Domains of attraction of neural networks at finite temperature

1991 J. Phys. A: Math. Gen. 24 1103

(http://iopscience.iop.org/0305-4470/24/5/024)

J. Phys. A: Math. Gen. 24 (1991) 1103–1117. Printed in the UK

Domains of attraction of neural networks at finite temperature*

G Nardulli† and G Pasquariello‡

† Dipartimento di Fisica dell'Universita' di Bari, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Bari, Italy

‡ Istituto Elaborazione Segnali e Immagini-Consiglio Nazionale delle Ricerche, Bari, Italy

Received 11 June 1990, in final form 28 November 1990

Abstract. We consider neural networks trained by the symmetric Edinburgh algorithm at finite temperature. We show that, while the introduction of the thermal noise affects the dynamics in a computable way, it does not change the domains of attraction of stored patterns.

1. Introduction

Neural networks are systems consisting of a large number N of interconnected neurons, where a set of $P=\alpha N$ patterns $\{\xi_i^\mu\}$ is stored ($\mu=1,\ldots,P$; $i=1,\ldots,N$); under suitable hypotheses they behave as content-addressable memories as the stored patterns can be dynamically retrieved from an initial configuration containing some distortion.

Neural networks have been intensively studied in the last few years using statistical physics methods (Hopfield 1982, Amit *et al* 1987, Mezard *et al* 1987). In these models the state of the neuron is described by an Ising variable $S_i=\pm 1$ ($S_i=+1$ when the ith neuron is active, $S_i=-1$ when it is quiescient) and the synaptic couplings among different neurons are provided by a real matrix J_{ik}. At each time step neurons are updated; assuming parallel dynamics, as we shall do in the present paper, the neuron state at time $t+1$ is given by the deterministic rule

$$S_i'=\operatorname{sgn}\left(\sum_{k=1}^{N} J_{ik}S_k\right) \tag{1.1}$$

where $S_i=S_i(t)$ and $S_i'=S_i(t+1)$ are states of the neuron i at the times t and $t+1$, respectively. One can also introduce a random noise, parametrized in terms of temperature $T=\beta^{-1}$ (Little 1974), so that the updating rule becomes probabilistic and (1.1) is substituted by

$$\begin{aligned} P(\{S_i'\}|\{S_j\}) &= \prod_{i=1}^{N} P(S_i'|\{S_j\}) \\ &= \prod_{i=1}^{N} [1+\exp(-2\beta S_i' H_i)]^{-1} \end{aligned} \tag{1.2}$$

* Research supported in part by MPI (Ministero della Pubblica Istruzione—Italy).

0305-4470/91/051103+15$03.50

where $P(\{S_i'\}|\{S_j\})$ is the probability for neurons having the values $\{S_i'\}$ at time $t+1$, given the network configuration $\{S_i\}$ at time t, and

$$H_i = \sqrt{\frac{\alpha}{N}} \sum_{j=1}^{N} J_{ij} S_j. \tag{1.3}$$

For symmetric synaptic couplings ($J_{ik} = J_{ki}$) one can show (Peretto 1984) that the master equation based on the transition rate (1.2) satisfies detailed balance and tends to a stationary Gibbs distribution; therefore, the standard methods of equilibrium statistical mechanics can be successfully applied. In this paper we shall consider a different problem, i.e. the effect of the random noise on the domains of attraction of the stored patterns.

The domain of attraction of one of the patterns $\{\xi_i\}$ is defined in terms of the initial overlap between $\{\xi_i\}$ and $\{S_i(0)\}$:

$$m_0 = m(t=0) = \frac{1}{N} \sum_{i=1}^{N} \xi_i S_i(0). \tag{1.4}$$

One says that the domain of attraction of $\{\xi_i\}$ is measured by the number m_c if any initial configuration $\{S_i(0)\}$ having

$$m_0 \geqslant m_c \tag{1.5}$$

is attracted towards the fixed point $\{\xi_i\}$.

The crucial parameter describing the ability of the network to retrieve the stored pattern $\{\xi_i\}$ from the initial configuration is

$$K = \min_i \gamma_i \tag{1.6}$$

with

$$\gamma_i = \frac{1}{\sqrt{N}} \sum_{j=1}^{N} J_{ij} \xi_j \xi_i \tag{1.7}$$

where we choose the normalization

$$\sum_{i,j=1}^{N} J_{ij}^2 = N^2 \tag{1.8}$$

and assume

$$J_{ii} = 0. \tag{1.9}$$

One can show that ξ_i is a fixed point of the dynamics (1.1) if and only if $K > 0$; moreover, larger values of K correspond to larger basins of attraction (Forrest 1988).

A systematic study of the K dependence of the basins of attraction has been recently made (Kepler and Abbott 1988) by considering networks trained by the so-called Edinburgh algorithm (Wallace 1985, Bruce *et al* 1987, Gardner 1988), which is a local iterative learning algorithm that can be used to construct matrices J_{ik} implementing the condition $\gamma_i \geqslant K$. The purpose of the present paper is to study the thermal dependence of the domains of attraction and to generalize the $T=0$ results to the case of finite temperature. The main result of our analysis is that, for saturated networks (i.e. networks having the maximum number of patterns allowed by a given value of K), the size of the domains of attraction is basically independent of temperature and is determined only by K: the random noise slows down the retrieval of the memories,

but does not change the critical value m_c of the initial overlap that fixes the domain of attraction.

The plan of the paper is as follows. In section 2 we describe the dynamics at $T=0$. Since we are interested in the generalization to finite temperature we consider symmetric synaptic matrices: first we describe the algorithm that generates J_{ik}; second we discuss the properties of saturated networks; finally we review the argument used to determine the domain of attraction at zero temperature. In section 3 we discuss the dynamics at $T\neq 0$ and we compare the theoretical expectations for $m_1(m_0, \beta)$ with computer data. In section 4 we present our results, showing the independence from temperature of the domains of attraction and we draw our conclusions. Finally, in the appendix we show that the Gardner formula (Gardner 1988) relating, for saturated networks, K to α can be generalized to symmetric synaptic strengths.

2. Dynamics at $T=0$

First of all we describe the symmetric learning algorithm we shall use in the present paper. We wish to generate a symmetric matrix of synaptic couplings J_{ik} having ξ^μ as fixed points of the dynamics and satisfying, for each i, μ,

$$\gamma_i^\mu = \frac{\sqrt{N}}{\|J\|} \sum_{j=1}^{N} J_{ij}\xi_j^\mu \xi_i^\mu \geqslant K. \tag{2.1}$$

The norm of matrix J_{ik} is defined in terms of the scalar product

$$\|J\| = \sqrt{(J, J)} \tag{2.2}$$

and

$$(J, U) = \sum_{l,m=1}^{N} J_{lm}U_{lm}. \tag{2.3}$$

We note that (2.1) reduces to (1.6) and (1.7) if the normalization condition (1.8) is satisfied.

In order to implement (2.1) we consider a modified version of the Edinburgh algorithm: one starts from a symmetric random matrix $J_{ij}^{(0)}$ and at each time step a parallel updating in the sites i, j is performed as follows:

$$J_{ij}^{(m)} \to J_{ij}^{(m+1)} = J_{ij}^{(m)} + \delta J_{ij} \tag{2.4}$$

$$\delta J_{ij} = \frac{\|J^{(m)}\|}{N^{3/2}} \sum_{\mu=1}^{\alpha N} \frac{f(\gamma_i^\mu)\varepsilon_i^\mu + f(\gamma_j^\mu)\varepsilon_j^\mu}{2} \xi_i^\mu \xi_j^\mu (1-\delta_{ij}) \tag{2.5}$$

where $J_{ij}^{(m)}$ is the matrix of synaptic couplings after m iterations, $f(\gamma)$ is a function to be described below and ε_i^μ is the mask

$$\varepsilon_i^\mu = \theta\left(\frac{\|J^{(m)}\|}{\sqrt{N}} K - \sum_{j=1}^{N} J_{ij}^{(m)}\xi_j^\mu \xi_i^\mu\right). \tag{2.6}$$

The algorithm is iterated until $\varepsilon_i^\mu = \varepsilon_j^\mu = 0$, in which case the condition (2.1) is fulfilled. This learning algorithm for $f=$ constant represents a generalization of the Hebb rule (Hebb 1949) and coincides with the symmetric version of the Edinburgh algorithm (Bruce *et al* 1987, Gardner 1988). Different choices can be done for the function f; in particular one can show (Abbott and Kepler 1989) that, for

$$f(\gamma) = K + \delta - \gamma + \sqrt{(K+\delta-\gamma)^2 - \delta^2} \tag{2.7}$$

and asymmetric matrices, a convergence theorem similar to the perceptron convergence theorem (Rosenblatt 1962, Minsky and Papert 1969) exists and, in addition, a remarkable increase of the speed of convergence is obtained. The choice (2.7) is called the nonlinear rule; for small δ the function f is then approximately given by the maximum value compatible with the convergence of the algorithm: $f = 2(K + \delta - \gamma)$. The convergence theorem in the present case states that if a matrix J^* exists such that

$$\sum_{j=1}^{N} J_{ij}^* \xi_j^\mu \xi_i^\mu > (K+\delta)\frac{\|J^*\|}{\sqrt{N}} \tag{2.8}$$

then the algorithm defined by (2.4) and (2.5) converges in a finite number of steps; the proof of the theorem is similar to the one valid for the asymmetric case and it will not be repeated here. We have trained networks with $N = 200$ and $\delta = 0.01$ for several values of K, from $K \simeq 0.70$ to $K \simeq 4.9$, with αN independent memory states, where α is the maximum value of the capacity parameter allowed by K; in other words we have considered (almost) saturated networks. For any value of K there is a maximum number of memories $P = \alpha_s N$ that can be stored; the saturation value α_s is obtained by considering the limit $q \to 1$ in the equation:

$$q = \frac{\alpha}{2\pi}(1-q)\int_{-\infty}^{+\infty} \mathrm{D}t \frac{\exp(-x^2)}{H^2(x)} \tag{2.9}$$

where $\mathrm{D}t = \mathrm{d}t \exp(-t^2/2)/\sqrt{2\pi}$, $x = (\sqrt{q}\, t + K)/\sqrt{1-q}$ and

$$H(x) = \int_x^{+\infty} \mathrm{D}t. \tag{2.10}$$

Equation (2.9) has been first obtained for asymmetric matrices (Gardner 1988); its extension to symmetric synaptic couplings is discussed in the appendix, where we also compare our results with some results previously obtained on this subject (Gardner *et al* 1989). For the saturation value one obtains

$$\alpha_s = \left(\int_{-K}^{+\infty} \mathrm{D}t (t+K)^2\right)^{-1}. \tag{2.11}$$

The results (2.9)-(2.11) are obtained in the thermodynamical limit ($N \to \infty$). For finite N the maximum number of patterns that can be stored is smaller than $\alpha_s N$; for example, $\alpha_s = 0.25$ would correspond to $K = 1.74$ (from (2.11)), whereas the maximum value of K that can be reached is $K \simeq 1.67$ (after 79 iterations). However, it should be noted that numerical results do not depend on the precise determination of K around the critical value. By way of example in figure 1 we consider the overlap between $\{\xi_i\}$ and the network configuration after one time step:

$$m_1 = m(t=1) = \frac{\Sigma_i \xi_i S_i(1)}{N} \tag{2.12}$$

and the final overlap

$$m_f = m(t=\infty) = \frac{\Sigma_i \xi_i S_i(\infty)}{N}. \tag{2.13}$$

In figure 1(a) we plot the value of m_1 as a function of K obtained by two simulations with the same value of α and different initial overlaps m_0; the limiting values of m_1,

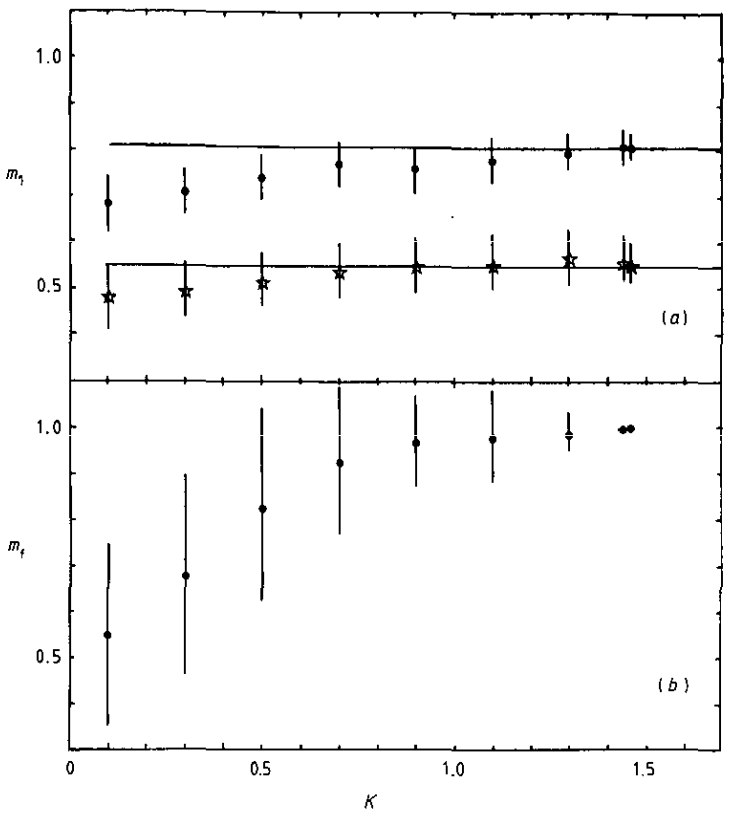

Figure 1. (*a*) First overlap m_1 as a function of the parameter K for two different values of m_0: $m_0 = 0.60$ (upper curve) and $m_0 = 0.40$ (lower curve); data are taken at $\alpha = 0.25$ which corresponds to a saturation value $K = 1.74$. The straight lines are the theoretical expectations corresponding to $K = 1.74$. (*b*) Final overlap m_f as a function of K (for $\alpha = 0.25$); $m_f = 1$ is the theoretical expectation for $K = 1.74$.

corresponding to $K = 1.74$ are computed by using usual formulae (Kepler and Abbott 1988) and one clearly sees that they are reached, within the errors, already for $K \approx 1.44$. The same pattern can be recognized in figure 1(*b*), where the value $m_f = 1$ is obtained for $K \geqslant 1.40$. Since dynamical quantities are more sensitive to α than K, we prefer to use α instead of K as the parameter describing the domain of attraction; however, it is understood that for each value of α the corresponding K can be obtained by solving (2.11).

Let us conclude this section by briefly summarizing the approach which fixes the basin of attraction at $T = 0$ and for K not too small ($K \geqslant 0.7$) (Kepler and Abbott 1988). First of all one observes empirical evidence for the crucial role played by the parameter

$$\frac{m_1 - m_0}{1 - m_0}$$

in determining whether the network is attracted towards $\{\xi_i\}$, starting from an initial configuration $\{S_i(0)\}$ with overlap m_0 with $\{\xi_i\}$. As a matter of fact one finds that the

probability for the convergence is approximately given by

$$P_{S\to\xi}=\theta\left(\frac{m_1-m_0}{1-m_0}-a_0\right). \tag{2.14}$$

We have checked that this result also holds for exactly symmetric matrices J_{ik}; a_0 has the value

$$a_0\approx 0.5 \tag{2.15}$$

which is almost independent of α, as can be seen from figure 2. On the basis of the empirical evidence found in Kepler and Abbott (1988), (2.14) is expected to hold exactly in the thermodynamical limit; for finite N ($N=200$ in our case) we conventionally define a_0 as the value at which the probability of a final overlap $m_f>0.97$ is larger than 97%.

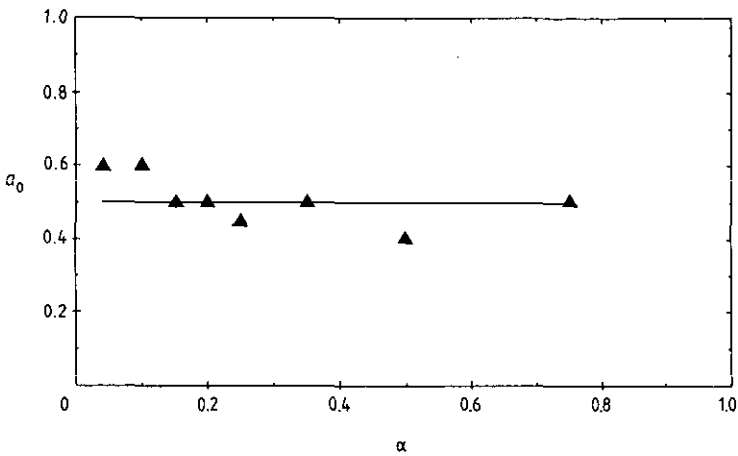

Figure 2. The parameter a_0 as a function of α.

In order to determine the basin of attraction at $T=0$ one derives a formula relating m_1 to m_0 for given K (Forrest 1988, Kepler and Abbott 1988, Krauth *et al* 1988b):

$$m_1(m_0)=\int_{-\infty}^{+\infty}\mathrm{d}\gamma\,\rho(\gamma)\,\mathrm{erf}\frac{m_0\gamma}{\sqrt{2(1-m_0^2)}} \tag{2.16}$$

where, for saturated networks, one finds:

$$\rho(\gamma)=\theta(\gamma-K)\frac{\exp(-\gamma^2/2)}{\sqrt{2\pi}}+\tfrac{1}{2}\delta(\gamma-K)\left(1+\mathrm{erf}\frac{K}{\sqrt{2}}\right). \tag{2.17}$$

The value of the domain of attraction m_c is determined by solving the equation

$$\frac{m_1(m_c)-m_c}{1-m_c}=a_0. \tag{2.18}$$

In the next section we shall generalize this procedure to finite temperature; we observe here that (2.17) also holds for symmetric matrices J_{ik}; as a matter of fact the only difference in the definition of $\rho(\gamma)$ (Kepler and Abbott 1988) for the asymmetric and the symmetric cases is the normalization of the matrix, $\Sigma_j J_{ij}^2=N$ and $\Sigma_{i,j} J_{ij}^2=N^2$ respectively; however, in the limit $N\to\infty$ the two conditions become equivalent, so

that (2.17) holds for symmetric synaptic couplings too. Let us finally observe that the effect of the symmetry of J_{ik} on the domains of attraction has been also studied by Krauth *et al* (1988a), in a different context, i.e. in the one-pattern model.

3. Dynamics at $T \neq 0$

In this section we generalize previous results to finite temperature, when the dynamics is provided by the probabilistic rule (1.2). From computer data, obtained by almost saturated symmetrical networks with various values of α, we first obtain a generalization of (2.14) in the form

$$P_{S \to \xi} = \theta\left(\frac{m_1 - m_0}{1 - m_0} - a(\alpha, T)\right) \tag{3.1}$$

where

$$a(\alpha, 0) = a_0. \tag{3.2}$$

For each value of α, $a(\alpha, T)$ decreases with T, as can be seen from figure 3 (the curves in the figure correspond to a least-squares fit to computer data by a second-degree polynomial, i.e. $a(\alpha, T) \approx a_0 - a_1 T - a_2 T^2$). The introduction of the temperature does not change the shape of the probability function $P_{S \to \xi}$: it simply shifts the curve by the amount $a(\alpha, T) - a(\alpha, 0)$, as can be seen from figure 4 where we plot $P_{S \to \xi}$ as a function of $(m_1 - m_0)/(1 - m_0)$ for different values of T. It should be noted that, for each α, there is a maximal temperature T_m above which retrieval of a stored pattern ξ is impossible even starting with an initial configuration having overlap $m_0 = 1$ with ξ. This result appears clearly in figure 5, where we have plotted computer data for the final overlap m_f as a function of T for different values of α. T_m is the abscissa of the intersection point between the straight line $m_f = 0.97$ with the magnetization curve; similar curves were also obtained for the Hopfield model (Amit *et al* 1987).

In order to check if the changes of the dynamics we have described above result in a variation of the domains of attraction, we now compute the first overlap at finite

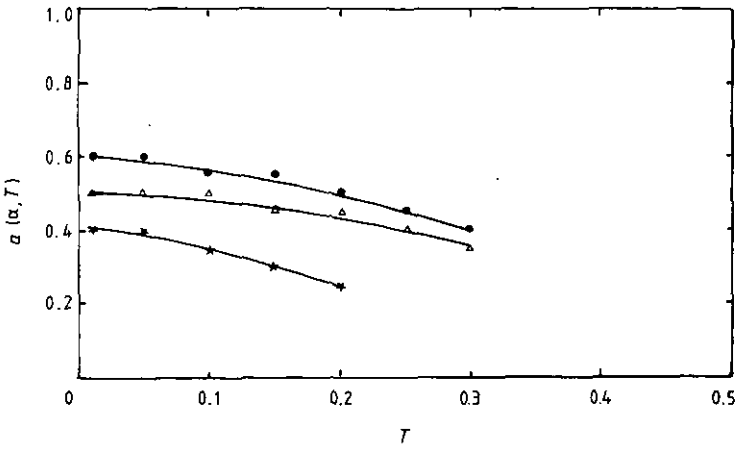

Figure 3. Computer data for the parameter $a(\alpha, T)$ as a function of the temperature T for different values of α: going from top to bottom corresponds to $\alpha = 0.15$, 0.25 and 0.50. The curves are polynomial fits to the data.

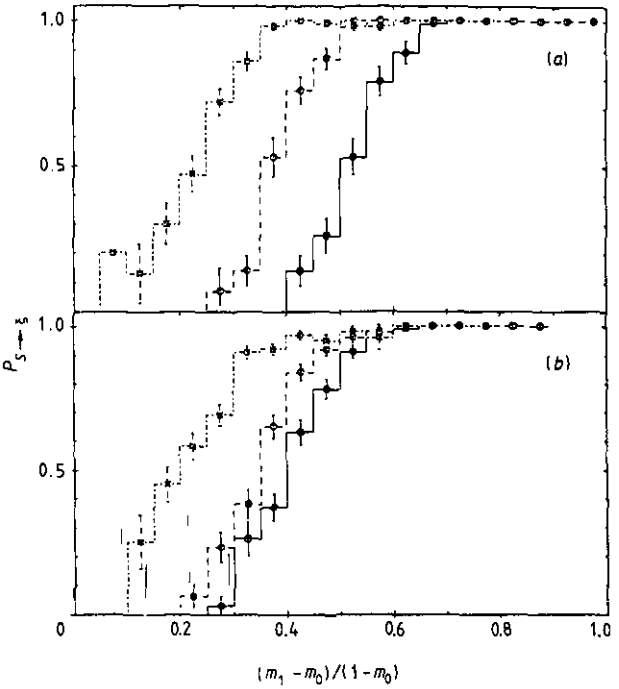

Figure 4. The probability $P_{S\to\xi}$ of almost perfect retrieval ($m_f \geqslant 0.97$) as a function of $(m_1-m_0)/(1-m_0)$ for two values of α: $\alpha=0.04$ (a) and $\alpha=0.15$ (b). From the right to the left the histograms correspond to $T=0$, 0.15 and 0.30.

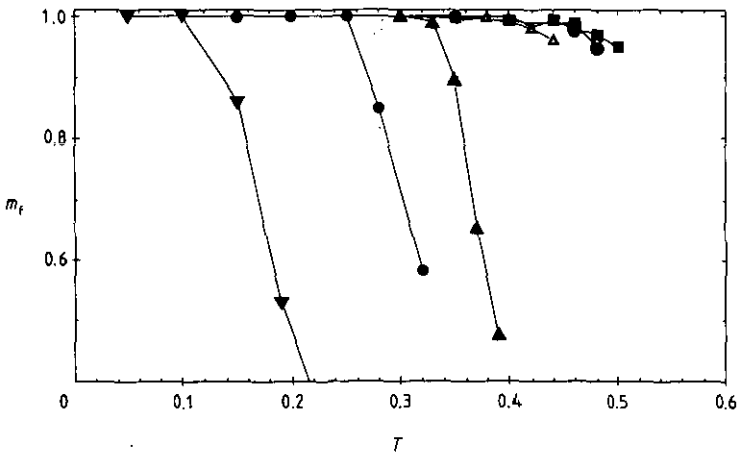

Figure 5. Final overlap m_f as a function of T for different values of α; going from right to left the data correspond to $\alpha=0.04$, 0.10, 0.15, 0.35, 0.50 and 0.75.

temperature. First we write the probability distribution of m_1, for a given value of m_0 at $T\neq 0$, as follows:

$$P_T(m_1|m_0)=\frac{\mathrm{Tr}_{S'}\,\mathrm{Tr}_S\,\delta(m_0-1/N\,\Sigma_i\,\xi_i S_i)P(\{S'\}|\{S\})\delta(m_1-1/N\,\Sigma_i\,\xi_i S_i')}{\mathrm{Tr}_S\,\delta(m_0-1/N\,\Sigma_i\,\xi_i S_i)} \tag{3.3}$$

where $S_i = S_i(0)$, $S_i' = S_i'(t=1)$ and $\{\xi_i\}$ is one of the stored patterns. We now compute (3.3) by an approach similar to the one used by Kepler and Abbott (1988). By using (1.2) and introducing the independent variables $h_j = 1/\sqrt{N}\,\Sigma_k J_{jk}\xi_j S_k$ we get

$$P_T(m_1|m_0) = \mathrm{Tr}_{S'}\,\mathrm{Tr}_S \int \prod_j \left[\mathrm{d}h_j\, \delta\left(h_j - \frac{1}{\sqrt{N}} \sum_{k\neq j} J_{jk}\xi_j S_k\right) \frac{\exp(\beta\sqrt{\alpha}\; S_j'\xi_j h_j)}{2\cosh(\beta\sqrt{\alpha}\;\xi_j h_j)} \right]$$
$$\times \frac{\delta(m_0 - 1/N\,\Sigma_i\,\xi_i S_i)\,\delta(m_1 - 1/N\,\Sigma_i\,\xi_i S_i')}{\mathrm{Tr}_S\,\delta(m_0 - 1/N\,\Sigma_i\,\xi_i S_i)}$$
$$= \int \left(\prod_j \mathrm{d}h_j\right) \mathrm{Tr}_{S'}\,\delta\left(m_1 - \frac{1}{N}\sum_i \xi_i S_i'\right) \prod_l p_l(h, \{S'\}|m_0) \tag{3.4}$$

where

$$p_l(h, \{S'\}|m_0) = \frac{\exp(\beta\sqrt{\alpha}\; S_l'\xi_l h_l)}{2\cosh(\beta\sqrt{\alpha}\;\xi_l h_l)}\, p(h_l|m_0) \tag{3.5a}$$

$$p(h_l|m_0) = \frac{\mathrm{Tr}_S\,\delta(m_0 - 1/N\,\Sigma_i\,\xi_i S_i)\,\delta(h_l - 1/\sqrt{N}\,\Sigma_k\, J_{lk}\xi_l S_k)}{\mathrm{Tr}_S\,\delta(m_0 - 1/N\,\Sigma_i\,\xi_i S_i)}. \tag{3.5b}$$

$p(h_l|m_0)$ is independent of temperature and in the thermodynamical limit reduces to

$$p(h_l|m_0) = \frac{1}{\sqrt{2\pi(1-m_0^2)}} \exp\left(-\frac{(h_l - m_0\gamma_l)^2}{2(1-m_0^2)}\right) \tag{3.6}$$

with γ_l defined in (1.7). We use an exponential representation for $\delta(m_1 - 1/N\,\Sigma_i\,\xi_i S_i')$ in (3.4) and we get

$$P_T(m_1|m_0) = [2\pi(1-m_0^2)]^{-N/2} \int \frac{\mathrm{d}x}{2\pi} \exp(\mathrm{i}m_1 x) \prod_j \left[\mathrm{d}h_j \exp\left(-\frac{(h_j - m_0\gamma_j)^2}{2(1-m_0^2)}\right) \right.$$
$$\left. \times \left(\frac{\exp(-\mathrm{i}x/N\ \mathrm{sgn}\ h_j)}{1+\exp(-2\beta\sqrt{\alpha}|h_j|)} + \frac{\exp(\mathrm{i}x/N\ \mathrm{sgn}\ h_j)}{1+\exp(2\beta\sqrt{\alpha}|h_j|)}\right)\right]. \tag{3.7}$$

The integrals over the variables h_j can be performed in series with the result

$$P_T(m_1|m_0) = \int \frac{\mathrm{d}x}{2\pi} \exp\left\{ \mathrm{i}m_1 x + \sum_{j=1}^{N} \ln\left[\cos\frac{x}{N} - \mathrm{i}\sin\frac{x}{N}\,\mathrm{erf}\frac{m_0\gamma_j}{\sqrt{2(1-m_0)^2}} \right.\right.$$
$$\left.\left. + \mathrm{i}\sin\left(\frac{x}{N}\right) f(\gamma_j, m_0, \beta)\right]\right\} \tag{3.8}$$

where

$$f(\gamma_j, m_0, \beta) = \sum_{n=1}^{\infty} (-1)^n \{\mathrm{erfc}(u_n^+)\exp[(u_n^+)^2 - u^2] - \mathrm{erfc}(u_n^-)\exp[(u_n^-)^2 - u^2]\} \tag{3.9}$$

with

$$\mathrm{erfc}(z) = 1 - \mathrm{erf}(z) \tag{3.10a}$$

$$u = \frac{\gamma m_0}{\sqrt{2(1-m_0^2)}} \tag{3.10b}$$

and

$$u_n^{\pm} = n\beta\sqrt{2\alpha(1-m_0^2)} \pm u. \tag{3.11}$$

Therefore in the thermodynamical limit we obtain

$$P_T(m_1|m_0)=\delta\left[m_1-\frac{1}{N}\sum_{i=1}^{n}\left(\operatorname{erf}\frac{m_0\gamma_i}{\sqrt{2(1-m_0^2)}}-f(\gamma_i, m_0, \beta)\right)\right] \tag{3.12}$$

so that we derive

$$m_1(m_0, T)=\int_{-\infty}^{+\infty} d\gamma\,\rho(\gamma)\left(\operatorname{erf}\frac{m_0\gamma}{\sqrt{2(1-m_0^2)}}-f(\gamma, m_0, \beta)\right) \tag{3.13}$$

with $\rho(\gamma)$ given in (2.17). Equation (3.13) shows a clear dependence of m_1 on T; moreover, it is easily seen that for $T\to 0$ one again obtains (2.16).

In figure 6 we compare the analytical result (3.13) with the computer data obtained for different T and α. For all values of α we observe clear deviations from the zero temperature results: for fixed m_0, m_1 decreases as T increases. The overall agreement between analytical predictions and computer data is good.

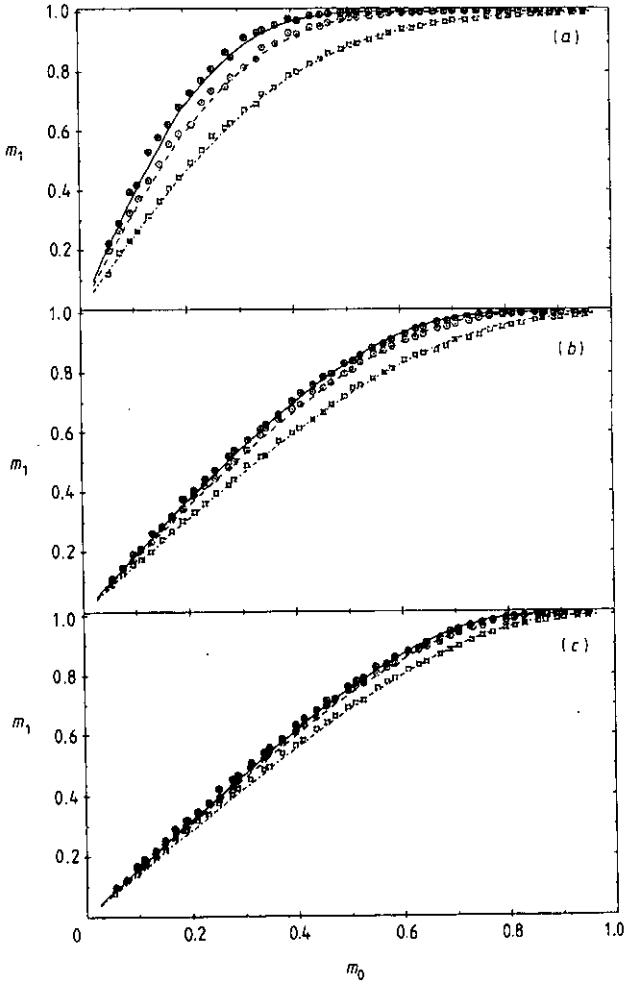

Figure 6. Comparison between the predicted value of m_1 as a function of m_0 and computer data for three different values of α: $\alpha=0.04$ (a), $\alpha=0.15$ (b) and $\alpha=0.20$ (c) and different temperatures. Going from the top curve to the bottom corresponds to $T=0$, 0.15 and 0.30.

4. Conclusions

In order to compute the domains of attraction of the stored patterns $\{\xi_i\}$, we substitute $a(\alpha, T)$ and $m_1(m_0, T)$ in the equation

$$\frac{m_1(m_c, T) - m_c}{1 - m_c} = a(\alpha, T). \tag{4.1}$$

For $m_1(m_c, T)$ we use (3.13), whereas for $a(\alpha, T)$ we employ the polynomial approximations to the computer data. By solving (4.1) one obtains $m_c(\alpha, T)$, i.e. the size of the domain of attraction as a function of the capacity parameter α and the temperature. These results are expressed by the curves of figure 7 which agree fairly well with numerical data obtained by the computer simulation. The interesting result we find, in the range of values of T and α where we have data, is that m_c depends only on α: in other words the T dependence of RHS and LHS of (4.1) cancels out. This is shown by the almost straight lines appearing in the figure, which expresses the independence of the domain of attraction from the temperature.

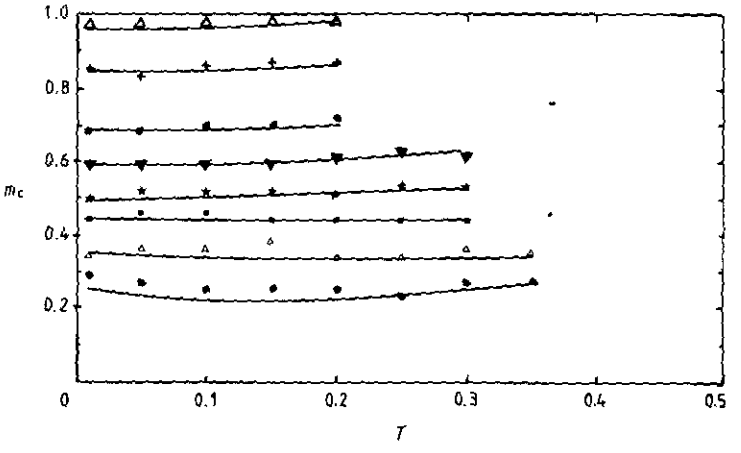

Figure 7. Comparison between the predicted values for the domains of attraction m_c and the computer data. Curves from bottom to top correspond to $\alpha = 0.04$, 0.10, 0.15, 0.20, 0.25, 0.35, 0.50 and 0.75.

The result we find is far from being trivial because, as we have shown, it follows from a well-defined change in the dynamics induced by the thermal noise. The main effect of the introduction of the temperature is to slow down the retrieval process, as is shown by the decrease of m_1 with T in figure 6. The convergence towards the fixed point ξ needs more time; nevertheless, the thermal noise does not change the size of the basin of attraction which is only measured, for saturated networks, by the capacity parameter α.

Acknowledgments

We thank A Maritan for helpful suggestions and G Cicuta, P Colangelo, N Cufaro-Petroni and M Villani for useful discussions.

Appendix

The proof of (2.9) is an extension to symmetric synaptic couplings of the results valid for the asymmetric case (Gardner 1988). Basically one has to compute the fraction V_T of the phase space occupied by matrices satisfying (1.8), (1.9), (2.1) and the symmetric condition

$$J_{ij} = J_{ji}. \tag{A1}$$

V_T is given by

$$V_T = \frac{M}{D} \tag{A2}$$

with

$$M = \int \prod_{i,j,(i\neq j)} \mathrm{d}J_{ij}\, \delta\left(\sum_{s,t,(s\neq t)} J_{st}^2 - N^2\right) \prod_{l<k} \delta(J_{lk} - J_{kl}) \prod_{\mu,i} \theta(\gamma_i^\mu - K) \tag{A3}$$

$$D = \int \prod_{i,j(i\neq j)} \mathrm{d}J_{ij} \prod_{l<k} \delta(J_{lk} - J_{kl}) \delta\left(\sum_{s,t,(s\neq t)} J_{st}^2 - N^2\right) \tag{A4}$$

and

$$\gamma_i^\mu = \frac{1}{\sqrt{N}} \sum_{i\neq j} J_{ij} \xi_i^\mu \xi_j^\mu. \tag{A5}$$

In order to compute M we write

$$\prod_{l<k} \delta(J_{lk} - J_{kl}) = \int \prod_{l<k} \frac{\mathrm{d}w_{lk}}{2\pi} \exp\left(\mathrm{i} \sum_{l<k} (J_{lk} - J_{kl}) w_{lk}\right). \tag{A6}$$

Therefore (A3) can be recast in the form

$$M = \int \frac{\mathrm{d}E}{4\pi \mathrm{i}} \exp\left(\frac{EN^2}{2}\right) \prod_{l<k} \frac{\mathrm{d}w_{lk}}{2\pi} \exp\left(\sum_i \ln z_i\right) \tag{A7}$$

where

$$z_i = \int \prod_{j\neq i} \mathrm{d}J_{ij} \exp\left(-\frac{E}{2} \sum_{j\neq i} J_{ij}^2 + \mathrm{i} \sum_{j\neq i} X_{ij} J_{ij}\right) \prod_\mu \theta(\gamma_i^\mu - K) \tag{A8}$$

and

$$X_{ij} = \begin{cases} +w_{ij} & (i<j) \\ -w_{ji} & (i>j). \end{cases} \tag{A9}$$

We now assume that $\ln z_i$ is self-averaging, so that we only have to compute the average $\langle \ln z \rangle$ over the quenched variables ξ_i^μ. In order to do that we introduce n replicas and make use of the usual identity

$$\langle \ln z \rangle = \lim_{n\to 0} \frac{\langle z^n \rangle - 1}{n}. \tag{A10}$$

The average $\langle z^n \rangle$ is given by

$$\langle z^n \rangle = \left\langle \prod_{\alpha=1}^{n} \int \prod_{i\neq j} \mathrm{d}J_{ij}^\alpha \exp\left[\sum_\alpha \sum_{j\neq i} \left(-\frac{E}{2} (J_{ij}^\alpha)^2 + \mathrm{i} X_{ij} J_{ij}^\alpha\right)\right] \right.$$
$$\left. \times \prod_\mu \theta\left(\frac{\xi_i^\mu}{\sqrt{N}} \sum_{i\neq j} J_{ij}^\alpha \xi_j^\mu - K\right) \right\rangle. \tag{A11}$$

We now use the identity

$$\theta\left(\frac{1}{\sqrt{N}}\xi_i^\mu \sum_{j\neq i} J_{ij}^\alpha \xi_j^\mu - K\right)$$

$$= \int_K^{+\infty} \frac{\mathrm{d}\lambda_\alpha^\mu}{2\pi} \int_{-\infty}^{+\infty} \mathrm{d}y_\alpha^\mu \exp\left[\mathrm{i}y_\alpha^\mu\left(\lambda_\alpha^\mu - \frac{1}{\sqrt{N}}\xi_i^\mu \sum_{j\neq i} J_{ij}^\alpha \xi_j^\mu\right)\right] \tag{A12}$$

and introduce the parameters $q_{\alpha\beta}$ and its conjugate variables $F_{\alpha\beta}$ as in the Gardner calculation

$$1 = \int \prod_{\alpha<\beta} \mathrm{d}q_{\alpha\beta}\, \delta\left(q_{\alpha\beta} - \frac{1}{N}\sum_{j\neq i} J_{ij}^\alpha J_{ij}^\beta\right)$$

$$= \int \prod_{\alpha<\beta} \mathrm{d}q_{\alpha\beta} \frac{\mathrm{d}F_{\alpha\beta}}{(2\pi\mathrm{i}/N)} \exp\left[-\sum_{\alpha<\beta} F_{\alpha\beta}\left(Nq_{\alpha\beta} - \sum_{j\neq i} J_{ij}^\alpha J_{ij}^\beta\right)\right] \tag{A13}$$

with the result

$$\langle z^n\rangle = \int \prod_{\alpha<\beta} \mathrm{d}q_{\alpha\beta} \frac{\mathrm{d}F_{\alpha\beta}}{(2\pi\mathrm{i}/N)} \exp\left[N\left(\alpha G_1(q) + G_2(F, E, X) - \sum_{\alpha<\beta} F_{\alpha\beta}q_{\alpha\beta}\right)\right] \tag{A14}$$

where

$$G_1(q) = \ln \prod_{\alpha=1}^{n} \int_{-\infty}^{+\infty} \mathrm{d}y_\alpha \int_K^{+\infty} \frac{\mathrm{d}\lambda_\alpha}{2\pi} \exp\left(\mathrm{i}\sum_\alpha y_\alpha\lambda_\alpha - \tfrac{1}{2}\sum_\alpha y_\alpha^2 - \sum_{\alpha<\beta} q_{\alpha\beta}y_\alpha y_\beta\right)$$

$$G_2(F, E, X) = \frac{1}{N}\ln \prod_{\alpha=1}^{n} \prod_{j\neq i} \int \mathrm{d}J_{ij}^\alpha$$

$$\times \exp\left[\sum_{i\neq j}\left(\mathrm{i}X_{ij}\sum_\alpha J_{ij}^\alpha - \frac{E}{2}\sum_\alpha (J_{ij}^\alpha)^2 + \sum_{\alpha<\beta} F_{\alpha\beta}J_{ij}^\alpha J_{ij}^\beta\right)\right]. \tag{A15}$$

The RHS of (A14) can be computed by the saddle-point method; the saddle-point equations are

$$\alpha\frac{\partial G_1}{\partial q_{\alpha\beta}} = F_{\alpha\beta}$$

$$\frac{\partial G_2}{\partial F_{\alpha\beta}} = q_{\alpha\beta}. \tag{A16}$$

Assuming replica symmetry one has

$$F_{\alpha\beta} = F$$

$$q_{\alpha\beta} = q. \tag{A17}$$

Equations (A16) become, in the $n\to 0$ limit,

$$F = \frac{\alpha}{2\pi(1-q)} \int \mathrm{D}t \frac{\exp(-x^2)}{H^2(x)} \tag{A18}$$

$$q = \frac{F-\lambda}{(E+F)^2} \tag{A19}$$

where $Dt = dt \exp(-t^2/2)/\sqrt{2\pi}$, $x = (\sqrt{q}\, t + K)/\sqrt{1-q}$, $H(x) = \int_x^\infty Dz$ and

$$\lambda = \frac{1}{N} \sum_{j \neq i} X_{ij}^2 \tag{A20}$$

where X_{ij} *is defined in* (A9).

In conclusion, M in (A2) has the form

$$M = \int \frac{dE}{4\pi i} \exp\left(\frac{EN^2}{2}\right) \prod_{l<k} \frac{dw_{lk}}{2\pi} \exp(N^2 S[E, q]) \tag{A21}$$

where

$$S[E, q] = \alpha \int Dt \ln H(x) + \tfrac{1}{2} \ln \frac{2\pi}{E+F} + Fq + \frac{Eq}{2}. \tag{A22}$$

In order to compute (A20) we again use the saddle-point method and obtain the equations

$$w_{lk} = 0$$

$$\frac{1}{2} + \frac{\partial}{\partial E}\left(\alpha \int Dt \ln H(x) - \frac{1}{4} \ln \frac{F}{q} + \frac{Fq}{2} + \frac{\sqrt{Fq}}{2}\right) = 0. \tag{A23}$$

Assuming that both $E(q)$ and $F(q)$ are regular for $q \to 0$, (A23), together with (A18), give the result

$$E = -F + \sqrt{\frac{F}{q}} \tag{A24}$$

$$F = \frac{q}{(1-q)^2} \tag{A25}$$

so that we finally obtain from (A18) the following formula:

$$q = \frac{\alpha}{2\pi}(1-q) \int Dt \frac{\exp(-x^2)}{H^2(x)}. \tag{A26}$$

It is easily shown that the limit $q \to 1$ corresponds to $M \to 0$, i.e. the vanishing of the fraction of phase space V_T defined in (A2)-(A4). Therefore, in this limit one obtains the maximum capacity of the network for a given value of K, i.e. the value α_s given in (2.11). These results coincide with the formulae obtained by Gardner (1988).

Let us now compare our result with the analysis performed by Gardner *et al* (1989). Gardner and coworkers have tried to compute the space of interaction in neural networks in the general case, i.e. for any symmetry property of J_{ik}. They parametrize the symmetry properties of J_{ik} in terms of a parameter η and observe that in the fully connected case one cannot write the fractional volume V_T as a product of volumes for each site, due to the fact that different rows of the matrix J_{ik} are correlated; therefore, they consider the average over ξ_i^μ as given by

$$\left\langle \prod_{\alpha,\mu,i} \theta(\gamma_i^\mu - K) \right\rangle \tag{A27}$$

(see (8) in their paper) and they find that the cumulant expansion does not converge.

The main difference between the present work and the one by Gardner *et al* is that we consider

$$\left\langle \prod_{\alpha,\mu} \theta(\gamma_i^\mu - K) \right\rangle \tag{A28}$$

(see (A11)), which can be computed exactly as we have already shown. The reason why we consider the quantity (A28) instead of (A27) is that, since we only consider the special case of symmetric synaptic couplings, corresponding to $\eta = 1$, we must take into account the condition $\delta(J_{jk} - J_{kj})$ and not the general one $\delta(\Sigma_j J_{ij} J_{ji} - \eta N)$. Therefore we are able to write (A7) and compute $\ln z$ assuming self-averaging. This is different from computing $\ln V_T$, which, as stressed by Gardner *et al*, cannot be done by the Gardner method.

Owing to the difficulty in the treatment of the general fully connected case, Gardner and coworkers compute the phase space of interaction in a diluted case, where each site is connected, on the average, to C other sites ($C \ll N$). They find analytic expressions for $\alpha_c(K, \eta)$ that, for $\eta = 1$, are lower than the Gardner well-known result. On the contrary we have found that the Gardner result is also valid in the symmetric fully connected case. It is clear that there is no conflict since in this paper we have considered a fully connected network and not a diluted one. It is easy to be convinced of such a difference by considering (25) and (26) in the paper by Gardner *et al*. By way of example, for $K = 1.67$ these equations give the maximum storage capacity $\alpha_c = 0.219$, whereas, as we have mentioned in section 2, we are able to reach numerically the value $\alpha = 0.25$ with the same value of K: in other words we find 'experimentally' that for fully connected networks the maximum storage capacity α_c is larger than the corresponding value obtained for diluted nets and is actually very close to the theoretical value, valid for $N \to \infty$, predicted by (2.11): $\alpha_c = 0.265$ for $K = 1.67$.

References

Abbott L F and Kepler T B 1989 *J. Phys. A: Math. Gen.* **22** L711

Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys., NY* **173** 30

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

Forrest B 1988 *J. Phys. A: Math. Gen.* **21** 245

Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257

Gardner E, Gutfreund H and Yekutieli I 1989 *J. Phys. A: Math. Gen.* **22** 1995

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

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

Kepler T B and Abbott L F 1988 *J. Physique* **49** 1657

Krauth W, Nadal J-P and Mezard M 1988a *J. Phys. A: Math. Gen.* **21** 2995

Krauth W, Mezard M and Nadal J-P 1988b *Complex Syst.* **2** 387

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

Mezard M, Parisi G and Virasoro M *Spin Glass Theory and Beyond* (Singapore: World Scientific)

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

Peretto P 1984 *Biol. Cybern.* **50** 51

Rosenblatt F 1962 *Principles of Neurodynamics* (New York: Spartan)

Wallace D J 1985 *Advances in Lattice Theory* ed D W Duke and J F Owens (Singapore: World Scientific)