!\left\langle \tanh \beta \left(\sqrt{\alpha r}\, z + \sum_{m, \bar{\alpha}_m} x^{\bar{\alpha}_m} \delta^{\bar{\alpha}_m} + cM + h \right) \right\rangle\!\!\right\rangle \tag{37}$$

$$q = \left\langle\!\!\left\langle \tanh^2 \beta \left(\sqrt{\alpha r}\, z + \sum_{m, \bar{\alpha}_m} x^{\bar{\alpha}_m} \delta^{\bar{\alpha}_m} + cM + h \right) \right\rangle\!\!\right\rangle \tag{38}$$

$$r = q / [1 - \beta(1 - q)]^2 \tag{39}$$

where M is the magnetisation per spin.

Based on experience from other applications [2, 5] we expect that the replica symmetry approximation works fairly well, with only small errors at low temperatures. This is confirmed by numerical simulations.

4. Retrieval states

4.1. Order parameter equations

The order parameters for the retrieval states are of the form given by (29). To study the retrieval of a pattern $\xi^{\bar{\alpha}_s}$ on level s we sum from $m=1$ to s on both sides of (36) and get

$$x(a_s^2-a_0^2)=\langle\!\langle(\xi^{\bar{\alpha}_s}-a_0)\tanh\beta[\sqrt{\alpha r}\,z+x(\xi^{\bar{\alpha}_s}-a_0)+cM+h]\rangle\!\rangle \tag{40}$$

$$M=\langle\!\langle\tanh\beta[\sqrt{\alpha r}\,z+x(\xi^{\bar{\alpha}_s}-a_0)+cM+h]\rangle\!\rangle \tag{41}$$

$$q=\langle\!\langle\tanh^2\beta[\sqrt{\alpha r}\,z+x(\xi^{\bar{\alpha}_s}-a_0)+cM+h]\rangle\!\rangle \tag{42}$$

and r given by (39).

The structure of these equations is the same for patterns at all levels in the hierarchy. This is more obvious when (40) is written in the following way:

$$(a_s x)(1-\mu_s^2)=\langle\!\langle(\sigma^{\bar{\alpha}_s}-\mu_s)\tanh\beta[\sqrt{\alpha r}\,z+(xa_s)(\sigma^{\bar{\alpha}_s}-\mu_s)+cM+h]\rangle\!\rangle \tag{43}$$

with (41) and (42) similar. xa_s lies between zero and unity (it is shown below that it is the overlap) and $\mu_s=a_0/a_s$ is the magnetisation of the normalised states $\sigma^{\bar{\alpha}_s}$. So, as a function of μ_s, α, h and c, all properties are the same for states at different levels.

As mentioned before (§ 2.3) the ancestor patterns $\sigma^{\bar{\alpha}_s}$ are as stable as the 'final' patterns $\sigma^{\bar{\alpha}_l}$ $(=\xi^{\bar{\alpha}_l})$. This is a consequence of the way the storage prescription is derived. Therefore it is not very surprising that (40)-(42) are almost identical for patterns at different levels of the hierarchy.

In the limit of zero temperature the saddle-point equations become (see [2])

$$x(a_s^2-a_0^2)=\left\langle\!\!\left\langle(\xi^{\bar{\alpha}_s}-a_0)\operatorname{erf}\frac{1}{\sqrt{2\alpha r}}[x(\xi^{\bar{\alpha}_s}-a_0)+cM+h]\right\rangle\!\!\right\rangle \tag{44}$$

$$M=\left\langle\!\!\left\langle\operatorname{erf}\frac{1}{\sqrt{2\alpha r}}[x(\xi^{\bar{\alpha}_s}-a_0)+cM+h]\right\rangle\!\!\right\rangle \tag{45}$$

$$C\equiv\beta(1-q)=\left(\frac{2}{\pi\alpha r}\right)^{1/2}\left\langle\!\!\left\langle\exp\left(-\frac{1}{2\alpha r}[x(\xi^{\bar{\alpha}_s}-a_0)+cM+h]^2\right)\right\rangle\!\!\right\rangle \tag{46}$$

$$r=1/(1-C)^2. \tag{47}$$

If $c=0$ we arrive at the equations of Amit *et al* [4] (the unconstrained dynamics), and when $c=-h/a$ it is equivalent to the case of constrained dynamics.

4.2. $h=0$, $c=1$

As mentioned before, the case $c=1$ corresponds to our original couplings J_{ij}^0.

If

$$X=x(a_s-a_0)+M \qquad Y=-x(a_s-a_0)+M \tag{48}$$

the average over the patterns in (40) and (41) can be performed (using (5)) and we get

$$x = (1/2a_s)\langle \tanh \beta(\sqrt{\alpha r}\, z + X) - \tanh \beta(\sqrt{\alpha r}\, z + Y)\rangle_z \tag{49}$$

$$M = \langle \tfrac{1}{2}(1 + a_0/a_s) \tanh \beta(\sqrt{\alpha r}\, z + X) - \tfrac{1}{2}(1 - a_0/a_s) \tanh \beta(\sqrt{\alpha r}\, z + Y)\rangle_z. \tag{50}$$

From this it is found

$$X = \langle \tanh \beta(\sqrt{\alpha r}\, z + X)\rangle_z \tag{51}$$

$$Y = \langle \tanh \beta(\sqrt{\alpha r}\, z + Y)\rangle_z. \tag{52}$$

The solution $X = Y$ does not fit (48), so we must have $X = -Y$, and (48) then gives

$$M = a_0 x = a_0 X / a_s. \tag{53}$$

This important relation says that when x is maximum ($x = a_s^{-1}$) then $M = a_0/a_s$, which is the magnetisation of the normalised state $\sigma^{\tilde{\alpha}_s}$ ($= \xi^{\tilde{\alpha}_l}$ when $s = l$). When the temperature rises and x decreases, M follows it linearly. This drift towards lower magnetisation is due to 'the force of entropy'—the system seeks a higher density of states. It hints that a field to lean against would increase the stability of the patterns.

If X is introduced in equation (42) we get

$$q = \langle \tanh^2 \beta(\sqrt{\alpha r}\, z + X)\rangle_z. \tag{54}$$

Now (51), (54) and (39) are identical to the retrieval state equations of the Hopfield model [2], when X plays the role of the overlap. From (26), (48) and (53) it is easily seen that X is indeed the overlap with the normalised patterns $\sigma^{\tilde{\alpha}_s}$:

$$X = \frac{1}{N} \sum_i \sigma_i^{\tilde{\alpha}_s} \langle S_i \rangle. \tag{55}$$

Anything that is found for the retrieval states of the Hopfield model (in zero fields) therefore also apply to the retrieval states (and ancestor states) in this case. The only difference is the magnetisation, which is zero in the Hopfield model and given by (53) in our case.

For example, the load capacity $\alpha_c(T)$ is the same, and the zero-temperature value in particular is

$$\alpha_c(T = 0) = \alpha_c^{\mathrm{Hop}} = 0.138.$$

This result was also found by Feigelman and Ioffe for zero magnetisation [6].

4.3. $h > 0$

When the overlap decreases, the magnetisation follows it. So one could try to put on a field to keep the magnetisation of the system constant.

Solving equations (44)–(47) numerically for $\alpha_c(0)$ with $c = 1$ and $s = l$ gives the dependence of $\alpha_c(0)$ on h shown in figure 1 for some values of a_0. For large a_0, α_c can be raised considerably. This shows that the pseudo-inverse method does not necessarily give the highest capacity.

Choosing J_{ij} according to (18) means

$$c = 1 - h/a_0. \tag{56}$$

The effect on α_c is shown in figure 2 (full curves). These curves all stabilise around $h = 10$ and for high a we even see a maximum. When h is high this is like putting a potential well (parabolic) around magnetisation $M = a_0$, and is equivalent to the constrained dynamics of Amit *et al* [4].

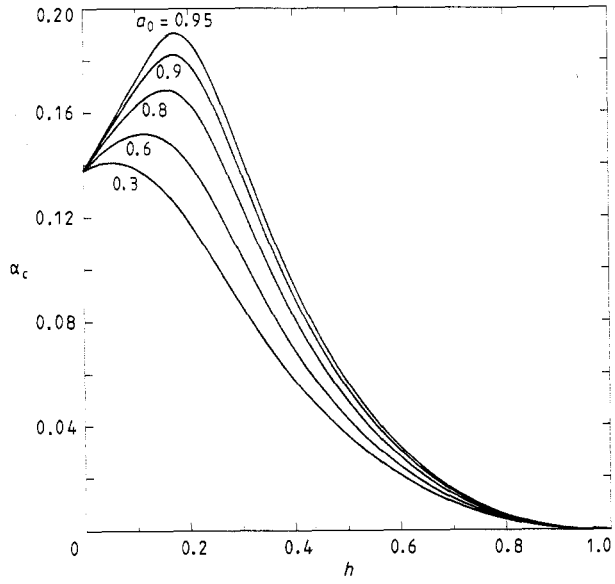

Figure 1. The dependence of α_c on the field h at zero temperature and $c = 1$ for different values of a_0.

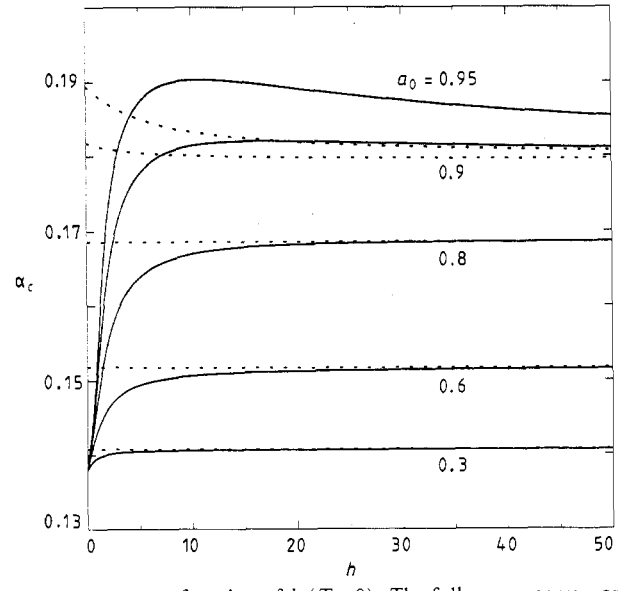

Figure 2. α_c as a function of h ($T = 0$). The full curves correspond to c given by (56) and the broken curves to c given by (57).

From figure 1 we have found that the highest α_c occurs around $h \simeq 0.2a_0$. A field of this size balances the drift towards smaller magnetisation. Instead of (56), which can be written as $h = a_0(1-c)$, we will therefore try $h = a_0(1-c) + 0.2a_0$ or equivalently

$$c = 1.2 - h/a_0. \tag{57}$$

In figure 2 α_c is also shown for this choice of c, h (broken curves). The curves are almost horizontal lines, so picking a pair c, h according to (57) will always give a near-optimal value of α_c.

This way of stabilising patterns is much more flexible than constraining the dynamics to a small part of phase space. It shows that for better storage capacity it is more important to balance the drift towards lower magnetisation than to impose constraints. (On the other hand constrained dynamics may have some advantages, because it destabilises some of the spurious states.)

4.4. Transition temperatures

The different levels of patterns in the hierarchy may have different critical temperatures—say T_m for level m—for the transition from the retrieval phase to the spin-glass phase.

When $c=1$ and $h=0$ all the retrieval states have the same free energy. But when c and h change according to (56) for example, the energy of any state with magnetisation different from a_0 is raised. This is most easily seen from the Hamiltonian (22):

$$\begin{aligned} H &= -\tfrac{1}{2}\sum_{ij} J_{ij}^0 S_i S_j + \tfrac{1}{2}(1-c)NM^2 - hNM \\ &= -\tfrac{1}{2}\sum_{ij} J_{ij}^0 S_i S_j + NhM(M/2a_0 - 1) \end{aligned} \tag{58}$$

where c has been fixed by (56) according to (19). The minimum of the last term is at $M=a_0$. The transition temperatures will in this case increase with level number: $T_0 < T_1 < \ldots < T_l$, because the magnetisations of the states on higher levels are larger than a_0.

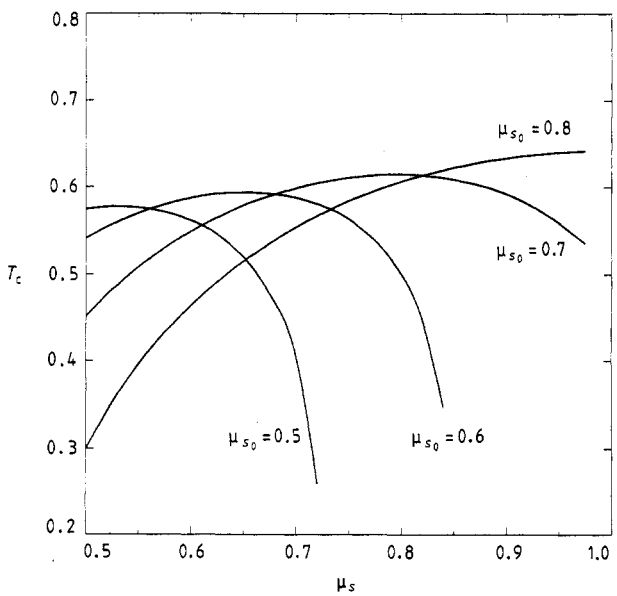

Figure 3. The critical temperature, T_c, as a function of magnetisation when c is given by (59). $\mu_s = a_0/a_s$ is the magnetisation of an ancestor state and μ_{s_0} is the magnetisation of the level in focus ($a_0=0.5$, $\alpha=0.05$, $h=1.0$, $c=1-h/\mu_{s_0}$).

Similarly, one can pick out patterns at a particular level $s = s_0$ by setting

$$c \simeq 1 - h/\mu_{s_0} \tag{59}$$

where μ_{s_0} is the magnetisation of the normalised states $\sigma^{\tilde{\alpha}_{s_0}}$, $\mu_{s_0} = a_0/a_{s_0}$. This will lead to transition temperatures $T_0 < \ldots < T_{s_0}$ and $T_{s_0} > \ldots > T_l$.

In figure 3 we show an example where the mean-field equations have been solved numerically to find the critical temperature for specific values of a_0, h, α, μ_{s_0}. The figure applies to a hierarchy with any number of levels.

If c is kept fixed it is possible to focus on different levels of the hierarchy by tuning the field h according to (59) ($h = \mu_s(1-c)$). If h is high ($h > 1-c$) it is most likely that the system 'sleeps': $\langle S_i \rangle \simeq 1$ for all i. When h is gradually lowered (and $T > 0$) the system can climb down the hierarchical tree and find more and more detailed information.

5. Discussion

The network described here is capable of storing hierarchically correlated magnetised patterns with a storage capacity (α_c) at the same size as—or a little higher than—that of the Hopfield model. Both the patterns and their ancestor states (group averages) are remembered in one network, which is different from the model of Gutfreund [5], where l identical networks are used to do the same.

The model contains two important parameters c and h. c is a constant added to all the synaptic strengths, so it is analogous to a uniform ferro- (or antiferro-) magnetic interaction between all pairs of spins, and h is equivalent to a threshold of the neurons. The pseudo-inverse method gives a simple connection between c and h but we have shown that a small departure from that may increase the storage capacity of the network. This is in agreement with the results of Gardner and others [11, 12].

This system is very flexible. By tuning the parameters β, h and c it is possible to 'focus' on different levels of the hierarchy. Therefore, by varying h for example, it is possible to climb up and down the hierarchical tree.

It is not necessary to constrain dynamics to states around a certain value of magnetisation to get a high storage capacity. But constraining of any strength is possible—again by tuning c and h—and it actually excludes a lot of spurious states [4]. This easy way to get rid of spurious states is one reason that could make these magnetised networks interesting for applications.

For practical use and biological relevance there are obvious difficulties with a model like this. To the well known difficulties of the Hopfield model we have to add at least two. First, to learn something new it must fit into an existing group, or one has to teach the network a whole new group (that must fit into the tree). Second, even though our brain may organise information in some hierarchical manner, the ultrametric structure is probably too rigid.

Feigelman and Ioffe [6] did the mean-field calculation for this network ($c = 1$, $h = 0$) with zero magnetisation ($a_0 = 0$) and found the storage capacity. It is a bit disappointing that the ultrametric structure does not lead to a higher storage capacity, as they point out, because of the analogy of the SK spin glass with an exponential number of stable states. On the other hand, it is clear from the approach taken here that $\alpha_c \leqslant 1$ because the patterns have to be linearly independent [3].

Appendix

In this appendix we calculate the average over the non-condensed states in (25). It takes the form

$$\left\langle\!\!\left\langle \exp\left(\beta N \sum_{\rho m \tilde{\alpha}_m} x_\rho^{\tilde{\alpha}_m} \frac{1}{N} \sum_i \delta_i^{\tilde{\alpha}_m} S_i^\rho\right)\right\rangle\!\!\right\rangle = \prod_i \left\langle\!\!\left\langle \prod_{m\tilde{\alpha}_m} \exp\left(\sum_\rho \beta x_\rho^{\tilde{\alpha}_m} S_i^\rho \delta_i^{\tilde{\alpha}_m}\right)\right\rangle\!\!\right\rangle. \tag{A1}$$

Putting $d^{\tilde{\alpha}_m} = \Sigma_\rho \beta x_\rho^{\tilde{\alpha}_m} S^\rho$ and expanding to second order in the d (as is done in [2]) yields

$$\begin{aligned}
\left\langle\!\!\left\langle \prod_{m\tilde{\alpha}_m} \exp(d^{\tilde{\alpha}_m}\delta^{\tilde{\alpha}_m})\right\rangle\!\!\right\rangle &= \left\langle\!\!\left\langle \prod_{\alpha_1} \exp(d^{\alpha_1}\delta^{\alpha_1}) \prod_{\alpha_2} \exp(d^{\tilde{\alpha}_2}\delta^{\tilde{\alpha}_2}) \ldots \prod_{\alpha_l} \exp(d^{\tilde{\alpha}_l}\delta^{\tilde{\alpha}_l})\right\rangle\!\!\right\rangle \\
&= \left\langle\!\!\left\langle \prod_{m\tilde{\alpha}_m} \exp(d^{\tilde{\alpha}_m}\delta^{\tilde{\alpha}_m}) \prod_{\alpha_l} \exp(-d^{\tilde{\alpha}_l}\xi^{\tilde{\alpha}_{l-1}})\right.\right. \\
&\quad \left.\left. \times [\tfrac{1}{2}(1+\xi^{\tilde{\alpha}_{l-1}}/a_l)\exp(a_l d^{\tilde{\alpha}_l}) + \tfrac{1}{2}(1-\xi^{\tilde{\alpha}_{l-1}}/a_l)\exp(-a_l d^{\tilde{\alpha}_l})]\right\rangle\!\!\right\rangle \\
&= \left\langle\!\!\left\langle \prod_{m\tilde{\alpha}_m} \exp(d^{\tilde{\alpha}_m}\delta^{\tilde{\alpha}_m}) \prod_{\alpha_l} \exp(-d^{\tilde{\alpha}_l}\xi^{\tilde{\alpha}_{l-1}})(1+\tfrac{1}{2}(a_l d^{\tilde{\alpha}_l})^2 + \xi^{\tilde{\alpha}_{l-1}} d^{\tilde{\alpha}_l})\right\rangle\!\!\right\rangle \\
&= \left\langle\!\!\left\langle \prod_{m\tilde{\alpha}_m} \exp(d^{\tilde{\alpha}_m}\delta^{\tilde{\alpha}_m}) \prod_{\alpha_l} \exp[\tfrac{1}{2}(a_l d^{\tilde{\alpha}_l})^2 - \tfrac{1}{2}(\xi^{\tilde{\alpha}_{l-1}} d^{\tilde{\alpha}_l})^2]\right\rangle\!\!\right\rangle \\
&= \exp\left(\tfrac{1}{2}(a_l^2 - a_{l-1}^2)\sum_{\tilde{\alpha}_l} (d^{\tilde{\alpha}_l})^2\right)\left\langle\!\!\left\langle \prod_{m<l,\tilde{\alpha}_m} \exp(d^{\tilde{\alpha}_m}\delta^{\tilde{\alpha}_m})\right\rangle\!\!\right\rangle \\
&\;\;\vdots \\
\left\langle\!\!\left\langle \prod_{m\tilde{\alpha}_m} \exp(d^{\tilde{\alpha}_m}\delta^{\tilde{\alpha}_m})\right\rangle\!\!\right\rangle &= \exp\left(\tfrac{1}{2}\sum_{m,\tilde{\alpha}_m} (a_m^2 - a_{m-1}^2)(d^{\tilde{\alpha}_m})^2\right).
\end{aligned} \tag{A2}$$

Combining (A1) and (A2) and including the quadratic term (x^2) from (25) yields

$$\begin{aligned}
E &\equiv \left\langle\!\!\left\langle \exp\left[-\beta N \sum_{\rho m \tilde{\alpha}_m} \left(\frac{1}{2}\frac{\eta_m}{k_m}(x_\rho^{\tilde{\alpha}_m})^2 + x_\rho^{\tilde{\alpha}_m}\frac{1}{N}\sum_i \delta_i^{\tilde{\alpha}_m} S_i^\rho\right)\right]\right\rangle\!\!\right\rangle \\
&= \exp\left[-\tfrac{1}{2}\beta N \sum_{m,\tilde{\alpha}_m} (a_m^2 - a_{m-1}^2) x_\rho^{\tilde{\alpha}_m} x_\sigma^{\tilde{\alpha}_m}\left(\delta_{\rho\sigma} - \frac{\beta}{N}\sum_i S_i^\rho S_i^\sigma\right)\right].
\end{aligned} \tag{A3}$$

Here (10) is used. Integration of (A3) gives

$$\int \left(\prod_{\rho m \tilde{\alpha}_m} \left(\frac{\beta N \eta_m}{2\pi k_m}\right)^{1/2} \mathrm{d}x_\rho^{\tilde{\alpha}_m}\right) E = \prod_{m\tilde{\alpha}_m} \left(\prod_\rho \lambda_\rho^{-1/2}\right) = \exp\left(-\tfrac{1}{2}\sum_{m=1}^{l} \frac{p}{k_m} \sum_\rho \ln \lambda_\rho\right) \tag{A4}$$

where λ_ρ are the eigenvalues of

$$\Lambda_{\rho\sigma} = \delta_{\rho\sigma} - \frac{\beta}{N}\sum_i S_i^\rho S_i^\sigma. \tag{A5}$$

If $k_m = b^{l-m}$ then

$$\lim_{b\to\infty} \sum_m p/k_m = \lim_{b\to\infty} \sum_m b^m = b^l = p$$

so finally (A4) can be written:

$$\int \left(\prod_{\rho m \tilde{a}_m} \left(\frac{\beta N \eta_m}{2\pi k_m} \right)^{1/2} \mathrm{d}x_\rho^{\tilde{a}_m} \right) E = \exp\left(-\tfrac{1}{2} p \sum_\rho \ln \lambda_\rho \right). \tag{A6}$$

This expression is exactly the same as one finds in the Hopfield model [2].

References

[1] Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554
[2] Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys., NY* **173** 30
[3] Cortes C, Krogh A and Hertz J A 1987 *J. Phys. A: Math. Gen.* **20** 4449
[4] Amit D J, Gutfreund H and Sompolinsky H 1987 *Phys. Rev.* A **35** 2293
[5] Gutfreund H 1988 *Phys. Rev.* A **37** 570
[6] Ioffe L B and Feigelman M V 1986 *Zh. Eksp. Teor. Fiz. Pis. Red.* **44** 148 (*JETP Lett.* **44** 189)
[7] Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380
[8] Parga N and Virasoro M A 1986 *J. Physique* **47** 1857
[9] Personnaz L, Guyon I and Dreyfus G 1985 *J. Physique Lett.* **46** L359
[10] Kohonen T 1984 *Self Organization and Associative Memory* (Berlin: Springer)
[11] Gardner E and Derrida B 1988 *J. Phys. A: Math. Gen.* **21** 271
Gardner E 1987 *Europhys. Lett.* **4** 481
[12] Krauth W and Mézard M 1987 *J. Phys. A: Math. Gen.* **20** L745
[13] Rammal R, Toulouse G and Virasoro M A 1986 *Rev. Mod. Phys.* **58** 765