EUROPHYSICS LETTERS 15 July 1987
Europhys. Lett., 4 (2), pp. 167-173 (1987)

An Exactly Solvable Asymmetric Neural Network Model.

B. DERRIDA (*), E. GARDNER (**) and A. ZIPPELIUS (***)

(*) *Service de Physique Théorique, CEN-Saclay - 91191 Gif-sur-Yvette Cedex, France*
(**) *Dept. of Phys., University of Edinburgh - Edinburgh EH93JZ, U.K.*
(***) *IFF der KFA Julich, Postfach 1913 - 5170 Julich, F.R.G.*

(received 2 February 1987; accepted in final form 17 March 1987)

PACS. 05.40. – Fluctuation phenomena, random processes, and Brownian motion.
PACS. 05.50. – Lattice theory and statistics: Ising problems.

Abstract. – We consider a diluted and nonsymmetric version of the Little-Hopfield model which can be solved exactly. We obtain the analytic expression of the evolution of one configuration having a finite overlap on one stored pattern. We show that even when the system remembers, two different configurations which remain close to the same pattern never become identical. Lastly, we show that when two stored patterns are correlated, there exists a regime for which the system remembers these patterns without being able to distinguish them.

Spin glass models for associative memory have found increasing interest in the last few years. As first proposed by Little [1] and Hopfield [2], these models are based on an Ising Hamiltonian and hence can be treated by equilibrium statistical mechanics. A detailed discussion of the equilibrium properties of the Hopfield model is given in Amit *et al.* [3].

Two assumptions are crucial to allow for an exact solution of the equilibrium properties of the model: the synaptic connections are taken to be symmetric and each neuron is connected to an infinite number of other neurons. In biological networks the synapses are known to be asymmetric and on the average a neuron is connected only to a fraction $\rho \simeq 10^{-6}$ of all neurons. Hence it is important to study the effects of asymmetry and dilution.

Nonsymmetric models have been investigated by several groups [4-8]. One possible way to introduce asymmetry is to keep the «learning rules» of Hebb, but cut out some of the synaptic connections. As long as the fraction ρ remains finite, the memory states are not seriously degraded [4]. On the other hand the effect of extreme dilution, $\rho = O(1/N)$ is expected to be much more drastic. Random systems with long-range interactions, but finite coordination number have been discussed in the context of diluted spin glasses [8], graph optimization [9] and random networks of automata [10-12].

In this paper, we give an exact solution for the dynamics of a dilute nonsymmetric version of the Little-Hopfield model [1, 2]. The model consists of a system of N Ising spins $\sigma_1 = \pm 1$, whose interactions J_{ij} depend on p stored patterns. By definition of the model the J_{ij} are given by

$$J_{ij} = C_{ij} \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu , \tag{1}$$

where ξ_i^μ $(=\pm 1)$ is the value of site i in pattern μ and the C_{ij} are random independent parameters which represent the dilution and the asymmetry. For each pair i, j, a C_{ij} is chosen at random according to the distribution $\rho(C_{ij})$

$$\rho(C_{ij}) = \frac{C}{N}\delta(C_{ij}-1) + \left(1-\frac{C}{N}\right)\delta(C_{ij}). \tag{2}$$

Notice that the interactions J_{ij} are not symmetric because for each pair (i,j), C_{ij} and C_{ji} are independent random variables. (However, for the pairs i,j such that $C_{ij}=C_{ji}$, one has $J_{ij}=J_{ji}$.)

For this model the following two dynamics can be considered.

1) Parallel dynamics for which at time t, all spins are updated simultaneously in the following way [13]: on each site i the field $h_i(t)$ is computed

$$h_i(t) = \sum_j J_{ij}\sigma_j(t) \tag{3}$$

and then the spins are updated according to

$$\sigma_i(t+\Delta t) = \begin{cases} +1 \text{ with probability } (1+\exp[-2h_i(t)/T_0])^{-1}, \\ -1 \text{ with probability } (1+\exp[2h_i(t)/T_0])^{-1}. \end{cases} \tag{4}$$

The parameter T_0 which appears in (4) is by definition the temperature.

For parallel dynamics, the natural time scale is

$$\Delta t = 1. \tag{5}$$

2) Random sequential dynamics for which at time t, one chooses at random a site i among the N sites and one updates this site according to (3) and (4). Since at each time step, only one spin is updated, one should scale the time with the system size N

$$\Delta t = \frac{1}{N}. \tag{6}$$

We will later compare the time evolution of two different configurations. In that case, one can decide either that the random sequence of the updated spins is the same for both configurations or not. We will describe here only the case where it is the same.

In this letter, we obtain exact results for the dynamical properties of this model in the thermodynamic limit $(N \to \infty)$.

We will first consider the evolution of a configuration $\{\sigma_i(t)\}$ having a macroscopic overlap on one stored pattern and microscopic overlaps on the other $p-1$ random patterns. We will show that the evolution of $m(t)$ defined by

$$m(t) = \frac{1}{N}\sum_{i=1}^{N}\langle \xi_i^1 \sigma_i(t)\rangle \tag{7}$$

is given for parallel dynamics by

$$m(t+1) = f(m(t)) \tag{8a}$$

and for random sequential updating by

$$\frac{\mathrm{d}m(t)}{\mathrm{d}t} = f(m(t)) - m(t)\,, \tag{8b}$$

where $f(m)$ is given by

$$f(m) = \sum_{K=0}^{\infty} \frac{C^K e^{-c}}{K!} \sum_{n=0}^{K} \sum_{s=0}^{K(p-1)} \frac{(1+m)^{K-n}(1-m)^n}{2^{Kp}} \binom{K}{n} \binom{K(p-1)}{s} \operatorname{tgh}\left[\frac{Kp-2n-2s}{T_0}\right]. \tag{9}$$

$\left(\text{We use the notation } \binom{K}{n} = \frac{K!}{n!(K-n)!}.\right)$

We will then study the evolution of two configurations having a finite overlap on one pattern. This will show that when the system remembers, the attractor is more complicated than a single fixed configuration near the stored pattern.

Lastly we will consider the evolution of a configuration having finite overlaps on two stored correlated patterns. We will see that in general there are three regimes: 1) the system remembers the two patterns as distinct patterns, 2) the system remembers the 2 patterns, but cannot distinguish them, 3) the system does not remember.

Let us start by deriving formulae (8) and (9). The reason that the model is solvable is the same as the reason [12] which was already applied to the problem of random networks of automata [12, 14, 15] to show that the annealed and the quenched models are identical in the limit $N \to \infty$. The argument is the following. Consider a site i and let us call $j_1, j_2, \ldots, j_K$ the K sites j such that $J_{ij} \neq 0$. Assume that at time t

$$\langle \xi_{j_r}^1 \sigma_{j_r}(t) \rangle = m_{j_r}(t)\,, \tag{10}$$

where the average in (10) means both a thermal average at temperature T and over an ensemble of initial conditions at time $t = 0$ (for example all initial configurations having a fixed overlap $m(0)$ over the first pattern). Equation (10) means that

$$\sigma_{j_r}(t) = \begin{cases} \xi_{j_r}^1 & \text{with probability } \dfrac{1 + m_{j_r}(t)}{2}\,, \\[2ex] -\xi_{j_r}^1 & \text{with probability } \dfrac{1 - m_{j_r}(t)}{2}\,. \end{cases} \tag{11}$$

In principle the spins $\sigma_{j_1}(t), \sigma_{j_2}(t), \ldots, \sigma_{j_K}(t)$ could be correlated because they might have ancestors in common. However, one can show [12, 15] that as long as the constant C is finite or more precisely if

$$C \ll \log N \tag{12}$$

for almost all sites i, the spins $\sigma_{j_1}(t), \ldots, \sigma_{j_K}(t)$ are uncorrelated. The calculation of spin $\sigma_i(t)$ involves a tree of ancestors which connects site i to the initial conditions at time $t = 0$. The typical number of sites in this tree is less than C^t and as long as $C^t \ll N^{1/2}$, all the sites in this tree are different. So, if condition (12) is satisfied, the spins $\sigma_{j_1}(t), \ldots, \sigma_{j_K}(t)$ are uncorrelated for almost all sites i. The field $h_i(t)$ is given by

$$h_i(t) = \xi_i^1 \left(\sum_{r=1}^{K} \xi_{j_r}^1 \sigma_{j_r}(t) \right) + \sum_{\mu=2}^{p} \sum_{r=1}^{K} \xi_i^\mu \xi_{j_r}^\mu \sigma_{j_r}(t) \tag{13}$$

and the probability $P(\{\tau_{j_r}\}, s)$ that

$$h_i(t) = \xi_i^1 \left[\sum_{r=1}^{K} \tau_{j_r} + (p-1)K - 2s \right] \tag{14a}$$

is

$$P(\{\tau_{j_r}\}, s) = \frac{1}{2^{K(p-1)}} \binom{K(p-1)}{s} \prod_{r=1}^{K} \left(\frac{1 + \tau_{j_r} m_{j_r}(t)}{2} \right), \tag{14b}$$

where $\tau_{j_r} = \sigma_{j_r}(t)$ and $(p-1)K - 2s = \xi_i^{(1)} \sum_{\mu=2}^{p} \sum_{r=1}^{K} \xi_i^\mu \xi_{j_r}^\mu \sigma_{j_r}(t)$.
It is then easy to compute the average $m_i(t + \Delta t)$. One sees that $m_i(t + \Delta t)$ is a linear function of the $m_{j_r}(t)$. Therefore, when one averages over all the shapes of the tree of ancestors (in particular over K), all the $m_{j_r}(t)$ have the same average $m(t)$ and one finds

$$m_i(t + \Delta t) = f(m(t)), \tag{15}$$

where f is given by (9).

For parallel dynamics, this proves formula (8a). For random sequential updating, one can easily establish (8b) by considering that

$$m(t + \Delta t) = \frac{N-1}{N} m(t) + \frac{1}{N} m_i(t + \Delta t) = m(t) + \Delta t[f(m(t)) - m(t)]. \tag{16}$$

The time evolution given by (8) and (9) is valid for random uncorrelated patterns $\{\xi_i^\mu\}$ and for an initial configuration having a finite projection on only one pattern. The system remembers if there is a nonzero attractive fixed point m^* of the map (8a) or of the flow (8b). One can determine the temperature $T^*(C, p)$ below which the system remembers stored patterns. We did not find a closed expression of T^* for finite p and C, but we think that there should be no problem to determine T^* numerically.

In the limit C and $p \to \infty$, keeping in mind that $N \to \infty$ first (see condition 12), if one defines α and a reduced temperature T by

$$\alpha = \frac{p-1}{C} \quad \text{and} \quad T = T_0/C, \tag{17}$$

the expression (9) of $f(m)$ becomes

$$f(m) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{+\infty} \mathrm{d}y \exp[-y^2] \operatorname{tgh}\left(\frac{1}{T} (m - y\sqrt{2\alpha}) \right). \tag{18}$$

The critical temperature T^* is given by $f'(0) = 1$. The transition is second order because m^* vanishes as $T \to T^*$. At 0 temperature, the critical value α_c of α is

$$\alpha_c = 2/\pi = 0.6366\ldots \tag{19}$$

and the transition is second order $m^* \sim (\alpha_c - \alpha)^{1/2}$. One can also see from (18) that for $T = 0$ and $\alpha \to 0$

$$m^* \simeq 1 - \sqrt{\frac{2\alpha}{\pi}} \exp\left[-\frac{1}{2\alpha} \right]. \tag{20}$$

One should notice that the dilute asymmetric model studied here has a second-order transition at α_c instead of the first-order transition predicted for the symmetric nondiluted case [3]. Also the value α_c is larger than 0.14 (even if one takes into account that the nonsymmetry forces one to store two bonds instead of one).

In order to understand the nature of the attractor near a stored pattern, we are now going to study the evolution of two configurations $\sigma_i(t)$ and $\tilde{\sigma}_i(t)$ having a finite projection on one pattern (say pattern 1) and zero projections on the other $p-1$ patterns. For the same reason as in the case of one configuration, the spins $\sigma_{j_1}(t), \ldots, \sigma_{j_K}(t)$ are uncorrelated and the spins $\tilde{\sigma}_{j_1}(t), \ldots, \tilde{\sigma}_{j_K}(t)$ are uncorrelated. However, $\sigma_{j_r}(t)$ and $\tilde{\sigma}_{j_r}(t)$ are correlated. One can then consider 3 quantities: $m(t)$, $\tilde{m}(t)$ defined by (7) and $q(t)$ defined by

$$q(t) = \frac{1}{N} \sum_{i=1}^{N} q_i(t) , \tag{21}$$

where

$$q_1(t) = \langle \sigma_i(t) \tilde{\sigma}_i(t) \rangle \; . \tag{22}$$

The evolution of $m(t)$ and $\tilde{m}(t)$ are still given by (8) and (9). The only new information is the time evolution of $q(t)$. If one defines $h_i(t)$ and $\tilde{h}_i(t)$ by

$$\left\{ \begin{aligned} h_i(t) &= \xi_i^1 \sum_{r=1}^{K} \xi_{j_r}^1 \sigma_{j_r}(t) + \sum_{r=1}^{K} \sum_{\mu=2}^{p} \xi_i^\mu \xi_{j_r}^\mu \sigma_{j_r}(t) , \\ \tilde{h}_i(t) &= \xi_i^1 \sum_{r=1}^{K} \xi_{j_r}^1 \tilde{\sigma}_{j_r}(t) + \sum_{r=1}^{K} \sum_{\mu=2}^{p} \xi_i^\mu \xi_{j_r}^\mu \tilde{\sigma}_{j_r}(t) , \end{aligned} \right. \tag{23}$$

these two fields are correlated. One can compute the probability $P(n_1, n_2, n_3, n_4, s_1, s_2)$ that

$$\left\{ \begin{aligned} \xi_i^1 h_i(t) &= n_1 + n_2 - n_3 - n_4 + (p-1)(n_1 + n_2 + n_3 + n_4) - 2s_1 - 2s_2 , \\ \xi_i^1 \tilde{h}_i(t) &= n_1 - n_2 + n_3 - n_4 + (p-1)(n_1 - n_2 - n_3 + n_4) - 2s_1 + 2s_2 , \end{aligned} \right. \tag{24}$$

is given by

$$P(n_1, n_2, n_3, n_4, s_1, s_2) = \frac{1}{2^{K(p-1)}} \frac{K!}{n_1!\, n_2!\, n_3!\, n_4!} \binom{(n_1+n_4)(p-1)}{s_1} \binom{(n_2+n_3)(p-1)}{s_2} \cdot$$

$$\cdot\, \delta_{n_1+n_2+n_3+n_4, K} \left(\frac{1+m+\tilde{m}+q}{4}\right)^{n_1} \left(\frac{1+m-\tilde{m}-q}{4}\right)^{n_2} \left(\frac{1-m+\tilde{m}-q}{4}\right)^{n_3} \left(\frac{1-m-\tilde{m}+q}{4}\right)^{n_4} , \tag{25}$$

where m, $\tilde{m}$ and q are the values at time t. These expressions are rather easy to understand. They express the fact that among the K ancestors j of site i, n_1 are such that $\xi_j^1 = \sigma_j = \tilde{\sigma}_j$, n_2 such that $\xi_j^1 = \sigma_j = -\tilde{\sigma}_j$, n_3 such that $\xi_j^1 = -\sigma_j = \tilde{\sigma}_j$, n_4 such that $\xi_j^1 = -\sigma_j = -\tilde{\sigma}_j$. One can then compute $q_i(t + \Delta t)$ by averaging over n_1, n_2, n_3, n_4, s_1 and s_2:

$$q_1(t + \Delta t) = \overline{\operatorname{tgh}\left(\frac{h_i(t)}{T}\right) \operatorname{tgh}\left(\frac{\tilde{h}_i(t)}{T}\right)} \tag{26}$$

and from that obtain the time evolution of $q(t)$ in the parallel and the sequential dynamics. Let us write here the evolution of $q(t)$ in the limit p and $C \to \infty$ (see eq. (17))

$$q(t+\Delta t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \mathrm{d}y \, \mathrm{d}z \frac{\exp[-y^2 - z^2]}{\pi} \operatorname{tgh}\left(\frac{m - \sqrt{\alpha(1+q)}\, y - \sqrt{\alpha(1-q)}\, z}{T}\right) \cdot$$

$$\cdot \operatorname{tgh}\left(\frac{\tilde{m} - \sqrt{\alpha(1+q)}\, y + \sqrt{\alpha(1-q)}\, z}{T}\right), \tag{27}$$

where on the right-hand side of (27), m, $\tilde{m}$ and q are the values at time t. In the limit $t \to \infty$, $m(t)$ and $\tilde{m}(t) \to m^*$ given by (18), whereas $q(t)$ converges to the fixed point q^* of (27). For α small, one finds

$$q^* \simeq 1 - \frac{8}{\pi^2} \exp[-1/\alpha]. \tag{28}$$

It is interesting to compare (20) and (28) for α small $1 - q^* \ll 1 - m^*$. Since q^* is not 1, this means that near a stored pattern there is not a single attractive fixed configuration but either a more complicated attractor or a cloud of attractors as in the symmetric case [3, 16]. So the picture of a single valley near a stored pattern is certainly too simplified.

A similar result has been obtained recently by Feigelman and Ioffe [8] for a nondiluted and nonsymmetric version of the Hopfield model that they can solve in the limit $\alpha \to 0$.

Lastly let us consider briefly the case of a configuration $\{\sigma_i(t)\}$ having projections $m_1(t)$ and $m_2(t)$ on two patterns $\{\xi_i^1\}$ and $\{\xi_i^2\}$ (see eq. (7)) and a zero projection of the $p-2$ other random patterns. We shall consider the case for which the two patterns 1 and 2 have a finite overlap Q

$$Q = \frac{1}{N} \sum_{i=1}^{N} \xi_i^1 \xi_i^2. \tag{29}$$

The other $p-2$ patterns being random, they have, with probability 1, zero projection on patterns 1 and 2. One can repeat calculations similar to those given above to obtain the evolution of $m_1(t)$ and $m_2(t)$. Let us just give here the result in the limit p and $C \to \infty$ (see eq. (17)):

$$m_1(t+\Delta t) + \varepsilon m_2(t+\Delta t) = (1 + \varepsilon Q) \int_{-\infty}^{+\infty} \frac{\mathrm{d}y \exp[-y^2]}{\sqrt{\pi}} \operatorname{tgh}\left(\frac{(m_1 + \varepsilon m_2) - \sqrt{2\alpha}\, y}{T}\right) \tag{30}$$

for $\varepsilon = \pm 1$.

In the limit $T \to 0$ and $t \to \infty$, one finds two critical values of α:

$$\alpha_c^{(1)} = \frac{2}{\pi}(1+Q)^2 \qquad \text{and} \qquad \alpha_c^{(2)} = \frac{2}{\pi}(1-Q)^2. \tag{31}$$

For $\alpha > \alpha_c^{(1)}$ the system does not remember anything ($m_1^* = m_2^* = 0$). For $\alpha_c^{(2)} < \alpha < \alpha_c^{(1)}$, one finds an attractive fixed point $m_1^* = m_2^* \neq 0$. The system remembers patterns 1 and 2, but cannot distinguish them. For $\alpha < \alpha_c^{(2)}$, one finds two attractive fixed points with $0 \neq m_1^* \neq m_2^* \neq 0$ which means that the system remembers the two patterns and can distinguish them.

In this letter, we have seen that the dynamics of a diluted and asymmetric version of the

Little-Hopfield model can be solved exactly. As for other models we find a critical α_c above which the system does not remember. For $\alpha < \alpha_c$, two initial configurations close to a stored pattern, remain close to the pattern, but do not become identical. This shows that the attractor near a stored pattern has a more complex structure than a single attractive configuration. Lastly, we have seen that when some of the stored patterns are correlated, there exist regimes for which the system remembers the patterns, but cannot distinguish them.

We think that our approach could be generalized to other situations like for example the case of time-dependent stored patterns. However, the case of symmetric bonds J_{ij} is probably impossible to treat by this approach because after two times steps the same site appears several times in the tree of ancestors.

REFERENCES

[1] LITTLE W. A., *Math. Biosci.*, **19** (1974) 101.
[2] HOPFIELD J. J., *Proc. Nat. Acad. Sci. USA*, **79** (1982) 2554; **81** (1984) 3088.
[3] AMIT D. J., GUTFREUND H. and SOMPOLINSKY H., *Phys. Rev. A*, **32** (1985) 1007; *Phys. Rev. Lett.*, **55** (1985) 1530; *Ann. Phys.* (*N.Y.*), **173** (1987) 30.
[4] HERTZ J. A., GRINSTEIN G. and SOLLA S. A., preprint (1986).
[5] BAUSCH R., JANSSEN H. K., KREE R. and ZIPPELIUS A., *J. Phys. C*, **19** (1986) L-779.
[6] PARISI G., *J. Phys. A*, **19** (1986) L-675.
[7] SOMPOLINSKY H. and KANTER I., *Phys. Rev. Lett.*, **57** (1986) 2861.
[8] FEIGELMAN M. V. and IOFFE L. B., preprint (1986).
[9] VIANA L. and BRAY A. J., *J. Phys. C*, **18** (1985) 3037; KANTER I. and SOMPOLINSKY H., *Phys. Rev. Lett.*, 58 (1987) 164; MOTTISHAW P. and DE DOMINICIS C., preprint (1986); MÉZARD M. and PARISI G., preprint (1986).
[10] FU Y. and ANDERSON P. W., *J. Phys. A*, **19** (1986) 1605.
[11] KAUFFMAN S. A., *J. Theor. Biol.*, 22 (1969) 437 and *Physica D*, **10** (1984) 145.
[12] DERRIDA B. and WEISBUCH G., *J. Phys.* (*Paris*), **47** (1986) 1297.
[13] GARDNER E., DERRIDA B. and MOTTISHAW P., to appear in *J. Phys.* (*Paris*) (1987).
[14] DERRIDA B. and POMEAU Y., *Europhys. Lett.*, **1** (1986) 45.
[15] HILHORST H. J. and NIJMIJER M., *J. Phys.* (*Paris*), **48** (1987) 185.
[16] GARDNER E., *J. Phys. A*, **19** (1986) L-1047.