

Dual processes in neural network models I. Neural dynamics versus dynamics of learning

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1992 J. Phys. A: Math. Gen. 25 2577

(<http://iopscience.iop.org/0305-4470/25/9/028>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.62

The article was downloaded on 01/06/2010 at 18:30

Please note that [terms and conditions apply](#).

Dual processes in neural network models: I. Neural dynamics versus dynamics of learning

A C C Coolen†‡§ and L G V M Lenders†

† Department of Medical Physics and Biophysics, University of Nijmegen, Geert Grooteplein Noord 21, NL-6525 EZ Nijmegen, The Netherlands

‡ Utrecht Biophysics Research Institute, University of Utrecht, Princetonplein 5, NL-3584 CC Utrecht, The Netherlands

Received 21 August 1991

Abstract. The dynamics of learning in an unsupervised formulation of the Kohonen model is shown to be equivalent to the dynamics of local order parameters in an attractor neural network with short-range Hebbian interactions and long-range anti-Hebbian interactions. This duality can be used to study analytically the creation of topology conserving maps.

1. Introduction

Neural networks can, in crude approximation, be viewed as stochastic spin systems with time-dependent interactions. The spins (or neurons) are fast variables, which are considered to be the information processing elements (varying on a time-scale of typically a few milliseconds). The interactions (or synapses) are slow variables, which are responsible for the distributed storage of both the 'program' of the system at hand and the 'data' to be recalled or processed (with a typical time-scale of at least a few hundred milliseconds). In this picture 'learning' is the process of modifying the spin interactions. There is a complicated interplay between the two types of dynamic variables. The spin dynamics is (at zero temperature) a local field alignment, involving the (generally long-range and non-symmetric) interactions. The exact laws governing the interaction dynamics, however, have not yet been determined. All that is known is that in biological neural networks they must have a local character: the modification of a given pair interaction may depend only on states, local fields and current interaction strengths of the two spins involved. Because of the different time-scale of the two dynamical processes in neural networks and of our lack of detailed knowledge of the rules involved in the dynamics of learning, the first step in statistical mechanical studies has been to study the two processes separately.

Following the papers by Little [1] and Hopfield [2] many types of models have been studied in which the interactions are fixed and the spin states are the only degrees of freedom. The first systematic equilibrium statistical mechanical study of systems with fixed symmetric Hebbian [3] interactions was carried out by Amit *et al* [4-6].

§ Present address: Department of Theoretical Physics, University of Oxford, 1 Keble Road, Oxford OX1 3NP, UK.

Extensions to more general symmetric non-Hebbian interaction matrices were developed in [7]. If the interactions are non-symmetric, the local field alignment can no longer be written as a Glauber dynamics, and therefore equilibrium statistical mechanics is no longer applicable. In this case, however, one can often derive, from the microscopic dynamics, evolution equations for suitably defined order parameters [8–13]. A detailed overview of the statistical mechanics of Ising spin neural networks can be found in textbooks such as [14–18]. Following the work by Gardner [19] many physicists have now turned their attention towards the complementary problem: how to study analytically the process of learning if the statistics of the spins are given. This appears to be a completely different type of problem. Again the first step was to apply equilibrium statistical mechanics, which has given us insight into aspects of learning like information storage capacity [19, 20], convergence times [21, 22] and generalization [23]. Also the outcome of nonlinear learning processes like the perceptron could be calculated in some cases [24]. An attempt to study analytically the interplay between neural dynamics and the dynamics of learning was presented in [25].

We are now in a situation where the field of neural network theory appears to cover two distinct classes of problems: the analysis of neural dynamics in systems with fixed interactions (class A) and the analysis of the dynamics of learning in systems where the spin statistics are given (class B). It is the purpose of this paper to show that there are non-trivial problems in class B which can be mapped onto problems in class A (about which much more is known). In particular we will set up a duality between the dynamics of interactions in an unsupervised formulation of the Kohonen [26, 27] model and the dynamics of local order parameters in attractor networks with (fixed) short-range Hebbian interactions in combination with (fixed) long-range anti-Hebbian interactions.

This paper is organized as follows. In section 2 we derive the evolution equations of local order parameters in attractor networks with spatial structure and interactions of both the Hebbian and anti-Hebbian type (which is a generalization of [28, 29]). In section 3 we show how the very same field equations describe the evolution of interactions (and thus the creation of topology conserving maps) in an unsupervised formulation of the Kohonen [26, 27] model. We present some numerical results describing the creation of topology conserving maps. In section 4 the field equations are studied in more detail: we calculate the critical temperature and a Liapunov functional which, in equilibrium, can be identified with the free energy. We derive some of the constraints that must be imposed upon the spatial structure of the attractor network. An intensive study of the fixed-point equations is considered to be beyond the scope of this paper and will be published in a forthcoming paper. Finally we discuss the results obtained. Appendix B contains an overview of the notation introduced and a brief description of the main variables.

2. Ising spin neural networks with spatial structure

In this section we derive evolution equations for local order parameters in Ising spin neural networks with spatial structure and fixed neural interactions. We start with a network of N neurons, each of which can be in two states, $s_i = \pm 1$, $i = 1, \dots, N$. The N -neuron network state will be written as s . A finite number of N -component vectors ξ^μ ($\mu = 1, \dots, p$) represent the information which is built in using Hebbian rules [3] and which determines the connection matrix J_{ij} for the effect of neuron j on neuron

i by

$$J_{ij} = \begin{cases} \frac{1}{S} \sigma_{ij} \sum_{\mu=1}^p \xi_i^\mu \xi_j^\mu & \text{if } j \in S_i \\ 0 & \text{if } j \notin S_i. \end{cases} \tag{1}$$

Here we have denoted by S_i the set of neurons which can send a signal to neuron i . The variables $\sigma_{ij} \in \{-1, 1\}$ indicate the type of the interactions: J_{ij} is Hebbian if $\sigma_{ij} = 1$, J_{ij} is anti-Hebbian if $\sigma_{ij} = -1$. S is the average number of neurons contributing to an input, i.e. $S = 1/N \sum_j |S_j|$, where $|S_j|$ is the number of neurons in the set S_j . If, furthermore, the components ξ_j^μ are drawn from the set $\{-1, 1\}$, then the vectors ξ^μ are by definition microscopic network states and can be thought of as patterns. Throughout this paper, however, we will impose no restrictions on the values of the components ξ_j^μ . Note that the interaction matrix (1) need not be symmetric; its symmetry properties still depend on the choice made for the sets S_i and the variables σ_{ij} .

As in statistical mechanics we will consider an ensemble of states, so we can speak about the probability $p_t(s)$ of finding the system at time t in the state s . The evolution of this probability is now assumed to be governed by a stochastic process, in which the probability per unit time for neuron j to flip from s_j to $-s_j$ is some given function of the total state s , which will be denoted by $w_j(s)$. The master equation for this process is

$$\frac{dp_t(s)}{dt} = \sum_j w_j(F_j s) p_t(F_j s) - p_t(s) \sum_j w_j(s) \tag{2}$$

where F_j is an operator, defined by

$$F_j \Phi(s_1, \dots, s_j, \dots, s_N) = \Phi(s_1, \dots, -s_j, \dots, s_N).$$

For $w_j(s)$ we make the usual choice

$$w_j(s) = \frac{1}{2} [1 - \tanh(\beta s_j h_j(s))] \tag{3}$$

where $\beta = 1/T$ (the ‘temperature’ T being a measure of the amount of noise) and the input $h_j(s)$ (or local field) acting upon neuron j is

$$h_j(s) = \sum_{i \in S_j} J_{ji} s_i. \tag{4}$$

In general it will be impossible to find the solution of equation (2). However, as in statistical mechanics, we are not interested in the microscopic details of a network state, but rather in the question of whether the values of certain *macroscopic* features can be calculated. In appendix A we analyse the evolution in time of any set of linear order parameters Ω_μ :

$$\Omega_\mu(s) = \sum_i \omega_i^{(\mu)} s_i \quad (\mu = 1, \dots, n). \tag{5}$$

For the case in which the local fields $h_i(s)$ depend on s , only through the values of the order parameters, $h_i(s) = h_i[\Omega(s)]$, and if the number n of order parameters is not too large, we find deterministic evolution equations for the order parameters in the limit $N \rightarrow \infty$:

$$\frac{d}{dt} \Omega_\mu(t) = \lim_{N \rightarrow \infty} \sum_j \omega_j^{(\mu)} \tanh(\beta h_j[\Omega(t)]) - \Omega_\mu(t). \tag{6}$$

Here we have assumed $\omega_i^{(\mu)} = O(|V_\mu|^{-1})$ for all i, μ (in order to have well defined order parameters in the limit $N \rightarrow \infty$). The support V_μ of Ω_μ is defined as $V_\mu \equiv \{i | \omega_i^{(\mu)} \neq 0\}$, $|V_\mu|$ being the number of elements in V_μ (for simplicity we take all $|V_\mu|$ to be of the same order in N , i.e. $\exists V: \lim_{N \rightarrow \infty} |V_\mu|/V = O(1)$). The restriction on the number of order parameters is

$$\lim_{N \rightarrow \infty} n/\sqrt{V} = 0. \tag{7}$$

A special case of a set of order parameters satisfying all these requirements will now be considered.

In order to describe the macroscopic dynamical behaviour of a network with a given spatial structure, as given by (1), we divide the system into a large number of non-overlapping clusters $\lambda = 1, 2, \dots, \Lambda$ of adjacent neurons. The indices of the neurons in cluster λ form the set I_λ , which contains $|I_\lambda| = N/\Lambda \gg 1$ indices. We now define macroscopic variables by

$$Q_{\lambda\mu}(s) = \frac{1}{|I_\lambda|} \sum_{i \in I_\lambda} s_i \xi_i^\mu. \tag{8}$$

A macroscopic state is now defined by the vector $\mathbf{Q} = (Q_{11}, \dots, Q_{\Lambda p})$, where each $Q_{\lambda\mu}$ ranges between -1 and 1 . In order to show that the inputs $h_j(s)$ can also be considered as a function of $\mathbf{Q}(s)$, we write $\sum_{i \in S_j}$ in equation (4) as $\sum_{\lambda'} \sum_{i \in D_{j\lambda'}}$, where $D_{j\lambda'}$ is the intersection of S_j and $I_{\lambda'}$ (see figure 1). Then we obtain

$$h_j(s) = \frac{1}{S} \sum_{\mu'} \xi_j^{\mu'} \sum_{\lambda'} \sum_{i \in D_{j\lambda'}} \sigma_{ji} s_i \xi_i^{\mu'}.$$

If S is taken to be sufficiently large (i.e. $S \gg N/\Lambda$) and if on the cluster scale the connections are of uniform type (i.e. $\sigma_{ji} = \sigma_{\lambda\lambda'}$ for all $j \in I_\lambda, i \in I_{\lambda'}$) it is seen from the definition (8) that

$$h_j(s) \simeq \sum_{\lambda'} \frac{|D_{j\lambda'}|}{S} \sigma_{\lambda\lambda'} Q_{\lambda'\mu'}(s) \xi_j^{\mu'}.$$

In the limit $N \rightarrow \infty$ this becomes an equality (if S_j is chosen such that the boundaries ∂S_j coincide with the boundaries of the sets I_λ , then the equality holds without any

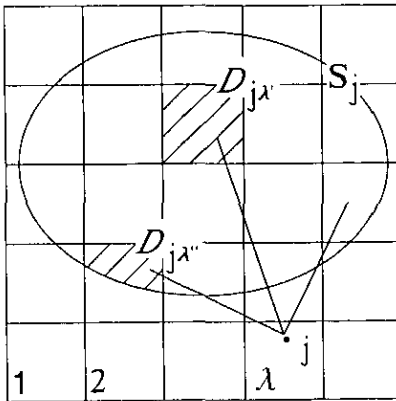


Figure 1. The partitioning of the sum over all contributions to the input of neuron j ; $D_{j\lambda'}$ denotes the intersection of cluster λ' and the set S_j of all neurons contributing to this input.

restriction on S). Equation (6) can now be used to describe the evolution in time of Q . However, in order to simplify the final result we will also assume that for all $i, j \in I_\lambda: |D_{i\lambda}| = |D_{j\lambda}| \equiv |D_{\lambda\lambda}|$ for all λ' (put differently: the interaction density is assumed to be constant on the cluster scale). This allows us to define

$$n(\lambda, \lambda') \equiv \sigma_{\lambda\lambda'} |D_{\lambda\lambda'}| S^{-1}$$

$|n(\lambda, \lambda')|$ is the density of connections from $I_{\lambda'}$, to any neuron in I_λ . By definition it is normalized according to

$$\Lambda^{-1} \sum_{\lambda\lambda'} |n(\lambda, \lambda')| = 1.$$

Since equation (6) can now be applied to the order parameters $Q_{\lambda\mu}$ we can write for $N \rightarrow \infty$ (using $|I_\lambda| = N/\Lambda$):

$$\frac{d}{dt} Q_{\lambda\mu}(t) = \lim_{N \rightarrow \infty} \frac{\Lambda}{N} \sum_{j \in I_\lambda} \xi_j^\mu \tanh \left(\beta \sum_{\lambda'\mu'} n(\lambda, \lambda') \xi_j^{\lambda'} Q_{\lambda'\mu'}(t) \right) - Q_{\lambda\mu}(t). \quad (9)$$

Finally we take a continuum limit and replace the cluster labels λ and λ' by the continuous position vectors x and y ($x, y \in D$); for $Q_{\lambda\mu}(t)$ we write $q_\mu(x, t)$. By taking this limit, equation (9) becomes

$$\frac{\partial q(x, t)}{\partial t} = A(x, q) - q(x, t) \quad \text{with } q(x, 0) = q_0(x) \quad (10)$$

where

$$A(x, q) = \left\langle \xi(x) \tanh \beta \left(\xi(x) \cdot \int_D dy n(x, y) q(y, t) \right) \right\rangle_{\xi(x)} \quad (11)$$

and q, A and $\xi(x)$ are p -dimensional vectors. The average $\langle \rangle_{\xi(x)}$ is computed over the distribution of the vectors $(\xi_1^1, \dots, \xi_1^p)$ at position x ; it therefore depends on their specific realization. The function $n(x, y)$ contains all information regarding the network's structure: $|n(x, y)|$ is the density of connections from position y to position x , $\text{sgn}[n(x, y)]$ indicates the type ($\text{sgn} = 1$: Hebbian, $\text{sgn} = -1$: anti-Hebbian). The structure function n is (by definition) normalized according to

$$|D|^{-1} \int_D \int_D dx dy |n(x, y)| = 1 \quad |D| \equiv \int_D dx. \quad (12)$$

We assumed p to be fixed. The number n of order parameters in (9) is $p\Lambda$. The size V of their supports is N/Λ . In order to write local inputs as functions of Q we have assumed $S \gg N/\Lambda$; according to (7) the conditions under which equations (10) and (11) will hold are therefore found to be

$$\lim_{N \rightarrow \infty} \frac{N}{\Lambda S} = 0 \quad \lim_{N \rightarrow \infty} \frac{\Lambda^3}{N} = 0.$$

One might choose $\Lambda = N^\epsilon$ and $S = N^\gamma$, such that $0 < \epsilon < \frac{1}{3}$ and $1 - \epsilon < \gamma < 1$.

If the components ξ_j^μ do not vary within the clusters I_λ , equation (11) reduces to

$$A(x, q) = \xi(x) \tanh \beta \left(\xi(x) \cdot \int_D dy n(x, y) q(y, t) \right).$$

In this case we can make a decomposition: $q(x, t) = q(x, t)\xi(x) + q^\perp(x, t)$, where $q^\perp(x, t) \cdot \xi(x) = 0$. It then follows from (10) that

$$q^\perp(x, t) = q^\perp(x, 0) e^{-t}$$

$$\frac{\partial}{\partial t} q(x, t) = \tanh\left(\beta \int_D dy \hat{n}(x, y) q(y, t)\right) - q(x, t)$$

where $\hat{n}(x; y) \equiv n(x; y)\xi(x) \cdot \xi(y)$ and $q(x, t)$ turns out to be the average activity (or magnetization) at position x at time t . If, on the other hand, the vectors ξ^μ are drawn independently at each position x (from a given distribution $\rho(\xi)$), then

$$A(x, q) = \left\langle \xi \cdot \tanh \beta \left(\xi \cdot \int_D dy n(x, y) q(y, t) \right) \right\rangle_\xi. \quad (13)$$

It is important to stress that there are no further restrictions on the distribution $\rho(\xi)$. Only if all ξ^μ are drawn from the set $\{-1, 1\}$ can we think in terms of patterns being stored; applications of equations (10) and (11) where input vectors correspond to patterns can be found in [29]. If, however, the vectors ξ have components which are not restricted to a discrete set, the picture of stored patterns might have to be abandoned.

3. The dual process: creation of topology conserving maps

In this section we will define an unsupervised form of the Kohonen model [26, 27] for the creation of topology conserving maps. In this two-layer system the dynamic quantities are the neural interactions. In an input layer, consisting of p real-valued neurons, input vectors s are presented in random order (drawn from a given probability distribution). These vectors are taken to be normalized: $|s| = 1$ for all s . The second layer consists of a (usually 2D) array of neurons, which receive signals h_i from the input layer:

$$h_i \equiv \sum_j J_{ij} s_j. \quad (14)$$

A set B_i of neighbours is associated with each neuron i in the second layer ($i \in B_i$). The interactions J_{ij} between the two layers evolve in time according to the following stochastic procedure:

- (i) choose at random an input vector s ;
- (ii) determine k such that $h_k \geq h_i$ for all i ;
- (iii) $\forall l \in B_k: \Delta J_{lj} = \epsilon s_j$ (for all j) ($0 < \epsilon \ll 1$);
- (iv) $\forall l \in B_k: J_{lj} \rightarrow J_{lj} (\sum_m J_{lm}^2)^{-1/2}$ (for all j);
- (v) return to (i).

As was shown by Kohonen [26, 27], the net result of this procedure is that the neurons in the second layer tend to become 'feature detectors' (tuned to one specific input vector). Furthermore, because of the incorporation of the sets B_k in the updating procedure, neighbouring neurons are forced to detect similar features. If the set of input vectors from which the examples are drawn is topologically equivalent to the physical array of the second layer, an internal representation will be formed of the set of input vectors. In this case one can associate with each input vector a position on the array, such that presentation of this input vector causes an activity peak in the array exactly at this position. The mapping from the set of input vectors to the positions of the corresponding activity peaks in the array is topology conserving [26, 27]. From now on we will refer to the second layer as the map (in accordance with the literature).

From a physiological point of view the Kohonen model has some unattractive features. Firstly, a supervisor is needed to determine which neuron is activated most upon presentation of an input vector. Secondly, connections are normalized by hand. From a theoretical point of view the problem with Kohonen's model is that the interaction dynamics is extremely discontinuous, which makes analysis rather difficult.

As an alternative we propose the following. First we replace the normalization step by a decay; the modification step now reads

$$(\forall m, j): \Delta J_{mj} = \varepsilon (s_j f_m[\mathbf{h}(s)] - J_{mj})$$

where

$$\begin{aligned} f_m(\mathbf{h}) &= 1 && \text{if } m \in B_{k(\mathbf{h})} \\ f_m(\mathbf{h}) &= 0 && \text{if } m \notin B_{k(\mathbf{h})} \end{aligned}$$

$k(\mathbf{h})$ is defined as

$$(\forall l \neq k): h_k > h_l.$$

The function of f_m is twofold: firstly, through the introduction of the sets $B_{k(\mathbf{h})}$ a local convolution is introduced that ensures neighbouring neurons modify their weights in approximately the same way; secondly, f_m introduces nonlinearity into the weight dynamics. In order to arrive at a continuous dynamical law we will make the replacement

$$f_m(\mathbf{h}) \rightarrow f \left[\sum_l n_{ml} h_l \right].$$

Here f is a monotonic nonlinear function to be specified ($f'(x) \geq 0$); n_{ml} defines a local convolution which monotonically decreases as a function of the distance between the physical locations x_m and x_l of the neurons m and l in the array. Now all neurons update their interactions, nonlinearly weighted by the strength of their (convoluted) inputs. The system's dynamics is now given by:

- (i) choose at random an input vector s ;
- (ii) $\forall i, j: \Delta J_{ij} = \varepsilon (s_j f[\sum_k n_{ik} h_k] - J_{ij})$ ($0 < \varepsilon \ll 1$);
- (iii) return to (i).

By taking the limit $\varepsilon \rightarrow 0$ we obtain

$$\frac{d}{dt} J_{ij} = \left\langle s_j f \left[\sum_{kl} n_{il} J_{lk} s_k \right] \right\rangle_s - J_{ij}. \tag{15}$$

The average in (15) is defined over the probability distribution of the input vectors. To establish contact with the laws governing the dynamics of order parameters in the model of section 2, we need only make the choice $f(x) \equiv \tanh(\beta x)$, take a continuum limit and switch notation: $i \rightarrow x, j \rightarrow \mu, s \rightarrow \xi, n_{il} \rightarrow n(x, y)$ and $J_{ij}(t) \rightarrow J_\mu(x, t)$ (now $J_\mu(x, t)$ represents the interaction strength at time t from input channel μ to position x in the map). The result is

$$\frac{\partial \mathbf{J}(x, t)}{\partial t} = \left\langle \xi \tanh \beta \left(\xi \cdot \int_D dy n(x, y) \mathbf{J}(y, t) \right) \right\rangle_\xi - \mathbf{J}(x, t). \tag{16}$$

If we identify the convolution kernel n with the structure function n of the attractor network and if we also identify the nonlinearity parameter β with the inverse temperature in the attractor network, equation (16) shows that the laws governing the processes in sections 2 and 3 are exactly the same. Table 1 shows an overview of this duality.

Table 1. The duality between the dynamics of interactions in topology conserving maps and the dynamics of local order parameters in suitably defined attractor networks.

$$\partial_t \psi_\mu(x, t) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_i \xi_i^\mu \tanh\left(\beta \xi_i \cdot \int dy n(x, y) \psi(y, t)\right) - \psi_\mu(x, t)$$

	Topology conserving maps (dynamic interactions in a layered network)	Attractor networks (dynamic neurons in a recurrent network)
$\psi_\mu(x, t)$	Interaction from input channel μ to position x in the map	Value of μ th local order parameter
$n(x, y)$	Convolution in the map	Structure function
ξ_i^μ	i th sample from μ th input channel	i th component of μ th stored vector
N	Number of input examples	Number of neurons
p	Number of input channels	Number of stored vectors
β	Nonlinearity parameter	Noise parameter

A number of assumptions have been made in section 2 in order to demonstrate the equivalence between the processes in sections 2 and 3 (apart from the definitions of the microscopic dynamical laws and the Hebbian-type interaction matrix). The assumptions that are essential for having deterministic evolution of order parameters in the attractor model (in the thermodynamic limit) are the scaling requirements

$$N/\Lambda S \rightarrow 0 \quad \text{and} \quad \Lambda^3/N \rightarrow 0 \quad (\text{for } N \rightarrow \infty).$$

If these requirements were to be violated, we would be forced to take into account the fluctuations in the values of the order parameters. The remaining assumptions made in section 2, such as

$$(\forall ij \in I_\lambda)(\forall \lambda'): |I_\lambda \cap S_i| = |I_\lambda \cap S_j|$$

are restrictions on the smoothness of the spatial density of interactions (introduced for computational convenience), which might be eliminated.

The formation of topology conserving maps by a process similar to (16) has already been advocated by Amari [30]. Whether or not equation (16) describes the creation of topology conserving maps will depend on the choice made for the convolution kernel n and for the parameter $T \equiv \beta^{-1}$. We will show that n needs to be positive at a short range and negative at a long range (in accordance with a proposal in [26, 27]). This is the reason why the derivation in section 2 had to include both Hebbian and anti-Hebbian interactions, for in terms of the attractor model the above-mentioned restriction on the allowed convolution kernels implies that the dual attractor network has short-range Hebbian interactions in combination with long-range anti-Hebbian interactions. Figure 2 shows the result of a numerical iteration of equation (16) (starting from a random initial configuration $J(x, 0)$), with $p = 3$, $T = 0.0$, $x \in D \equiv [-\frac{1}{2}, \frac{1}{2}]^2$ and n chosen to be

$$n(x, y) \equiv (2\pi\sigma^2)^{-1} e^{-|x-y|^2/2\sigma^2} - 1.$$

The width σ was taken to be 0.01. Since a system configuration $J(x)$ in this case is a mapping from the square $[-\frac{1}{2}, \frac{1}{2}]^2$ onto \mathbb{R}^3 , we choose (following Kohonen [26, 27]) as a graphical representation of the equilibrium configuration the projections of $J(x, \infty)$

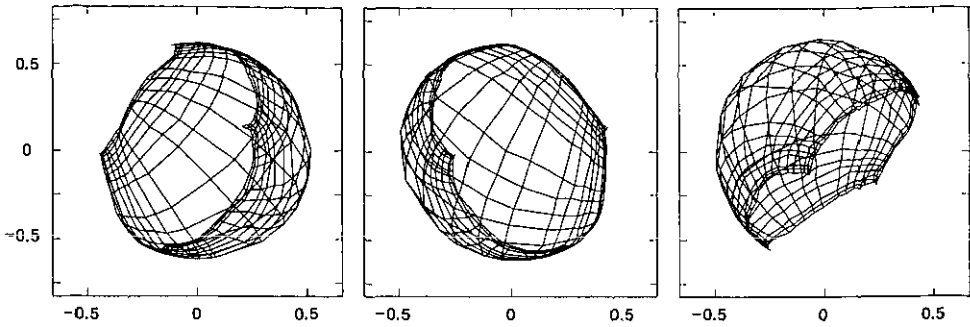


Figure 2. Projections of the equilibrium configuration $J(x, \infty)$ onto the planes $J_1 = 0$ (left), $J_2 = 0$ (middle) and $J_3 = 0$ (right), obtained by numerical iteration of the interaction dynamics.

onto the planes $J_1 = 0$, $J_2 = 0$ and $J_3 = 0$. The distribution of input vectors in this example was chosen to be

$$\rho(\xi) \equiv \frac{3}{4\pi} \left(\frac{4}{3}\right)^{-3} \theta \left[\frac{4}{3} - |\xi| \right].$$

Figure 2 shows that an internal representation is formed of the orientations of the input set.

4. General properties of the field equations

The field equations we arrived at in the previous sections are of the form

$$\partial_t \psi = \langle \xi \tanh(\beta \xi \cdot n \otimes \psi) \rangle_\xi - \psi \tag{17}$$

where $\psi(x, t) \in \mathbb{R}^p$, $x \in D \subset \mathbb{R}^n$ (bounded) and

$$\begin{aligned} (n \otimes f)(x) &\equiv \int_D dy n(x, y) f(y) \\ \langle \Phi(\xi) \rangle_\xi &\equiv \int d\xi \rho(\xi) \Phi(\xi) \end{aligned}$$

(where $\langle \xi^2 \rangle_\xi < \infty$). To simplify our notation we will write spatial averages over D as

$$\langle f \rangle_D \equiv |D|^{-1} \int_D dx f(x) \quad |D| \equiv \int_D dx.$$

First we show that the fields ψ will remain bounded, by writing the solution of (17) in the form

$$\begin{aligned} \psi(x, t) &\equiv \psi_0(x) e^{-t} + \langle \xi \gamma_\xi(x, t) \rangle_\xi \\ \partial_t \gamma_\xi &= \tanh(\beta \xi \cdot n \otimes \psi_0 e^{-t} + \beta \xi \cdot n \otimes \langle \xi' \gamma_{\xi'} \rangle_\xi) - \gamma_\xi \\ \gamma_\xi(x, 0) &= 0 \quad \text{for all } \xi. \end{aligned} \tag{18}$$

Clearly $\gamma_\xi(x, t) \in (-1, 1)$ (for all x, t), which implies that the fields ψ will remain bounded. Throughout the rest of this paper we will assume that n is a compact

symmetric operator with non-negative eigenvalues λ_i . As a consequence we can identify a Liapunov functional F :

$$F[\psi] \equiv \frac{1}{2} \langle \psi \cdot n \otimes \psi \rangle_D - \beta^{-1} \langle \langle \log(2 \cosh[\beta \xi \cdot n \otimes \psi]) \rangle \rangle_{\xi D} \tag{19}$$

since

$$\frac{d}{dt} F[\psi] = |D| \sum_{\mu} \left\langle \frac{\delta F}{\delta \psi_{\mu}} \partial_t \psi_{\mu} \right\rangle_D = - \langle \partial_t \psi \cdot n \otimes \partial_t \psi \rangle_D \leq 0.$$

If n is positive definite $(d/dt)F[\psi] = 0$ implies that ψ is a fixed point. If, on the other hand, n has zero eigenvalues $(d/dt)F[\psi] = 0$ implies that $n \otimes \psi$ no longer varies with time; according to (17) the system will now decay exponentially towards a fixed point. The functional F has a lower bound, since

$$\begin{aligned} F[\psi] &\geq -\beta^{-1} \langle \langle \log(2 \cosh[\beta \xi \cdot n \otimes \psi]) \rangle \rangle_{\xi D} \\ &\geq - \langle \langle \xi \cdot n \otimes \psi \rangle \rangle_{\xi D} \\ &\geq -\lambda_{\max} \langle \langle \xi \rangle \rangle_{\xi} \langle \psi^2 \rangle_D^{1/2}. \end{aligned}$$

Here λ_{\max} is the largest eigenvalue of the operator n . In equilibrium the functional F can be interpreted as the (rescaled) free energy of the attractor network in section 2, for which the (rescaled) Hamiltonian is given by

$$H = -\frac{1}{2} \langle \psi \cdot n \otimes \psi \rangle_D. \tag{20}$$

For a microscopic realization s (corresponding to a macroscopic state ψ) to exist in the attractor network of section 2, the possible values of the fields will have to be restricted (according to (8)) to

$$(\forall x \in D) (\exists \gamma_{\xi}(x) \in [-1, 1]): \quad \psi(x) = \langle \xi \gamma_{\xi}(x) \rangle_{\xi}$$

which, if not true at $t = 0$, will certainly be true near equilibrium (according to (18)). In equilibrium we can write $F = H - TS$, with

$$S[\psi] \equiv \langle \langle S(\beta \xi \cdot n \otimes \psi) \rangle \rangle_{\xi D} \quad S(x) \equiv \log 2 + \log \cosh(x) - x \tanh(x). \tag{21}$$

Since $S(x) \in \langle 0, \log 2 \rangle$ the quantity $S[\psi]$ may be interpreted as a rescaled entropy. If we choose $n(x, y)$ to be a positive constant, expression (19) reduces to the free energy per neuron of the fully connected model as derived in [4].

Next we turn to the temperature dependence of equation (17). Since $\langle \xi^2 \rangle_{\xi}$ is finite the covariance matrix $C_{\mu\nu} \equiv \langle \xi_{\mu} \xi_{\nu} \rangle_{\xi}$ is well defined (with eigenvalues $c_{\mu} \in [0, c_{\max}]$), which allows us to calculate an upper bound for the critical temperature T_c . For $T > T_c$ the only equilibrium solution of (17) is the trivial state $\psi(x) = 0$; for $T < T_c$ there are non-trivial fixed-point solutions of (17). The operator n is again taken to be symmetric and semipositive definite, to ensure that a fixed point will be reached. First we rewrite the fixed-point equation:

$$\phi = n \otimes \langle \xi \tanh[\beta \xi \cdot \phi] \rangle_{\xi} \quad \phi \equiv n \otimes \psi. \tag{22}$$

Equation (22) can also be written as

$$\phi = \beta n \otimes \left\langle \xi (\xi \cdot \phi) \int_0^1 ds (1 - \tanh^2[s \beta \xi \cdot \phi]) \right\rangle_{\xi}.$$

Using the symmetry of n , it now follows that for a fixed point of (17)

$$\langle \psi \cdot \phi \rangle_D = \beta \left\langle \left\langle (\xi \cdot \phi)^2 \int_0^1 ds (1 - \tanh^2[s \beta \xi \cdot \phi]) \right\rangle \right\rangle_{\xi D} \leq \beta \langle \phi \cdot C \phi \rangle_D.$$

In terms of ψ this means

$$\langle \psi \cdot (n \otimes \psi - \beta C n^2 \otimes \psi) \rangle_D \leq 0. \tag{23}$$

On the other hand we know that if $T > c_{\max} \lambda_{\max}$

$$\langle \psi \cdot (n \otimes \psi - \beta C n^2 \otimes \psi) \rangle_D \geq \langle \psi^2 \rangle_D \min_{i,\mu} \lambda_i [1 - \beta c_\mu \lambda_i] \geq 0. \tag{24}$$

Here c_{\max} is the largest eigenvalue of the covariance matrix C and λ_{\max} is the largest eigenvalue of the operator n . Combining the two inequalities (23) and (24) shows that the left-hand side of (23) and (24) equals zero as soon as $T > c_{\max} \lambda_{\max}$. This, in turn (in combination with $T > c_{\max} \lambda_{\max}$) implies $\psi = 0$. At $T_c \equiv c_{\max} \lambda_{\max}$ the system undergoes a second-order phase transition, which can be seen after expanding the solution ψ of the fixed-point equation in powers of $\varepsilon \equiv \beta - \beta_c$:

$$\psi = \psi_0 \varepsilon^{1/2} + \psi_1 \varepsilon + \psi_2 \varepsilon^{3/2} + O(\varepsilon^2). \tag{25}$$

The result of substituting this expansion into the fixed-point equation of (17) is

$$\begin{aligned} n \otimes \psi_0 &= \lambda_{\max} \psi_0 & C \psi_0 &= c_{\max} \psi_0 \\ n \otimes \psi_1 &= \lambda_{\max} \psi_1 & C \psi_1 &= c_{\max} \psi_1 \\ \psi_2 &= \beta_c C n \otimes \psi_2 + \beta_c^{-1} \psi_0 - \frac{1}{3} c_{\max}^{-3} \langle \xi (\xi \cdot \psi_0)^3 \rangle_\xi \end{aligned}$$

from which we can deduce

$$\langle \psi_0^2 \rangle_D = \frac{1}{3} \beta_c c_{\max}^{-3} \langle \langle \xi \cdot \psi_0 \rangle^4 \rangle_{D\xi}. \tag{26}$$

Using (26) we can express the rescaled free energy F in terms of the expansion (25):

$$\begin{aligned} T > T_c: \quad F &= -T \log 2 \\ T < T_c: \quad F &= -T \log 2 - \frac{1}{4} (\beta - \beta_c)^2 \beta_c^{-1} \lambda_{\max} \langle \psi_0^2 \rangle + O(\beta - \beta_c)^{5/2}. \end{aligned}$$

The non-trivial solutions of the fixed-point equations which bifurcate at $T = T_c$ are eigenfunctions of n and C , corresponding to the eigenvalues λ_{\max} and c_{\max} , respectively (since n and C commute).

Finally, we will address the problem of what restrictions are to be imposed on the allowed functions $n(x, y)$. In order to arrive at topology conserving maps in which the structure is not influenced by arbitrary properties of n , a natural choice is to take for n a translation invariant kernel: $n(x, y) \equiv n(|x - y|)$. Since fixed-point solutions of (17) are required to be highly correlated at short distances the function $n(z)$ must obey: $n(0) > 0$, $n'(z) < 0$ for all z . If $n(z)$ were to be positive for all $z \leq \max(|x - y|)$ ($x, y \in D$), the ground state (according to (20)) would be a configuration where, away from the boundary ∂D , ψ would be a constant. The above-mentioned considerations naturally lead to the choice

$$\begin{aligned} n(|x - y|) &= n_+(|x - y|) - J|D|^{-1} \\ n_+(0) > 0 \quad n'_+(z) < 0 \quad \lim_{z \rightarrow \infty} n_+(z) &= 0 \end{aligned} \tag{27}$$

$$0 < J \leq J^* \equiv |D|^{-1} \int_D \int_D dx dy n_+(|x - y|).$$

The reason for the extra restriction $J \leq J^*$ is that for $J > J^*$ the operator n will certainly have negative eigenvalues, since

$$\begin{aligned} \lambda_{\min} &= \min_f \int_D \int_D dx dy f(x) f(y) n(|x-y|) \left(\int_D dx f^2(x) \right)^{-1} \\ &\leq |D|^{-1} \int_D \int_D dx dy n(|x-y|) = J^* - J. \end{aligned}$$

Since the choice (27) implies allowing for an infinite range of the inhibitory interactions in the Kohonen map (or for an infinite range in the attractor model of section 2, respectively), a more natural choice might be to choose for the shape of the interaction kernel a 'Mexican hat'-type function such as

$$n(|x-y|) = n_+ (|x-y|) - J (2\pi\Sigma^2)^{-d/2} \exp[-\frac{1}{2}(x-y)^2/\Sigma^2]$$

(where $d \equiv \dim x$). The calculational disadvantage of the latter choice is clearly that for finite Σ and $J > 0$ the interaction kernel will no longer be positive semidefinite, which implies that the function F (19) will no longer be a Liapunov function. However, in the limit $T=0$ the Hamiltonian H is itself a Liapunov function for all symmetric choices of the interaction kernel n , whatever the eigenvalue spectrum. Therefore, the analysis of the zero temperature limit of the dynamic equations (17), presented in [31], can in principle be extended to the case of having interaction kernels with negative eigenvalues (such as the one above). Physically one expects that choosing 'Mexican hat'-type interactions with $\Sigma \ll |D|^{1/d}$ will lead to an increase in the frequency of occurrence of metastable states (topological defects of the 'butterfly' type [26, 27]), since regions in D with a large spatial separation will no longer be able to interact.

5. Discussion

We have shown that two seemingly unrelated processes, the dynamics of local order parameters in Ising spin neural networks with spatial structure on the one hand and the dynamics of learning in an unsupervised formulation of Kohonen's model on the other, are in fact described by the same partial differential equation. This duality has some interesting consequences. Firstly, it allows one to apply to the problems associated with the dynamics of learning the techniques and intuition developed for describing and analysing the neural dynamics in attractor networks. The process of learning can now also be understood in terms of free energy minima and phase transitions in well known types of systems. Problems like finite size effects in attractor networks, for instance, turn out to be directly related to having only a finite number of input examples in the Kohonen model. Secondly, the duality shows that it might be interesting to study not only attractor networks in which the vectors ξ^μ are microscopic system states (patterns), stored with Hebb's rule, but also systems where these vectors are of a different kind and where both Hebbian and anti-Hebbian interactions are present. We have shown that attractor networks of the latter kind can even be used for creating topology conserving maps.

A next step will be to study the fixed-point equations in more detail (which will be the subject of a forthcoming paper) and to find out to what extent this duality between neural dynamics and dynamics of learning might be a more general property of neural systems with Hebbian-type learning rules.

Acknowledgments

We would like to thank Theo Ruijgrok and Harm Jonker for interesting discussions on this subject.

Appendix A. The evolution of linear order parameters

In this section we derive flow equations for order parameters Ω_μ of the form

$$\Omega_\mu(s) = \sum_i \omega_i^{(\mu)} s_i.$$

The vector $\Omega(s) \equiv (\Omega_1(s), \dots, \Omega_n(s))$ constitutes our macroscopic level of description. The support V_μ of Ω_μ is defined as: $V_\mu \equiv \{i | \omega_i^{(\mu)} \neq 0\}$. In order to find deterministic equations we assume $\omega_i^{(\mu)} = O(|V_\mu|^{-1})$ for all i, μ ($|V_\mu|$ is the number of elements in V_μ). For simplicity we assume all $|V_\mu|$ to be of the same order in N . The probability $P_t(\Omega)$ that the order parameters will have the value $\Omega = (\Omega_1, \dots, \Omega_n)$ at time t is given by

$$P_t(\Omega) = \sum_s p_t(s) \delta(\Omega - \Omega(s)).$$

Our purpose is to find a differential equation for $P_t(\Omega)$ using the microscopic master equation. This will be possible only if the local fields $h_i(s)$ depend on s only through the values of the order parameters: $h_i(s) = h_i[\Omega(s)]$. Using (2) one finds

$$\frac{d}{dt} P_t(\Omega) = \sum_s \sum_j w_j(s) p_t(s) (\delta(\Omega - \Omega(s) + 2s_j \omega_j) - \delta(\Omega - \Omega(s)))$$

where $\omega_j \equiv (\omega_j^{(1)}, \dots, \omega_j^{(n)})$. The average value of any function $\Phi[\Omega]$ is $\langle \Phi \rangle_t \equiv \int d\Omega P_t(\Omega) \Phi[\Omega]$. Its time derivative is given by

$$\begin{aligned} \frac{d}{dt} \langle \Phi \rangle_t &= \sum_s \sum_j w_j(s) p_t(s) (\Phi[\Omega(s) + 2s_j \omega_j] - \Phi[\Omega(s)]) \\ &= \sum_s \sum_j w_j(s) p_t(s) \left(\sum_{m \geq 1} \frac{2^{2m}}{(2m)!} \sum_{k_1=1}^n \dots \sum_{k_{2m}=1}^n \omega_j^{(k_1)} \dots \omega_j^{(k_{2m})} \frac{\partial^{2m} \Phi[\Omega(s)]}{\partial \Omega_{k_1} \dots \partial \Omega_{k_{2m}}} \right. \\ &\quad \left. - s_j \sum_{m \geq 0} \frac{2^{2m+1}}{(2m+1)!} \sum_{k_1=1}^n \dots \sum_{k_{2m+1}=1}^n \omega_j^{(k_1)} \dots \omega_j^{(k_{2m+1})} \frac{\partial^{2m+1} \Phi[\Omega(s)]}{\partial \Omega_{k_1} \dots \partial \Omega_{k_{2m+1}}} \right). \end{aligned}$$

Inserting the unit operator $\int d\Omega \delta(\Omega - \Omega(s))$ and performing partial integrations gives

$$\begin{aligned} \frac{d}{dt} \langle \Phi \rangle_t &= \int d\Omega \Phi[\Omega] \sum_{m \geq 1} \frac{2^{2m}}{(2m)!} \sum_{k_1=1}^n \dots \sum_{k_{2m}=1}^n \omega_j^{(k_1)} \dots \omega_j^{(k_{2m})} \frac{\partial^{2m}}{\partial \Omega_{k_1} \dots \partial \Omega_{k_{2m}}} \\ &\quad \times \sum_s \sum_j w_j(s) \delta(\Omega - \Omega(s)) p_t(s) + \int d\Omega \Phi[\Omega] \sum_{m \geq 0} \frac{2^{2m+1}}{(2m+1)!} \\ &\quad \times \sum_{k_1=1}^n \dots \sum_{k_{2m+1}=1}^n \omega_j^{(k_1)} \dots \omega_j^{(k_{2m+1})} \frac{\partial^{2m+1}}{\partial \Omega_{k_1} \dots \partial \Omega_{k_{2m+1}}} \\ &\quad \times \sum_s \sum_j s_j w_j(s) \delta(\Omega - \Omega(s)) p_t(s). \end{aligned}$$

From this, since $\langle \Phi \rangle_t = \int d\Omega P_t(\Omega) \Phi[\Omega]$, we deduce

$$\begin{aligned} \frac{d}{dt} P_t(\Omega) &= \sum_{m \geq 1} \frac{2^{2m-1}}{(2m)!} \sum_{k_1=1}^n \dots \sum_{k_{2m}=1}^n \sum_j \omega_j^{(k_1)} \dots \omega_j^{(k_{2m})} \frac{\partial^{2m}}{\partial \Omega_{k_1} \dots \partial \Omega_{k_{2m}}} \\ &\quad \times P_t(\Omega) (1 - \bar{s}_j(\Omega) \tanh(\beta h_j[\Omega])) \\ &+ \sum_{m \geq 0} \frac{2^{2m}}{(2m+1)!} \sum_{k_1=1}^n \dots \sum_{k_{2m+1}=1}^n \sum_j \omega_j^{(k_1)} \dots \omega_j^{(k_{2m+1})} \frac{\partial^{2m+1}}{\partial \Omega_{k_1} \dots \partial \Omega_{k_{2m+1}}} \\ &\quad \times P_t(\Omega) (\bar{s}_j(\Omega) - \tanh(\beta h_j[\Omega])) \end{aligned}$$

where

$$\bar{s}_j(\Omega) \equiv \sum_s s_j \delta(\Omega - \Omega(s)) p_t(s) / \sum_s \delta(\Omega - \Omega(s)) p_t(s).$$

We may now conclude that for finite β

$$\frac{d}{dt} P_t(\Omega) = \sum_{\mu} \frac{\partial}{\partial \Omega_{\mu}} \left(P_t(\Omega) \left(\Omega_{\mu} - \sum_j \omega_j^{(\mu)} \tanh(\beta h_j[\Omega]) \right) \right) + \sum_{m \geq 2} O\left(V \left(\frac{n}{V} \right)^m \right) \quad (A.1)$$

(where $\lim_{N \rightarrow \infty} |V_{\mu}|/V = O(1)$). For $N \rightarrow \infty$ we find deterministic evolution equations for the order parameters if $\lim_{N \rightarrow \infty} n/\sqrt{V} = 0$. In the latter case the solution of equation (A.1) is given by $P_t(\Omega) = \int d\Omega_0 P_0(\Omega_0) \delta(\Omega - \Omega(\Omega_0, t))$, in which $\Omega(\Omega_0, t)$ is the solution of

$$\Omega(0) = \Omega_0 \quad \frac{d}{dt} \Omega_{\mu}(t) = \lim_{N \rightarrow \infty} \sum_j \omega_j^{(\mu)} \tanh(\beta h_j[\Omega(t)]) - \Omega_{\mu}(t). \quad (A.2)$$

Appendix B. Notation

This appendix contains an overview of the notation introduced and a brief description of the main variables.

Ising spin neural networks with spatial structure

$s \in \{-1, 1\}^N$	microscopic spin state
$\xi^{\mu} \in \mathbb{R}^N$ ($\mu = 1 \dots p$)	stored vectors
J_{ij}	spin interactions
$\sigma_{ij} \in \{-1, 1\}$	structure variables (indicating type of a given interaction J_{ij})
S_i	support of neuron i (contains $ S_i $ elements)
$p_t(s)$	microscopic state probability
$w_j(s)$	transition rates
$\hat{h}_i(s) \equiv \sum_{j \in S_i} J_{ij} s_j$	local alignment fields
$\Omega_{\mu}(s)$ ($\mu = 1 \dots n$)	linear order parameters, $h_i(s) = h_i[\Omega(s)]$
V_{μ}	support of Ω_{μ} (contains $ V_{\mu} $ elements)
I_{λ} ($\lambda = 1 \dots \Lambda$)	non-overlapping clusters of adjacent neurons
$Q_{\lambda\mu}(s)$	overlap between $s _{I_{\lambda}}$ and $\xi^{\mu} _{I_{\lambda}}$
$n(\lambda, \lambda')$	coarse-grained spatial structure variables
$x, y \in D$	position vectors (continuum limit: $\lambda \rightarrow x$)
$q_{\mu}(x, t)$ ($\mu = 1 \dots p$)	local linear order parameter (overlaps)
$n(x, y)$	structure function (defining density and type of interactions)

Topology conserving maps

$s \in \mathbb{R}^p, s = 1$	input vectors of the Kohonen model
$h_i \equiv \sum_j J_{ij} s_j$	local fields in the Kohonen map (size map: N)
$B_i (i = 1 \dots N)$	sets of neural neighbours in Kohonen map
ε	learning rate
f	monotonic nonlinear function
β	nonlinearity parameter
$n_{ml}, n(x, y)$	linear convolution on local fields in the map
$J_\mu(x, t)$	interaction strength at time t from input channel μ to position x in the map

The field equations

$\psi(x, t) \in \mathbb{R}^p$	p dynamic fields (representing either local order parameters or dynamic interactions)
$x \in D \subset \mathbb{R}^n$ (bounded)	physical location of neurons
$\langle f \rangle_D \equiv D ^{-1} \int_D dx f(x)$	spatial average
$(n \otimes f)(x)$	integral operator n operating on function f (eigenvalues of n : $\lambda \leq \lambda_{\max}$)
J	strength of background inhibition in n
$\xi \in \mathbb{R}^p$	randomly drawn vectors (probability distribution: ρ)
$C_{\mu\nu} \equiv \langle \xi_\mu \xi_\nu \rangle_\xi$	covariance matrix (eigenvalues: $c_\mu \in [0, c_{\max}]$)
$F[\psi]$	Liapunov function (free energy)
$H[\psi]$	Hamiltonian
$S[\psi]$	entropy

References

- [1] Little W A 1974 *Math. Biosci.* **19** 101
- [2] Hopfield J J 1982 *Proc. Natl Acad. Sci. USA* **79** 2554
- [3] Hebb D O 1949 *The Organization of Behaviour* (New York: Wiley)
- [4] Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. A* **32** 1007
- [5] Amit D J, Gutfreund H and Sompolinsky H 1985 *Phys. Rev. Lett.* **55** 1530
- [6] Amit D J, Gutfreund H and Sompolinsky H 1987 *Phys. Rev. A* **35** 2293
- [7] van Hemmen J L, Grensing D, Huber A and Kuehn R 1986 *Z. Phys. B* **65** 53
- [8] Derrida B, Gardner E and Zippelius A 1987 *Europhys. Lett.* **4** 167
- [9] Coolen A C C and Ruijgrok Th W 1988 *Phys. Rev. A* **38** 4253
- [10] Derrida B and Nadal J P 1987 *J. Stat. Phys.* **49** 993
- [11] Meir R and Domany E 1988 *Phys. Rev. A* **37** 608
- [12] Gardner E, Derrida B and Mottishaw P 1987 *J. Physique* **48** 741
- [13] Riedel U, Kuehn R and van Hemmen J L 1988 *Phys. Rev. A* **38** 1105
- [14] Amit D J 1989 *Modeling Brain Function* (Cambridge: Cambridge University Press)
- [15] Geszti T 1990 *Physical Models of Neural Networks* (Singapore: World Scientific)
- [16] Hertz J, Krogh A and Palmer R G 1991 *Introduction to the Theory of Neural Computation* (Redwood City: Addison-Wesley)
- [17] Kamp Y and Hasler M 1990 *Recursive Neural Networks for Associative Memory* (Chichester: Wiley)
- [18] Mueller B and Reinhardt J 1990 *Neural Networks, an Introduction* (Berlin: Springer)
- [19] Gardner E 1988 *J. Phys. A: Math. Gen.* **21** 257
- [20] Abbott L F 1990 *Network* **1** 105
- [21] Oppen M 1988 *Phys. Rev. A* **38** 3824

- [22] Hertz J A, Thorbergsson G I and Krogh A 1989 *J. Phys. A: Math. Gen.* **22** 2133
- [23] Kinzel W 1990 *Statistical Mechanics of Neural Networks* ed L Garrido (Berlin: Springer)
- [24] Wiegerinck W and Coolen A C C 1992 Analytical expressions for connections of perceptrons *J. Phys. A: Math. Gen.* submitted
- [25] Jonker H J J and Coolen A C C 1991 *J. Phys. A: Math. Gen.* **24** 4219
- [26] Kohonen T 1982 *Biol. Cybern.* **44** 135
- [27] Kohonen T 1988 *Self-Organization and Associative Memory* (Berlin: Springer)
- [28] Noest A J 1989 *Phys. Rev. Lett.* **63** 1739
- [29] Coolen A C C 1990 *Statistical Mechanics of Neural Networks* ed L Garrido (Berlin: Springer)
- [30] Amari S 1990 *Proc. IEEE* **78** 1443
- [31] Coolen A C C and Lenders L G V M 1992 Dual processes in neural network models II: Analysis of zero-temperature fixed-point equations *J. Phys. A: Math. Gen.* **25** 2593.