Image evolution in Hopfield networks

A. C. C. Coolen
Department of Medical and Physiological Physics, Princetonplein 5, 3584 CC Utrecht, The Netherlands

Th. W. Ruijgrok
Institute for Theoretical Physics, Princetonplein 5, P.O. Box 80 006, 3508 TA Utrecht, The Netherlands
(Received 18 February 1988)

We consider neural networks of the Hopfield type with couplings J_{ij} which need not be symmetric. From the master equation for microscopic states we derive an evolution equation for the probability density of the macroscopic parameters q_μ, which measure the overlap of the instantaneous microscopic state (or image) with one of the built-in patterns. No restrictions are imposed on the choice of the patterns. For three different temperatures this equation is used to illustrate retrieval in the standard Hopfield network and limit-cycle behavior in nonsymmetric models.

I. INTRODUCTION

In this paper we consider a Hopfield model for neural networks.[1,2] The case of symmetric connections J_{ij} between two neurons i and j has been extensively studied by Amit, Gutfreund, and Sompolinsky.[2] The study of the temporal development of such networks seems to be restricted to diluted networks[3-5] (where many bonds are cut), to feed-forward networks,[6] or to fully connected networks at zero temperature.[7]

It is the purpose of the present paper to show that none of these restrictions is necessary. In particular we will derive a flow equation for the macroscopic parameters q_μ, which measure the overlap of the instantaneous microscopic state with one of the built-in patterns. The starting point for this derivation is the master equation for the probability at any time to find the neural network in a certain microscopic state, defined by specifying for each neuron whether it is on or off. We want to stress that the transition probability per unit time for a neuronal state flip of course depends on the weighted sum of arriving potentials on that specific neuron, but is calculated without first introducing a Hamiltonian, which, by its nature, would have to be symmetric in the neural connections.

Another point in which our work is different from what is known to us in the literature is the fact that the patterns to be retrieved can be defined as we wish, and need not be drawn at random from a uniform distribution. Only at the very end, and only by way of illustration, did we make this latter choice when the flow diagrams of the figure were calculated.

II. DERIVATION OF THE FLOW EQUATIONS

As in all neural networks of the Hopfield type,[1,2] we model the neurons as Ising spins s_j $(j=1,\ldots,N)$. If neuron j fires we set $s_j=+1$; if it is at rest $s_j=-1$. The local fields h_i are defined by $h_i=\sum_j J_{ij}(s_j+1)$, where J_{ij} is the strength of the synaptic connection from neuron j to neuron i. It is excitatory if J_{ij} is positive and inhibitory if J_{ij} is negative. Neuron j contributes to the local field at the position of neuron i only if the former is firing. For the interactions we take

$$J_{ij}=\frac{1}{N}\sum_{\mu,\nu}\xi_i^\mu A_{\mu\nu}\xi_j^\nu \quad (\mu,\nu=1,\ldots,p)\ , \tag{1}$$

where each ξ_k^α is either $+1$ of -1. $A_{\mu\nu}$ are the elements of a $p\times p$ matrix, which need not be symmetric. Therefore, nonsymmetric couplings J_{ij} will also be considered. However, only with a symmetric A can the equilibrium properties of the system be described in terms of a Hamiltonian. The standard couplings[1,2] are obtained by taking the identity matrix for A. The p vectors $(\xi_1^\mu,\ldots,\xi_N^\mu)$ represent the patterns that are anchored in the network.

Our problem is now to give a rule according to which an arbitrary initial state of the spins $(s_1,\ldots,s_N)$ will change in time and to ascertain to what extent, if at all, one of the built-in patterns will be approached.

In order to do so we assume that for each spin there exists a probability per unit time $w(s_j\to -s_j)$ to flip and that this w only depends on the value of this spin and on the local field. We take the standard form for w and write for the probability that spin j will flip in the next unit of time

$$w(s_j\to -s_j)=\tfrac{1}{2}[1-\tanh(\beta s_j h_j)]\ , \tag{2}$$

where $T=1/\beta$ is a measure for the rate of spontaneous spin flips. With the choice for w made, we now write down the master equation for $p(\mathbf{s},t)$, which is the probability to find the system at time t in the state $\mathbf{s}=(s_1,\ldots,s_N)$:

$$\frac{\partial p(\mathbf{s},t)}{\partial t}=\sum_{j=1}^{N} w(-s_j\to s_j)p(F_j\mathbf{s},t) - p(\mathbf{s},t)\sum_{j=1}^{N} w(s_j\to -s_j)\ , \tag{3}$$

where we have made use of the spin flip operator F_j, defined by

$$F_j\phi(s_1,\ldots,s_j,\ldots,s_N)=\phi(s_1,\ldots,-s_j,\ldots,s_N)\ .$$

Since N is very large, the solution of this equation for an arbitrary initial state is impossible. This, however, is hardly a disadvantage because, like in statistical mechanics, we are not interested in the microscopic details of a state, but rather in the question whether the development of certain macroscopic features can be calculated. For these features we take the overlap with the built-in patterns, defined by[2]

$$q^\mu(\mathbf{s})=\frac{1}{N}\sum_{j=1}^{N}\xi_j^\mu s_j \quad \text{or} \quad \mathbf{q}(\mathbf{s})=\frac{1}{N}\sum_{j=1}^{N}\boldsymbol{\xi}_j s_j\ , \tag{4}$$

where $\mathbf{q}(\mathbf{s})$ and $\boldsymbol{\xi}_j$ are vectors with p components. For the ideal retrieval of pattern μ_0 one should have $q^\mu=\delta_{\mu\mu_0}$. Our aim is to derive from (3) an equation for the probability of finding the system at time t in a state with macroscopic correlation parameters $\mathbf{q}=(q_1,\ldots,q_p)$. This probability is defined as

$$P(\mathbf{q},t)=\sum_{\mathbf{s}}p(\mathbf{s},t)\delta(\mathbf{q}-\mathbf{q}(\mathbf{s}))\ . \tag{5}$$

Using the master equation (3) it is straightforward to show that

$$\frac{\partial P(\mathbf{q},t)}{\partial t}=\tfrac{1}{2}\sum_{\mathbf{s}}\int d\mathbf{q}'p(\mathbf{s},t)\delta(\mathbf{q}'-\mathbf{q}(\mathbf{s}))\int d\mathbf{x}\sum_{j=1}^{N}\delta(\mathbf{x}-\boldsymbol{\xi}_j s_j)[1-\tanh(\beta\mathbf{x}A\mathbf{q}')]\left[\delta\left[\mathbf{q}-\mathbf{q}'+\frac{2\mathbf{x}}{N}\right]-\delta(\mathbf{q}-\mathbf{q}')\right]\ . \tag{6}$$

The summation over the spin index j will now be performed, using the partition introduced by van Hemmen *et al.*[8] There the set of all indices is divided into subsets I_η, which depend on the built-in patterns ξ_i^μ in the following way:

$$\{i\le N\}=\bigcup_{\boldsymbol{\eta}}I_{\boldsymbol{\eta}} \quad \text{where } I_{\boldsymbol{\eta}}=\{i\le N\mid\boldsymbol{\eta}=\boldsymbol{\xi}_i\}\ . \tag{7}$$

The number of different p-dimensional vectors $\boldsymbol{\eta}$ is 2^p, which is much smaller than the number N of vectors $\boldsymbol{\xi}_i$. Therefore, the number of indices in each of these sets, to be denoted by $|I_\eta|$, will almost always be the same for all $\boldsymbol{\eta}$, if all ξ_k^α are chosen randomly:

$$|I_\eta|=2^{-p}N+O(N^{1/2})\ . \tag{8}$$

For the time being, however, we will not make use of this fact and consider $|I_\eta|$ as numbers which can be calculated for each particular realization of the built-in patterns ξ_j^μ.

It is seen immediately that the overlap functions $\mathbf{q}(\mathbf{s})$ defined in Eq. (4) can be written as

$$\mathbf{q}(\mathbf{s})=\frac{1}{N}\sum_{\boldsymbol{\eta}}|I_\eta|\,m_\eta(\mathbf{s})\boldsymbol{\eta} \quad \text{with } m_\eta(\mathbf{s})=\frac{1}{|I_\eta|}\sum_{j\in I_\eta}s_j\ . \tag{9}$$

With this relation it is straightforward, although rather tedious, to prove that Eq. (6) can be written in the form

$$\frac{\partial P(\mathbf{q},t)}{\partial t}=-\sum_{\mu}\frac{\partial}{\partial q_\mu}\{[-q_\mu+\langle\eta^\mu\tanh(\beta\boldsymbol{\eta}A\mathbf{q})\rangle_\eta]\times P(\mathbf{q},t)\}\ , \tag{10}$$

where the following abbreviation has been introduced:

$$\langle f(\boldsymbol{\eta})\rangle_\eta\equiv\frac{1}{N}\sum_{\boldsymbol{\eta}}|I_\eta|f(\boldsymbol{\eta})\ . \tag{11}$$

In the process of deriving Eq. (10) terms of order $1/N$ have been neglected, but terms of order $N^{-1/2}$ are still present because of Eq. (8).

Equation (10) is the continuity equation for the density of a flow in the p-dimensional $\mathbf{q}$ space. The equation for the flow itself becomes

$$\frac{d\mathbf{q}}{dt}=-\mathbf{q}+\langle\boldsymbol{\eta}\tanh(\beta\boldsymbol{\eta}A\mathbf{q})\rangle_\eta \tag{12}$$

This equation still depends on the particular realization of chosen patterns ξ_j^μ. If we would be interested in the average flow obtained after averaging over these patterns, we would run into all difficulties connected with nonlinear stochastic differential equations.[9] In this paper we will avoid these problems by restricting ourselves to the cases where $|I_\eta|=2^{-p}N$.

The method described in this paper can be generalized in different ways. One is to include a neuronal threshold h_t, so that the expression for the local field becomes

$$h_i=\sum_j J_{ij}(s_j+1)-h_t\ .$$

It is also possible to add a constant ferromagnetic or antiferromagnetic coupling J/N to the J_{ij} of Eq. (1). In this case Eq. (10) takes the form

$$\frac{\partial P(\mathbf{q},m,t)}{\partial t}=-\sum_{\mu=1}^{p}\frac{\partial}{\partial q_\mu}(F_\mu P)-\frac{\partial}{\partial m}(F_0 P)\ , \tag{13}$$

where the flow fields are given by

$$F_\mu=-q_\mu+\langle\eta^\mu\tanh\beta(\boldsymbol{\eta}A\mathbf{q}+mJ+J-h_t)\rangle_\eta \tag{14}$$

and

$$F_0=-m+\langle\tanh\beta(\boldsymbol{\eta}A\mathbf{q}+mJ+J-h_t)\rangle_\eta\ . \tag{15}$$

The variable m is the average spin

$$m=\frac{1}{N}\sum_j s_j\ .$$

The equations for the flow become

$$\frac{dq_\mu}{dt}=F_\mu \quad \text{and} \quad \frac{dm}{dt}=F_0\ . \tag{16}$$

In principle the variable m should also have been con-

sidered as a macroscopic variable in deriving Eq. (10). However, since $J=h_t=0$ it would have had no effect on the $\mathbf{q}(t)$. The equation for $m(t)$ itself would have become

$$\frac{dm}{dt}=-m+\langle\tanh(\beta\eta A\mathbf{q})\rangle_\eta ,$$

so that $m(t)$ can be written in terms of $m(0)$ and $\mathbf{q}(t)$.

III. A SIMPLE APPLICATION

We will now discuss the results of some numerical solutions of Eq. (12) for the case of two patterns, i.e., $p=2$. In Fig. 1 we show the flow lines in the q_1-q_2 plane for three different temperatures and with three different forms for the matrix A. The first choice

$$A=\begin{bmatrix}1 & 0\\ 0 & 1\end{bmatrix}$$

corresponds to the standard Hebb connections. The equilibrium properties of this model have been studied extensively by Amit, Gutfreund, and Sompolinsky.[2]

We see that for all cases with large T the vector $\mathbf{q}(t)$ approaches the origin, so that no correlations with the built-in patterns remain. The strongest correlation is obtained for low temperatures. For the third (asymmetric) A these correlations switch back and forth between the two patterns and their negatives. This is manifested by the way in which the state point $\mathbf{q}(t)$ traverses the limit cycle, spending most time near the corners where the pure states $(\pm1,0)$ and $(0,\pm1)$ are found.

A remarkable feature is the fact that for low temperatures the flow lines seem to consist of straight lines connected by sharp bends. This can be explained by observing that for large β the function $\tanh(\beta\ldots)$ is either $+1$ or -1, so that in Eq. (12) the second term in the right-hand side assumes only a few different values $\mathbf{c}_l$ in a number of regions R_l in the $\mathbf{q}$ plane. In each region $\mathbf{R}_l$ the solution therefore decays exponentially towards $\mathbf{c}_l$, until it enters another region. Hence the sharp bends.

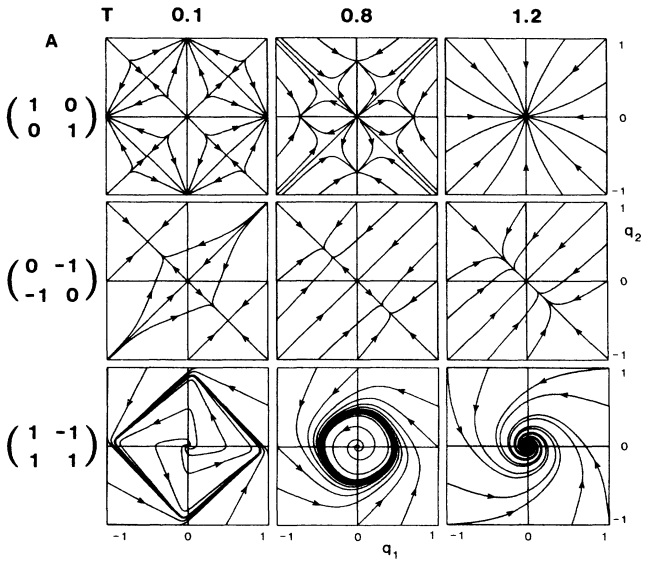

FIG. 1. Examples of the flow described by Eq. (12).

We want to close this paper with a few remarks about fluctuations. The equations (12) and (16) are true macroscopic equations when the p built-in patterns are fixed. If, however, the patterns themselves are fluctuating for some reason, these equations are nonlinear stochastic differential equations, which only serve as starting point, from which the average behavior of $\mathbf{q}(t)$ and its fluctuations should be calculated.

Another remark is that a description in terms of $\mathbf{q}(t)$ of the evolution of an image may be too crude. It involves, after all, a reduction of the number of variables from N to p, which may be too drastic. An intermediate description is obtained by using the average magnetizations $m_\eta(\mathbf{s})$ [Eq. (9)] of the spins of the index sets I_η, of which there are 2^p. On this level it then becomes interesting to ask how many images and which ones correspond to a final state in $\mathbf{q}$ space.

[1]J. J. Hopfield, Proc. Natl. Acad. Sci. USA **79**, 2554 (1982).
[2]D. J. Amit, H. Gutfreund, and H. Sompolinsky, Phys. Rev. A **32**, 1007 (1985); **35**, 2293 (1987).
[3]B. Derrida, E. Gardner, and A. Zippelius, Europhys. Lett. **4**, 167 (1987).
[4]R. Kree and A. Zippelius, Phys. Rev. A **36**, 4421 (1987).
[5]B. Derrida and J. P. Nadal, J. Stat. Phys. **49**, 993 (1987).
[6]R. Meir and E. Domany, Phys. Rev. A **37**, 608 (1988).
[7]E. Gardner, B. Derrida, and P. Mottishaw, J. Phys. (Paris) **48**, 741 (1987).
[8]J. L. van Hemmen, D. Grensing, A. Huber, and R. Kühn, Z. Phys. B **65**, 53 (1986).
[9]N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).