Structure of metastable states in the Hopfield model

LETTER TO THE EDITOR

Structure of metastable states in the Hopfield model

E Gardner

Department of Physics, University of Edinburgh, Mayfield Road, Edinburgh EH9 3JZ, UK

Received 30 May 1986

Abstract. An upper bound for the number of metastable states in the Hopfield model is calculated as a function of the Hamming fraction from an input pattern. For all finite values of α, the ratio of number of patterns to nodes, the Hamming fraction from the input pattern to the nearest metastable state is finite. When $\alpha < 0.113$, the bound also implies that there is a gap between a set of states close to the input pattern and another set centred around the Hamming fraction 0.5 from it.

The Hopfield model (Hopfield 1982) is a pattern recognition model which stores a set of n input vectors as N bit numbers. The number of patterns which can be stored assuming single spin-flip dynamics behaves as $N/2 \ln N$ for large N (Weisbuch and Fogelman Soulie 1985, Bruce *et al* 1986). This means that for finite values of $\alpha = n/N$ no input pattern will be perfectly recalled. However, a mean-field analysis of the thermodynamics of the model (Amit *et al* 1985) shows that for sufficiently low temperature and for α less than a critical value α_c, there exists a metastable state close to the input pattern which is perfectly recalled and that at a lower value of α, α_1, this state has a lower energy than that of the spin glass state (which exists for all α and is uncorrelated with the input pattern) and so becomes a ground state. At zero temperature, the replica symmetric ansatz gives $\alpha_c = 0.138$ and $\alpha_1 = 0.052$. Although this ansatz is known to be incorrect at zero temperature, replica breaking effects in α_c and α_1 are expected to be small. For $\alpha < \alpha_c$ the model is expected to have associative memory and so be useful as a pattern recognition model: iteration from a state within the basin of attraction of the correlated state will increase its correlation with the input vector.

However, there are also other metastable states which do not appear in the thermodynamic calculation but which can be important in the dynamics. The number of such states in a spin glass is exponentially large in N (Bray and Moore 1980, 1981). Therefore in the Hopfield model one would expect that, in addition to the spin glass state at Hamming distance 0.5 from the input pattern, there are other non-equilibrium states at distances centred around this value. Similarly there should be another exponentially large set of states with distances approximately centred around the correlated state which appears in the thermodynamic calculation.

Non-equilibrium states are relevant, for example, to the results of iteration from an input pattern. If one assumes the existence of only two thermodynamic states, one correlated with the input pattern and one at a distance 0.5 from it, then assuming one iterates to the correlated states for $\alpha < \alpha_c$, the final Hamming distance would have a sharp jump at $\alpha = \alpha_c$ from the correlated state value 0.015 to the uncorrelated value 0.5. However, numerical results (Amit *et al* 1986, Bruce *et al* 1986) using single spin-flip

0305-4470/86/161047+06$02.50 © 1986 The Institute of Physics

dynamics show that the jump is not sharp. There are large finite-size effects which persist up to values of α much larger than α_c. The simplest possiblity is that these effects can be analysed as a first-order phase transition. For sufficiently large values of N, there is a clear gap between two bands of possible final states; there is a relative probability $\sim e^{NF(\alpha)}$ of iteration to a non-equilibrium state in the correlated band relative to the uncorrelated band and the phase transition is defined to be at the value α_0 of α where $F(\alpha_0)=0$. Only in the limit $N\to\infty$ does the jump in the Hamming distance at α_0 become sharp. It is possible that α_0 is not equal to α_c; it could be at a lower value if the iteration allows one to jump over the nearest band of states or it could be at a higher value if the correlated band of non-equilibrium states exists and has higher weight than the uncorrelated band for $\alpha>\alpha_c$. Using the above dynamics, α_0 turns out to be ~ 0.145, just above the replica symmetric prediction and so the combination of replica breaking and dynamical effects seems to be small. The final Hamming distances are equal to their remanent values (Kinzel 1985) corresponding to the large entropy of non-equilibrium states. For the spin glass state the final value is less than 0.5 (Amit *et al* 1986). For the correlated band, the effect is too small to be seen numerically at least for the above single spin-flip dynamics. Another possiblity is that there are values of α for which the probability of iterating to each band remains finite as $N\to\infty$; it could be that sample to sample and pattern to pattern fluctuations in the final Hamming distance remain non-zero in this limit.

In this letter we will calculate the expectation of the number of metastable states as a function of the Hamming distance Ng from an input pattern. This is an upper bound for the typical number of such states. Since, for any particular realisation of the other input patterns, the number of metastable states, $\mathcal{N}(N,g)$ behaves as $e^{NS(g)}$ for large N, the quantity $\log\mathcal{N}(N,g)$ is extensive and so should be self-averaging. This means that the value of $\mathcal{N}(N,g)$ for a given realisation is (for large N) almost always equal to $e^{N\langle S(g)\rangle}$ where $\langle\,\rangle$ represents an average over all possible realisations; other values of $\mathcal{N}(N,g)$ occur with probabilities which are exponentially small in N. Therefore, by convexity of the distribution of $S(g)$, the expectation of $\mathcal{N}(N,g)$, $\langle e^{NS(g)}\rangle$ is an upper bound for this typical value.

The calculation of $\langle e^{NS(g)}\rangle$ will be done for the $(1,-1)$ model although it can also be generalised to other dynamics. This model is defined in terms of Ising spins S_i which take the values $+1$ or -1 at each site i, $i=1,\ldots,N$. A spin is given the value $+1$ if

$$\sum_{j\neq i} T_{ij}S_j>0 \tag{1}$$

and the value -1 otherwise. This dynamics can be done either in series or in parallel. The values of T_{ij} on each link are given by the storage prescription (Hopfield 1982)

$$T_{ij}=\frac{1}{N}\sum_{r=1}^{n} S_i^r S_j^r \tag{2}$$

where the $\{S_i^r, i=1,\ldots,N\}$ $r=1,\ldots,n$ are the patterns one wants to store.

If a state, $\{S_i, i=1,\ldots,N\}$, at distance Ng from an input vector, r, is to be perfectly recalled, it must be a stable state of the dynamics defined by (1). This means that the quantity

$$R_i^r=S_i\sum_{j\neq i} T_{ij}S_j \tag{3}$$

must be positive on each site i.

Separation of the term coming from the input vector r from the interference term coming from the other vectors gives

$$R_i^r=\begin{cases}1-2g+\dfrac{1}{N}\sum\limits_{j\neq i}\sum\limits_{s\neq r} S_i^sS_j^sS_iS_j & \text{for the } Ng \text{ values of } i \text{ for which } S_i^r=S_i\\ -1+2g+\dfrac{1}{N}\sum\limits_{j\neq i}\sum\limits_{s\neq r} S_i^sS_j^sS_iS_j & \text{otherwise.}\end{cases}\tag{4}$$

The expectation, over all possible realisations of the other input patterns,

$$\begin{aligned}\langle \mathrm{e}^{NS(g)}\rangle &= \mathrm{e}^{NF(g)}\\ &=\left\langle\prod_i \theta(R_i^r)\right\rangle\end{aligned}\tag{5}$$

can be obtained using the integral representation for the θ functions

$$\theta(\tau)=\int_0^\infty \frac{\mathrm{d}\mu}{2\pi}\int_{-\infty}^{\infty}\mathrm{d}x\,\exp[\mathrm{i}x(\mu-\tau)]$$

as a saddle point over two parameters a and b:

$$\begin{aligned}F(g,\alpha)=\alpha\bigg[&b+\tfrac{1}{2}\ln a-\frac{1}{2}+\frac{(1-b)^2}{2a}+(1-g)\ln\left(\frac{1}{\sqrt{2\pi}}\int_{-t}^{\infty}\mathrm{d}\mu\,\exp(-\mu^2/2)\right)\\ &+g\ln\left(\frac{1}{\sqrt{2\pi}}\int_{u}^{\infty}\mathrm{d}\mu\,\exp(-\mu^2/2)\right)-g\ln g-(1-g)\ln(1-g)\bigg]\end{aligned}\tag{6}$$

where

$$\begin{aligned}t&=\frac{1-2g-b\alpha}{\sqrt{a\alpha}}\\ u&=\frac{1-2g+b\alpha}{\sqrt{a\alpha}}.\end{aligned}\tag{7}$$

The mean-field equations for a and b can be solved numerically and the results for $\alpha=0.1$, 0.113 and 0.2 are plotted in figures 1(a), (b) and (c), respectively. If $F(g,\alpha)$ is negative, then since the typical value of $\mathcal{N}(N,g)$ is bounded above by zero in the thermodynamic limit, there are no states in this limit at these values of g.

For $g=0$, $F(g,\alpha)$ is negative for all finite values of α and so the probability that an input vector is perfectly stored vanishes in the thermodynamic limit. At $g=0$, the typical value of $\mathcal{N}(N,g)$ is equal to its expectation and expansion of $F(0,\alpha)$ around $\alpha=0$ allows one to prove that the storage capacity n_0 of the network behaves as $N/2\ln N$ (Bruce *et al* 1986).

The Hamming fraction $g_0(\alpha)$ from the input vector to the nearest metastable state is finite for all values of α since $F(g,\alpha)$ is negative in a finite region around $g=0$. For small α, expansion around $\alpha=0$ gives

$$g_0(\alpha)=\left(\frac{\alpha}{2\pi}\right)^{1/2}\mathrm{e}^{-1/2\alpha}.\tag{8}$$

For values of $\alpha<0.113$ there is a narrow band of values of $g>g_0(\alpha)$ for which $F(g,\alpha)$ is positive and this band contains the correlated state found in the thermodynamic calculations. The width of the band $\Delta g(\alpha)$ behaves as

$$\left(\frac{7}{2\alpha}\right)^{1/2}(g_0(\alpha))^{3/2}\tag{9}$$

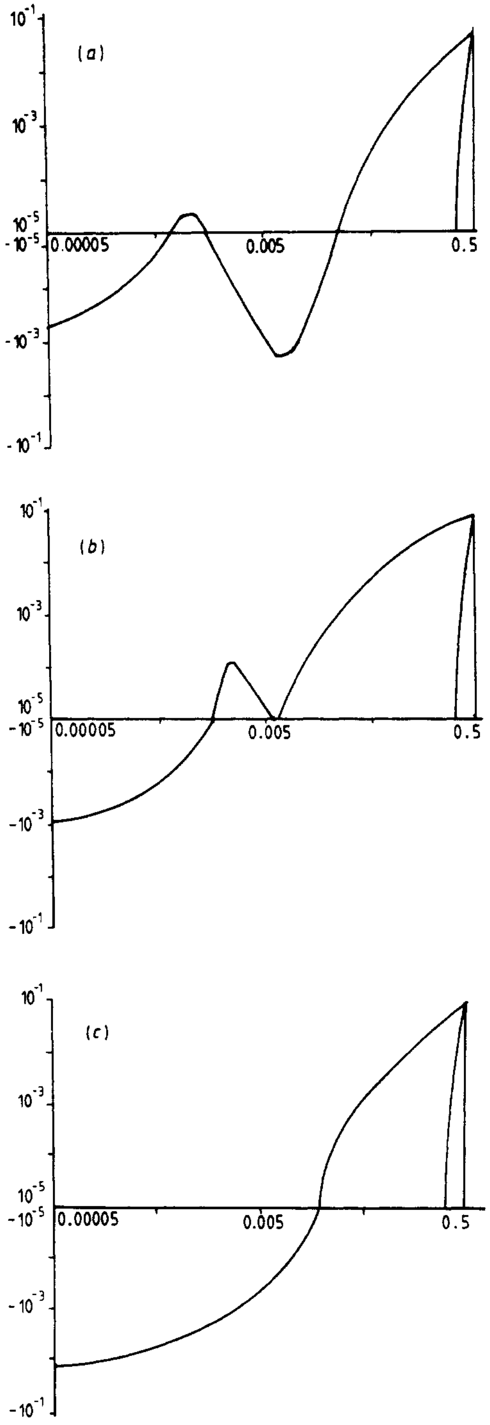

Figure 1. The function $F(g)$ for (a) $\alpha = 0.1$ where the height of the first peak is 2.5×10^{-5}, (b) $\alpha = 0.113$ where the height of the first peak is 1.1×10^{-4} and (c) $\alpha = 0.2$. The lower curve shows $S(g)$ if the states were distributed according to phase space.

for small α and the maximum value of F in this region behaves as

$$\frac{7}{4}\frac{g_0^2(\alpha)}{\alpha} \tag{10}$$

near $\alpha = 0$.

For values of $\alpha < 0.113$ there is a gap between this band of states and the more distant band of metastable states centred around $g = 0.5$ which exist for all values of α. The bound is, however, not strong enough to prove the existence of a gap for values of α as large as α_c. The second band is much broader than would be expected if the states were distributed according to phase space. The bound is therefore consistent with the existence for all values of α of additional states with macroscopic correlations with the input vector. The existence of these states would explain the larger remanent magnetisation obtained in numerical simulations by iterating from an input vector relative to that obtained by iterating from a random vector (Toulouse *et al* 1986).

For $g = 0.5$ the bound should give the exact number of metastables states (Bray and Moore 1980). The total number of metastable states is given by this value since $g = 0.5$ is the saddle point in the integral over g:

$$\mathcal{N}(N) = \int \mathrm{d}g\, \mathcal{N}(N, g). \tag{11}$$

For large α, the result of Bray and Moore (1980) for the Sherrington-Kirkpatrick model is recovered and for small α one has

$$\mathcal{N}(N) = \mathrm{e}^{N}[\tfrac{1}{2}\alpha(\ln(2/\pi\alpha) - 1) + \mathrm{O}(\alpha^2)]. \tag{12}$$

it is also possible to include the energy

$$E = -\tfrac{1}{2} \sum_{\substack{i,j \\ (i \neq j)}} T_{ij}S_iS_j = N\varepsilon \tag{13}$$

in the above calculations. The expectation of the number of metastable states at distance Ng and of the energy per site ε, $\langle\mathcal{N}(N, g, \varepsilon)\rangle$, is given by

$$\langle\mathcal{N}(N, g, \varepsilon)\rangle = \left\langle \prod_i \delta\left(\varepsilon + \frac{1}{2N} \sum_{\substack{i,j \\ (i \neq j)}} T_{ij}S_iS_j\right)\theta(R_i')\right\rangle. \tag{14}$$

For finite values of g, the number of such states decreases from the value at the saddle point over ε, at $\varepsilon_0(g, \alpha)$, which gives $\langle\mathcal{N}(N, g)\rangle$ and the gap between the two bands of metastable states widens. So, as ε decreases the maximum value of α at which one can prove the existence of a gap increases. Since the maximum of $\mathcal{N}$ (N, g, ε) is again at $g = 0.5$, there is an upper bound for the total number of states as a function of energy

$$\begin{aligned}\langle\mathcal{N}(N, \varepsilon)\rangle &= \mathrm{e}^{NG(\varepsilon)} \\ &= \langle\mathcal{N}(N, 0.5, \varepsilon)\rangle\end{aligned} \tag{15}$$

where

$$G(\varepsilon) = G_0(\varepsilon) + \frac{\alpha}{2}\left(\frac{1}{[(1+(\varepsilon/\sqrt{\alpha})^2)^{1/2} - \varepsilon/\sqrt{\alpha}]^2} + 2\ln\left\{\left[1+\left(\frac{\varepsilon}{\sqrt{\alpha}}\right)^2\right]^{1/2} - \frac{\varepsilon}{\sqrt{\alpha}}\right\}\right) - \varepsilon^2 \tag{16}$$

and $G_0(\varepsilon)$ is given by the extremum over the variable x of

$$\tfrac{1}{2}x^2+2x\varepsilon+\ln\left(2\int_{-x}^{\infty}\frac{\mathrm{d}\mu}{\sqrt{2\pi}}\exp(-\mu^2/2)\right)+\varepsilon^2. \qquad (17)$$

It would be interesting to repeat these calculations using replicas. This would allow one to calculate the typical value of $\mathcal{N}(N, g, \varepsilon)=\exp(N\langle S(g, \varepsilon)\rangle)$. For $g=0.5$ the replica symmetric solution is identical to the above solution and the replica symmetric solution should remain correct at least for a finite region around $\varepsilon=\varepsilon_0(0.5, \alpha)$ and $g=0.5$. However, replica symmetry breaking will probably be important for the lower energy states (Bray and Moore 1981). In this case, the replica symmetric solution would provide a better upper bound for $\mathcal{N}(N, g, \varepsilon)$. Dynamical effects and non-equilibrium states should also be important for other properties of neural networks. Associative memory properties, in particular, do depend on the kind of dynamics which is used and can be studied using similar methods to those described here (Derrida *et al* 1986).

I would like to thank B Derrida and G Toulouse for useful comments. Financial support was from the Science and Engineering Research Council, UK.

References

Amit, D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. Lett.* **55** 1530
—— 1986 *Racah Institute of Physics preprint*
Bray A J and Moore M A 1980 *J. Phys. C: Solid State Phys.* **13** L469
—— 1981 *J. Phys. C: Solid State Phys.* **14** 1313
Bruce A D, Gardner E and Wallace D J 1986 *Edinburgh preprint 387*
Derrida B, Gardner E and Mottishaw P 1986 *Saclay preprint*
Hopfield J J 1972 *Proc. Natl Acad. Sci. USA* **79** 2554
Kinzel W 1985 *Julich preprint*
Toulouse G, Dehaene S and Changeux J P 1986 *Proc. Natl Acad. Sci. USA* in press
Weisbuch G and Fogelman Soulie F 1985 *J. Physique Lett.* **46** L623