Interaction asymmetry and hierarchy of patterns in associative-memory models

L. B. Ioffe and M. V. Feĭgel'man

L. D. Landau Institute of Theoretical Physics, Academy of Sciences of the USSR

(Submitted 15 May 1986)

Pis'ma Zh. Eksp. Teor. Fiz. **44**, No. 3, 148–150 (10 August 1986)

Some models of an associative memory are proposed and studied. These models permit an asymmetry of the connections between neurons and also a storage of correlated patterns forming hierarchical structures similar to that of a spin glass.

1. Many papers which have recently appeared in the physics literature report use of the methods of statistical physics to study the physical properties of a slightly unusual object: a neuron network. Since such a network has $\sim 10^{10}$ elements, a statistical description would seem unavoidable. Furthermore, the properties of neuron net-

 0021-3640/86/150189-04$01.00

works are analogous in many ways to those of spin glasses, so that the methods of the theory of spin glasses can be used to study neuron networks. In Hopfield's model,[1,2] the most popular, the neurons are replaced by elements (σ_i) which take on basically two values (± 1; these are "soft" Ising spins), while the patterns which are stored in this model are the set $\{\xi_i^{(p)}\}(\xi_i^{(p)} = \pm 1\ 1 \leqslant i \leqslant N, 1 \leqslant p \leqslant k)$, where i specifies the neuron, and p the pattern. It is assumed that the equation of motion of the variables σ_i is the same as that of the purely relaxational Langevin dynamics of spins with an energy $H_t\{\sigma_i\}$:

$$\frac{\partial \sigma_i}{\partial t} = -\frac{\partial H_t}{\partial \sigma_i} + f_i(t), \quad \langle f_i(t) f_j(t') \rangle = 2T\delta_{ij}\delta(t-t'), \quad H_t = H_0 + H$$
$$H_0 = \sum_i \lambda(\sigma_i^2 - 1)^2, \qquad H = -\frac{1}{2}\sum_{i,j} J_{ij}^{(0)}\sigma_i\sigma_j, \qquad \lambda \gg 1. \tag{1}$$

The matrix $J_{ij}^{(0)}$ must be chosen in such a way that the stationary solutions of Eq. (1) be the same as the set of patterns to be stored, $\{\xi_i^{(p)}\}$. If this condition is satisfied, the solution $\sigma_j(t)$ with initial conditions $\sigma_i(0)$ sufficiently close to one of the patterns $\{\xi_i^{(p)}\}$ will relax toward this pattern, since a "recalling" is occurring. In Hopfield's model, $J_{ij}^{(0)}$ is chosen in the extremely simple form $J_{ij}^{(0)} = (1/N)\ \Sigma_p\ \xi_i^{(p)}\xi_j^{(p)}$. This model has now been studied in detail by the methods of statistical physics.[3–5]

Hopfield's model has two serious deficiencies: First, it works well (i.e., the number of patterns that can be stored is large) only when uncorrelated patterns $\xi_i^{(p)}$ are being stored. Second, it embodies the symmetry of the matrix J_{ij}, an assumption which is totally illogical from the standpoint of real neuron networks. In the present letter we will discuss how these deficiencies might be remedied (the detailed calculations will be published separately). We begin with the asymmetry of the matrix J_{ij}.

2. Let us examine the simplest generalization of Eq. (1) with an asymmetric matrix J_{ij}:

$$\frac{\partial \sigma_i}{\partial t} = -\frac{\partial H_0}{\partial \sigma_i} + \sum_i J_{ij}\sigma_j + f_i(t), \qquad \langle f_i(t) f_j(t') \rangle = 2T\delta_{ij}\delta(t-t'). \tag{2}$$

We assume that not all of the couplings (i,j) can be implemented; i.e., we choose J_{ij} in the form $J_{ij} = (1+\epsilon_{ij})J_{ij}^{(0)}$, where ϵ_{ij} takes on the values ± 1 with equal probabilities and for all couplings (i,j), while we have $J_{ij}^{(0)} = (1/N)\Sigma_{p=1}^{k}\ \xi_i^{(p)}\xi_j^{(p)}$. In order to study the solution of Eq. (2), we use the method of a dynamic generating functional (Ref. 6, for example). After taking an average over ϵ_{ij}, we find

$$\langle \sigma_i(t)\sigma_i(t')\rangle = \int \sigma_i(t)\sigma_i(t')\exp(iS\{\sigma_i(t), \psi_i(t)\})D\sigma_i(t)D\psi_i(t),$$
$$S = \int dt\{\sum_i \psi_i(t)\left(\dot{\sigma}_i + \frac{\partial H_0}{\partial t} + \sum_j J_{ij}^{(0)}\sigma_j\right) + \sum_i T\psi_i^2(t)\} + \tilde{S}, \tag{3}$$
$$\tilde{S} = \frac{\alpha q}{2}\sum_i\left(\int \psi_i(t)dt\right)^2 + \frac{\alpha}{2}\sum_i \iint dt dt'\,\psi_i(t)\psi_i(t')D(t-t').$$

Here $\alpha = k/N$ is the relative number of patterns; q is the Edwards-Anderson parameter; $D(t)$ is the spin correlation function $[\overline{\langle\sigma_i(0)\sigma_i(t)\rangle} = q + D(t)(\lim D(t) = 0]$; and the term $\widetilde{S}$ appears in S as a result of the averaging over ϵ_{ij}. The last term in $\widetilde{S}$ corresponds to the presence of an intrinsic noise $\eta_i(t)$ with the correlation function $\langle\eta_i(t)\eta_j(t')\rangle = \alpha\delta_{ij}D(t-t')$, which means a violation of the fluctuation-dissipation theorem (Ref. 7). Equations (3) are actually self-consistency equations for $D(t)$. We will examine them only in the case $\alpha \ll 1$. In this case, for a state of the system close to one of the patterns $\xi_i^{(p)}$ (say $\xi_i^{(1)}$) it is convenient to make the change of variables $\sigma_i \to \sigma_i \xi_i^{(1)}, \psi_i \to \psi_i \xi_i^{(1)}, J_{ij}^{(0)} \to J_{ij}^{(0)} \xi_i^{(1)} \xi_i^{(1)}$, which does not change the form of (3). We then break up $J_{ij}^{(0)}$ into two parts:

$$J_{ij}^{(0)} = 1/N + \widetilde{J}_{ij}, \widetilde{J}_{ij} = (1/N) \sum_{p=1}^{k} \xi_i^{(1)} \xi_i^{(p)} \xi_j^{(1)} \xi_j^{(p)}.$$

The condition $\alpha \ll 1$ allows us to treat the $\widetilde{J}_{ij}$ as random and independent quantities. We therefore average Eq. (3) over $\widetilde{J}_{ij}$ and proceed as we would in a study of the dynamics of a spin glass.[8] As a result, we find the equations of motion of a spin in a static external field h and a variable $\eta(t)$. At a temperature $T \gg \exp(-1/4\alpha)$ we can transform to Glauber equations for $\varphi = p_+ - p_-$, where $p_\pm$ is the probability for finding a spin in the state ± 1 in the given field $h + \eta(t)$:

$$\dot{\varphi} = \tanh[(h + \eta(t))/T] - \varphi. \tag{4}$$

In this temperature range, the static fields h have a Gaussian distribution with a mean value $\bar{h} = m, m \simeq 1, (h - \bar{h})^2 = \alpha r, r \simeq 2$. Expressing the correlation function $\langle\sigma(0)\sigma(t)\rangle$ in terms of the solutions of Eq. (4), and taking an average over the noise $\eta(t)$, we find an expression for $D(0)$, which determines the intensity of the nonthermal noise:

$$1 - D(0) = \int \frac{dh}{\sqrt{2\pi\alpha r}} \exp\left(-\frac{(h-m)^2}{2\alpha r}\right) \left[\int \tanh\left(\frac{h+\eta}{T}\right) \exp\left(-\frac{\eta^2}{2\alpha D(0)}\right) \frac{d\eta}{\sqrt{2\pi\alpha D(0)}}\right]^2 \tag{5}$$

It follows that for $\exp(-1/4\alpha) \ll T \ll \alpha$ we have $D(0) \simeq \exp(-1/4\alpha)$, in this temperature range, the only consequence of the asymmetry of the matrix J_{ij} is an additional noise, small in comparison with the thermal noise, which does not affect the stationary states of the system and thus does not degrade the operation of the model.

3. Working in the symmetric model, let us examine the writing of correlated patterns $\xi_i^{(p)} = \xi_i(1 - 2\beta_i^{(p)})$, where $\beta_i^{(p)}$ take on the values 0 and 1 $[\text{Prob}(\beta = 1) = c \ll 1]$. When the customary writing algorithm[1] is used, the various patterns $\xi_i^{(p)}$ are indistinguishable, and the stationary states of the system correspond to the "base" pattern ξ_i. In order to resolve the "fine structure" of the patterns, we need to take the interaction matrix in the form

$$J_{ij}^{(0)} = \frac{1}{N} \xi_i \xi_j \left[1 + \frac{1}{\epsilon} \sum_{p=1}^{k} (\beta_i^{(p)} - c)(\beta_j^{(p)} - c)\right], \tag{6}$$

where $\epsilon < 2c$. We have studied the properties of a memory system of this type for finite α by the methods of Ref. 4. In contrast with Ref. 4, we have two types of order parameters here:

$$m = \frac{1}{N}\sum_{i=1}^{N} \xi_i \langle \sigma_i \rangle ; \qquad u^{(p)} = \frac{1}{N\epsilon}\sum_{i=1}^{N} (\beta_i^{(p)} - c)\langle \sigma_i \rangle \xi_i . \qquad (7)$$

The pattern $\xi_i^{(p)}$ corresponds to a stationary state with $m \neq 0, u^{(p)} \neq 0$. If we have $m \neq 0$ but $u^{(p)} = 0$ for all p, a base pattern ξ_i arises. With $\epsilon = c$, the states of the two types have the same free energy and are realized at a "temperature" $T < T_c(\alpha)$, where the functional dependence $T_c(\alpha)$ is the same as that found in Ref. 4. Furthermore the maximum relative number of reproducible (at $T=0$) patterns is $\alpha_c \simeq 0.14$. For $c < \epsilon < 2c$ there is an intermediate region $\alpha_2 < \alpha < \alpha_1$ in which only the base pattern ξ_i is stable; we have $\alpha_2 < \alpha_c < \alpha_1$. In the limit $\epsilon \to 2c$ we have $\alpha_2 \to 0$. For $\alpha < \alpha_2$ and $T < T_2(\alpha)$ the various patterns $\xi_i^{(p)}$ are resolvable. Interestingly, there is an intermediate "temperature" range, $T_2(\alpha) < T < T_1(\alpha)$, in which only the base pattern is resolvable. The functional dependences $\alpha_{1,2}(\epsilon)$ are found through a numerical solution of the mean-field equations and will be published separately.

4. The algorithm for writing information in (6) can be generalized to the case of a memory with a hierarchical structure with k_1 uncorrelated base patterns $\xi_i^{(\lambda)}$, each having k_2 "satellites" $\xi_i^{(\lambda,p)}$:

$$J_{ij}^{(0)} = \frac{1}{N}\sum_{\lambda=1}^{k_1} \xi_i^{(\lambda)}\xi_j^{(\lambda)}\left[1 + \frac{1}{\epsilon}\sum_{p=1}^{k_2} (\beta_i^{(\lambda,p)} - c)(\beta_j^{(\lambda,p)} - c)\right]. \qquad (8)$$

The maximum total number of stably reproducible patterns is $k_1 k_2 \sim N$. A hierarchical organization of the memory allows a rapid recognition of the base patterns (classes) $\xi_i^{(\lambda)}$ at a higher noise level T, with a subsequent resolution of the individual patterns $\xi_i^{(\lambda,p)}$ with decreasing T. It would be possible to generalize (8) to the case of a hierarchy with $n > 2$ levels, but the total number of patterns, $k = k_1 \cdot k_2 \ldots \cdot k_n$, remains on the order of N. A construction of the same type as in (8) was proposed in Ref. 9 for the particular case $\epsilon = c$.

[1]J. J. Hopfield, Proc. Nat. Acad. Sci. USA **79**, 2554 (1982).
[2]J. J. Hopfield, Proc. Nat. Acad. Sci. USA **81**, 3088 (1984).
[3]D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985).
[4]D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. Lett. **55**, 1530 (1985).
[5]M. V. Feigel'man and L. B. Ioffe, Europhys. Lett. **1**, 197 (1986).
[6]H. Sompolinsky and A. Zippelius, Phys. Rev. B **25**, 6860 (1982).
[7]J. R. L. de Almeida and D. J. Thouless, J. Phys. **A11**, 983 (1978).
[8]H. Sompolinsky, Phys. Rev. Lett. **47**, 935 (1981).
[9]N. Parga and M. A. Virasoro, Preprint Trieste IC 1985-186.

Translated by Dave Parsons