Neural network with hierarchical clustering near saturation

J. Phys. A: Math. Gen. 25 (1992) 779–795. Printed in the UK

Neural network with hierarchical clustering near saturation

M A Pires Idiart and Alba Theumann

Instituto de Física, Universidade Federal do Rio Grande do Sul, Caixa Postal 15051,
91500 Porto Alegre, RS, Brazil

Received 8 April 1991, in final form 11 October 1991

Abstract. We perform a detailed investigation of the storage properties of a model for neural networks that exhibits the same organization into clusters as Dyson's hierarchical model for ferromagnetism, combined with Hebb's learning algorithm for an extensive number of stored patterns $p = \alpha N$, where N is the size of the network. In a previous publication we presented results for the retrieval properties of the model in the case of finite p, showing that together with the original stored pattern or 'ancestor' the system also retrieves a hierarchy of 'descendants'. Here we first perform a signal-to-noise analysis, obtaining a succession of critical storage capacities for the 'ancestor' and its 'descendants' that are below the Hopfield value. Afterwards we apply the statistical mechanics formulation of Amit, Gutfreund and Sompolinsky, to obtain also in this case a succession of critical storage capacities that are below the corresponding value for Hopfield's model. In both cases we consider the ratio of the critical storage capacity for the 'ancestor' to the same quantity as evaluated in Hopfield's model, and we prove rigorously that the signal-to-noise method provides a lower bound for this ratio, that is bounded from above by unity. We present the phase diagram in the α–T plane for the particular case of two clusters and one descendant. We observe the existence of two lines $T_c^1(\alpha) \leqslant T_M^1(\alpha)$ such that at $T = T_M^1(\alpha)$ the 'ancestor' orders continuously but for $T_c^1(\alpha) < T < T_M^1(\alpha)$ the global minimum is still given by the spin–glass phase, while for $T < T_c^1(\alpha)$ the free energy of the retrieved ancestor becomes a global minimum, just as in Hopfield's model. A new feature of the model studied here is the existence of a third line $T_M^2(\alpha) < T_c^1(\alpha)$ such that at $T = T_M^2(\alpha)$ the 'descendant' orders discontinuously. The existence of a fourth line $T_c^2(\alpha) < T_M^2(\alpha)$ depends on the strength of the interaction.

1. Introduction

In a previous publication [1] we presented a neural network model where the neurons, and not the patterns, were grouped into hierarchical clusters as in Dyson's model [2] for ferromagnetism, only the ferromagnetic interactions were here replaced by Hebb's learning algorithm. One interesting feature of this model is that, although it bears some resemblance to Dotsenko's cluster model [3], it is more tractable and it allows detailed mathematical investigations. The retrieval properties for a finite number of stored patterns were discussed in a second paper [4] (in the following, referred to as I), where we showed that if the number of clusters $l = 2^r$ remains small, $r < 3$, there is perfect retrieval of a family of 'descendants' together with the originally embedded pattern or 'ancestor'. The descendants differ from the ancestor in the relative sign of the cluster overlaps, and they are local minima of the energy while the ancestor always remains as a global minimum. Introducing as usual a 'temperature' T as a

0305-4470/92/040779+17$04.50 © 1992 IOP Publishing Ltd

measure of synaptic noise we obtain that, for a given pattern, the ancestor and its descendants order succesively by reducing T. There is a series of critical temperatures $T_1^* > T_2^* > \cdots > T_r^*$ at which order first the 'ancestor' or original pattern and then the successive 'descendants' obtained by increasing the number of partitions in clusters with different signs for the overlaps. Although the ancestor pairs through a second-order transition, the other transitions are discontinuous. At and below the ordering temperature the ancestor solution of the saddle-point equations remains as a global minimum and the descendants as local minima of the free energy. The importance of these local minima is that they may act as attractors in a dynamical relaxation process. When the number of clusters l increases 'blurred' solutions start to appear that mix the descendants of a given ancestor and may hinder perfect retrieval, the number of these spurious solutions increasing exponentially [4] with large values of l.

In the present paper we present a detailed study of the storage capacity of the model which has an extensive number of stored patterns $p = \alpha N$, where N is the size of the network. Although all theories coincide in predicting a critical value α_c for the storage capacity such that for $\alpha > \alpha_c$ there is no possible retrieval, the actual value obtained for α_c varies according to the different criteria used in the definition of good retrieval. In the signal-to-noise analysis if we make the strong requirement of perfect retrieval at every site, then α_c is itself a decreasing function of N. By using statistical methods [5] we make instead the weaker requirement of having good retrieval on average. In the case of Hopfield's model the signal-to-noise analysis gives $\alpha_H^{SN} = (2\ln N)^{-1}$ while the work of Amit, Gutfreund and Sompolinsky [5] gives $\alpha_H^c = 0.14$, which coincides with the estimates obtained previously by Hopfield in numerical simulations.

This paper is organized as follows. In section 2 we describe the model and in section 3 we perform a signal-to-noise analysis, obtaining a succession of critical storage capacities $\alpha_H^{SN} > \alpha_1^{SN} > \alpha_2^{SN} > \ldots > \alpha_r^{SN}$ for the original pattern and its 'descendants', where the first inequality indicates that all the critical capacities are below the Hopfield value. In section 4 we perform an analysis of the storage capacity by using the statistical mechanics techniques introduced by Amit *et al* [5]. We also obtain a succession of critical values $\alpha_H^c > \alpha_1^c > \alpha_2^c > \cdots$, such that for $\alpha > \alpha_\gamma^c$ there is no retrieval of the γ descendant. We showed rigorously for the original pattern ($\gamma = 1$) that the ratio of critical capacities $\alpha_1^{SN}/\alpha_H^{SN}$ calculated by the signal-to-noise method is a lower bound for α_1^c/α_H^c, the same quantity calculated by statistical mechanics methods. A complete phase diagram in the α–T plane is also presented in this section, where we show for the particular case of two clusters that there exist lines $T_M^\gamma(\alpha)$ and $T_c^\gamma(\alpha)$ such that for a given α and $T < T_M^\gamma(\alpha)$ orders the γ descendant, while for $T < T_c^\gamma(\alpha) < T_M^\gamma(\alpha)$ the free energy of the retrieved descendant is lower than the free energy of the spin-glass phase. The numerical values obtained for α_1^c agree with the rigorous bounds previously calculated. Section 5 is dedicated to conclusions and we comment on the relation with stochastic models with modulated or restricted interaction range. In the appendix we use the linked cluster theorem to evaluate the average values over the random embedded patterns.

2. The model

We consider a network of N neurons represented by Ising spin variables $\sigma_i = \pm 1$, $i = 1, \ldots, N$, where there are stored p patterns $\{\xi_i^\mu\}$, $\mu = 1, \ldots, p$, with the independent random variables ξ_i^μ taking values ±1 with equal probability. The hierarchical clusters are organized as follows [1, 4]: the N neurons are partitioned in $l = 2^r$ clusters of N_0 sites each, where r is an integer and we introduce the cluster overlaps

$$S_a^\mu = \frac{1}{N_0} \sum_i^{\{a\}} \xi_i^\mu \sigma_i \qquad a = 1, 2, \ldots, l \tag{1}$$

where the sum in equation (1) indicates that the index i runs over the sites in cluster $\{a\}$. To this partition we associate an interaction energy:

$$\mathcal{H}_{(0)} = -\frac{JN_0}{2l} \epsilon^r \sum_{a=1}^{l} \sum_\mu^{p} [S_a^\mu]^2 \tag{2}$$

where ϵ is an arbitrary positive coupling.

At the second level every two consecutive clusters are joined into a larger cluster of $2N_0$ sites and by continuing this process we have, at the kth level, 2^{r-k} clusters with $N_k = N_0 2^k$ sites each. To every partition we associate a cluster overlap $S_{a'}^\mu(k)$, $a' = 1, 2, \ldots, 2^{r-k}$ as in equation (1) and an interaction energy $\mathcal{H}_k$ as in equation (2) with coupling strength ϵ^{r-k}. The total energy is obtained by adding the $\mathcal{H}_k$ from $k = 0$ to $k = r$, with the result that

$$\mathcal{H} = -\frac{JN_0}{2l} \sum_{a,b}^{l} A_{ab}(l) \sum_\mu S_a^\mu S_b^\mu \tag{3a}$$

which can also be written as

$$\mathcal{H} = -\frac{J}{2} \sum_{a,b}^{l} A_{ab}(l) \sum_i^{\{a\}} \sum_j^{\{b\}} J_{ij} \sigma_i \sigma_j \tag{3b}$$

where the J_{ij} are given by Hebb's learning rule:

$$J_{ij} = \frac{1}{N} \sum_\mu \xi_i^\mu \xi_j^\mu . \tag{4}$$

The coefficients $A_{ab}(l)$ are the elements of a $l \times l$ matrix $\mathbf{A}$ that turns out to be of ultrametric form. The explicit expression for $\mathbf{A}(l)$ and the eigenvectors $v^\gamma(l)$ were discussed in I where we showed the following properties:

(i) They can be obtained through the recursion relations

$$\begin{aligned} v^{2\eta-1}(l) &= \begin{bmatrix} v^\eta(l/2) \\ v^\eta(l/2) \end{bmatrix} \\ v^{2\eta}(l) &= \begin{bmatrix} v^\eta(l/2) \\ -v^\eta(l/2) \end{bmatrix} v^1 = 1 \qquad \eta = 1, 2, \ldots, l/2 \end{aligned} \tag{5}$$

with the corresponding eigenvalues

$$\begin{aligned} \lambda_1(l) &= l + \epsilon\lambda_1(l/2) \\ \lambda_2(l) &= \epsilon\lambda_1(l/2) \\ \lambda_{2\eta-1}(l) &= \lambda_{2\eta}(l) = \epsilon\lambda_\eta(l/2) \qquad \eta \geqslant 2 . \end{aligned} \tag{6}$$

It follows that $v_a^\gamma = \pm 1$.

(ii) Except for $\boldsymbol{v}^1(l)$ and $\boldsymbol{v}^2(l)$, all the other eigenvectors fall into degenerate groups with decreasing eigenvalues:

$$\lambda_1 > \lambda_2 > \lambda_3 = \lambda_4 > \cdots > \lambda_{l/2+1} = \lambda_{l/2+2} = \cdots = \lambda_l . \tag{7}$$

We may write the energy in an alternative way

$$\mathcal{H} = -\frac{JN_0}{2l} \sum_\mu^p \sum_\gamma^l \lambda_\gamma \left[\sum_a^l v_a^\gamma S_a^\mu \right]^2 \tag{8}$$

and it follows that, for each embedded pattern, the energy function will be minimized by l configurations of overlaps that satisfy:

$$\operatorname{sgn}(S_a^\mu) = v_a^\gamma \qquad \gamma = 1, 2, \ldots, l. \tag{9}$$

It is interesting to point out that in the form of equation (8) the energy function is reminiscent of some 'palimpsestic' schemes formulated to store working or short-term memories [6]. The learning rule of [6] is obtained from equation (8) if we make the correspondence

$$p = 1 \qquad l = \alpha N \qquad (\xi_i v_a^\gamma)_{i \in a} = \eta_i^\gamma \qquad \eta_i^\gamma = \pm 1 \tag{10}$$

and let λ_γ be some positive, integrable function of γ that would play the role of a 'time'. The analogy stops there, but some of our equations can be compared with those in [6] with the correspondence in equation (10).

Dyson's hierarchical model is obtained by writing $J_{ij} \equiv 1$, $N_0 = 2$, $r = \ln(N)/\ln(2)$ and it simulates decaying power law interactions $|i-j|^{-(1+\sigma)}$, with the range parameter $\sigma = \ln(\epsilon)/\ln(2)$, then the limit of long- (short-)-range interactions corresponds to $\epsilon \to 0$ ($\epsilon \to \infty$). We analyse these two limits.

When $\epsilon = 0$ the present model reduces to Hopfield's model with long-range, uniform interactions. In this limit only the largest eigenvalue $\lambda_1 = l$ differs from zero in equation (6) and the only retrieval states are the 'ancestors' or 'pure' states parallel to the encoded patterns in every block together with the completely antiparallel states, that means two states by pattern. To discuss the opposite limit $\epsilon \to \infty$ we should normalize $J = \epsilon^{-r}$ in equation (2) or equation (3) to keep the interaction energy finite and then take the limit. In this case the only interaction that survives is within a single block and the matrix **A** becomes diagonal, hence the system splits into l independent neural networks with N_0 neurons each. The encoded patterns also split in l independent cluster patterns, and each cluster state can be either parallel or antiparallel to its own pattern, then we have in total 2^l possible states for each global pattern. We conclude that by increasing ϵ and by shortening the interaction range we favour the appearance of 'mixed' states, with the clusters partially parallel and partially antiparallel to the encoded patterns. These are the 'descendants' in equation (9) with $\gamma \geqslant 2$.

3. Signal-to-noise analysis

The condition for the stability of a given configuration $\{\sigma\}$ of the network is that the neuron σ_i, for the site $i \in a$, be aligned with the local field h_i at the same site. This gives from equation ($3b$)

$$(\sigma_i h_i)_{i\in a} = \sigma_i \sum_b A_{ab} \sum_j^{\{b\}} J_{ij}\sigma_j > 0. \tag{11}$$

According to our previous results [1, 4] for finite values of p, for each embedded pattern $\{\xi_\mu\}$ the system retrieves a family of patterns $\{\eta^{\mu\gamma}\}$, $\gamma = 1, 2, \ldots, l$, where

$$\eta_i^{\mu\gamma} \equiv v_a^\gamma \xi_i^\mu \qquad i \in a \tag{12}$$

and it is clear that perfect alignment along these patterns will minimize the energy in equation (8). However, when p is extensive the retrieval of a given pattern may be hindered by the interference with all the others [7], then we determine the maximum value of p such that equation (11) holds true for every site when $\sigma_i = \eta_i^{\mu\gamma}$ in equation (12). We obtain, by separating the term with $\nu = \mu$,

$$\lambda_\gamma + \sum_b A_{ab} r_b > 0 \tag{13}$$

where

$$r_b = v_a^\lambda v_b^\lambda \frac{1}{N_0} \sum_j^{\{b\}} \sum_{\nu\neq\mu} \xi_i^\nu \xi_i^\mu \xi_j^\nu \xi_j^\mu \tag{14}$$

is a Gaussian noise with zero mean and variance $\langle r_b^2\rangle = \alpha l$. From equation (13) we obtain the probability of having the neuron at site $i \in a$ aligned with its local field [7], when the whole network is in the configuration $\{\sigma\} = \{\eta^{\mu\gamma}\}$

$$P_i^{\mu\gamma} = \pi^{-l/2} \int_{-\infty}^{\infty} \prod_b \mathrm{d}\chi_b \mathrm{e}^{-\Sigma_b \chi_b^2} \theta\left(\lambda_\gamma + (2\alpha l)^{1/2} \sum_b A_{ab}\chi_b\right). \tag{15}$$

It is straightforward to perform the integral in equation (15) by means of the identity:

$$1 = \int_{-\infty}^{\infty} \mathrm{d}\xi \int_{-\infty}^{\infty} \frac{\mathrm{d}\rho}{2\pi} \mathrm{e}^{\mathrm{i}\rho(\xi - \Sigma_b A_{ab}\chi_b)} \tag{16}$$

with the result

$$P_i^{\mu\gamma} = \frac{1}{2}\left\{1 + \Phi\left[\lambda_\gamma \Big/ \left(2\alpha l \sum_b A_{ab}^2\right)^{1/2}\right]\right\} \tag{17}$$

where

$$\Phi(x) = \frac{2}{\sqrt{\pi}} \int_0^x \mathrm{d}z\, \mathrm{e}^{-z^2}. \tag{18}$$

In this analysis the critical value of α is obtained by requiring that every site should be perfectly aligned with its local field, which means from equation (17):

$$\prod_i^N P_i^{\mu\gamma} = 2^{-N}\left\{1 + \Phi\left[\lambda_\gamma \Big/ \left(2\alpha l \sum_b A_{ab}^2\right)^{1/2}\right]\right\}^N \approx 1. \tag{19}$$

In the limit $N \to \infty$, $\alpha \ll 1$, we obtain from equation (19) by using the asymptotic expression for the probability integral [8],

$$\alpha_\gamma^{\mathrm{SN}} = \frac{1}{2\ln N}\frac{\lambda_\gamma^2}{l\Sigma_b A_{ab}^2} \tag{20}$$

which can be written as a ratio:

$$\alpha_\gamma^{\mathrm{SN}}/\alpha_{\mathrm{H}}^{\mathrm{SN}} = \lambda_\gamma^2 \Big/ \left(\sum_{\delta=1}^{l} \lambda_\delta^2\right) \tag{21}$$

where we used the signal-to-noise result for Hopfield's model, $\alpha_{\mathrm{H}}^{\mathrm{SN}} = (2\ln N)^{-1}$.

4. Mean field theory

Here we study the statistical mechanics of the Hamiltonian in equations (3) by following the method of Amit *et al* [5]. We assume from the start that a finite number s of patterns condense macroscopically, and these are treated by introducing the thermal averages of the cluster overlaps in equation (1) as order parameters, while the other memories $\nu = s+1, \ldots, p$ will form the order parameter in the spin–glass phase.

We use the replica method to write the free energy per site:

$$f = -\lim_{N_0\to\infty}\lim_{nl\to 0}\frac{1}{\beta N_0 nl}(Z_n - 1) \tag{22}$$

where the replicated partition function is obtained from equation (3*a*):

$$Z_n = \langle Z^n \rangle = \left\langle \mathrm{Tr}_{\{\sigma_\rho\}} \exp\left[\frac{\beta J N_0}{2l}\sum_{a,b} A_{ab} \sum_{\rho=1}^{n}\sum_{\mu=1}^{p} S_a^{\mu\rho} S_b^{\mu\rho}\right]\right\rangle \tag{23}$$

and we indicate by a bracket the quenched average over the ξs. We also have in equation (23) the replicated overlaps

$$S_a^{\mu\rho} = \frac{1}{N_0}\sum_i^{\{a\}} \xi_i^\mu \sigma_i^\rho \tag{24}$$

with the replica index $\rho = 1, 2, \ldots, n$.

For the ‘condensed’ memories, $\mu = 1, 2, \ldots, s$, the standard procedure of Gaussian integration gives [5]

$$Z_n = \left[|\mathbf{A}| \left(\frac{\beta J N_0}{2\pi l} \right)^l \right]^{ms/2} \int \prod_a \prod_{\mu\rho} \mathrm{d} m_a^{\mu\rho}$$

$$\times \exp \left(-\frac{\beta J N_0}{2l} \Sigma_{a,b} A_{ab} \Sigma_{\rho=1}^{n} \Sigma_{\mu=1}^{s} m_a^{\mu\rho} m_b^{\mu\rho} \right)$$

$$\times \mathrm{Tr}_{\{\sigma_\rho\}} \left\{ \exp \left(-\frac{\beta J N_0}{l} \Sigma_{a,b} A_{ab} \Sigma_{\mu=1}^{s} \Sigma_{\rho=1}^{n} m_a^{\mu\rho} S_b^{\mu\rho} \right) \Lambda^{p-s} \right\} \qquad (25)$$

where

$$\Lambda = \left\langle \exp \left(\frac{\beta J N_0}{2l} \Sigma_{a,b}^{l} A_{ab} \Sigma_{\rho}^{n} S_a^{\nu\rho} S_b^{\nu\rho} \right) \right\rangle_{\{\xi'\}} . \qquad (26)$$

In the derivation of equations (25) and (26) we exploited the statistical independence of the ξ_i^ν for different νs and the average is over only one set $\{\xi_i^\nu\}$ for fixed ν. The evaluation of the quantity Λ in equation (26) in the thermodynamic limit is presented in the appendix, with the result that

$$\Lambda = \exp \left\{ \frac{1}{2} \sum_{k=2}^{\infty} \frac{1}{k} \left(\frac{\beta J}{l} \right)^k \sum_{a_1 \ldots a_k}^{l} A_{a_1 a_2} A_{a_2 a_3} \ldots A_{a_k a_1} \mathrm{Tr}_{\{\rho\}} [\mathbf{Q}_{a_1} \mathbf{Q}_{a_2} \ldots \mathbf{Q}_{a_k}] \right\} \qquad (27)$$

where we introduced the $n \times n$ matrices $\mathbf{Q}_a$ with elements

$$Q_a^{\rho\rho'} = \frac{1}{N_0} \sum_i^{\{a\}} \sigma_i^\rho \sigma_i^{\rho'} \qquad \rho \neq \rho'$$

$$Q_a^{\rho\rho} = 1. \qquad (28)$$

We call $q_a^{\rho\rho'}$ the spin–glass order parameter conjugate to the operator $Q_a^{\rho\rho'}$ in equation (28) and we introduce it in equation (25) by using the identity:

$$\prod_a \prod_{\rho \neq \rho'} \left[\frac{\alpha \beta^2 N_0}{2} \int_{-i\infty}^{i\infty} \frac{\mathrm{d} r_a^{\rho\rho'}}{2\pi \mathrm{i}} \int_{-\infty}^{\infty} \mathrm{d} q_a^{\rho\rho'} \exp \left(\frac{\alpha \beta^2 N_0}{2} r_a^{\rho\rho'} (q_a^{\rho\rho'} - Q_a^{\rho\rho'} \right) \right] = 1 \qquad (29)$$

to obtain the result

$$Z_n = C_n \int_{-\infty}^{\infty} \prod_a \left\{ \prod_{\mu\rho} \mathrm{d} m_a^{\mu\rho} \prod_{\rho < \rho'} \mathrm{d} r_a^{\rho\rho'} \, \mathrm{d} q_a^{\rho\rho'} \right\} \exp(-\beta N n f(m, q, r)) \qquad (30)$$

where we have grouped together the multiplicative constants in C_n and we have in the exponent

$$f(m, q, r) = \frac{J}{2l^2 n} \sum_{\mu\rho} \sum_{a,b} A_{ab} m_a^{\mu\rho} m_b^{\mu\rho} + \frac{1}{2ln} \beta\alpha \sum_{\rho \neq \rho'} \sum_a r_a^{\rho\rho'} q_a^{\rho\rho'}$$

$$- \frac{\alpha}{2\beta n} \sum_{k=2}^{\infty} \frac{1}{k} \left(\frac{\beta J}{l} \right)^k \sum_{a_1 \ldots a_k} A_{a_1 a_2} \ldots A_{a_k a_1}$$

$$\mathrm{Tr}_{\{\rho\}} [\mathbf{q}_{a_1} \mathbf{q}_{a_2} \ldots \mathbf{q}_{a_k}] - \frac{1}{\beta N n} \ln \left\{ \mathrm{Tr}_{\{\sigma\}} \exp^{-\beta H_{\mathrm{eff}}} \right\} \qquad (31)$$

with

$$H_{\text{eff}} = \frac{JN_0}{l}\sum_{\mu\rho}\sum_{ab} A_{ab} m_a^{\mu\rho} S_b^{\mu\rho} + \frac{\alpha\beta N_0}{2}\sum_{a}\sum_{\rho\neq\rho'} r_a^{\rho\rho'} Q_a^{\rho\rho'}. \tag{32}$$

The $n \times n$ matrix $\mathbf{q}_a$ in equation (31) has elements $q_a^{\rho\rho'}$ for $\rho \neq \rho'$, $q_a^{\rho\rho} = 1$, which are the order parameters conjugate to the operators $\mathbf{Q}_a$ in equation (28).

In the thermodynamic limit we evaluate the integral in equation (30) by using the saddle-point method and to proceed further with the calculation we assume replica symmetric solutions, i.e.

$$m_a^{\mu\rho} = m_a^{\mu} \qquad q_a^{\rho\rho'} = q_a \qquad r_a^{\rho\rho'} = r_a. \tag{33}$$

Even within the assumption of replica symmetry for the solutions of the saddle-point equations, we cannot find a closed expression for the free energy in equation (31) if the $\mathbf{q}_a$ matrices differ for different clusters. However, we are mainly interested here in the retrieval of the 'pure' memories and its 'descendants' given in equation (12) and discussed in [4]

$$\begin{aligned} m_a^{\mu} &= \delta_{\mu_1} m_a \\ m_a &= m_\gamma v_a^\gamma \qquad \gamma = 1, 2, \ldots, l \end{aligned} \tag{34}$$

where v^γ are the eigenvectors of $\mathbf{A}$ in equation (15). The solutions in equation (34) will minimize the energy according to the criterion in equation (9), and they are homogeneous in magnitude. Taking into account that q_a is the spin–glass order parameter within the cluster 'a' and a positive definite quantity, it is then natural to think that it will be homogeneous for the retrieval solutions. Hence we restrict ourselves in the following to saddle-point solutions that satisfy equation (34) together with:

$$q_a = q \qquad r_a = r. \tag{35}$$

The assumption of homogeneity does not hold for solutions that mix eigenvectors with different eigenvalues, like

$$\begin{aligned} m_a &= m_\gamma v_a^\gamma \qquad \text{if } 1 \leqslant a < l/2 \\ m_a &= m_{\gamma'} v_a^{\gamma'} \qquad \text{if } l/2 \leqslant a \leqslant l \end{aligned}$$

for $\lambda_\gamma \neq \lambda_{\gamma'}$.

It was shown in I that these mixed solutions exist for $l \geqslant 8$ and their number grows exponentially for sufficiently large values of l.

The free energy is obtained by taking the limit $n \to 0$ of $f(m, q, r)$ in equation (31) at the saddle-point in equation (34) and equation (35):

$$\begin{aligned} f_{\text{SP}}(\gamma) = {} & \frac{1}{2\beta_\gamma} m_\gamma^2 + \frac{1}{2}(1-q) r\alpha\beta \\ & + \frac{\alpha}{2\beta}\sum_{\delta=1}^{l}\left\{\ln\left[1 - \frac{\beta}{\beta_\delta}(1-q)\right] - \frac{\beta}{\beta_\delta} q \left[1 - \frac{\beta}{\beta_\delta}(1-q)\right]^{-1}\right\} \\ & - \frac{1}{\beta}\int_{-\infty}^{\infty}\frac{\mathrm{d}z}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}z^2} \ln\left[2\cosh\left(\frac{\beta}{\beta_\gamma} m_\gamma + z\beta\sqrt{\alpha r}\right)\right] \end{aligned} \tag{36}$$

where the eigenvalues λ_γ are given in equation (6) and $\beta_\gamma = l/J\lambda_\gamma$. We have also used the property that the matrix $\mathbf{q}$ has one eigenvalue equal to $[1 + (n-1)q]$, and $(n-1)$ degenerate eigenvalues equal to $(1-q)$.

The self-averaging process involved in the derivation of equation (36) when the cluster size $N_0 \to \infty$ is trivial because we have only one memory present in equation (34).

The order parameters m_γ, q, r in equations (34) and (35) are given by the solutions of the saddle-point equations:

$$m_\gamma = \int_{-\infty}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}z^2} \tanh\left(\frac{\beta}{\beta_\gamma} m_\gamma + z\beta\sqrt{\alpha r}\right) \tag{37}$$

$$r = \sum_{\delta=1}^{l} \frac{q}{[\beta_\delta - \beta(1-q)]^2} \tag{38}$$

$$q = \int_{-\infty}^{\infty} \frac{\mathrm{d}z}{\sqrt{2\pi}} \mathrm{e}^{-\frac{1}{2}z^2} \tanh^2\left(\frac{\beta}{\beta_\gamma} m_\gamma + z\beta\sqrt{\alpha r}\right). \tag{39}$$

4.1. Storage capacity at $T = 0$: upper and lower bounds

By taking the limit $\beta \to \infty$ in equations (37)–(39) we obtain the equations

$$m_\gamma = \Phi\left(\frac{1}{\sqrt{2\alpha r_\gamma}} \frac{1}{\beta_\gamma} m_\gamma\right) \tag{40}$$

$$r_\gamma = \sum_{\delta=1}^{l} [\beta_\delta - C_\gamma]^{-2} \tag{41}$$

$$C_\gamma = \lim_{\beta\to\infty} \beta(1-q) = \left(\frac{2}{\pi\alpha r_\gamma}\right)^{1/2} \exp\left\{-\frac{1}{2\alpha r_\gamma}\left(\frac{m_\gamma}{\beta_\gamma}\right)^2\right\} \tag{42}$$

where the function $\Phi(x)$ is given in equation (18). We are looking for the values of α that allow for solutions $m_\gamma \neq 0$, then equations (40)–(42) can be combined into

$$\alpha = \frac{1}{2}\left\{\sum_{\delta=1}^{l}\left[\frac{\lambda_\gamma}{\lambda_\delta}\frac{\Phi_{(U)}}{U} - \frac{2}{\sqrt{\pi}}\mathrm{e}^{-U^2}\right]^{-2}\right\}^{-1} = \chi_\gamma(U) \tag{43}$$

where $U = m_\gamma(\beta_\gamma\sqrt{2\alpha r})^{-1}$. The critical value of the storage capacity $\alpha_\gamma^{\mathrm{c}}$ is determined from the condition that for $\alpha > \alpha_\gamma^{\mathrm{c}}$ there is no possible retrieval of the γ descendant in equation (34), then $\alpha_\gamma^{\mathrm{c}}$ is given by the maximum value of the function at the right-hand side of equation (43). From equation (7) we can derive the inequality:

$$\frac{\lambda_1}{\lambda_\delta}\Phi_{(U)} \geqslant \Phi_{(U)} > \frac{2}{\sqrt{\pi}} U \mathrm{e}^{-U^2} \tag{44}$$

and it follows from equation (44) that α_1^c in equation (43) has the lower and upper bounds

$$\left(\lambda_1^2/\sum_\delta \lambda_\delta^2\right)\max\{H_{(U)}\} \leqslant 2\alpha_1^c \leqslant \max\{H_{(U)}\} \tag{45}$$

where we defined

$$H_{(U)} = \left[\frac{\Phi_{(U)}}{U} - \frac{2}{\sqrt{\pi}}\mathrm{e}^{-U^2}\right]^2 . \tag{46}$$

The equality in equation (45) only holds in the Hopfield limit $l = 1$ where it coincides with the result in [5]

$$2\alpha_{\mathrm{H}}^c = \max\{H_{(U)}\}. \tag{47}$$

Equations (45) and (47) can be combined as follows.

$$\frac{\alpha_1^{\mathrm{SN}}}{\alpha_{\mathrm{H}}^{\mathrm{SN}}} \leqslant \frac{\alpha_1^c}{\alpha_{\mathrm{H}}^c} \leqslant 1 \tag{48}$$

where we used equation (21) to prove that the ratio of the critical storage capacities calculated by the signal-to-noise method gives a lower bound of $\alpha_1^c/\alpha_{\mathrm{H}}^c$. Similar bounds cannot be derived for the 'descendants' with $2 \leqslant \gamma \leqslant l$ because $(\lambda_\gamma/\lambda_\delta) < 1$ for $\delta < \gamma$ and equation (44) no longer holds, but the numerical results discussed later indicate that $\alpha_{\mathrm{H}}^c > \alpha_1^c > \alpha_2^c > \cdots$.

We present numerical results for the critical storage capacities as a function of the parameter ϵ in equation (6) for the two cases $l = 2$ and $l = 8$.

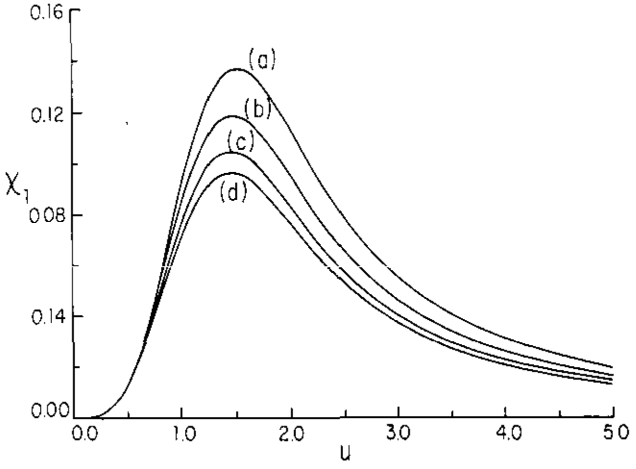

Figure 1. Plot of $\chi_1(U)$ in equation (43) for two clusters and different values of ϵ. The index $\gamma = 1$ corresponds to the 'ancestor'. (*a*), (*b*), (*c*) and (*d*) correspond to $\epsilon = 0.1, 1.6, 3.1, 4.6$.

In the case $l = 2$ the system is divided into two clusters and we have from equation (6)

$$\lambda_1 = 2 + \epsilon \qquad \lambda_2 = \epsilon. \tag{49}$$

The function on the right-hand side of equation (43) is plotted in figure 1 for $\gamma = 1$ and in figure 2 for $\gamma = 2$, for several values of ϵ. The function in figure 1 has a single maximum and the determination of α_1^c is straightforward. The function in figure 2, however, presents two maxima and in order to decide which one determines α_2^c we have to look at the stability conditions. The non-vanishing solution of equation (40) is stable if the inequality

$$1 - \frac{1}{\beta_1} C_\gamma > 0 \tag{50}$$

holds at $T = 0$, where C_γ is given in equation (42), then equation (50) can also be written as

$$\frac{2}{\sqrt{\pi}} \mathrm{e}^{-U^2} < \frac{\lambda_\gamma}{\lambda_1} \frac{1}{U}. \tag{51}$$

Now, the two maxima in figure 2 are separated by a minimum at U^* such that

$$\frac{\lambda_2}{\lambda_1} \frac{\Phi_{U^*}}{U^*} = \frac{2}{\sqrt{\pi}} \mathrm{e}^{-U^{*2}} \tag{52}$$

while for all $U > U^*$ the inequality in equation (51) will hold. We conclude that α_2^c is determined by the maximum value of the function that occurs for $U > U^*$. We show in figure 3 the values for $\alpha_1^c(\epsilon)$ and $\alpha_2^c(\epsilon)$ when $l = 2$, and a similar analysis was performed for $l = 8$ with the results displayed in figure 4.

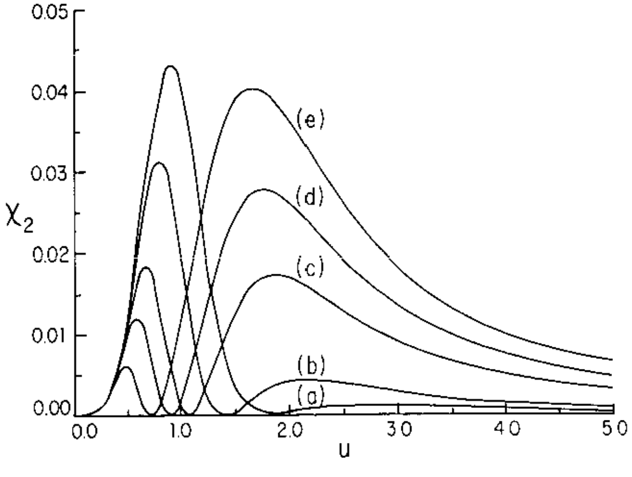

Figure 2. Plot of $\chi_2(U)$ in equation (43) for two clusters and different values of ϵ. The index $\gamma = 2$ corresponds to the 'descendant'. (*a*), (*b*), (*c*), (*d*) and (*e*) correspond to $\epsilon = 0.1, 0.6, 1.6, 3.1, 4.6$.

It is interesting to compare our results with those obtained in the model of Feigelman and Ioffe [9]. They considered a learning algorithm that stores a 'basic' pattern ξ_i together with K 'satellites' through the correlated images:

$$\xi_i^{(p)} = \xi_i(1 - 2\beta_i^{(p)}) \tag{53}$$

where the β_i^p, $p = 1, \ldots, K$ are also random variables. Equation (53) should be compared with the expression for the patterns stored in our model given in equation (12), with the conclusion that the main difference between both models is that our 'satellite' or 'descendant' variables $v_i^\gamma = v_a^\gamma$ if $i \in a$, are not random but determined by the eigenvectors in equation (6). In their case the number K of 'satellites' is also extensive, while we have a finite number of descendants and an extensive number of basic patterns or 'ancestors'. Some results are also analogous to those obtained by us in I, as the existence of a second-order transition for the basic pattern and a first-order transition at lower temperature for the satellites. Their phase diagram, however, is different from ours as discussed later.

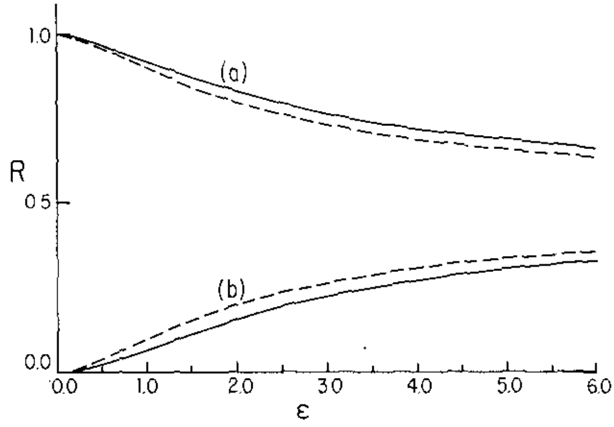

Figure 3. Plot of the critical storage capacity ratios $\alpha_\gamma^c(\epsilon)/\alpha_H^c$ (full curve) and the corresponding signal-to-noise quantity $\alpha_\gamma^{SN}(\epsilon)/\alpha_H^c$ (broken curve) for two clusters: (*a*) $\gamma = 1$, ancestor; and (*b*) $\gamma = 2$, descendant.

4.2. Phase diagram in the α–T plane

We recall first the results on the $\alpha = 0$ line that were derived in I: for a number $l = 2^r$ of clusters and coming from high temperatures there is first a second-order transition at the inverse temperature $\beta_1 = l/J\lambda_1$ where the 'pure' pattern or 'ancestor' in equation (34) orders continuously:

$$m_a^\mu = \delta_{\mu 1} m_a$$

$$m_a = m_1 \simeq \left(\frac{\beta}{\beta_1} - 1\right)^{1/2} .$$

For lower temperatures each group of ' descendants' in equation (6) with eigenvalue λ_γ, $2^{s-2} + 1 \leqslant \gamma \leqslant 2^{s-1}$ orders at $\beta = \beta_s^* > \beta_s$ with a discontinuity $m_s^* \neq 0$ in the order parameter.

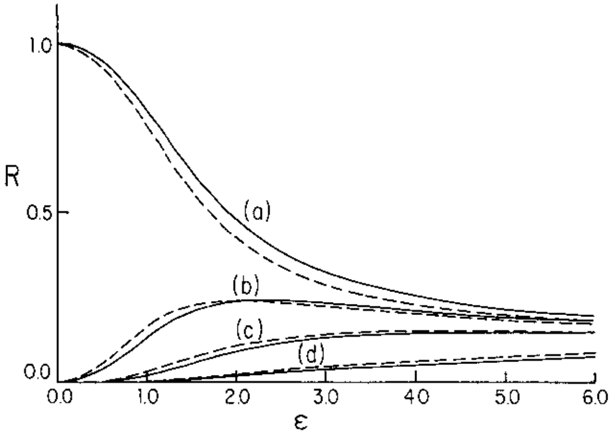

Figure 4. Plot of the critical storage capacity ratios $\alpha^c_\gamma(\epsilon)/\alpha^c_H$ (full curve) and the corresponding signal-to-noise quantity $\alpha^{SN}_\gamma(\epsilon)/\alpha^c_H$ (broken curve) for eight clusters. (*a*), (*b*), (*c*) and (*d*) correspond to $\gamma = 1, 2, 3, 4$, with $\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4$. The index $\gamma = 1$ corresponds to the ancestor.

For α finite and high temperatures we only have the paramagnetic solution $m_\gamma = 0$, $q = 0$ of the saddle-point equations (37)–(39). By lowering the temperature we hit first a continuous transition to a pure spin–glass phase with $m_\gamma = 0$. The transition temperature is found by expanding equation (38) and (39) for small values of q and is the solution of the equation:

$$\frac{1}{\alpha} = \sum_\delta \left[\frac{\beta_\delta}{\beta_G} - 1\right]^{-2} \qquad \beta_G \leqslant \beta_1 \tag{54}$$

where the last condition is dictated by stability. For $\alpha \to 0$ we obtain the solution:

$$\beta_G \approx \beta_1(1 - \sqrt{\alpha}) \tag{55}$$

similar to [4].

To facilitate the comparison with Amit *et al* we use here the same notation as in [5] to describe the numerical solution of equations (37)–(39) for the particular case $l = 2$, when we have only the 'ancestor' ($\gamma = 1$) and one 'descendant' ($\gamma = 2$). On lowering the temperature and coming from the spin–glass phase one reaches the line $T^{(1)}_M(\alpha)$ at which the 'ancestor' orders with a discontinuity $m_1 \neq 0$. This transition becomes continuous only at the T-axis. In this region the retrieval solution is locally stable but the free energy $f_{SP}(1)$ calculated from equation (36) is higher than the free energy $f_{SP}(SG)$ for the spin–glass solution. The line $T^{(1)}_M(\alpha)$ intersects the α-axis at the critical α^1_c. From equation (48) we obtain the lower bounds 0.110, 0.093 and 0.084 for α^1_c when λ_2/λ_1 equals 0.5, 0.7 and 0.8, respectively, what is consistent with the numerical results in figures 5, 6 and 7.

There exists a second line $T^{(1)}_c(\alpha) < T^{(1)}_M(\alpha)$ at which $f_{SP}(1) = f_{SP}(SG)$ while for $T < T^{(1)}_c(\alpha)$ we have $f_{SP}(1) < f_{SP}(SG)$ and the retrieval phase for the ancestor is a global minimum. Up until now the phase diagram looks the same as in [5], but for still lower temperatures one hits a third line $T^{(2)}_M(\alpha)$ at which the 'descendant'

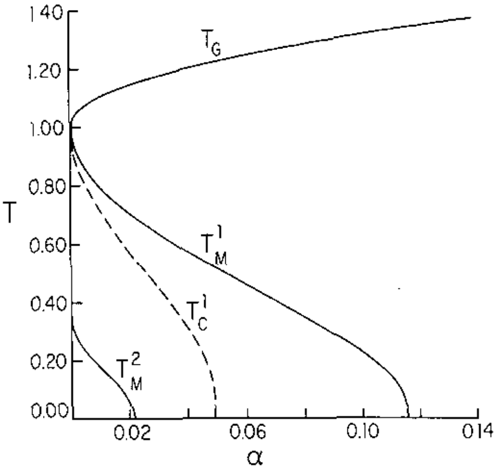

Figure 5. Plot of critical temperatures as a function of α for two clusters and $\epsilon = 2$. T_G is the critical temperature of the spin-glass phase. For $T_c^1(\alpha) < T < T_M^1(\alpha)$ the ancestor ($\gamma = 1$) orders, but it becomes a global minimum for $T < T_c^1(\alpha)$. For $T < T_M^2(\alpha)$ the descendant ($\gamma = 2$) orders but its free energy is always higher than the spin-glass phase. We fixed the scale $2/J\lambda_1 = 1$.

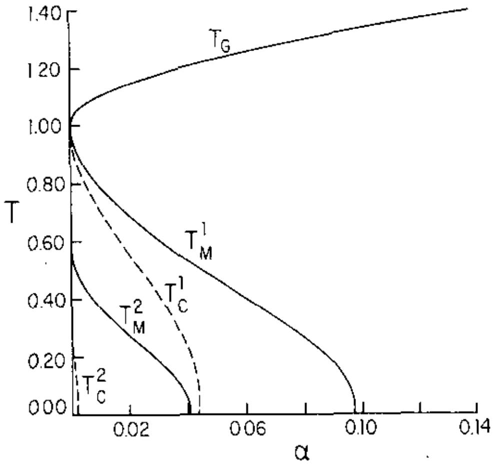

Figure 6. Same as in figure 5 for $\epsilon = 4.6$. We can see the appearance of a fourth line $T_c^2(\alpha)$ that was not present in figure 5. For $T < T_c^2(\alpha)$ the free energy for the descendant is below the free energy for the spin-glass phase.

solution with $\gamma = 2$ in equations (37)–(39) orders discontinuously with $m_2 \neq 0$. Below this line the new solution is locally stable with $f_{SP}(2)\rangle f_{SP}(SG)\rangle f_{SP}(1)$, and the intersection of $T_M^{(2)}(\alpha)$ with the α-axis determines the value α_c^2.

The existence of a fourth line $T_c^{(2)}(\alpha)$, such that for $T < T_c^{(2)}(\alpha)$ one has $f_{SP}(SG) > f_{SP}(2) > f_{SP}(1)$, depends on the relative value λ_2/λ_1. From the numerical results in figures 5, 6 and 7 we conclude that this phase does not exist for $\lambda_2/\lambda_1 \leqslant 0.6$ or $\epsilon < 3$. This is consistent with the discussion at the end of section 2, where we showed that the existence of the 'descendants' is favoured by large values of ϵ.

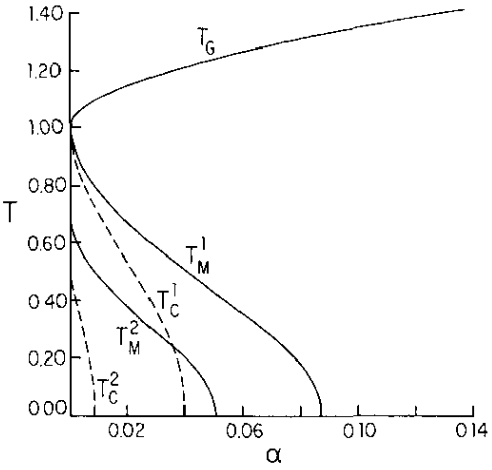

Figure 7. Same as in figure 6 for $\epsilon = 8$.

5. Conclusions

We have presented a detailed analysis of the retrieval and storage properties of a model for neural networks introduced previously by us where the neurons, and not the patterns, are organized in hierarchical clusters [1, 4]. Our results indicate that the space organization of the neurons induces an organization of the retrieved memories, as for each embedded memory or 'ancestor' the system is able to retrieve a family of 'descendants' that differ from each other and from the ancestor in the sign of the cluster overlaps.

Although reminiscent of cluster models discussed previously by other authors [3], the model studied here presents the great advantage of its tractability, that allows for a detailed investigation of its efficiency for retrieving and storing information.

Understanding the properties of neural network models with modulated or restricted range connections is very important for optimizing the hardware realization of attractor neural networks. These realizations suffer from severe problems, unless the assumption of full-range connectivity is broken. On the other hand, strictly short-range networks are biologically unrealistic. The consideration of this problem led Coolen [10] and Noest [11] to study models of neural networks with spatial structure by means of stochastic equations. Although their method is different from ours, they also describe the evolution of the system by using cluster overlaps similar to our equation (1).

In [11] it is pointed out that the occurrence of domains is the distinguishing feature of models with restricted range connections. In an analogous way we show in the present work that the retrieval of 'descendants' with mixed alignments in different clusters originates in the modulation of the interaction, as was discussed at the end of section 2.

Acknowledgments

Supported in part by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Financiadora de Estudos e Projetos (FINEP).

Appendix

Here we present a derivation of equation (27) in the text. We start by writing from equation (26)

$$\Lambda = \left\langle \exp\left\{ \frac{\beta J}{2N} \sum_{i \neq j} T_{ij} \xi_i \xi_j \right\} \right\rangle_{\{\xi\}} \tag{A1}$$

where $N = lN_0$ and we have from equation (24)

$$T_{ij} = A_{ab} \sum_{\rho}^{n} \sigma_i^{\rho} \sigma_j^{\rho} \qquad \text{if } i \in a, j \in b$$
$$\text{or if } i \in b, j \in a. \tag{A2}$$

The bracket in equation (A1) indicates an average over the independent variables ξ_i at each site that take values ±1 with equal probability, then we can use the cumulant expansion to write:

$$\ln \Lambda = \sum_{k=1}^{\infty} \frac{1}{k!} \left(\frac{\beta J}{2N} \right)^k \left\langle \left({\sum_{ij}}' T_{ij} \xi_i \xi_j \right)^k \right\rangle_{\mathrm{c}} \tag{A3}$$

where ${\sum}'_{ijk\ldots}$ indicates $\sum_{i \neq j \neq k \ldots}$ and $\langle \ldots \rangle_{\mathrm{c}}$ means a cumulant average

$$\left\langle {\sum_{ij}}' T_{ij} \xi_i \xi_j \right\rangle_{\mathrm{c}} = \sum_{i \neq j} T_{ij} \langle \xi_i \xi_j \rangle = 0$$

$$\left\langle \left({\sum_{ij}}' T_{ij} \xi_i \xi_j \right)^2 \right\rangle_{\mathrm{c}} = {\sum_{ij}}' {\sum_{kl}}' T_{ij} T_{kl} [\langle \xi_i \xi_j \xi_k \xi_l \rangle - \langle \xi_i \xi_j \rangle \langle \xi_k \xi_l \rangle] \tag{A4}$$

and so on.

The cumulant expansion ensures that only sites in connected clusters will contribute to the sums. In addition, the factor N^{-k} in front of the kth average ensures that the only non-vanishing contribution in the thermodynamic limit comes from terms that involve the largest number of independent sums. Then we have from equation (A3) by taking into account the weights for each average

$$\ln \Lambda = \left(\frac{\beta J}{2N} \right)^2 \frac{2}{2!} {\sum_{ij}}' T_{ij}^2 + \left(\frac{\beta J}{2N} \right)^3 \frac{8}{3!} {\sum_{ijk}}' T_{ij} T_{jk} T_{kl}$$
$$+ \left(\frac{\beta J}{2N} \right)^4 \frac{48}{4!} \left\{ {\sum_{ijkl}}' T_{ij} T_{jk} T_{kl} T_{li} + 3 {\sum_{ijk}}' T_{ij} T_{jk} T_{kl}^2 + {\sum_{ij}}' T_{ij}^4 \right\} + \cdots \tag{A5}$$

but the last two sums in equation (A5) are $\mathrm{O}(N^3)$ and $\mathrm{O}(N^2)$ respectively, and they can be neglected. By continuing this process it can be seen that for every order in perturbation theory there corresponds one dominant contribution and one obtains

$$\ln \Lambda = \frac{1}{2}\sum_{k=2}^{\infty}\frac{1}{k}\left(\frac{\beta J}{lN_0}\right)^k {\sum_{i_1 i_2 \ldots i_k}}' T_{i_1 i_2} T_{i_2 i_3} \ldots T_{i_k i_1}. \qquad \text{(A6)}$$

In the thermodynamic limit one may consider unrestricted sums in equation (A6), which by using equation (A2) may be rewritten as

$$\begin{aligned} {\sum_{i_1 i_2 \ldots i_k}}' T_{i_1 i_2} T_{i_2 i_3} \ldots T_{i_k i_1} &= \sum_{a_1}^{l}\sum_{i_1}^{\{a_1\}}\sum_{a_2}^{l}\sum_{i_2}^{\{a_2\}}\cdots\sum_{a_k}^{l}\sum_{i_k}^{\{a_k\}} A_{a_1 a_2} A_{a_2 a_3} \ldots A_{a_k a_1} \\ &\quad \times \sum_{\rho_1 \rho_2 \ldots \rho_k}^{n} \sigma_{i_1}^{\rho_1}\sigma_{i_2}^{\rho_1}\sigma_{i_2}^{\rho_2}\sigma_{i_3}^{\rho_2}\ldots\sigma_{i_k}^{\rho_k}\sigma_{i_1}^{\rho_k} \end{aligned} \qquad \text{(A7)}$$

and the result in equation (27) with the definition in equation (28) is obtained by introducing equation (A7) into equation (A6).

References

[1] Pires Idiart M A and Theumann A 1990 *Neural Networks and Spin-Glasses, Proc. STATPHYS 17 Workshop (Porto Alegre, Brazil, 1989)* ed W K Theumann and R Köberle (Singapore: World Scientific)

[2] Dyson F J 1969 *Commun. Math. Phys.* **12** 91

[3] Dotsenko V S 1985 *J. Phys. C: Solid State Phys.* **18** L1017

[4] Pires Idiart M A and Theumann A 1991 *J. Phys. A: Math. Gen.* **24** L649–58

[5] Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys.* **173** 30

[6] Mézard M, Nadal J P and Toulouse G 1986 *J. Physique* **47** 1457
Viana L, 1988 *J. Physique* **49** 167

[7] Lautrup B 1988 *Preprint* NBI-HE-88-06, Niels Bohr Institute

[8] Gradshteyn S and Ryzhik I M 1965 *Table of Integrals, Series and Products* (New York: Academic)

[9] Feigelman M V and Ioffe L B 1987 *Int. J. Mod. Phys.* B **1** 51

[10] Coolen A C C 1991 *Statistical Mechanics of Neural Networks* ed L Garrido (Berlin: Springer)

[11] Noest A J 1989 *Phys. Rev. Lett.* **63** 1739