Modified pseudo-inverse neural networks storing correlated patterns

1992 J. Phys. A: Math. Gen. 25 2843

(http://iopscience.iop.org/0305-4470/25/10/013)

J. Phys. A: Math. Gen. 25 (1992) 2843-2857. Printed in the UK

Modified pseudo-inverse neural networks storing correlated patterns

R Der†, V S Dotsenko‡ and B Tirozzi

Dipartimento di Matematica, Universita' di Roma 'La Sapienza', Piazzale Aldo Moro 2, 00185 Roma, Italy

Received 22 July 1991

Abstract. Neural networks with symmetric couplings which have an intermediate form between the Hebb learning rule and the pseudo-inverse one, storing strongly correlated patterns, are studied. Signal-to-noise analysis is made and replica-symmetric thermodynamic calculations are performed. Both approaches show that both in the Hopfield model limit and in the pseudo-inverse model limit the maximal capacity of the order of $(2p/\ln(1/p)^{-1}$ (where $p \ll 1$ is the average neural activity) can be achieved by appropriate adjustment of the threshold term of the Hamiltonian.

1. Introduction

In this paper we consider the problem of the maximal capacity of fully connected neural networks consisting of N ($N \to \infty$) binary units with symmetric couplings.

We consider the model in which the couplings depend on a continuous parameter λ such that it could be considered as an intermediate between two well studied systems: the Hopfield model (see e.g. Amit *et al* 1987) and the so-called pseudo-inverse model (Personaz *et al* 1985, Kanter and Sompolinsky 1987).

In the Hopfield model storing $M = \alpha N$ uncorrelated patterns the couplings are defined as follows:

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

where (ξ_i^μ) are the stored patterns. In the pseudo-inverse model the couplings are

$$J_{ij}^{(\mathrm{PS})} = \frac{1}{N} \sum_{\mu,\nu=1}^{M} \xi_i^\mu (C^{-1})_{\mu\nu} \xi_j^\nu \tag{2}$$

where

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

In the model under consideration the couplings are defined according to the rule:

$$J_{ij} = \frac{1}{N} \sum_{\mu\nu}^{M} \xi_i^\mu (\hat{1} + \lambda \hat{C})_{\mu\nu}^{-1} \xi_j^\nu. \tag{4}$$

† On leave from Leipzig University, Federal Republic of Germany.

‡ On leave from Landau Institute for Theoretical Physics, Academy of Sciences of the USSR, Moscow, Russia.

0305-4470/92/102843+15$04.50

For $\lambda = 0$ we recover the Hopfield model (1) and in the limit $\lambda \to \infty$ the structure of the J_{ij} tends to the pseudo-inverse one (2). For the case of uncorrelated patterns such a system has already been studied by Dotsenko *et al* (1991). It was shown that the maximal capacity of the model α_c depends continuously on the parameter λ, giving the value $\alpha_c = 0.14$ of the Hopfield model (Amit *et al* 1987) for $\lambda = 0$, and getting close to $\alpha_c = 1$ of the pseudo-inverse model (Kanter and Sompolinsky 1987) for $\lambda \to \infty$.

Here we address the question: what should be the structure of such a model and what is its maximal capacity in the case of strong correlations among the stored patterns? The Hopfield model storing strongly correlated patterns has been considered by Tsodyks and Feigelmann (1988) and Buhmann *et al* (1989). They showed that the maximal capacity:

$$\alpha_c^{(\max)} \sim (2p \ln(1/p))^{-1} \tag{5}$$

predicted by Gardner (1988) is achieved if one considers the system consisting of binary units v_i taking values 0 and 1 and described by the Hamiltonian:

$$H = -\frac{1}{2}\sum_{ij} J_{ij} v_i v_j + u \sum_i v_i \tag{6}$$

where

$$J_{ij} = \frac{1}{N}\sum_{\mu=1}^{M} (\eta_i^\mu - p)(\eta_j^\mu - p). \tag{7}$$

Here the (η_i^μ) are the patterns described by the probability distribution:

$$P(\eta) = p\delta(\eta - 1) + (1-p)\delta(\eta)$$

and the parameter $p \ll 1$ describes the average activity of the neurons: $\langle\langle \eta \rangle\rangle = p$.

It was shown that the result (5) is obtained if one adjusts the value of the threshold u in the Hamiltonian (6) such that $u = u_0 p$, where $u_0 = 1 - \delta$, and $p \ll \delta \ll 1$.

Straightforward generalization of the model (6), (7) will be done by taking the couplings in the form:

$$J_{ij} = \frac{1}{N}\sum_{\mu\nu}^{M} \xi_i^\mu (\hat{1} + \lambda \hat{C})^{-1}_{\mu\nu} \xi_j^\nu \tag{8}$$

where $\xi_i^\mu \equiv \eta_i^\mu - p$ and

$$C_{\mu\nu} = \frac{1}{N}\sum_i \xi_i^\mu \xi_i^\nu. \tag{9}$$

The random variables ξ_i^μ are described by the distribution:

$$P(\xi) = p\delta(\xi - (1-p)) + (1-p)\delta(\xi + p) \tag{10}$$

such that $\langle\langle \xi \rangle\rangle = 0$ and $\langle\langle \xi^2 \rangle\rangle = p(1-p)$. In the limit $\lambda = 0$ we recover the Tsodyks–Feigelmann model (6), (7), and in the limit $\lambda \to \infty$ we obtain the pseudo-inverse variant of the model storing strongly correlated patterns.

In section 2 we make a signal-to-noise analysis for the model (6) with the couplings given by (8). It will be shown that by appropriate adjustment of the threshold u the maximal capacity of the model reaches Gardner's value (5) for any λ including the pseudo-inverse limit $\lambda \to \infty$. It will also be shown that the 'pure' pseudo-inverse model, when $\lambda = \infty$ exactly, is a special point and in such a case $\alpha_c^{(\max)} = 1$.

In section 3 the results of section 2 will be confirmed by a direct replica-symmetric calculation of the free energy and the solutions of the corresponding saddle-point equations. In section 4 we discuss the obtained results.

2. Signal-to-noise analysis

The dynamics of the system is described by the equations:

$$v_i(t+1) = \theta\left(\sum_{i\neq j} J_{ij} v_j(t) - u\right). \tag{11}$$

The local field produced at site i by the spin configuration (v_i) is:

$$h_i = \sum_{i\neq j} J_{ij} v_j(t) - u. \tag{12}$$

In order to study the stability of the stored patterns one has to consider the local fields produced by these patterns. Consider the local field produced by e.g. the pattern number 1:

$$h_i^1 = \sum_{i\neq j} J_{ij} \eta_j^1 - u. \tag{13}$$

These fields can be represented in the form:

$$h_i^1 = s\xi_i^1 - u + R_i. \tag{14}$$

Here s is the signal term, and random noise term R_i contains the contribution from all the other patterns.

Using the definition for the J_{ij}, equation (8), one can compute (see the appendix) the value of the signal and of the noise explicitly:

$$s = \frac{1}{\lambda}\chi(1-\alpha\chi) \tag{15}$$

$$\langle\langle R^2 \rangle\rangle = \frac{p(1-p)}{\lambda^2}(1-\alpha\chi)^2(1-\chi)^2 \frac{\alpha\chi^2}{1-\alpha\chi^2} \tag{16}$$

where

$$\chi = \frac{\lambda}{\alpha} \operatorname{Tr} \hat{A}(\hat{1}+\lambda\hat{A})^{-1} \tag{17}$$

and $\hat{A}_{ij}$ is the Hopfield model interaction matrix:

$$A_{ij} = \frac{1}{N}\sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu. \tag{18}$$

The calculations (see the appendix) give:

$$\chi = \frac{1+\alpha+1/(\lambda p(1-p)) - \sqrt{(1+\alpha+1/(\lambda p(1-p)))^2 - 4\alpha}}{2\alpha} \tag{19}$$

The signal-to-noise ratio is then:

$$\frac{s}{\sigma} = \frac{1}{\sqrt{\alpha p(1-p)}}\varphi(\chi) \tag{20}$$

where

$$\varphi(\chi)=\frac{\sqrt{1-\alpha\chi^2}}{(1-\chi)}. \tag{21}$$

For $p \ll 1$ the maximum value of α will be shown to be large ($\sim 1/p$). Using the explicit expression for χ, equation (19), one can then easily show that the value of the factor $\varphi(\chi)$ for all values of λ is restricted by

$$1 \leqslant \varphi(\chi) \leqslant 1+1/\alpha. \tag{22}$$

Therefore in the first order approximation in p one may take $\varphi(\chi) \simeq 1$ and

$$\frac{s}{\sigma} \simeq \frac{1}{\sqrt{\alpha p}}. \tag{23}$$

The local field h_i^1 produced by the pattern 1, equation (14), can be described by the Gaussian distribution:

$$P(h|\xi^1)=\frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{(h-s\xi^1-u)^2}{2\sigma^2}\right) \tag{24}$$

The pattern ξ_i^1 can be said to be stored perfectly if, according to the dynamics (11), the equation

$$\eta_i^1=\theta(h_i^1) \tag{25}$$

is fulfilled at all sites. Taking into account the distribution of the local fields (24) and all the distribution of the ξ_i^1, equation (10), one can easily derive the total number of mistakes, the number of sites at which equation (25) is not fulfilled:

$$\varepsilon = Np\int_{-\infty}^{0} \mathrm{d}h P(h|\xi=1-p)+N(1-p)\int_{0}^{\infty} \mathrm{d}h P(h|\xi=-p) \tag{26}$$

or

$$\varepsilon = Np \operatorname{erfc}\left(\frac{(1-p)s-u}{\sigma}\right)+N(1-p)\operatorname{erfc}\left(\frac{ps+u}{\sigma}\right) \tag{27}$$

where

$$\operatorname{erfc}(x) \equiv \sqrt{\frac{2}{\pi}}\int_{x}^{\infty} \mathrm{d}x \exp\left(\frac{x^2}{2}\right). \tag{28}$$

In order to minimize the amount of errors one can adjust the value of the threshold u which is the free parameter of the model. The appropriate value of u can be obtained from the equation

$$\frac{\partial\varepsilon}{\partial u}=0$$

which corresponds to the minimization of ε with respect to u:

$$p\exp\left(-\frac{((1-p)s-u)^2}{2\sigma^2}\right)-(1-p)\exp\left(-\frac{(ps+u)^2}{2\sigma^2}\right)=0. \tag{29}$$

The solution of this equation gives the optimal value of the threshold:

$$u=s\left(\frac{1}{2}-p+\left(\frac{\sigma}{s}\right)^2\ln\left(\frac{1-p}{p}\right)\right) \tag{30}$$

Using this result, the total number of mistakes, equation (27), can be represented as:

$$\frac{\varepsilon}{N}=p\,\mathrm{erfc}\left[\frac{s}{\sigma}\left(\frac{1}{2}-\left(\frac{\sigma}{s}\right)^2\ln\left(\frac{1-p}{p}\right)\right)\right]+(1-p)\,\mathrm{erfc}\left[\frac{s}{\sigma}\left(\frac{1}{2}+\left(\frac{\sigma}{s}\right)^2\ln\left(\frac{1-p}{p}\right)\right)\right]. \qquad (31)$$

The pattern is getting unstable if ε/N reaches the value p. Assuming that $p \ll 1$, the condition $\varepsilon/N=p$ gives the limiting value for the signal-to-noise ratio:

$$\left(\frac{\sigma}{s}\right)^2\ln\frac{1}{p}\simeq\frac{1}{2} \qquad (32)$$

or, using (23),

$$\alpha_{\max}\simeq\frac{1}{2p\ln(1/p)} \qquad (33)$$

Therefore in the main order in p the maximal capacity of our model (6)-(8) does not depend on λ and coincides with that of the Hopfield model storing strongly correlated patterns (Tsodyks and Feigelmann 1988). The optimal value of threshold (30) depends on λ. From (15) and (19) (for $\alpha \gg 1$ and $p \ll 1$) one gets:

$$s\simeq\frac{p(1+\lambda p)}{(1+\lambda\alpha p)^2}.$$

Therefore the optimal value of the threshold at which the capacity may reach its maximal value (33) is:

$$u\simeq\frac{p(1-p)(1+\lambda p)}{\left(1+\dfrac{\lambda}{2\ln(1/p)}\right)^2}. \qquad (34)$$

If λ is not very large $\lambda p \ll 1$:

$$u\simeq\frac{p(1-p)}{\left(1+\dfrac{\lambda}{2\ln(1/p)}\right)^2}. \qquad (35)$$

In the limit $\lambda\to\infty$ when the structure of the couplings (8) becomes close to the pseudo-inverse model the threshold is:

$$u\simeq\frac{p(1-p)4(\ln(1/p))^2}{\lambda^2}. \qquad (36)$$

Let us remark that, with such an optimal choice of the threshold, equation (34), the α maximal (33) remains the same even in the limit $\lambda\to\infty$.

Note, however, that if one takes $\lambda=\infty$ exactly, then after the scaling $J_{ij}\to J_{ij}/\lambda$, for the signal (15) and for the noise (16) one gets:

$$s=\begin{cases}(1-\alpha) & \text{if } \alpha\leqslant 1\\ 0 & \text{if } \alpha\geqslant 1.\end{cases}$$

and $\sigma\equiv 0$. Since at $\alpha\geqslant 1$ there is no signal, such a 'pure' pseudo-inverse model can retrieve patterns only at $\alpha<1$. However for any $\lambda\neq\infty$ there is *non-zero signal* and there is non-zero noise and their ratio (23) remains finite for any α even if $\lambda\to\infty$.

3. The mean field solution

The model under consideration is described by the Hamiltonian:

$$H=-\frac{1}{2N}\sum_{i\neq j}^{N}\sum_{\mu\nu}^{M} v_i\xi_i^\mu(\hat{1}+\lambda\hat{C})^{-1}_{\mu\nu}\xi_j^\nu v_j+u\sum_i v_i \tag{37}$$

where the variables (v_i) take two values $v=0,1$, and the matrix $\hat{C}$ is defined by:

$$\hat{C}_{\mu\nu}=\frac{1}{N}\sum_i^N \xi_i^\mu\xi_i^\nu \tag{38}$$

and the random variables (ξ_i^μ) are described by the probability distribution:

$$P(\xi)=p\delta(\xi-(1-p))+(1-p)\delta(\xi+p). \tag{39}$$

The free energy is calculated using a standard replica trick:

$$F=-\frac{1}{\beta}\lim_{n\to 0}\frac{\langle\langle Z^n\rangle\rangle-1}{n}=-\frac{1}{\beta}\ln Z \tag{40}$$

where the replicated partition function is:

$$Z^n=\sum_{v_i^\rho}\exp\left(\frac{\beta}{2N}\sum_{i\neq j}^{N}\sum_{\mu\nu}^{M}\sum_{\rho=1}^{n} v_i^\rho\xi_i^\mu(\hat{1}+\lambda\hat{C})^{-1}_{\mu\nu}\xi_j^\nu v_j^\rho-\beta u\sum_{i\rho}v_i^\rho\right). \tag{41}$$

Introducing the field a_μ^ρ and Φ_i^ρ one gets:

$$Z^n=\int \mathrm{D}\boldsymbol{a}\int \mathrm{D}\boldsymbol{\Phi}\sum_{v_i^\rho}\exp\left(-\frac{\lambda\beta}{2}\sum_{i\rho}(\Phi_i^\rho)^2-\frac{\beta}{2}\sum_{\mu\rho}(a_\mu^\rho)^2\right)$$
$$\times\exp\left(+\frac{\beta}{\sqrt{N}}\sum_{i,\mu,\rho}a_\mu^\rho\xi_i^\mu(v_i^\rho+\mathrm{i}\lambda\Phi_i^\rho)-\beta\left(u+\frac{\alpha}{2\lambda}\chi\right)\sum_{i\rho}v_i^\rho\right) \tag{42}$$

where:

$$\frac{\alpha}{\lambda}\chi=\frac{1}{N}\operatorname{Tr}\hat{A}(\hat{1}+\lambda\hat{A})^{-1} \tag{43}$$

and $\hat{A}$ is the Hopfield model interaction matrix (18). (In (42) the term containing $\det(\hat{1}+\lambda\hat{C})$, which contributes an irrelevant constant into the free-energy, is omitted.) Following standard calculations (see e.g. Amit *et al* 1987 and Dotsenko *et al* 1991) one makes the average over all the patterns ξ_i^μ but one. We define the pattern which is expected to condense to $\xi_i^{\mu=1}\equiv\xi_i$, and correspondingly redefine $a_{\mu=1}^\rho\equiv a^\rho$.

Then introducing:

$$Q_{\gamma\rho}=\frac{1}{N}\sum_i^N(v_i^\rho+\mathrm{i}\lambda\Phi_i^\rho)(v_i^\gamma+\mathrm{i}\lambda\Phi_i^\gamma) \tag{44}$$

one obtains for the free-energy density:

$$-\beta Nnf(a,\hat{Q},\hat{R})$$
$$=\frac{1}{2}\beta N\sum_\rho(a_\rho)^2-\frac{1}{2}\alpha N\operatorname{Tr}\ln(\hat{1}-\beta p(1-p)\hat{Q})-\frac{\alpha\beta^2N}{2}\sum_{\rho\gamma}R_{\rho\gamma}Q_{\rho\gamma}$$
$$+N\left\langle\left\langle\ln\sum_{v^\rho}\int \mathrm{d}\Phi^\rho\exp\left(-\frac{\lambda\beta}{2}\sum_\rho(\Phi^\rho)^2+2\beta\sum_\rho a^\rho\xi(v^\rho\mathrm{i}\lambda\Phi^\rho)\right)\right.\right.$$

$$\times\exp\left(-\beta\left(u+\frac{\alpha}{2\lambda}\chi\right)\sum_\rho v^\rho+\frac{\alpha\beta^2}{2}\sum_{\rho\gamma}R_{\rho\gamma}(v^\rho+\mathrm{i}\lambda\Phi^\rho)(v^\gamma+\mathrm{i}\lambda\Phi^\gamma)\right)\right\rangle\!\!\right\rangle_\xi \tag{45}$$

where the $R_{\rho\gamma}$ is the matrix conjugate to (44) and $\langle\!\langle\ldots\rangle\!\rangle_\xi$ means the average over the pattern ξ.

Assuming the replica symmetry:

$$Q_{\rho\gamma}=\begin{cases}Q & \rho\neq\gamma\\ Q_0 & \rho=\gamma\end{cases} \tag{46a}$$

$$R_{\rho\gamma}=\begin{cases}R & \rho\neq\gamma\\ R_0 & \rho=\gamma\end{cases} \tag{46b}$$

and introducing:

$$\Delta=1+\lambda\alpha\beta(R_0-R) \tag{47}$$

$$C=\beta(Q_0-Q) \tag{48}$$

one finally gets:

$$\begin{aligned}f(a,Q_0,C,R,\Delta)&=\frac{1}{2}\left(1+\frac{\lambda p(1-p)}{\Delta}\right)a^2+\frac{\alpha}{2\beta}\ln(1-p(1-p)C)-\frac{\alpha}{2}\frac{p(1-p)(Q_0-C/\beta)}{1-p(1-p)C}\\&\quad+\frac{1}{2}\alpha RC+\frac{\Delta-1}{2\lambda}Q_0+\frac{\lambda\alpha R}{2\Delta}+\frac{\ln\Delta}{2\beta}\\&\quad-\frac{1}{\beta}\left\langle\!\!\left\langle\overline{\ln\left[1+\exp\left(\frac{\beta}{\Delta}(a\xi+\sqrt{\alpha R}z)-\beta u+\frac{\beta(\Delta-1)}{2\lambda\Delta}-\frac{\beta\alpha}{2\lambda}\chi\right)\right]}\right\rangle\!\!\right\rangle_\xi\end{aligned} \tag{49}$$

where $\overline{(\ldots)}$ means Gaussian averaging over z.

In the zero-temperature limit ($\beta\to\infty$), the corresponding saddle-point equations are:

$$a=\frac{1}{\Delta+p(1-p)\lambda}\left\langle\!\!\left\langle\xi\,\mathrm{erfc}\left(\frac{u\Delta-a\xi-(\Delta-1)/2\lambda+(\alpha/2\lambda)\chi\Delta}{\sqrt{2\alpha R}}\right)\right\rangle\!\!\right\rangle_\xi \tag{50}$$

$$R=\frac{p^2(1-p)^2Q_0}{(1-p(1-p)C)^2} \tag{51}$$

$$\Delta=1+\frac{\lambda\alpha p(1-p)}{1-p(1-p)C} \tag{52}$$

$$C\Delta=\sqrt{\frac{2}{\pi\alpha R}}\left\langle\!\!\left\langle\exp\left(-\frac{(u\Delta-a\xi-(\Delta-1)/2\lambda+(\alpha/2\lambda)\chi\Delta)^2}{2\alpha R}\right)\right\rangle\!\!\right\rangle_\xi-\lambda \tag{53}$$

$$\begin{aligned}Q_0\Delta^2&=\left\langle\!\!\left\langle\mathrm{erfc}\left(\frac{u\Delta-a\xi-(\Delta-1)/2\lambda+(\alpha/2\lambda)\chi\Delta}{\sqrt{2\alpha R}}\right)\right\rangle\!\!\right\rangle_\xi+\alpha\lambda^2R\\&\quad-\lambda p^2(1-p)^2(2\Delta+p(1-p)\lambda)a^2\\&\quad-2\lambda\alpha\sqrt{\frac{2R}{\pi\alpha}}\left\langle\!\!\left\langle\exp\left(-\frac{(u\Delta-a\xi-(\Delta-1)/2\lambda+(\alpha/2\lambda)\chi\Delta)^2}{2\alpha R}\right)\right\rangle\!\!\right\rangle_\xi.\end{aligned} \tag{54}$$

In what follows we consider the limit of strong correlations among patterns: $p\ll 1$.

Consider first the case when λ is not too large (which includes the Hopfield model limit).

(a) $\lambda p \ll 1$. It can be shown *a posteriori* that in (51) and (52) the term $Cp \ll 1$, and therefore it can be omitted. It can also be shown that in the RHS of (54) all but the first term can also be omitted. Therefore (50)–(54) can be reduced to:

$$a = \frac{1}{\Delta}\left\langle\!\!\left\langle \xi \operatorname{erfc}\left(\frac{u\Delta - a\xi - (\Delta-1)/2\lambda + (\alpha/2\lambda)\chi\Delta}{\sqrt{2\alpha R}}\right)\right\rangle\!\!\right\rangle_{\xi} \tag{55}$$

$$\Delta = 1 + \lambda\alpha p \tag{56}$$

$$R\Delta^2 = 2p^2\left\langle\!\!\left\langle \operatorname{erfc}\left(\frac{u\Delta - a\xi - (\Delta-1)/2\lambda + (\alpha/2\lambda)\chi\Delta}{\sqrt{2\alpha R}}\right)\right\rangle\!\!\right\rangle_{\xi}. \tag{57}$$

The factor χ (see (19)) in the limit $\lambda p \ll 1$ is

$$\chi = \frac{\lambda p}{1 + \lambda p\alpha}. \tag{57}$$

Together with (56) for Δ one gets:

$$-\frac{\Delta - 1}{2\lambda} + \frac{\alpha}{2\lambda}\chi\Delta = 0$$

and therefore (55)–(57) (in the leading order in p) can be reduced to:

$$a = \frac{p}{\Delta}\left[\operatorname{erfc}\left(\frac{u\Delta - a}{\sqrt{2\alpha R}}\right) - \operatorname{erf}\left(\frac{u\Delta}{\sqrt{2\alpha R}}\right)\right] \tag{59}$$

$$R = \frac{p^2}{\Delta^2}\left[p \operatorname{erf}\left(\frac{u\Delta - a}{\sqrt{2\alpha R}}\right)\operatorname{erf}\left(\frac{u\Delta}{\sqrt{2\alpha R}}\right)\right]. \tag{60}$$

Redefining

$$a = pa_0 \qquad R = p^2 r \qquad u = pu_0 \qquad \alpha = \frac{1}{p}\alpha_0 \tag{61}$$

and introducing

$$x = \frac{u_0\Delta}{\sqrt{2\alpha R}} \qquad z = x - \frac{a_0}{\sqrt{2\alpha R}} \tag{62}$$

one gets:

$$z = \frac{1}{\sqrt{2\alpha_0}} \frac{u_0\Delta^2 - \operatorname{erf}(z) + \operatorname{erf}(x)}{\sqrt{\operatorname{erf}(z) + \operatorname{erf}(x)/p}} \tag{63}$$

$$z = \frac{u_0\Delta^2}{\sqrt{2\alpha_0}} \frac{1}{\sqrt{\operatorname{erf}(z) + \operatorname{erf}(x)/p}}. \tag{64}$$

If we choose the free parameter u_0 such that

$$u_0\Delta^2 = 1 - \delta \tag{65}$$

where

$$p \ll \delta \ll \frac{1}{\sqrt{\ln(1/p)}} \tag{66}$$

then the solutions of (63), (64) exist until

$$\alpha_0 \ll \frac{1}{2\ln(1/p)} \ll 1. \tag{67}$$

These solutions are:

$$x \simeq \frac{1}{\sqrt{2\alpha_0}} \gg 1 \tag{68}$$

$$z \simeq \frac{\delta}{\sqrt{2\alpha_0}} \ll 1 \tag{69}$$

or, in terms of the original variables:

$$a = \frac{p}{1+1\lambda\alpha p} \tag{70}$$

$$R = \frac{p^3}{(1+1\lambda\alpha p)^2}. \tag{71}$$

Note that according to the definition (see (42)) the variable a is connected with the overlaps:

$$m_\mu = \frac{1}{N}\sum_i \xi_i^\mu \langle\sigma_i\rangle \tag{72}$$

as follows:

$$a = \sum_\mu (\hat{1}+1\lambda\hat{C})^{-1}_{1\mu} m_\mu. \tag{73}$$

Therefore the overlap in the retrieval state $m \equiv m_{\mu=1}$ is equal to:

$$m = \frac{(1+1\lambda p)p}{(1+\lambda\alpha p)} \tag{74}$$

According to (65), (66) such retrieval states exist if we choose the threshold in the original Hamiltonian (37) as follows:

$$u = \frac{p}{(1+\lambda\alpha p)^2}(1-\delta) \tag{75}$$

where $p \ll \delta \ll 1/\sqrt{\ln(1/p)}$. Then the maximal value of α is

$$\alpha_{\max} \simeq \frac{1}{2p\ln(1/p)} \tag{76}$$

Inserting the solution (70), (71) in (53) and (54) one can easily check that $pC \ll 1$ and that all the terms but the first in the RHS of (54) are of higher order in p. Now consider the following case.

(b) $\lambda p \gg 1$. Again it can be checked *a posteriori* that the first term in the RHS of (53) is much smaller than λ, and all but the first term in (54) can be omitted.

Then (50)-(54) are reduced to:

$$a = \frac{1}{\Delta + p\lambda} \left\langle\!\!\left\langle \xi \operatorname{erfc}\left(\frac{u\Delta - a\xi - (\Delta - 1)/2\lambda + (\alpha/2\lambda)\chi\Delta}{\sqrt{2\alpha R}} \right) \right\rangle\!\!\right\rangle_\xi \tag{77}$$

$$R = \frac{p^2 Q_0}{1 + (\lambda p/\Delta)} \tag{78}$$

$$\Delta = 1 + \frac{\lambda\alpha p}{1 + (\lambda p/\Delta)} \tag{79}$$

$$Q_0\Delta^2 = 2\left\langle\!\!\left\langle \operatorname{erfc}\left(\frac{u\Delta - a\xi - (\Delta - 1)/2\lambda + (\alpha/2\lambda)\chi\Delta}{\sqrt{2\alpha R}} \right) \right\rangle\!\!\right\rangle_\xi . \tag{80}$$

Assuming that $\alpha \gg 1$ one obtains from (79):

$$\Delta \simeq \lambda\alpha p \tag{81}$$

The (19) for χ gives

$$\chi \simeq \frac{1}{\alpha}. \tag{82}$$

Therefore the term

$$-\frac{\Delta - 1}{2\lambda} + \frac{\alpha}{2\lambda}\chi\Delta$$

in (77) and (80) can be omitted, and (77)-(80) are reduced to

$$a = \frac{1}{\lambda\alpha}\left[\operatorname{erf}\left(\frac{u\Delta - a}{\sqrt{2\alpha R}} \right) - \operatorname{erf}\left(\frac{u\Delta}{\sqrt{2\alpha R}} \right) \right] \tag{83}$$

$$R = \frac{1}{\lambda^2\alpha^2}\left[p \operatorname{erf}\left(\frac{u\Delta - a}{\sqrt{2\alpha R}} \right) + \operatorname{erf}\left(\frac{u\Delta}{\sqrt{2\alpha R}} \right) \right]. \tag{84}$$

Introducing $x = u\Delta/\sqrt{2\alpha R}$ and $z = x - a/\sqrt{2\alpha R}$, one again obtains (63), (64) in which

$$u\frac{(1 + 1\lambda\alpha p)^2}{p} \rightarrow u\lambda^2\alpha^2 p. \tag{85}$$

Therefore the threshold should be chosen in the form:

$$u = \frac{1}{\lambda^2\alpha^2 p}(1 - \delta) \tag{86}$$

where $p \ll \delta \ll 1/\sqrt{\ln(1/p)}$, and then until

$$\alpha < \alpha_{\max} \simeq \frac{1}{2p\ln(1/p)} \tag{87}$$

one again obtains the retrieval solutions (68), (69). In terms of the original variables:

$$R \simeq \frac{p}{\lambda^2\alpha^2} \tag{88}$$

$$a \simeq \frac{1}{\lambda\alpha^2}. \tag{89}$$

For the overlap in the retrieval state one gets:

$$m \simeq \frac{p}{\alpha} \simeq p^2 \ln \frac{1}{p}. \tag{90}$$

Note that the results of the above thermodynamic solution for the maximal capacity and for the value of the threshold, equations (75) and (86), perfectly fit the predictions of the signal-to-noise analysis, equations (35) and (36).

4. Conclusions

We have considered the fully connected neural network defined by the symmetric coupling matrix:

$$\hat{J} = J\hat{A}(\hat{1} + \lambda\hat{A})^{-1}$$

where $\hat{A}$ is the Hopfield model coupling matrix.

The results of the present and the previous studies (Dotsenko *et al* 1991) show that such a model is very rich in its behaviour and theoretically very robust.

For the model storing uncorrelated patterns the maximal capacity can be moved from $\alpha_c = 0.14$ (at $\lambda = 0$) up to $\alpha_c = 1$ (for $\lambda \to \infty$). For the model storing strongly correlated patterns the maximal capacity can reach the Gardner limiting value $(p \ln(1/p))^{-1}$ for any value of the parameters λ if the threshold term in the Hamiltonian is chosen in the optimal way.

Although due to the replica symmetry breaking a new structure of the metastable states may appear in the system at finite λ (Dotsenko and Tirozzi 1991) it does not seem to produce a strong effect on the retrieval properties of the model.

Acknowledgments

Two of the authors (RD and VD) are grateful for hospitality of the Department of Mathematics of the University of Rome I.

Appendix: the diagram technique

A.1. Calculation of $Tr\,\hat{A}(\hat{1}+\lambda\hat{A})^{-1}$

Consider first the quantity

$$G_i \equiv (\hat{1} + \lambda\hat{A})^{-1}_{ii} \tag{A1}$$

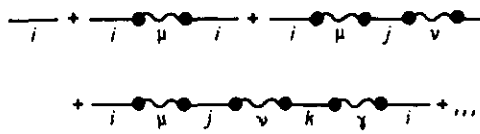
where

$$\hat{A}_{ij} = \frac{1}{N}\sum_{\mu} \xi_i^{\mu}\xi_j^{\mu}$$

and the random ξ's are described by the distribution function (10). Expanding the RHS of (A1) one gets:

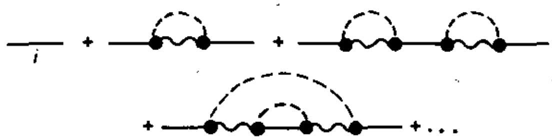
$$G_i = \delta_{ii} - \frac{\lambda}{N}\sum_{\mu} \xi_i^{\mu}\xi_i^{\mu} + \frac{\lambda^2}{N^2}\sum_{j}\sum_{\mu\nu} \xi_i^{\mu}\xi_j^{\mu}\xi_j^{\nu}\xi_i^{\nu} - \frac{\lambda^3}{N^3}\sum_{jk}\sum_{\mu\nu\gamma} \xi_i^{\mu}\xi_j^{\mu}\xi_j^{\nu}\xi_k^{\nu}\xi_k^{\gamma}\xi_i^{\gamma} + \dots \tag{A2}$$

In terms of the diagrams (see Hertz *et al* 1991) it can be represented as

$G_i =$ 

(A3)

where the straight lines carry the space indices, while wavy lines carry the pattern indices and the additional factor $(-\lambda/N)$. Assuming self-averaging, one has to make all possible pairings of the ξ's. The result can be represented as follows:

$G_i =$ 

(A4)

Here the dashed line causes all the indices of the vertices it connects to be equal and carry additional factor $\langle\langle\xi^2\rangle\rangle = p(1-p)$. In the representation of (A4) one may omit all the diagrams with cross sections of the dashed lines since they are of higher order in $1/N$.

The diagrams can be summed up in terms of a self-energy Σ, corresponding to the Dyson equation:

$G_i \equiv$  (A5)

The solution of this equation is:

$$G = \frac{1}{1-\Sigma} \tag{A6}$$

where:

$\Sigma =$  $\equiv -\dfrac{\lambda p(1-p)}{N} \sum_{\mu=1}^{M} A_\mu$ (A7)

and the double wavy line is defined by the equation:

 (A8)

The solution of this equation is

$$\mathscr{A}_\mu \equiv \mathscr{A} = \frac{1}{1+\lambda p(1-p)G} \tag{A9}$$

and therefore

$$\Sigma = -\frac{\alpha\lambda p(1-p)}{1+\lambda p(1-p)G} \tag{A10}$$

Equation (A6) for G reads:

$$G = \left(1 + \frac{\alpha\lambda p(1-p)}{1+\lambda p(1-p)G}\right)^{-1}. \tag{A11}$$

Its solution is

$$G=\frac{1-\alpha-\frac{1}{\lambda p(1-p)}+\sqrt{\left(1-\alpha-\frac{1}{\lambda p(1-p)}\right)^2+\frac{4}{\lambda p(1-p)}}}{2}. \tag{A12}$$

For the factor

$$\chi=\frac{\lambda}{\alpha}\operatorname{Tr}\hat{A}(\hat{1}+\lambda\hat{A})^{-1} \tag{A13}$$

one easily gets:

$$\chi=\frac{1}{\alpha}(1-G) \tag{A14}$$

which gives the result:

$$\chi=\frac{1+\alpha+\frac{1}{\lambda p(1-p)}-\sqrt{\left(1-\alpha-\frac{1}{\lambda p(1-p)}\right)^2+\frac{4}{\lambda p(1-p)}}}{2}. \tag{A15}$$

A.2. Signal and noise

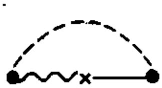
The signal term in the local fields produced by the pattern number 1 can be represented as follows:

$$\sum_{i\neq j} J_{ij}\xi_j^1=\sum_{i\neq j}(\hat{A}-\lambda\hat{A}^2+\lambda^2\hat{A}^3-\ldots)_{ij}\xi_j^1 \tag{A16}$$

In terms of the diagram introduced above the lowest order term is

$$\left(-\frac{1}{\lambda}\right)\times$$  $$=\frac{1}{N}\sum_{i\neq j}\xi_i^1\xi_j^1\xi_j^1=\xi_i^1 p(1-p). \tag{A17}$$

Here the cross represents the fixed pattern ξ_i^1. The diagram



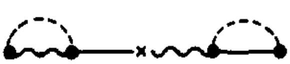
(A18)

can be omitted because it is of the order of $1/N$, and the diagram

(A19)

is forbidden because $j\neq i$.

Higher order corrections are included by 'dressing' the diagram (A17), i.e. by replacing the bare diagram elements with the double wavy line or the heavy straight line. However, in the higher orders there will also contribute diagrams like:



(A20)

and the corresponding diagrams obtained by dressing (A20). Therefore the signal, which is the sum of all contributions may be represented as:

$$S=\left(-\frac{1}{\lambda}\right)$$  (A21)

where

 = 1 +  (A22)

Since (see (A13) and (A5)):

 $= -\chi$ (A23)

and

$1+$  $= G = 1-\alpha\chi$ (A24)

one obtains the result:

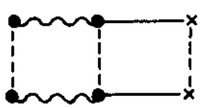
$$S=\frac{1}{\lambda}\chi(1-\alpha\chi). \tag{A25}$$

The noise term can be represented as follows:

$$\sigma=\left\langle\!\!\left\langle\left(\sum_{i\neq i} J_{ij}\xi_j^1 - s\xi_i^1\right)^2\right\rangle\!\!\right\rangle \tag{A26}$$

where the average $\langle\!\langle\ldots\rangle\!\rangle$ is over all the patterns but the first one.

In the lowest order one gets

$\frac{1}{\lambda^2}$  $=\frac{1}{N^2}\sum_{\nu}\sum_{i\neq j}(\xi_i^\nu)^2(\xi_j^\nu)^2(\xi_j^1)^2=\alpha p^3(1-p)^3.$ (A27)

Then the sum over all the diagrams of all order can be represented as

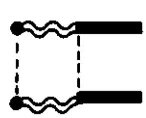
$\sigma^2=$  (A28)

where

 $=1+$  $=1-\alpha\chi$ (A29)

 $=1+$  $=1-\chi$ (A30)

and

$L\equiv$  =  +  (A31)

or

$$L=\alpha\chi^2+(\alpha\chi^2)^2L. \tag{A32}$$

Taking into account (A29), (A30) and (A32) one finally gets the result:

$$\sigma^2 = 2\frac{p(1-p)}{\lambda^2}(1-\chi)^2(1-1\alpha\chi)^2\frac{\alpha\chi^2}{1-\alpha\chi^2}. \tag{A33}$$

The additional factor $p(1-1p)$ comes from the dashed line to the right in the diagram (A28).

References

Buhmann J, Divko R and Schulten K 1989 *Phys. Rev.* A **39** 2689
Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys.* **173** 30
Dotsenko V and Tirozzi B 1991 *J. Phys. A: Math. Gen.* **24** 5163
Dotsenko V, Yarunin N and Dorotheyev E 1991 *J. Phys. A: Math. Gen.* **24** 2419
Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
Hertz J A, Krough A and Thorbergsson G I 1989 *J. Phys. A: Math. Gen.* **22** 2133
Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380
Personaz L, Guyon I and Dreyfus G 1985 *J. Physique Lett.* **46** L359
Tsodyks M and Feigelmann M 1988 *Europhys. Lett.* **6** 101