The goal of this work is to characterize mathematically the essential mechanisms and principles of operation of the mamalian olfactory neural network and evaluate its computation and pattern recognition capabilities. The intent is to explicate novel design principles that may underly the superior performance of biological systems in pattern recognition through detailed study of a particular system.
Investigation of Dynamic Algorithm for Pattern Recognition in Cerebral Cortex

1 September 1987 - 31 August 1988

Dr. Walter J. Freeman
2571 Life Science Building
Department of Physiology-Anatomy
University of California
Berkeley, California 94720
THEORETICAL RESEARCH

OBJECTIVE

The goal of this work is to characterize mathematically the essential mechanisms and principles of operation of the mammalian olfactory neural network and evaluate its computation and pattern recognition capabilities. The intent is to explicate novel design principles that may underly the superior performance of biological systems in pattern recognition through detailed study of a particular system. This research will be carried out in collaboration with the systems theory group of the Berkeley Electrical Engineering department headed by Shankar Sastry and Pravin Varaiya who are already funded by the Air Force JSEP program for the purpose of establishing a theoretical framework for the evaluation of architectures and algorithms for parallel computation — with particular emphasis on neural networks.

APPROACH

The specific strategy, in line with the program of the systems group, is to carry out a study of the pattern recognition dynamics of a simplified version the model of the first olfactory sensory cortex developed by Freeman and shown to accurately simulate the observed temporal behavior of the olfactory EEG. This will involve detailed application of the numerical and analytic tools of dynamical systems theory, and bifurcation theory in particular — both center manifold theory and the newly emerging singularity theory of bifurcations with symmetry — in an effort to obtain the deepest possible mathematical insight into the intrinsic mechanisms of pattern recognition and learning. Particular examination will be made of the functional importance of the unusual network features contained in the natural design of this system — bifurcation, oscillation, chaos, and sigmoid gain variation.

CONTRIBUTION

This laboratory is in a unique position to exploit the historically fertile interaction between concepts of natural and artificial design, since this is the only laboratory to have demonstrated correlations between spatial patterns of cortical network activity and the pattern recognition behavior of an animal. The olfactory system is one of the best of the candidates presently being investigated for a biological system which may actually function as an "associative memory" according to the general principles of current engineering "neural networks". This allows both the evaluation of the biological validity of present theoretical network constructs, and the generation of new principles from the experimental observation of a real neural network. The biological experimentation is encouraging us to study the properties of networks with continuous dynamics in layers with crosscoupling feedback. There is little work to date emphasizing the special pattern recognition capabilities of feedback nets with smooth nonlinearities. In addition, the biological net inspires network models with properties that are quite foreign to any present engineering effort. The net is driven in and out of its pattern recognition mode through a bifurcation and works with oscillatory and possibly chaotic attractors.
RELEVANCE

The neural network model under investigating may be considered a nonsymmetrically coupled variant of the model of analog neural dynamics popularized by Hopfield. These investigations will thus be relevant to the mainstream of neural network research. The challenge of achieving a theoretical understanding of this real biological situation poses many problems of mathematical analysis which are in fact on or beyond the present frontier of network research. Examination of the complex dynamics and advantages of nonsymmetrically coupled nets is already cited as a research goal by the systems group. The prominence of bifurcation and oscillatory and possibly chaotic dynamics observed in the olfactory system, and the need to explore learning algorithms for the storage of attracting trajectories in order to model its hypothesized operation, places these efforts on the frontier of network research. In the other direction, study can be made of the biological validity in this experimental system, and the impact on the pattern recognition performance of this model of concepts already developed in the network literature such as adaptive resonance, hidden units, and higher order correlations.
Theory: Status

Experimental observation

Spatial pattern of oscillation correlates with odor input after animal conditioning

Input

\[
\begin{array}{ccc}
1 & \rightarrow & a_{11} \rightarrow 1 \\
2 & \rightarrow & a_{21} \rightarrow 2 \\
3 & \rightarrow & a_{31} \rightarrow 3 \\
\end{array}
\]

Output

\[
\begin{array}{ccc}
1 & \rightarrow & a_{12} \rightarrow 2 \\
2 & \rightarrow & a_{22} \rightarrow 3 \\
3 & \rightarrow & a_{32} \rightarrow 1 \\
\end{array}
\]

Model - Neural Network of coupled columns

Multiple Hopf bifurcation

Single equilibrium in reset state

Multiple attractors after inspiration

More attractors after arousal

gives multiple basins of attraction for periodic attractors that classify input
ACCOMPLISHMENTS

NORMAL FORM PROJECTION ALGORITHM

Introduction

A new class of learning algorithms for the storage of static and periodic attractors in recurrent analog neural networks has been developed. The method allows programming of the ambient vector field independent of the patterns to be stored. The stability of cycles and equilibria, geometry of basins of attraction, rates of convergence to attractors, and the location in parameter space of primary and secondary bifurcations can be programmed in a prototype vector field - the normal form. Fixed points on the axes of this system of nonlinear ordinary differential equations are then linearly transformed into desired spatial or spatio-temporal patterns by projecting the system into network coordinates - the standard basis - using the desired vectors as columns of the transformation matrix. This method of network synthesis is roughly the inverse of the usual procedure in bifurcation theory for analysis of a given physical system. It differs from the numerical approach of previous papers (Baird 88) in that the last stage of analysis, the calculation of normal form coefficients, is inverted as well.

In the general case, the resulting network has forth order correlations, but the use of restrictions on the detail of vector field programming and the types of patterns to be stored result in network architectures requiring only second order correlations. The network can be constructed with the usual crossbar interconnections by implementing the higher order interactions within the units. For biological modeling, where possibly the patterns to be stored are sparse and nearly orthogonal, the learning rule for periodic patterns becomes a complex outer product rule which is local, additive, commutative, and incremental. It reduces to the usual Hebb-like rule for static attractors. Networks with excitatory and inhibitory units that approximate known biological architectures such as the olfactory bulb and cortex can also be constructed by proper choice of constraints. Standing or traveling waves may be stored to mimic the different patterns of neurophysiological activity observed in the olfactory system.

Although developed from the bifurcation theory approach discussed in previous papers (Baird 88), the results on these networks are global and exact. They are not restricted to the neighborhood of any bifurcation, and can be discussed without reference to bifurcation theory. There is however a multiple bifurcation in each system at critical values of different bifurcation parameters that can be used to generate the attractors from a single equilibrium at the origin. The projection algorithm provides one solution to the problem of storing exact analog fixed points in a neural network - at a capacity of 2N attractors for an N node network. Periodic attractors which are simple cycles can be stored at a capacity of N/2, and arbitrarily complicated periodic attractors can probably be constructed from any subset of these by a kind of Fourier synthesis. If the stability of fixed points on some axes of the normal form can be transferred to an interior fixed point by vector field programming, then the temporal pattern in the network corresponding to this "mixed mode" is a linear combination of the axis patterns at their different frequencies. The vectorfield can be programmed so that the axis attractors are the only attractors in the system - there are no "spurious attractors". If symmetric normal form coefficients are chosen to program the vector field, then the normal form becomes a gradient vector field. It is exactly the gradient of an explicit potential function which is also a strict Lyapunov function for the system.
When there are symmetries chosen for the normal form equations, then the network equations are equivariant, and the network attractors are interrelated by a group of transformations. These are given from the learning process even though there are no restrictions on the patterns being learned. With such symmetries, the basins of attraction can be made into hyperplane bounded wedges. These hyperplane basin boundaries can be positioned by the learning algorithm to realize optimal clustering of inputs into categories and retrieval by nearest Euclidean distance. This is an analog network which can be restricted to categorize like a binary network of linear threshold logic units, and still produce a specified analog output. While analog behavior can usually be approximated by a sufficiently high dimensional digital system, that increased dimensionality is one of the prices paid.

An autoassociative memory like this one can always be used as a heteroassociative memory by wiring input to some of the components of the state vector and taking output from the rest. Because the basins of attraction are known exactly in this system, it will be possible to explore in detail its mapping and generalization behavior when it is used for heteroassociation. The work of James Anderson demonstrates that almost all of the applications of the feedforward backpropagation networks can be approached as well with a feedback network used in this fashion. There is however little mathematical analysis of these systems, and we expect that analysis of the present networks may reveal superior generalization performance over feedforward networks in many task domains.

A unique and important application of the projection algorithm is to construct networks for the exploration of the computational capabilities of more exotic vector fields involving, for example, quasiperiodic or chaotic flows. A neural network can be synthesized to realize any nonlinear behavior that is known in the catalogue of well analysed normal forms. The normal form theorem asserts that there is a large class of ode's and pde's which give local vector fields that are topologically equivalent to that of any normal form under consideration. This implies that there are many other possible network architectures to be discovered that can realize the performance of that produced by the projection algorithm.

**Projection Algorithm**

**Fundamental Theorem**

The central result is most compactly stated as the following theorem: Any set $S$,

\[ s = 1, 2, \ldots, N/2 \]

of cycles $rs \times e$ of linearly independent vectors of relative component amplitudes $\alpha s$ of $R$N and phases $\phi s$ of $SN$, with frequencies $\omega s$ of real $R$ and global amplitudes $\alpha s$ of $R$, may be established in the vector field of the analog third order network:

\[ X_i = -t x_i + \sum E_{Tij} x_j + \sum E_{kij} T_{ijkl} x_j x_k x_l + d_i d(t) \quad (3.1) \]

by some variant of the learning operation:

\[ T_{ij} = \sum \sum \sum \sum \sum \sum P-im P-jn P-inj \quad (3.2) \]
where the NxN matrix $P$ contains the real and imaginary components of the complex eigenvectors $x e^{\iota \theta}$ as columns, $J$ is an NxN matrix of complex conjugate eigenvalues in diagonal blocks, and $A_n$ is an Nxn matrix of 2x2 blocks of repeated coefficients of the normal form equations. The vector field of the dynamics of the global amplitudes $r_s$ and phases $\theta_s$ is given exactly by the normal form equations:

$$
\begin{align*}
\dot{r}_s &= u_s r_s - r_s E_j a_{sj} r_{j2} \\
\dot{\theta}_s &= w_s + E_j b_{sj} r_{j2}
\end{align*}
$$

(3.3)

In particular, for $a_{sk} > 0$, and $ass/aks < 1$, for all $s$ and $k$, the cycles $s = 1, 2, ..., N/2$ are stable, and have amplitudes $r_s = (us/ass)^{1/2}$, where $us = 1 - t$.

**Proof**

The proof of this is instructive since it is a constructive proof. An informal version proceeds by showing first that there are always fixed points on the axes of these amplitude equations, whose stability is given by the coefficients of the nonlinear terms. Then the network above is constructed from these equations by two coordinate transformations. The first is from polar to Cartesian coordinates, and the second is a linear transformation from canonical "mode" coordinates into the standard basis $e_1, e_2, ..., e_N$, or "network coordinates". This second transformation constitutes the "learning algorithm", because it transforms the simple fixed points of the amplitude equations into the specific spatio-temporal memory patterns desired for the network.

**Amplitude fixed points**

Because the amplitude equations are independent of the rotation $\theta$, the fixed points of the amplitude equations characterize the asymptotic states of the underlying oscillations. The stability of these cycles is therefore given by the stability of the fixed points of the amplitude equations. On each axis $r_s$, the other components $r_j$ are zero, by definition, hence

$$
\dot{r}_j = r_j \left( u_j - E_k a_{jk} r_{k2} \right) = 0, \quad \text{for } r_j = 0, \quad \text{which leaves}
$$

$$
\dot{r}_s = r_s \left( u_s - a_{ss} r_{s2} \right), \quad \text{and } r_s = 0, \quad \text{when } r_{s2} = us/ass. \quad \text{Thus we have an equilibrium on each axis } s, \quad \text{at } r_s = (us/ass)^{1/2} \text{ as claimed.}
$$

Now the Jacobian of the amplitude equations at some fixed point $r^*$ has elements
\[ J_{ij} = -2 \delta_{ij} r^{-1} r^{-j} \quad J_{ii} = u_1 - 3 \delta_{i1} r^{-1} - E_j \delta_{ij} r^{-j} \quad (3.4) \]

For a fixed point \( r^s \) on axis \( s \), \( J_{ij} = 0 \), since \( r^{-1} \) or \( r^{-j} = 0 \), making \( J \) a diagonal matrix whose entries are therefore its eigenvalues. Now \( J_{ii} = \delta_{i1} - \delta_{ii} r^{-1} r^{-s} \) for \( i \neq s \), and \( J_{ss} = u_s - 3 \delta_{ss} r^{-s} \). Since \( r^{-s} = u_s/\delta_{ss} \), \( J_{ss} = -2 u_s \), and \( J_{ii} = u_1 - \delta_{ii} (u_s/\delta_{ss}) \). This gives \( \delta_{ss} > u_1/u_s \) as the condition for all negative eigenvalues that assures the stability of \( r^s \). Choice of \( \delta_{ij}/\delta_{ii} > u_j/u_i \), for all \( i, j \), therefore guarantees stability of all axis fixed points.

Coordinate transformations

We now construct the neural network from these well behaved equations by the following transformations.

First: polar to Cartesian, \((r_s, \theta_s)\) to \((v_{2s-1}, v_{2s})\): Using

\[ v_{2s-1} = r_s \cos \theta_s \quad v_{2s} = r_s \sin \theta_s \]

and differentiating these gives,

\[ v_{2s-1} = r_s \cos \theta_s - r_s \sin \theta_s \theta_s \]

\[ v_{2s} = r_s \sin \theta_s + r_s \cos \theta_s \theta_s \]

by the chain rule. Now substituting \( \cos \theta_s = v_{2s-1}/r_s \), and

\[ r_s \sin \theta_s = v_{2s} \]

gives:

\[ v_{2s-1} = v_{2s-1}/r_s r_s - v_{2s} \theta_s \]

\[ v_{2s} = v_{2s} r_s + v_{2s-1}/r_s \theta_s \]

Entering the expressions of the normal form for \( r_s \) and \( \theta_s \), gives:

\[ v_{2s-1} = v_{2s-1}/r_s (u_s r_s + r_s E_j \delta_{sj} r_{j2}) - v_{2s} (w_s + E_j b_s j r_{j2}) \]

and since \( r_{s2} = v_{2s-1} + v_{2s2} \),

\[ n/2 \]

\[ v_{2s-1} = u_s v_{2s-1} - w_s v_{2s} + E_j [v_{2s-1} \delta_{sj} - v_{2s} b_s j] (v_{2j-12} + v_{2j2}) \]
Similarly,
\[ v_{2s} = u_s v_{2s} + w_s v_{2s-1} + E_j [v_{2s} a_{sj} + v_{2s-1} b_{sj}] (v_{2j-12} + v_{2j2}) \]

Setting the \( b_{sj} = 0 \) for simplicity, choosing \( u_s = -t + 1 \) to get a standard network form, and reindexing \( i,j = 1,2,...,N \), we get one Cartesian equivalent of the polar normal form equations.

\[ v_i = -t v_i + E_j J_{ij} v_j + v_i E_j A_{ij} v_j \]  

(3.4)

Here \( J \) is a matrix containing 2x2 blocks along the diagonal of the local couplings of the linear terms of each pair of the previous equations \( v_{2s-1}, v_{2s} \), with \( -t \) separated out of the diagonal terms. The matrix \( A \) has 2x2 blocks of identical coefficients \( a_{ij} \) of the nonlinear terms from each pair.

\[
J = \begin{pmatrix}
1 & -w_1 & & & \\
& 1 & -w_2 & & \\
& & 1 & -w_2 & \\
& & & 1 & -w_1 \\
& & & & 1
\end{pmatrix}
\]

\[
A = \begin{pmatrix}
a_{11} & a_{12} & a_{13} & & \\
& a_{11} & a_{12} & a_{13} & \\
& & a_{11} & a_{12} & a_{13} & \\
& & & a_{11} & a_{12} & a_{13} &
\end{pmatrix}
\]

Learning Transformation - Linear Term

\( J \) can be viewed as the canonical form of a real matrix with complex conjugate eigenvalues, where the conjugate pairs appear in blocks along the diagonal as shown. The equations (3.4) describe the interaction of these linearly uncoupled conjugate node pairs due to the coupling of the nonlinear terms. They can be viewed as local oscillators with frequencies chosen as \( w_i \) and nonlinear competition chosen by the \( a_{ij} \) (shunting eqns?) The second coordinate transformation introduces further off diagonal coupling of the linear terms so that each local oscillator becomes a desired spatially distributed pattern or "mode". We can interpret the normal form equations as network equations in eigenvector or "memory coordinates". In other words, we can think of them as having been produced from some network by a diagonalizing coordinate transformation \( P \), so that each axis pair \( v_{2s-1}, v_{2s} \) is a complex eigenvector basis, and \( J \) displays the complex conjugate eigenvalues of some coupling matrix \( T \) in network coordinates. In this case, for the linear term,

\[ J = P^{-1} T P \]

Then it is clear that \( T = P J P^{-1} \).

The matrix \( P \) is usually constructed of columns which are the eigenvectors calculated from the matrix \( T \) to be diagonalized. For the present purpose, we work backwards, choosing eigenvectors for \( P \) and eigenvalues
for \( J \), then constructing \( T \) by the formula above. To satisfy the
theorem, we choose as columns the real and imaginary vectors
\[ \begin{pmatrix} x \cos \alpha & x \sin \alpha \end{pmatrix} \]
of the cycles \( x e^{i\alpha} \) of the set \( S \) to be learned. Any linearly independent set of complex eigenvectors in \( P \) can be
chosen to transform the system and become the patterns "stored" in the
vectorfield of network dynamics. (diagonalization?) This "learning" process
might also be thought of as giving a vector "representation" to "Platonic"
oscillators in idealized coordinates by "projecting" out of that space onto
the vectors of the standard basis - hence the name "projection algorithm".
The transformation exchanges simple vectors or "spatial patterns" in a
complicated coordinate basis for complicated patterns in a simple coor-
dinate basis. If we write the matrix expression for \( T \) above in component
form, we recover the expression given in the theorem for \( T_{ij} \).

\[ T_{ij} = \sum m P_{im} a_{mn} P^{-1}n_j \]

Nonlinear term projection

The nonlinear terms are transformed as well, but the expression
cannot be easily written in matrix form. Using the component form of the
transformation,

\[ x_i = \sum j P_{ij} v_j \quad x_i = \sum j P_{ij} v_j \quad v_j = E_k P^{-1}j_k x_k \]

and substituting into (3.4), gives:

\[ x_i = \left( -t + 1 \right) \sum j P_{ij} (E_k P^{-1}j_k x_k) + \sum j P_{ij} E_k j_k (E_l P^{-1}k_l x_l) \]

\[ + \sum j P_{ij} (E_k P^{-1}j_k x_k) E_l a_{lj} (E_m P^{-1}m_l x_m) (E_n P^{-1}n_l x_n) \]

Rearranging the orders of summation gives,

\[ x_i = \left( -t + 1 \right) E_k (E_j P_{ij} P^{-1}j_k) x_k + E_l (E_k E_j P_{ij} j_k P^{-1}k_l) x_l \]

\[ + E_n E_m E_k (E_l E_j P_{ij} P^{-1}j_k a_{lj} P^{-1}m_l P^{-1}n_l) x_k x_m x_n \]

Finally, performing the bracketed summations and relabeling indices
gives us the network (3.1.1) of the theorem,

\[ x_i = - t x_i + \sum j T_{ij} x_j + \sum j E_{ijkl} T_{ijkl} x_j x_k x_l \]

with the expression for the tensor of the nonlinear term,

\[ T_{ijkl} = \sum m P_{im} a_{mn} P^{-1}n_j P^{-1}n_k P^{-1}n_l \]

Q.E.D.
We designed, built and tested an electronic neural network for classification of patterns that are defined with a "learning" algorithm. The non-linear dynamics of the network are motivated by experimental findings in EEG recordings of the olfactory bulb. These findings indicate that a massively parallel architecture can be utilized to best describe the bulb's dynamics. The electronic design is a digital/analog hybrid approach, utilizing the speed and flexibility of random access memory (RAM) for the storing of synaptic strengths that are modified in "learning", while still preserving the analog computational power of neural networks. This approach also includes a multiplexing scheme which decreases the number of connections, in the hardware, from order $N^2$ to order $N$, where $N$ is the number of neural units. The key operation is a bias that is induced by an step input, causing the system to switch from a low-gain equilibrium state to a high-gain limit cycle state. The output is read as a spatial pattern of oscillatory amplitudes.

Central Pattern Generating and Recognizing in Olfactory Bulb: A Correlation Learning Rule

A learning rule called an input correlation rule that can simplify exploration of the behavior of learning, generating, and classifying patterns in the vertebrate olfactory system is proposed. We apply this correlation rule to a set of fully interconnected coupled oscillators that comprises a dynamical model of the olfactory bulb so as to form "templates" of oscillators with strengthened interconnection in respect to inputs classed as "learned." We obtain a content addressable memory in which phase coherent oscillation provides for central pattern generation and recognition. We use this analog model neural network to simulate dynamic features of the olfactory bulb in detail by numerical integration and multivariate analysis. The model classifies 100\% correctly for incomplete inputs, testing inputs about their training centroids, and distortion by noise that is defined as input to nontemplate elements. The model also allows substantially overlapping templates, which implies that it possesses a large information capacity. For multiple inputs the model gives correct output of the forms A and B, A and not B, B and not A, or neither. The initial conditions of the model at the time of onset of input play no role in classification. Classification is achieved within 20 to 50 ms of simulated run time, even though convergence to a limit cycle requires up to 10 cycles (200ms). The repetition rate of convergence from one pattern to the next in the model exceeds 10 patterns/s.
Pattern Recognition in Industry

100% classification as well as grouping for artificial data

Real problem solving: the data set consists of 64 channels of phase values of 20 bad screws and 50 good screws; Using 10 screws for each group as the learning screws to change Kee, then forming 0 and 1 data.

100% classification for both groups are achieved, and the T-value is 11.129.

*The system is driven by its input, not by its initial condition.

*The information capacity is larger than N/m, N—the number of the neurons, m—the number of the "on" channels.

*The repetition rate of convergence exceeds 10 patterns/sec.

*For input and output correlation learning rule, the system behaves almost the same.

*Classification is achieved within 20 to 50 ms of simulated run time.

*The system also allows substantially overlapping templates.
MULTIPLEXING TO ADDRESS FUNDAMENTAL PROBLEM

- Biological brains are very densely connected
- This is hard to achieve in silicon
- Our multiplexing scheme solves this problem
- Reduces the number of connections from $N^2$ to $N$
Objective
To build a real time massively parallel neural network in order to achieve pattern recognition using physiological design principles

Digital/Analog Hybrid System
Has major advantages compared to conventional neural networks
- Digital Control, analog processing
- Easy and reliable manipulation of synaptic weights through digital means
- Millisecond convergence of solution through analog computation, with re-setting.
- Repetitive structure allows for easy VLSI implementation.
OUTSIDE CONTACTS

AFOSR 86-0271 (Freeman/Fregly)
Center for nonlinear dynamics
of the brain. $225,000
URIP equipment grant for computer
hardware for EEG analysis

MH 43324-01 (A.S. Gevins/Freeman)
Mass Action of the Human Neocortex.
$135,000 Three year grant from
NIMH for the analysis of human
EEG during pattern recognition

MH 06686-25 (W.J. Freeman)
Correlation of EEG and behavior
$101,000 Studies in animals
supported by NIMH

Hewlett Packard Corp. $30,000
Postdoctoral Fellowship grant
to X.-J. Wang

PUBLICATIONS

Freeman, W.J., Yao, Y., Burke, B. (1987)
Central Pattern Generating and Recognizing
in Olfactory Bulb: A Correlation Learning

Freeman, W.J., Why Neural Networks Don't Fly:
Inquiry into the Neurodynamics of
Biological Intelligence. Proc. IEEE Int.
Conf. on Neural Networks, 1988, II 1-8.

For Programming Static or Periodic
Attractors and Their Bifurcations in
Dynamic Neural Networks. Proc. IEEE Int.
Conf. on Neural Networks, 1988: I 9-16.

Baird, B (1988) A Normal Form Projection
Algorithm for Programming Analog Networks
with Dynamic or Static Attractors in
Specified Vector Fields. (in preparation)