Biol. Cybernetics 22, 85—95 (1976)

On Global Properties of Neuronal Interaction

T. J. Sejnowski*
Department of Physics, University of California at Santa Barbara, USA

Received: September 10, 1975

Abstract

Rate-coded interaction between neurons is governed by a nonlinear equation, of which the McCulloch-Pitts model for a neural network and the Hartline-Ratliff model of lateral inhibition are special limiting cases. Feature detection is treated as a variational problem and conditions are derived on the connections between neurons at a trigger feature. Perturbations from continual background activity are examined as a possible means for coding and processing information.

Introduction

A single neuron in the first stages of sensory processing responds optimally to a particular pattern of sensory stimulation; each neuron apparently detects a different characteristic feature of the sensory environment. Quite apart from what implications feature detection may have for the problem of pattern recognition, the experimental data contain valuable information about the architecture of the nervous system. One aim of this article is to examine how feature detection arises from neuronal interaction.

There is substantial spontaneous activity in neurons throughout the nervous system. Information processing appears to proceed in the form of variation from a continual background. This suggests that two types of input to a field of neurons may be distinguished, one which provides a background and another which acts as a perturbation.

The second half of this article is concerned with information processing by the perturbation of membrane potentials. The average membrane potentials of neurons, which can be defined over a much shorter time than average rates, are taken as the background.

The analysis of perturbations from a background requires that the response of not one but many neurons be examined. There are circumstances when neurons act collectively, through long-range interaction, to special patterns of input perturbation. These special patterns can be selectively modified for each background.

I. Neuronal Interaction

A nonlinear equation for stationary interaction between neurons is derived and compared with other models for neural networks.

1.

Neurons process information by the spatial summation and temporal integration of electrical activity from afferents. Some neurons have a threshold membrane potential beyond which an action potential is released; others are only capable of passive electrotonic conduction. The influence of one neuron on another takes place at synaptic junctions, which can be either excitatory or inhibitory, and either electrically or chemically mediated (Bennett, 1974).

For a neuron capable of producing an action potential, information about its internal state is contained in the spike train it produces. Although the stochastic nature of neuronal spike trains has been extensively studied (Moore, Perkel and Segundo, 1966), there is as yet remarkably little known about what types of information are encoded in the nervous system (Perkel and Bullock, 1968). In many experiments the average rate of firing is of primary concern. Although rate coding is likely to be widely exploited throughout the nervous system, other more sophisticated forms of coding are no doubt used as well.

Inasmuch as many neurons do not produce an action potential, a measure of average activity more general than rate of firing is required. One possibility which reflects the integration of excitatory and inhibitory synaptic events is the average membrane potential. In neurons which produce an action potential, however, the membrane potential above the threshold for spike generation may depart from passive integration. Therefore, let ϕ represent the average membrane potential induced in a neuron, not including the action potential.

A field of interacting neurons is in equilibrium if the average potential of each neuron is constant. Neurons

---

* Present Address: Joseph Henry Laboratories, Princeton University, Princeton, N.J. 08540, USA.

which produce an action potential fire at a constant average rate, but not necessarily at a constant rate. Assume that a neuron's average potential depends only on the average potential of all other neurons, as expressed by

$$\phi_a = \Phi_a(\phi_1, \phi_2, \phi_3 \ldots \phi_N). \tag{1}$$

An equilibrium is a fixed point of this general transformation, a particular example of which is given in the following section.

2.

Assume that the average rate of firing of a neuron r_b affects the average potential ϕ_a linearly through a connection matrix C_{ab} so that

$$\phi_a = \eta_a + \sum_b C_{ab} r_b ,$$

where η_a is the average external input, including the resting potential. Recurrent collaterals are taken into account through the diagonal terms of the connection matrix.

Since the average rate of firing of a neuron depends on its average potential, let

$$r_a = r_a^* P((\phi_a - \theta_a)/\sigma_a) ,$$

where r_a^* is the maximum rate of firing, θ_a is the threshold for firing and σ_a is a normalization. The function $P(x)$ can be interpreted as the probability that a neuron with average normalized potential x will fire during a time interval $1/r_a^*$, which is approximately the refractory period. There is a low probability of firing when the average potential is low and a high probability when the average potential is high. The firing probability distribution should therefore satisfy

1) $\lim_{x \to -\infty} P(x) = 0 .$

2) $\lim_{x \to \infty} P(x) = 1 .$

3) $P(x)$ increases smoothly and monotonically near $x = 0$ such that $P'(x) \leqq 1$.

A typical firing probability distribution is shown in Fig. 1. The distribution is normalized so that σ_a is approximately the width of the transition region between a low and high rate of firing. Although σ_a may be different for each neuron, it is convenient to set them all equal to λ. All results in this article can be generalized to arbitrary widths, as demonstrated in Appendix 1.

The transition width λ is not an intrinsic property of a neuron, but rather depends on the statistical nature of the membrane potential. The firing probability distribution is a function of the average potential and should not be confused with the threshold function of a neuron, which is much narrower and depends on the membrane potential.

The equilibrium equation for interaction between pairs of neurons

$$\phi_a = \eta_a + \sum_b K_{ab} P((\phi_b - \theta_b)/\lambda) \tag{2}$$

with an effective connection matrix

$$K_{ab} = C_{ab} r_b^*$$

has the form of the general equilibrium Eq. (1). Although the equilibrium equation was motivated by rate-coded interaction, only the average potential of each neuron enters explicitly. As shown in Appendix 1, the equilibrium equation and all results based on it can be generalized to neurons which do not produce an action potential.

Approximate equations for the average rate of firing of a neuron have been proposed which are based on diffusion models for stochastic activity (Stein, 1967; Cowan, 1971). They are formally equivalent to the equilibrium equation. A more realistic time-dependent model (Stein *et al.*, 1974) has a steady state solution which is equivalent to an equilibrium solution.

Although the average potential and average rate of firing are formally equivalent variables, they are not physically equivalent. The rate of firing of a single neuron is a coarse measure, averaged over a time long compared to the interspike intervals. The membrane potential of a neuron, on the other hand, is the result of a far greater number of synaptic events, and can be averaged over a much shorter time. Moreover, many neurons do not produce an action potential. *The membrane potential is therefore considered a more fundamental variable.*

3.

Limiting cases of the equilibrium equation are familiar models of neural networks. The probability for a neuron to fire is given by

$$p_a = P((\phi_a - \theta_a)/\lambda) .$$

Upon eliminating ϕ_a this becomes

$$p_a = P\left((\eta_a + \sum_b K_{ab} p_b - \theta_a)/\lambda\right),$$

which is formally equivalent to the equilibrium equation.

As the width of the probability distribution is decreased, it approaches a step function

$$\lim_{\lambda \to 0} P(x/\lambda) = H(x) = \begin{cases} 1 & x > 0 \\ 0 & x < 0 . \end{cases}$$

In this limit

$$p_a = H\left(\eta_a + \sum_b K_{ab} p_b - \theta_a\right),$$

which may not have any solution since $H(x)$ is discontinuous. This is the McCulloch-Pitts model (1943) for neural networks, with the following interpretation: Introduce a discrete time variable. The right hand side evaluated at some moment determines whether or not an idealized neuron fires at the next time step.

Another important limiting case of the equilibrium equation arises when the firing probability distribution is approximately linear over a range of average potential near threshold. Equivalently, this approximation can be thought of as the limit $\lambda \to \infty$. Expand $P(x)$ near $x=0$ and retain terms only up to first order. Since

$$r_a = r_a^* p_a$$

the linearized equation can be rewritten in the form given by Hartline and Ratliff (1957) for lateral inhibition in the Limulus eye

$$r_a = e_a - \sum_b K_{ab}^I (r_b - r_b^0),$$

where, according to their terminology, e_a is the external stimulus, K_{ab}^I are the inhibitory coefficients, and r_b^0 is the threshold rate for inhibition. These are linearly related to the variables η_a, K_{ab}, and θ_b of the equilibrium equation.

II. Expansion Around Equilibrium

The expansion of the equilibrium equation given here is used in Part III to find variational conditions, in Part IV to examine the stability of an equilibrium solution, and in Part V to study background perturbations.

4.

Let us prove that for sufficiently large λ there is a unique solution to the equilibrium equation. Consider the map

$$F\phi = \eta + KP((\phi - \theta)/\lambda),$$

for which a fixed point

$$\phi = F\phi$$

is an equilibrium solution.

Take any two vectors $\phi_1, \phi_2 \in \mathbb{R}^N$ and consider

$$\|F\phi_1 - F\phi_2\| \leqq \|K\| \, \|P((\phi_1 - \theta)/\lambda) - P((\phi_2 - \theta)/\lambda)\|$$
$$\leqq \frac{1}{\lambda} \|K\| \, \|\phi_1 - \phi_2\|,$$

where the last step follows from the normalization condition on $P(x)$. If $\|K\| < \lambda$ then the map contracts, and by the contraction mapping theorem there is a unique solution to the equilibrium equation. In fact, as shown in Appendix 2, the equilibrium equation always has at least one solution.

Since λ is the transition width between a low and high rate of firing, the condition $\|K\| < \lambda$ means that no single connection can dominate the rate of firing of any neuron. This case will be called weak coupling.

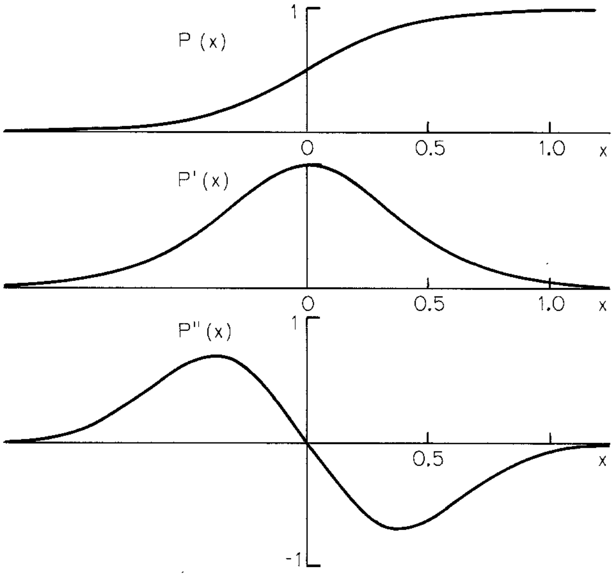

Fig. 1. Typical probability distribution for neuron firing and its derivatives as a function of normalized average potential

5.

Consider the equilibrium equation as a map of η_a into ϕ_a. For weak coupling between neurons, which is assumed throughout this part, and for a sufficiently smooth $P(x)$, the equilibrium equation is a diffeomorphism. Let the average external input $\eta_a(u)$ vary along a smooth curve in $\mathbb{R}^N$ parameterized by u. Expand the image of the input curve $\phi_a(u)$ around an equilibrium solution $\phi_a(0)$ so that

$$\phi_a(u) = \phi_a(0) + u\phi_a'(0) + \frac{u^2}{2}\phi_a''(0) + \dots .$$

The first order variation ϕ_a' satisfies

$$\phi_a' = \eta_a' + \frac{1}{\lambda}\sum_b K_{ab}' \phi_b', \tag{3}$$

where

$$K_{ab}' = K_{ab} P'((\phi_b(0) - \theta_b)/\lambda).$$

As shown in Fig. 1, the derivative of the firing probability distribution is only significant when the average potential of a neuron is near threshold. Neurons in this critical region are the ones most sensitive to input perturbations; the interactions between such critical neurons dominate the equation of first variation, or perturbation equation. The number of critical neurons for most backgrounds is likely to be small in comparison with all the neurons in the field, at least for those areas in which there is a low spontaneous activity. Consequently, the interaction matrix K'_{ab} may be only a small subset of all connections between neurons.

Some synapses are known to habituate upon repeated stimulation. When habituation moves the average potential of a neuron closer to threshold (for example, by decreasing the strength of excitatory afferents when the neuron is firing at maximum rate), the neuron's sensitivity to changes of the average potential is enhanced, and the neuron makes a larger contribution to the interaction matrix.

Let the solution of the interaction equation be given by

$$\phi'_a = \sum_b R_{ab} \eta'_b ,$$

where R_{ab} is called the interaction resolvent. The Neumann series

$$R_{ab} = \delta_{ab} + \frac{1}{\lambda} K'_{ab} + \frac{1}{\lambda^2} \sum_c K'_{ac} K'_{cb} + \ldots \tag{4}$$

converges when

$$\|K'\| < \lambda .$$

This expansion is a sum over interaction paths between critical neurons.

The second order variation of the equilibrium equation is

$$\phi''_a = \left(\eta''_a + \sum_b K''_{ab} \phi'^2_b \right) + \frac{1}{\lambda} \sum_b K'_{ab} \phi''_b ,$$

where

$$K''_{ab} = K_{ab} P''((\phi_b(0) - \theta_b)/\lambda) .$$

The solution of the second order equation is given by

$$\phi''_a = \sum_b R_{ab} \left(\eta''_b + \sum_c K''_{bc} \phi'^2_c \right) ,$$

which depends on the solution of the first order equation.

As shown in Fig. 1, those neurons for which $P''(x)$ is appreciable are either just above or just below threshold. They will be called border neurons and K''_{ab} will be called the border matrix. The critical and border neurons are nearly disjoint subsets of all neurons.

The n-th order variation equation has the same form as that of the first and second order. The solution is given by

$$\phi^{(n)}_a = \sum_b R_{ab} (\eta^{(n)}_b + L^{(n)}_b(\phi', \phi'', \ldots \phi^{(n-1)})) ,$$

where $L^{(n)}_b$ is a function of the previous $n-1$ variations. The nonlinear equilibrium equation can therefore be replaced by an infinite series of linear equations, the n-th order depending only on the solution of the preceding $n-1$ orders. The first few terms of the expansion suffice to explore local properties and small perturbations.

III. Feature Detection

The concept of a feature detector has proven helpful in understanding the response of single neurons to sensory stimuli. An analytic description of feature detection is developed in this part. Only the static case is considered, for which neither the connection matrix nor the external stimulus varies with time.

6.

The receptive field of a neuron is the set of sensory receptors which significantly affects its rate of firing. The stimulus pattern within the receptive field which is most effective at eliciting a response from a neuron is called its trigger feature. There is a collection of interconnected maps of a sensory field within the brain; the neurons in each representation possess different characteristic trigger features. In the visual system, for example, trigger features progress from spots of light to edges and slits as the sensory field ascends from the retina to visual cortex. Parameters of a feature to which neurons in striate cortex are known to be sensitive include position, size, orientation, direction of movement, occularity, disparity and color contrast.

The response of a feature detector is sensitive to some stimulus parameters, which produce a significant decrement when varied around the trigger feature, while other parameters leave the response invariant. For example, complex cells in striate cortex are tuned to an optimal slit width and orientation, but give a constant response as the slit moves across the receptive field in a particular direction.

7.

Let $\mathscr{A}$ be the set of all neurons in a field and $\mathscr{B}$ the subset which receive external input. As in the previous part, $\phi_a(u)$ is the average potential of the a-th neuron along a curve in $\mathbb{R}^N$ parameterized by u.

The receptive field of a neuron is the subset of sensory input $\mathscr{B}_a \subset \mathscr{B}$ which affects its response. The trigger feature of a neuron is experimentally determined by varying stimulus parameters to optimize the rate of firing. Define a local trigger feature as a sensory stimulus $\eta_b(0)$ such that

$$\phi_a(0) \geqq \phi_a(u)$$

for all nearby $\eta_b(u)$ in the receptive field which pass through the trigger feature. Feature detection has the form of a variational problem.

The results of the preceding part on expansion of the equilibrium equation provide necessary and sufficient conditions for a stimulus to be a trigger feature. The average potential $\phi_a(0)$, and hence the response of a neuron, is an extremum if and only if $\phi'_a(0)=0$ for all variations around $\eta_b(0)$. Since

$$\phi'_a = \sum_b R_{ab}\eta'_b$$

and η'_b is arbitrary in the receptive field $\mathscr{B}_a$, the extremum condition implies that

$$R_{ab}=0\,, \quad b\in\mathscr{B}_a\,.$$

An extremum may be a maximum, minimum, or saddle point depending on the sign of the second order variation around the trigger feature. There is a maximum if

$$\phi''_a = \sum_{b\in\mathscr{B}_a} R_{ab}\left(\eta''_b + \frac{1}{\lambda^2}\sum_c K''_{bc}\phi'^2_c\right) \leqq 0\,.$$

The first term on the right vanishes by virtue of the condition on the first order variation. Also, since $\phi' = R\eta'$, the second order condition becomes

$$\sum_{kl} F^a_{kl}\eta'_k\eta'_l \leqq 0\,,$$

where

$$F^a_{kl} = \sum_{bc} R_{ab}K''_{bc}R_{ck}R_{cl}\,.$$

8.

Variations from a trigger feature decrease a neuron's rate of firing, more so in some directions than others. The directions of maximum and minimum decrement lie in the tangent space of variations η'_a, and can be determined from the quadratic form in the second variation condition.

The eigenvectors ξ^b of the symmetric matrix F^a_{kl} satisfy

$$\sum_l F^a_{kl}\xi^b_l = \nu_b\xi^b_k$$

and form a basis in which the matrix is diagonal. The second variation condition in this basis is

$$\sum_b \nu_b(\eta', \xi^b)^2 \leqq 0\,,$$

from which it follows that a necessary and sufficient condition for a maximum is

$$\nu_b \leqq 0$$

for

$$(\eta', \xi^b) \neq 0\,.$$

That is, the spectrum of F^a_{kl} in the receptive field must be non-positive. (For an inhibitory feature the spectrum in the receptive field must be non-negative.)

The eigenvectors of the most negative eigenvalues in the spectrum are those directions along which the response of a neuron is most sensitive to variations from a trigger feature. For example, the orientation of an edge parameterizes such a direction for complex cells in striate cortex. Similarly, eigenvectors with eigenvalues near zero give the directions which leave the response of a neuron invariant.

IV. Strong Coupling and Stability

The equilibrium equation has a unique solution for weak coupling between neurons, a result which may not hold when the coupling is strong. The characteristics of strong coupling are examined in this part.

9.

A topological fixed point theorem is used in Appendix 2 to prove that the equilibrium equation always has at least one solution for arbitrary η, θ, K, and $\lambda \neq 0$. Although the result is remarkably general, there is no suggestion, as in the case of the contraction mapping theorem, of how to construct a solution.

More than one solution of the equilibrium equation may exist for a given η when the weak coupling condition is not satisfied. There is, however, always a unique η for every ϕ. The example in Appendix 3 demonstrates how multiple solutions arise as the coupling is increased, or alternatively, the transition width λ is decreased.

Ambiguous figures, such as the Necker cube, can be perceived in more than one way. The striking shift of perception may be due to a stage of visual processing with more than one stable equilibrium (Wilson and Cowan, 1972).

The strength of coupling separates solutions of the equilibrium equation into two qualitatively different classes. Local interactions dominate when coupling is weak, whereas long-range interactions become significant for strong coupling. The difference is reflected in the interaction path expansion (4): As the transition width is decreased, longer chains of neurons become increasingly more important. Long-range is not used here is the sense of physical distance, but rather in the sense of distance through synapses.

Feature detection was based on the existence of a unique solution for every input, which may not be the case when coupling is strong. Long-range interaction requires that neurons be examined not individually, but as a coordinated ensemble.

The example in Appendix 3 demonstrates that as external inputs are varied, new families of solutions may appear and others disappear. Bifurcation theory is concerned with the conditions under which nonlinear equations exhibit such behavior.

The spectrum of the first order equation of variation contains information on the location of bifurcation points. The perturbation Eq. (3) can be written

$$A\phi' = \eta' ,$$

where

$$A = I - \frac{1}{\lambda} K' .$$

The solutions of the characteristic equation $\det A = 0$ are the eigenvalues of K', to which there correspond eigenvectors that satisfy

$$K'\psi^n = \lambda_n \psi^n .$$

The interaction resolvent $R = A^{-1}$ exists if and only if λ is not one of the eigenvalues. Consequently, if $\lambda \neq \lambda_n$ then there is a unique solution to the perturbation equation.

If there is a bifurcation point of the equilibrium equation at (λ, ϕ) then λ is an eigenvalue of the interaction matrix. If λ is an eigenvalue with odd multiplicity, then it must be a bifurcation point (Nirenberg, 1974). There may or may not be a bifurcation point when the multiplicity is even.

10.

An equilibrium solution is stable if small displacements from equilibrium are restored. A time-dependent equation for the average potential, which will be examined in more detail in a forthcoming article, is given by

$$\tau \frac{d}{dt} \phi + \phi = \eta + KP((\phi - \theta)/\lambda) ,$$

where τ is the membrane time constant. An equilibrium solution satisfies

$$\frac{d}{dt} \phi^0 = 0 .$$

If a perturbation ϕ' is added so that

$$\phi = \phi^0 + \phi'$$

then

$$\tau \frac{d}{dt} \phi' + A\phi' = \eta' + O(\|\phi'\|^2) .$$

The solution of the linearized equation is asymptotically stable if and only if

$$\mathrm{Re}\, \lambda_n < \lambda .$$

That is, the spectrum of the interaction matrix must lie to the left of the transition width for an asymptotically stable equilibrium.

11.

The interaction resolvent has poles on the spectrum of the interaction matrix (Dunford and Schwartz, 1958). An explicit spectral resolution of the resolvent is given here and interpreted in the next part.

It is always possible to find a similarity transform which reduces an arbitrary matrix to Jordan form

$$J' = S^{-1} J S ,$$

where J' has entries on the diagonal or just above it. Furthermore, the diagonal entries are the eigenvalues of the matrix while above the diagonal only 0 or 1 appears. For the purpose of dimensional scaling, it is convenient to have λ rather than 1 above the diagonal. Thus, J' is a direct sum of blocks in the form

$$\begin{pmatrix} \lambda_n & \lambda & & 0 \\ & \lambda_n & \lambda & \\ & & \ddots & \ddots \\ & & & \ddots & \lambda \\ 0 & & & & \lambda_n \end{pmatrix} .$$

Several blocks may have the same eigenvalue. Let the l-th block of λ_n have dimension m_{nl}. The multiplicity of λ_n is

$$m_n = \sum_l m_{nl} .$$

Define ψ_m^{nl} as the m-th column in the (n, l)-block of a matrix S which transforms the interaction matrix to Jordan form. Similarly, let ψ_m^{*nl} be the m-th row of S^{-1}. These vectors satisfy the following equations for generalized eigenvectors:

$$(\lambda_n I - K')^m \psi_m^{nl} = 0$$

$$\psi_m^{*nl}(\lambda_n I - K')^{m_{nl}-m+1} = 0\,.$$

The eigenvectors of K' are the ψ_m^{nl} with $m=1$ and the eigenvectors of the adjoint K'^* are the ψ_m^{*nl} with $m=m_{nl}$.

By virtue of their definition the two sets of generalized eigenvectors are dimensionless and biorthonormal

$$(\psi_m^{*nl}, \psi_{m'}^{n'l'}) = \delta_{nn'}\delta_{ll'}\delta_{mm'}\,,$$

but are not themselves necessarily orthonormal.

A spectral resolution of the interaction resolvent is obtained by inverting each block of the interaction equation in a biorthonormal system. The result is

$$R = \sum_{nl} \sum_{\substack{m,m' \\ m' \geqq m}}^{m_{nl}} \frac{\psi_m^{nl} \otimes \psi_{m'}^{*nl}}{\left(1 - \dfrac{\lambda_n}{\lambda}\right)^{m'-m+1}}\,. \tag{5}$$

The interaction path expansion of the interaction resolvent

$$R = \sum_{n=0}^{\infty} (K'/\lambda)^n$$

converges if λ is greater than the spectral radius of the interaction matrix

$$r(K') = \max|\lambda_n|\,.$$

V. Information Processing

The nervous system most likely processes and stores information through some form of collective activity. The global properties of neuronal interaction which were analyzed in the previous part are here given an interpretation.

12.

The considerable level of spontaneous activity in neurons throughout the nervous system has often been noted. The continual background may provide a context against which variations in activity are analyzed.

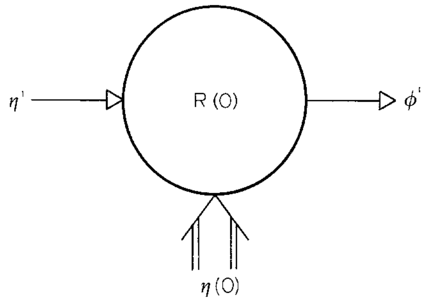

Fig. 2. The nonlinear equilibrium equation determines the background activity of neurons according to the background input $\eta(0)$. Input perturbations η' are transformed by the interaction resolvent $R(0)$ into output perturbations ϕ' relative to the background

Distinguish two sources of input to a field of neurons, one which produces a background, and another which acts as a perturbation. The two may arise from different classes of input afferents or from the modulation of only one set. The perturbation input must be modulated in any case to separate it from the more slowly varying background. For example, bursts of firing or firing in a correlated pattern may serve as forms of perturbation. The modulation of background activity may be related to the prominent rhythms on a gross scale in EEG recording and to the observed regularity of evoked potentials.

The interaction resolvent maps input perturbations into output perturbations, as illustrated in Fig. 2. The dominant poles in the spectral resolution of the interaction resolvent (5) are from those eigenvalues of the interaction matrix nearest to the transition width. A background near a bifurcation is particularly sensitive to perturbations, but is also nearly unstable.

The high order poles at an eigenvalue dominate when

$$\left|1 - \frac{\lambda_n}{\lambda}\right| < 1\,,$$

which is the interior of a circle of radius λ centered at λ on the real axis. The low order poles are more important outside the circle. The eigenvectors of the interaction matrix are attached to the highest order pole in each block.

The output perturbation is a particular combination of generalized eigenvectors ψ_m^{nl}. If one of these special patterns or features is present in the input, only features with the same eigenvalue are produced in the output. To be more precise, define the supspace

$$\mathscr{F}^n = (\psi_m^{nl})_{m=1,2\,.\,m_{nl}}^{l=1,2\,\ldots}$$

Then $\mathscr{F}^n$ is an invariant subspace of the resolvent

$$R:\mathscr{F}^n\rightarrow\mathscr{F}^n .$$

The map of input features into associated output features, which will be called feature association, involves many neurons acting collectively through long-range interaction.

The output perturbation ϕ' of one field of neurons is reflected as an input perturbation η' to another field according to

$$\eta'=K'\phi' ,$$

where K' is the interaction matrix linking the two fields. Not all the information in ϕ' is transmitted since only efferents from critical neurons contribute to η'.

13.

Several components from several modalities may participate in the background. Combinations of input produce a complex variety of solutions to the nonlinear equilibrium equation. The general solution of the perturbation equation with input perturbations from several sources is a linear superposition of solutions.

The mode of information processing is different for background and perturbation inputs. Sensory input when applied as a background effectively picks out a subset of connections for the interaction matrix. The resulting features reflect the symmetries of the sensory pattern. If, on the other hand, sensory input is applied as a perturbation, output associations are produced relative to the context determined by the background. A sensory pattern can be analyzed for different features, in a manner not unlike that of attention, by changing the background.

The symmetries of the interaction matrix are reflected in the structure of the interaction resolvent. Assume that the interaction matrix is invariant under a group G of transformations

$$K'=D(g)K'D(g)^{-1} , \qquad g\in G .$$

The transformation $D(g)$ mixes only generalized eigenvectors $\{\psi^n\}$ corresponding to the same eigenvalue λ_n. Hence

$$D(g)\psi^n_a=\sum_b^{m_n} D^n_{ba}(g)\psi^n_b ,$$

and the matricies $D^n(g)$ form a representation of the group with basis vectors $\{\psi^n\}$ for each n.

Features contain information about the symmetry of the connection matrix and background input. Analysis of background perturbations rather than the background itself could lead to a classification of patterns on the basis of their symmetry.

The interaction matrix determines the features of a background. If the critical neurons are only a small subset of all the neurons, then the features for that background may be altered without significantly affecting the features of other backgrounds. Of course, interference between backgrounds becomes more serious as more are developed, thereby limiting the number which are effectively independent.

It is well-known that human memory is associative and that context is an important variable. Moreover, new information can be selectively stored without destroying information already in memory. These characteristics of long-term memory are shared by feature association.

Models for memory have been proposed (Steinbuch, 1961; Anderson, 1972; Kohonen, 1972) which are similar to feature association. These models are based on linear maps of the form

$$T=\sum_n p^n\otimes q^n .$$

The spectral resolution of the interaction resolvent (5), which has a physiological basis, is a special class of such linear maps. Nonlinear interaction permits many different backgrounds to be embedded in the same field of neurons, thereby giving feature association a contextual structure.

VI. Information Storage

This part examines how plastic change to the connection matrix affects feature association. The strength of a connection may depend on many physiological variables, such as the characteristics of synapses, dendritic spines, and dendritic processes.

14.

The interaction matrix depends on both the strength of connection between critical neurons and on the background

$$K'_{ab}=K_{ab}P'((\phi_b(0)-\theta_b)/\lambda) .$$

When the strength of a connection is changed, both factors contribute to the first order perturbation

$$\delta K'_{ab}=\delta K_{ab}P'_b+\frac{1}{\lambda}K''_{ab}\delta\phi_b ,$$

where

$$P'_b=P'((\phi_b(0)-\theta_b)/\lambda) .$$

The first contribution is due to change in the connections from critical neurons. The second term, owing to the slight shift in the background caused by δK_{ab},

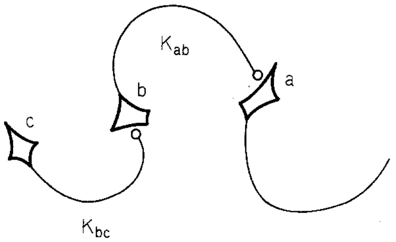

Fig. 3. Neuron b makes efferent connections K_{ab} and has afferent connections K_{bc}. Modifications of the interaction matrix K'_{ab} depend mainly on efferent connections if b is a critical neuron, and afferent connections if b is a border neuron

gives the migration between critical neurons and border neurons. The shift can be found from the equilibrium equation

$$\delta\phi_a = \sum_d \delta K_{ad} P_d + \frac{1}{\lambda} \sum_b K'_{ab} \delta\phi_b \,,$$

the solution of which is

$$\delta\phi_b = \sum_{cd} R_{bc} \delta K_{cd} P_d \,.$$

The term arising from border neurons therefore becomes

$$\frac{1}{\lambda} K''_{ab} \sum_{cd} R_{bc} \delta K_{cd} P_d \,.$$

If the interaction path expansion (4) is used for R_{bc}, then the leading term to the border contribution is

$$\frac{1}{\lambda} K''_{ab} \sum_d \delta K_{bd} P_d + O\left(\frac{1}{\lambda^2}\right).$$

As shown in Fig. 3, the first term of $\delta K'_{ab}$ represents plastic change of the efferent connections from a critical neuron, whereas the dominant part of the second term is given by plastic change of afferent connections to a border neuron. Different mechanisms for the alteration of connections may exist to exploit the two possibilities.

15.

The interaction resolvent depends on the interaction matrix

$$R(\lambda, K') = \left(I - \frac{1}{\lambda} K'\right)^{-1}$$

and is meromorphic on the λ-plane. Furthermore, if the interaction matrix depends analytically on a parameter, $K'(\varepsilon)$, then the resolvent is analytic for sufficiently small ε, except for λ on the spectrum of $K'(0)$ (Kato, 1966).

Decompose $K'(\varepsilon)$ into an unperturbed part and a perturbation

$$K'(\varepsilon) = K_0 + \kappa(\varepsilon)$$

with

$$\kappa(0) = 0 \,.$$

Rewrite

$$A = I - \frac{1}{\lambda} K'$$

in the form

$$A = A_0 \left(I - \frac{1}{\lambda} R_0 \kappa(\varepsilon)\right).$$

Take the inverse of both sides to obtain

$$R(\varepsilon) = \left(I - \frac{1}{\lambda} R_0 \kappa(\varepsilon)\right)^{-1} R_0 \,.$$

If ε is chosen sufficiently small so that

$$\|\kappa(\varepsilon)\| < \lambda / \|R_0\|$$

then the interaction resolvent has the expansion

$$R(\varepsilon) = R_0 + \frac{1}{\lambda} R_0 \kappa R_0 + \frac{1}{\lambda^2} R_0 \kappa R_0 \kappa R_0 + \dots$$

which makes explicit the analyticity of the resolvent.

The eigenvalues and eigenfunctions of $K'(\varepsilon)$ are not, in general, analytic functions. There is, however, a perturbation expansion when K_0 is semisimple

$$K_0 = \sum_a \lambda_a \psi_a \otimes \psi_a^*$$

and there are no degenerate eigenvalues. Then to first order:

$$\delta\lambda_a = (\psi_a^*, \kappa\psi_a)$$

$$\delta\psi_a = \sum_{b \neq a} \psi_b \frac{(\psi_b^*, \kappa\psi_a)}{\lambda_a - \lambda_b}$$

$$\delta\psi_a^* = \sum_{b \neq a} \frac{(\psi_a^*, \kappa\psi_b)}{\lambda_a - \lambda_b} \psi_b^* \,.$$

The spectrum is shifted according to the diagonal components of κ in an eigenvector basis. The off-diagonal elements give the mixing of features between eigenvalues, the closest eigenvalues making the largest contribution. Feature association can be selectively modified for each background by discreetly altering the interaction matrix.

Discussion

How might the preceding treatment of neuronal interaction be applied to practical problems in neurobiology? Not without considerable difficulty, for the present idealization falls far short of the extraordinary richness of the nervous system. Realistic detail was sacrificed in order to study neuronal interconnections; the main contribution of this article is at the global level of organization.

All information about the external world available to an organism is derived, at one time or another, from patterns of activity on sensory receptors; all actions in the world are effected by patterns of motor activity. A sensory pattern may have special significance for some animals and trigger a stereotyped motor response. Internal variables, which may depend on experience, serve to coordinate and temper behavior. In man, the perception of special patterns is augmented by a highly developed long-term memory.

The diversity of behavior among living creatures is most likely due to an equal diversity of design principles. Feature association, one such possible design principle, has two basic parts: first, a background which determines the features and serves as a context, and second, the map of input features into associated output features. Feature association is embedded in the nonlinear interaction between neurons and can be selectively modified for each background.

Memory, from a global point of view, is a collective response of many neurons. However suggestive, feature association cannot be taken as a serious model of memory until there is evidence for the modulation and detection of perturbations from the average background activity. One class of afferents to an area may contribute only to the background while another may act only as a perturbation. Alternatively, a single input may be modulated, forming both background and perturbation.

Although feature detection and feature association merit study in their own right as techniques for information processing, there remains the question of how far the results of this article can be extended. Time-dependent neuronal interaction will be treated elsewhere. Despite the formidable complexity of the brain, there is no compelling reason why its design principles must be as complicated.

Acknowledgement. Discussions with David Bender have been a source of clarity and inspiration. I am indebted to Professor Charles Gross for his support and encouragement.

Appendix

1. Generalizations

The equilibrium equation can be generalized for neurons with different transition widths, for neurons which do not produce an action potential, and for arbitrary branching of input afferents.

The equilibrium equation for neurons with unequal transition widths is

$$\phi_a = \eta_a + \sum_b K_{ab} P((\phi_b - \theta_b)/\sigma_b) .$$

It is convenient to scale all widths

$$\sigma_b = \lambda \Gamma_b$$

by the common factor λ. The variation equations in Part II remain valid if the variation matricies are redefined as

$$K_{ab}^{(n)} = K_{ab} P_b^{(n)} / \Gamma_b^n .$$

The expansion of the equilibrium equation and the structure of feature detection and feature association are the same as before.

Further generalization of the equilibrium equation can be made for an arbitrary distribution $P_b(x)$, different for each neuron, which need no longer be interpreted as the firing probability. The equation is then sufficiently general to include any interaction between a pair of neurons depending only on average potential. Neurons which do not produce an action potential are included.

The afferents to an area may branch, thereby innervating more than one neuron in the field. Let M afferent fibers branch and connect with N neurons according to an $N \times M$ branching matrix B. All results follow as before with η everywhere replaced by $B\eta$.

2. Existence of a Solution

A fixed point theorem is used to prove the existence of a solution to the equilibrium equation (Smart, 1974).

Theorem (Brouwer). *Every continuous mapping of a closed n-ball into itself has a fixed point.*

Define

$$\phi^* = \phi - \eta .$$

The equilibrium equation can be written as the fixed point

$$\phi^* = F\phi^* = KP((\phi^* + \eta - \theta)/\lambda)$$

of a bounded operator

$$\|F\| \leqq \|K\| .$$

Choose ϕ^* in the ball

$$\mathscr{B} = \{\phi^* \mid \|\phi^*\| \leqq \|K\|\} .$$

Then

$$F : \mathscr{B} \to \mathscr{B}' \subset \mathscr{B}$$

and the map is continuous for $\lambda \neq 0$. By the above theorem, F has a fixed point and the equilibrium equation always has at least one solution.

3. The Two Neuron Model

The graphical method used by Wilson and Cowan (1972) can be applied to the two neuron model. The firing probability of a neuron is given by

$$p_a = P((\phi_a - \theta_a)/\lambda)$$

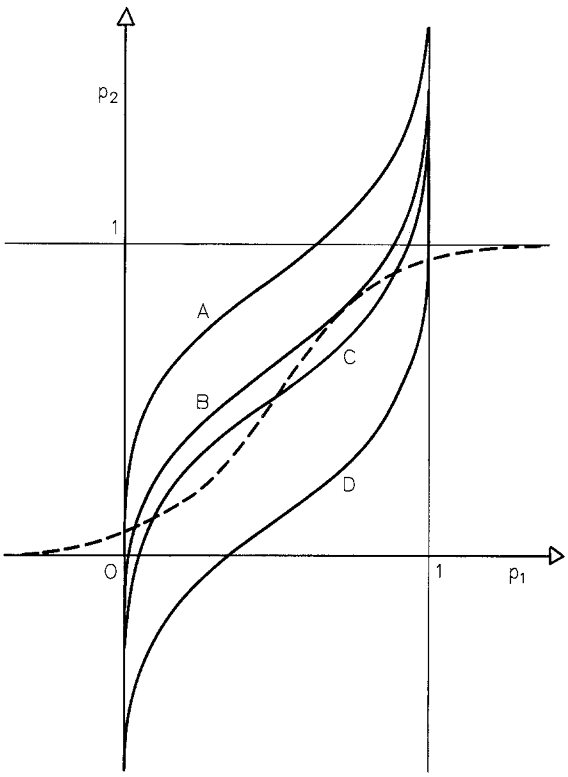

Fig. 4. Solutions for the two neuron model are given by the intersections of $p_1(p_2)$ (broken line) and $p_2(p_1)$ (solid lines), shown here for the case of reciprocal excitation without recurrent collaterals. A variety of equilibrium solutions is exhibited by the four different values of η_1 chosen

which, when inverted, becomes

$$\eta_a + \sum_b K_{ab} p_b - \theta_a = \lambda P^{-1}(p_a).$$

For $N=2$ the pair of equations is

$$K_{12} p_2 = \theta_1 - \eta_1 + \lambda P^{-1}(p_1) - K_{11} p_1$$

$$K_{21} p_1 = \theta_2 - \eta_2 + \lambda P^{-1}(p_2) - K_{22} p_2$$

which gives $p_2(p_1)$ and $p_1(p_2)$. The intersection of these two equations is a graphical solution, as illustrated in Fig. 4.

For large λ there is clearly one and only one solution. As λ decreases, there is a point at which two solutions are possible and beyond which there may be more, depending on η_a. The new solution which appears on curve *B* of Fig. 4 is a bifurcation point. Of the three solutions shown on Curve *C*, the outer two are stable equilibria while the inner solution is unstable.

References

Anderson, J. A.: A simple neural network generating an interacting memory. Math. Biosci. **14**, 197—220 (1972)

Bennett, M. V. L. (Ed.): Synaptic transmission and neuronal interaction. New York: Raven Press 1974

Cowan, J. D.: Stochastic models of neuroelectric activity. In: Rice, S., Light, J., Freed, K. (Eds.): Proceedings of the international union of pure and applied physics conference on statistical mechanics, pp. 109—127. Chicago: University of Chicago Press 1971

Dunford, N., Schwartz, J. T.: Linear operators, Part I. General theory. New York: Interscience 1958

Hartline, H. K., Ratliff, F.: Inhibitory interaction of receptor units in the eye of the Limulus. J. gen. Physiol. **40**, 357—376 (1957)

Kato, T.: Perturbation theory for linear operators. Berlin-Heidelberg-New York: Springer 1966

Kohonen, T.: Correlation matrix memories. IEEE Trans. Computers **21**, 353—366 (1972)

McCulloch, W. S., Pitts, W. H.: A logical calculus of ideas immanent in nervous activity. Bull. math. Biophys. **5**, 115—133 (1943)

Moore, G. P., Perkel, D. H., Segundo, J. P.: Statistical analysis and functional interpretation of neuronal spike data. Ann. Rev. Phys. **28**, 493—522 (1966)

Nirenberg, L.: Topics in nonlinear functional analysis. New York: Courant Institute of Mathematical Sciences 1974

Perkel, D. H., Bullock, T. H.: Neural coding. Neurosci. Res. Progr. Bull. **6**, No. 3 (1968)

Smart, D. R.: Fixed point theorems. Cambridge: Cambridge University Press 1974

Stein, R. B.: The frequency of nerve action potentials generated by applied currents. Proc. roy. Soc. Ser. B **167**, 64—86 (1967)

Stein, R. B., Leung, K. V., Oguztoreli, M. N., Williams, D. W.: Properties of small neural networks. Kybernetik **14**, 223—230 (1974)

Steinbuch, K.: Die Lernmatrix. Kybernetik **1**, 36—45 (1961)

Wilson, H. R., Cowan, J. D.: Excitatory and inhibitory interactions in localized populations of model neurons. Biophys. J. **12**, 1—24 (1972)

T. J. Sejnowski
Princeton University
Dept. of Physics
Joseph Henry Laboratories
P.O. Box 708
Princeton, N.J. 08540, USA