Journal of Statistical Physics, Vol. 50, Nos. 1/2, 1988

Nonlinear Neural Networks. I. General Theory

J. L. van Hemmen,[1] D. Grensing,[2] A. Huber,[2] and R. Kühn[2,3]

Received April 14, 1987; revision received August 7, 1987

A neural network is called nonlinear if the introduction of new data into the synaptic efficacies has to be performed through a *non*linear operation. The original Hopfield model is linear, whereas, for instance, clipped synapses constitute a nonlinear model. Here a general theory is presented to obtain the statistical mechanics of a neural network with finitely many patterns and arbitrary (symmetric) nonlinearity. The problem is reduced to minimizing a free energy functional over all solutions of a fixed-point equation with synaptic kernel Q. The case of clipped synapses with bimodal and Gaussian probability distribution is analyzed in detail. To this end, a simple theory is developed that gives the spectrum of Q and thereby all the solutions that bifurcate from the high-temperature phase.

KEY WORDS: Neural networks; spin glasses; learning rules; clipped and other nonlinear synapses; synaptic kernel; spectral theory.

1. INTRODUCTION

1.1. Recollection As Collective Behavior

One of the fascinating aspects of a neural network is its function as an associative memory with a surprising fault tolerance with respect to both input data errors and internal failures. This fault tolerance, which has also been implemented in electronic hardware, has attracted a great deal of attention.[1-5] In contrast to previous work, it has been modeled as a *collective* action of a large, densely interconnected network of (formal) neurons.

[1] Sonderforschungsbereich 123, Universität Heidelberg, D-6900 Heidelberg, West Germany.
[2] Institut für Theoretische Physik und Sternwarte der Universität Kiel, D-2300 Kiel, West Germany.
[3] Present address: Sonderforschungsbereich 123, Universität Heidelberg, D-6900 Heidelberg, West Germany.

0022-4715/88/0100-0231$06.00/0

The basic idea[1] is to introduce an energy function or Hamiltonian

$$H_N = -\tfrac{1}{2}\sum_{i,j} J_{ij} S(i)\, S(j) \tag{1.1}$$

with suitable, symmetric couplings $J_{ij} = J_{ji}$, to model[6] the neurons by Ising spins $S(i)$, $1 \leqslant i \leqslant N$, and to endow the system with a Monte Carlo (or Glauber) dynamics. It then performs a downhill motion in the (free) energy landscape associated with H_N. Since the Monte Carlo dynamics is a Markov process which converges to equilibrium[7] as time proceeds, the long-time behavior of the neural network is determined by the equilibrium statistical mechanics, i.e., the free energy and the ergodic components,[8] of the Ising spins system with Hamiltonian (1.1).

For suitable J_{ij} (see below) and at low enough temperature, the Ising spin system performs a phase transition, a collective phenomenon *par excellence*. Below the critical temperature, it then possesses several ergodic components (free energy valleys), some of which can be associated directly with the stored memories. So some of the stable points of the flow are related to the memories of the system and if mistakes are introduced in one or several digits of a specific memory, this will result in a point near to it but still *in its basin of attraction*. That is, the system functions as an associative memory.

1.2. Nonlinearity and Locality

Typically, a neuron is connected to about 10^3–10^4 other neurons and, though there is some long-range interaction, the dominant connectivity is to the numerous nearby cells. As a first and crude approximation to reality, we therefore assume a neural network to be fully interconnected.

The patterns to be stored[9] in the synaptic couplings J_{ij} are N-bit words $\{\xi_{i\alpha};\ 1 \leqslant i \leqslant N\}$, which represent specific spin configurations. They are labeled by $1 \leqslant \alpha \leqslant q$, where, in the present paper, q is taken to be (essentially) finite. To facilitate the modeling, the $\xi_{i\alpha}$ are independent, identically distributed random variables, which assume the values ± 1 with equal probability. This assumption corresponds to optimal coding.[1]

Hopfields' original choice was to put[1]

$$J_{ij} = JN^{-1} \sum_{\alpha=1}^{q} \xi_{i\alpha}\xi_{j\alpha} \equiv JN^{-1}\xi_i \cdot \xi_j \tag{1.2}$$

More generally, it is highly desirable to study models with[10,11],[4]

$$J_{ij} = JN^{-1}\phi(\xi_i \cdot \xi_j) \tag{1.3}$$

[4] Unless stated otherwise we will assume that $J = 1$.

for some synaptic function ϕ. The original Hopfield model with couplings (1.2) has $\phi(x)=x$ and is therefore called *linear*. The linearity greatly simplifies the ensuing analysis.

Clipping, also proposed by Hopfield,[1] is another example of (1.3). Clipped synapses have $\phi(x)=\mathrm{sgn}(x)$. This synaptic function, which is highly *non*linear, corresponds to the *minimal* amount of information that can be stored. In practical work it is of great interest. The point is that the original Hopfield model combines binary and analog data processing elements. A pattern presented to the network is given in a binary representation, whereas the storing of this N-bit digital word via (1.2) needs an analog process. The advantage of the function $\phi(x)=\mathrm{sgn}(x)$ is that computing can be performed mainly through logical operations. This also greatly speeds up computer simulations. Another important reason for considering this type of function is that it is far easier to implement in silicon versions than the original, linear synapses (1.2). In what follows we therefore analyze the case of clipped synapses in great detail.

As we will see shortly, when studying forgetful memories, there are ample reasons to consider nonlinear models more general than (1.3). In so doing we will constantly bear in mind two general principles, locality and symmetry. It is hard to imagine that synapses need detailed global information on the patterns. In fact, the prescriptions (1.2) and (1.3) are *local*, i.e., a synapse connecting neurons i and j needs only information on ξ_i and ξ_j, which is locally available to it. Moreover, both (1.2) and (1.3) are *symmetric* in i and j. This is the reason behind the Hamiltonian formulation of the dynamics. Under these two assumptions, locality and symmetry, the most general synaptic interaction is given by[10]

$$J_{ij}=N^{-1}Q(\xi_i;\xi_j) \tag{1.4}$$

for some *synaptic kernel* $Q(\mathbf{x};\mathbf{y})=Q(\mathbf{y};\mathbf{x})$ defined on $\mathbb{R}^q\times\mathbb{R}^q$. Locality is extremely important in practical work; symmetry can be dispensed with.

1.3. Outline and Summary

This paper and the following[18] are devoted to a comprehensive study of *nonlinear* neural networks *à la* (1.3) and (1.4). The free energy of the model (1.1) with interaction (1.4) and arbitrary synaptic kernel Q is obtained in Section 2. The ξ_i are taken to be random vectors in $\mathbb{R}^q$ (q fixed), whose components need not necessarily be ± 1. For the moment, the number of patterns q is finite. There are several reasons for studying the finite-q case in detail. First, since our aim is to isolate generic properties, it deserves attention in its own right, the more since the nonlinearity gives

rise to new phenomena as compared to the original Hopfield model.[13] Second, one need not assume any restriction on either the synaptic kernel Q or the spins. Third, the case of extensively many patterns can be reduced[10] to the previous, finite-q case by singling out *finitely* many patterns, taking advantage of the spectral theory developed in the present paper, and averaging over the remaining patterns through the replica method.

To determine the ergodic components of the Hamiltonian (1.1) with the interaction (1.4), one has to solve a fixed-point equation of the form

$$m(\mathbf{x}) = \tanh\left\{\beta \int d\mu(\mathbf{y})\, Q(\mathbf{x};\mathbf{y})\, m(\mathbf{y})\right\} \tag{1.5}$$

where μ is a measure on $\mathbb{R}^q$ which specifies the probability distribution of ξ. As usual, β is the inverse temperature. We note that $m \equiv 0$ is always a solution to (1.5). It corresponds to a trivial (paramagnetic) state. Bifurcations away from zero and associated with them a phase transition and new phases are determined by the spectrum of the integral operator in (1.5) with kernel $Q(\mathbf{x};\mathbf{y})$. So it is important to resolve the associated spectral problem. In general this is highly nontrivial.

Considerable progress is obtained if one becomes more specific. In Section 3 we return to neural networks with random *binary* input data and assume that the components $\xi_{i\alpha}$ of the vectors ξ_i are independent and ± 1 with equal probability. Then $\mathbf{x}$ and $\mathbf{y}$ in (1.5) have equal weight 2^{-q} and range through $\{-1, 1\}^q$, the corners of the (hyper)cube $[-1, 1]^q$ in $\mathbb{R}^q$. Let $\mathcal{G}_q$ be the Abelian group generated by the q inversions $x_\alpha \to -x_\alpha$, $1 \leqslant \alpha \leqslant q$ (we invert only one component), and let g be an element of $\mathcal{G}_q$. The group $\mathcal{G}_q$ contains 2^q elements and g^2 is the identity, whatever g. If for all $\mathbf{x}$ and $\mathbf{y}$ in $\{-1, 1\}^q$ and for all g in $\mathcal{G}_q$

$$Q(g\mathbf{x}; g\mathbf{y}) = Q(\mathbf{x};\mathbf{y}) \tag{1.6}$$

then the spectral problem associated with (1.5) can be solved completely. This is done in Section 3. There it is proven that all Q belonging to the class (1.6) have the *same* eigenvectors v_ρ, though the corresponding eigenvalues λ_ρ may differ. The reason behind this remarkable property is that the corners of the (hyper)cube $[-1, 1]^q$ may be identified with the elements of the group $\mathcal{G}_q$. Then the 2^q characters of $\mathcal{G}_q$ are the eigenvectors of Q. This is a distinguished set, which will be used throughout what follows.

The class (1.6) comprises nearly all known models of a neural network. The point is that these have Q's of the form

$$Q(\mathbf{x};\mathbf{y}) = \phi(\{x_\alpha y_\alpha;\, 1 \leqslant \alpha \leqslant q\}) \tag{1.7}$$

where, by abuse of notation, $\phi(\{z_\alpha\})$ may be, but need not be, a function of q variables z_α. For instance, in (1.3), $\phi(x)$ is a function of a single variable x and $Q(\mathbf{x};\mathbf{y})=\phi(\mathbf{x}\cdot\mathbf{y})$. On the other hand, the synaptic function of most memories that forget[1,15-17] is defined *iteratively*. One starts with $\phi_1(x)=\phi(\mu_1 x_1 y_1)$ for some function $\phi(x)$ of a single variable x and after q iterations ends up with a function ϕ_q of q variables,

$$\phi_q(\{x_\alpha y_\alpha;\, 1\leqslant\alpha\leqslant q\})=\phi(\mu_q x_q y_q+\phi_{q-1}(\{x_\alpha y_\alpha;\, 1\leqslant\alpha\leqslant q-1\})) \quad (1.8)$$

One easily verifies that the function defined by (1.8) satisfies (1.6). In fact, we will show that (1.6) and (1.7) are equivalent. It is to be noted, though, that in solving this general type of model we exploit in an essential way the fact that all corners of $[-1,1]^q$ have equal weight.

In the present case the positive eigenvalues of Q determine the bifurcations from $m(\mathbf{x})\equiv 0$. The first solutions to (1.5) that branch off are associated with the largest positive eigenvalue $\lambda_{\max}$ of Q. Under suitable conditions (see Section 3) they correspond to the q stored patterns. The bifurcation temperature T_c, which is also critical, will be shown[18] to be given by $\beta_c 2^{-q}\lambda_{\max}=1$ or, equivalently, $T_c=2^{-q}\lambda_{\max}$. Let $\tilde{\lambda}$ denote the second largest eigenvalue. The Hopfield model[1] has $\tilde{\lambda}=0$, but, in general, nonlinearity gives rise to a positive $\tilde{\lambda}$. New, originally unstable solutions bifurcate from zero at a temperature $\tilde{T}_q=2^{-q}\tilde{\lambda}$ and they become metastable soon thereafter. So one would like to minimize the fraction $\tilde{T}_q/T_c=\tilde{\lambda}/\lambda_{\max}$. In Section 3 it is shown that for *clipped* synapses with $\phi(x)=\mathrm{sgn}(x)$ this fraction *goes to zero* at least as fast as q^{-1}; in fact, as q^{-2} for every other large q.

In Section 4 we replace the bimodal distribution of the $\xi_{i\alpha}$ by a Gaussian one, solve the spectral problem associated with $\phi(x)=\mathrm{sgn}(x)$ exactly, and prove that $\tilde{\lambda}/\lambda_{\max}\propto q^{-2}$ as q becomes large. So in both cases one can find an extensive temperature range, namely $\tilde{T}_q<T<T_c$, where no other bifurcations have occurred than the ones related to the largest eigenvalue $\lambda_{\max}$. Since the bimodal and Gaussian cases are typical representatives of a *discrete* and a *continuous* probability distribution, respectively, the common features of their spectra suggest some universality.

Finally, Section 5 comprises a discussion of our results and possible extensions. Among other things, we propose a two-stage simulated annealing procedure that gives optimal retrieval. We also note that the number of patterns may increase with N, though rather slowly ($q\ll\ln N$).

In the following paper[18] we study information retrieval and show how the present general considerations can be used to solve various models of a nonlinear neural network, how the associated bifurcation and stability analysis can be performed, which kind of pattern will dominate, and how forgetfulness of a memory can be interpreted.

2. STATISTICAL MECHANICS OF NONLINEAR NEURAL NETWORKS

In this section we determine the free energy of the Ising spin-glass Hamiltonian (1.1) with

$$J_{ij} = N^{-1} Q(\xi_i; \xi_j) \tag{2.1}$$

for some function $Q(\mathbf{x}; \mathbf{y}) = Q(\mathbf{y}; \mathbf{x})$ on $\mathbb{R}^q \times \mathbb{R}^q$. The $\xi_{i\alpha}$ have fixed values, randomly chosen according to their distribution. They need not necessarily be ± 1. The model (2.1) can be solved by using a simple large-deviations argument. For full details the reader is referred to Refs. 10 and 19.

Let us suppose first that we had to solve a much simpler problem: Deriving the free energy

$$-\beta f(\beta) = \lim_{N \to \infty} N^{-1} \ln \operatorname{tr} \exp(-\beta H_N) \tag{2.2}$$

of the Curie–Weiss Hamiltonian

$$H_N = -\tfrac{1}{2} JN \left[N^{-1} \sum_{i=1}^{N} S(i) \right]^2 \equiv -\tfrac{1}{2} JN m_N^2 \tag{2.3}$$

without using the well-known linearization trick.[20] The trace in (2.2) is a normalized one, i.e., for N spins it is 2^{-N} times the usual trace, which is a sum over the 2^N Ising spin configurations. To evaluate it, we note that the whole expression depends on the magnetization m_N. It therefore seems reasonable to take m_N as a new "integration" variable with values between -1 and 1. Suppose we had found the corresponding density, to be called $\mathscr{D}_N(m)$. Then, as $N \to \infty$,

$$\operatorname{tr} \exp(\tfrac{1}{2} N \beta J m_N^2) = \int_{-\infty}^{+\infty} dm\, \mathscr{D}_N(m) \exp[N(\tfrac{1}{2} \beta J m_N^2)] \tag{2.4}$$

$\mathscr{D}_N(m)$ is easily found. It is the fraction of the 2^N spin configurations that have magnetization m. Let $m_N(k) = N^{-1}[-(N-k)+k] = N^{-1}[2k-N]$ be the magnetization for $(N-k)$ spins down and k spins up. Then $k = \frac{1}{2}N(1+m)$ and, by Stirling's formula,

$$\mathscr{D}_N(m) = 2^{-N} \binom{N}{k} = 2^{-N} \binom{N}{\frac{1}{2}N(1+m)} \sim \exp[-Nc^*(m)] \tag{2.5}$$

where

$$c^*(m) = \tfrac{1}{2}[(1+m)\ln(1+m) + (1-m)\ln(1-m)] \tag{2.6}$$

if $|m| \leqslant 1$, and $+\infty$ elsewhere.

In fact, $c^*(m)$ is the Legendre transform

$$c^*(m) = \sup_t [mt - c(t)] \tag{2.7}$$

of the c-function

$$c(t) = \ln \operatorname{tr} e^{tS} = \ln[\cosh(t)] \tag{2.8}$$

For n-vector or soft spins the combinatorial argument (2.5) does not work, but the final result still holds, in that $\mathscr{D}_N(m) = \exp[-Nc^*(m)]$, where $c^*(m)$ is the Legendre transform of a slightly more complicated c-function.[21] This has interesting implications (to be reported elsewhere) if one replaces the step-function approximation of the voltage firing rate of a typical neuron by a smooth sigmoid.

Combining (2.4) and (2.5), we can easily evaluate the free energy (2.2). We get, using a Laplace argument,

$$\begin{aligned} -\beta f(\beta) &= \lim_{N \to \infty} N^{-1} \ln \int_{-\infty}^{+\infty} dm \exp\{N[\tfrac{1}{2}\beta J m^2 - c^*(m)]\} \\ &= \sup_m [\tfrac{1}{2}\beta J m^2 - c^*(m)] \end{aligned} \tag{2.9}$$

The supremum is realized for those m that satisfy the fixed-point equation

$$\beta J m = dc^*(m)/dm = \tanh^{-1}(m) \to m = \tanh(\beta J m) \tag{2.10}$$

We now return to our problem.

Let us suppose first that the ξ's have a discrete probability distribution. Say, the vector ξ assumes, with probability p_γ, n different positions γ, where γ denotes a q-vector. Now the index set $\{1 \leqslant i \leqslant N\}$ may be divided[10,19,22] into n disjoint subsets

$$I_\gamma = \{i: \xi_i = \gamma\} \tag{2.11}$$

whose sizes become deterministic[23] as $N \to \infty$,

$$N^{-1} |I_\gamma| = p_\gamma \tag{2.12}$$

With each I_γ we associate a magnetization or order parameter

$$m_\gamma = |I_\gamma|^{-1} \sum_{i \in I_\gamma} S(i) \tag{2.13}$$

If $\gamma \neq \gamma'$, then these order parameters are not directly correlated.

Using (2.1), (2.12), and (2.13), we rewrite (1.1) in the form

$$-\beta H_N = \tfrac{1}{2}\beta N \sum_{\gamma\gamma'} m_\gamma [p_\gamma Q(\gamma;\gamma')\, p_{\gamma'}]\, m_{\gamma'} \equiv NQ(\mathbf{m}) \tag{2.14}$$

where $\mathbf{m}$ is a vector in $\mathbb{R}^n$ with components m_γ. We have to evaluate the trace of $\exp(-\beta H_N)$. As before, it seems natural to take the m_γ as new integration variables in the limit $N \to \infty$. Since they are not directly correlated (in fact, as integration variables they are independent) the corresponding density is

$$\mathscr{D}_N(\mathbf{m}) = \prod_\gamma \exp[-|I_\gamma|\, c^*(m_\gamma)] = \exp\left\{-N\left[\sum_\gamma p_\gamma c^*(m_\gamma)\right]\right\} \tag{2.15}$$

and thus, by another Laplace argument,

$$\begin{aligned} -\beta f(\beta) &= \lim_{N\to\infty} N^{-1} \ln \int d^n m \exp\left\{N\left[Q(\mathbf{m}) - \sum_\gamma p_\gamma c^*(m_\gamma)\right]\right\} \\ &= \sup_{\mathbf{m}} \left[Q(\mathbf{m}) - \sum_\gamma p_\gamma c^*(m_\gamma)\right] \end{aligned} \tag{2.16}$$

where $c^*(m)$ is defined by (2.6).

The maximum in (2.16) is realized among the $\mathbf{m}$ that satisfy the fixed-point equation [cf. Eq. (2.10)]

$$m_\gamma = \tanh\left[\beta \sum_{\gamma'} Q(\gamma;\gamma')\, p_{\gamma'} m_{\gamma'}\right] \equiv \tanh(x_\gamma) \tag{2.17}$$

A fixed point $\mathbf{m}$ is stable, i.e., gives rise to a (local) maximum, if the second derivative of (2.16) is negative-definite. [We will always say that we "maximize" the free energy functional (2.16), though, of course, the free energy $f(\beta)$ itself is minimized.] Using (2.14) and (2.17), we then find that the matrix with elements

$$\beta p_\gamma Q(\gamma;\gamma')\, p_{\gamma'} - p_\gamma \delta_{\gamma\gamma'} (1 - m_\gamma^2)^{-1} \tag{2.18}$$

should have negative eigenvalues only.

Some comments are in order. An absolute maximum of (2.16) corresponds thermodynamically to a stable phase and a local maximum to a *meta*stable phase, whereas a saddle point or a minimum is to be related to an unstable phase. This distinction is of particular relevance to the Monte Carlo dynamics: Once the system is in a stable or metastable phase, it will never get out in a finite amount of time ($N \to \infty$). On the other hand, unstable phases are left at a finite speed. To see why this is so, we

look at (2.16) and (2.17) from a slightly different point of view. By (2.16), a maximum of the free energy functional corresponds to a minimum of $f(\beta)$ itself. The asymptotic behavior of the Monte Carlo dynamics is determined by the structure of the *ergodic components* or, more loosely formulated, the free energy valleys. These are labeled by the solutions m_γ of the fixed-point equation (2.17) and their thermodynamic stability tells us that a valley is "really a valley," i.e., concave *upward.* So once the system is in it, it will never get out, since the barriers have a height proportional to N and $N \to \infty$; cf. Ref. 24.

As is shown in Section 2.3 of the following paper,[18] for small enough β (high enough temperature), the only solution to (2.17) is $m_\gamma = 0$ for all γ. Let Q be the matrix with elements $Q(\gamma; \gamma')$ and P the diagonal matrix $\{p_\gamma\}$. Moreover, and in contrast to Ref. 10, where this quantity is called λ_1, let $\Lambda_{\max} > 0$ (if any) be the largest eigenvalue of QP. (One easily shows that the eigenvalues of QP coincide with the ones of $P^{1/2}QP^{1/2}$, including multiplicity; hence they are real.) One or several nontrivial solutions to (2.17) branch off into the direction of certain eigenvectors belonging to $\Lambda_{\max}$ and a phase transition occurs as T reaches $T_c = \Lambda_{\max}$.

The expression (2.16) may be simplified so as to avoid an explicit calculation of $c^*(m)$, the Legendre transform of $c(t)$; cf. Eqs. (2.7) and (2.8). Using the fact that $\mathbf{m} = \{m_\gamma\}$ satisfies the fixed-point equation (2.17), one easily verifies, in the present case by explicit calculation, that $c^*(m_\gamma) = m_\gamma x_\gamma - c(x_\gamma)$ and thus

$$-\beta f(\beta) = -\tfrac{1}{2}\beta \sum_{\gamma\gamma'} m_\gamma p_\gamma Q(\gamma; \gamma') p_{\gamma'} m_{\gamma'} + \sum_\gamma p_\gamma c(x_\gamma) \tag{2.19}$$

where we take the solution(s) $\mathbf{m}$ of (2.17) that maximize(s) (2.19). This expression generally holds for any strictly convex and differentiable c-function.[19,25]

What are the modifications needed for a *continuous* probability distribution μ of the ξ's? Simply reinterpret m_γ as a *function* $m(\gamma)$ or, more explicitly, $m(\mathbf{x})$ on the probability space. Instead of (2.17) we now get

$$m(\mathbf{x}) = \tanh\left[\beta \int d\mu(\mathbf{y})\, Q(\mathbf{x}; \mathbf{y})\, m(\mathbf{y})\right] \tag{2.20}$$

while

$$-\beta f(\beta) = \int d\mu(\mathbf{x})\, c\left(\beta \int d\mu(\mathbf{y})\, Q(\mathbf{x}; \mathbf{y})\, m(\mathbf{y})\right) - \tfrac{1}{2}\beta \iint d\mu(\mathbf{x})\, d\mu(\mathbf{y})\, m(\mathbf{x})\, Q(\mathbf{x}; \mathbf{y})\, m(\mathbf{y}) \tag{2.21}$$

replaces (2.19). The proof[19] only requires a mild regularity condition on Q.

In fact, the representation (2.20)–(2.21) holds for any probability distribution of the ξ's. For instance, specializing to the case of random binary input data, we get

$$m(\mathbf{x}) = \tanh\left[\beta 2^{-q} \sum_{\mathbf{y}} Q(\mathbf{x};\mathbf{y})\, m(\mathbf{y})\right] \tag{2.22}$$

which is (2.17). Here $\mathbf{x}$ and $\mathbf{y}$ range through $\{-1, 1\}^q$, the corners of the (hyper)cube $[-1, 1]^q$, which all have equal weight 2^{-q}.

3. SPECTRAL THEORY

For a finite number q of stored patterns, Eq. (2.21) and the associated set of transcendental fixed-point equations (2.20) provide a complete, albeit implicit, description of the equilibrium statistical mechanics of a nonlinear neural network with *arbitrary* synaptic kernel $Q(\mathbf{x};\mathbf{y})$. To study the performance of some particular model, its synaptic function ϕ or, more generally, its kernel $Q(\mathbf{x};\mathbf{y})$ has to be specified.

In this section we concentrate on the case where the components $\xi_{i\alpha}$ are independent and assume the values ± 1 with equal probability (bimodal distribution). The associated fixed-point equation (2.22) already indicates that our first task consists in solving the spectral problem for the $2^q \times 2^q$ matrix $Q(\mathbf{x},\mathbf{y})$, where $\mathbf{x}$ and $\mathbf{y}$ range through the corners $\mathscr{C}^q = \{-1, 1\}^q$ of the unit hypercube $[-1, 1]^q$. This is a formidable task, which has to be performed analytically, since $2^q \geqslant 2048$ for $q \geqslant 11$, out of the reach of most computers.

In Section 3.1 we present a general solution for a rather large class of Q's that satisfy a simple invariance condition. In Section 3.2 we study the properties of the eigenvalues in more detail, while in Section 3.3 we specialize to the important case of clipped synapses.

3.1. General Theory

Let $(\mathbf{x})_i$ denote the component x_i of the vector $\mathbf{x}$ in $\mathbb{R}^q$. Furthermore, let g_α be the inversion with respect to the coordinate axis α. That is, $g_\alpha \mathbf{x}$ has the same coordinates as $\mathbf{x}$ except for $(g_\alpha \mathbf{x})_\alpha = -x_\alpha$. The q inversions g_α generate an Abelian group $\mathscr{G}_q$ with 2^q elements, to be denoted by g. Note that g^2 is always the identity. We now show that all the Q's that satisfy the condition

$$Q(g_\alpha \mathbf{x};\, g_\alpha \mathbf{y}) = Q(\mathbf{x};\mathbf{y}), \qquad 1 \leqslant \alpha \leqslant q \tag{3.1}$$

or, equivalently,[5]

$$Q(g\mathbf{x};\mathbf{y}) = Q(\mathbf{x};g\mathbf{y}) \tag{3.2}$$

for all g in $\mathcal{G}_q$, have a *common* set of eigenvectors, though the corresponding eigenvalues may, and in general will, be different.

Through the operation $\circ$ defined by

$$(\mathbf{x}\circ\mathbf{y})_i = x_i y_i, \qquad 1 \leqslant i \leqslant q \tag{3.3}$$

the corners $\mathcal{C}^q$ form an Abelian group themselves, with $\mathbf{e} = (1, 1, \ldots, 1)$ as unit element. This group is also denoted by $\mathcal{C}^q$. As in $\mathcal{G}_q$, every element of $\mathcal{C}^q$ is its own inverse, i.e., $\mathbf{x}\circ\mathbf{x} = \mathbf{e}$. In fact, $\mathcal{G}_q$ and $\mathcal{C}^q$ are isomorphic,

$$\mathcal{G}_q \cong \mathcal{C}^q \tag{3.4}$$

To see this, note that for each $\mathbf{x}$ there is a *unique* g in $\mathcal{G}_q$ such that $\mathbf{x} = g\mathbf{e}$. Identify $\mathbf{x}$ and g.

We now determine the 2^q characters of $\mathcal{C}^q$. Let ρ be one of the 2^q subsets of $\{1, \ldots, q\}$ and let

$$v_\rho(\mathbf{x}) = \prod_{i\in\rho} x_i \tag{3.5}$$

The empty product ($\rho = \varnothing$) is always one. Plainly,

$$v_\rho(\mathbf{x}\circ\mathbf{y}) = v_\rho(\mathbf{x})\, v_\rho(\mathbf{y}) \tag{3.6}$$

so v_ρ is a character. Moreover,

$$\sum_{\mathbf{x}} v_\rho(\mathbf{x})\, v_{\rho'}(\mathbf{x}) = 2^q \delta_{\rho,\rho'} \tag{3.7}$$

so the v_ρ are orthogonal. Finally, because of (3.6), (3.2), and the group property of $\mathcal{C}^q$, they are eigenvectors of Q,

$$\begin{aligned}
\sum_{\mathbf{y}} Q(\mathbf{x};\mathbf{y})\, v_\rho(\mathbf{y}) &= \sum_{\mathbf{y}} Q(\mathbf{x}\circ\mathbf{e};\mathbf{y})\, v_\rho(\mathbf{x}\circ\mathbf{y})\, v_\rho(\mathbf{x}) \\
&= \left[\sum_{\mathbf{y}} Q(\mathbf{e};\mathbf{x}\circ\mathbf{y})\, v_\rho(\mathbf{x}\circ\mathbf{y})\right] v_\rho(\mathbf{x}) \\
&= \left[\sum_{\mathbf{z}} Q(\mathbf{e};\mathbf{z})\, v_\rho(\mathbf{z})\right] v_\rho(\mathbf{x}) \\
&= \lambda_\rho v_\rho(\mathbf{x})
\end{aligned} \tag{3.8}$$

[5] To obtain (3.2) we have exploited the fact that *here* $g_\alpha^{-1} = g_\alpha$.

where

$$\lambda_\rho = \sum_{\mathbf{x}} Q(\mathbf{e}; \mathbf{x})\, v_\rho(\mathbf{x}) \tag{3.9}$$

is an explicit representation of the eigenvalue λ_ρ corresponding to the eigenvector v_ρ.

Stepping back for a first overview, we see that by virtue of (3.7) we have found all the 2^q characters of $\mathscr{C}^q$, i.e., all its irreducible representations (which are bound to be one-dimensional, since the group is Abelian[6]). Whatever Q, as long as it obeys (3.1) or (3.2), it has the v_ρ as eigenvectors. The eigenvalues, however, do depend on Q, as is evident from (3.9).

It is time to harvest some corollaries. Quite a few eigenvalues λ_ρ may vanish. If Q is odd (even) in the sense that

$$Q(\mathbf{e}; -\mathbf{x}) = \pm Q(\mathbf{e}; \mathbf{x}) \tag{3.10}$$

where the minus sign stands for odd (and plus for even), then $\lambda_\rho = 0$ for $|\rho|$, the number of elements in ρ, being even (odd). This directly follows from (3.5), (3.9), and (3.10).

We can also determine the most general form of the Q that satisfy (3.1) or (3.2). By the spectral theorem,

$$Q(\mathbf{x}; \mathbf{y}) = \sum_\rho \lambda_\rho 2^{-q} v_\rho(\mathbf{x})\, v_\rho(\mathbf{y}) \tag{3.11}$$

Here the 2^{-q} comes from an additional normalization of the v_ρ. Then, by virtue of (3.5), this may be rewritten

$$Q(\mathbf{x}; \mathbf{y}) = \phi(\{x_\alpha y_\alpha; 1 \leqslant \alpha \leqslant q\}) \tag{3.12}$$

for some function $\phi\colon \mathbb{R}^q \to \mathbb{R}$; cf. (1.7). We only need its restriction to $\mathscr{C}^q = \{-1, 1\}^q$.

The representation (3.11) is not as uncommon as it looks. Take, for instance, $\lambda_\varnothing 2^{-q} = c_0$ and $\lambda_\rho 2^{-q} = c_1$ for all ρ with one element ($|\rho| = 1$) while $\lambda_\rho = 0$ for $|\rho| \geqslant 2$. Then an "evident" extension of (3.11) to $\mathbb{R}^q \times \mathbb{R}^q$ is

$$Q(\mathbf{x}; \mathbf{y}) = c_0 + c_1 \sum_{\alpha=1}^{q} x_\alpha y_\alpha = c_0 + c_1 \mathbf{x} \cdot \mathbf{y} \tag{3.13}$$

and we have regained the *Hopfield model* ($c_1 > 0$) without ($c_0 = 0$) or in combination with ($c_0 < 0$) an antiferromagnetic background. Another interesting case, to which we now turn, is provided by $Q(\mathbf{x}; \mathbf{y}) = \phi(\mathbf{x} \cdot \mathbf{y})$ for some function $\phi(x)$ of a single variable x, as in (1.3).

[6] Mathematically, the characters constitute a complete set of eigenfunctions for the convolution kernel $k(\mathbf{x} \circ \mathbf{y}^{-1})$ on a compact Abelian group with Haar measure μ. Here μ is the normalized counting measure.

3.2. Properties of Eigenvalues

Scalar product models *à la* (1.3) allow for a particularly simple and explicit representation of the eigenvalues λ_ρ. A scalar product model is defined by the condition that the synaptic kernel $Q(\mathbf{x};\mathbf{y})$ be of the form

$$Q(\mathbf{x};\mathbf{y}) = \phi(\mathbf{x}\cdot\mathbf{y}) \tag{3.14}$$

for some function ϕ. In this case Eq. (3.9) may be written

$$\begin{aligned}\lambda_\rho &= \sum_{\mathbf{x}} \phi(\mathbf{e}\cdot\mathbf{x})\, v_\rho(\mathbf{x}) \\ &= \sum_{\mathbf{x}} \phi\left(\sum_{j=1}^{q} x_j\right) \prod_{i\in\rho} x_i\end{aligned} \tag{3.15}$$

Here we have used (3.5) and the definition of $\mathbf{e} = (1, 1, \ldots, 1)$. Let us now take an arbitrary vector $\mathbf{x}$ from $\mathscr{C}^q$ and, given the index set ρ, let us denote by k (or l) the number of negative components of $\mathbf{x}$ with label inside (or outside) ρ. Then we get

$$\sum_{j=1}^{q} x_j = q - 2(k+l), \qquad \prod_{i\in\rho} x_i = (-1)^k \tag{3.16}$$

and thus, by a simple combinatorial argument,

$$\lambda_\rho = \sum_{k=0}^{|\rho|} \sum_{l=0}^{q-|\rho|} (-1)^k \binom{|\rho|}{k} \binom{q-|\rho|}{l} \phi(q-2(k+l)) \tag{3.17}$$

Hence λ_ρ only depends on the *size* $|\rho|$ of the set ρ. This entails that, except for possible accidental degeneracies (see below), the multiplicity of $\lambda_\rho = \lambda_{|\rho|}$ is

$$\binom{q}{|\rho|}$$

which is a property we will use repeatedly.

As is evident from the explicit representation (3.17), the eigenvalues λ_ρ also depend on q. Let us denote this dependence by $\lambda_\rho^{(q)}$. In Appendix A it is shown that

$$\lambda_{|\rho|+1}^{(q)} = \lambda_{|\rho|-1}^{(q)} - 4\lambda_{|\rho|-1}^{(q-2)} \tag{3.18}$$

whatever the synaptic function ϕ. The only proviso is that ϕ should not depend on q.

Before turning to clipped synapses, we illustrate the use of (3.17) for three particular cases. First we take

$$\phi(x) = J \sinh(\zeta x) \tag{3.19}$$

where J and ζ will be specified shortly. Expressing $\sinh(x)$ in terms of $\exp(\pm x)$, one easily can perform the sum in (3.17) so as to get

$$\lambda_\rho = \lambda_{|\rho|}^{(q)} = 2^{q-1} J \cosh^q(\zeta)[\tanh^{|\rho|}(\zeta) - \tanh^{|\rho|}(-\zeta)] \tag{3.20}$$

The largest eigenvalue is λ_1. It is q-fold degenerate. For even $|\rho|$ the λ_ρ vanish, as they should by the parity argument associated with (3.10). We also see that if we want to fix $T_c = 2^{-q}\lambda_1^{(q)}$ for large q, we have to rescale J and ζ by putting $J \to J\sqrt{q}$ and $\zeta \to \zeta/\sqrt{q}$. Then

$$2^{-q}\lambda_1^{(q)} = J \cosh^q(\zeta/\sqrt{q})\,[\sqrt{q}\tanh(\zeta/\sqrt{q})] \to J\zeta \exp(\tfrac{1}{2}\zeta) \tag{3.21}$$

as q becomes large. The other (nonzero) $2^{-q}\lambda_\rho^{(q)}$, with $|\rho| \neq 1$, converge to zero. This behavior is generic for scalar product models.[10]

The second example is provided by an even function,

$$\phi(x) = J \cosh(\zeta x) \tag{3.22}$$

Now

$$\lambda_\rho = \lambda_{|\rho|}^{(q)} = 2^{q-1} \cosh^q(\zeta)\,[\tanh^{|\rho|}(\zeta) + \tanh^{|\rho|}(-\zeta)] \tag{3.23}$$

For odd $|\rho|$ the λ_ρ vanish, and the same scaling as before applies. The largest eigenvalue is λ_0. It is nondegenerate. The corresponding microscopic states are ferromagnetic and hence not suitable for storing patterns with magnetization zero.

Finally, we turn to the Hopfield model. For this linear model, $Q(\mathbf{x};\mathbf{y}) = \mathbf{x}\cdot\mathbf{y}$ and there are at least three ways, direct and indirect, to obtain the eigenvalues and the corresponding eigenvectors of the synaptic kernel Q. First, we may use (3.17). It is, however, simpler and also more instructive to return to (3.9). One then easily verifies that $\lambda_\rho = 1$ for all sets ρ with one element, i.e., $|\rho| = 1$, and $\lambda_\rho = 0$ for $|\rho| \neq 1$. The very same result also follows from (3.13) with $c_0 = 0$. In summary, for the Hopfield model the *only* nonzero eigenvalue is $\lambda_1 = 1$ and it is q-fold degenerate. As we will see later,[18] the q eigenvectors x_α, $1 \leqslant \alpha \leqslant q$, correspond to the q stored patterns.

3.3. Clipped Synapses

Clipping means that $\phi(x) = \mathrm{sgn}(x)$ or, equivalently,

$$\phi(x) = -1 + 2\Theta(x) \tag{3.24}$$

where $\Theta(x)$ is the Heaviside function: $\Theta(x) = 0$ for $x < 0$, $\Theta(0) = 1/2$ and $\Theta(x) = 1$ for $x > 0$. Inserting (3.24) into (3.17), we obtain

$$\lambda_\rho = \lambda^{(q)}_{|\rho|} = -2^q \delta_{|\rho|,0} + 2 \sum_{k=0}^{|\rho|} \sum_{l=0}^{q-|\rho|} (-1)^k \binom{|\rho|}{k} \binom{q-|\rho|}{l} \Theta(q-2(k+l)) \tag{3.25}$$

If q is odd, the argument of the Heaviside function in (3.25) never vanishes. If q is even, however, it does and this complicates to some extent the evaluation of the double sum in (3.25), since $\Theta(0) = 1/2$. The details are given in Appendix B. Due to parity, only the $\lambda^{(q)}_\rho$ with $|\rho|$ *odd* are nonzero.

If q is odd, we find

$$\lambda^{(q)}_{|\rho|} = 2 \sum_{l=0}^{|\rho|-1} (-1)^l \binom{|\rho|-1}{l} \binom{q-|\rho|}{\frac{1}{2}(q-1)-l}, \qquad 1 \leqslant |\rho| \leqslant \tfrac{1}{2}(q+1) \tag{3.26}$$

while the remaining eigenvalues follow from

$$\lambda^{(q)}_{|\rho|} = (-1)^{(q-1)/2} \lambda^{(q)}_{q+1-|\rho|}, \qquad \tfrac{1}{2}(q+1) < |\rho| \leqslant q \tag{3.27}$$

If q is even, then

$$\lambda^{(q)}_{|\rho|} = \sum_{l=0}^{|\rho|-1} (-1)^l \binom{|\rho|-1}{l} \binom{q+1-|\rho|}{\frac{1}{2}q-l}, \qquad 1 \leqslant |\rho| \leqslant \tfrac{1}{2}(q+2) \tag{3.28}$$

while the "mirror terms" are given by

$$\lambda^{(q)}_{|\rho|} = (-1)^{q/2} \lambda^{(q)}_{q+2-|\rho|}, \qquad \tfrac{1}{2}(q+2) < |\rho| \leqslant q-1 \tag{3.29}$$

For not too large q, there is an alternative, rather efficient way of calculating the eigenvalues. The $\lambda^{(q)}_1$ can be obtained directly through (3.26) and (3.28) or also (3.17). The result is

$$\lambda^{(q)}_1 = \begin{cases} 2\binom{q-1}{\frac{1}{2}(q-1)}, & \text{if } q \text{ is odd} \\ 2\binom{q-1}{\frac{1}{2}(q-2)}, & \text{if } q \text{ is even} \end{cases} \tag{3.30}$$

Starting with $\lambda_1^{(q)}$, one can evaluate the other $\lambda_{|\rho|}^{(q)}$ by using the recursion relation (3.18) and proceeding via a "Pascal triangle"; cf. Table I.

The $\lambda_n^{(q)}$ vanish for even n. For n odd, their absolute value monotonically decreases with n as long as $n \leqslant \frac{1}{2}(q-1)$ for odd q or $n \leqslant \frac{1}{2}(q-2)$ for even q; see Table I. Since $\lambda_1^{(q)}$ is positive, it is the *largest* eigenvalue. Its multiplicity is q, except for $q = 4k+1$, where we get $q+1$ due to an accidental degeneracy: $\lambda_1^{(4k+1)} = \lambda_{4k+1}^{(4k+1)}$. This follows from (3.27) and is illustrated by Table I. The underlying physics is discussed in Section 3 of the following paper.[18]

For future purposes we also have to calculate the second largest eigenvalue, which was called $\tilde{\lambda}$ in the Introduction. Turning to Table I once again, we see that $\tilde{\lambda}$ may be found either at the low-$|\rho|$ or at the high-$|\rho|$ end, or at both ends, of the relevant branch of the tree in the table. The fraction $\tilde{\lambda}/\lambda_1$ is most easily calculated by using (3.18) and (3.30) once or twice. This then gives, provided $q \geqslant 5$,

$$\lambda_5^{(q)} = \lambda_{q-4}^{(q)} = \frac{3}{(q-2)(q-4)} \lambda_1^{(q)} \qquad \text{if} \quad q = 4k+1 \tag{3.31}$$

$$\lambda_{q-2}^{(q)} = \frac{1}{q-2} \lambda_1^{(q)} \qquad \text{if} \quad q = 4k+3 \tag{3.32}$$

and

$$\lambda_5^{(q)} = \lambda_{q-3}^{(q)} = \frac{3}{(q-1)(q-3)} \lambda_1^{(q)} \qquad \text{if} \quad q = 4k \tag{3.33}$$

$$\lambda_{q-1}^{(q)} = \frac{1}{q-1} \lambda_1^{(q)} \qquad \text{if} \quad q = 4k+2 \tag{3.34}$$

For small q we refer to Table I. The degeneracy corresponding to (3.31)–(3.34) is

$$\binom{q}{5} + \binom{q}{4}, \quad \binom{q}{2}, \quad \binom{q}{5} + \binom{q}{3}, \quad \binom{q}{1} \tag{3.35}$$

respectively. For $q = 4k$ and $4k+1$ we have an accidental degeneracy and $\tilde{\lambda}/\lambda_{\max} \propto q^{-2}$ for large q, whereas for $q = 4k+2$ and $4k+3$ we have no accidental degeneracy and $\tilde{\lambda}/\lambda_{\max} \propto q^{-1}$.

In summary, for $|\rho| \neq 1$,

$$\lambda_{|\rho|}^{(q)}/\lambda_1^{(q)} \to 0 \qquad \text{as} \quad q \to \infty \tag{3.36}$$

The only, trivial, exception is provided by $|\rho| = q = 4k+1$. The property (3.36) will turn out to be instrumental in proving[18] that for an extensive

Table I. Eigenvalues $\lambda_n^{(q)}$ for Clipped Synapses[a]

								q							
							2	1							
								2	2						
						−4	4	3							
								4	6	−2					
					12	−4	12	5			−4 ×				
								6	20	−4	→ 4				
				−40	8	−8	40	7							
								8	70	−10	6	−10			
			140	−20	12	−20	140	9							
								10	252	−28	12	−12	28		
		−504	56	−24	24	−56	504	11							
								12	924	−84	28	−20	28	−84	
	1848	−168	56	−40	56	−168	1848	13							
								14	3432	−264	72	−40	40	−72	264
−6864	528	−144	80	−80	144	−528	6864	15							
								16	12870	−858	198	−90	70	−90	198	−858
15	13	11	9	7	5	3	1	n	1	3	5	7	9	11	13	15

[a] The middle column gives the numbers q of embedded patterns, the bottom row the values of the index $n = |\rho|$, i.e., the cardinalities of the sets $\rho \subseteq \{1,..., q\}$. The left and right triangles correspond to odd and even q, respectively. The working of the recursion relation (3.18) has been indicated once for even q.

temperature range below T_c and for q large the performance of the clipped synapses reduces to that of the (linear) Hopfield model with the same number of patterns, despite the clipping. Furthermore, it will be shown[18] that the q stored patterns are among the states that bifurcate from zero at $T_c = 2^{-q}\lambda_1$. Other, new states branch off at a lower temperature $\tilde{T}_q$, with $\tilde{T}_q/T_c = \tilde{\lambda}/\lambda_1$, where $\tilde{\lambda}$ is the second largest eigenvalue. According to (3.31)–(3.34), the fraction $\tilde{T}_q/T_c$ becomes negligibly small as q gets large.

4. GAUSSIAN DISTRIBUTION

The fixed-point equation (2.20) and the expression for the free energy (2.21) hold for an arbitrary probability distribution μ of the independent random vectors ξ_i. In the previous section we concentrated on the case where the components $\xi_{i\alpha}$, $1 \leqslant \alpha \leqslant q$, are also independent, assuming the values ± 1 with equal probability (bimodal distribution). In this section we exactly solve the spectral problem associated with (2.20) for a Gaussian distribution of the $\xi_{i\alpha}$. Moreover, we take the synaptic function ϕ to be clipped, i.e., $\phi(x) = \operatorname{sgn}(x)$.

We want to determine the eigenvalues and eigenfunctions of the integral operator Q defined by

$$(Q\psi)(\mathbf{x}) = \int_{\mathbb{R}^q} d\mu(\mathbf{y}) \operatorname{sgn}(\mathbf{x}\cdot\mathbf{y})\, \psi(\mathbf{y}) \tag{4.1}$$

where

$$d\mu(\mathbf{y}) = \prod_{\alpha=1}^{q} \frac{1}{\sqrt{2\pi}} \exp\frac{-y_\alpha^2}{2}\, dy_\alpha = (2\pi)^{-q/2} \exp(-\tfrac{1}{2}\mathbf{y}^2)\, d^q y \tag{4.2}$$

is the *rotational*-invariant, Gaussian measure on $\mathbb{R}^q$ with mean zero and total mass one. Q operates on all functions that are square integrable with respect to μ [i.e., $L^2_\mu(\mathbb{R}^q)$].

To solve the eigenvalue problem $Q\psi = \lambda\psi$, we note that $\operatorname{sgn}(\mathbf{x}\cdot\mathbf{y}) = \operatorname{sgn}(\hat{\mathbf{x}}\cdot\hat{\mathbf{y}})$, where $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are unit vectors parallel to $\mathbf{x}$ and $\mathbf{y}$. In view of future applications we are only interested in eigenfunctions belonging to *non*zero eigenvalues. Then $\psi(\mathbf{x}) = \lambda^{-1}(Q\psi)(\mathbf{x})$ only depends on $\hat{\mathbf{x}}$, the direction of the vector $\mathbf{x}$, and we can reduce the problem to the surface of the unit sphere in $\mathbb{R}^q$. That is, we may rewrite $(Q\psi)(\mathbf{x}) = \lambda\psi(\mathbf{x})$ in the form

$$\lambda\psi(\hat{\mathbf{x}}) = \left[\int_0^\infty dR\, (2\pi)^{-q/2}\, e^{-R^2/2} R^{q-1}\right] \int d\Omega(\hat{\mathbf{y}}) \operatorname{sgn}(\hat{\mathbf{x}}\cdot\hat{\mathbf{y}})\, \psi(\hat{\mathbf{y}}) \tag{4.3}$$

Here $d\Omega(\hat{\mathbf{y}})$ denotes the Lebesgue measure on the unit sphere and the expression between square brackets equals[26]

$$\tfrac{1}{2}\Gamma(q/2)\,\pi^{-q/2} \equiv \omega_q^{-1} \tag{4.4}$$

Since ω_q is the total area[26] of the unit sphere, $\omega_q^{-1}\, d\Omega(\hat{\mathbf{y}})$ is normalized to one, as was the original measure μ.

By virtue of (4.3) we are left with the spectral problem $\hat{Q}\psi = \hat{\lambda}\psi$ associated with the operator

$$(\hat{Q}\psi)(\hat{\mathbf{x}}) = \int d\Omega(\hat{\mathbf{y}})\, \mathrm{sgn}(\hat{\mathbf{x}}\cdot\hat{\mathbf{y}})\, \psi(\hat{\mathbf{y}}) \tag{4.5}$$

on the unit sphere. The eigenvalues λ of the original problem are related to $\hat{\lambda}$ by $\lambda = \hat{\lambda}\omega_q^{-1}$. The solution to (4.5) is given by the *Funk–Hecke theorem.*[26] For a given $p = q-2$, let $S_n(\hat{\mathbf{y}})$ be any of the[26]

$$h(n, p) = (2n+p)\,\frac{(n+p-1)!}{p!\,n!} \tag{4.6}$$

surface harmonics of degree n. These are linearly independent harmonic polynomials of degree n of the $q = p+2$ variables $x_1, x_2, \ldots, x_q$. Then the content of the Funk–Hecke theorem is that

$$\int d\Omega(\hat{\mathbf{y}})\, \mathrm{sgn}(\hat{\mathbf{x}}\cdot\hat{\mathbf{y}})\, S_n(\hat{\mathbf{y}}) = \hat{\lambda}_n S_n(\hat{\mathbf{x}}) \tag{4.7}$$

where

$$\hat{\lambda}_n = \frac{\omega_{p+1}}{C_n^{p/2}(1)} \int_{-1}^{1} dx\, \mathrm{sgn}(x)\, C_n^{p/2}(x)(1-x^2)^{p/2-1/2} \tag{4.8}$$

and $C_n^{p/2}(x)$ is a *Gegenbauer* polynomial[26] of degree n and order $p/2$. Since[26]

$$C_n^{\nu}(-x) = (-1)^n\, C_n^{\nu}(x) \tag{4.9}$$

$\hat{\lambda}_n$ vanishes for $n =$ even. For odd n we proceed as follows.

For $\nu \neq 0$ (here $\nu = \frac{1}{2}p$) there exists a useful relation[26]

$$C_n^{\nu}(x) = (-2)^{-n}(1-x^2)^{-\nu+1/2}\,\frac{(2\nu)_n}{(\nu+1/2)_n\, n!}\,\frac{d^n}{dx^n}\,(1-x^2)^{n+\nu-1/2} \tag{4.10}$$

with

$$(a)_0 = 1, \qquad (a)_n = a(a+1)\cdots(a+n-1), \qquad n = 1, 2, \ldots \tag{4.11}$$

Using (4.10), we now can do the integration in (4.8). After some algebra we get

$$\hat{\lambda}_n = \frac{\omega_{p+1}}{C_n^{p/2}(1)} \frac{p(p+2n+1)}{(p+n)(p+n+1)} \frac{1}{n} C_{n-1}^{q/2}(0) \tag{4.12}$$

If $n = \text{even}$, $C_{n-1}^{\nu}(0)$ vanishes, whereas for $n = 2s+1$ (odd) we find

$$C_{2s}^{\nu}(0) = (-1)^s \frac{\Gamma(\nu+s)}{\Gamma(s+1)\,\Gamma(\nu)} \tag{4.13}$$

Note the alternating sign.

Returning to the original eigenvalue problem, we then obtain, if $n = 1$ ($s = 0$),

$$\lambda_1 = \frac{\omega_{q-1}}{\omega_q} \frac{p(p+3)}{(p+1)(p+2)} [C_1^{p/2}(1)]^{-1} \tag{4.14}$$

where $C_1^{p/2}(1) = p = q-2$. Moreover, as $q \to \infty$,

$$\omega_{q-1}/\omega_q \sim (q/2\pi)^{1/2} \tag{4.15}$$

so that for large q we end up with

$$\lambda_1 \sim \frac{1}{(2\pi)^{1/2}} \left(\frac{q+1}{q-1}\right) \frac{1}{q^{1/2}} \tag{4.16}$$

To fix T_c (see below), one therefore has to rescale J in (1.4) by putting $J \to q^{1/2}J$. According to (4.6), there are q independent surface harmonics of degree 1.

The eigenvalue λ_3 is negative, but the ratio λ_5/λ_1 is positive and easily determined to be, for large q,

$$\frac{\lambda_5}{\lambda_1} \sim \frac{3(p+11)(p+2)^2}{(p+3)^2(p+4)(p+5)(p+6)} \tag{4.17}$$

so that

$$\lambda_5/\lambda_1 \propto q^{-2} \tag{4.18}$$

as $q \to \infty$, precisely as in Section 3.3, Eqs. (3.31) and (3.33). Moreover, one can show that $\lambda_{2s+1}/\lambda_1 \propto (-1)^s q^{-s}$, so that λ_5 is the second largest eigenvalue. At $T_c = \lambda_1 J$ the q original patterns bifurcate from zero and at a lower temperature $\tilde{T}_q = \lambda_5 J$ other solutions, corresponding to λ_5, branch off.[18] Due to (4.18), we have $\tilde{T}_q/T_c = \lambda_5/\lambda_1 \propto q^{-2}$ as q becomes large.

The integral kernel $\operatorname{sgn}(\hat{\mathbf{x}} \cdot \hat{\mathbf{y}})$ is invariant under the full orthogonal group in $\mathbb{R}^q$. At T_c, determined by $\beta_c J \lambda_1 = 1$, this symmetry is broken, nontrivial solutions to (2.20) bifurcate away from zero, and we are left with the subgroup that leaves invariant a surface harmonic of degree 1. As already noted, there are q of them. Each is oriented along a coordinate axis. For instance, if $q = 3$, then one easily verifies, by taking suitable linear combinations of $Y_{\pm 1,1}$ and $Y_{0,1}$, that x_1/r, x_2/r, and x_3/r are the three surface harmonics of degree 1.

In contrast to the bimodal distribution of Section 3, a scalar multiple ($\neq 0$) of a surface harmonic is *not* a solution to the fixed-point equation (2.20). At T_c we only get a bifurcation into the *direction* of one of the eigenfunctions. Because they can be transformed into each other by the orthogonal group, we get in fact a whole "umbrella" of solutions that bifurcate from zero. As the temperature is lowered and $\beta \to \infty$, the solutions that started as a surface harmonic converge to

$$m(\mathbf{x}) = \operatorname{sgn}(\mathbf{x} \cdot \hat{\mathbf{e}}_\alpha), \qquad 1 \leqslant \alpha \leqslant q \tag{4.19}$$

where $\hat{\mathbf{e}}_\alpha$ is the unit vector in the Cartesian α direction. These correspond to the microscopic spin configurations

$$S(i) = \operatorname{sgn}(\xi_{i\alpha}), \qquad 1 \leqslant \alpha \leqslant q \tag{4.20}$$

See Section 3, in particular Eq. (3.8), of the following paper.[18]

Because of the rotational invariance of the Gaussian distribution and therefore of the free energy functional (2.21), patterns may be transformed continuously into each other at hardly any cost of (free) energy. So there are no large free energy barriers between the ergodic components. The rotation we referred to is a global transformation, but so is the destruction of a pattern by noise. It therefore turns out that the Gaussian model is not suitable for storing data. In spite of that, it is extremely convenient to illustrate some general features of a neural network with clipped synapses and independent (unbiased) random patterns:

1. There is a critical temperature T_c proportional to the maximal eigenvalue λ_1. Hence $T_c \propto q^{-1/2}$. To fix T_c, we have to rescale J [cf. (1.3)] by putting $J \to q^{1/2} J$.

2. As $q \to \infty$, there exists a very large temperature range $\tilde{T}_q < T < T_c$ where the original patterns are stable and no other metastable states have appeared yet, except for the ones associated with λ_1. Rescaling J does not alter the fraction $\tilde{T}_q / T_c \propto q^{-2}$. For the Gaussian distribution this fraction approaches zero as q^{-2} for all q, whereas for the bimodal distribution the same holds true for every other q; cf. (4.18) and (3.31)–(3.34).

5. DISCUSSION

For finitely many patterns and arbitrary synaptic kernel Q we have determined the equilibrium statistical mechanics which governs the asymptotic behavior of the nonlinear neural network associated with Q. In addition, a complete spectral theory has been given for all Q that satisfy the invariance condition (3.1). Taking advantage of these considerations, we show in the next paper[18] how information can be processed and retrieved in various specific models. The main idea is simple.

The synaptic kernel Q is chosen in such a way that, when a critical temperature T_c associated with its largest eigenvalue is reached, the q stored patterns bifurcate first.[7] Usually, they are the *only* stable states directly below T_c. As the temperature is lowered, more stable states, not quite resembling the original patterns, and with them more basins of attraction appear. These all function as "wastebaskets," into which a noisy or incomplete pattern may disappear as time proceeds. Though not wanted, they cannot be eliminated. In the Hopfield model, which is linear, these spurious states are directly related to the q stored patterns, but, as we will see,[18] a nonlinear model allows much more unwanted states which become stable at low temperatures—despite, or, better, just because of other desirable properties of the model. How can one get rid of this huge amount of unwanted states?

Since we have a Monte Carlo dynamics, the solution is provided by a simple, two-stage procedure. In a temperature window just below T_c there are only q ergodic components,[8] namely the ones associated with the q stored patterns. If we are given a noisy or incomplete pattern, we pick a temperature T in this window and let the Monte Carlo dynamics run for quite a while. Since there are no other basins of attraction than the ones associated with the q original patterns, it is to be expected that with a high probability the system converges to the right ergodic component. There are still too many errors, though, because of the thermal noise inherent to a high temperature. We therefore slowly cool the system down to $T=0$. Because the ergodic components associated with the original patterns remain stable, there is no bifurcation, and, once the system is in the right component, it now *must* converge to the right pattern. That is, we have "recalled a memory" through a simulated annealing procedure which starts just *below* T_c. Of course, we have to pay a price: Slow cooling requires quite a bit of (computer) time. A cheaper way out[18] is possible, however, if one is willing to accept the spurious states associated with the q original

[7] Here it is implicitly assumed that the largest eigenvalue is (at least) q-fold degenerate. This degeneracy holds for all inner-product models; cf. (3.14) and (3.17). Weighted patterns or forgetful memories (Ref. 18, Section 6) do not have this degeneracy, however.

patterns. One then simply chooses a temperature above the one where the states associated with the *second* largest eigenvalue of Q bifurcate from zero. This temperature may be rather low (cf. Section 3.3), so that thermal noise is nearly eliminated.

Do the considerations of Section 2 also apply when q *increases* with N? To answer this question, we return to (2.12) and see that the answer is yes as long as (2.12), i.e., the strong law of large numbers for the size of the sets (2.11), holds. There are $n = 2^q$ sets of about equal size[27] and each of them should contain $\mathcal{O}(N)$ sites. Hence $q \ll \log_2(N)$ will do. If one requires the solution to be exact, this is as far as general theory can go.

APPENDIX A

In this Appendix we prove the recursion relation (3.18),

$$\lambda^{(q)}_{|\rho|+1} = \lambda^{(q)}_{|\rho|-1} - 4\lambda^{(q-2)}_{|\rho|-1} \tag{A.1}$$

for the eigenvalues (3.17),

$$\lambda^{(q)}_{|\rho|} = \sum_{k=0}^{|\rho|} \sum_{l=0}^{q-|\rho|} (-1)^k \binom{|\rho|}{k} \binom{q-|\rho|}{l} \phi(q-2(k+l)) \tag{A.2}$$

of scalar product models with *arbitrary* synaptic function ϕ. As yet, ϕ does not depend on q. To simplify the notation, in what follows we write n for $|\rho|$, and as a rule do not explicitly specify the ranges of the various summations, taking them, as in (A.2), to be extended over those subsets of $\mathbb{Z}$ where the binomial coefficients do not vanish. Then, using the recursion relation

$$\binom{m}{n} = \binom{m-1}{n} + \binom{m-1}{n-1} \tag{A.3}$$

for the binomial coefficients, we obtain

$$\begin{aligned}
\lambda^{(q)}_{n-1} - \lambda^{(q)}_{n+1} = \sum_k \sum_l (-1)^k \binom{n-1}{k} \Bigg\{ & \binom{q-n-1}{l} \\
& + 2\binom{q-n-1}{l-1} + \binom{q-n-1}{l-2} \Bigg\} \phi(q-2(k+l)) \\
- \sum_k \sum_l (-1)^k \Bigg\{ & \binom{n-1}{k} + 2\binom{n-1}{k-1} \\
& + \binom{n-1}{k-2} \Bigg\} \binom{q-n-1}{l} \phi(q-2(k+l))
\end{aligned} \tag{A.4}$$

The sums corresponding to the first terms in the braces in (A.4) cancel and we get, introducing shifted summation variables for the remaining terms,

$$\begin{aligned}\lambda_{n-1}^{(q)} - \lambda_{n+1}^{(q)} = {} & 2\sum_k \sum_{\tilde{l}} (-1)^k \binom{n-1}{k}\binom{q-n-1}{\tilde{l}} \phi(q-2-2(k+\tilde{l})) \\ & + \sum_k \sum_{\tilde{l}} (-1)^k \binom{n-1}{k}\binom{q-n-1}{\tilde{l}} \phi(q-4-2(k+\tilde{l})) \\ & - 2\sum_{\tilde{k}} \sum_l (-1)^{\tilde{k}+1} \binom{n-1}{\tilde{k}}\binom{q-n-1}{l} \phi(q-2-2(\tilde{k}+l)) \\ & - \sum_{\tilde{k}} \sum_l (-1)^{\tilde{k}+2} \binom{n-1}{\tilde{k}}\binom{q-n-1}{l} \phi(q-4-2(\tilde{k}+l))\end{aligned} \tag{A.5}$$

The second and the last sum in (A.5) cancel, the first and the third give equal contributions, so that we are left with

$$\begin{aligned}\lambda_{n-1}^{(q)} - \lambda_{n+1}^{(q)} &= 4\sum_k \sum_l (-1)^k \binom{n-1}{k}\binom{q-2-(n-1)}{l} \\ &\quad \times \phi(q-2-2(k+l)) \\ &= 4\lambda_{n-1}^{(q-2)}\end{aligned} \tag{A.6}$$

which is (A.1).

APPENDIX B

We are going to compute the nonzero eigenvalues of the synaptic kernel for clipped synapses,

$$\begin{aligned}\lambda_n^{(q)} &= 2\sum_{k=0}^{n} \sum_{l=0}^{q-n} (-1)^k \binom{n}{k}\binom{q-n}{l} \Theta(q-2(k+l)) \\ &= 2\sum_{k=0}^{n} \sum_{l=0}^{q-n} (-1)^k \binom{n}{k}\binom{q-n}{l} \Theta(q/2-(k+l)), \qquad n \text{ odd}\end{aligned} \tag{B.1}$$

To simplify the notation, instead of $|\rho|$ we write n. So $0 \leqslant n \leqslant q$. The cases q odd and q even will be treated separately. We first turn to the case of odd q.

1. If q is odd, the argument of the Θ function in (B.1) is never zero, so the Θ function simply imposes the constraint

$$k + l \leqslant (q-1)/2 \tag{B.2}$$

on the range of the k and l summations. Therefore, (B.1) can be rewritten

$$\lambda_n^{(q)} = 2 \sum_{l=0}^{q-n} \sum_{k=0}^{\min\{n,(q-1)/2-l\}} \binom{q-n}{l} (-1)^k \binom{n}{k} \tag{B.3}$$

All terms in (B.3) with $\frac{1}{2}(q-1)-l \geqslant n$ vanish, since the sum over k gives zero. For the remaining terms we also can perform the sum over k and get[28]

$$\lambda_n^{(q)} = 2 \sum_{l=0}^{q-n} \binom{q-n}{l} (-1)^{(q-1)/2-l} \binom{n-1}{\frac{1}{2}(q-1)-l} \tag{B.4}$$

where the second binomial coefficient in (B.4) is zero whenever $\frac{1}{2}(q-1)-l$ exceeds $n-1$ or l exceeds $\frac{1}{2}(q-1)$. If we put $\tilde{l} = \frac{1}{2}(q-1)-l$, (B.4) gives

$$\lambda_n^{(q)} = 2 \sum_{\tilde{l}=0}^{n-1} (-1)^{\tilde{l}} \binom{n-1}{\tilde{l}} \binom{q-n}{\frac{1}{2}(q-1)-\tilde{l}}, \qquad 1 \leqslant n \leqslant \tfrac{1}{2}(q+1) \tag{B.5}$$

while a comparison of (B.4) and (B.5) yields

$$\lambda_n^{(q)} = (-1)^{(q-1)/2} \lambda_{q+1-n}^{(q)}, \qquad \tfrac{1}{2}(q+1) < n \leqslant q \tag{B.6}$$

This completes the discussion of the case where q is odd.

2. If q is even, then

$$\Theta(q/2-(k+l)) = \begin{cases} 1 & \text{if } k+l \leqslant \frac{1}{2}(q-2) \\ 1/2 & \text{if } k+l = \frac{1}{2}q \end{cases} \tag{B.7}$$

Thus extra boundary terms occur in the evaluation of (B.1) when the argument of the Θ function vanishes. With (B.7), Eq. (B.1) gives

$$\begin{aligned} \lambda_n^{(q)} = & \sum_{k=0}^{n} (-1)^k \binom{n}{k} \binom{q-n}{\frac{1}{2}q-k} \\ & + 2 \sum_{l=0}^{q-n} \sum_{k=0}^{\min\{n,(q-2)/2-l\}} \binom{q-n}{l} (-1)^k \binom{n}{k} \end{aligned} \tag{B.8}$$

The first sum in (B.8) comes from the boundary terms, the second from the terms where $k+l \leqslant \frac{1}{2}(q-2)$. In the second sum, the same manipulations can be performed as in (B.3)–(B.5). This gives

$$\begin{aligned} \lambda_n^{(q)} = & \sum_{k=0}^{n} (-1)^k \binom{n}{k} \binom{q-n}{\frac{1}{2}q-k} \\ & + 2 \sum_{\tilde{l}=0}^{n-1} (-1)^{\tilde{l}} \binom{n-1}{\tilde{l}} \binom{q-n}{\frac{1}{2}(q-2)-\tilde{l}}, \qquad 1 \leqslant n \leqslant q/2 \end{aligned} \tag{B.9}$$

We now use the recursion relation (A.3) for $\binom{n}{k}$ in the first sum of (B.9) to get

$$\lambda_n^{(q)} = \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k}\binom{q-k}{\frac{1}{2}q-k} + \sum_{k=1}^{n} (-1)^k \binom{n-1}{k-1}\binom{q-n}{\frac{1}{2}q-k}$$
$$+ 2\sum_{7=0}^{n-1} (-1)^7 \binom{n-1}{7}\binom{q-n}{\frac{1}{2}(q-2)-7}, \qquad 1 \leqslant n \leqslant q/2 \tag{B.10}$$

The second sum in (B.10) is $-1/2$ times the third, so that we end up with (renaming $7 \to k$)

$$\lambda_n^{(q)} = \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k}\left\{\binom{q-n}{\frac{1}{2}q-k} + \binom{q-n}{\frac{1}{2}(q-2)-k}\right\}$$
$$= \sum_{k=0}^{n-1} (-1)^k \binom{n-1}{k}\binom{q+1-n}{\frac{1}{2}q-k}, \qquad 1 \leqslant n \leqslant q/2 \tag{B.11}$$

In the last step we have used (A.3) again. As in case 1, one obtains mirror terms of the form

$$\lambda_n^{(q)} = (-1)^{q/2}\, \lambda_{q+2-n}^{(q)}, \qquad \tfrac{1}{2}(q+2) \leqslant n \leqslant q-1 \tag{B.12}$$

ACKNOWLEDGMENTS

The authors thank E. Pehlke for an illuminating discussion concerning Section 3.1. They also gratefully acknowledge the hospitality of the Institut für Theoretische Physik der Universität Kiel and the Sonderforschungsbereich 123 der Universität Heidelberg during their stays in Kiel (J.L.v.H.) and Heidelberg (D.G., A.H., and R.K.). Part of this work has been supported by the Deutsche Forschungsgemeinschaft (DFG).

REFERENCES

1. J. J. Hopfield, *Proc. Natl. Acad. Sci. USA* **79**:2554 (1982); **81**:3088 (1984).
2. W. A. Little, *Math. Biosci.* **19**:101 (1974); W. A. Little and G. L. Shaw, *Math. Biosci.* **39**:281 (1978).
3. P. Peretto, *Biol. Cybernet.* **50**:51 (1984).
4. G. Toulouse, S. Dehaene, and J.-P. Changeux, *Proc. Natl. Acad. Sci. USA* **83**:1695 (1986).
5. J. J. Hopfield and D. W. Tank, *Biol. Cybernet.* **52**:141 (1985).
6. W. A. McCulloch and W. Pitts, *Math. Biophys.* **5**:115 (1943).
7. K. Binder, in *Monte Carlo Methods in Statistical Physics*, by K. Binder, ed. (Springer, New York, 1979), p. 145.
8. A. C. D. van Enter and J. L. van Hemmen, *Phys. Rev. A* **29**:355 (1984).
9. D. O. Hebb, *The Organization of Behavior* (Wiley, New York, 1949).

10. J. L. van Hemmen and R. Kühn, *Phys. Rev. Lett.* **57**:913 (1986); J. L. van Hemmen, *Phys. Rev. A* **36**:1959 (1987).
11. H. Sompolinsky, *Phys. Rev. A* **34**:2571 (1986).
12. D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Phys. Rev. Lett.* **55**:1530 (1985); *Ann. Phys.* **173**:30 (1987).
13. D. J. Amit, H. Gutfreund, and H. Sompolinsky, *Phys. Rev. A* **32**:1007 (1985).
14. W. Kinzel, *Z. Phys. B* **60**:205 (1985).
15. J. J. Hopfield, in *Modelling and Analysis in Biomedicine*, C. Nicolini, ed. (World Scientific, Singapore, 1984), pp. 369–389, in particular p. 381.
16. G. Parisi, *J. Phys. A: Math. Gen.* **19**:L617 (1986).
17. J. P. Nadal, G. Toulouse, J.-P. Changeux, and S. Dehaene, *Europhys. Lett.* **1**:535 (1986).
18. J. L. van Hemmen, D. Grensing, A. Huber, and R. Kühn, *J. Stat. Phys.*, this issue, following paper.
19. J. L. van Hemmen, D. Grensing, A. Huber, and R. Kühn, *Z. Phys. B* **65**:53 (1986).
20. H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Oxford University Press, Oxford, 1971), Section 6.5.
21. J. L. van Hemmen, *Phys. Rev. Lett.* **49**:409 (1982); in *Lecture Notes in Physics*, No. 192 (1983), pp. 203–233, in particular the Appendix.
22. D. Grensing and R. Kühn, *J. Phys. A: Math. Gen.* **19**:L1153 (1986).
23. J. Lamperti, *Probability* (Benjamin, New York, 1966), Section 7.
24. R. B. Griffiths, Chi-Yuan Weng, and J. S. Langer, *Phys. Rev.* **149**:301 (1966).
25. J. L. van Hemmen, *Phys. Rev. A* **34**:3435 (1986).
26. A. Erdélyi *et al.*, *Higher Transcendental Functions*, Vol. 2 (McGraw-Hill, New York, 1953), Chapter XI, in particular pp. 232–251.
27. J. L. van Hemmen and A. C. D. van Enter, *Phys. Rev. A* **34**:2509 (1986), in particular the Appendix.
28. M. Abramowitz and I. A. Stegun (eds.), *Handbook of Mathematical Functions* (Dover, New York, 1965), Section 24.1.1.