The number of metastable states in the projection rule neural network

1994 J. Phys. A: Math. Gen. 27 5857

(http://iopscience.iop.org/0305-4470/27/17/018)

J. Phys. A: Math. Gen. 27 (1994) 5857–5870. Printed in the UK

The number of metastable states in the projection rule neural network

Peter Kuhlmann† and Joachim K Anlauf‡

† Physikalisches Institut, Theoretische Physik III, Julius-Maximilians-Universität Würzburg Am Hubland, D-97074 Würzburg, Germany

‡ Siemens AG, ZFE STSN 43, Otto-Hahn-Ring 6, D-81739 München, Germany

Received 17 May 1994

Abstract. We calculate the direct average $\langle\langle \mathcal{N} \rangle\rangle$ of the number of metastable states in the projection rule neural network by means of the saddle-point method. $\langle\langle \mathcal{N} \rangle\rangle$ is obtained as a function of the Hamming distance g to a given pattern, stability γ and energy E. The critical storage capacity of the model is calculated. The examination of the energy dependence of $\langle\langle \mathcal{N} \rangle\rangle$ leads to lower bounds for the energies of the spin-glass states.

1. Introduction

In spin glasses [1–3], and attractor neural networks [4, 5], information about the ground state can be gained by two methods. The first one deals with the calculation of the thermodynamic properties. It is assumed that the spin system is in thermodynamic equilibrium with a heat bath at temperature T. So the free energy and the order parameters can be calculated, in the limit $T \to 0$ the ground state energy is obtained.

The second method consists of the calculation of the number $\mathcal{N}$ of metastable states. These are states in which the spin system can be trapped during a zero-temperature dynamic process. If $\mathcal{N}$ is calculated as a function of energy E, the zero of

$$h(E) = \frac{1}{N} \ln \mathcal{N}(E) \tag{1}$$

yields the ground state energy. In this paper this method is applied to the projection rule neural network [6–8] for which the thermodynamic properties have been calculated in [9]. The paper is organized as follows. In the remainder of the introduction some known results about the calculation of $\mathcal{N}$ are presented in the context of the Sherrington–Kirkpatrick model of spin glasses (SK model) and of the Hopfield model. In section 2 the projection rule neural network (PRNN) is presented. The mean number of metastable states is calculated in section 3. The results are given in section 4, where conclusions about the retrieval properties of the PRNN will be drawn. Finally the results are discussed in section 5.

1.1. The definition of a metastable state

In general, spin-glass models and attractor neural networks are systems consisting of N Ising spins $S_i \in \{-1, +1\}$, with couplings J_{ij} between them. It is a well known result that the

0305-4470/94/175857+14$19.50

serial Glauber heat-bath process [10] converges to a Boltzmann distribution with an energy function

$$E = -\tfrac{1}{2}\sum_{i,j} J_{ij} S_i S_j \tag{2}$$

if J_{ij} is symmetric and there are no self-couplings ($J_{ii} = 0\ \forall i$), e.g., see [4]. If the temperature T of the heat bath is zero, the serial Glauber dynamics becomes deterministic: in a time step Δt one single randomly or sequentially chosen spin S_i is updated according to

$$S_i(t + \Delta t) = \operatorname{sign}(u_i(t)) \tag{3}$$

where

$$u_i(t) = \sum_{j=1}^{N} J_{ij} S_j(t) \tag{4}$$

is the local field at site i.

A state

$$\underline{S} = (S_1, \ldots, S_N)^T$$

is called metastable, if it is a fixed point of the dynamics (3)

$$S_i = \operatorname{sign} u_i \qquad \forall i.$$

Hence the local energies λ_i satisfy equivalently

$$\lambda_i = S_i u_i > 0 \qquad \forall i. \tag{5}$$

If J_{ij} is symmetric and there are no self-couplings ($J_{ii} = 0\ \forall i$), the serial $T = 0$ Glauber dynamics can only flow to metastable states. There are no cycles in this case [4]. Therefore a calculation of the number of metastable states is strongly connected to serial dynamics and the corresponding energy function E (see equation (2)). In parallel dynamics the cycles outnumber the metastable states [11]. In this case a different energy function has to be considered [12, 13] and the number of cycles will be most important for parallel dynamics.

1.2. The density of metastable states

The density of metastable states can be calculated by setting

$$\mathcal{N}(\epsilon) = \sum_{\{S_i\}} \left(\prod_{i=1}^{N} \Theta(\lambda_i) \right) \delta\left(\epsilon + \frac{E}{N} \right) \tag{6}$$

where the Θ-function is defined by

$$\Theta(x) = \begin{cases} 1 & \text{for } x > 0 \\ 0 & \text{for } x \leqslant 0. \end{cases}$$

As required above, $\mathcal{N}$ is a function of energy, because ϵ is a parameter that denotes the negative energy per site:

$$\epsilon = -\frac{E}{N}. \tag{7}$$

In the case of spin glasses and attractor neural networks the couplings J_{ij} have a probability distribution $P(\{J_{ij}\})$. Due to the effect of frustration in these models, we expect the behaviour

$$\mathcal{N} = \mathrm{e}^{Nh} \tag{8}$$

(see equation (1)), where h is self-averaging in the limit $N \to \infty$. If $\langle\langle h \rangle\rangle$ denotes the average of h over the J_{ij}, $\mathrm{e}^{N\langle\langle h \rangle\rangle}$ can be regarded as an asymptotic expression for $\mathcal{N}$. Jensen's inequality [14] states that $\langle\langle h \rangle\rangle$ is bounded by the direct average:

$$\mathrm{e}^{N\langle\langle h \rangle\rangle} \leqslant \langle\langle \mathrm{e}^{Nh} \rangle\rangle = \langle\langle \mathcal{N} \rangle\rangle. \tag{9}$$

1.3. $\mathcal{N}$ in the SK model

In the SK model of spin glasses the couplings J_{ij} are Gaussian distributed with zero mean and standard deviation $\sigma = 1/\sqrt{N}$. In their original work [15], Sherrington and Kirkpatrick calculated the thermodynamic properties in replica symmetry. By introducing a scheme of replica symmetry breaking, Parisi constructed what is believed to be the exact solution of the SK model [16]. The direct average of the number of metastable states $\langle\langle \mathcal{N} \rangle\rangle$ has been calculated by Tanaka and Edwards [17], and an upper bound for the negative energy (per spin) ϵ of the spin-glass ground state has been gained by Bray and Moore [18],

$$\epsilon = -\frac{E}{N} \leqslant \epsilon_{\mathrm{B}} = 0.791.$$

The replica symmetric theory for $\langle\langle h \rangle\rangle$ was solved by Roberts [19]. He discovered that only in a region of low ϵ-values does the inequality (9) become an equality:

$$\epsilon = -\frac{E}{N} \leqslant \epsilon_0 = 0.672.$$

Hence the direct average could not be expected to yield a good approximation for the ground-state energy. But in his replica symmetric calculation of $\langle\langle h \rangle\rangle$ Roberts obtained

$$\epsilon_{\mathrm{G}} \sim 0.77$$

for the negative energy (per spin) of the spin-glass ground state. This was in good agreement with the Parisi result ($\epsilon_{\mathrm{P}} = 0.7633 \pm 1.E-4$) and results from computer simulations ($\epsilon_{\mathrm{S}} = 0.76 \pm 0.01$) [16].

1.4. $\mathcal{N}$ in the Hopfield model

In the Hopfield model [20] $p = \alpha N$ patterns

$$\underline{\xi^\nu} = (\xi_1^\nu, \ldots, \xi_N^\nu)^T$$

with

$$\xi_i^\nu \in \{-1, +1\} \qquad \nu = 1, \ldots, p \qquad i = 1, \ldots, N$$

should be stored by Hebb's coupling matrix

$$J_{ij} = \frac{1}{N} \sum_{\nu=1}^{p} \xi_i^\nu \xi_j^\nu. \tag{10}$$

Since the J_{ij} are correlated, the average over the independent

$$\xi_i^\nu \qquad p(\xi_i^\nu = \pm 1) = \tfrac{1}{2}$$

has to be taken.

For a constant $\alpha = p/N$, the direct average $\langle\langle \mathcal{N} \rangle\rangle$ has been calculated by Gardner [21] in the limit $N \to \infty$. Details of the calculation may be found in [22] and in [23]. Gardner included the Hamming distance g to a specified (e.g. the first) pattern in her calculation. g is defined by the overlap m^1 to the first pattern,

$$m^1 = \frac{1}{N} \sum_{i=1}^{N} \xi_i^1 S_i = 1 - 2g. \tag{11}$$

So g denotes the fraction of the wrong bits in a state.

For $\alpha < 0.113$, Gardner observed a distinct band of metastable states near the pattern. The band is responsible for the retrieval, and the thermodynamic ground state [24] is believed to be part of the band.

For $\alpha \to \infty$, the correlations between the J_{ij} vanish and the SK result

$$\lim_{N \to \infty} \frac{1}{N} \ln \langle\langle \mathcal{N} \rangle\rangle = 0.1992$$

is recovered.

2. The model

In the projection rule neural network we are considering, $p = \alpha N$ patterns

$$\underline{\xi^\nu} = (\xi_1^\nu, \ldots, \xi_N^\nu)^T$$

with

$$\xi_i^\nu \in \{-1, +1\} \qquad \nu = 1, \ldots, p \qquad i = 1, \ldots, N$$

have to be stored [7–9]. The patterns are chosen according to the independent probability distribution ($p(\xi_i^\nu = \pm 1) = \frac{1}{2}$).

If $\alpha < 1$ is kept fixed and the limit $N \to \infty$ is taken, they are linearly independent with probability 1 [9, 25]. Hence the correlation matrix

$$C_{\mu\nu} = \frac{1}{N} \sum_{i=1}^{N} \xi_i^\mu \xi_i^\nu \tag{12}$$

is positive-definite, and the projection rule coupling matrix

$$J_{ij} = \frac{1}{N} \sum_{\mu,\nu} \xi_i^\mu (C^{-1})_{\mu\nu} \xi_j^\nu \tag{13}$$

is well defined. For $\alpha < 1$, J_{ij} projects every state $\underline{S} \in \{+1, -1\}^N$ into the space of the patterns: if we consider the unique decomposition of a general state

$$\underline{S} = \sum_{\mu=1}^{p} a_\mu \underline{\xi^\mu} + \underline{\delta S} \tag{14}$$

where $\underline{\delta S}$ is orthogonal to all the patterns, then the action of the projection matrix (13) on $\underline{S}$ is

$$\sum_{j=1}^{N} J_{ij} S_j = \sum_{\mu=1}^{p} a_\mu \xi_i^\mu \qquad \forall i. \tag{15}$$

In particular, the patterns have a high stability, because the local energies (5) of the patterns obey

$$\lambda_i^\mu = \xi_i^\mu \sum_{j=1}^{N} J_{ij} \xi_j^\mu = 1 \qquad \forall i, \mu. \tag{16}$$

Indeed, it has been shown that the p patterns are the stable ground states for all $\alpha < 1$ [9]. So the projection rule neural network can be used as an associative memory. To establish the relationship between the thermodynamic calculation in [9] and our calculation of the direct average $\langle\langle \mathcal{N} \rangle\rangle$ the self-couplings J_{ii} should be excluded in the local fields in equation (4). Thus, equation (5) reads

$$\lambda_i = S_i \sum_{j=1}^{N} J_{ij} S_j > J_{ii} \qquad \forall i. \tag{17}$$

Instead of equation (17) we impose the constraint

$$\lambda_i = S_i \sum_{j=1}^{N} J_{ij} S_j > \gamma \qquad \forall i \tag{18}$$

in our analytical calculation. The stability parameter γ is a threshold for the local energies λ_i. $\gamma = \alpha$ can be related to the network without self-couplings as follows. We calculate the probability distribution $P(J)$ of J_{ii} in the limit $N \to \infty$. A beta distribution [26] with mean $\langle\langle J_{ii} \rangle\rangle = \alpha$ and standard deviation $\sigma = \sqrt{\alpha(1-\alpha)/(N/2+1)}$ is gained:

$$P(J) = \langle\langle \delta(J - J_{ii}) \rangle\rangle \tag{19}$$

$$= \frac{\Gamma(N/2)}{\Gamma[(N/2)\alpha]\Gamma[(N/2)(1-\alpha)]} J^{(N/2)\alpha - 1} (1 - J)^{(N/2)(1-\alpha)-1}. \tag{20}$$

Hence it is shown that J_{ii} is self-averaging in the limit $N \to \infty$, because $\sigma \to 0$ ($N \to \infty$).

3. The mean number of metastable states in the PRNN

Using equation (6) we obtain

$$\mathcal{N}(\alpha, g, \gamma, \epsilon) = \sum_{\{S_i\}}{}' \left(\prod_{j=1}^{N} \Theta\left(S_j \sum_{k=1}^{N} J_{jk} S_k - \gamma \right) \right) \delta\left(\epsilon - \frac{1}{2N} \sum_{j,k} J_{jk} S_j S_k \right). \tag{21}$$

The prime denotes the restricted sum over all the states $\{S_i\}$ that have a fixed Hamming distance g to the first pattern and are uncorrelated to all the other patterns (see equation (11)). γ is the stability parameter specified above in equation (18). $\epsilon = -E/N$ denotes the negative energy E per site. Note that

$$\epsilon = +\frac{1}{2N} \sum_{j,k} J_{jk} S_j S_k = \frac{1}{2N} \sum_{j,k\,(j \neq k)} J_{jk} S_j S_k + \frac{1}{2}\alpha \tag{22}$$

because the relation $\sum_{j=1}^{N} J_{jj} = N\alpha$ holds for all N. Since Kanter and Sompolinsky also included the constant term in the energy, their results [9] can be compared with our results for the dependence of $\mathcal{N}$ on ϵ.

We perform the average $\langle\langle \mathcal{N} \rangle\rangle = (1/2^{Np}) \sum_{\{\xi_i^\mu\}} \mathcal{N}$ by means of the saddle-point method, which becomes exact in the limit $N \to \infty$. Details of the computation may be found in the appendix and in [27].

We obtain the result

$$\lim_{N\to\infty} \frac{1}{N} \ln \langle\langle \mathcal{N}(\alpha, g, \gamma, \epsilon) \rangle\rangle = f(\alpha, g, \gamma, \epsilon). \tag{23}$$

$f(\alpha, g, \gamma, \epsilon)$ is the saddle-point value with respect to the variables X, Y, Z, a_1, l_1, w of

$$\begin{aligned} f(\alpha, g, \gamma, \epsilon, X, Y, Z, a_1, l_1, w) &= -g \ln g - (1-g)\ln(1-g) - \alpha \ln \alpha - (1-\alpha)\ln(1-\alpha) \\ &\quad - w\epsilon + \frac{\alpha}{2} \ln(Y^2 - XZ) - \frac{1}{2} \ln K + \frac{w^2 Z}{8K} \\ &\quad + \frac{w}{2K} \{(K - 1 + Y) + (2g-1)[-l_1 Z - a_1(1-Y)]\} \\ &\quad + \frac{1}{2K} [l_1^2 Z + 2 l_1 a_1 (1-Y) + a_1^2 X + X] \\ &\quad + \frac{2g-1}{K} [l_1(1-Y) + a_1 X] + g \ln \Phi(t_1) + (1-g) \ln \Phi(t_2) \end{aligned} \tag{24}$$

where

$$K = (1-Y)^2 - XZ$$

$$\Phi(t) = \int_{-\infty}^{t} \frac{\mathrm{d}\lambda}{\sqrt{2\pi}} \mathrm{e}^{-\lambda^2/2}$$

$$t_1 = \sqrt{\frac{K}{Z}} \left[1 - \gamma + \frac{1}{K} \left(\frac{1}{2} w Z - l_1 Z - a_1(1-Y) + Y - 1 \right) \right]$$

and

$$t_2 = \sqrt{\frac{K}{Z}}\left[1-\gamma+\frac{1}{K}\left(\frac{1}{2}wZ+l_1Z+a_1(1-Y)+Y-1\right)\right].$$

The order parameters Z and w can be interpreted as follows. $Z = \sum_{\mu=2}^{p} a_\mu^2$, where the a_ν ($\nu \in \{1,\ldots,p\}$) are the coefficients in the linear combination (15) that results from the projection of a state into the space of the patterns. w indicates where f is at its maximum $\epsilon_{\max}$ with respect to ϵ: $w = 0$ if $\epsilon = \epsilon_{\max}$.

4. Results

The saddle-point equations for the order parameters X, Y, Z, a_1, l_1 and w have been solved in various regions of the space of the four parameters capacity α, Hamming distance g, stability γ and negative energy per site ϵ. For the stability γ we will only be interested in $\gamma = 0$ and $\gamma = \alpha$. As stated above (see section 2) in the $\gamma = 0$ case the self-couplings are included in the local fields, whereas $\gamma = \alpha$ is put in relation to the network without self-couplings.

4.1. The total number of metastable states

If $f = \lim_{N\to\infty}(1/N)\ln\langle\langle\mathcal{N}\rangle\rangle$ is maximized with respect to energy ϵ and Hamming distance g we obtain $a_1 = 0$, $l_1 = 0$ and $w = 0$ in equation (24). Of course the Hamming distance to the first pattern is $g = \frac{1}{2}$ at the maximum, and almost all the metastable states are uncorrelated to all the patterns. We shall call such a state that resides at the maximum of f a *typical* metastable state. The corresponding curves for $\gamma = 0$ and $\gamma = \alpha$ are shown in figure 1. The curve for $\gamma = \alpha$ looks similar to a curve of simulation results for the mean remanent magnetization m_r at $g = \frac{1}{2}$ in the model without self-couplings as obtained by Henkel and Opper [28]. Henkel and Opper defined m_r by the overlap

$$m_r = \frac{1}{N}\underline{S}^s \cdot \underline{S}^m \tag{25}$$

where $\underline{S}^s$ denotes a random starting state and $\underline{S}^m$ denotes the corresponding stopping state of the serial $T = 0$ Glauber dynamics, i.e. a particular metastable state. Although the exact relationship between m_r and f is not known, it can be seen from figure 2 that they increase similarly. It is obvious that a stopping state will be closer to a starting state, if there are more traps for the dynamics.

In the limit $\alpha \to 1$ we observe $f \to \ln 2$ in figure 1 for both $\gamma = 0$ and $\gamma = \alpha$. It is easily seen from equation (16) that the coupling matrix J_{ij} converges to the identity matrix as $\alpha \to 1$ and that for $\gamma = 0$ all the states will be metastable in this limit. Nevertheless one has to note that for all $\alpha < 1$ a typical metastable state has a component in the space that is orthogonal to the space spanned by the patterns. If the converse were true, i.e. if a typical metastable state lay fully in the space of the patterns, then $\epsilon = \frac{1}{2}$ should hold for it because of equations (16) and (22). But $\epsilon_{\max} < \frac{1}{2}$ is observed as a solution of the saddle-point equations for all $\alpha < 1$.

In the $\gamma = \alpha$ case, even the patterns are unstable for $\alpha = 1$. So the limit $f(\gamma = \alpha) \to \ln 2$ ($\alpha \to 1$) is not reached; for $\alpha = 1$ there are no metastable states in the $\gamma = \alpha$ case.

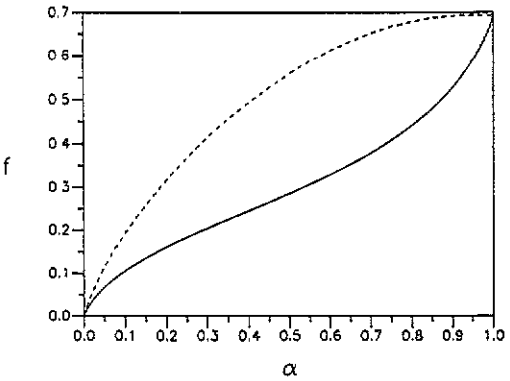

Figure 1. The exponent f in $\langle\langle \mathcal{N} \rangle\rangle \sim \exp(Nf)$ as a function of α at the Hamming distance $g = \frac{1}{2}$ and $\epsilon = \epsilon_{\max}$. The broken line refers to stability $\gamma = 0$, while the full line corresponds to $\gamma = \alpha$.

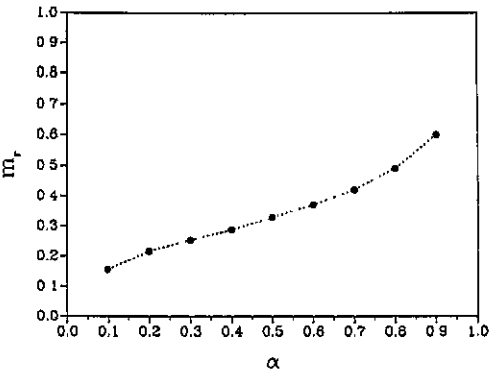

Figure 2. The remanent magnetization m_r in a serial $\tau = 0$ Glauber process. This simulation for the PRNN without self-couplings has been performed by Henkel and Opper [28].

In the limit $\alpha \to 0$, f can be calculated analytically as

$$f(\alpha, g = \tfrac{1}{2}, \gamma = 0, \epsilon_{\max}) = -\frac{\alpha}{2}\ln\alpha + \frac{\alpha}{2}\left[\ln\left(\frac{2}{\pi}\right) + 1 - \ln\left(1 - \frac{2}{\pi}\right)\right] + \mathcal{O}(\alpha^2)$$

and

$$f(\alpha, g = \tfrac{1}{2}, \gamma = \alpha, \epsilon_{\max}) = -\frac{\alpha}{2}\ln\alpha + \frac{\alpha}{2}\left[\ln\left(\frac{2}{\pi}\right) - 1 - \ln\left(1 - \frac{2}{\pi}\right)\right] + \mathcal{O}(\alpha^2).$$

Note that f is greater than the corresponding value in the Hopfield model and in the lowest order in α they coincide to $f \to -(\alpha/2)\ln\alpha$ [21, 27].

4.2. A general Hamming distance g to a pattern

We still consider $f = \lim_{N\to\infty}(1/N)\ln\langle\langle \mathcal{N} \rangle\rangle$ at its maximum with respect to ϵ. The Hamming distance g is a free parameter now. If the metastable states were equally distributed in the phase space $\{-1, +1\}^N$ our result at general g would be

$$f_{\rm ph}(\alpha, g, \gamma, \epsilon_{\max}) = \lim_{N\to\infty}\frac{1}{N}\ln\left[\exp(Nf(\alpha, g = \tfrac{1}{2}, \gamma, \epsilon_{\max}))\frac{\binom{N}{Ng}}{2^N}\right]$$

$$= f(\alpha, g = \tfrac{1}{2}, \gamma, \epsilon_{\max}) - g\ln g - (1-g)\ln(1-g) - \ln 2. \tag{26}$$

But as in the Hopfield model [21], we observe

$$f_{\rm ph}(\alpha, g, \gamma, \epsilon_{\max}) < f(\alpha, g, \gamma, \epsilon_{\max})$$

for all $g < \frac{1}{2}$. Especially at lower values of g there are many more metastable states than predicted by equation (26).

As a key result of our paper we observe areas around the patterns in which nearly no metastable states exist (see figure 3). This is because of Jensen's inequality (9): the upper

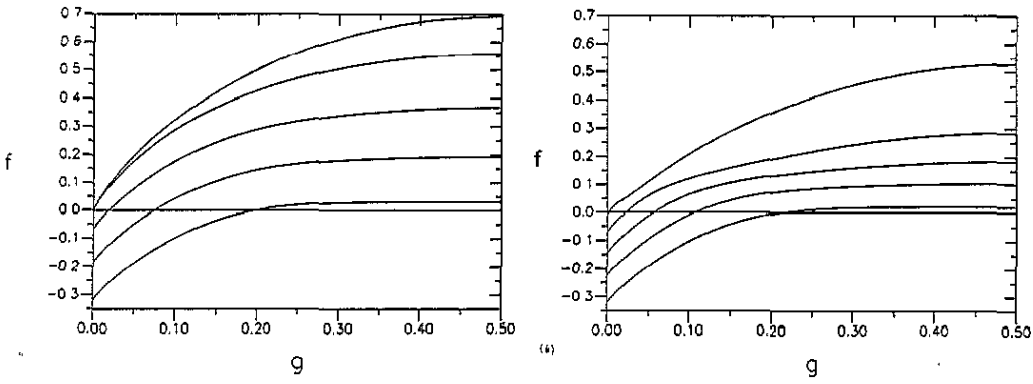

Figure 3. $f(\alpha, g, \gamma, \epsilon_{max})$ as a function of g for $\alpha = 0.01$, 0.1, 0.25, 0.5 and 0.9 from bottom to top for (*a*) $\gamma = 0$ and (*b*) $\gamma = \alpha$.

bound for the number of metastable states is already zero in these areas. We denote the zeros of f with respect to g by g_0. If one starts a serial $T = 0$ Glauber process at $g < g_0$, nearly no metastable state prevents the system from flowing to the pattern. Therefore the zeros g_0 in figure 3 can serve as lower bounds for the sizes of the basins of attraction (see [9]). These lower bounds are shown in figure 4. If we define α_c to be the critical storage capacity where the radius of attraction becomes zero, we can conclude from the lower bounds in figure 4 that $\alpha_c \geqslant \frac{1}{2}$ for $\gamma = 0$ and $\alpha_c = 1$ for $\gamma = \alpha$. This can also be derived analytically [27]. The thermodynamic calculation in [9] also yields the result $\alpha_c = 1$. This is an important check for our ansatz $\gamma = \alpha$ in equation (18). The result $\alpha_c(\gamma = 0) = \frac{1}{2}$ has already been derived by Kanter and Sompolinsky [9]. They showed that states, which differ from a pattern in only one site, are also metastable. We can support their reasoning by noting that for $\gamma = 0$ and $\alpha \geqslant \frac{1}{2}$ there are exponentially many metastable states in each vicinity of the pattern. Hence it is likely that the system is prevented from flowing to the ground state.

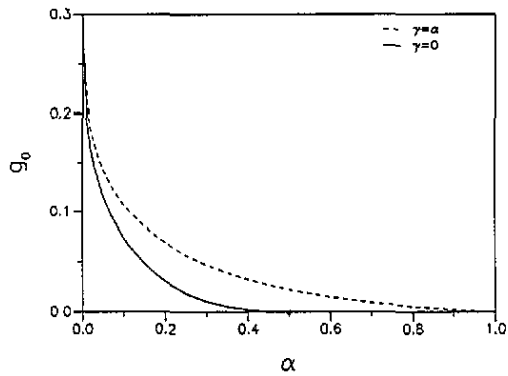

Figure 4. The zeros g_0 of the function $f(\alpha, g, \gamma, \epsilon_{max})$ with respect to g are shown in the cases $\gamma = \alpha$ (broken line) and $\gamma = 0$ (full line). Note that $g_0 \to \frac{1}{2}(\alpha \to 0)$ in both cases and that $g_0(\alpha) = 0$ for $\alpha \geqslant \frac{1}{2}$ in the $\gamma = \alpha$ case.

4.3. The energy dependence of f

We examine the dependence of

$$f(\alpha, g, \gamma, \epsilon)$$

on ϵ at the maximum $g = \frac{1}{2}$. Hence only metastable states that are uncorrelated to all the patterns are considered. Since e^{Nf} provides an upper bound for the number of metastable states $\mathcal{N}$, we can adopt the same reasoning as in the SK model to gain information about the energy of the spin-glass state. In the PRNN the spin-glass state is a solution of the replica symmetric thermodynamic calculation in [9]. It is uncorrelated to all the patterns and should therefore be a particular metastable state at $g = \frac{1}{2}$. Consequently, if we denote its energy by $E_G = -N\epsilon_G$, the zero of f with respect to ϵ should yield an upper bound for ϵ_G.

As outlined above in section 2 we must set this stability parameter to $\gamma = \alpha$ in order to establish a relationship between our calculation and the thermodynamic calculation in [9]. For each α the dependence of $f(\alpha, g = \frac{1}{2}, \gamma = \alpha, \epsilon)$ on ϵ can be plotted. We shall call these curves energy spectra. They are shown in figure 5. The spectra are similar to the analogous curve in the SK model [18]. The right and left zeros of f and the ϵ_{max} are illustrated in figure 6. The right zeros provide upper bounds for the values of ϵ_G.

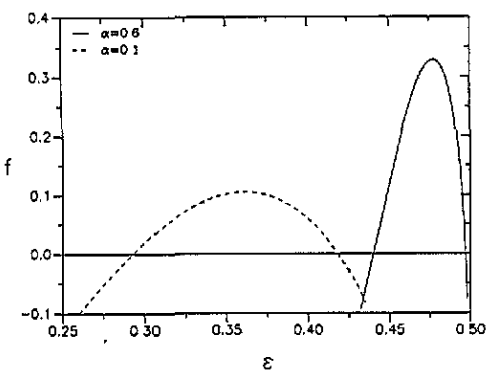

Figure 5. The energy spectra $(\epsilon, f(\alpha, g = \frac{1}{2}, \gamma = \alpha, \epsilon))$ for $\alpha = 0.1$ and $\alpha = 0.6$.

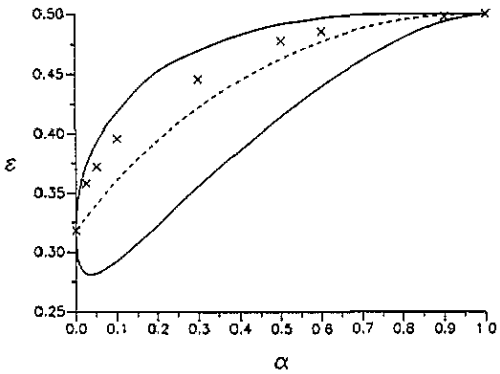

Figure 6. Full lines refer to the upper and the lower zeros in the spectra $(\epsilon, f(\epsilon))$ for different α and at $g = \frac{1}{2}, \gamma = \alpha$. The simulation results [9] for the mean energies of the final states in a serial $T = 0$ Glauber process are represented by crosses. They are found to lie in between the curve of the upper zeros and the curve of the maxima ϵ_{max} in the spectra (broken line).

In contradiction to that, the replica symmetric solution of [9] yields

$$\epsilon_G \to \frac{1}{2}\left(\alpha \to \alpha_G = 1 - \frac{2}{\pi}\right)$$

and for $\alpha > \alpha_G$, it does not predict any spin-glass state. We believe that the thermodynamic spin-glass state exists for all $\alpha < 1$ and that $\alpha_G < 1$ is an artifact of the replica symmetric ansatz in [9]. The corresponding curve of the ϵ_G should be lower than the curve of the right zeros. In fact, numerical results of Kanter and Sompolinsky for the mean energy of the final states in a $T = 0$ serial Glauber process [9] are found to be located between the curve of the right zeros and the curve of the ϵ_{max}. Similar results have already been obtained for the SK spin glass [29].

5. Discussion

Although we calculated an exact upper bound for the number of metastable states in the projection rule neural network the question remains whether the equal sign holds in Jensen's inequality (9). The regions of the parameter space in which

$$\lim_{N\to\infty}\frac{1}{N}\ln\mathcal{N}=\lim_{N\to\infty}\frac{1}{N}\ln\langle\!\langle\mathcal{N}\rangle\!\rangle \tag{27}$$

holds can only be determined by a replica calculation of the self-averaging quantity

$$h=\lim_{N\to\infty}\frac{1}{N}\ln\mathcal{N}=\lim_{N\to\infty}\frac{1}{N}\langle\!\langle\ln\mathcal{N}\rangle\!\rangle. \tag{28}$$

In order to prove equation (27) it has to be shown that all the overlaps $Q_{\rho\sigma}$ between different replica ρ and σ are zero. As is known from the SK model [19] and from other models of neural networks [30], the replica symmetric calculation of h yields good results for the ground-state energy. They are even comparable to a replica symmetry breaking (RSB) thermodynamic calculation.

It will also be interesting to examine whether h reaches the upper bound f in the limit $\alpha\to 1$. Even in the $\gamma=\alpha$ case where the patterns are unstable for $\alpha=1$ we obtain the results

$$f(\alpha, g=\tfrac{1}{2}, \gamma=\alpha, \epsilon=\epsilon_{\max})\to\ln 2 \qquad (\alpha\to 1). \tag{29}$$

Thus it should be clarified whether h has the same limit.

Finally we address two problems that arise in section 4.2 where f is considered as a function of the Hamming distance g.

First the discrepancy between the actual value for the radius of attraction and its lower bound is considered. Obviously the calculation of the number $\mathcal{N}$ of metastable states cannot predict this dynamic property exactly, because $\mathcal{N}$ is much smaller than the number of general states. So during the serial $T=0$ Glauber process the system walks along general states to lower energies. Therefore the dynamics does not notice the existence of exponentially many metastable states.

The second problem is the discontinuity

$$f(g=0)\neq f(g\to 0+)$$

(see figure 3). Since the patterns are metastable, $\mathcal{N}=1$ and $f(g=0)\geqslant 0$ must hold. The discontinuity can be illustrated by noting that the projection matrix discerns between general states and states that lie in the space spanned by the patterns. To make this more explicit, we consider the order parameter

$$Z=\sum_{\mu=2}^{p}a_\mu^2$$

where the a_μ are the coefficients in the superposition (14) for a metastable state $\underline{S}$. In the limit $g\to 0$ we obtain from an analytic treatment of the saddle-point equations

$$Z=4g\frac{1+\gamma}{2(1-\alpha)}+\mathcal{O}(g^2).$$

In general the order parameter Z is a monotonously increasing function of g. Hence for any $g>0$ it follows that $Z>0$, and thus the patterns $\mu=2,\ldots,p$ should contribute to a metastable state $\underline{S}$. When f passes its zero g_0 this requirement and the requirement of metastability can no longer be satisfied simultaneously.

6. Summary

For the projection rule neural network the following main results about the structure of the metastable states have been gained:

(i) An upper bound e^{Nf} for the total number of metastable states $\mathcal{N}$ has been calculated in the limit $N \to \infty$. The remanent magnetization m_r is found to behave similarly to f.

(ii) By investigating the dependence of f on the Hamming distance g we observe that there are areas around the patterns in which nearly no metastable states exist. Since these areas are lower bounds for the sizes of the basins of attraction in a serial dynamic process, the critical storage capacity $\alpha_c(\gamma = \alpha) = 1$ is proved.

(iii) In the $\gamma = \alpha$ case the energy dependence of f has been examined. We show that the replica symmetric ansatz in [9] yields incorrect results for the negative energy ϵ_G of the thermodynamic spin-glass states. Our upper bound for ϵ_G is checked with respect to the mean negative energy ϵ_F of the final states in a serial $T = 0$ Glauber process. As expected, our upper bound ϵ_B fulfils

$$\epsilon_B(\alpha) > \epsilon_F(\alpha)$$

for $\alpha < 1$.

Acknowledgments

We would like to thank M Opper, R D Henkel, M Biehl, H Eissfeller and W Kinzel for many helpful discussions. PK would like to thank the Deutsche Forschungsgemeinschaft (DFG) for their financial support.

Appendix

We calculate the mean number of metastable states in the PRNN by formula (21). By applying the Fourier representation of the δ function and of the Θ function

$$\delta(x) = \int_{-\infty}^{+\infty} \frac{dk}{2\pi} \exp(ikx) \tag{A.1}$$

$$\Theta(x) = \int_0^{\infty} d\lambda\, \delta(\lambda - x) \tag{A.2}$$

we obtain

$$\begin{aligned}\mathcal{N}(\alpha, g, \gamma, \epsilon) &= \sum_{\{S_i\}}{}' \left\{ \prod_{j=1}^{N} \int_0^{\infty} d\lambda_j' \, \delta\left[\lambda_j' - \left(S_j \sum_{k=1}^{N} J_{jk} S_k - \gamma \right) \right] \right\} \\ &\quad \times \delta\left[-\epsilon + \left(\frac{1}{2N} \sum_{j,k} J_{jk} S_j S_k \right) \right] \\ &= \sum_{\{S_i\}}{}' \left(\prod_{j=1}^{N} \int_{\gamma}^{\infty} d\lambda_j \int_{-i\infty}^{+i\infty} \frac{d\mu_j}{2\pi i} \right) \exp\left[\sum_{j=1}^{N} \mu_j \left(\lambda_j - S_j \sum_{k=1}^{N} J_{jk} S_k \right) \right] \\ &\quad \times \int_{-i\infty}^{+i\infty} \frac{d\omega}{2\pi i} \exp\left(-\omega\epsilon + \frac{1}{2N} \omega \sum_{j=1}^{N} \lambda_j \right). \end{aligned} \tag{A.3}$$

The inverse of the correlation matrix is removed by noting that

$$\sum_{k=1}^{N} J_{jk}S_k = \sum_{k=1}^{N}\frac{1}{N}\sum_{\mu,\nu}\xi_j^\mu (C^{-1})_{\mu\nu}\xi_k^\nu S_k = \sum_{\mu,\nu}(C^{-1})_{\mu\nu}\left(\frac{1}{N}\sum_{k=1}^{N}\xi_k^\nu S_k\right)\xi_j^\mu$$
$$= \sum_{\mu=1}^{p}\left(\sum_{\nu=1}^{p}(C^{-1})_{\mu\nu}m_\nu\right)\xi_j^\mu = \sum_{\mu=1}^{p} a_\mu \xi_j^\mu. \tag{A.4}$$

Thus we first introduce

$$1 = \left(\prod_{\mu=1}^{p}\int_{-\infty}^{+\infty} \mathrm{d}m_\mu\right)\delta\left(m_\mu - \frac{1}{N}\sum_{k=1}^{N}\xi_k^\mu S_k\right) \tag{A.5}$$

in equation (A.3) using the Fourier representation (A.1). In a second step the transformation of variables

$$m_\mu = \sum_{\nu=1}^{p} C_{\mu\nu}a_\nu \tag{A.6}$$

is applied and we obtain

$$\mathcal{N}(\alpha, g, \gamma, \epsilon)$$
$$= \sum_{\{S_i\}}{}'\left(\prod_{j=1}^{N}\int_{\gamma}^{\infty}\mathrm{d}\lambda_j\int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty}\frac{\mathrm{d}\mu_j}{2\pi\mathrm{i}}\right)\left(\prod_{\nu=1}^{p}\int_{-\infty}^{+\infty}\mathrm{d}a_\nu\int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty}\frac{\mathrm{d}l_\nu}{2\pi\mathrm{i}}\right)\det C$$
$$\times\exp\left[\sum_{j=1}^{N}\mu_j\left(\lambda_j - \sum_{\mu=1}^{p}a_\mu S_j\xi_j^\mu\right)\right]\int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty}\frac{\mathrm{d}\omega}{2\pi\mathrm{i}}\exp\left(-\omega\epsilon + \frac{1}{2N}\omega\sum_{j=1}^{N}\lambda_j\right)$$
$$\times\exp\left[\sum_{\mu=1}^{p}l_\mu\left(\frac{1}{N}\sum_{j=1}^{N}S_j\xi_j^\mu\lambda_j - \frac{1}{N}\sum_{k=1}^{N}\xi_k^\mu S_k\right)\right]. \tag{A.7}$$

We rescale

$$l_1' = \frac{l_1}{N} \qquad a_\mu' = \sqrt{N}a_\mu \qquad l_\mu' = \frac{1}{\sqrt{N}}l_\mu \qquad \text{for } \mu \geqslant 2$$

and omit the primes in the following. It can be shown by a replica calculation that $\det C$ is self-averaging [27] for $N \to \infty$ and that we can set

$$\det C \sim \exp N[-\alpha - (1-\alpha)\ln(1-\alpha)]. \tag{A.8}$$

The direct average over the patterns $\mu = 2, \ldots, p$ can then easily be performed, and we obtain

$$\langle\!\langle \mathcal{N}(\alpha, g, \gamma, \epsilon)\rangle\!\rangle = \binom{N}{Ng}\left(\prod_{j=1}^{N}\int_{\gamma}^{\infty}\mathrm{d}\lambda_j\int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty}\frac{\mathrm{d}\mu_j}{2\pi\mathrm{i}}\right)$$
$$\times\left(\prod_{\nu=1}^{p}\int_{-\infty}^{+\infty}\mathrm{d}a_\nu\int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty}\frac{\mathrm{d}l_\nu}{2\pi\mathrm{i}}\right)\int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty}\frac{\mathrm{d}\omega}{2\pi\mathrm{i}}N\det C$$
$$\times\exp\left[\sum_{j=1}^{N}\mu_j\lambda_j + a_1\left(\sum_{j=1}^{Ng}\mu_j - \sum_{Ng+1}^{N}\mu_j\right) + l_1\left(\sum_{j=1}^{Ng}(1-\lambda_j) + \sum_{Ng+1}^{N}(\lambda_j - 1)\right)\right.$$
$$\left. - N\omega\epsilon + \tfrac{1}{2}\omega\sum_{j=1}^{N}\lambda_j + \sum_{j=1}^{N}(\tfrac{1}{2}X(\lambda_j-1)^2 - Y\mu_j(\lambda_j-1) + \tfrac{1}{2}Z\mu_j^2)\right]$$

where

$$X = \frac{1}{N}\sum_{\mu=2}^{p} l_\mu^2 \qquad Y = \frac{1}{N}\sum_{\mu=2}^{p} l_\mu a_\mu \qquad Z = \frac{1}{N}\sum_{\mu=2}^{p} a_\mu^2. \tag{A.9}$$

Finally we introduce these abbreviations as order parameters by

$$1 = \int_{-\infty}^{+\infty} \mathrm{d}X \int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty} \frac{\mathrm{d}\hat{x}}{2\pi\mathrm{i}} N \exp\left[N\hat{x}\left(X - \frac{1}{N}\sum_{\mu=2}^{p} l_\mu^2\right)\right] \tag{A.10}$$

and similarly for Y and Z.

It is then possible to decouple the variables a_ν, l_ν from the variables λ_j, μ_j and the saddle-point method can be applied to calculate

$$f(\alpha, g, \gamma, \epsilon) = \lim_{N\to\infty} \frac{1}{N} \ln\langle\langle \mathcal{N} \rangle\rangle. \tag{A.11}$$

The saddle-point equations for the variables $\hat{x}$, $\hat{y}$, $\hat{z}$ become algebraic. After the insertion of $\hat{x}$, $\hat{y}$, $\hat{z}$, equation (24) is obtained.

References

[1] Binder K and Young A P 1986 *Rev. Mod. Phys.* **58** 801
[2] Mézard M, Parisi G and Virasoro M A 1987 *Spin Glass Theory and Beyond* (Singapore: World Scientific)
[3] Fischer K H and Hertz J 1991 *Spin Glasses* (Cambridge: Cambridge University Press)
[4] Amit D J 1989 *Modelling Brain Function: The World of Attractor Neural Networks* (Cambridge: Cambridge University Press)
[5] Hertz J, Krogh A and Palmer R G 1991 *Introduction to the Theory of Neural Computation* (Redwood City: Addison-Wesley)
[6] Kohonen T 1988 *Self-organisation and Associative Memory* (Berlin: Springer)
[7] Personnaz L, Guyon I and Dreyfus G 1985 *J. Physique* **46** L359
[8] Personnaz L, Guyon I and Dreyfus G 1986 *Phys. Rev.* A **34** 4217
[9] Kanter I and Sompolinsky H 1987 *Phys. Rev.* A **35** 380
[10] Glauber R J 1963 *J. Math. Phys.* **4** 294
[11] Gutfreund H, Reger J D and Young A P 1988 *J. Phys. A: Math. Gen.* **21** 2775
[12] Frumkin A and Moses E 1986 *Phys. Rev.* A **34** 714
[13] Fontanari J F and Köberle R 1987 *Phys. Rev.* A **36** 2475
[14] Feller W 1957 *An Introduction to Probability Theory and its Applications* (New York: Wiley)
[15] Sherrington D and Kirkpatrick S 1975 *Phys. Rev. Lett.* **35** 1792
Sherrington D and Kirkpatrick S 1978 *Phys. Rev.* B **17** 4384
[16] Parisi G 1980 *J. Phys. A: Math. Gen.* **13** L115
[17] Tanaka F and Edwards F S 1980 *J. Phys. F: Met. Phys.* **10** 2769
[18] Bray A J and Moore M A 1980 *J. Phys. A: Math. Gen.* **22** L469
[19] Roberts S A 1981 *J. Phys. C: Solid State Phys.* **14** 3015
[20] Hopfield J J 1982 *Proc. Natl. Acad. Sci. USA* **79** 2554
[21] Gardner E 1986 *J. Phys. A: Math. Gen.* **19** L1047
[22] Fukai T 1990 *J. Phys. A: Math. Gen.* **23** 249
[23] Fukai T and Shiino M 1990 *Phys. Rev.* A **42** 7459
[24] Amit D J, Gutfreund H and Sompolinsky H 1987 *Ann. Phys.* **173** 30
[25] Odlyzko A M 1988 *J. Combin. Theory* A **17** 124
[26] Abramovitz M and Stegun I 1965 *Handbook of Mathematical Functions* (Washington, DC: National Bureau of Standards)
[27] Kuhlmann P 1990 *Diploma Thesis* University of Giessen
[28] Henkel R D and Opper M 1991 *J. Phys. A: Math. Gen.* **24** 2201
[29] Henkel R D and Kinzel W 1987 *J. Phys. A: Math. Gen.* **20** L727
[30] Meir R 1992 *Preprint*