Journal of Statistical Physics, Vol. 67, Nos. 1/2, 1992

Growth of Correlation in the Hopfield Model

P. Sen[1]

Received June 24, 1991; final October 23, 1991

Introducing a finite correlation ρ_0 between any two learned patterns (others remaining uncorrelated), we observe in a numerical simulation that the Hopfield model stores these two patterns with correlation ρ_f such that $\rho_f \geqslant \rho_0$ for any loading capacity α. The patterns are memorized perfectly (with $\rho_f = \rho_0$) up to $\alpha \simeq 0.05$ for finite correlations ρ_0 not exceeding a value $\rho_c(\alpha)$, where $\rho_c(\alpha)$ decreases continuously to zero at $\alpha \simeq 0.05$.

KEY WORDS: Correlated patterns; Hopfield model; noise; dynamics.

In the Hopfield model of a neural network,[1] each neuron is represented by two-state (active/passive) Ising spins $S_i = \pm 1$ and the synaptic connections are represented by spin–spin interactions J_{ij} constructed following the Hebb rule: $J_{ij} = \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu / N$. Here, ξ_i^μ represent the patterns (configurations) to be learnt or memorized (p is the number of patterns and N is the total number of neurons). Each ξ_i^μ is taken to be an independent random quenched variable in the Hopfield model.

The Hopfield model is a solvable one in the thermodynamic limit. In the exact solution of this model,[2] it is necessary to assume that the learned patterns are all completely random and uncorrelated. Models with correlated patterns have been studied (both numerically and analytically), where the average correlation between all the patterns has been kept finite[3] (i.e., each pattern has a nonzero overlap with all other patterns on average). Models with hierarchically correlated patterns have also been studied.[4] We consider a different problem in which only two of the patterns are correlated (have finite overlap) so that the average correlation is still irrelevant. Such finite correlations between a specific number of patterns have also been considered.[5–9] In particular, Derrida *et al.*[5] solved

[1] Saha Institute of Nuclear Physics, 1/AF Bidhannagar, Calcutta 700064, India.

this problem for the randomly diluted asymmetric version (where only two patterns are correlated) and found three phases: in one the patterns could be clearly distinguished, in the second, they are memorized with the same final overlap and could not be distinguished. In the third, none of them could be recalled. Similar phases where obtained for the Hopfield model[6] and also for other learning rules.[7] Fontanari and Koberle[6] assumed the solution of Amit *et al.*[2] to be valid even with this finite correlation (a numerical study with rather small system size was also done). The restriction of the correlation between two patterns only ensures that the critical memory-loading capacity remains unaffected at about 14%, i.e., $\alpha_c \approx 0.14$, where $\alpha = p/N$. Here, we are interested in investigating, in the context of the Hopfield model, the variation of the final correlation ρ_f between the corresponding learned patterns, with the initial correlation ρ_0 between a randomly chosen pair of patterns to be learned. In the earlier studies,[5-7] the dynamics of a single configuration having finite overlap with this pair was studied. We study the evolution of *two* configurations, corresponding to the two correlated patterns, and measure their overlap after they reach their respective fixed points. In comparison to the earlier results,[5,6] one expects a region in the α–ρ_0 space where the ratio $\rho_f/\rho_0 = 1$ and another with $\rho_f/\rho_0 \neq 1$. (We restrict ourselves to $\alpha < \alpha_c$, so that the patterns are always recalled.) In short, we try to understand the ability of the Hopfield model to store the correlation between two patterns.

We introduce a correlation ρ_0 between two patterns (denoted by 1 and 2) selected randomly. All other pairs of patterns are completely uncorrelated. Then the initial correlation between the patterns are given by

$$\rho^{\mu\nu} = \sum_i \xi_i^\mu \xi_i^\nu / N$$

where

$$\begin{aligned} \rho^{\mu\nu} &= \rho_0 \qquad \text{for } \mu = 1 \quad \text{and} \quad \nu = 2 \\ &= 0 \qquad \text{for other } \mu,\ \nu \text{ values} \end{aligned}$$

The system is then allowed to evolve from both the initial states $S_i^{(1)}(0) = \xi_i^1$ and $S_i^{(2)}(0) = \xi_i^2$. The dynamics followed is the usual Hopfield dynamics

$$S_i(t+1) = \operatorname{sgn}\left(\sum_j J_{ij} S_j(t)\right)$$

The correlation ρ_f between the states $S^{(1)}$ and $S^{(2)}$ after reaching their respective fixed points (denoted by t^*) is measured by

$$\rho_f = \sum_i S_i^{(1)}(t^*)\, S_i^{(2)}(t^*)/N$$

The internal field h_i for the ith neuron in the state $S_i = \xi_i^\mu$ when $\mu \neq 1$ or 2 is given by

$$h_i = \xi_i^\mu \left(1 + \sum_{\substack{i \neq j \\ \mu \neq \nu}} \xi_i^\mu \xi_j^\mu \xi_i^\nu \xi_j^\nu / N\right) \tag{1}$$

The noise factor

$$\sum_{\substack{i \neq j \\ \mu \neq \nu}} \xi_i^\mu \xi_j^\mu \xi_i^\nu \xi_j^\nu / N$$

is a random variable with variance $\delta = (p-1)/N$. For $\mu = 1$ or 2, the weight factor for ξ_i^μ becomes $1 \pm \rho_0$ (according to whether $\xi_i^1 = \pm \xi_i^2$) and the noise term is a sum over $p-2$ patterns in the expression of h. For example, for $\rho_0 = 1$, there is effectively a single pattern learned twice. The noise factor is a random variable with variance $\delta \cong (p-2)/N$ for two correlated patterns. Hence the value of the critical loading capacity is unchanged for large p and N.

Special attention is given to the limit $\rho_0 \to 0$ to see whether the Hopfield model generates any correlation in this limit.

In our simulation, we have taken $N = 500$ for measuring ρ_f/ρ_0. The patterns between which correlation is introduced are selected randomly and averaging is done over 25 configurations for each value of ρ_0 and α. For different values of α, the ratio ρ_f/ρ_0 is calculated. For $\rho_0 = 1$, this ratio is trivially unity.

The ratio ρ_f/ρ_0 is seen (see Fig. 1) to be greater than or equal to 1 for all values of α and ρ_0, implying that the Hopfield model can retain correlation between a particular pair of patterns either retaining its exact value or by enhancing it. For small values of α, $\rho_f/\rho_0 = 1$ when the initial correlation ρ_0 is also small. It then increases from 1, reaches a maximum value, and again decreases to 1. For higher values of α, the same behavior is observed: the value of ρ_f/ρ_0 is always greater than one, however small is ρ_0. The value of ρ_0 at which the ratio reaches the maximum is also seen to shift toward $\rho_0 = 0.5$ from higher values as α increases.

For studying the behavior of this system when $\rho_0 \to 0$, we consider a larger system with $N = 1000$. The critical value of $\rho_0 = \rho_c$ below which $\rho_f/\rho_0 = 1$ and above which it starts deviating from 1 is located for each α (with averaging over about 50 configurations for each α). As α is increased, ρ_c decreases continuously and reaches a value $\leqslant 0.07$ beyond $\alpha \cong 0.055$. It maybe recalled that in the Hopfield model, the retrieval states become absolute minima of the free energy function up to $\alpha = 0.05$ for completely random patterns. Our results indicate that this feature is probably retained even when two patterns are correlated up to a maximum correlation $\rho_c(\alpha)$,

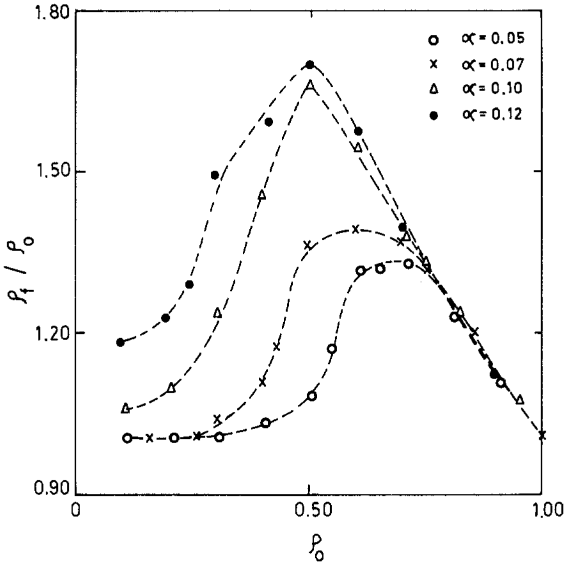

Fig. 1. The variation with α of the ratio ρ_f/ρ_0 of the final and initial correlations between any two patterns in the Hopfield model (simulation data for $N = 500$). The dashed lines are guides to the eye.

where $\rho_c(\alpha)$ is found to assume nonzero values for $0 \leqslant \alpha \leqslant 0.055$. [That $\rho_c(\alpha)$ does not really go to zero may be due to the finite system size: e.g., ρ_c is about 0.15 for $\alpha = 0.07$ for 500 neurons (Fig. 1) and this value is definitely greater than that for 1000 neurons.] One can guess that ρ_c goes to zero beyond $\alpha = 0.05$ in the thermodynamic limit from the behavior of $\rho_c(\alpha)$ in Fig. 2.

Hence, the Hopfield model generates some correlations, however small, for completely uncorrelated patterns above a certain α. This is also indicated in Fig. 1, in which the extrapolated value ρ_f/ρ_0 does not go to 1 for $\rho_0 \to 0$ for higher values of α. It may be mentioned in this context that in the layered feedforward neural network model, the value of correlation between the learned patterns abruptly jumps to unity from a negligible value at a critical $\alpha \approx 0.18$[10] (this correlation grows continuously from zero in the diluted model[11]; however, the correlation in this model is built up during the learning process). Obviously, no such behavior is expected in the Hopfield model for $\rho_0 \to 0$, as patterns are memorized with a maximum error of 3% for $\alpha < 0.14$. The correlations generated by the Hopfield model in this limit, therefore, are much lesser than unity.

In conclusion, we have observed that the Hopfield model indeed has the property of recognizing and enhancing the correlation between two

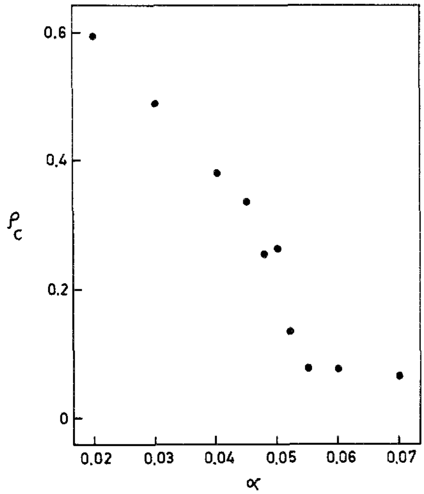

Fig. 2. The variation of the critical value of initial correlation ρ_c (up to which $\rho_f/\rho_0 = 1$) versus α (simulation data for $N = 1000$).

initially correlated patterns. With increasing α and ρ_0, it generates more and more correlation, while the ratio of the final correlation to the initial correlation reaches a maximum at $\rho_0 \cong 0.5$ for large values of α. When the patterns are highly correlated, this ratio becomes independent of α.

ACKNOWLEDGMENTS

I am grateful to B. K. Chakrabarti for his comments and suggestions. I am also grateful to M. Ghosh for bringing ref. 8 to my notice and to K. Y. M. Wong for sending a preprint of ref. 7 prior to publication.

REFERENCES

1. J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**:2554 (1982).
2. D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Ann. Phys.* **170**:30 (1987).
3. D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Phys. Rev. A* **35**:2293 (1987); F. Pazmandi and T. Geszti, *Europhys. Lett.* **13**:673 (1990).
4. H. Gutfreund, *Phys. Rev. A* **37**:570 (1987); M. V. Feigelman and B. Ioffe, *Int. J. Mod. Phys. B* **1**:51 (1987).
5. B. Derrida, E. Gardner, and A. Zippelius, *Europhys. Lett.* **4**:167 (1987).
6. J. F. Fontanari and R. Koberle, *J. Phys. A* **21**:2477 (1988).
7. A. Rau, K. Y. M. Wong, and D. Sherrington, preprint (1991).
8. T. J. P. Penna and P. M. C. de Oliviera, *J. Phys. A* **22**:L719 (1989).
9. F. A. Tamarit and E. M. F. Curado, *J. Stat. Phys.* **62**:473 (1991).
10. R. Meir and E. Domany, *Phys. Rev. A* **37**:2660 (1988).
11. P. Sen and B. K. Chakrabarti, *Phys. Rev. A* **40**:4700 (1989).

Communicated by J. L. Lebowitz