Network: Comput. Neural Syst. **10** (1999) 111–121. Printed in the UK PII: S0954-898X(99)94664-X

Solutions of the BCM learning rule in a network of lateral interacting nonlinear neurons

G C Castellani†§, N Intrator‡‖, H Shouval‡¶ and L N Cooper‡+

† Physics Department, Bologna University, Via Berti Pichat 6/2, 40127 Bologna, Italy
‡ Physics Department, Department of Neuroscience and Institute for Brain and Neural Systems, Brown University, Providence, RI 02912, USA

Received 28 May 1998, in final form 17 September 1998

Abstract. We introduce a new method for obtaining the fixed points for neurons that follow the BCM learning rule. The new formalism, which is based on the objective function formulation, permits analysis of a laterally connected network of nonlinear neurons and allows explicit calculation of the fixed points under various network conditions. We show that the stable fixed points, in terms of the postsynaptic activity, are not altered by the lateral connectivity or nonlinearity. We show that the lateral connectivity alters the probability of attaining different states in a network of interacting neurons. We further show the exact alteration in presynaptic weights as a result of the neuronal nonlinearity.

1. Introduction

The BCM theory of cortical plasticity has been introduced by Bienenstock, Cooper and Munro (BCM) (Bienenstock *et al* 1982) to account for the changes observed in cell response of visual cortex due to changes in visual environment.

The BCM synaptic modification rule has the form

$$\dot{\boldsymbol{m}}(t) = \boldsymbol{d}\,\phi(c,\theta)$$

where $\boldsymbol{m}$ is the synaptic weight vector, ϕ is a nonlinear function of the postsynaptic activity c which has two zero crossings, one at $c = 0$ and the other at $c = \theta$, and $\boldsymbol{d}$ is the presynaptic activity vector. The variable θ, also called the moving threshold, is a super-linear function of the history of cell activity.

It was shown that a variant of this theory performs exploratory projection pursuit using a projection index that measures multi-modality (Intrator and Cooper 1992). This learning model allows modelling and theoretical analysis of various visual deprivation experiments such as monocular deprivation (MD), binocular deprivation (BD) and reversed suture (RS) (Intrator and Cooper 1992) and is in agreement with the many experimental results on visual cortical plasticity (Clothiaux *et al* 1991, Law and Cooper 1994, Shouval *et al* 1996, 1997). Recently, it was shown that the consequences of this theory are consistent with experimental results on long term potentation (LTP) and long term depression (LTD) (Dudek and Bear 1992, Kirkwood *et al* 1993, 1996). A network implementation which can find several projections in

§ E-mail address: gasto@alma.unibo.it
‖ On leave from: School of Mathematical Sciences, Tel-Aviv University. E-mail address: nin@cns.brown.edu
¶ E-mail address: hzs@cns.brown.edu
+ E-mail address: lnc@cns.brown.edu

0954-898X/99/020111+11$19.50

parallel while retaining its computational efficiency, was found to be applicable for extracting features from very high-dimensional vector spaces (Intrator *et al* 1991, 1996, Huynh *et al* 1996).

Up to now, we have analysed the properties of the BCM rule using two different methods. In the initial BCM paper (Bienenstock *et al* 1982) a general form of the ϕ function was assumed. A fixed-point method with linear stability analysis was used to analytically find the stable fixed points in two simple cases: (a) when the inputs are two linearly independent vectors, in the positive quadrant of two-dimensional space; (b) for N orthogonal vectors in the positive quadrant of an N-dimensional space. In both cases the stable fixed point was shown to be the maximally selective one, i.e the weight vector ($\boldsymbol{m}$) becomes orthogonal to all the inputs but one. Later an objective function formulation was used in the case of the quadratic form of ϕ (Intrator and Cooper 1992). This method has made it possible to extend these results to the case of N linearly independent vectors that are not restricted to lie in the positive quadrant.

In this paper we extend our earlier analysis (Intrator and Cooper 1992) to a *laterally* connected network of *nonlinear* neurons. We characterize the space of solutions, their stability properties and their temporal evolution with different configurations of the cortico-cortical synapses (lateral interactions).

The method we use here is a novel direct method in which we study a matrix form of the dynamics using D^{T}, the matrix spanned by the individual input vectors ($\boldsymbol{d}$). We solve a deterministic matrix equation (e.g., equation (3)) rather than an equation averaged over the inputs. This method enables analysis of different more realistic cases such as nonlinear (sigmoidal) neurons and networks with various forms of lateral interactions.

The results are surprising; we find that the same fixed points exist for the activity c in all these different cases. Neither lateral interactions nor the nonlinearity change the nature of the stable fixed points in terms of the postsynaptic activity. However, the fixed points in terms of the synaptic weight vectors ($\boldsymbol{m}$) are altered (exact solution is given). A network of N interacting BCM neurons has solutions that are combinations of all possible single cell solutions. There are two basic types of solutions; The first is that different cells become selective to different input patterns, this is termed *network selective states*. The other possibility is that different cells become selective to the same input patterns, this is termed *network associative states*. In both cases the solutions are exactly the same as those of a network with non-interacting neurons. Thus, surprisingly, the lateral connectivity does not change the fixed points (in terms of c) or their stability. Instead, it was found that the lateral connectivity affects the probability of attaining selective and associative states. As expected, inhibition favours selective states while excitation favours associative states. This analysis is supported by simulations of a BCM network in averaged (batch learning mode) or stochastic form.

2. A single neuron

The mathematical technique and notation used throughout are introduced in this section using a simple example of a single linear neuron in a two-dimensional (2D) space receiving two inputs. Intrator and Cooper (1992) presented an objective function formulation for the theory which indicates what the neuronal goal is and enables simple analysis of the dynamics. Their objective function is given by

$$R(m) = -\tfrac{1}{3}E\big[\boldsymbol{m}\cdot\boldsymbol{d}\big]^3 + \tfrac{1}{4}E^2\big[\boldsymbol{m}\cdot\boldsymbol{d}\big]^2 \tag{1}$$

where E denotes the expectation with respect to the input environment, $\boldsymbol{m}$ is the presynaptic weight vector and $\boldsymbol{d}$ is a stochastic input vector taken from the distribution of the environment. This function is bounded from below, and it thus has (local) minima which can be attained

by gradient descent $\dot{\boldsymbol{m}} = -\nabla R$. This leads to an approximate solution via the stochastic differential equation

$$\dot{\boldsymbol{m}} = \boldsymbol{d}\,\phi(c, \theta) \tag{2}$$

where $c = \boldsymbol{m}\cdot\boldsymbol{d}$ is the neuronal activity, $\boldsymbol{m}$ and $\boldsymbol{d}$ are the synaptic strength and incoming signal vectors, respectively. The function $\phi(c, \theta) = c(c - \theta)$ is a quadratic function that changes sign at a dynamic threshold that is a nonlinear function of some time-averaged measure of cellular activity, which is replaced (under the slow-learning assumption) by the expectation over the environment $\theta = E[c^2] = \sum_{i=0}^{n} p_i(\boldsymbol{m}\cdot\boldsymbol{d}_i)^2$ (Intrator and Cooper 1992), where p_i is the probability of choosing vector d_i from the data set. The averaged version of this equation can be written as

$$\dot{\boldsymbol{m}}(t) = P D^{\mathrm{T}}\boldsymbol{\phi}(\boldsymbol{c}, \theta) \tag{3}$$

where the matrix of inputs D is composed of the different input vectors, and P is a diagonal matrix of the probabilities. In the simple, 2D case, $D = \left(\begin{smallmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{smallmatrix}\right)$ and $P = \left(\begin{smallmatrix} p_1 & 0 \\ 0 & p_2 \end{smallmatrix}\right)$, where p_i is the probability of choosing the ith vector from the data set. The neuronal activity due to input environment is given by $\boldsymbol{c} = \left(\begin{smallmatrix} c_1 \\ c_2 \end{smallmatrix}\right)$ whose elements are the neuron's outputs in response to the two input and $\boldsymbol{\phi} = \left(\begin{smallmatrix} \phi(c_1,\theta) \\ \phi(c_2,\theta) \end{smallmatrix}\right)$ is the vector of the corresponding neuronal activation function. Using this notation, the averaged form of the BCM rule is given by

$$\begin{aligned}
\dot{m}_1 = {} & (m_1d_{11} + m_2d_{12})((m_1d_{11} + m_2d_{12}) - (p_1(m_1d_{11} + m_2d_{12})^2 \\
& + p_2(m_1d_{21} + m_2d_{22})^2))d_{11}p_1 + (m_1d_{21} + m_2d_{22})((m_1d_{21} + m_2d_{22}) \\
& - (p_1(m_1d_{11} + m_2d_{12})^2 + p_2(m_1d_{21} + m_2d_{22})^2))d_{21}p_2 \\
\dot{m}_2 = {} & (m_1d_{11} + m_2d_{12})((m_1d_{11} + m_2d_{12}) - (p_1(m_1d_{11} + m_2d_{12})^2 \\
& + p_2(m_1d_{21} + m_2d_{22})^2))d_{12}p_1 + (m_1d_{21} + m_2d_{22})((m_1d_{21} + m_2d_{22}) \\
& - (p_1(m_1d_{11} + m_2d_{12})^2 + p_2(m_1d_{21} + m_2d_{22})^2))d_{22}p_2.
\end{aligned} \tag{4}$$

We are interested in finding the stationary states of (2) or equivalently of (3). From (3) we note that the condition $\dot{\boldsymbol{m}}(t) = 0$ implies that $P D^{\mathrm{T}}\boldsymbol{\phi}$ must be zero, and this is possible if and only if $\boldsymbol{\phi} = 0$, because we require that the input vectors are linearly independent (i.e. $|D| \neq 0$). The condition $\boldsymbol{\phi} = 0$ gives

$$\begin{aligned}
&(m_1d_{11} + m_2d_{12})((m_1d_{11} + m_2d_{12}) - (p_1(m_1d_{11} + m_2d_{12})^2 + p_2(m_1d_{21} + m_2d_{22})^2)) = 0 \\
&(m_1d_{21} + m_2d_{22})((m_1d_{21} + m_2d_{22}) - (p_1(m_1d_{11} + m_2d_{12})^2 + p_2(m_1d_{21} + m_2d_{22})^2)) = 0
\end{aligned} \tag{5}$$

namely

$$\begin{aligned}
&c_1(c_1 - (p_1c_1^2 + p_2c_2^2)) = 0 \\
&c_2(c_2 - (p_1c_1^2 + p_2c_2^2)) = 0.
\end{aligned} \tag{6}$$

The fixed points are given by

$$(c_1, c_2) = \left\{(0,0), \left(\frac{1}{p_1}, 0\right), \left(0, \frac{1}{p_2}\right), \left(\frac{1}{p_2 + p_1}, \frac{1}{p_2 + p_1}\right)\right\}.$$

The $\boldsymbol{m}$ solutions can be obtained through the inverse transformation $\boldsymbol{m} = D^{-1}\boldsymbol{c}$. Stability analysis (see the appendix as well as Bienenstock *et al* (1982) and Intrator and Cooper (1992)) shows that the stable solutions are the *selective fixed points* $\left(\frac{1}{p_1}, 0\right)$ and $\left(0, \frac{1}{p_2}\right)$.

2.1. A single neuron with n inputs

The extension to a single neuron with n inputs is quite simple. We can consider the input matrix

$$D = \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{pmatrix}$$

whose rows $\boldsymbol{d}_i$ represent n-input vectors. The determinant of D is non-zero since the inputs are linearly independent. Using previous notation we have

$$\boldsymbol{c} = (c_1, c_2, \ldots, c_n) = (\boldsymbol{m} \cdot \boldsymbol{d}_1, \boldsymbol{m} \cdot \boldsymbol{d}_2, \ldots, \boldsymbol{m} \cdot \boldsymbol{d}_n)$$

$$\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_n) = (\phi(c_1, \theta), \phi(c_2, \theta), \ldots, \phi(c_n, \theta)).$$

The threshold θ is given by

$$\theta \equiv E[c^2] = \sum_{j=1}^{n} p_j (\boldsymbol{m} \cdot \boldsymbol{d}_j)^2 = \sum_{j=1}^{n} p_j c_j^2. \tag{7}$$

The dynamics is given by

$$\begin{pmatrix} \dot{m}_1 \\ \dot{m}_2 \\ \vdots \\ \dot{m}_n \end{pmatrix} = \begin{pmatrix} p_1 & 0 & 0 & \cdots & 0 \\ 0 & p_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & p_n \end{pmatrix} \begin{pmatrix} d_{11} & d_{12} & \cdots & d_{1n} \\ d_{21} & d_{22} & \cdots & d_{2n} \\ \vdots & \vdots & & \vdots \\ d_{n1} & d_{n2} & \cdots & d_{nn} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_n \end{pmatrix}$$

or simply

$$\dot{\boldsymbol{m}}(t) = P D^{\mathrm{T}} \boldsymbol{\phi}. \tag{8}$$

As in the two-dimensional case, stationary points are characterized by $\boldsymbol{\phi} = 0$, to give

$$c_1 \left(c_1 - \left(\sum_{j=1}^{n} p_j c_j^2 \right) \right) = 0$$

$$c_2 \left(c_2 - \left(\sum_{j=1}^{n} p_j c_j^2 \right) \right) = 0$$

$$\vdots$$

$$c_n \left(c_n - \left(\sum_{j=1}^{n} p_j c_j^2 \right) \right) = 0$$

or in a more compact form

$$c_i \left(c_i - \left(\sum_{j=0}^{n} p_j c_j^2 \right) \right) = 0 \qquad i = 1, \ldots, n \tag{9}$$

The solutions of (9) are (we report only the equivalence classes)

$$S = \begin{cases} (0,0,0,\ldots,0) \\ \left(0,0,\ldots,\dfrac{1}{p_i},\ldots,0\right) \\ \left(0,\ldots,0,\dfrac{1}{(p_i+p_j)},0,\ldots,0,\dfrac{1}{(p_i+p_j)},0,\ldots,0\right) \\ \left(0,\ldots,0,\dfrac{1}{(p_i+p_j+p_k)},0,\ldots,0,\dfrac{1}{(p_i+p_j+p_k)},0,\ldots,0,\dfrac{1}{(p_i+p_j+p_k)},\ldots,0\right) \\ \qquad\vdots \\ (1,1,1,\ldots,1). \end{cases}$$

The corresponding $\boldsymbol{m}$ solutions are $\boldsymbol{m} = D^{-1}\boldsymbol{c}$. Stability analysis (see the appendix as well as Intrator and Cooper 1992) shows that the only stable states (in the Lyapunov sense) are those with maximum selectivity given by $\left(0,0,\ldots,1/p_i,\ldots,0\right),\ i = 1,\ldots,n$.

It turns out that a well known immunological network model (Weisbuch *et al* 1990) has the same set of solutions S (Castellani *et al* 1998).

3. A network with lateral interactions

When a neuron is in a network, the incoming inputs can arise from the thalamus, for instance the LGN if these neurons are in V1. Another set of inputs can arise from other cortical neurons. In a network setup, the vector of synaptic weights $\boldsymbol{m}$ for a single neuron now becomes a matrix M for all the network neurons. The vector of neuronal activities $\boldsymbol{c}$ (due to the matrix of inputs D) now becomes a matrix of neuronal activities, however we shall treat it as a super-vector. We start with the formulation as given in Cooper and Scofield (1988) for a network with a single input $\boldsymbol{d}$, the network activity is therefore given by

$$\boldsymbol{c} = M\boldsymbol{d} + L\boldsymbol{c} \tag{10}$$

where L is the cortico–cortical connectivity matrix in which l_{ij} is the interaction between neuron i (the target) and neuron j (the source); $\boldsymbol{d}$ is a single input vector and M is the matrix of the feedforward (thalamocortical) synapses; m_{ij} represents the feedforward connections to cell i arising from input channel j.

For a two-neuron network we can write the cortico–cortical matrix as

$$L = \begin{pmatrix} 0 & l_{12} \\ l_{21} & 0 \end{pmatrix}.$$

From equation (10) we obtain

$$\boldsymbol{c} = (I - L)^{-1}M\boldsymbol{d} \tag{11}$$

namely

$$\begin{pmatrix} c_1 \\ c_2 \end{pmatrix} = \frac{1}{1 - l_{12}l_{21}} \begin{pmatrix} 1 & l_{12} \\ l_{21} & 1 \end{pmatrix} \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{pmatrix} \begin{pmatrix} d_{11} \\ d_{12} \end{pmatrix}. \tag{12}$$

From these equations, which represent network activity due to a single input vector, we switch notation as to characterize a solution that represents network activity resulting from the *full input environment*. For this, we define a super-vector with components which represent the activity of neuron i due to input d_j, $\boldsymbol{c} = (c_{11}, c_{12}, c_{21}, c_{22})$. $\boldsymbol{m}$ also becomes a super-vector which

now represents the synaptic weight of all the neurons in the network (m_{ij} is the jth synapse of neuron i), so that the dynamics is given by

$$\begin{aligned}
\dot{m}_{11} &= \phi(c_{11}, \theta_1)d_{11}p_1 + \phi(c_{12}, \theta_1)d_{21}p_2 \\
\dot{m}_{12} &= \phi(c_{12}, \theta_1)d_{12}p_1 + \phi(c_{12}, \theta_1)d_{22}p_2 \\
\dot{m}_{21} &= \phi(c_{21}, \theta_2)d_{11}p_1 + \phi(c_{22}, \theta_1)d_{21}p_2 \\
\dot{m}_{22} &= \phi(c_{21}, \theta_2)d_{11}p_1 + \phi(c_{22}, \theta_2)d_{22}p_2
\end{aligned} \tag{13}$$

or

$$\dot{\boldsymbol{m}}(t) = \mathcal{P}_2\mathcal{D}_2^{\mathrm{T}}\boldsymbol{\phi} \tag{14}$$

where $\mathcal{D}_2^{\mathrm{T}}$ is the direct product of the input matrices: $\mathcal{D}_2^{\mathrm{T}} = D^{\mathrm{T}} \otimes D^{\mathrm{T}}$, $\mathcal{P}_2$ is the direct product of probability matrices $\mathcal{P}_2 = P \otimes P$ and $\boldsymbol{\phi}$ is the vector of the neuronal activation function: $\boldsymbol{\phi} = (\phi_{11}, \phi_{12}, \phi_{21}, \phi_{22})$. For the fixed-point equation we require again that $\boldsymbol{\phi} = 0$. Using the definition of $\boldsymbol{c}$ the neuronal activity takes the form

$$\boldsymbol{c} = \mathcal{L}_2\mathcal{D}_2\boldsymbol{m}$$

where the two matrices $\mathcal{L}_2$ and $\mathcal{D}_2$ are

$$\mathcal{L}_2 = \frac{1}{1 - l_{12}l_{21}}\begin{pmatrix} 1 & 0 & l_{12} & 0 \\ 0 & 1 & 0 & l_{12} \\ l_{21} & 0 & 1 & 0 \\ 0 & l_{21} & 0 & 1 \end{pmatrix} \qquad \mathcal{D}_2 = \begin{pmatrix} d_{11} & d_{12} & 0 & 0 \\ d_{21} & d_{22} & 0 & 0 \\ 0 & 0 & d_{11} & d_{12} \\ 0 & 0 & d_{21} & d_{22} \end{pmatrix}$$

The fixed-point equation associated with the system (13) becomes

$$\begin{aligned}
c_{11}(c_{12} - (p_1c_{11}^2 + p_2c_{12}^2)) &= 0 \\
c_{12}(c_{12} - (p_1c_{11}^2 + p_2c_{12}^2)) &= 0 \\
c_{21}(c_{21} - (p_1c_{21}^2 + p_2c_{22}^2)) &= 0 \\
c_{22}(c_{22} - (p_1c_{21}^2 + p_2c_{22}^2)) &= 0
\end{aligned} \tag{15}$$

As the system is now decoupled, the solutions are the direct product of the solutions of the two-dimensional system, namely,

$$\begin{aligned}
(c_{11}, c_{12}, c_{21}, c_{22}) = &\left\{(0,0), \left(\frac{1}{p_1}, 0\right), \left(0, \frac{1}{p_2}\right), \left(\frac{1}{p_2 + p_1}, \frac{1}{p_2 + p_1}\right)\right\} \\
&\otimes \left\{(0,0), \left(\frac{1}{p_1}, 0\right), \left(0, \frac{1}{p_2}\right), \left(\frac{1}{p_1 + p_2}, \frac{1}{p_1 + p_2}\right)\right\}.
\end{aligned}$$

As before, the $\boldsymbol{m}$ solutions can be obtained by the inverse transformation $\boldsymbol{m} = \mathcal{D}_2^{-1}\mathcal{L}_2^{-1}\boldsymbol{c}$.

Stability analysis (see the appendix) shows that the previously characterized set of solutions is stable as long as $\| L \| < 1$.

Similar analysis holds for a network of n neurons with lateral connections that receive n input vectors. It is easy see that for n neurons the matrix $\mathcal{L}_n$ takes the following form:

$$\mathcal{L}_n \equiv \begin{pmatrix} \mathcal{L}_{n_{11}} & \mathcal{L}_{n_{12}} & \cdots & \mathcal{L}_{n_{1n}} \\ \mathcal{L}_{n_{21}} & \mathcal{L}_{n_{22}} & \cdots & \mathcal{L}_{n_{2n}} \\ \vdots & & & \vdots \\ \mathcal{L}_{n_{n1}} & \mathcal{L}_{n_{n2}} & \cdots & \mathcal{L}_{n_{nn}} \end{pmatrix}$$

where $\mathcal{L}_{n_{ij}}$ is a diagonal $n \times n$ matrix with diagonal elements from $(I - L)_{ij}^{-1}$; or equivalently $\mathcal{L}_{n_{ij}} = (I - L)_{ij}^{-1} I$ with I the identity $n \times n$ matrix. For the input matrix we have the same form

of the two-neurons network, based on an iterated direct product: $\mathcal{D}_n = D \otimes D \otimes D \otimes \cdots D$; and the linear substitution that permits the exact solution of the network is $\boldsymbol{c} = \mathcal{L}_n \mathcal{D}_n \boldsymbol{m}$ with obvious inverse if $|\mathcal{D}| \neq 0$ and $|\mathcal{L}_n| \neq 0$†.

4. Basins of attraction

As can be seen above and in the appendix, all the stable solutions in the single-neuron case are also stable in the network case. For example, one possible stable solution is when all neurons become selective to the same input. In the two-dimensional case, the stable solutions (in terms of c) are (see the appendix)

$$\left(0, \frac{1}{p_2}, 0, \frac{1}{p_2}\right) \qquad \left(0, \frac{1}{p_2}, \frac{1}{p_1}, 0\right) \qquad \left(\frac{1}{p_1}, 0, \frac{1}{p_1}, 0\right) \qquad \left(\frac{1}{p_1}, 0, 0, \frac{1}{p_2}\right).$$

While network interactions do not change the stability of the possible solutions, they do change the basins of attraction associated with different solutions. Numerical integration of the system (13), both in averaged and stochastic form, shows that the size of the basin of attraction is changed. When all the interaction terms l_{ij} are set to zero, the probability of reaching one of the four stable states is the same (equal attracting power). This symmetric situation can be broken by setting the cortico–cortical connections l_{ij} to non-zero values; When l_{ij} are negative, there is an increase in the probability of reaching the states $\left(0, \frac{1}{p_2}, \frac{1}{p_1}, 0\right)$ or $\left(\frac{1}{p_1}, 0, 0, \frac{1}{p_2}\right)$ that correspond to different states of selectivity for neurons 1 and 2 (we call them *network selective states*). When the connections l_{ij} are positive, there is an increase in the probability to reach the states $\left(0, \frac{1}{p_2}, 0, \frac{1}{p_2}\right)$ or $\left(\frac{1}{p_1}, 0, \frac{1}{p_1}, 0\right)$ which correspond to equal selectivity states (we call them *network associative states*).

Table 1. Simulation results for a two-neuron network. The system (14) is integrated over initial conditions spanning a hypercube of size 0.1 for different values of the lateral connections L. The table gives the distribution of the different solutions.

l values	% of selective states	% of associative states
−0.2	98.1	1.9
−0.1	88.87	11.13
−0.05	76.9	23.1
0	50.45	49.45
0.05	42	58
0.1	15	85
0.2	6	94

The numerical simulations on system (13) are summarized in table 1. They show that the probability of selectivity or association monotonically varies with the magnitude of the L connections. There is an apparent asymmetry in the effect of positive or negative values of the interactions on the probabilities. We think that this results from using only positive values as inputs, so that negative interactions which tend to reduce cell activity have a more coherent effect than positive interactions. This is currently being further investigated (Castellani *et al* in preparation).

The extension to a network with n neurons is shown first on an $n = 3$ network with a subsequent generalization to $n > 3$. In a three-neuron network, the number of stable

† If we derive the BCM from the objective function we obtain a slightly different equation from (14), namely $\dot{\boldsymbol{m}}(t) = \mathcal{L}_2 \mathcal{P}_2 \mathcal{D}_2 \boldsymbol{\phi}$. The study of the fixed points of this equation is the same as in the previous case because the matrix $\mathcal{L}_n$ is non-singular.

solutions is given by the product of the single-neuron stable solutions, namely, 3^3. The number of completely selective solutions is 3!, which corresponds to the number of bijective functions between the set of three-input vectors and the three-output neurons. The associative solutions can be divided into *completely associative* and *partially associative*, where *completely associative* refers to those solutions that associate all the neurons to a single input pattern. It is clear that the number of such solutions coincides with the number of neurons. The other solutions exhibit an incomplete associativity; for example a typical solution in this class has two neurons which are selective to the same input and the third neuron is selective to another. The number of such solutions is found as the difference between the total number of solutions, the completely selective and associative: $3^3 - 3! - 3$. For a general n-size network the number of stable solutions is n^n, the number of completely selective solutions is $n!$ and the number of completely associative solutions is n. Thus, the number of solutions with incomplete associativity is $n^n - n! - n$.

If all the lateral connections are negative, different neurons reach different stable states (selective state) with higher probability, while if the elements $l_{i,j}$ are positive, different neurons are more likely to reach a similar stable state (associative state). The situation of intermixed lateral interaction where some are positive and some are negative (as in the Mexican hat lateral profile) are still being worked out. Simulations show that a combination of associative and selective states emerge.

5. Nonlinear neurons

5.1. A single neuron

The BCM rules can be extended to nonlinear neurons in which the neuron's activity is defined to be $c = \sigma(\boldsymbol{m}\cdot\boldsymbol{d})$ where σ is a smooth sigmoidal function. The exact derivation of the learning procedure by using the the minimization of the objective function (Intrator and Cooper 1992) gives

$$\dot{\boldsymbol{m}}(t) = \mu E\big[\phi(\sigma(\boldsymbol{m}\cdot\boldsymbol{d}), \theta)\sigma'(\boldsymbol{m}\cdot\boldsymbol{d})\boldsymbol{d}\big] \tag{16}$$

From this we can write the analogue to equation (3) as

$$\dot{\boldsymbol{m}}(t) = \Sigma P D\boldsymbol{\phi} \tag{17}$$

where D is the input matrix, $\boldsymbol{\phi}$ is the vector of ϕ calculated at the points $(\sigma(\boldsymbol{m}\cdot\boldsymbol{d}_1), \sigma(\boldsymbol{m}\cdot\boldsymbol{d}_2), \ldots, \sigma(\boldsymbol{m}\cdot\boldsymbol{d}_n))$, and Σ is a matrix containing the derivatives of σ at the points $(\boldsymbol{m}\cdot\boldsymbol{d}_1, \boldsymbol{m}\cdot\boldsymbol{d}_2, \ldots, \boldsymbol{m}\cdot\boldsymbol{d}_n)$:

$$\Sigma = \begin{pmatrix} \sigma_1' & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2' & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_n' \end{pmatrix} \tag{18}$$

The matrix Σ is positive definite as σ is smooth and monotonic, thus, the search for the stationary states of (17) leads to $\boldsymbol{\phi} = 0$. For convenience we define the variable ζ such that $\zeta = \sigma(\boldsymbol{m}\cdot\boldsymbol{d})$ and $\boldsymbol{\zeta} = (\sigma_1, \sigma_2, \ldots, \sigma_n)$. Thus the fixed-point solutions in terms of ζ are equivalent to the solutions of (9). It follows that the solutions for m result from solving an equation of the form $\boldsymbol{m} = D^{-1}\sigma^{-1}(\boldsymbol{\zeta})$; with $\zeta \in S$ (see section 2.1). The case of n nonlinear neurons without interaction, follows from the combination of the linear case and the nonlinear case.

5.2. A nonlinear neuron with lateral interactions

The case of nonlinear neurons with lateral interactions is also tractable, but requires the complete derivation from the objective function (Intrator and Cooper 1992). This function in the case of a nonlinear neuron takes the form (compare with equation (1))

$$R(m) = -\tfrac{1}{3}E\big[\sigma^3(\xi)\big] + \tfrac{1}{4}E^2\big[\sigma^2(\xi)\big] \tag{19}$$

where the variable ξ is the inhibited activity of the neurons prior to applying the nonlinearity σ. Using the type of network described in section 3 we would define ξ as $\xi = (I - L)^{-1}M\boldsymbol{d}$. This leads to the following gradient descent dynamics:

$$\begin{aligned} E[-\nabla_m R_m] &= \big\{E\big[\sigma^2(\xi)\sigma'\nabla_m\xi\big] - E\big[\sigma^2(\xi)\big]E\big[\sigma(\xi)\sigma'\nabla_m\xi\big]\big\} \\ &= E\big[\phi(\sigma(\xi), \theta_m)\sigma'\nabla_m\xi\big] \\ &= \Sigma\mathcal{P}\mathcal{L}\mathcal{D}\phi(\sigma(\xi), \theta_m). \end{aligned} \tag{20}$$

From equation (20) we can see that the stationary solutions arise from the equation $\boldsymbol{\phi}(\sigma(\xi)) = 0$, because the matrices Σ, $\mathcal{L}$, $\mathcal{D}$ are positive definite; hence the solutions are

$$\boldsymbol{m} = D^{-1}\mathcal{L}^{-1}\Sigma^{-1}(\zeta)$$

with $\zeta \in S$.

6. Discussion

Full characterization of a laterally connected network of nonlinear BCM neurons has been given for linearly separable input environments. As in the previous linear, single-cell cases analysed, the fixed points are such that each neuron responds to only one of the input patterns. The solution generates neuronal activity distribution with a large mass at zero in a similar manner to the single-neuron case. A network with similar neuronal distribution, the rectified-Gaussian belief network has recently been shown to be useful for generative models that discover sparse distributed representations of objects (Hinton and Ghahramani 1997).

For a network of interacting neurons all different combinations of stable single cell solutions are also stable. We performed simulations of the system (14) by integrating over initial conditions for different values of the lateral connections. This allows us to characterize the probability of finding the system in either an associative or selective solution. The lateral connections alter the size of the basin of attraction, so that inhibition increases the probability of obtaining selective states and excitation increases the probability of obtaining associative states. This behaviour holds for different values of the lateral connections L, with the property that increased values (of inhibitory or excitatory) connections increase the probability of falling into the selective or associative regime, respectively.

This analysis is the first step in analysing a network of nonlinear BCM neurons under a realistic visual environment. The next step is to move from the linearly independent environment that we have been using so far in our analysis to a realistic environment, where inputs are dependent as we have been recently using in our simulation (Blais *et al* 1998).

Acknowledgments

The authors thank the members of the Institute for Brain and Neural Systems for many fruitful conversations. This research was supported by the Charles A Dana Foundation, the Office of Naval Research and the National Science Foundation; GCC was partially supported by an exchange program between Bologna and Brown Universities.

Appendix. Stability analysis

To analyse the stability of the solutions, we examine the Jacobian matrix, which in a two-dimensional case is

$$\begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} \dfrac{\partial \phi_1}{\partial c_1} & \dfrac{\partial \phi_1}{\partial c_2} \\ \dfrac{\partial \phi_2}{\partial c_1} & \dfrac{\partial \phi_2}{\partial c_2} \end{pmatrix} \begin{pmatrix} \dfrac{\partial c_1}{\partial m_1} & \dfrac{\partial c_1}{\partial m_2} \\ \dfrac{\partial c_2}{\partial m_1} & \dfrac{\partial c_2}{\partial m_2} \end{pmatrix}$$

or

$$\begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}^{\mathrm{T}} \begin{pmatrix} 2c_1 - \theta - 2p_1c_1^2 & -2p_2c_2c_1 \\ -2p_1c_1c_2 & 2c_2 - \theta - 2p_2c_2^2 \end{pmatrix} \begin{pmatrix} \dfrac{\partial c_1}{\partial m_1} & \dfrac{\partial c_1}{\partial m_2} \\ \dfrac{\partial c_2}{\partial m_1} & \dfrac{\partial c_2}{\partial m_2} \end{pmatrix}.$$

Note that the third matrix is the input matrix D since $\boldsymbol{c}$ is defined as $\boldsymbol{c} = D\boldsymbol{m}$. At the critical point $\left(\frac{1}{p_1}, 0\right)$ or $\left(0, \frac{1}{p_2}\right)$, the central matrix becomes diagonal with respectively $-\frac{1}{p_1}$ or $-\frac{1}{p_2}$ as diagonal elements, thus, the Jacobian is negative definite†. For the other critical points; $(0, 0)$ is unstable in a Lyapunov sense (it is neutrally stable); the point $\left(\frac{1}{p_1+p_2}, \frac{1}{p_1+p_2}\right)$ is unstable because the Jacobian $D^{\mathrm{T}}\frac{\partial \phi}{\partial c}D$ is a quadratic form non-negative definite in the case of linearly independent vectors and $p_1 + p_2 = 1$:

$$J_2\big|_{\left(\frac{1}{p_1+p_2}, \frac{1}{p_1+p_2}\right)} = D^{\mathrm{T}}D - 2D^{\mathrm{T}} \begin{pmatrix} p_1 & p_2 \\ p_1 & p_2 \end{pmatrix} D.$$

The generalization to the n-dimensional case lead to the n-dimensional Jacobian

$$J_n = D^{\mathrm{T}} \begin{pmatrix} 2c_1 - \theta - 2p_1c_1^2 & -2p_2c_2c_1 & \cdots & -2p_nc_nc_1 \\ -2p_1c_1c_2 & 2c_2 - \theta - 2p_2c_2^2 & \cdots & -2p_nc_nc_2 \\ \vdots & \vdots & & \vdots \\ -2p_1c_1c_n & -2p_2c_2c_n & \cdots & 2c_n - \theta - 2p_nc_n^2 \end{pmatrix} D.$$

It is clear that the points with one non-zero coordinate set all the off-diagonal terms to zero, and in this way we obtain a diagonal matrix with diagonal elements $\left(-\frac{1}{p_i}, -\frac{1}{p_i}, \ldots, -\frac{1}{p_i}\right)$ for $i = 1, \ldots, n$. Therefore, the stability is guaranteed from the above considerations. The instability of all the other points follows from considerations as in the two-dimensional case. This analysis, performed for the case of one neuron with n inputs, is also applicable in all the other cases; the n neurons with n inputs, the network with lateral connections and the network of nonlinear neurons with lateral connections. The corresponding Jacobian matrices in each case become

$$\mathcal{D}_n^{\mathrm{T}}\left(\frac{\partial \phi}{\partial c}\right)\left(\frac{\partial c}{\partial m}\right) = \mathcal{D}_n^{\mathrm{T}}\left(\frac{\partial \phi}{\partial c}\right)\mathcal{D}_n. \tag{A1}$$

$$\mathcal{D}_n^{\mathrm{T}}\left(\frac{\partial \phi}{\partial c}\right)\left(\frac{\partial c}{\partial m}\right) = \mathcal{D}_n^{\mathrm{T}}\left(\frac{\partial \phi}{\partial c}\right)\mathcal{L}_n\mathcal{D}_n. \tag{A2}$$

$$\mathcal{D}_n^{\mathrm{T}}\left(\frac{\partial \phi}{\partial c}\right)\left(\frac{\partial c}{\partial m}\right) = \mathcal{D}_n^{\mathrm{T}}\left(\frac{\partial \phi}{\partial m}\right)\Sigma\mathcal{L}_n\mathcal{D}_n. \tag{A3}$$

In the first case (A1), the argument used in the the case of a single neuron with n inputs still holds because the matrix $\mathcal{D}_n$ is a direct product of $\mathcal{D}$ matrices. For the second case (A2) we

† The eigenvalues of the matrix $D^{\mathrm{T}}D$ are all positive and real because this matrix is symmetric and positive definite, and the product of diagonal matrices is commutative.

observe that the matrix $\mathcal{L}_n$ is a direct product of $\mathcal{L}$ matrices, and for each of these matrices, $L_{i,j} < 1$. Thus we can use a series expansion for the inverse matrix:

$$\frac{1}{1-L} \approx 1 + L + L^2 + \cdots .$$

For the critical points with one non-zero component, we obtain the following approximation:

$$\mathcal{D}_n^{\mathrm{T}} \left(\frac{\partial \phi}{\partial c} \right) \mathcal{L}_n \mathcal{D}_n \approx \left(\frac{\partial \phi}{\partial c} \right) \left(D^{\mathrm{T}} D + D^{\mathrm{T}} L D + D^{\mathrm{T}} L^2 D + \cdots \right).$$

From the power expansion (note that also $\|L\| < 1$) it is clear that the positive nature of the matrix $D^{\mathrm{T}} D$ does not change by introducing the matrix L; in other words, all the critical points with one non-zero component are asymptotically stable. These considerations are also valid for the case of a network of nonlinear neurons (A3) because the matrix Σ is diagonal and positive definite, thus, one can obtain the same expansion. The demonstration of the instability of the points with more than one non-zero component is obtained using the same method as in the two-dimensional case. The stability results were also confirmed using a direct integration of the system (14).

References

Bienenstock E L, Cooper L N and Munro P W 1982 Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex *J. Neurosci.* **2** 32–48

Blais B S, Intrator N, Shouval H and Cooper L N 1998 Receptive field formation in natural scene environments: comparison of single cell learning rules *Neural Comput.* **10** 1797–813

Castellani G, Giberti C, Franceschi C and Bersani F 1998 Stable state analysis of an immune network model *Int. J. Bifurc. and Chaos* **8** 1285–301

Clothiaux E E, Cooper L N and Bear M F 1991 Synaptic plasticity in visual cortex: comparison of theory with experiment *J. Neurophysiol.* **66** 1785–804

Cooper L N and Scofield C L 1988 Mean-field theory of a neural network *Proc. Natl Acad. Sci. USA* **85** 1973–7

Dudek S M and Bear M F 1992 Homosynaptic long-term depression in area CA1 of hippocampus and the effects on NMDA receptor blockade *Proc. Natl Acad. Sci. USA* **89** 4363–7

Hinton G E and Ghahramani Z 1997 Generative models for discovering sparse distributed representations *Phil. Trans. R. Soc.* B **352** 1177–90

Huynh Q, Cooper L N, Intrator N and Shouval H 1998 Classification of underwater mammals using feature extraction based on time-frequency analysis and BCM theory *IEEE Trans. Signal Process.* **46** 1202–7

Intrator N and Cooper L N 1992 Objective function formulation of the BCM theory of visual cortical plasticity: statistical connections, stability conditions *Neural Networks* **5** 3–17

Intrator N, Gold J I, Bülthoff H H and Edelman S 1991 Three-dimensional object recognition using an unsupervised neural network: understanding the distinguishing features *Proc. 8th Israeli Conf. on Artificial Intelligence Computer Vision* ed Y Feldman and A Bruckstein (Amsterdam: Elsevier) pp 113–23

Intrator N, Reisfeld D and Yeshurun Y 1996 Face recognition using a hybrid supervised/unsupervised neural network *Pattern Recognition Lett.* **17** 67–76

Kirkwood A, Duden S M, Gold J T, Aizenman C and Bear M F 1993 Common forms of synaptic plasticity in Hippocampus and Neocortex in *vitro Science* **260** 1518–21

Kirkwood A, Rioult M G and Bear M F 1996 Experience-dependent modification of synaptic plasticity in visual cortex *Nature* **381** 526–8

Law C and Cooper L 1994 Formation of receptive fields according to the BCM theory in realistic visual environments *Proc. Natl Acad. Sci. USA* **91** 7797–801

Shouval H, Intrator N, Law C C and Cooper L N 1996 Effect of binocular cortical misalignment on ocular dominance and orientation selectivity *Neural Comput.* **8** 1021–40

Shouval H, Intrator N and Cooper L N 1997 BCM network develops orientation selectivity and ocular dominance from natural scenes environment *Vision Res.* **37** 3339–42

Weisbuch G, Boer R D and Perelson A 1990 Localized memories in idiotypic networks *J. Theor. Biol.* **146** 483–99