Hebbian Neural Networks which Topographically Self-Organize

Olivier C. Martin and Juan Carlos Letelier

[1] Division de Physique Théorique#,
Institut de Physique Nucléaire, Orsay Cedex 91406, France

[2] Departamento de Biologia, Facultad de Ciencias,
Universidad de Chile, Casilla 653, Santiago, Chile

Abstract

We show that certain modifications of Hebbian synaptic dynamics solve the problem of epigenetic formation of bicontinuous maps or neurotopies between two layers of neurons. Our models of topographic self-organization are biologically motivated, and incorporate competition between all the synapses of a given neuron. This allows for a wide range of arborization and receptive field sizes.

Unité de recherche des Universités Paris XI et Paris VI, associée au CNRS

Introduction

A well established fact in neural systems is that anatomical interconnections between two neuronal nuclei or layers are usually topographically ordered, i.e., nearby neurons in one layer connect, via axons and synapses, to nearby neurons in the other. The best examples of these topographically ordered connections, called neurotopies, are the retinotopy (projection from the retina to the visual cortex) and the somatotopy (projection from the skin to the sensory cortex). A neurotopy forms a topographically ordered bicontinous map from layer 1 (input layer) to layer 2 (the "processing" layer). Experimentally, a neurotopy is detected by the following facts: a) each neuron in layer 2 responds to stimuli from a subset of neurons in layer 1 forming the "receptive field"; b) each receptive field is localized; c) if the geometrical center of the receptive field of a cell at position (x,y) is denoted by (a^*, b^*), the map $\phi(x,y) = (a^*(x,y), b^*(x,y))$ is bicontinous, i.e., ϕ and its inverse are continuous.

Two mechanisms, preformism and epigenesis, have been advanced to explain the formation of these maps. The oldest theory, a preformist concept due to Sperry (Sperry, 1963), hypothesizes that the map is produced by a set of chemical markers in layer 2 which are under genetic control and which direct the neural fibers from layer 1 to their "correct" targets. However this theory has two important weaknesses. First, it requires the existence of a family of marker molecules which have not been detected. Second, it does not explain the reorganization (plasticity) of neurotopies found in the adult after disruption of a layer (Scharma, 1972). A refinement of Sperry's theory replaces the family of markers by just a few chemical species, called morphogens, which create through concentrations gradients an implicit reference frame in layer 2. These models have had some success in explaining experimental data, but they are still rather ad-hoc. In a completely different approach, a number of workers (Kohonen 1982), (Linsker, 1989), (Ritter and Schulten, 1986), have exhibited computer algorithms which will topographically order an initial map from layer 2 to layer 1 without using the existence of an implicit or preformed map. While these computer models embody the central idea of epigenesis, creation of order from a random or disordered state as the result of the application of a few rules, they have no biological basis: a) they usually don't deal with a network of neurons with synapses; b) they require isolated pointwise stimulii; and c) they use an external global ranking mechanism.

There are some epigenetic models which try to induce the self-organization of connections using local and biologically plausible rules only (Amari, 1980), (Willshaw and von der Malsburg, 1976), (Zhang, 1991). These attempts have used Hebb's rule (Hebb, 1949), in which the synaptic interconnection strength varies with the correlation of pre and post-synaptic events. In these models some amount of topographic ordering occurs because the Hebbian dynamics encourages neighboring neurons to develop similar receptive fields. However, unless the initial conditions are biassed appropriately, the map becomes piecewise continuous only, global topographic ordering is not obtained. Extra non epigenetic elements, such as pre-wiring to "proper" targets, must be introduced to achieve true neurotopic maps. There are however two works (Bienenstock, 1983 and 1983) in which global topographic ordering occurs for general initial conditions; the price paid is that the models order only one dimensional layers with periodic boundary conditions and require that one have equal numbers of neurons in the two layers. In this paper, we show that one can remove these restrictions by appropriate choices of neural dynamics: we have been able to exhibit a class of local epigenetic models which are biologically realistic, and which self-organize both one and two-dimensional layers into globally ordered maps for generic initial conditions. An important feature of our models is the modification of the basic Hebbian mechanism by a term which incorporates competition between all the synapses belonging to a given neuron. The omission of this term in previous neural network models explains in part the long-standing failure to obtain global neurotopies.

Network Architecture and Dynamics

The architecture of our models is as follows. The two neural layers are composed of neurons in a two-dimensional geometry with (discrete) cartesian coordinates (a, b) (layer 1) and (x, y) (layer 2). The axon for each neuron in layer 1 arborizes to have excitatory synaptic connections with all of the neurons in layer 2. We denote the strength of the synapse from (a,b) to (x,y) by $W_{xy,ab}$; this quantity is assumed to be modifiable and positive. The full connectivity guarantees that no bias or ordering is imposed on the connections in the initial conditions or otherwise. This is in contrast to works which study plasticity questions (Bienenstock 1982), (Miller, keller, and Stryker, 1989). Thus, in our models topographic ordering can appear only though a process of self-organization. The neurons in layer 2 are laterally connected by non-modifiable synaptic connections; we label the corresponding strengths $S_{x'y',xy}$. We take these to be positive, symmetric, and short ranged.

Now for the neuronal dynamics. Denote the average firing rate or "potential" of a neuron in layer 1 by U_{ab} and in layer 2 by V_{xy}. We take V to be linear in each of its inputs. These inputs come from all of the cell's incoming synapses so that we have

$$V_{xy} = \sum_{ab} W_{xy,ab} U_{ab} + \sum_{x'y'} S_{xy,x'y'} V_{x'y'}$$

or in matrix notation: $V = (I - S)^{-1} WU$. We can think of the U's (e.g., if layer 1 is a retina) as the response to an image stimulus. In the ensembles of biological interest, nearby neurons will have correlated stimulii, and so we take U_{ab} and U_{cd} to be correlated if the cells (a, b) and (c, d) are physically close. We thus define the correlation matrix $Q_{ab,cd} =< U_{ab} U_{cd} > - < U_{ab} >< U_{cd} >$ where $<>$ is the average over the ensemble of stimulii. We take Q to be short range with positive matrix elements. Consider now the the classic formulation of Hebb's rule:

$$\tau \dot{W}_{xy,ab} = V_{xy} U_{ab}$$

(τ is a time scale and a dot denotes a time derivative). This rule strengthens a synapse whenever the pre and post-synaptic neurons are "active" or fire in near synchrony. It is biologically plausible because it takes into account only the pre and post-synaptic states, i.e., it is local. This is in contrast to the Kohonen-like rules which are completely non-local. Many authors (Bienenstock, 1982), (Linsker, 1986), (Miller, Keller, and Stryker, 1989) have modified the above bare Hebb equation. One generalization is

$$\tau \dot{W}_{xy,ab} = V_{xy} U_{ab} - < V_{xy} >< U_{ab} > - g W_{xy,ab}$$

The average $< V >$ is to be take over some time interval. Since τ, which gives the time scale associated with changes in W, is very large, we can replace $< V >$ by the average over the ensemble of images assuming the W's are fixed. Similarly, for the long-time behavior, VU may be replaced by its average. U can then be eliminated, and the first two terms of the right hand side of the above equation become $(I - S)^{-1} WQ$. This dynamics tends to develop local order, as shown in (Amari, 1980).

To obtain global order, it is necessary to go beyond this type of Hebbian dynamics. We do this by introducing competition between the synapses of a given neuron. Such a competition is biologically justified: the biochemical reactions in the synaptic regions require among other things ATP and nerve growth factor which must be actively transported from the main body of the neuron to the synapses. When a neuron has only weak synaptic connections, there is very little competition because there are enough resources for all synapses of the cell, but when the neuron has many strong connections, the cell body cannot cope with the demand, and competition among the synapses becomes strong. This leads us to consider models which have the following synaptic dynamics:

$$\tau \dot{W}_{xy,ab} = [(I-S)^{-1}WQ]_{xy,ab} - g_1 \sum_{a'b'} W_{xy,a'b'} - g_2 \sum_{x'y'} W_{x'y',ab} - g_3 W_{xy,ab}$$

The g's are parameters which may vary with the level of activity of the cells (see later). These quantities and the constraints $W_{xy,ab} > 0$ are the only source of non-linearities of this model. We have found that for a broad range of g functions, the W's self-organize to form a topographic map. We begin by explaining how the network self-organizes in one dimension, and then proceed to the two-dimensional case.

Case of One-Dimensional Layers

We label the neurons in the one-dimensional input (respectively processing) layer according to the natural order $a = 1, 2, \cdots M' - 1$ (resp. $x = 1, 2, \cdots M - 1$)). This labelling reflects the fact that there are extremities to the layers; the ring geometry is not appropriate. The $W_{x,a}$ are given initial values which are small, positive, and random. Thus at the beginning of the self-organization, the competition between the synapses is weak: g_1 and g_2 should be small, so we first consider the limit $g_1 = g_2 = 0$. For the propose of the analysis, we take g_3 to be the same for all neurons. This has the advantage of leading to dynamics which can be treated analytically because the equations for synaptic change are "quasi-linear", i.e., are simple in a certain basis. Consider decomposing W (thought of as a vector) in terms of the eigenvectors of the Hebbian operator: $H(W) = (I - S)^{-1}WQ$. For ease of presentation, we shall work with special choices of S and Q so that the eigenfunctions can be written down explicitly, but the final results will be insensitive to the details of S and Q. If S and Q are short range, positive, symmetric, and translation invariant, up to an additive constant they essentially correspond to second derivatives. For zero boundary conditions (no inputs from outside the layers), the relevant eigenfunctions are approximately sines in each variable: $\Psi_{m,m'}(x,a) = sin(m\pi x/M)sin(m'\pi a/M')$ with m (resp. m') $= 1, 2, \cdots M-1$ (resp. $M'-1$). Then at all times, $W(t) = \sum A_{m,m'}(t)\Psi_{m,m'}$ with $\dot{A}_{m,m'}(t) = (\lambda_m \mu_{m'} - g_3(t))A_{m,m'}(t)$ in the quasi-linear regime ($W > 0, g_1 = g_2 = 0$, and g_3 location-independent). λ_m and $\mu_{m'}$ are the eigenvalues of the linear operators in x and a space from which the Hebbian operator is constructed. For the models of interest, these eigenvalues are strictly positive, and the fastest growing modes are $\Psi_{1,1}$, $\Psi_{1,2}$, $\Psi_{2,1}$, $\Psi_{2,2}$, etc... The above dynamics continuously "purify" W to make it more and more like $\Psi_{1,1}$. If g_3 is an increasing function of the total synaptic strength or activity in the network, then the system converges to a time independent state given by a multiple of $\Psi_{1,1}$. (In a similar spirit, Bienenstock et al. chose the growth rate of synapses to depend on the maturity of the entire network.) The state $\Psi_{1,1}$ does not correspond to a topographic map because $W_{x,a}$ has no correlations between x and a. To get correlations, W must be the sum of several eigenmodes. What would be the ideal linear combination? We can map layer 1 to layer 2 by a simple stretching; in continuum notation, this would correspond to

$$W_{x,a} = \delta_{x/M,a/M'} = \sum_m sin(\frac{m\pi x}{M})sin(\frac{m'\pi a}{M'})$$

where δ is the delta function and we have dropped irrelevant multiplicative factors. If at some point W is given by the first few terms in this series, the large scale features of the connections will be correct; one can then simply refine the receptive fields to obtain the desired neurotopy.

Above, we saw that the purification enabled one to obtain the first term in the above series. The parameters g_1 and g_2 give the second term in the following way. As the W's grow, the competition between synapses increases, so g_1 and g_2 turn on. If the Hebbian eigenfunctions are exactly the sines used above, the g_1 and g_2 terms do not affect the growth of $\Psi_{2,2}$, but they suppress the growth of $\Psi_{1,1}$, $\Psi_{1,2}$, and $\Psi_{1,1}$. Even if the eigenfunctions are not exactly sines, the qualitative behavior remains the same: the growth of these first three modes is suppressed much more than for the analogue of $\Psi_{2,2}$. By increasing g_1 and g_2, eventually $\Psi_{1,1}$ becomes unstable to $\Psi_{2,2}$ perturbations. Such perturbations will grow, and W will become essentially of the form $A_{1,1}(t)\Psi_{1,1} + A_{2,2}(t)\Psi_{2,2}$; in addition, $A_{1,1}$ decreases whereas $A_{2,2}$ increases with time. When $A_{2,2}$ reaches a large enough value, the neurons at the edge of the cortex see their W's connecting to one of the edges of the retina vanish (since $\Psi_{2,2}$ has negative components, whereas $\Psi_{1,1}$ is positive everywhere). Thereafter, the region of zero W's grows like a moving front and these neurons develop narrow receptive fields. The two-dimensional analogue is illustrated in Fig. 1. This process then affects the neurons further away from the edge of layer 2, and soon all layer 2 neurons have narrow receptive fields. If $A_{1,1}$ and $A_{2,2}$ are of the same sign, we obtain the "direct map" as given by the above delta function. If they are of opposite sign, we obtain an inverted map with one layer reflected. These are the only two classes of topographic maps in 1 dimension. Here the existence of several classes of topographic maps follows from symmetry considerations.

One can also understand the self-organization mechanism using an argument based on energetics. For fixed g's, the synaptic dynamics of our model is of gradient descent type. Then the W's converge to local minima of an energy function. Because of the Hebbian term, there is an energy cost for discontinuities in W. It is thus natural to expect the absolute minimum of energy to correspond to a bicontinuous map, giving one of the allowed topographic mappings. We know from simulations that the energy function also has local minima, and that these correspond to piecewise-continuous maps. The slow growth of the coupling constants g_1 and g_2 allow the energy landscape to change smoothly and drives one into the global minimum. Our procedure is thus analogous to Hopfield and Tank's suggestion for finding energy minima (Hopfield and Tank, 1986). Note that the g_1 term controls the receptive field size and the g_2 term controls the arborization size of retinal cells.

Case of Two-Dimensional Layers

The above self-organization also occurs in two-dimensional layers. Consider rectangular layers with zero boundary conditions, taking (a, b) and (x, y) to be the cartesian coordinates of the neurons. Again, we begin with $g_1 = g_2 = 0$, and g_3 identical for all neurons, we work in the eigenspace of the Hebbian operator, and we take S and Q to act as Laplace operators (the results hold for more general operators also). If the dimensions of the input (resp. processing) layer are $M'-1, N'-1$ (resp. $M-1, N-1$), the eigenfunctions are:

$$\Psi_{mn,m'n'}(x,a) = sin(m\pi x/M)sin(n\pi y/N)sin(m'\pi a/M')sin(n'\pi b/N')$$

The fastest growing modes are $\Psi_{11,11}$, $\Psi_{11,12}$, $\Psi_{11,21}$, $\Psi_{12,11}$, $\Psi_{21,11}$, $\Psi_{12,12}$, $\Psi_{12,21}$, $\Psi_{21,12}$,

$\Psi_{21,21}$, ... One of the topographic maps we wish to obtain is the "direct" map, which we represent symbolically as

$$W_{xy,ab} = \delta_{x/M,a/M'}\delta_{y/N,b/N'} = \Psi_{11,11} + \Psi_{12,12} + \Psi_{21,21} + \ldots$$

where again multiplicative factors have been dropped. Just as in the one-dimensional case, if we can get W to be given by the first few (e.g., three) of these modes, then further refinement will lead to the desired neurotopy. The first term of the sum can be obtained from the purification process when g_1 and g_2 are small. At the end of this quasi-linear regime, W is a multiple of $\Psi_{11,11}$. Then we need to make $\Psi_{12,12} + \Psi_{21,21}$ appear at the expense of the other modes. As one increases g_1 and g_2, the growth rate of $\Psi_{11,11}$ is decreased by (an amount proportional to) $g_1 + g_2$, that of $\Psi_{21,11}$ and $\Psi_{12,11}$ by g_1, that of $\Psi_{11,21}$ and $\Psi_{11,12}$ by g_2, while $\Psi_{12,12}$, $\Psi_{21,21}$, $\Psi_{21,12}$, and $\Psi_{12,21}$ are unaffected. (Even if the wave functions are not exactly products of sines, the qualitative effect of g_1 and g_2 on the growth rates remains the same.) Thus eventually $\Psi_{11,11}$ destabilizes and the synaptic strengths become of the form

$$W = A_{11,11}\Psi_{11,11} + A_{12,12}\Psi_{12,12} + A_{21,21}\Psi_{21,21} + A_{12,21}\Psi_{12,21} + A_{21,12}\Psi_{21,12}$$

We see that there are still too many modes present. This is because there exists several topographic maps other than the "direct" one: for instance $W = \Psi_{11,11} + \Psi_{12,21} + \Psi_{21,12}$ leads to a map where one of the layers has been reflected about the diagonal. There are a total of eight topographic maps for this geometry. Which map is "selected" varies with the geometry of the problem as the above Ψ's usually have different growth rates (Letelier and Martin, 1992).

In a generic rectangular case, all the eigenvalues of the Hebbian operator are non-degenerate. Assume for instance that $\Psi_{21,21}$ grows faster than $\Psi_{12,21}$ or $\Psi_{21,12}$ or $\Psi_{12,12}$. Then W becomes of the form $A_{11,11}\Psi_{11,11} + A_{21,21}\Psi_{21,21}$. As $\Psi_{21,21}$ grows at the expense of $\Psi_{11,11}$, some W's begin to vanish exactly as in the one-dimensional case. Eventually all cortical neurons develop receptive fields in the shape of stripes parallel to the b direction as in Figure 1. The $W = 0$ values effectively restrict the coordinate domain, so the system almost becomes a collection of one-dimensional networks. As g_1 and g_2 increase further, the W's develop a new instability in the b and y directions. Then the cortical cells in different stripes collapse their receptive fields similarly because there is some amount of coupling between these stripes. The most unstable mode of the striped configuration is essentially $\Psi_{12,12}$ which gives rise to the desired neurotopy, and the final (a^*, b^*) map is bicontinuous as long as the g's were not increased too quickly. Figure 2 shows the final map for a run from a random initial start where g_1 and g_2 were slowly increased. The 17x7 dots represent the coodinates of the cells in the retina. The nodes of the wire mesh give the value of (a^*, b^*) for each of the 13x8 neurons in the cortex (layer 2). Links have been draw connecting the nodes corresponding to neighboring neurons in the cortex allowing a visualization of the neurotopy. It is interesting to note that if the g's are fixed from the begining one can fall into a local minimum of the energy function. Such maps can be obtained from Figure 2 by twists (thereby forming a bow-tie) or by cuts and stretches. The slow growth of our g's can be said to be analogous to the annealing schedule used in Kohonen-like networks. Instead of taking the g's as given functions of time, it is also possible to consider g's which are functions of the cell's local activity, or total synaptic strength, and even to change the form of the terms multiplying the g's, e.g., by raising them to a power greater than one. Such models are local in the strictest sense, and we found by simulation that their qualitative behavior is the same as the cases explained here.

The attentive reader will have notived that the self-organization rest primarily on the non-degeneracy of the Hebbian operator. Non-degeneracy is generic, so that irregular or biologically realistic network geometries will give rise to neurotopies in the context of our synaptic dynamics. On the other hand, if there is a degeneracy in the leading eigenvalues, the creation of neurotopies is much less robust. In particular, we have found that for a square geometry, the system will select one of the topographic maps only if the g's are increased extremely slowly.

In summary, we have modified Hebb's equation by including terms which expresses the competition among all the synapses of a given cell. This enabled us to exhibit a family of local neural networks which self-organize into topographically ordered maps without the need for markers or special initial conditions: the self-ordering is purely epigenetic.

We thank David Kleinfeld, Ralph Linsker, Ken Miller, and Haim Sompolinsky for comments. This work was supported in part by the National Science Foundation (NSF-ECS- 8909127) and by an award from the Professional Staff Congress of City University

of New York.

References

[1] S.I. Amari, Bull. Math. Bio. 42, 339 (1980)

[2] E. Bienenstock, L.N. Cooper, and P. Munro, J. Neurosci. 2, 32 (1982)

[3] E. Bienenstock, in "Synergetics of the Brain", Ed. E. Basar et. al. (1983).

[4] A.F. Haussler and C. von der Malsburg, J. Theor. Neurobiol. 2 (1983) 47.

[5] D.O. Hebb, "The Organization of Behavior", (1949), John Wiley and Sons, NY

[6] J.J. Hopfield and D.W. Tank, Science 233, 625 (1986)

[7] T. Kohonen, Bio. Cybern. 43, 59 (1982)

[8] J.C. Letelier and O.C. Martin, submitted to "Network".

[9] R. Linsker, Proc. Natl. Acad. Sci. USA 83, 7508 (1986)

[10] R. Linsker, Neural Comp. 1, 402 (1989)

[11] K.D. Miller, J.B. Keller, and M.P. Stryker, Science 245, 605 (1989)

[12] J.C. Pearson, L.H. Finkel, and G.M. Edelman, J. Neurosci. 7, 4209 (1987)

[13] H. Ritter and K. Schulten, Bio. Cybern. 54, 99 (1986)

[14] S.C. Sharma, Exp. Neurol. 34, 171 (1972)

[15] R.W. Sperry, Proc. Natl. Acad. Sci. USA 50, 703 (1963)

[16] D.J. Willshaw and C. von der Malsburg, Proc. R. Soc. B 194, 431 (1976)

[17] J. Zhang, Neural Comp. 3, 54 (1991)

Figure Captions

Figure 1: Receptive field of a cortical neuron after the first instability for a network with two-dimensional layers. Plotted are the strengths of the synaptic connections.

Figure 2: Representation of the final state of the map between a 7x17 "retina" and an 8x13 "cortex".