Effects of pattern bias on retrieval dynamics in $0, 1$ and $-1, 1$ Hopfield networks

1998 J. Phys. A: Math. Gen. 31 L171

(http://iopscience.iop.org/0305-4470/31/9/002)

J. Phys. A: Math. Gen. **31** (1998) L171–L176. Printed in the UK PII: S0305-4470(98)89373-6

LETTER TO THE EDITOR

Effects of pattern bias on retrieval dynamics in $\mathbf{0,1}$ and $\mathbf{-1,1}$ Hopfield networks

K F Wilson and C F Osborne
Department of Physics, Monash University, Melbourne, Australia

Received 17 November 1997

Abstract. The effects of pattern bias on the retrieval of patterns stored in $-1,1$ and $0,1$ Hopfield networks is examined. Pattern bias is the relative proportion of -1 and 1 pattern bits. We find both analytically and via numerical simulation, that $-1,1$ networks are not affected by pattern bias. In contrast, $0,1$ networks will form asymmetric basins of attraction for a pattern and its inverse depending on the net bias of all stored patterns.

1. Analytic results

Hopfield [1] networks consist of units which may take possible states $S_i = -1, 1$, ($i = 1, 2, \ldots, N$ where N is the number of units in the net). For a $0,1$ network, the possible unit states are $s_i = 0, 1$. The connections between the units are described by a weight matrix, w_{ij}, ($i, j = 1, 2, \ldots, N$) which is symmetric and has positive or zero diagonal. The weight w_{ij} is the strength with which unit j affects unit i. Patterns consist of N bits, $\xi_i = \pm 1$ ($i = 1, 2, \ldots, N$). Pattern storage is most simply done using the Hebb [2] rule, equation (1), which makes the stored patterns energy minima in the configuration space of the network,

$$w_{ij} = 1/N \sum_{\mu} \xi_i^{\mu} \xi_j^{\mu}. \tag{1}$$

The probability of a unit taking state $S_i = \pm 1$, in the $-1,1$ net, is

$$\mathrm{P}(S_i = \pm 1) = (1 + \exp(- \pm 2\beta h_i))^{-1} \tag{2}$$

where

$$h_i = \sum_j w_{ij} S_j \tag{3}$$

is the input to the unit from the rest of the network and $\beta = 1/T$ is the temperature or noise parameter of the system.

The mean-field approach gives the mean state of any unit as

$$\langle S_i \rangle = \tanh(\beta \langle h_i \rangle). \tag{4}$$

Using the condition

$$\langle S_i \rangle = m\xi_i^{\nu} \tag{5}$$

where m is the overlap of the network configuration with pattern ν, we arrive at the well known result

$$m = \tanh(\beta m) \tag{6}$$

0305-4470/98/090171+06$19.50

We now compare this with a 0, 1 network, where the possible unit states are $s_i = 0, 1$. The probability of a unit taking state 0 is

$$P(s_i = 0) = (1 + \exp(+2\beta h_i))^{-1} \tag{7a}$$

and the probability of a unit taking state 1 is

$$P(s_i = 1) = (1 + \exp(-2\beta h_i))^{-1}. \tag{7b}$$

From these probabilities the mean state for unit i is

$$\begin{aligned}\langle s_i \rangle &= P(s_i = 0) \cdot 0 + P(s_i = 1) \cdot 1 \\ &= (1 + \exp(-2\beta h_i))^{-1} \\ &= \left(1 + \exp\left(-2\beta \sum_{j\mu} \xi_i^\mu \xi_j^\mu \langle s_j \rangle\right)\right)^{-1}.\end{aligned} \tag{8}$$

The equivalent condition to (5) where the patterns are stored according (1) for this network is

$$\langle s_i \rangle = (m\xi_i^\nu + 1)/2. \tag{9}$$

Substituting this into (8) and expanding gives

$$m\xi_i^\nu = 2\left(1 + \exp\left(-\beta m\xi_i^\nu - \beta/N \sum_{j\neq i,\mu} \xi_i^\mu \xi_j^\mu m\xi_j^\nu - \beta/N \sum_{j\mu} \xi_i^\mu \xi_j^\mu\right)\right)^{-1} - 1. \tag{10}$$

The second term in the exponential will be negligible for small numbers of patterns, and hence may be ignored leaving

$$m\xi_i^\nu = \tanh[\beta/2(m\xi_i^\nu + b)] \tag{11}$$

where

$$b = 1/N \sum_{j\mu} \xi_i^\mu \xi_j^\mu. \tag{12}$$

This term is the bias of the patterns, indicating the relative proportions of −1's and 1's in the stored patterns. For the one-pattern case, $\mu = \nu = 1$, b is the bias of the only stored pattern. This means that there will be a shift in m when the stored pattern is not an equal mix of 1's and −1's. In particular, for a positively biased pattern ($b > 0$) the network will retrieve the pattern itself rather than the inverse, and vice versa for a negatively biased pattern ($b < 0$).

For the many pattern case, if b is greater than 0, then an unbiased pattern will still show retrieval of that pattern rather than its inverse. Hence it is possible to store a set of unbiased patterns, such that the only basins of attraction are the original patterns, by storing a biased pattern at the same time. The larger the bias, the smaller the basin of attraction for the pattern inverse, hence the greater the probability of retrieving the pattern in preference to the inverse.

Increasing the temperature will decrease the size of both basins of attraction, the smaller disappearing earlier than the larger. Hence at low pattern bias it is possible to remove an unwanted basin of attraction by an increase of temperature.

Figures 1(a) and (b) show the overlap for both a −1, 1 and 0, 1 network with $b = 0$, and overlap for a 0, 1 net with positive and zero bias. Note that figure 1(a) shows that the 0, 1 net is more sensitive to temperature than the −1, 1 net. There is no change in overlap with pattern bias for the −1, 1 network.

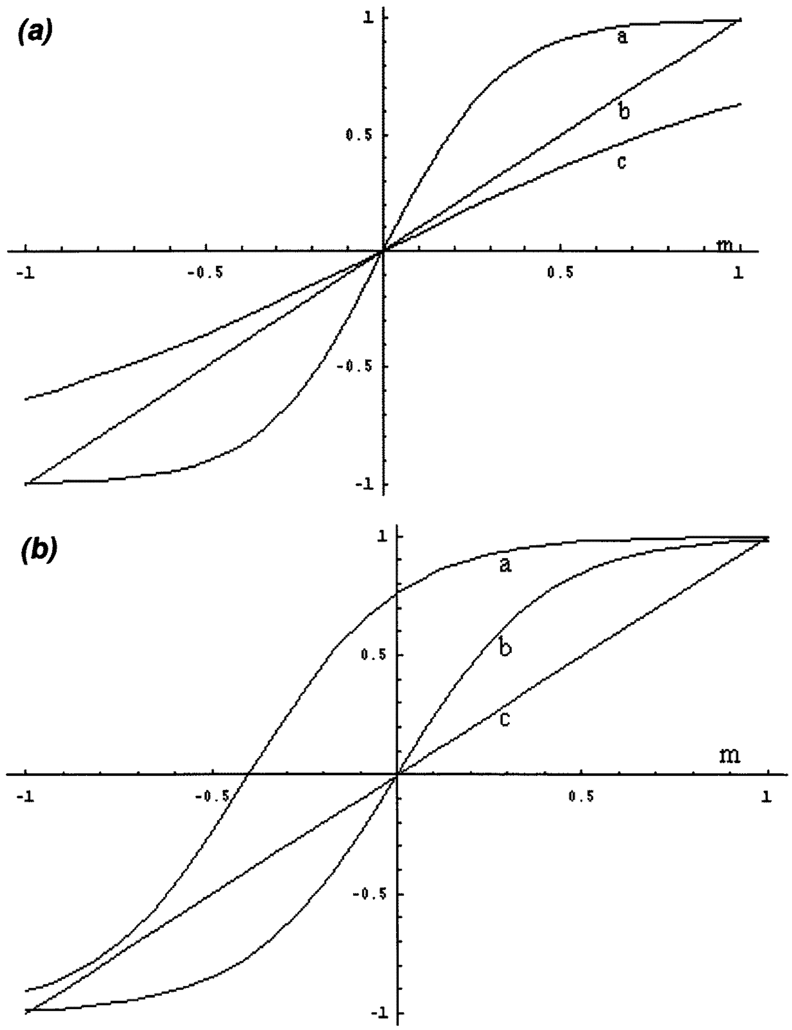

Figure 1. (a) Overlap m. Curve a is equation (6), curve b is the straight line $m = m$, and curve c is equation (12), at $\beta = 3$. The only solution at this temperature for function c, the overlap for the $0, 1$ network, is the trivial $m = 0$ solution, while function a, for the $-1, 1$ net still has non-zero solutions. (b) Curve a is equation (12) with $\beta = 10$ and $b = 0.2$ (biased), curve b is equation (12) with $b = 0$ (unbiased). Curve c is the straight line $m = m$. The biased curve has only one solution, while the unbiased has both the trivial $m = 0$ solution and the symmetric pattern and pattern-inverse retrieval solutions.

2. Simulation results

Simulations were run on a Silicon Graphics SGI 02 R1000 in C++. The networks used were of size $N = 200$ units.

Patterns were stored using the Hebb [2] rule, and the $0, 1$ and $-1, 1$ networks used the same weight matrix and started in equivalent states.

For a particular stored pattern which was slightly negatively biased ($b = -0.04$) it was found that at full connectivity the $0, 1$ network retrieved the inverse of the stored pattern

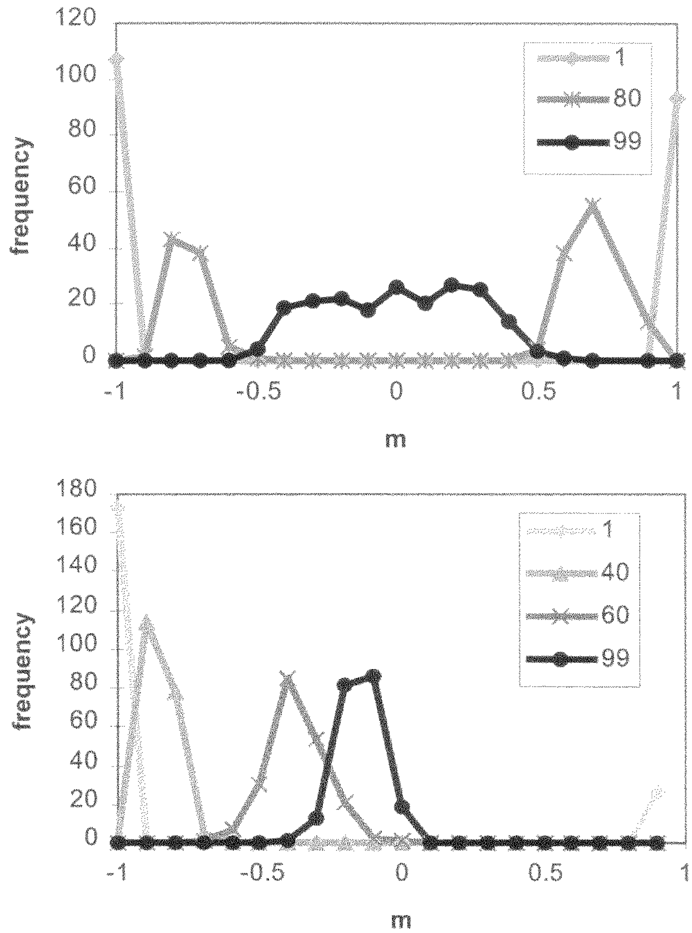

Figure 2. (a) Overlap, m, with stored pattern for $-1, 1$ network at temperatures 1°, 80° and 99°. Overlaps less than 0 indicate retrieval of inverse, overlaps greater than 0 indicate retrieval of the original pattern. Frequency is the number of occurrences out of 200 samples. (b) Overlap, m, with stored pattern for $0, 1$ network at temperatures 1°, 40°, 60° and 99°. Frequency is the number of occurrences out of 200 samples. There is retrieval of the original pattern only at low temperature, and with very low frequency.

far more frequently than the original. The $-1, 1$ network retrieved the pattern and its inverse with equal frequency. Figure 2(a) shows a frequency versus overlap plot for the $-1, 1$ network at low and high temperatures for a slightly biased pattern, figure 2(b) shows frequency versus overlap for the $0, 1$ network.

Simulations were run with single-pattern storage, with pattern bias varying between -0.3 and 0.3 in steps of 0.01. An average of 100 patterns with each possible bias was used and the network was started in 100 random configurations. The network was allowed to reach equilibrium and the final state compared with the stored pattern. Figure 3 shows a plot of average number of pattern recalls against b, for both $0, 1$ and $-1, 1$ networks. The two nets stored the same patterns (hence they had identical weight matrices) and started from equivalent random starting states after each storage. It is obvious that the retrieval properties of the -1,1 network are unaffected by pattern bias, while the 0,1 network is

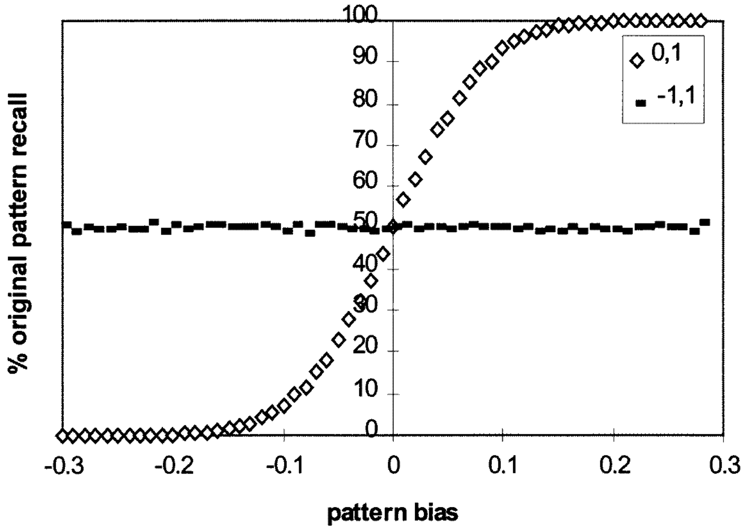

Figure 3. Frequency of original pattern retrieval as a function of bias, b, for both 0, 1 and $-1, 1$ networks. Approximately 100 patterns of each bias were used, with 100 random start states per pattern. The $-1, 1$ network show a standard deviation of 5% or less in pattern recall, while the 0, 1 network show standard deviation between 0% and 5%, the high deviations occurring for the less biased patterns.

affected as predicted by the analytic results.

When a mixture of unbiased and positively biased patterns were stored into a 0,1 network it was found that the biased patterns were retrieved entirely in their original form, while the unbiased patterns were retrieved mostly in original form, with the occasional retrieval of the pattern inverse. For example, two positively biased patterns with $b = 0.2$ and one unbiased pattern, $b = 0$, were stored. The retrievals of the biased patterns were entirely of the original pattern, while the retrievals of the unbiased pattern were 77% the original pattern, and the remaining 23% were retrievals of the inverse.

3. Conclusions

We have shown that for a $-1, 1$ network the overlap of the network configuration with stored patterns and hence the retrieval behaviour of the network, is unaffected by any pattern bias. In contrast, the 0, 1 network, usually considered to be equivalent to a $-1, 1$ net, will be affected by pattern bias. By slightly biasing a pattern it is possible to remove the basin of attraction for the pattern inverse.

In the multiple-pattern case, a series of unbiased patterns may be stored without their inverses by the storage of an additional biased pattern. Hence, a $-1, 1$ network with p patterns stored (where p is less than the network capacity) will have $2p$ basins of attraction. A 0, 1 network with a large enough overall pattern bias (or higher temperature) will have only p basins of attraction.

This work is especially significant for biological simulation. Biological neurons are either firing, and releasing transmitters, or not firing. This is closer to a 0, 1 state system, than a $-1, 1$ system in which a non-firing neuron has exactly the opposite effect to that when firing, rather than no effect at all. In addition, the manner in which the 0, 1 nets have a preferred memory for biased patterns is more like the way biological memory functions, in that we remember an event rather than both that event and its exact inverse.

References

[1] Hopfield J J 1982 *Proc. Natl Acad. Sci., USA* **79** 2554–8
[2] Hebb D O 1949 *Organisation of Behaviour* (New York: Wiley)