#1749

CONTRACTIVE NONLINEAR NEURAL NETWORKS: STABILITY AND SPATIAL FREQUENCIES

Douglas G. Kelly
Departments of Mathematics and Statistics
University of North Carolina
Chapel Hill, NC27599

January 1988

ABSTRACT

We consider models of the form $\mu\dot{x} = -x + p + WF(x)$, where $x = x(t)$ is a vector whose entries represent the electrical activities in the units of a neural network. W is a matrix of synaptic weights, F is a nonlinear function, and p is a vector (constant over time) of inputs to the units. If the map $WF(x)$ is a contraction, then the system has a unique equilibrium which is globally asymptotically stable; consequently the network's steady-state response to any constant input is independent of the initial state of the network. We consider also some relatively mild restrictions on W and $F(x)$, involving the eigenvalues of W and the derivative of F, that are sufficient to ensure that $WF(x)$ is a contraction. We show that in the case of spatially-homogeneous synaptic weights, the eigenvalues of W are simply related to the Fourier transform of the connection pattern. This relation makes it possible, given cortical activity patterns as measured by autoradiographic labeling, to construct a pattern of synaptic weights which produces steady state patterns showing similar frequency characteristics. Finally we bound the norm of the difference between the equilibrium of the model and that of the simpler linear model $\mu\dot{x} = -x + p + Wx$; this latter equilibrium can be computed simply from p in the homogeneous case using Fourier transforms.

This paper represents research supported by NIDR Program Grant DE-07509 (B. L. Whitsel, Principal Investigator).

Introduction.

In 1976 Holden [9] observed that the primary difficulty in the way of attempts to model real neural networks was the lack of data with which to compare models: "The results of these approaches – steady states, limit cycles and propagating waves – seem trivial compared to Sherrington's idea of 'meaningful patterns', and since there is so little relevant experimental evidence these is a danger that further development of these approaches might become ... less and less relevant to neurophysiology. In the problem of obtaining the behaviour of a net the more playsible the model, the less tractable the system of equations ... ."

Recent experimental techniques, however, most especially autoradiographic labeling (see [9] and [12]), have made possible observations relating to the electrical activity patterns thought to carry information in neocortex. In particular, the appearance in cortical functional activity patterns of spatial frequencies not present in anatomical observations has been noted [13].

In this paper we consider a relatively simple, though general, class of anatomically-based nonlinear neural network models, containing excitatory and inhibitory synaptic connections of arbitrary reach whose strengths do not change with time, and receiving input of arbitrary form from outside the system. We show that under the assumption of *contractiveness* (see Section 5 below) such systems respond to input by quickly reaching a steady state which uniquely encodes the input and is independent of the initial state. We show also that under the further assumption that the network is spatially homogeneous, the encoding of the input is equivalent to the action of a filter that enhances certain frequencies characteristic of the network. In addition, we demonstrate that it is possible to construct a synaptic weight pattern whose output spatial frequencies will match those observed to be present in the activity pattern produced by a real network (the forepaw area of S-I in cat) in response to repeated stimulation.

The general model we consider is essentially like those considered by Cowan and others (see [5] and [14]). It consists of N units which represent aggregates of neurons; for the purposes of comparison to the experimental data we view the units as cortical "minicolumns", which are approximately 35 μ (one neuron) in diameter and extend radially through the cortex from upper to lower surface. Associated with each unit, say unit j, is a time-varying quantity $x_j(t)$ representing the average electrical activity of the neurons in the unit over a short time interval containing time t. With each pair (j,k) of units is a synaptic weight W_{jk} which does not change with time; it represents an average strength of synaptic transmission from neurons in unit k to neurons in unit j. In realizations of the model the units are viewed as arranged on a two-dimensional grid or in a one-dimensional array, and the synaptic strengths may be near zero for units that are distant from each other. Synaptic transmission is nonlinear: activity in unit k has an effect on that in unit j proportional to $W_{jk}f_k(t)$, where f_k is a smooth sigmoid function. Input is constant over time but otherwise arbitrary: each unit, say unit j, receives input from outside the system that tends to change its activity by a fixed amount p_j per unit time, and the quantities p_j can form any pattern whatever, with no restrictions as to magnitude.

These assumptions result in the system (1.1) of differential equations governing the behavior of the network. Our results are as follows:

1. Under certain assumptions (primarily contractiveness; see Sections 2 and 5) involving the eigenvalues of the matrix $W = \{W_{jk}\}$ and the derivatives of the functions f_k, the activities $x_1(t)$, $x_2(t)$, ... , $x_N(t)$ converge at an exponential rate to steady-state activities $\bar{x}_1$, $\bar{x}_2$, ... , $\bar{x}_N$ that depend only on the inputs p_j and not on the initial conditions $x_j(0)$.

2. Under the assumption of spatial homogeneity of synaptic weights (that is, that W_{jk} is the same as W_{rs} if the direction and distance of unit k from unit j are the same as those of unit s from unit r), the conditions required in 1 above can be easily checked, and the relation of the steady state to the input can be approximately computed, using discrete Fourier analysis. In fact, the network under these assumptions acts as a filter, enhancing certain characteristic spatial frequencies present in the input pattern and suppressing others.

3. Discrete Fourier analysis of 2-deoxyglucose autoradiographic data ([9], [12]) shows spatial frequencies other than those predicted by known anatomical considerations. Appropriate choice of synaptic weights in the model can cause it to produce steady states with comparable spatial frequencies.

Acknowledgement. The author is greatly indebted to Dr. Barry Whitsel and his co-workers for generously providing data for analysis, and also for numerous illuminating and helpful conversations.

1. The basic model.

All the neural network models we consider in this paper are special cases of the following model: Units 1, 2, 3, ..., N represent single neurons or aggregates of neurons behaving as processing units. For each unit j, $x_j(t)$ represents the average membrane potential in unit j in a small time interval ending at time t. The quantities $x_j(t)$ change with time according to the differential equations

$$\mu x_j{}'(t) = -x_j(t) + p_j + \sum_{k=1}^{N} W_{jk} f_k(x_k(t)) \qquad (j=1,2,\ldots,N). \tag{1.1}$$

Here $f_k(x_k)$ is a function which converts average membrane potential into average firing rate for unit k; different units may have different conversion functions. W_{jk} is a constant representing the synaptic strength from unit k to unit j; multiplying by W_{jk} converts a firing rate in unit k to the resulting rate of increase or decrease in average membrane potential in unit j. Positive values of W_{jk} reflect excitatory synapses (or, more generally, units k whose net average effect on unit j is excitatory), and negative W_{jk} correspond to inhibition. p_j represents (constant) input to unit j at time t from outside the network, in the form of a rate of increase in average membrane potential. Finally, μ is a time constant governing the rate of change.

In one possible realization of this general model, the units are viewed as aggregates of neurons, say the 350-micron cortical columns widely thought to act as units of cortical information processing (see [11]). In such a realization an appropriate choice of $f_k(x_k)$ is a "smooth sigmoid" (see [14] and [5]), and it is no serious restriction to assume it is nondecreasing with a bounded derivative. In another possible realization, units represent single neurons; it may be desirable in such a case to build thresholds into the functions f_k (see [1], [2], [8]). In an intermediate case the units represent "minicolumns"; this produces a model appropriate to represent a small cylindrical patch of cortical tissue, say 450μ in diameter. See Section 3c below.

Because the quantities p_j are constant over time, the results we derive can be viewed as applicable to the short-term processing of input stimuli by primary sensory cortical regions, or to the processing of repeated input stimuli, such as that encountered in radiographic labeling experiments.

The model described above can be naturally incorporated into a more general model of neural network plasticity; by allowing the weights W_{jk} to change with time according to differential equations which reflect appropriate hypotheses about the effects of pre- and post-synaptic activity on synaptic weights, we can obtain any of a number of learning models which have been studied (see [4], [6], and [10]).

The use of vector-matrix notation results in considerable economy of expression in dealing with this model. Let

$$x = x(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{bmatrix} \quad \text{and} \quad p = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_N \end{bmatrix}$$

be vectors whose j^{th} components represent the average potential and average input, respectively, in unit j at time t. Let

$$W = \begin{bmatrix} W_{11} & \cdots & W_{1N} \\ \vdots & & \vdots \\ W_{N1} & \cdots & W_{NN} \end{bmatrix}$$

be the matrix of synaptic weights; W_{jk} is the entry in the j^{th} row and the k^{th} column. Finally let $F(x)$ be the function from $\mathbb{R}^N$ to $\mathbb{R}^N$ (that is, the n-dimensional function of an n-dimensional vector) defined by

$$F(x) = \begin{bmatrix} f_1(x_1) \\ \vdots \\ f_N(x_N) \end{bmatrix}.$$

Then the system (1.1) reads

$$\mu \dot{x} = -x + p + WF(x), \tag{1.2}$$

where $\dot{x}$ denotes $x'(t)$, the vector of time derivatives of the activity levels $x_j(t)$.

In using this notation we will occasionally need suffices in vector names; we will use superscripts in this case rather than subscripts, reserving subscripts for indicating the components of a vector. Thus we will refer to initial conditions for (1.2) of the form $x(0) = x^0$, and the j^{th} component of x^0 will be denoted x_j^0.

2. The linear case.

Here we consider the special case of the model described above in which $f_j(x_j) = x_j$ for each unit j, so that (1.2) becomes

$$\mu\dot{x} = -x + p + Wx. \tag{2.1}$$

This model is no less general than the one in which $f_j(x_j) = a_j + b_j x_j$, because the constants a_j and b_j can be absorbed into the components of p and W, respectively.

The most obvious advantage of this simplification is its theoretical tractability: as long as the matrix $I-W$ is nonsingular it has a unique equilibrium point and all solution curves are known (see (2.2) and (2.3) below). But despite this simplicity, and despite the general agreement that nonlinearities are necessary features of realistic brain models, the linear case can be used as an illuminating model of the processing of input by primary sensory cortex (see [13]).

If $I-W$ is nonsingular, then the system (2.1) has the unique equilibrium

$$\bar{x} = (I-W)^{-1}p. \tag{2.2}$$

Moreover, for initial conditions $x(0) = x^0$ the solution to (2.1), that is, the vector $x(t)$ of activities at any time t, is given by

$$x(t) = e^{(W-I)\frac{t}{\mu}}x^0 + (W-I)^{-1}e^{(W-I)\frac{t}{\mu}}p + (I-W)^{-1}p. \tag{2.3}$$

If all the eigenvalues of $W-I$ have negative real parts, then the first two terms in (2.3) approach 0 at an exponential rate as $t\to\infty$, and consequently $\lim_{t\to\infty} x(t) = \bar{x}$ regardless of x^0; in other words, the equilibrium $\bar{x}$ is globally asymptotically stable: it has all of $\mathbb{R}^N$ as its domain of attraction. We summarize these elementary results about linear systems:

If no eigenvalue of W equals 1, then there is a unique equilibrium, given by (2.2). If in addition the real parts of all eigenvalues of W are less than 1, then this equilibrium is the limiting state of the system for any initial state.

Another way of viewing the linear system in the stable case is as follows: given a matrix W of synaptic weights for which all eigenvalues have real parts less than 1, the system acts as a function which maps any input vector p in $\mathbf{R}^N$ to a steady state $\bar{x}$ in $\mathbf{R}^N$ (independent of the initial state x^0). This map is linear, and the system can be regarded as a linear filter; this will be more apparent in the examples given in the next section. In Section 5 below we will show that a large subclass of nonlinear models (1.1) also act as filters in a similar fashion.

In case W has the additional property of being diagonable, that is, if W has N linearly independent eigenvectors x^1, x^2, ..., x^N, we can view the operation of the linear system as a linear filter that enhances certain components of the input vector and suppresses others. For it is then possible to write the input vector p as a linear combination of the eigenvectors, say

$$p = \sum_{k=1}^{N} \pi_j x^j ; \tag{2.4}$$

and if λ_1, λ_2, ..., λ_N are the eigenvalues associated with x^1, x^2,...x^N, then the equilibrium can be computed as

$$\bar{x} = (I-W)^{-1} p = \sum_{k=1}^{N} \frac{\pi_j}{1-\lambda_j} x^j . \tag{2.5}$$

In the next section we give a class of models included in (2.1) in which the eigenvectors x^j are spatial sinusoids and the computation of the π_j and λ_j amounts to discrete Fourier analysis.

3. Spatially-homogeneous linear systems.

a. One dimension.

Here we view the units of the linear model (2.1) as arranged in a linear array, with equal spacing between adjacent units. We then impose the additional assumption of ***spatial homogeneity***: that W_{jk} depends only on the relative positions of the units j and k and not on their values. It is convenient for this purpose to number the units 0, 1, ..., $N-1$ instead of 1, 2, ..., N and to consider that unit $N-1$ is adjacent to unit 0. This is conceptually the simplest way of taking care of "end effects"; the artificiality of having the units at either end adjacent to each other is preferable to the alternatives: considering infinitely many units, clamping the activity levels of the end units 0 and $N-1$, or adding additional units with activities tapering to zero. If the number N of units is large, any of these ways of dealing with end effects has a negligible effect on the effectiveness of the model in reflecting real processes.

It is worthwhile to note that spatial homogeneity is not the same assumption as symmetry, nor as isotropy. Symmetry requires that $W_{jk} = W_{kj}$ for all j and k, while isotropy requires that $W_{jk} = W_{jl}$ for any units k and l at the same distance from j, regardless of their direction. We will assume symmetry in addition to spatial homogeneity in some of what follows; we will not have occasion to consider isotropy.

With the two end units adjacent, we are essentially viewing the units as though they were arranged on a circle. Spatial homogeneity then amounts to the assumption that there are constants w_0, w_1, ..., w_{N-1} such that $W_{jk} = w_{k-j}$, where $k-j$ is to be interpreted as the number of counterclockwise steps from unit j to unit k. Notice that spatial homogeneity does not necessarily imply symmetry: w_{k-j} need not equal w_{j-k}. The matrix W in the spatially-homogeneous case is a Toeplitz matrix, that is, a matrix of the form

$$W = \begin{bmatrix} w_0 & w_1 & & & w_{N-1} \\ w_{N-1} & w_0 & w_1 & & \\ & w_{N-1} & w_0 & \ddots & \\ & & \ddots & \ddots & w_1 \\ w_1 & & & w_{N-1} & w_0 \end{bmatrix}.$$

Proposition 1. In this case, for each $j = 0, 1, 2, \ldots, N-1$,

$$x^j = \begin{bmatrix} 1 \\ e^{\frac{2\pi ij}{N}} \\ e^{2\cdot\frac{2\pi ij}{N}} \\ \vdots \\ e^{(N-1)\cdot\frac{2\pi ij}{N}} \end{bmatrix} \tag{3.1}$$

is an eigenvector, and

$$\lambda_j = \sum_{s=0}^{N-1} w_s e^{+\frac{2\pi ijs}{N}} \tag{3.2}$$

is the eigenvalue associated with λ_j.

Proof. The k^{th} entry of Wx^j is

$$(Wx^j)_k = \sum_{m=0}^{N-1} W_{km} e^{\frac{2\pi ijm}{N}} = \sum_{s=0}^{N-1} w_s e^{\frac{2\pi ij(k+s)}{N}} = \lambda_j (x^j)_k,$$

the k^{th} entry of $\lambda_j x^j$. □

Proposition 2. The coefficients π_j in the expansion $p = \sum_{j=0}^{N-1} \pi_j x^j$ are given by

$$\pi_j = \frac{1}{N} \sum_{k=0}^{N-1} p_k e^{-\frac{2\pi ijk}{N}} \qquad (j = 0, 1, 2, \ldots, N-1). \tag{3.3}$$

Proof. The assertion is that (3.3) implies

$$p_k = \sum_{j=0}^{N-1} \pi_j e^{+\frac{2\pi ijk}{N}} \qquad (k = 0, 1, 2, \ldots, N-1). \tag{3.4}$$

But (3.3) and (3.4) are equivalent assertions: the former is the definition of the discrete Fourier transform and the latter is the discrete Fourier inversion formula. □

We note that the eigenvectors x^j of W do not depend on the values of the W_{jk}, but only on the fact that W is a Toeplitz matrix. They are always complex spatial sinusoids:

$$x^j = \begin{bmatrix} 1 \\ \cos(2\pi j/N) \\ \cos(2\cdot 2\pi j/N) \\ \vdots \\ \cos((N\text{-}1)\cdot 2\pi j/N) \end{bmatrix} + i \begin{bmatrix} 1 \\ \sin(2\pi j/N) \\ \sin(2\cdot 2\pi j/N) \\ \vdots \\ \sin((N\text{-}1)\cdot 2\pi j/N) \end{bmatrix}.$$

This is a sinusoid with wave number j; it completes j cycles as we move from unit 0 to unit $N-1$ and thus has a frequency of j/N cycles per unit and a period of N/j units.

The eigenvalues λ_j likewise are not real in general, but they are real if W is a symmetric matrix; that is, if we add the assumption of symmetry ($w_{N\text{-}j} = w_j$) to that of spatial homogeneity. (3.2) shows that the λ_j's constitute the inverse discrete Fourier transform of the w_j's.

In particular, given the weights $w_0, w_1, \ldots, w_{N\text{-}1}$ and the input activity levels $p_0, p_1, \ldots, p_{N\text{-}1}$, one can check stability and find the equilibrium activity levels $\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_{N\text{-}1}$ by the following sequence of steps:

1. Compute $\lambda_0, \lambda_1, \ldots, \lambda_{N\text{-}1}$ using (3.2).
   The system is stable if $\mathrm{Re}(\lambda_j)<1$ for all j.
2. Compute $\pi_0, \pi_1, \ldots, \pi_{N\text{-}1}$ using the discrete Fourier transform (3.3).
3. Compute $\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_{N\text{-}1}$ using

$$\bar{x}_k = \sum_{j=0}^{N\text{-}1} \frac{\pi_j}{1-\lambda_j} e^{+\frac{2\pi ijk}{N}} \quad (k = 0, 1, 2, \ldots, N-1). \tag{3.5}$$

The above discussion, involving the two propositions of this section along with (2.4) and (2.5) in the previous section, amounts to the implementation of the observation that in the stable, spatially-homogeneous linear case the model (insofar as we consider only the steady state $\bar{x}$ as a function of input p) acts by selectively enhancing certain spatial frequencies (those with λ_j near 1) and suppressing others. Figures 3.1 — 3.3 show examples of such systems; the enhancement of spatial frequencies is especially apparent when the input p contains a noise component.

b. Two dimensions.

Except for notational inconveniences the two-dimensional spatially-homogeneous linear model is not different in principle from its one-dimensional counterpart: the eigenvectors ("eigenpatterns") of W correspond to two-dimensional spatial sinusoids, the eigenvalues are the two-dimensional discrete inverse Fourier transform of the pattern of synaptic weights (in analogy with (3.2)); an input pattern p can be expressed as a linear combination of the eigenvectors by the two-dimensional discrete Fourier transform (in analogy with (3.3)); and the equilibrium pattern can be computed by dividing the coefficients in this expression by the numbers $1-\lambda_j$ and applying the inverse Fourier transform.

The units are now arranged on a two-dimensional grid, and we give the units double labels to indicate their positions; thus unit (j,k) is in the j^{th} row and the k^{th} column of the grid, for $j = 0, 1, 2, \ldots, N-1$ and $k = 0, 1, 2, \ldots, M-1$. For convenience in imposing spatial homogeneity we suppose that for each j, the unit $(j,M-1)$ at the end of row j is adjacent to $(j,0)$ and for each k, unit $(N-1,k)$ at the bottom of column k is adjacent to $(0,k)$. This amounts to arranging the units on a torus. The synaptic weights in the general, inhomogeneous case are now quantities with four indices: $W_{j,k,l,m}$ represents the synaptic strength from unit (l,m) to unit (j,k). But the assumption of homogeneity is that this number depends only on the numbers $l-j$ and $m-k$, which again are to be interpreted modulo N or M as being numbers of steps from j to l or from k to m. That is, we assume that there are numbers w_{rs} ($r = 0, 1, 2, \ldots, N$-1, $s = 0; 1, 2, \ldots, M$-1) such that

$$W_{j,k,j+r,k+s} = w_{rs}. \tag{3.6}$$

The vectors $x(t)$ of activity levels and p of inputs are now vectors in $\mathbf{R}^{N\times M}$ whose entries correspond to units on a grid, and it is convenient to represent them as two-dimensional arrays or patterns; thus

$$x(t) = \begin{bmatrix} x_{00}(t) & \cdots & x_{0,N-1}(t) \\ \vdots & & \vdots \\ x_{M-1,0}(t) & \cdots & x_{M-1,N-1}(t) \end{bmatrix} \quad \text{and} \quad p = \begin{bmatrix} p_{00} & \cdots & p_{0,N-1} \\ \vdots & & \vdots \\ p_{M-1,0} & \cdots & p_{M-1,N-1} \end{bmatrix}.$$

The matrix W is now an $MN \times MN$ matrix, whose eigenvectors are vectors of length MN which, like $x(t)$ and p, can best be viewed as two-dimensional patterns (the "characteristic patterns" or "characteristic states" of the system). It can be shown, in analogy with Proposition 1 above, that for each pair (j,k) the array x^{jk} whose (r,s) entry is

$$x^{jk}_{rs} = e^{2\pi i(\frac{jr}{M} + \frac{ks}{N})} \tag{3.7}$$

is an eigenvector, and the associated eigenvalue is

$$\lambda_{jk} = \sum_{r=0}^{M-1} \sum_{s=0}^{N-1} w_{rs}\, e^{+2\pi i(\frac{jr}{M} + \frac{ks}{N})} . \tag{3.8}$$

Moreover, to represent p as a linear combination of the x^{jk} with coefficients π_{jk}, that is, to find π_{jk} such that

$$p_{rs} = \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} \pi_{jk}\, e^{+2\pi i(\frac{jr}{M} + \frac{ks}{N})} , \tag{3.9}$$

one uses

$$\pi_{jk} = \frac{1}{MN} \sum_{r=0}^{M-1} \sum_{s=0}^{N-1} p_{rs}\, e^{-2\pi i(\frac{jr}{M} + \frac{ks}{N})} . \tag{3.10}$$

(3.10) is the definition of the discrete two-dimensional Fourier transform, and (3.9) is two-dimensional discrete Fourier inversion. Thus the steady state $\bar{x}$ can be computed from the weights w_{rs} and the input pattern values p_{rs} using a prescription like that given at the end of a. above, but with (3.2) and (3.3) replaced by (3.8) and (3.10), respectively, and with the (j,k) entry of the steady state pattern $\bar{x}$ computed as

$$\bar{x}_{jk} = \frac{1}{MN} \sum_{r=0}^{M-1} \sum_{s=0}^{N-1} \frac{\pi_{rs}}{1-\lambda_{rs}} e^{-2\pi i(\frac{jr}{M}+\frac{ks}{N})} . \tag{3.11}$$

Figure 3.4 shows an example of a two-dimensional homogeneous weight pattern and its eigenvalues; Figure 3.5 shows the steady-state response of this system to a hypothetical input pattern.

c. Inferring weight patterns from observed activity patterns.

Figure 3.6(a) is an average periodogram showing the spatial periodicities present in a single "patch" of the autoradiographic (2-deoxyglucose) labeling pattern evoked by repeated electrocutaneous stimulation to the forelimb in cat. (This figure is provided through the courtesy of Dr. B. L. Whitsel; see [9].) Each bar in this average periodogram represents the relative power in a frequency range of width 2.2 (actually 20/9) cycles per millimeter; thus, for example, the twelfth bar shows the power at frequencies between 26.67 and 28.89 cycles per millimeter, that is, to periods of 34.62 and 37.50 micrometers.

Some features shown by this average periodogram are: (i) The twelfth bar, corresponding to periods close to the width of a single cortical minicolumn. (ii) The first bar; the high power at the lowest frequencies is attributed to the fact that the patch shows more labeling in its center than at its edges. (iii) The power in the sixth, seventh, eighth, and ninth bars, corresponding to periods in the range 45 – 75 micrometers, or approximately 1.5 to 2.2 minicolumn widths. (iv) The fourth bar, corresponding to periods in the range 112.5 – 150 micrometers, or approximately 3.2 to 4.9 minicolumn widths. Higher frequencies shown by the average periodogram correspond to periods less than the width of a single minicolumn.

To account for the above observations with the one-dimensional linear model, we consider the units of the model to be single minicolumns, and we seek a synaptic weight pattern which will enhance frequencies corresponding to periods around two units and periods around four units. In a model with $n = 81$ units these will be Fourier frequencies around the 40^{th} and around the 21^{st}, respectively. The eigenvalues of such a synaptic-weight pattern are shown in Figure 3.6(b); the curve in that figure was constructed as a sum of two Gaussian curves with peaks at the desired frequencies. Next, we use the inverse of (3.2), namely

$$w_s = \frac{1}{N} \sum_{j=0}^{N-1} \lambda_j e^{-\frac{2\pi ijs}{N}} , \tag{3.12}$$

to find the synaptic weight pattern that produces the desired eigenvalues. This pattern is shown in Figure 3.6(c). Finally, as an example, Figure 3.6(d) shows the steady-state pattern produced by these weights from the noisy input pattern of Figure 3.3(a). Figure 3.6(e) is the periodogram of this steady state. It shows power at Fourier frequency 1, resulting from the overall shape of the input pattern, and power at frequencies corresponding to periods in ranges around two and four units. It is thus comparable to the power spectrum comprised in the first twelve bars of the average periodogram of Figure 3.6(a).

4. Linear approximation to the nonlinear model.

A standard approach to discussing stability of equilibria of nonlinear systems is via the following theorem (see [7], pp. 93 ff.):

> Let H be a function from $\mathbf{R}^N$ to $\mathbf{R}^N$ and let $\bar{x}$ be an equilibrium of the system $\dot{x} = H(x)$. Suppose all the eigenvalues of the Jacobian matrix of H at $\bar{x}$ have negative real parts. Then there is a neighborhood Ω of $\bar{x}$ such that if $x(0)$ is in Ω, then $x(t)$ approaches $\bar{x}$ exponentially fast as $t \to \infty$.

For the system (1.2) the function H is $H(x) = -x+p+WF(x)$ and its Jacobian matrix at $\bar{x}$ is

$$JH(\bar{x}) = WF'(\bar{x})-I, \tag{4.1}$$

where I is the $N \times N$ identity matrix and $F'(\bar{x})$ is the diagonal matrix whose (j,j) entry is $f_j{}'(\bar{x}_j)$. It follows that if the eigenvalues of $WF'(\bar{x})$ all have real parts less than 1, then all equilibria of the nonlinear system (1.2) are stable. Moreover, in case W is symmetric (but not in general), it can be shown that if the eigenvalues of W all have real parts less than 1 and if $|f_j{}'(\bar{x}_j)| \leq 1$ for all j, then the eigenvalues of $WF'(\bar{x})$ all have real parts less than 1.

This approach, however, does not settle the question of the existence of equilibria or of the number of equilibria; nor does it provide information about their domains of attraction, which are the neighborhoods Ω of the above theorem. In the next section we show that we can answer these questions and that of stability by assuming the network is contractive.

5. Contractive systems.

a. Definition.

An equilibrium of the system (1.2) is a fixed point of the map $G:\mathbb{R}^N \rightarrow \mathbb{R}^N$ defined by

$$G(x) = p + WF(x). \tag{5.1}$$

Because contraction maps (the definition is given in the next paragraph) have unique fixed points which can be approximated closely by iteration, is is natural to consider *contractive* neural networks: those for which the map $G(x)$ is a contraction. We will see that this assumption imposes little restriction on the network's pattern-forming properties. We will show also that the equilibrium of a contractive network is globally asymptotically stable; and that in some cases the equilibrium is well approximated by the equilibrium of the corresponding linear network model (2.1).

A map $G:\mathbb{R}^N \rightarrow \mathbb{R}^N$ is a *contraction* if there exists a real constant $\alpha < 1$ such that

$$\|G(x^1) - G(x^2)\| \leq \alpha \|x^1 - x^2\| \text{ for all } x^1, x^2 \in \mathbb{R}^N. \tag{5.2}$$

Here $\| \cdot \|$ denotes the usual Euclidean (L^2) norm:

$$\|x\| = \left(\sum_{k=1}^{N} x_k^{\,2}\right)^{\frac{1}{2}}.$$

Notice that $G(x)$ in (5.1) is a contraction if and only if $WF(x)$ is a contraction.

b. Unique stable equilibria.

The Contraction Mapping Theorem (see [3], pp. 161 ff.) is:

If $G:\mathbb{R}^N \rightarrow \mathbb{R}^N$ is a contraction, then G has one and only one fixed point $\bar{x}$. Given any $x^0 \in \mathbb{R}^N$, the sequence $\{x^n\}$ defined by $x^{n+1} = G(x^n)$ (n=0, 1, 2, ...) converges to $\bar{x}$ at a rate given by

$$\|x^n - \bar{x}\| \leq \frac{\alpha^n}{1-\alpha}\|x^1 - x^0\| \quad (n = 1, 2, \ldots)\,.$$

It follows that if G is a contraction, then the system (1.2) has a unique equilibrium, which can be approximated exponentially by iterating G from any starting value whatever. We shall see presently that this equilibrium must have all of $\mathbf{R}^N$ as its domain of attraction, so that the neural network will converge to it regardless of the initial state.

It is of interest to note the connection between the Euler approximations to the solution of (1.2) from a given initial state x^0 and the iteration of the map G starting from x^0. The Euler approximation for a system $\dot{x} = K(x)$ with step size Δt starting from $x(0) = x^0$ is given by

$$x^{n+1} = x^n + \Delta t\, K(x^n) \qquad (n = 0, 1, 2, \ldots). \tag{5.3}$$

For the system (1.2), $K(x)$ is equal to $\frac{1}{\mu}(-x + p + WF(x))$, and thus the Euler approximation is

$$x^{n+1} = x^n + \frac{\Delta t}{\mu}(-x^n + p + WF(x^n)). \tag{5.4}$$

If $\Delta t = \mu$, then this is precisely the n^{th} iterate of the map G. For any other value of Δt, (5.4) gives the n^{th} iterate of the map

$$G_{\Delta t}(x) = (1 - \frac{\Delta t}{\mu})x + \frac{\Delta t}{\mu}(p + WF(x)),$$

which is easily seen to be a contraction having the same fixed point $\bar{x}$ as does G, as long as $\Delta t < \mu$. What we have seen therefore is the following: If $G(x) = p + WF(x)$ is a contraction, then for any step size $\Delta t \leq \mu$ and any starting value x^0, the Euler approximation to the system (1.2) converges as $t \to \infty$ to its unique equilibrium, which is the fixed point of G.

Curiously, it does not follow directly from the convergence to $\bar{x}$ of all Euler approximations with sufficiently small step size from arbitrary initial points, that in general the network itself converges to $\bar{x}$. However, in the contractive case an application of Liapounov's direct method (see [7], pp. 291 ff.) provides this convergence, as we shall now show.

Given an arbitrary system $\dot{h} = K(h)$ with an equilibrium at $h=0$ (that is, such that $K(0)=0$) and a neighborhood Ω of $h=0$, a function $V:\Omega\rightarrow\mathbf{R}$ is a *Liapounov function* for the system if (i) V is a continuous function of h; (ii) $V(0)=0$ and $V(h)>0$ for all nonzero h; and (iii) $\dot{V}(0)=0$ and $\dot{V}(h) < 0$ for all nonzero h, where $\dot{V}:\mathbf{R}^N\rightarrow\mathbf{R}^N$ is defined by

$$\dot{V}(h) = \frac{\partial V}{\partial H}\cdot K(h) = \frac{\mathrm{d}}{\mathrm{dt}}V(h(t)).$$

The Liapounov theorem is that if such a function exists, then all solutions of $\dot{h} = K(h)$ for initial $h(0)$ in Ω approach 0 as $t\rightarrow\infty$.

We apply this method to the system (1.2) in the contractive case as follows: as before let $\bar{x}$ be the unique equilibrium of the system (1.2), and suppose $G(x) = p+WF(x)$ is a contraction. If we write h for $x-\bar{x}$, then (1.2) becomes

$$\dot{h} = \frac{1}{\mu}(-h + C(h)) \tag{5.5}$$

where

$$C(h) = p-\bar{x}+WF(\bar{x}+h) \tag{5.6}$$

is a contraction. We show that

$$V(h) = \|h\|^2 = h\cdot h$$

is a Liapounov function for the system (5.5) on $\Omega = \mathbf{R}^N$; it will then follow that $\dot{h}(t)\rightarrow 0$ for any $h(0)$; that is, that $x(t)\rightarrow\bar{x}$ for any $x(0)$, as $\mathrm{t}\rightarrow\infty$.

(i) and (ii) in the definition of a Liapounov function are obvious, and

$$\dot{V}(h) = \frac{1}{\mu}(h\cdot(-h+C(h))),$$

so that $\dot{V}(0) = 0$ and we need show only that $h\cdot h > |h\cdot C(h)|$ for nonzero h. But

$$|h\cdot C(h)| \leq \|h\|\cdot\|C(h)\| \leq \alpha\|h\|^2 < \|h\|^2 = h\cdot h,$$

and the result follows.

c. Sufficient conditions for a contractive system.

The *norm* of a matrix W is the nonnegative real number $\|W\|$ defined by

$$\|W\| = \max_{\|x\| \neq 0} \frac{\|Wx\|}{\|x\|};$$

thus $\|Wx\| \leq \|W\|\|x\|$ for all $x \in \mathbf{R}^N$, with equality for at least one x. If W is a symmetric matrix, then $\|\mathrm{W}\| = \max\{|\lambda| : \lambda$ is an eigenvalue of $W\}$, but in the absence of symmetry there is no simple general relationship between the norm and the eigenvalues of a matrix.

Now suppose the derivatives of the functions $f_j(x)$ are all bounded above by some number β: that is, $|f_j{}'(x)| < \beta$ for all $x \in \mathbf{R}$ and $j = 1, 2, \ldots, N$. It then follows immediately from the Mean Value Theorem that for any vectors x^1 and x^2 in $\mathbf{R}^N$,

$$\left\|G(x^1) - G(x^2)\right\| \leq \|W\|\left\|F(x^1) - F(x^2)\right\| \leq \beta\,\|W\|\|x^1 - x^2\|.$$

Consequently:

> *If $\beta\,\|W\| < 1$, where β is a bound on the $|f_j{}'(x)|$, then $G(x) = p + WF(x)$ is a contraction and therefore (1.2) has a unique equilibrium, which is globally asymptotically stable and can be approximated exponentially by iterating G from any starting value.*

Since the norm of a matrix W can be reduced by dividing all entries in the matrix by an appropriate constant, and since the functions used as "smooth sigmoids" in models whose units are aggregates of neurons typically have bounded derivatives, we see that insuring that a network of the form (1.2) is contractive is simply a matter of scaling the synaptic weights and the sigmoids f_j.

d. Estimating the equilibrium of a nonlinear system.

Finally we consider the possibility of using the equilibrium $\bar{x}$ of the linear system (2.1) as a basis for estimating the equilibrium of the nonlinear system. The purpose is not the approximate computation of the equilibrium of a nonlinear contractive system; this, after all, can be done efficiently by iterating G, because of the exponential convergence guaranteed by the Contraction Mapping

Theorem. Rather it is of interest to know when the equilibrium of the nonlinear system (1.2) is close to that of the linear system (2.1), so that it can be expected to exhibit similar characteristics, such as the spatial frequencies discussed in Section 3 above.

Accordingly let $\bar{x}$ and $\bar{y}$ be the unique equilibria of the systems (2.1) and (1.2) respectively; that is,

$$\bar{x} = p + W\bar{x} \quad \text{and} \quad \bar{y} = p + WF(\bar{y}).$$

We are interested in bounding $\|\bar{x}-\bar{y}\|$. If we iterate $G(y) = p + WF(y)$ starting from $y^0 = \bar{x}$, then it follows from the Contraction Mapping Theorem that

$$\|y^n-\bar{y}\| \le \frac{(\beta\|W\|)^n}{1-\beta\|W\|}\|y^1-\bar{x}\|,$$

provided $\beta\|W\| < 1$, where β is a bound on $|f_j{}'(x)|$. Therefore

$$\|\bar{y}-\bar{x}\| \le \frac{(\beta\|W\|)^n}{1-\beta\|W\|}\|y^1-\bar{x}\| + \|y^n-\bar{x}\|. \tag{5.7}$$

So we can bound $\|\bar{y}-\bar{x}\|$ by making a small number of iterations of G starting from $\bar{x}$ and applying (5.7). For $n = 1$, for example, we get

$$\|\bar{y}-\bar{x}\| \le \frac{1}{1-\beta\|W\|}\|p+WF(\bar{x})-\bar{x}\| = \frac{\|WF(\bar{x})-W\bar{x}\|}{1-\beta\|W\|} \le \frac{\|W\|\|F(\bar{x})-\bar{x}\|}{1-\beta\|W\|}. \tag{5.8}$$

This bound can be simplified further by making some regularity assumptions on the form of the sigmoids $f_j(x)$. In particular, if

$$f_j(0) = 0,\ f_j{}'(0) = 1, \text{ and } |f_j{}''(x)| \le \gamma \tag{5.9}$$

for all j and x, where γ is some positive constant, then for each j and any x we have

$$f_j(x) - x = \frac{x^2}{2}f_j{}''(\theta)$$

for some θ between 0 and x. Therefore

$$\| F(\bar{x}) - \bar{x} \| \leq \frac{\gamma}{2} \| \bar{x} \|_4^2 ,$$

where $\| \cdot \|_4$ denotes the L^4 norm, which is bounded by the L^2 norm:

$$\| x \|_4 = \Big(\sum_{k=1}^{N} x_k{}^4 \Big)^{\frac{1}{4}} \leq \| x \| .$$

Thus the bound (5.8) becomes, if (5.9) is assumed,

$$\| \bar{y} - \bar{x} \| \leq \frac{\gamma \| W \|}{2(1 - \beta \| W \|)} \| \bar{x} \|^2 . \tag{5.10}$$

Bounds like (5.8) and (5.10) reflect and quantify the intuitive observation that if the equilibrium of the linear system is within a region in which $F(x)$ is well approximated by x, then the equilibrium of the nonlinear system will be nearby. The bound (5.10) reflects the special case in which $F(x)$ is well approximated by x near the origin.

REFERENCES

[1] S. Amari: Field theory of self-organizing neural nets. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13*, 741–748, 1983.

[2] S. Amari and M. A. Arbib (eds.): *Competition and Cooperation in Neural Nets.* Vol 45 of *Lecture Notes in Biomathematics.* Berlin, Springer-Verlag, 1982.

[3] R. G. Bartle: *The Elements of Real Analysis* (second edition). New York, John Wiley and Sons, 1976.

[4] G. A. Carpenter and S. Grossberg: Neural Dynamics of Category Learning and Recognition: Attention, Memory Consolidation, and Amnesia. In J. Davis, R. Newburgh, and E. Wegman (eds.), *Brain Structure, Learning, and Memory*, AAAS Symposium Series, 1986.

[5] J. L. Feldman and J. D. Cowan: Large-Scale Activity in Neural Nets, I and II. *Biol. Cybernetics* 17, 29–51, 1975.

[6] S. Grossberg: *The Adaptive Brain* I and II. Amsterdam, Elsevier/North Holland, 1986.

[7] J. K. Hale: *Ordinary Differential Equations.* New York, John Wiley and Sons, 1969.

[8] A. V. Holden: *Models of the Stochastic Activity of Neurones.* Vol. 12 of *Lecture Notes in Biomathematics*, Berlin, Springer-Verlag, 1976.

[9] S. L. Juliano, P. J. Hand, and B. L. Whitsel: Patterns of increased metabolic activity in somatosensory cortex of monkeys *Macaca fascicularis*, subjected to controlled cutaneous stimulation: a 2-deoxyglucose study. *J. Neurophysiology* 46, 1260–1284, 1981.

[10] R. P. Lippman: An Introduction to Computing with Neural Nets. *IEEE ASSP Magazine*, April 1987.

[11] J. Szentagothai: The neurone network of the cerebral cortex: a functional interpretation. *Proc. Row. Soc. Lond.* B201, 219–249, 1978.

[12] B. L. Whitsel, O. Favorov, M. Tommerdahl, M. Diamond, S. Juliano, and D. Kelly: Dynamic processes govern the somatosensory cortical response to natural stimulation. To appear in J. S. Lund (ed.), *Organizational Principles of Sensory Processing: The Brain's View of the World.* New York, Oxford University Press, 1987.

[13] B. L. Whitsel and D. G. Kelly: Knowledge Acquisition ("Learning") by the Somatosensory Cortex. In J. Davis, R. Newburgh, and E. Wegman (eds)., *Brain Structure, Learning, and Memory.* AAAS Neuroscience Symposium, 1986.

[14] H. R. Wilson and J. D. Cowan: A Mathematical Theory of the Functional Dynamics of Cortical and Thalamic Nervous Tissue. *Kybernetik* 13, 55–80, 1973.

Figure 3.1: Two spatially-homogeneous one-dimensional synaptic weight patterns (a and c) and their squared eigenvalues (b and d), as computed from Equation (3.2), for a system of $N = 81$ units. All eigenvalues are real because the patterns are symmetric, and are less than 1 in absolute value. λ_3 is the largest of Pattern 2's eigenvalues; steady states will show enhanced spatial frequencies with periods of around 81/3 units. Those from Pattern 3 will show enhanced frequencies with periods of around 81/10 units.

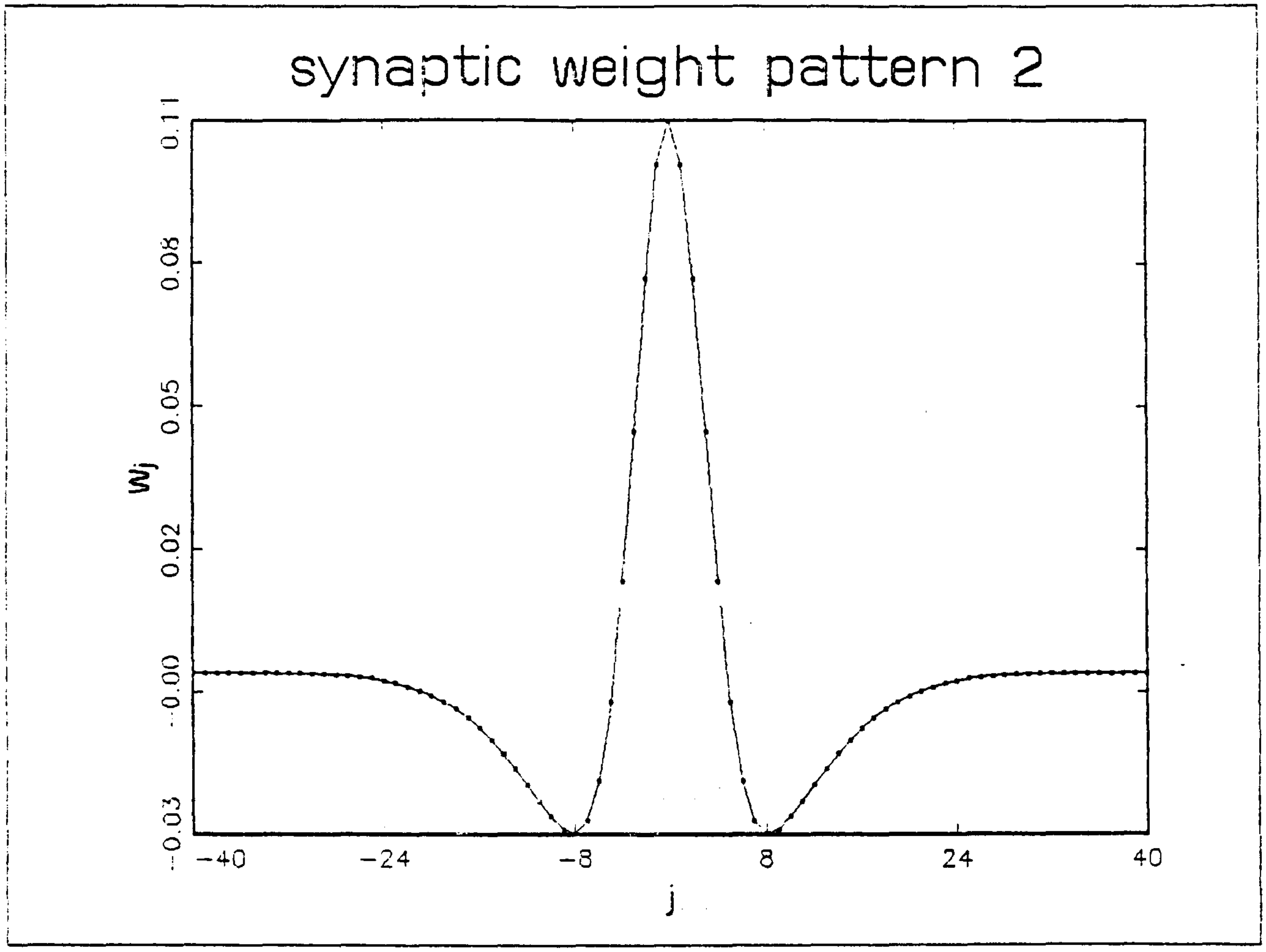
Figure 3.2: An input pattern (a) for a one-dimensional network of $N = 81$ units, and the steady states (b and c) produced from it by the two networks depicted in Figure 3.1. A change in the spread of excitatory and inhibitory connections produces a marked change in the network's encoding of an input stimulus; Pattern 2, with its wider spread of connections, enhances lower spatial frequencies than does Pattern 3.

Figure 3.3: See caption for Figure 3.2. Here the input pattern is a smooth curve with uncorrelated Gaussian noise added. Connection patterns 2 and 3 enhance different spatial frequencies in the noise, as described in the caption for Figure 3.1.

Figure 3.4: Spatially-homogeneous and symmetric two-dimensional weight pattern (a) and its squared eigenvalues (b) computed from Equation (3.8), for a 30×30 system. The eigenvalues are real and less than 1 in absolute value; the largest if λ_{03} (and seven others which equal it by symmetry).

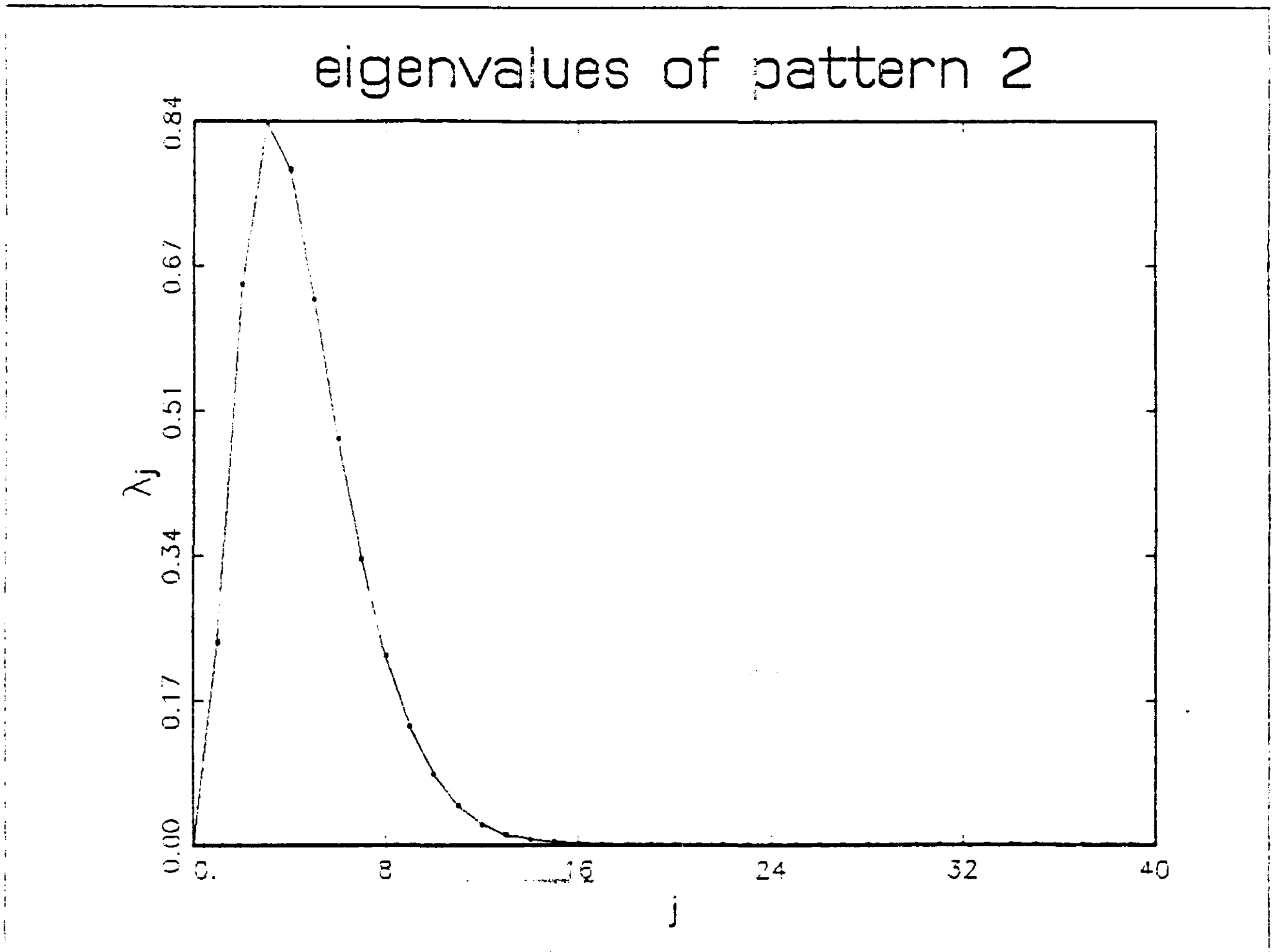
Figure 3.5: Input pattern (a) for a two-dimensional network of 30×30 networks, and the steady state (b) produced from it by the network whose connection pattern is shown in Figure 3.4. Spatial frequencies with wave numbers near 3 in the x- and y-directions are maximally enhanced.

Figure 3.6: Choice of synaptic weights in one-dimensional linear version of the model to show correspondence with observed frequencies in autoradiographic data. (a): Average of periodograms of eight serial sections taken from within a single "patch" (area of high labeling intensity roughly 450 μ in medio-lateral extent) in forelimb area of cat S-I. See text for description of notable features of the low-frequency spectrum, especially power in sixth to ninth bars (periods of 45 to 75 micrometers, or about two minicolumns), and power in fourth bar (periods of 112.5 to 150 micrometers, or about four minicolumns). (Courtesy of Dr. B. L. Whitsel.) (b): Eigenvalues which would enhance in the model the frequencies with periods around two and around four units; that is, those corresponding to the ones noted in (a). (c): Synaptic weight pattern constructed to have these eigenvalues. (d): Output of the linear model having the synaptic weights shown, using the input pattern shown in Fig. 3.3(a). (e): Periodogram of output, showing significant power at frequencies corresponding to those noted in (a).



synaptic weight pattern 2
0.11
0.08
0.05
0.02
-0.00
-0.03
Wj
-40
-24
-8
8
24
40
j

Fig. 3.1 (a)



eigenvalues of pattern 2
0.84
0.67
0.51
0.34
0.17
0.00
λ_j
0.
8
16
24
32
40
j

Fig. 3.1 (b)

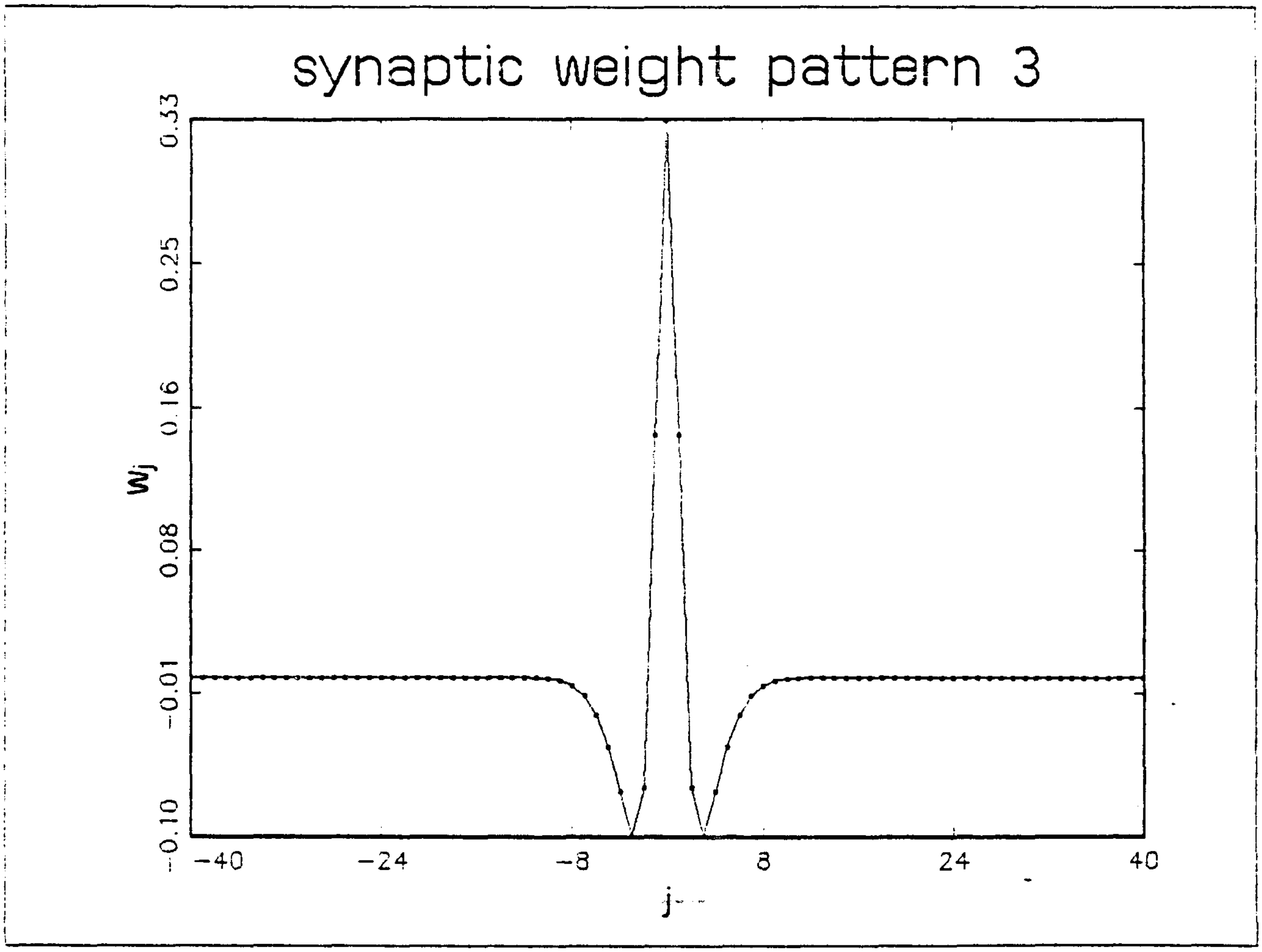
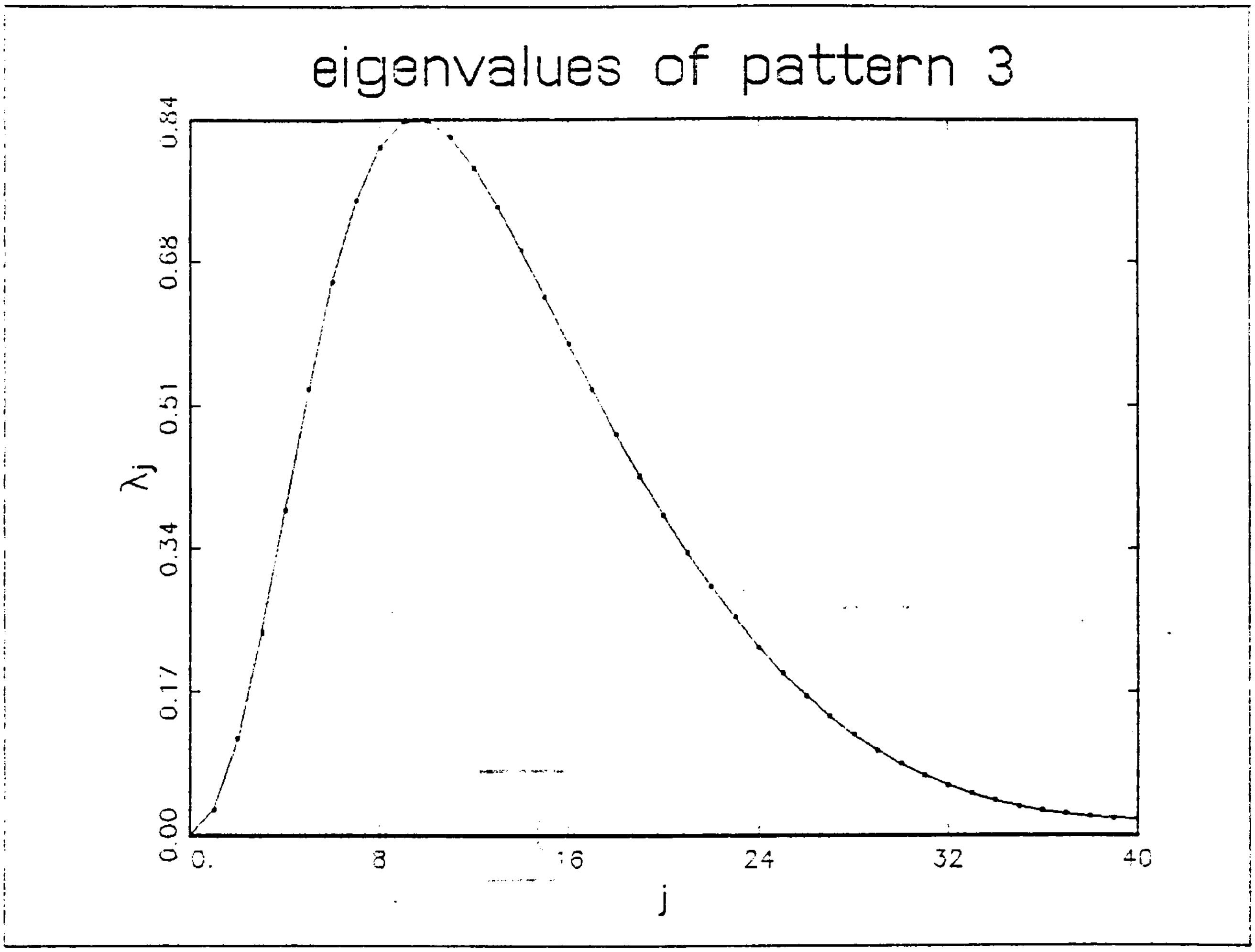

synaptic weight pattern 3
0.33
0.25
0.16
0.08
-0.01
-0.10
w_j
-40
-24
-8
8
24
40
j

Fig. 3.1 (c)



eigenvalues of pattern 3
0.84
0.68
0.51
0.34
0.17
0.00
λ_j
0.
8
16
24
32
40
j

Fig. 3.1 (d)

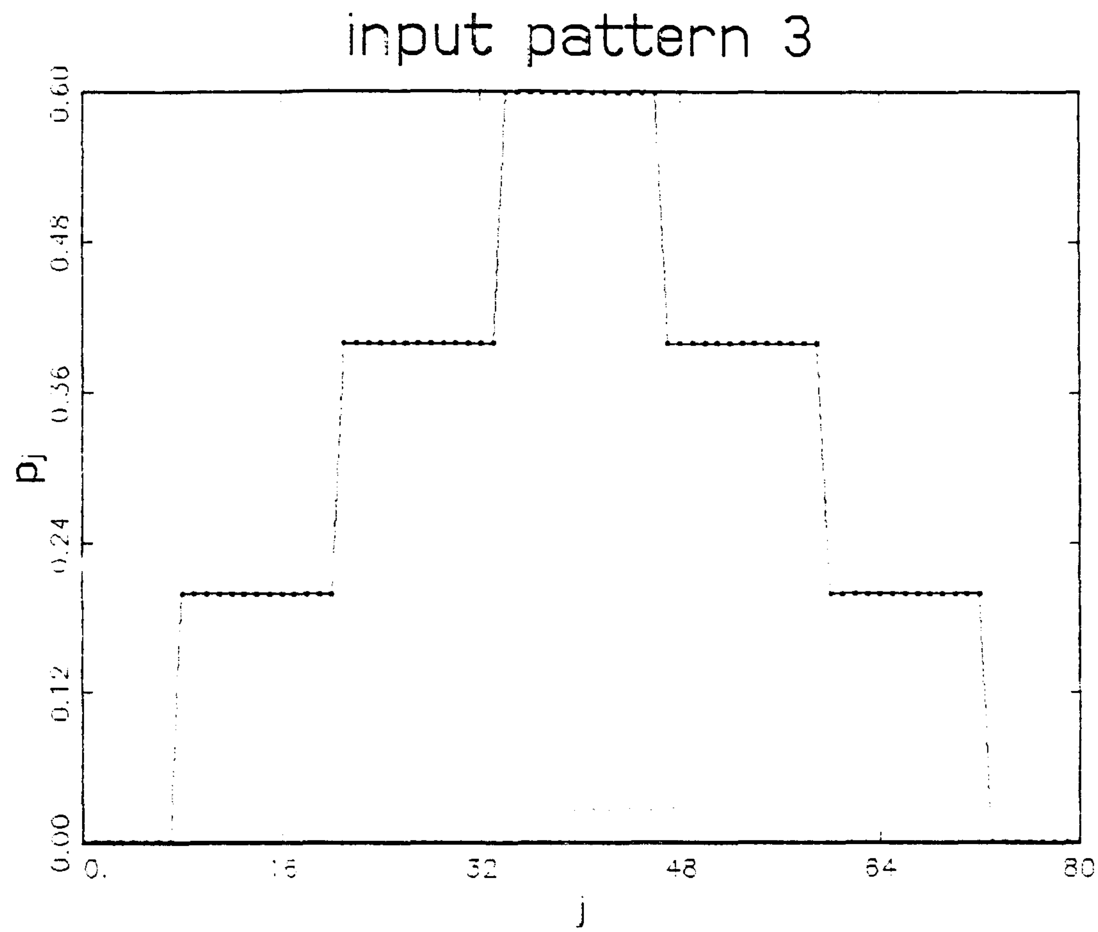
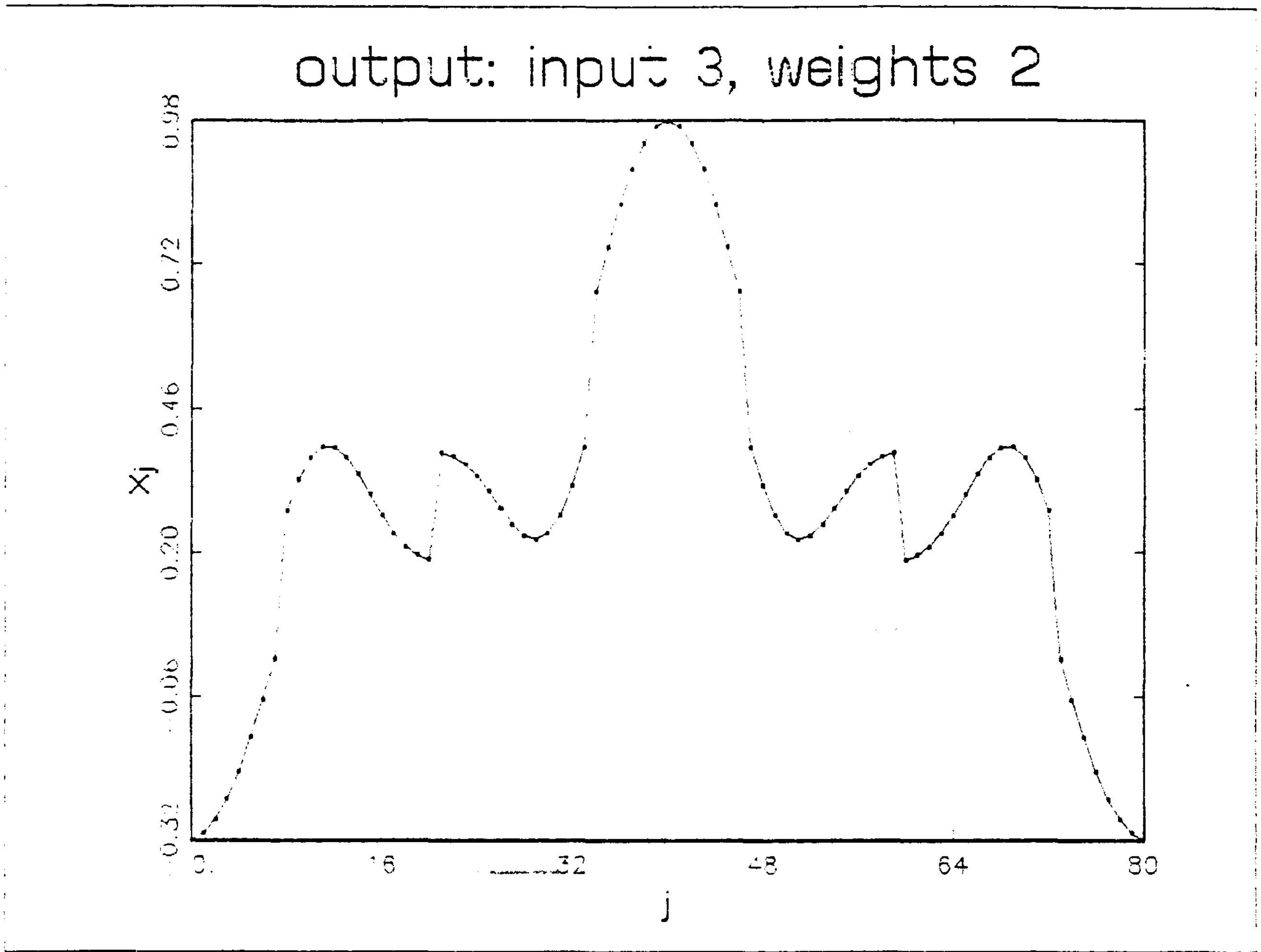
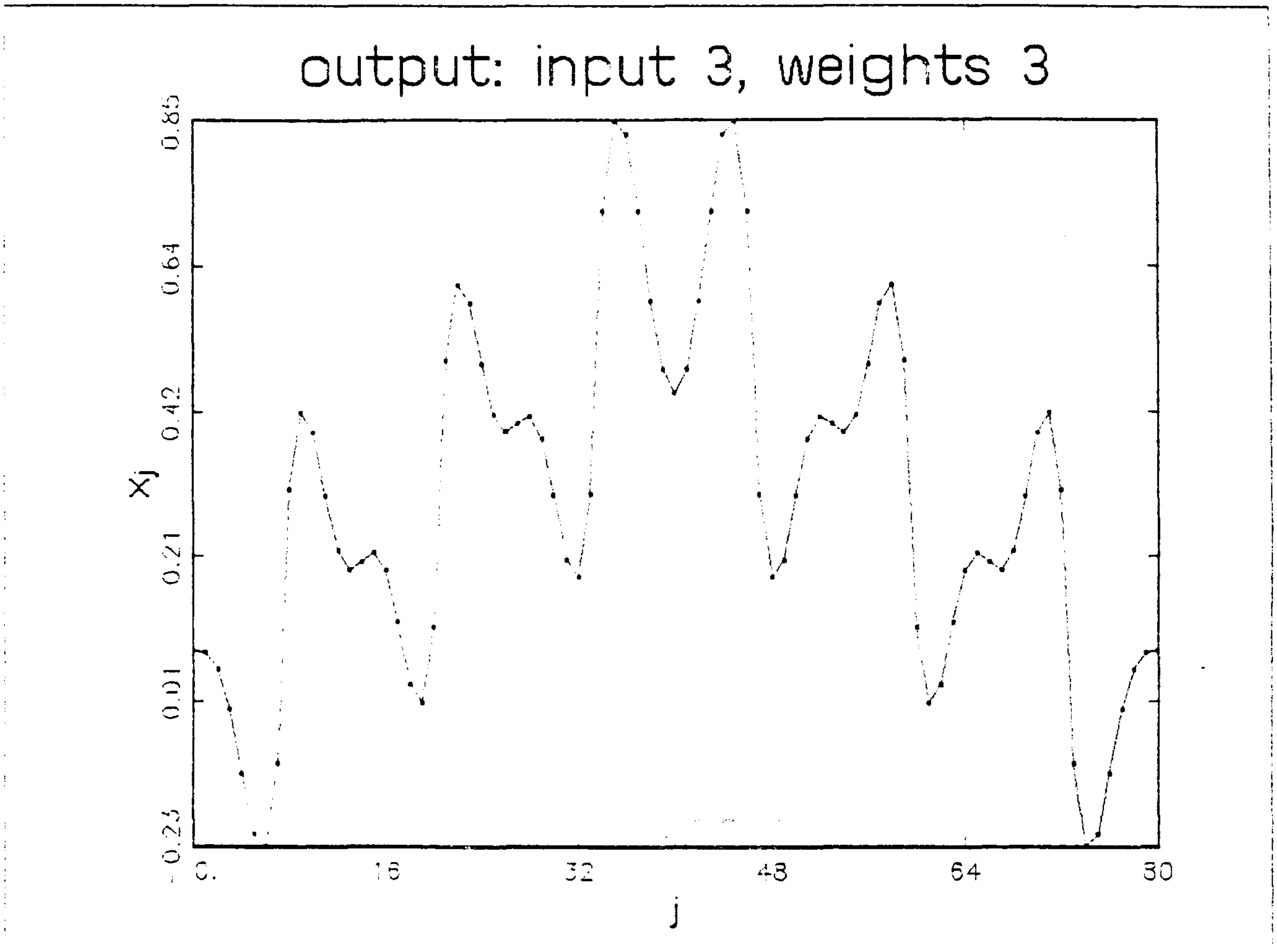


Fig. 3.2 (a)



output: input 3, weights 2
0.98
0.72
0.46
0.20
-0.06
-0.32
Xj
0.
16
32
48
64
80
j

Fig. 3.2 (b)



output: input 3, weights 3
0.85
0.64
0.42
0.21
0.01
-0.23
Xj
0.
16
32
48
64
80
j

Fig. 3.2(c)

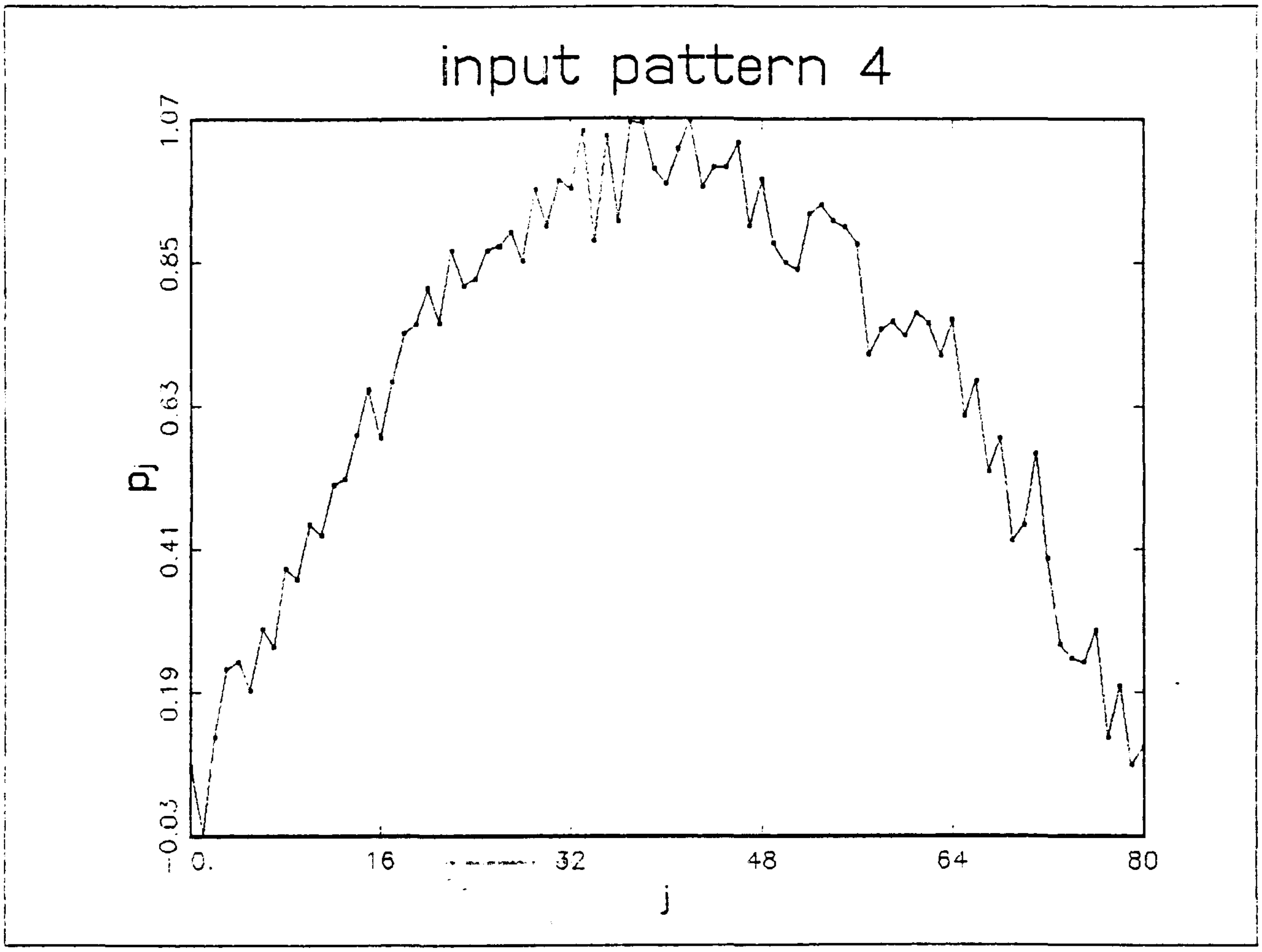


Fig. 3.3 (a)

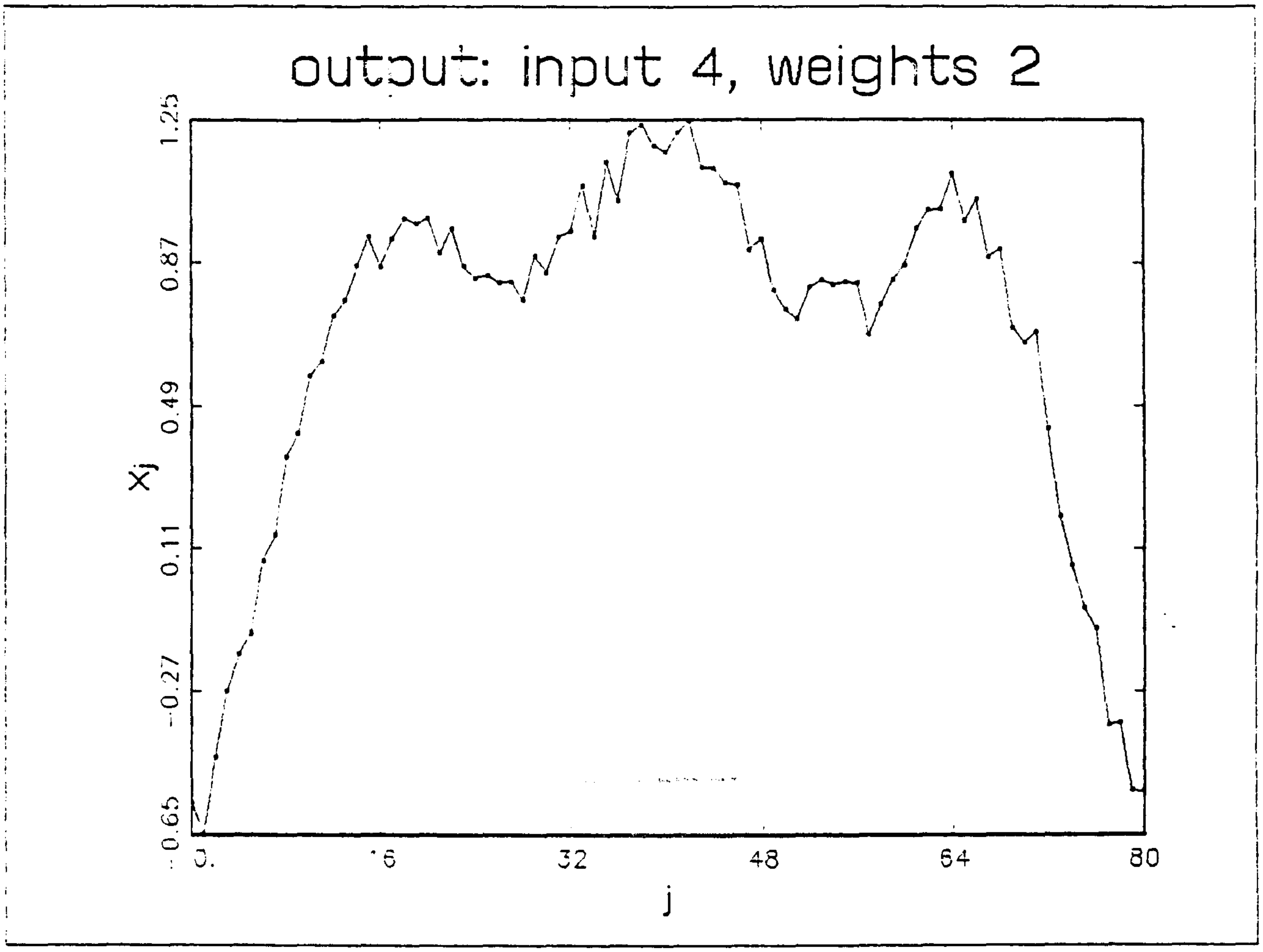
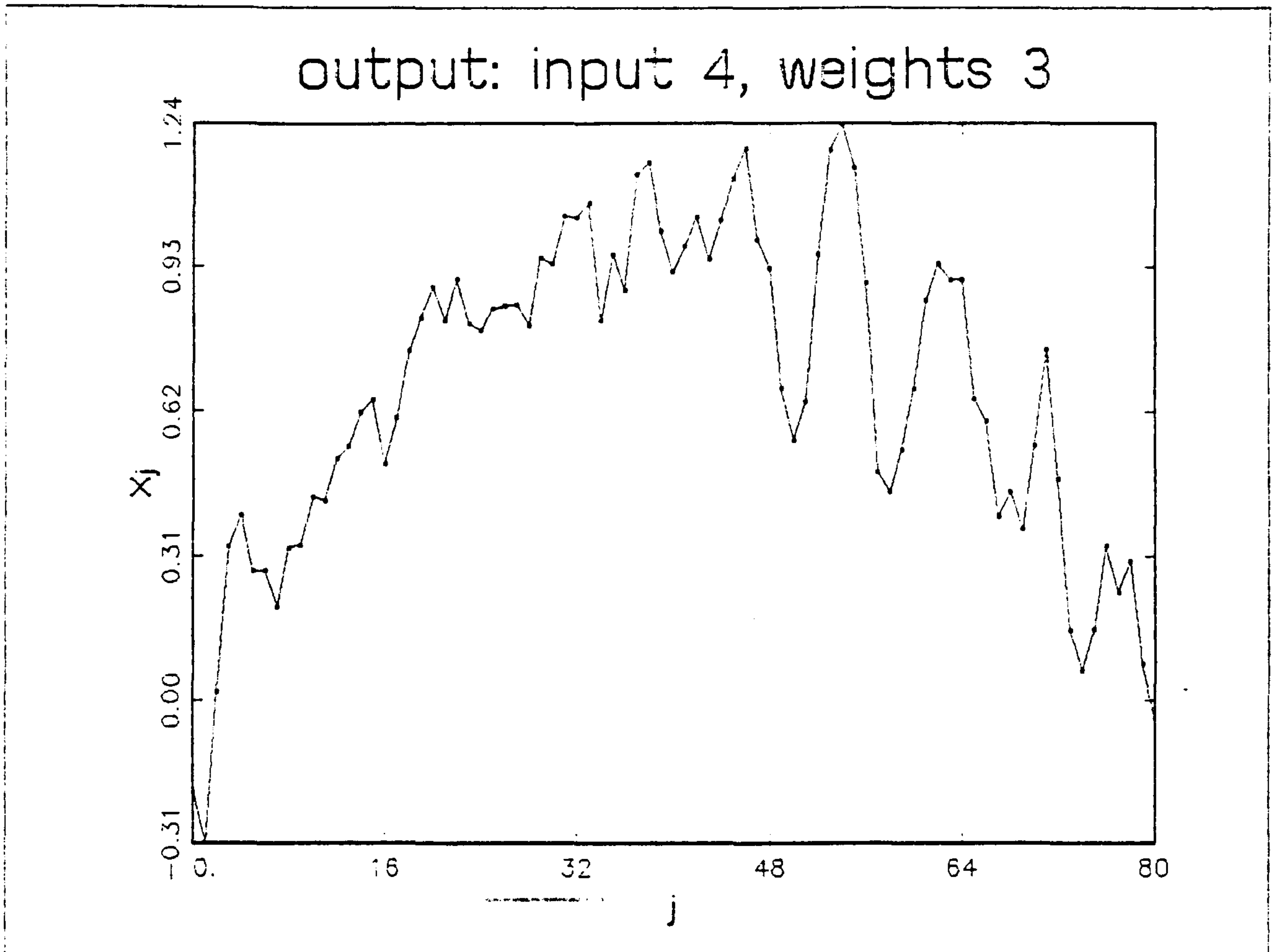

output: input 4, weights 2
1.25
0.87
0.49
0.11
-0.27
-0.65
Xj
0.
16
32
48
64
80
j

Fig. 3.3(b)



output: input 4, weights 3
1.24
0.93
0.62
0.31
0.00
-0.31
Xj
0.
16
32
48
64
80
j

Fig. 3.3 (c)

synaptic weight pattern 7

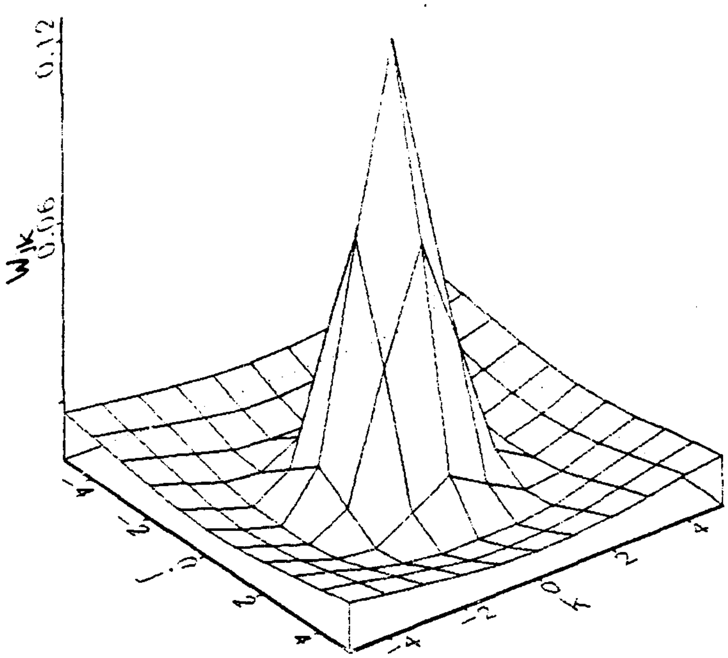


Fig. 3.4(a)

eigenvalues of pattern 7

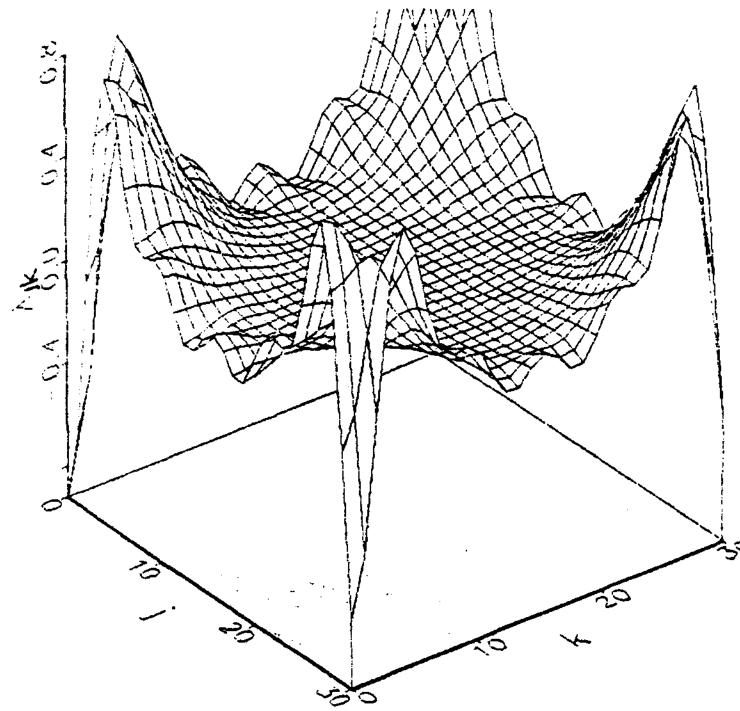


Fig. 3.4 (b)

input pattern 8

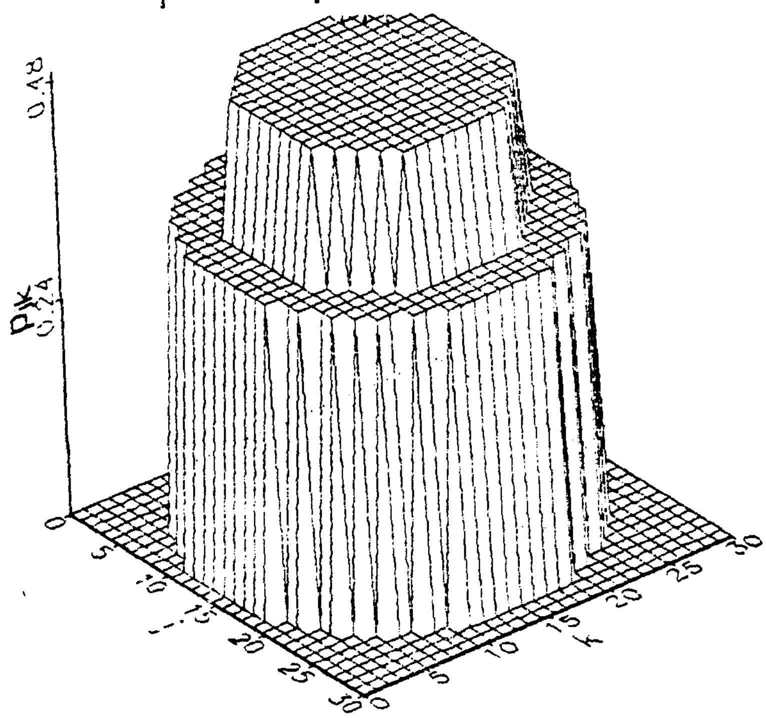


Fig. 3.5(a)

output: input 8, weights 7

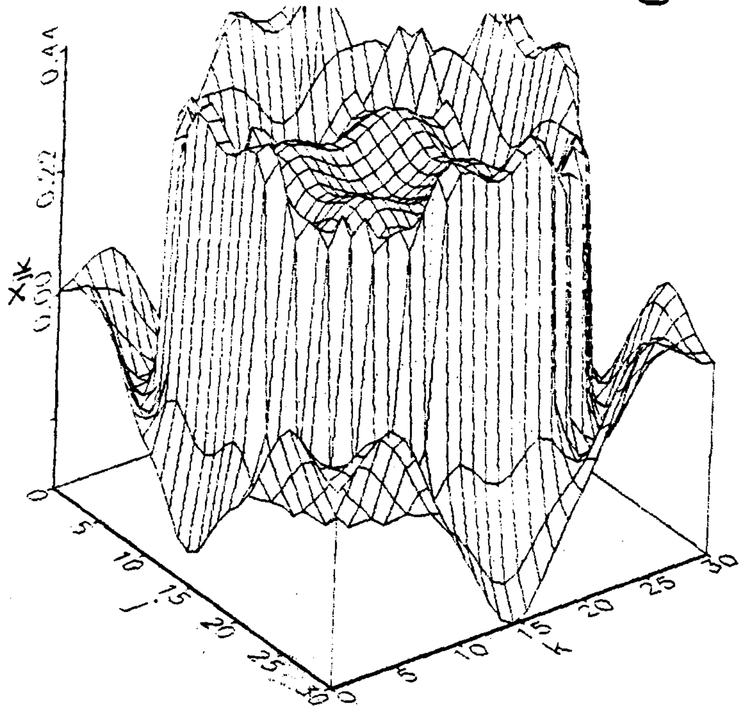
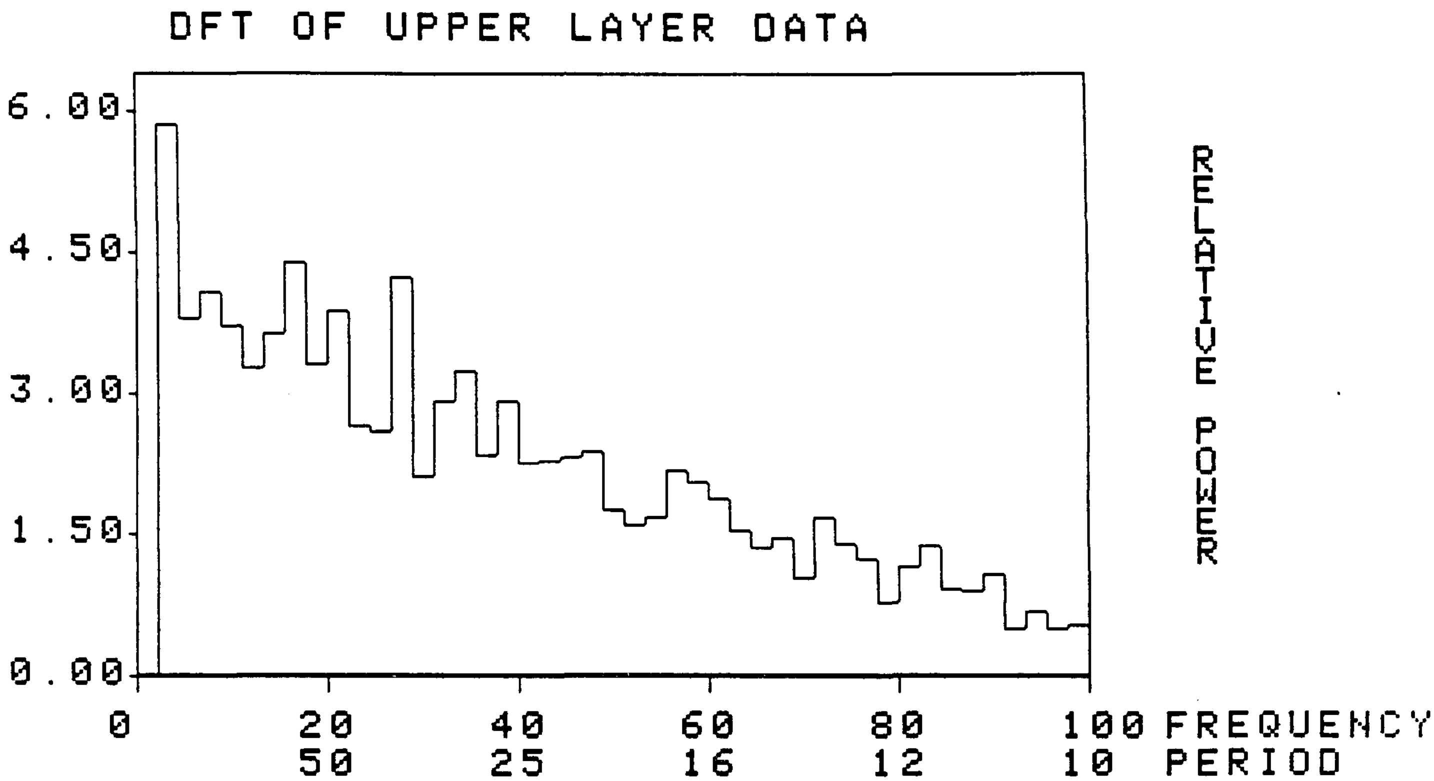


Fig. 3.5(b)



DFT OF UPPER LAYER DATA
6.00
4.50
3.00
1.50
0.00
0
20
40
60
80
100 FREQUENCY
50
25
16
12
10 PERIOD
RELATIVE POWER

Fig. 3.6(a)

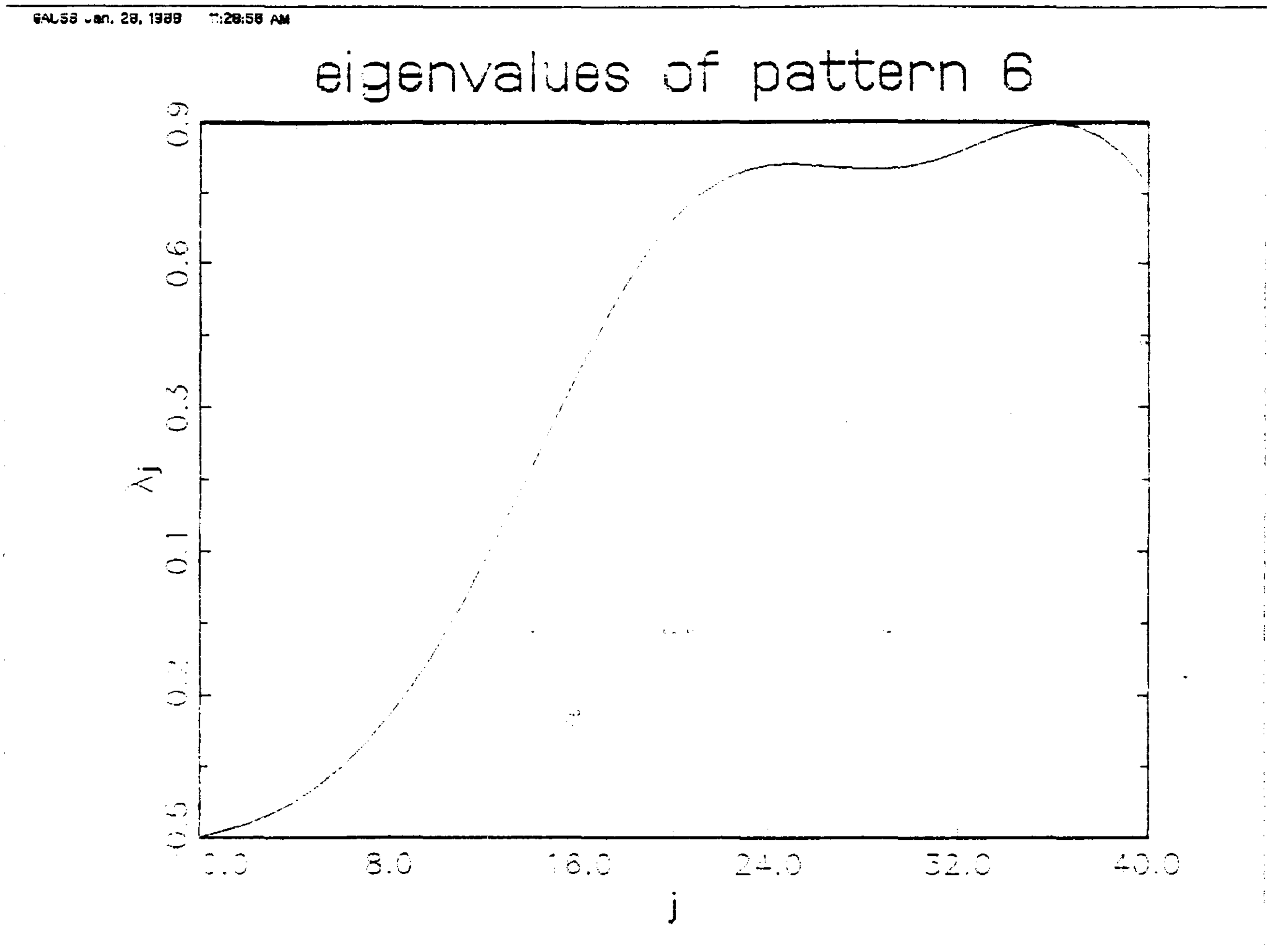

eigenvalues of pattern 6
0.9
0.6
0.3
0.1
8.0
16.0
24.0
32.0
40.0
j
λ_j

Fig. 3.6(b)

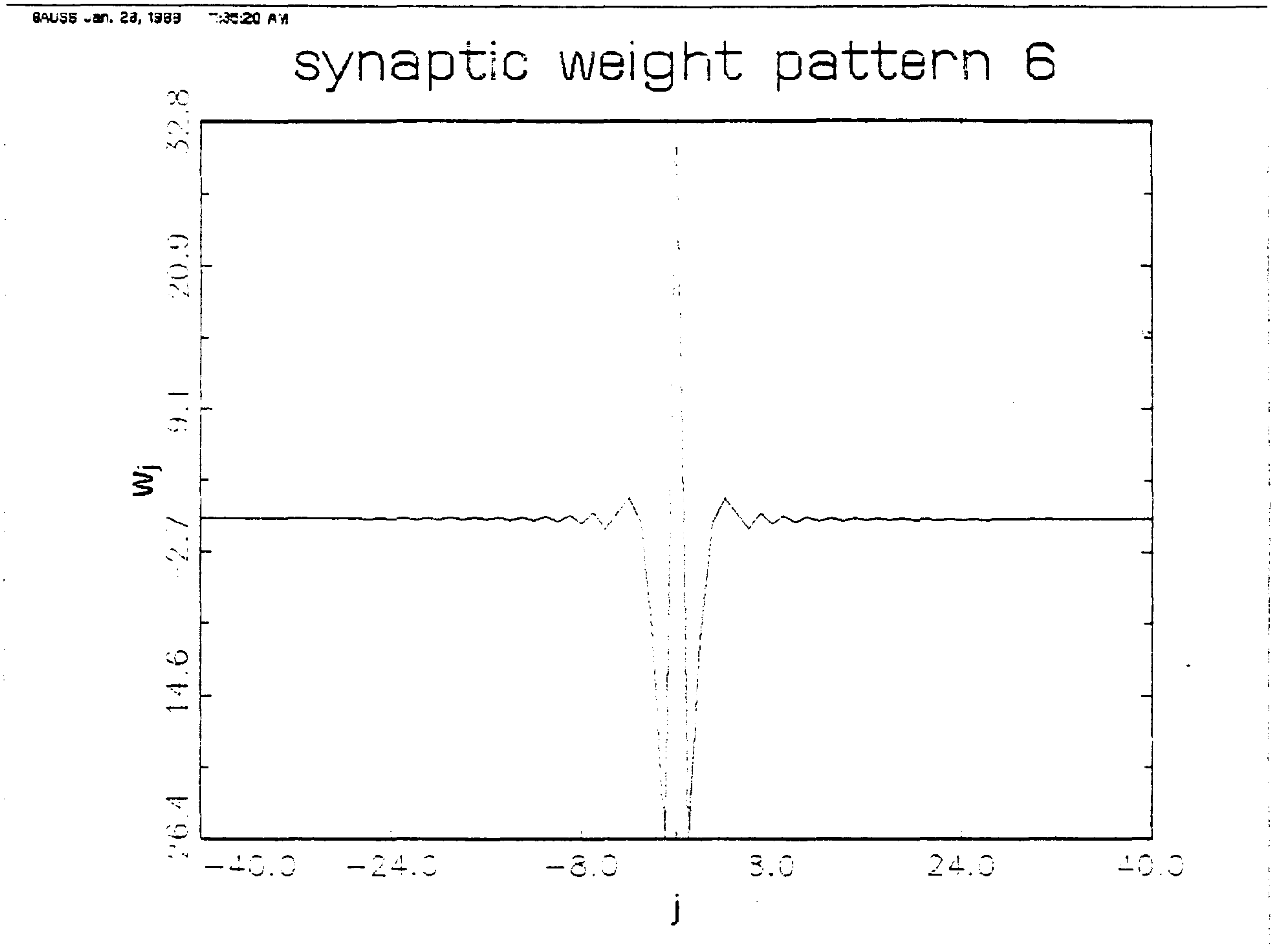

GAUSS Jan. 23, 1989 ?:35:20 AM
synaptic weight pattern 6
32.8
20.9
9.1
-2.7
14.6
26.4
Wj
-40.0
-24.0
-8.0
8.0
24.0
40.0
j

Fig. 3.6(c)

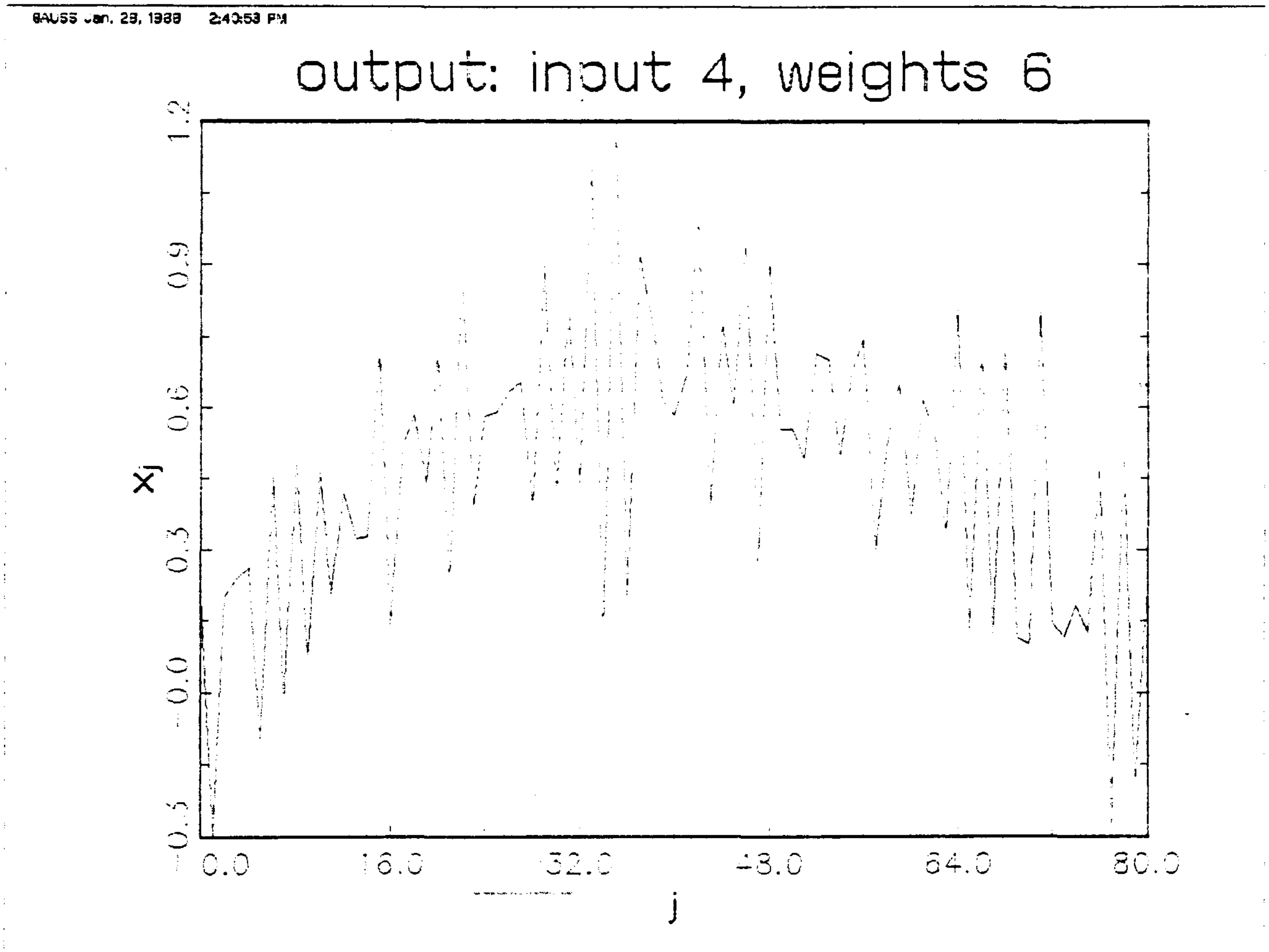

GAUSS Jan. 28, 1988 2:40:53 PM
output: input 4, weights 6
1.2
0.9
0.6
0.3
-0.0
-0.3
x_j
0.0
16.0
32.0
48.0
64.0
80.0
j

Fig. 3.6 (d)

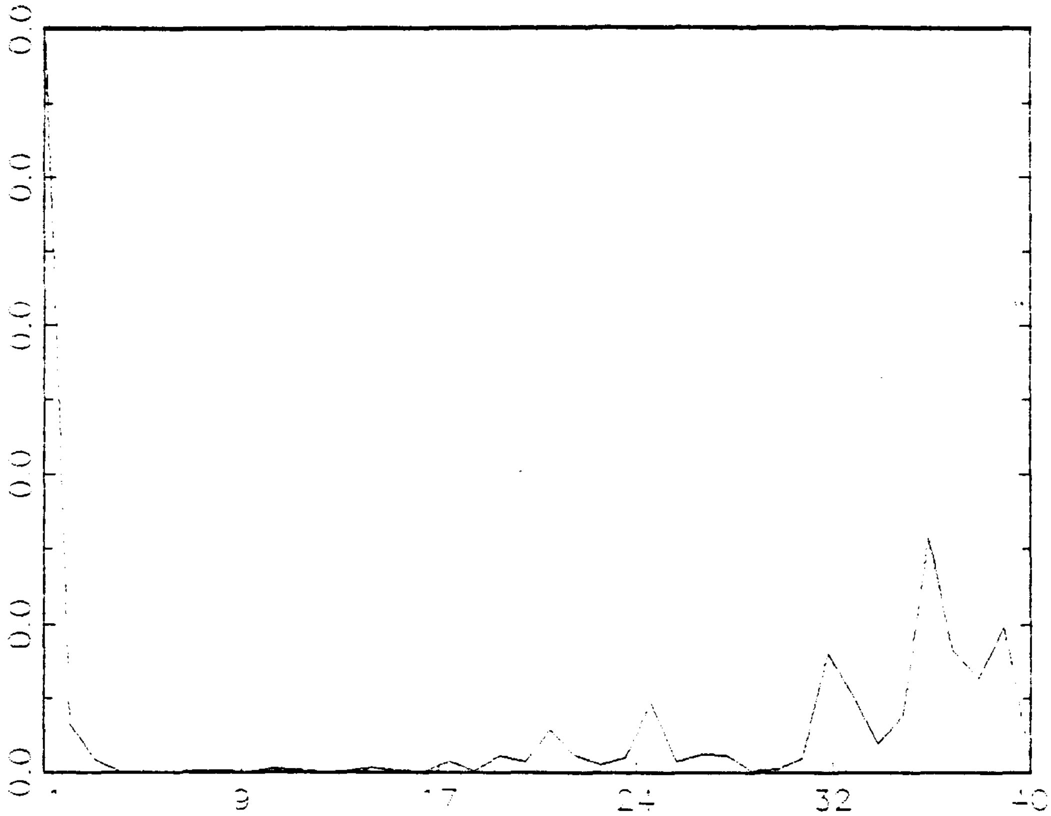


Fig. 3.6(e)