Statistical mechanics of Hopfield-like neural networks with modified interactions

1991 J. Phys. A: Math. Gen. 24 2419

(http://iopscience.iop.org/0305-4470/24/10/026)

J. Phys. A: Math. Gen. 24 (1991) 2419–2429. Printed in the UK

Statistical mechanics of Hopfield-like neural networks with modified interactions

Vik S Dotsenko†, N D Yarunin† and E A Dorotheyev‡

† L D Landau Institute for Theoretical Physics, Moscow, Kosygina 2, USSR
‡ N E Zhukovskii Central Airhydrodynamical Institute, Moscow, Zhukovskii, USSR

Received 4 September 1990

Abstract. Hopfield-like neural networks with modified interactions are studied by a mean-field theory. The modification of interactions is achieved during a special thermally noised iterative procedure. The resulting couplings have an intermediate form between the Hebb-like learning rule and the pseudo inverse one. Replica-symmetric free energy of the model is obtained. Statistical properties of the model depend on three parameters: reduced number of the stored patterns α, reduced number of iteration steps of the modification procedure λ and the temperature T. The phase diagram in the space of these parameters is obtained. The network can retrieve patterns at $T=0$ for $\alpha < \alpha_c(\lambda)$, where $\alpha_c(0) \simeq 0.14$ and $\alpha_c(\lambda \to \infty) \simeq 1.07$. As α decreases below $\alpha_0(\lambda)$ the FM retrieval states become ground states of the system, where $\alpha_0(0) = 0.05$ and $\alpha_0(\lambda \to \infty) = 2/\pi$.

1. Introduction

Among recent studies of neural networks there has been an upsurge of interest in the problem of finding the learning algorithm which would provide maximal storage capacity and maximal quality of retrieval.

Here we consider the fully connected network consisting of N Ising spins $\{\sigma\}$ $(i = 1, \ldots, N)$ and symmetric couplings J_{ij} which is supposed to store P uncorrelated patterns $\{\xi_i^{(\mu)}\}$ $(\mu = 1, \ldots, P)$. The model is studied in the limit when both $N \to \infty$ and $P \to \infty$ while the parameter $\alpha = P/N$ remains finite. Such system can be described by a Hamiltonian and it could be studied in terms of usual statistical mechanics. A classical example is the model introduced by Hopfield (1982) in which the interactions are defined according to the Hebbian rule:

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

The mean-field solution (Amit *et al* 1987) shows that below the line $\alpha_c(T)$ ($\alpha_c(T = 0) = 0.14$ and $\alpha_c(T \to 0) = 0$) the free energy has minima in which the spin states have non-zero macroscopic overlaps with one of the patterns. This system was shown to be highly robust in many respects, however the capacity $\alpha_c = 0.14$ for non-correlated patterns is obviously far from the maximum possible. The clear indication of this is given by the procedure of *unlearning* (Kleinfeld and Pendergraft (1987) and van Hemmen *et al* (1989)) which via special iterative redefinition of $J'_{i,j}$s makes possible a notable increase of a maximal capacity.

The other example is the model in which the couplings are defined in an essentially *non-local* way according to the pseudo inverse learning rule (Kohenen 1984):

$$J_{i,j}=\frac{1}{N}\sum_{\mu,\nu}\xi_i^{(\mu)}\xi_j^{(\nu)}(C)^{-1}_{\mu\nu} \tag{2}$$

where $C_{\mu\nu}$ is the overlap matrix:

$$C_{\mu\nu}=\frac{1}{N}\sum_i \xi_i^{(\mu)}\xi_i^{(\nu)}. \tag{3}$$

This model has been studied by Personaz *et al* (1985) and Kanter and Somplinsky (1987). The maximal storage capacity of such system storing uncorrelated patterns was shown to be $\alpha_c=1$ (at $\alpha>1$ the vectors $\xi^{(\mu)}$ cannot be linearly independent) and at $T=0$ the overlaps with the sorted patterns in the retrieval state is equal to 1 for all $\alpha<1$. However, the learning rule (2) 'breaks the rules of the game' since the couplings given by (2) are not defined via a *local* learning procedure.

Here we propose the model in which the couplings are defined via a *local* iterative learning procedure (section 2). It was noted on different occasions that introducing noise in the learning one can make things better (Gardner *et al* 1989, Wong and Sherrington 1990). Here we use *thermal* noise during the course of training. The resulting couplings could be defined explicitly, and in the course of iterations they are getting more and more close to those defined by the pseudo inverse learning rule (2).

In section 3 the mean-field solution for this model is obtained and replica symmetric free energy is calculated. The full phase diagram of the system in the space of parameters T, α and λ (where λ is the reduced number of iteration steps) is obtained in section 4. In section 5 the effect of replica symmetry breaking is briefly discussed. Since at $T=0$ the replica symmetry breaking is getting more and more strong with λ and α increasing, the results given by the replica symmetric solution may become unreliable at large λ.

2. The model

The model is formulated in the following way. We consider the system of N Ising spins $[\sigma_i]$ $(i=1,\ldots,N)$ which is described by the Hamiltonian:

$$H=-\tfrac{1}{2}\sum_{i,j}J_{i,j}\sigma_i\sigma_j. \tag{4}$$

For a given set of P non-correlated patterns $[\xi_i^{(\mu)}]$ $(\mu=1,\ldots,P)$ the couplings $J_{i,j}$ are defined according to the following iterative procedure. The starting values of the couplings are defined according to the Hebbian rule (1):

$$J_{i,j}(t=0)\equiv J_{i,j}^{(0)}=\frac{1}{N}\sum_{\mu=1}^{P}\xi_i^{(\mu)}\xi_j^{(\mu)}. \tag{5}$$

At each iteration time step t we impose a strong magnetic field h_i and define the quantity

$$f_t[h_i]=-\frac{1}{\beta}\log\left(\sum_{\sigma_i}\exp\left(-\tfrac{1}{2}\beta\sum_{i,j}J_{i,j}(t)\sigma_i\sigma_j-\beta\sum_i h_i\sigma_i\right)\right). \tag{6}$$

Here $\beta h \gg 1$ and $h \gg J(t)$ (h and $J(t)$ are the characteristic values of the fields and couplings) so that at the zero approximation $\langle\sigma_i\rangle \simeq \tanh(\beta h_i)$. Then we define

$$\chi_{i,j} = \frac{\delta^2 f_t}{\delta h_i \delta h_j} = \langle\sigma_i\sigma_j\rangle - \langle\sigma_i\rangle\langle\sigma_j\rangle. \tag{7}$$

Finally, taking

$$h_i = h\zeta_i \tag{8}$$

where $\zeta_i = \pm 1$ is an *arbitrary* fixed pattern, define a new couplings

$$J_{i,j}(t+1) = \frac{\chi_{i,j}}{(\beta(1-s^2))^2} \tag{9}$$

where $s = \tanh(\beta h)$.

Using the above definitions one easily finds the iteration equation explicitly:

$$J_{i,j}(t+1) = J_{i,j}(t) - \varepsilon \sum_k J_{i,k} J_{k,j} + \mathrm{O}(\varepsilon^2) \tag{10}$$

where $\varepsilon = \beta \cosh^{(-2)}(\beta h) \ll 1$.

Note, that although the iteration equation in its explicit form (10) is non-local, the iteration procedure itself can be formulated via *local* thermally noised process, (7)–(9).

In simple terms it could be interpreted as follows. We are fixing some arbitrary spin state $[\zeta_i]$ and then for a given spin–spin couplings $J_{i,j}(t)$ we allow thermal fluctuations in a close vicinity $\varepsilon \ll 1$ near this state. After that, a new Hamiltonian which contains new 'renormalized' couplings $J_{i,j}(t+1)$ is defined according to the ideology of the renormalization-group (RG) theory:

$$\exp\left(-\beta \sum_{i,j} J_{i,j}(t+1)\zeta_i\zeta_j\right) = \sum_{|\sigma-\zeta|<\varepsilon} \exp\left(-\beta \sum_{i,j} J_{i,j}(t)\sigma_i\sigma_j\right). \tag{11}$$

Using this RG equation one can easily get the iteration equation (10).

Now we explain why one can hope that it will make things better. Equation (10) can be easily solved to give:

$$J_{i,j}(t) = \sum_k J^{(0)}_{i,k}(1+\lambda \boldsymbol{J}^{(0)})^{(-1)}_{k,j} \tag{12}$$

where $\lambda = \varepsilon t$. Using (5) one gets:

$$J_{i,j}(t) = \sum_{\mu\nu} \xi_i^{(\mu)}(1+\lambda \boldsymbol{C})^{(-1)}_{\mu\nu}\xi_j^{(\nu)} \tag{13}$$

where the matrix $\boldsymbol{C}$ is given by (3).

Obviously, with increasing λ the structure of the couplings defined by (13) is getting close to that of the pseudo inverse couplings, (2). Using signal-to-noise analysis it can also be shown that at λ small and positive, the ratio of the characteristic value of the signal to that of the noise, grows with λ increasing.

Such a learning rule, (13), which has an intermediate form between the Hebbian, (1), and the pseudo inverse, (2), ones was studied for perceptron-type networks by Réfrégier and Vignolle (1989). They have shown that it provides the improvement of the generalization rate while preserving good learning performances.

3. The mean-field solution

In this section we calculate the free energy f of our system, (4) and (13), averaged over the random patterns $[\xi_i^{(\mu)}]$ using a standard replica trick:

$$-\beta N f = \lim_{n\to 0} \frac{\langle\langle Z^n \rangle\rangle - 1}{n}. \tag{14}$$

Here n is the number of the replicas, $\langle\langle \ldots \rangle\rangle$ is the averaging over the random patterns and

$$Z^n = \prod_{\rho=1}^{n} \sum_{\sigma_i^\rho} \exp\left(\tfrac{1}{2}\beta \sum_\rho \sum_{i,j} \sum_{\mu\nu} \sigma_i^\rho \xi_i^\mu (1+\lambda \boldsymbol{C})^{(-1)}_{\mu\nu} \sigma_j^\rho \xi_j^\nu \right). \tag{15}$$

Introducing the fields a_μ^ρ and ϕ_i^ρ one gets:

$$Z^n = \int \mathscr{D}\boldsymbol{a} \int \mathscr{D}\phi \sum_\sigma \exp\left(-\tfrac{1}{2}\beta N \sum_{\mu\rho} (a_\mu^\rho)^2 - \tfrac{1}{2}\beta\lambda \sum_{\rho i} (\phi_i^\rho)^2 + \beta \sum_{\mu\rho i} a_\mu^\rho \xi_i^{(\mu)} (\sigma_i^\rho + \mathrm{i}\lambda\phi_i^\rho) \right) \tag{16}$$

(here the term, containing $\det(1+\lambda \boldsymbol{C})$, which contributes an irrelevant constant into the free energy is omitted).

Then, following standard calculations similar to those of the Hopfield model (see e.g. Amit *et al* (1987)), one arrives at the following expression:

$$\langle\langle Z^n \rangle\rangle = \prod_\rho \left(\int \mathrm{d}a^\rho \right) \int \mathscr{D}\boldsymbol{Q} \int \mathscr{D}\boldsymbol{R} \exp\{-\beta N n f(\boldsymbol{a}, \boldsymbol{Q}, \boldsymbol{R})\} \tag{17}$$

where

$$\begin{aligned} -\beta N n f(\boldsymbol{a}, \boldsymbol{Q}, \boldsymbol{R}) &= -\tfrac{1}{2}\beta N \sum_\rho (a^\rho)^2 - \tfrac{1}{2}\alpha N \operatorname{Tr}\log(1-\beta \boldsymbol{Q}) - \tfrac{1}{2}\alpha\beta^2 N \sum_{\rho\gamma} R^{\rho\gamma} Q^{\rho\gamma} \\ &\quad + N \left\langle\!\!\left\langle \log\left(\prod_\rho \left(\sum_{\sigma^\rho} \int \mathrm{d}\phi^\rho \right) \right) \exp\left(-\tfrac{1}{2}\beta\lambda \sum_\rho (\phi^\rho)^2 + \beta \sum_\rho a^\rho \xi (\sigma^\rho + \mathrm{i}\lambda\phi^\rho) \right.\right.\right. \\ &\quad \left.\left.\left. + \tfrac{1}{2}\alpha\beta^2 \sum_{\rho\gamma} R^{\rho\gamma} (\sigma^\rho + \mathrm{i}\lambda\phi^\rho)(\sigma^\gamma + \mathrm{i}\lambda\phi^\gamma) \right) \right\rangle\!\!\right\rangle. \end{aligned} \tag{18}$$

Here

$$Q^{\rho\gamma} = \frac{1}{N} \sum_i (\sigma_i^\rho + \mathrm{i}\lambda\phi_i^\rho)(\sigma_i^\gamma + \mathrm{i}\lambda\phi_i^\gamma) \tag{19}$$

$R^{\rho\gamma}$ are the variables conjugate to (19) and $a^\rho \equiv a^\rho_{\mu=1}$, i.e. the pattern number one $\xi_i^{(\mu=1)} \equiv \xi_i$ is expected to condense.

Assuming the replica symmetry one takes:

$$R^{\rho\gamma} = \begin{cases} R & \text{if } \rho \neq \gamma \\ R_0 & \text{if } \rho = \gamma \end{cases} \tag{20}$$

$$Q^{\rho\gamma} = \begin{cases} Q & \text{if } \rho \neq \gamma \\ Q_0 & \text{if } \rho = \gamma \end{cases} \tag{21}$$

and $a^\rho = a$.

Then, taking the limit $n \to 0$ after some algebra one gets:

$$f = \tfrac{1}{2}a^2 + \tfrac{1}{2}\alpha\beta(R_0Q_0 - RQ) + \frac{\alpha}{2\beta}\left(\log(1-\beta(Q_0-Q)) - \frac{\beta Q}{1-\beta(Q_0-Q)}\right)$$
$$- \frac{1}{\beta}\int_{-\infty}^{+\infty} \mathrm{d}z\, \exp\left(-\frac{z^2}{2}\right)\left\langle\!\!\left\langle \log\left(\sum_\sigma \int_{-\infty}^{+\infty} \mathrm{d}\phi\, \exp\{-\beta H[\sigma, \phi; z]\}\right)\right\rangle\!\!\right\rangle \tag{22}$$

where

$$-\beta H[\sigma, \phi; z] = -\tfrac{1}{2}\beta\lambda\Delta(\phi - \phi_0(\sigma))^2 + \frac{\beta}{\Delta}(a\xi + \sqrt{\alpha R}\, z). \tag{23}$$

Here

$$\phi_0(\sigma) = \frac{\mathrm{i}}{\Delta}\left(a\xi + \sqrt{\alpha R}\, z + \frac{\Delta - 1}{\lambda}\sigma\right) \tag{24}$$

and

$$\Delta = 1 + \lambda\alpha\beta(R_0 - R). \tag{25}$$

Finally, from (22) one obtains:

$$f(a, Q_0, Q, R, \Delta)$$
$$= \tfrac{1}{2}\left(1 + \frac{\lambda}{\Delta}\right)a^2 + \frac{\alpha}{2\beta}\left(\log(1-\beta(Q_0-Q)) - \frac{\beta Q}{1-\beta(Q_0-Q)}\right)$$
$$+ \tfrac{1}{2}\alpha\beta R(Q_0 - Q) + \frac{\Delta - 1}{2\lambda}Q_0 + \frac{\lambda\alpha R}{2\Delta} + \frac{1-\Delta}{2\lambda\Delta} + \frac{\log\Delta}{2\beta}$$
$$- \frac{1}{\beta}\overline{\log\cosh[(\beta/\Delta)(a\xi + \sqrt{\alpha R}\, z)]}. \tag{26}$$

Here $\overline{(\ldots)}$ means the averaging over ξ and Gaussian z.

The corresponding saddle-point equations for the variables a, Q_0, Q, R, and Δ are:

$$a = \frac{1}{\Delta + \lambda}\overline{[\xi\tanh(\beta/\Delta)(a\xi + \sqrt{\alpha R}\, z)]} \tag{27}$$

$$R = \frac{Q}{(1 - \beta(Q_0 - Q))^2} \tag{28}$$

$$\Delta = 1 + \frac{\lambda\alpha}{1 - \beta(Q_0 - Q)} \tag{29}$$

$$\beta(Q_0 - Q) = \frac{\beta}{\Delta^2}\overline{\cosh^{(-2)}(\beta/\Delta)(a\xi + \sqrt{\alpha R}\, z)} - \frac{\lambda}{\Delta} \tag{30}$$

$$Q_0\Delta^2 = 1 + \lambda^2\alpha R - \lambda(\lambda + 2\Delta)a^2 - \frac{\lambda\Delta}{\beta}$$
$$- 2\lambda\alpha R\frac{\beta}{\Delta}\overline{\cosh^{(-2)}(\beta/\Delta)(a\xi + \sqrt{\alpha R}\, z)}. \tag{31}$$

Note that the variable a is directly connected with the overlap

$$m=\frac{1}{N}\sum_i \xi_i^{(1)}\langle\sigma_i\rangle \tag{32}$$

of the thermodynamic state with the pattern. According to (16):

$$a=\frac{1}{N}\sum_\nu\sum_i (1+\lambda C)^{(-1)}_{1\nu}\xi_i^{(\nu)}\langle\sigma_i\rangle \tag{33}$$

and therefore in the retrieval state

$$m=(1+\lambda)a. \tag{34}$$

4. The phase diagram

4.1. Zero temperature

At zero temperature (27)–(31) can be reduced to:

$$a=\frac{\operatorname{erf}(a/\sqrt{2\alpha R})}{\Delta+\lambda} \tag{35}$$

$$R=\frac{Q_0}{(1-C)^2} \tag{36}$$

$$\Delta=1+\frac{\lambda\alpha}{1-C} \tag{37}$$

$$C=\frac{1}{\Delta}\sqrt{\frac{2}{\pi\alpha R}}\exp\left(-\frac{a^2}{2\alpha R}\right)-\frac{\lambda}{\Delta} \tag{38}$$

$$Q_0\Delta=1+\lambda^2\alpha R-\lambda(\lambda+2\Delta)a^2-2\lambda\alpha\sqrt{\frac{2R}{\pi\alpha}}\exp\left(-\frac{a^2}{2\alpha R}\right). \tag{39}$$

Here $\operatorname{erf}(x)$ is the error function:

$$\operatorname{erf}(x)=\frac{2}{\sqrt{\pi}}\int_0^x \mathrm{d}t\,\exp(-t^2).$$

Introducing $y=a/\sqrt{2\alpha R}$ and $x=1/\sqrt{R}$ and excluding Q_0 one obtains:

$$y=\frac{x}{\Delta+\lambda}\frac{\operatorname{erf}(y)}{\sqrt{2\alpha}} \tag{40}$$

$$\Delta=1+\frac{\lambda\alpha}{1-C} \tag{41}$$

$$C\Delta=x\sqrt{\frac{2}{\pi\alpha}}\exp(-y^2)-\lambda \tag{42}$$

$$\Delta^2(1-C)^2=x^2+\lambda^2\alpha-2\lambda\alpha(\lambda+2\Delta)y^2-2\lambda\alpha x\sqrt{\frac{2}{\pi\alpha}}\exp(-y^2). \tag{43}$$

The results of the numerical solution of the above equations are shown in figure 1. The solutions with $y\neq 0$ which correspond to the retrieval states exist only for

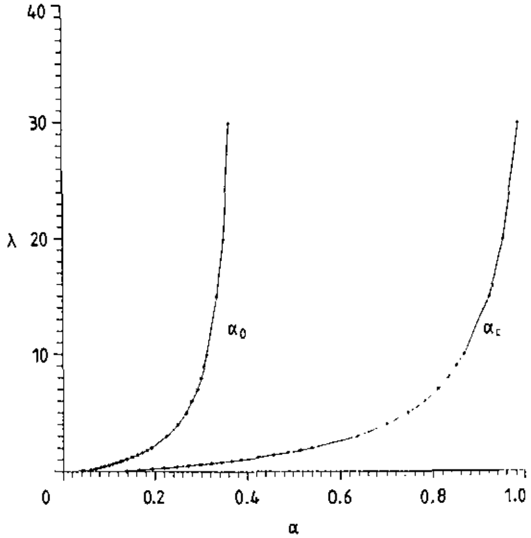

Figure 1. Zero-temperature phase diagram of the system.

$\alpha < \alpha_c(\lambda)$. The curve $\alpha_c(\lambda)$ starts at the point $\alpha_c(0) \simeq 0.14$ which correspond to the usual Hopfield model, and for $\lambda \to \infty$ asymptotically approaches the point $\alpha_c(\infty) \simeq 1.07$.

In the region $\alpha_0(\lambda) < \alpha < \alpha_c(\lambda)$ the retrieval states are metastable and the spin-glass solutions with $y = 0$ has lower energy. For $\alpha < \alpha_0(\lambda)$ the retrieval states are becoming global minima. Both $\alpha_c(\lambda)$ and $\alpha_0(\lambda)$ are the lines of the phase transitions of the first order.

The dependence of the overlaps (32), from α in the retrieval states for different values of λ is shown in figure 2. At $\alpha = \alpha_c(\lambda)$ the overlap m discontinuously jumps to zero.

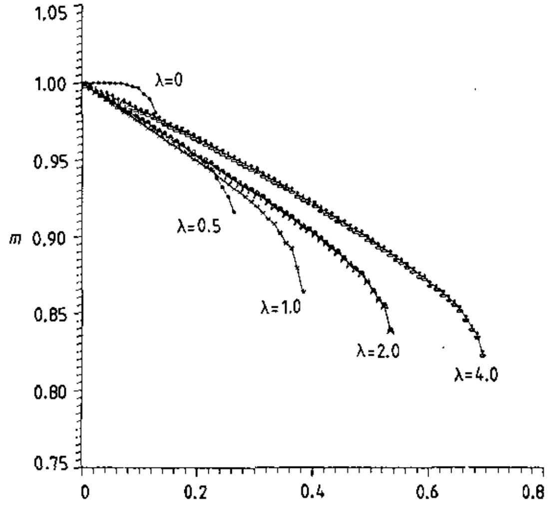

Figure 2. Dependence of the overlaps in the retrieval states from α for $\lambda = 0$; 0.5; 1; 2 and 4 at $T = 0$.

4.2. *Finite temperatures*

The results of the numerical solution of (27)–(31) at $T \neq 0$ are summarized in figure 3, where the curves of the phase transition $T_c(\alpha)$ are shown for several values of λ. The solutions with $y \neq 0$ exist only below the curve $T_c(\alpha)$. The dependence of the value of the overlaps at the critical curve $T_c(\lambda)$ from the reduced temperature T/T_c is shown in figure 4. Note that this dependence exhibits a sort of scaling as λ increases. The dependence of the value of the overlaps at the critical curve on α for several values of λ is shown in figure 5.

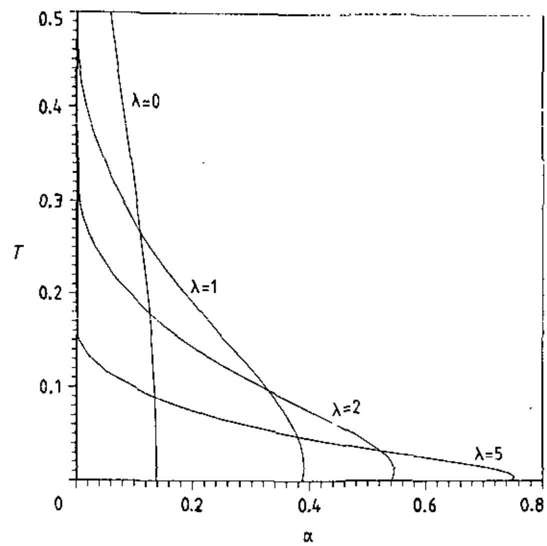

Figure 3. Critical temperature $T_c(\lambda)$ for $\lambda = 1, 2$ and 5.

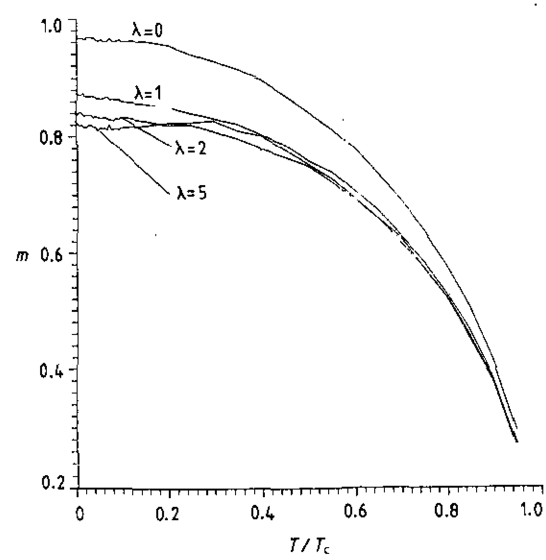

Figure 4. Dependence of the values of the overlaps at the critical curve $T_c(\lambda)$ from the reduced temperature T/T_c.

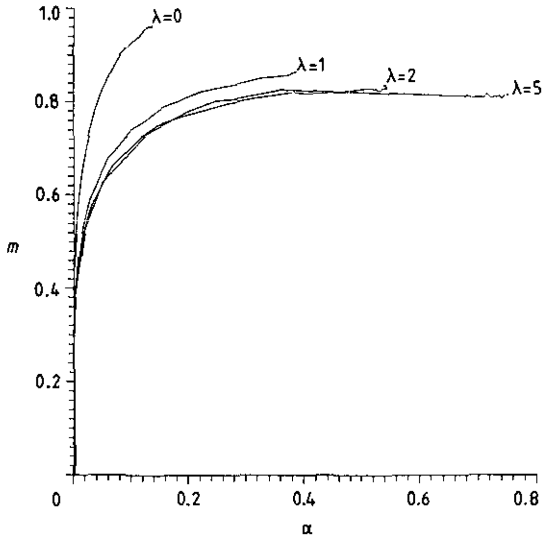

Figure 5. Dependence of the value of the overlaps at the critical curve from α for $\lambda=0$; 1; 2 and 5.

Critical temperature $T_c=1/(1+\lambda)$ is the point of the phase transition of the second order and the behaviour of the curves $\alpha_c(T)$ and $\alpha_0(T)$ near T_c can be found analytically. Expanding (27)-(31) over small α, a, R, Q, Q_0 and $t=T_c-T$ and introducing

$$\tau=\frac{1+\lambda}{\sqrt{\alpha}}\,t$$

$$y=\frac{a^2}{\sqrt{\alpha}}$$

after some algebra one obtains:

$$\tau=(\Delta+\lambda)^2(\tfrac{1}{3}y+\sqrt{\alpha}\,R) \tag{44}$$

$$\sqrt{\alpha}\,R=\frac{y+\sqrt{\alpha}\,R}{[((\Delta+\lambda)^3/\Delta)(y+\sqrt{\alpha}\,R)-((\Delta+\lambda)/\Delta)\tau]^2} \tag{45}$$

$$\Delta=1+\frac{\lambda\sqrt{\alpha}}{((\Delta+\lambda)^3/\Delta)(y+\sqrt{\alpha}\,R)-((\Delta+\lambda)/\Delta)\tau}. \tag{46}$$

It can be shown that the solutions of these equations with $a\neq 0$ exist only at $T<T_c(\alpha)$ where

$$T_c(\alpha)=T_c-\tau_0\frac{1+x_0\lambda\sqrt{\alpha}}{(1+\lambda)^2}\sqrt{\alpha} \tag{47}$$

where $\tau_0\simeq 1.95$ and $x_0\simeq 0.43$.

The curve $T_0(\alpha)$ is defined by the condition that the free energy of the solution of (44)-(46) with $a\neq 0$ becomes equal to the SG solution with $a=0$. For λ not very large: $\lambda\ll 1/\sqrt{\alpha}$, one obtains simple equations:

$$f(a)-f(0)=\tfrac{1}{2}T_c^3\alpha\left(\tfrac{1}{4}z^2-\tau'z+(1+\beta_c^2)\tau'+\tfrac{1}{2}+\beta_c^2\log z-\beta_c^2\frac{\tau'}{z}\right)=0 \tag{48}$$

$$\tfrac{1}{2}z^3-\tau'z^2+z+\tau'=0 \tag{49}$$

where

$$z = \tfrac{2}{3}(1+\lambda)^3 \frac{a^2}{\sqrt{\alpha}}$$

$$\tau' = \frac{(1+\lambda)^2}{\sqrt{\alpha}} t$$

(if $\lambda \ll 1/\sqrt{\alpha}$ equations (44)–(46) can be reduced into one equation (49)).

The solution of (48)–(49) gives the curve $T_0(\alpha)$:

$$T_0(\alpha) = T_c - \tau(\lambda)\sqrt{\alpha} \tag{50}$$

where

$$\tau(\lambda) \simeq 2.6(1-0.93\lambda) \tag{51}$$

for $\lambda \ll 1$, and

$$\tau(\lambda) \simeq 1 \tag{52}$$

for $\lambda \gg 1$ ($\lambda \ll 1/\sqrt{\alpha}$).

Therefore both curves $T_c(\alpha)$ and $T_0(\alpha)$ are getting up as λ increases, although the interval of temperatures where the retrieval is possible is getting smaller.

5. Discussion

A general approach for finding limits for the maximal storage capacity in neural networks was proposed by Gardner (1988, 1989). Its major advantage is that one does not need to know an explicit form of the learning rule. For a given set of P patterns $\xi_i^{(\mu)}$ in a system consisting of N Ising spins the couplings are chosen with the only condition that they provide stability for all patterns, i.e. the local fields $h_i = \Sigma_j J_{i,j}\xi_j^{\mu}$ are parallel to the patterns: $\xi_i^{\mu} h_i > k > 0$ in each site. It was shown that if all $J'_{i,j}$s are independent then for uncorrelated patterns such couplings exist only for $P < \alpha(k)N$, where $\alpha(k \to 0) = 2$ (if patterns are correlated, then $\alpha(k)$ grow with the degree of correlations). Moreover, a simple iterative learning algorithm was proved to converge to those couplings in a finite number of steps (provided that such couplings exist).

The question is to what extent the above results describe real retrieval properties of the networks. The doubts are in the following.

Although above $\alpha(k)$ one cannot provide stability of the patterns in each site, it is not impossible that, say at finite temperatures, the system would provide satisfactory retrieval with a finite percentage of errors and the phase transition to the phase with no retrieval at all occurs at quite a different value of α. In Gardner's approach one may also find a critical value of α allowing a finite percentage of errors, but the question is what percentage of errors is permissible?

On the other hand, although below $\alpha(k)$ one can be sure that all the patterns are stable, one can only hope that at finite $k > 0$ the patterns have a finite basin of attraction (in other words, which is the critical value of k (if any), which would provide a finite basin of attraction?)

The answers to the above questions have been found recently by Amit *et al* (1990) for the diluted networks with synaptic couplings taken to be optimal (at $\alpha = \alpha(k)$). The dynamical equations describing an evolution of the overlap of the current spin state with the given pattern has been solved numerically and the full phase diagram on the plane (α, T) has been obtained.

The maximal capacity for fully connected symmetric network is still not known, although there is a variant of the perceptron algorithm which allows us to find symmetric couplings whenever it is possible (Gardner 1988).

Here we have considered the fully connected Hopfield-like neural network with symmetric couplings which have an intermediate form between the Hebb learning rule, (1), and the pseudo inverse one, (2). An iterative thermally noised algorithm was proposed which makes it possible to obtain the modified couplings explicitly. It was shown that it provides a substantial increase of a storage capacity and the quality of retrieval.

The main results of the present study confirm a general idea that to improve the capacity and the functioning of a neural network one has to introduce some sort of a noise in the process of learning. Intuitively it seems that the noise shakes down the memories and makes them adjust to each other much better.

A somewhat puzzling result of the present study is that as $\lambda \to \infty$ when the structure of the interaction matrix J_{ij}, (13), becomes equivalent to that of the pseudo inverse model, (2) (Kanter and Sompolinsky 1987), the maximum storage capacity at the zero temperature is equal to 1.07 and not to 1. Actually this discrepancy is a consequence of the fact that the replica-symmetric solution found in this paper is not stable at large λ. The fact that the replica-symmetric solution is not stable against the replica symmetry breaking is clearly indicated by the zero-temperature entropy:

$$S = -\frac{\partial f}{\partial T}\bigg|_{T=0} = -\frac{\alpha}{2}\left(\frac{C}{1-C} + \log(1-C)\right) - \tfrac{1}{2}\log\Delta. \tag{53}$$

At $\lambda = 0$ (in the usual Hopfield model) the entropy is negative but its value is small (at $\alpha = \alpha_c$, $S \simeq 0.001$), which indicates that the replica-symmetry breaking in the retrieval states is very weak (Amit *et al* 1987). However, the solutions of (40)–(43) show that as λ (and α) increases, the value of the negative entropy, (53), also increases, indicating that the replica-symmetry breaking could become strong. Nevertheless, the replica-symmetric solution could become stable at finite temperatures. The detailed study of the replica-symmetry breaking in the considered model as well as the results of the numerical simulations will be reported elsewhere.

References

Amit D J, Evans M R, Horner H and Wong K Y M 1990 *J. Phys. A: Math. Gen.* **23** 3361–81

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

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

—— 1989 *J. Phys. A: Math. Gen.* **22** 1969

Gardner E, Stroud N and Wallace D J 1989 *J. Phys. A: Math. Gen.* **22** 2019

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

Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380

Kleinfeld D and Pendergraft D B 1987 *Biophys. J.* **51** 47

Kohonen T O 1984 *Self-Organization and Associative Memory* (Berlin: Springer)

Personaz L, Guyon I and Dreyfus G 1985 *J. Physique Lett.* **46** L359

Réfrégier Ph and Vignolle J-M 1989 *Europhys. Lett.* **10** 387

van Hemmen J L, Ioffe L B, Kuhn R and Vass M 1989 *Preprint*

Wong K and Sherrington D 1990 *J. Phys. A: Math. Gen.* **23** 4659