Formation and Organization of Receptive fields, with an input Environment Composed of Natural Scenes

by

Harel Shouval

B.Sc Tel-Aviv Universityy 1987

M.S Weizmann Institute 1989

Thesis

Submitted in partial fulfillment of the requirements for the

Degree of Doctor of Philosophy in the Department of

Physics at Brown University.

May 1995

This dissertation by **Harel Shouval**
is accepted in its present form by the Department of Physics
as satisfying the dissertation requirements for the degree of
Doctor of Philosophy.

Date _____ _____

Prof. Leon N Cooper

Recommended to the Graduate Council

Date _____ _____

Prof. Bradley J. Marston

Date _____ _____

Prof. Charles Elbaum

Approved by the Graduate Council

Date _____ _____

Vita

Born May 21 1961, Tel-Aviv, Israel

Education

- B.Sc Tel-Aviv University 1987

- M.Sc Weizmann Institute 1990

Publications

- Shouval H., I. Shariv ,T. Grossman, A. Friesem and E. Domany An All-Optical Hopfield Network: Theory and Experiment *International Journal of Neural Systems* Vol 1. p. 355-360

- Liu Y. and H. Shouval Localized Principal Components of Natural Images - an Analytic Solution *Network* Vol 5.2 p.

- Shouval H. and Y. Liu How does retinal preprocessing Affect the receptive field of a stabilized Hebbian Neuron *Proceedings of the CNS*94 conference* [To Appear].

- Shouval H. and Perrone Post Hebbian Learning Rules *The Handbook of Brain Theory and Neural Network* Editor M. Arbib [To Appear].

Acknowledgments

I would like to thank Professor Leon N Cooper for urging me to examine the effect of introducing a more realistic visual environments, and for asking me to make sense of some of the network models, and see how they relate to the BCM model. I would like to thank him also for sharing with me his ability to figure out what is important in a complex detailed situation, I hope some of it may have rubbed off on me.

I like to express my appreciation to Professor Eitan Domany who was my advisor at the Weizmann Institute, and who first introduced me to the world of scientific research, from him I learned the importance of a good understanding of the fundamentals.

I Would also like to thank all the members of the Institute of Brain and Neural Systems, and in particular to, Yong Liu with whom I cooperated a lot in the last three years, to Nathan Intrator and Mike Perrone for many fruitful discussions, to Charlie Law who was very helpful in setting up the natural images data base, and for Roger Blumberg for carefully reading a draft of this thesis. My thanks to Proffesors Bob Pelcovits David Gottlieb and to Igor Bukharev for help and comments on some of this work.

I would express my deepest thanks to my family; Rutie who helped even in the hardest times, though she would have preferred I did something comprehensible, to Ofer cause he is the cutest, and most curious kid in the world, and most of all to my parents who brought me up with a lot of love, and who encouraged my curiosity from the earliest age I can remember.

Contents

Vita iii

Acknowledgments iv

1 Introduction 1

1.1 Review of the Experimental Results 3

1.2 Review of previous models 7

1.3 Hebb's Rule and Its Instability 11

1.3.1 Stabilized Hebb Rule 12

1.3.2 The Two Channel Problem 14

1.3.3 Finding Multiple Principal Components 15

1.4 The BCM learning rule 15

1.4.1 BCM, and Natural Scenes 17

1.5 Comparing Oja's rule and BCM 17

1.6 An Environment Composed of Natural Images 20

2 A single Cell theory of PCA Neurons Trained with Natural Images 24

2.1 Introduction 24

2.1.1 The Method of Solving the Rotationally Symmetric problem 25

2.2 Principal Components of Natural Scenes 26

2.2.1 The Rotationally Symmetric Solution 27

2.2.2 Retrieving the Phase 28

2.2.3 Discussion 30

2.3 Including Retinal Preprocessing 30

2.3.1 The Radially Symmetric Solution 32

2.3.2 Results and comparison to simulations 33

2.3.3 An angular correction 34

2.3.4 Discussion 36

3 Synaptic Plasticity and the structure of the Visual Cortex **38**
3.1 Introduction 38
3.2 A self organizing network model 39
3.3 Analysis of the Common Input model: 41
3.3.1 The Excitatory case $I \geq 0$ 44
3.3.2 Interactions with Inhibition 47
3.4 Related Models 50
3.5 Discussion 52

4 Simultaneous formation of Binocularity and Orientation Selectivity **56**
4.1 Introduction 56
4.2 Model of Binocular inputs 57
4.3 BCM rule: results and robustness 59
4.4 PCA rule: results and robustness 62
4.5 Discussion 64

5 Conclusions **69**

A Appendix **72**
A.1 Representing the correlation matrix in a Bessel function Base 72
A.2 Finding the correlation function of the preprocessed images. 74

B Appendix **75**
B.1 BCM : Fixed Points in a Symmetric Distributions 75
B.1.1 Stability of the fixed points 76
B.1.2 Distributions Symmetric around $\mathbf{x}^0$ 78

Chapter 1

Introduction

In this thesis we will discuss the questions of the formation and organization of receptive fields in the visual cortex. The models discussed herein are based on abstract forms of synaptic plasticity. Much work has been done using such models, some of which will be reviewed later in this chapter. **The central theme of this work is the incorporation of a more realistic visual environment, composed of natural scenes in such Models.**

In this work we examine the simplest models, that can capture the phenomena we are trying to explain. However, such models should not be too simple. Simplifying assumptions can sometimes bypass the real difficulty encountered by the brain. In this thesis we have systematically tried to improve on one of the vast simplifications made in most models: the nature of the visual environment.

Many of the previous models of plasticity (von der Malsburg, 1973, eg), as is discussed in section 1.3, have used a highly simplified, low dimensional model of the environment. The use of such a simplified model environment is a valid first step in examining the performance of these models, however the choice of these input representations can alter the nature of the problem. Another set of models (Linsker, 1986, eg:) has assumed that what drives plasticity, is not patterned inputs, but rather that the inputs originate from spontaneous noise activity in the photoreceptor layer. These models attempt to model pre-eye-opening plasticity; however since much of the receptive field structure develops after eye opening, these models are incomplete.

We have chosen to model and analyze plasticity in an environment composed of natural scenes, which we believe is a relatively realistic assumption for the environment affecting post eye-opening plasticity. As can be seen in the next section, we have good reason to believe that post eye opening plasticity, plays an important, although not exclusive, role

in the formation and organization of receptive fields in the visual cortex.

Previous papers have employed information about the statistics of natural images (Atick and Redlich, 1992; Bialek et al., 1991, eg:) but these works are different since they take a more abstract, information theoretic point of view, as will be discussed later in this chapter.

This work addresses two related questions: Can plasticity explain the formation of oriented receptive fields, and can it explain the organization of receptive fields across the visual cortex. The minimal model for investigating formation of receptive fields is a single cell model, such models will be discussed in chapter 2. In order to examine how the organization of different types of receptive fields across the cortex comes about, we have to assume a network of interconnected neurons. Such a network will be analyzed in chapter 3. In chapter 3 we also show how network effects can be partially decoupled from the single cell effects and how some phenomenological models of receptive field organization, relate to the synaptic neuronal model which we analyze. In chapter 4 we address the formation of oriented binocular receptive fields; here again we have chosen a single cell model.

In order to make our models complete we have to determine the learning rule employed by these neurons. In this thesis we concentrate on the learning rule proposed by Oja (82), since it extracts the principal component of the images, it will often be called PCA. This rule is can be analyzed, and the results obtained for this rule are relevant to a whole family of Hebbian like learning rules. In chapter 4 we will compare the performance of this learning rule to the BCM rule (Bienenstock et al., 1982) in the case of binocular inputs.

The remaining section of this chapter will be devoted to reviewing some of the background material essential for understanding this work and how it relates to other work done in the field. In section 1.1, we give a brief review of the essential experimental results, and section 1.2 reviews much of the previous work on this subject. In section 1.3 we discuss the learning rule proposed by Hebb (Hebb, 1949), along with it's problematic features, and how these problems were overcome by some of the newer variant of the Hebbian rule, and in particular by the Oja (1982) rule. Another learning rule which was proposed, the BCM rule (Bienenstock et al., 1982), is explained in section 1.4, and the fundamental differences between this rule and the Oja rule are discussed in section 1.5. The last section of this chapter, section 1.6, will describe some of the properties of a learning environment composed of natural scenes.

1.1 Review of the Experimental Results

In this work we try to explain how receptive fields form in the primary visual cortex. To understand this it is necessary to have some knowledge about the anatomy and physiology of the visual pathway, leading from the eye to the visual cortex. In order to asses the results it is also important to know something about the physiology of the visual cortex itself. In this section we give a brief, over simplified and review of these subjects.

Figure 1.1 shows the anatomy of the visual pathway. When Light falls on receptors in the retina, the signal is then altered by the retinal circuitry and is projected by the ganglion cells onto the optic nerve. Signals from both eyes cross at the Optic Chiasm and continue to the Lateral Geniculate Nucleuos (LGN). In the LGN inputs from both eyes are segregated to different layers. From the LGN, signals project up to the visual cortex. Alongside this forward projection, there is also a massive backward projection from the cortex, onto the LGN, the function of which is unclear, and which we will not discuss further.

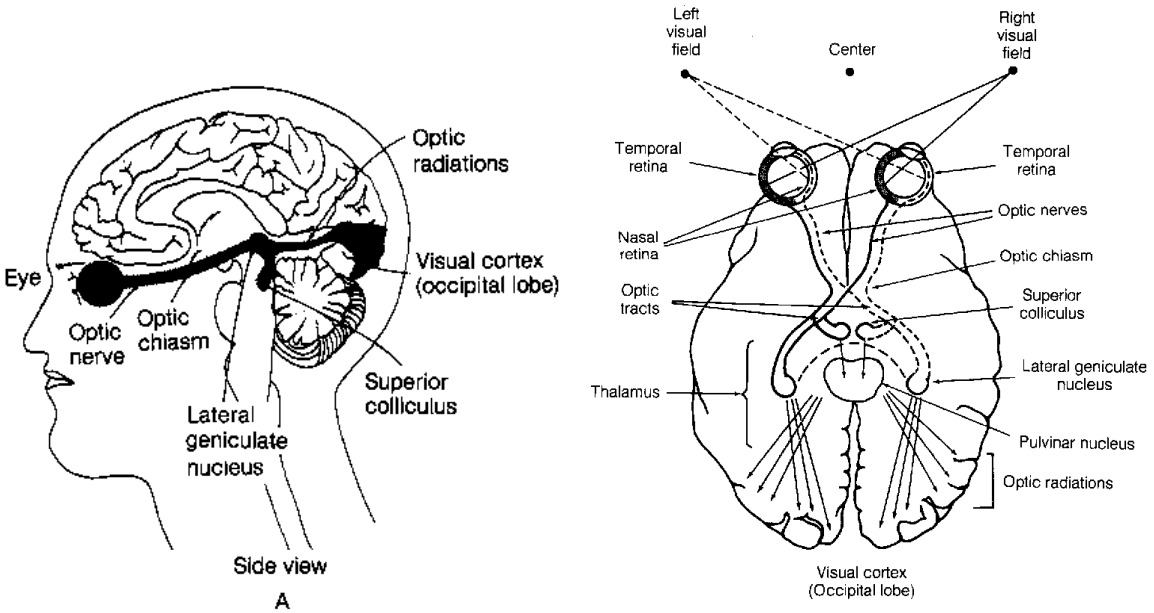

Figure 1.1: *The Visual pathway* (adapted fro: Sensation and Perception by S. Coren, C. Porac and L.M. Ward).

The receptive fields of both ganglion cell and LGN projections, have a center surround shape (figure 1.2). Some are "on center" which means that they are excited by spots of light falling on their centers, and inhibited by light on the surround, others are "off center" which are excited by light falling on their surround, and inhibited by light in their

center. There are several types of projections which differ in their size, velocity sensitivity, linearity and response time. In the cat the most abundant type near the fovea are the X cells which are linear, i.e: their response to an image composed composed of several components is the sum of the response to the components, and small (Orban., 1984).

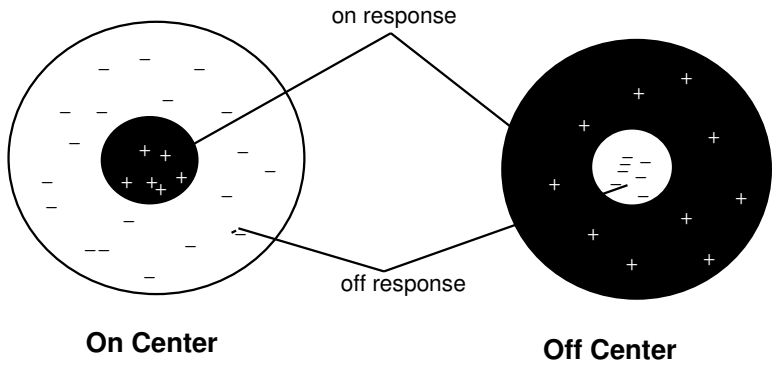

Figure 1.2: *Schematic drawing of On-Center and Off-Center, retinal ganglion cell receptive fields.*

In the cortex the receptive fields have a qualitatively different structure. They are no longer radially symmetric, but are oriented; they have can be shown to respond strongly to moving bars (Hubel and Wiesel, 1959), and sinosoidal gratings (Campbell and Robson, 1968). Cortical cells also come in different flavors: simple cells which are linear, and complex cells which are not. [1] The linear cells are called simple cells, and for them it is possible to plot the complete receptive field using spots of light. Several such receptive fields, as obtained by Palmer and Jones (87) are shown in figure 1.3.

Receptive fields in the visual cortex may respond to inputs from both eyes; those that do are called binocular. There are different degrees of binocularity in the cortex, In figure 1.4 an ocular dominance histogram of a normally reared cat is presented.

Many of the properties of the cortical neurons depend on the nature of the visual environment to which the neurons are exposed. It has been shown by Hubel and Weisel (65) that if cats are monocularly deprived, (e.g: if one of their eyes is sutured) then binocular cells will fail to develop.

But not only binocularity can be altered by experience. Many researchers have shown that lack of visual experience results in cells which are much less selective. Barlow and Pettigrew (71) found that there is a lack of specificity in visually inexperienced neurons. Blakemore and Van-Sluyters (75) found that in visually inexperienced cats about 20% of the cells are truly orientation selective, but they are typically much more broadly tuned then in

[1] The response of linear cells can be superimposed i.e: the response to two different stimuli is the sum of the responses to each separately.

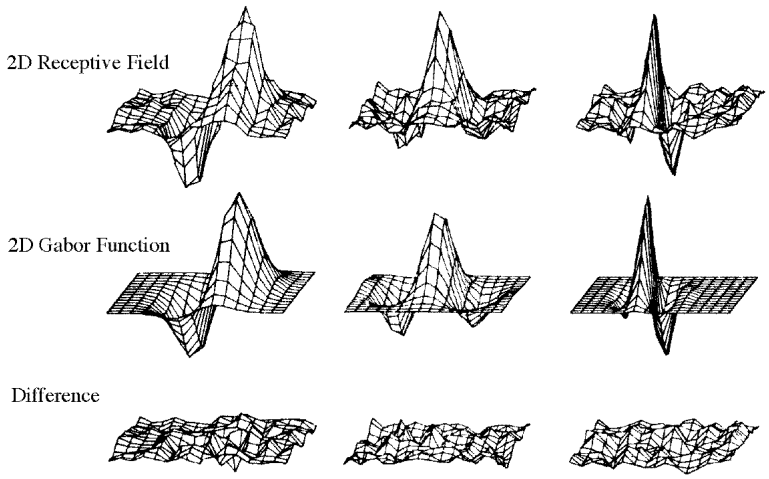

Figure 1.3: *(From Jones and Palmer [87]) The images on top represent the three dimensional structure of the actual receptive fields, the central row shows the best fit to a Gabor function, and the bottom row presents the difference between the first and second rows, in order to show that the difference is merely noise.*

visually experienced cats. They have also shown that by varying the type of environment to which the cats are exposed (e.g: Bars, spots), they can dramaticly alter the shape of the receptive fields.

Different results were obtained in a later study by Brasstad and Heggelund (85) who seemed to find results similar to those found initially (Hubel and Wiesel, 1963b), they reported that before eye opening, most of the cells were highly selective, however this study made no distinction between broadly and sharply tuned cells, or between orientation and direction selectivity

Imbert and Buisseret (75) observed that in dark reared cats, the orientation tuning initially increases after eye opening, as in normally reared cats, but then drops down below the initial levels. This stands in sharp contrast to what happens in normally reared cats.

As has been pointed out by Blakemore and Van-Sluyters (75), it seems that many of the differences between the experimental studies, may result from different definitions of selectivity, as well as from the different cortical layers that were recorded. Thus we think it sensible to adopt their conclusions about the matter, from which we will quote two of their five conclusions.

"(a) About a fifth of the cells we can record in young inexperienced cats, are truly orientation selective, though very broadly tuned by adult standards." "(d) There is little or no passive maturation of this system in the absence of patterned retinal images."

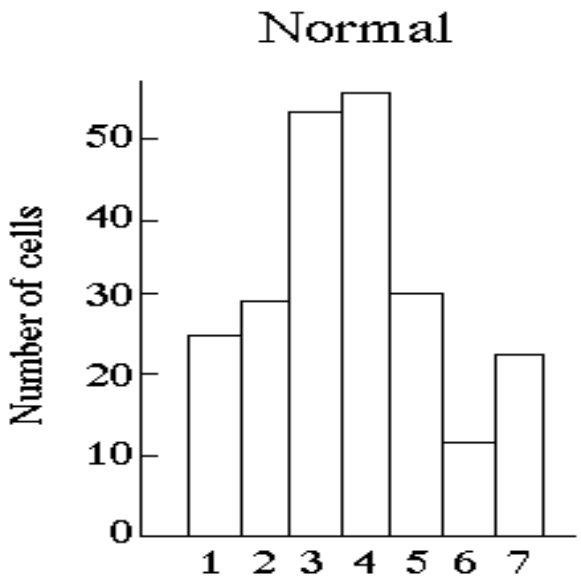

Figure 1.4: *A typical ocular dominance histogram for a cat, cells in the central bin (4) are binocular, whereas cells in bins 1 and 7 are monocular. Different methods of classification can result in slightly different shapes for this histogram, and the eccentricity at which the were measured can possibly change this shape as well. Adapted from (Wiesel and Hubel 62)*

In other words it seems clear that at least for cats, the nature of the visual environment is critically important in determineing the shape of the cortical receptive fields.

However for some other animals, such as rabbits, (Chow and Spear, 1974) the development of cortical receptive fields seems independent of visual experience. It is interesting to note that in rabbits the receptive fields from both eyes are not matched, e.g a cortical neuron in a rabbit is likely to have different optimal orientations in each eye.

Apart from the structure of single cell receptive fields, the cortex is characterized by the way the different receptive fields are organized across the cortex. With optical imaging, it is now possible to get a good characterization of this structure.

In figure 1.5 we can see how neurons with different preferred orientations are ordered across the cortex. The results displayed are for monkey but orientation maps for cats (Bonhoeffer and Gringald, 1991) were obtained by similar methods, and are qualitatively similar. In figure 1.6 we see how different ocular dominance preferences organize in the monkey, cat striate cortex is quite different in this respect yielding ocular dominance columns with smoother boundaries. These structures were termed orientation columns and ocular dominance columns respectively.

Careful analysis of maps of the visual cortex, obtained through optical imaging (Erwin et al., 1994) has revealed many detailed features of these maps, including the observation that both types of maps, the orientation, and ocular dominance maps are correlated.

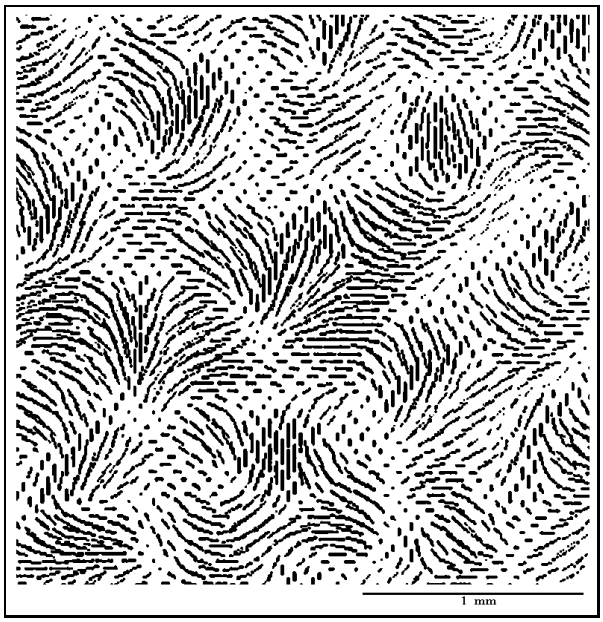

Figure 1.5: *Results of optical recording, from Blasdel '92 showing orientation preference of a small portion of striate cortex. The angle of the line segments show preferred orientation, and the segment lengths shows selectivity.*

We will not discuss these fine grained distinctions here.

1.2 Review of previous models

Over the years there have been numerous models, which attempt to explain the formation and organization of receptive fields in the cortex. Some are detailed synaptic models that assume different variants of the Hebb learning rule (Hebb, 1949); these rules will be discussed in detail in the next sections. In addition these models differ in their assumptions about the visual environment, and the network architectures. Other models which attempt to explain the organization of receptive fields in the cortex, are phenomenological, in that they do not explicitly define a detailed learning rule, and have neither synapses nor inputs. Instead the cortical sheet is defined as a sheet of a single, or several parameters, such as orientation of single cells, or their ocular dominance, these models differ in the organizing principles of such fields.

Many of the first attempts at simulating synaptic models have assumed a simplified model of the input environment, Malsburg (73) for example assumes a finite set of oriented bars (von der Malsburg, 1973). Others models assume a set of abstract overlapping patterns plus noise (Nass and Cooper, 1975; Cooper et al., 1979; Clothiaux et al., 1991).

A different family of models has assumed a more abstract, low dimensional input

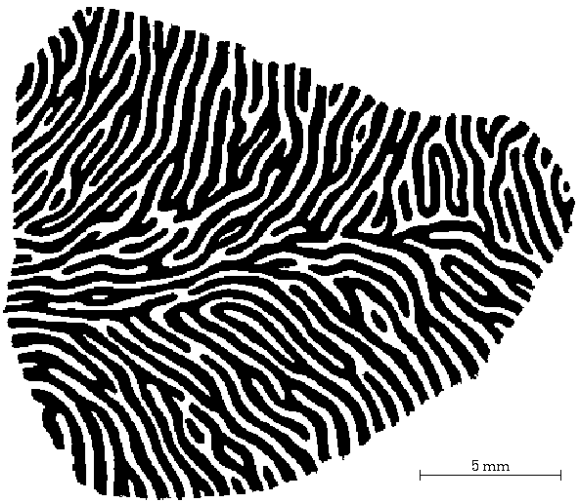

Figure 1.6: *An ocular dominance map for a monkey reconstructed from slices of striate cortex. Neurons were labeled with the radio-active dye, proline-H^3, which was transported to cortex after being injected into one eye.* (Adapted from Levay Et Al [78])

structure. Kohonen (82) assumes input patterns which are two dimensional, and intend to represent different orientations of bars. Others (Obermayer et al., 1992) have used inputs of slightly higher dimensions, where the added dimensions represent other attributes such as which eye the input comes from etc. The inputs in all of these models intend to model normal rearing, i.e., rearing in an environment composed of objects and patterns.

Another family of models (Linsker, 1986; Miller et al., 1989; Miller, 1994) has attempted to explain the receptive field structures which form before eye opening. In these models it is assumed that plasticity is driven by noise originating in retinal neurons, even in the absence of a patterned environment. Another feature common to these models is that they are presented in a correlational form, that is, their authors have averaged over the input ensemble, which results in a deterministic model, which is critically dependent on the form of the correlation function of the input layer.

In this thesis as in our recent publications, (Liu and Shouval, 1994; Shouval and Liu, 1994; Law and Cooper, 1994) we have attempted to explain plasticity after eye opening, when the input environment is composed of natural scenes. The models investigated by Linsker and Miller also employ high dimensional inputs, and attempt to give a relatively realistic model for inputs originating from the two dimensional retinal sheet before eye opening.

Most of the aforementioned models have attempted to show how the organization of receptive fields could emerge from plasticity (von der Malsburg, 1973; Nass and Cooper, 1975; Kohonen, 1982; Scofield and Cooper, 1985; Obermayer et al., 1992; Miller et al., 1989;

Inputs	Learning Rule	Interactions	Model
Symbolic (Low dimensional)	Hebb $\|L^2\|$ Normalization	Network Non Linear	Kohonen 82 Obermayer Et. Al 92
Simplified	Hebb $\|L^1\|$	Network, Linear	Malsburg 73
	Hebb fixed threshold	Single Cell + Network	Nass and Cooper 75
	BCM	Single Cell	BCM 82 Clothiaux Et. Al. 92
	BCM	Network	Scofield and Cooper 85
High Dimensional Pre eye opening	Hebb Saturating	Network Linear	Linsker 86 Miller Et. Al. 89 Miller Et. Al. 94
High Dimensional Post eye opening	Hebb $\|L^2\|$	Single Cell + Orthgonalizing Network	Hanckock Et. Al. 92 Liu and Shouval 94 Shouval and Liu 94
	BCM	Single Cell	Law and Cooper 94

Table 1.1: *Classification of different synaptic models of formation and organization of receptive fields in the cortex, the model*

Miller, 1994) and (Linsker 86 pg: 8779-83). In these models the cortex was assumed to be a sheet of interacting neurons.

Usually the interactions between the cortical neurons are assumed to be constant, i.e: not affected by plasticity. This is done primarily because modeling plastic intercortical synaptic weights would complicate the models, making them computationally much more expensive, and since fixed intercortical connections are usually sufficient for producing the desired organization. However, the Kohonen type models, (Kohonen, 1982; Obermayer et al., 1992) have assumed a shrinking interaction radius. This reduction in the interaction radius is not equivalent to plasticity in the cortical layer, since the change is independent of the cortical activity. The only published attempt far to model real intracortical plasticity, is by Sirosh and Miikkulainen (94) (Sirosh and Miikkulainen, 1994). Their model uses abstract inputs.

In table 1.1 we have a different aspects of these synaptic models.

A completely different approach has been taken by a family of models which we term phenomenological. What characterizes these models, is that the have no synapses defined, and no inputs as well. These models are meant to explain the organization of receptive fields across the cortex, rather then their formation. In these models (Swindale, 1982; Swindale, 1991; Cowan and Friedman, 1991; Rojer and Schwartz, 1990; Neibur and Wögötter, 1994) each neuron in the cortical sheet is defined by a parameter which represents the direction

of it's orientation preference. In some of these models (Swindale, 1982; Swindale, 1991; Cowan and Friedman, 1991) the directions are obtained by an iterative process, in which the preferred orientation of a neuron changes as a function of it's neighbors orientation. In Chapter 3 we show that some of these models are related to the synaptic models. The other models (Rojer and Schwartz, 1990; Neibur and Wögötter, 1994) are presented in Fourier space, where it is shown that in order to obtain an organization of receptive fields which are similar to cortical ones, it is sufficient that the power spectrum of the orientation fields have a toroidal shape, i.e: it should be zero as the frequency $k \rightarrow 0$ and drop fast enough to zero as the frequencies approach infinity.

Another distinct approach is to derive the receptive field structures from first principles that is to define some optimality criterion, and find the receptive fields which maximize it. Typically this sort of approach relies heavily on knowledge about the statistics of the environment.

Different optimality criteria have been used by different groups, in order to derive receptive fields in early visual processing. Bialek et. al (91) have used the maximum information transfer in the low SNR limit, to derive spatial receptive fields and van Hatern (van Hateren, 1992; van Hateren, 1993) used a similar more detailed theory to derive spatio-temporal receptive fileds of fly eyes. Works by Atick and his coworkers (Atick and Redlich, 1990; Atick and Redlich, 1992; Atick et al., 1992), have used the decorrelation principle proposed by Barlow (Barlow, 1961; Barlow, 1989), to explain retinal receptive fields. Recently Li and Atick (94) have made the additional assumption, in order to derive cortical receptive fields, that at the cortical level multiscale representation exists. This approach does not depend on plasticity, indeed these receptive fields may have formed across evolutionary time scales, and not as a result of plasticity.

In order to make a sensible choice of an optimality criterion, it is important that the function of the cortical system be known. In this respect there is a major difference between theories which attempt to explain retinal receptive fields, and those which attempt to explain cortical receptive fields. The anatomy of the cortical pathway suggests that it is indeed an information bottleneck, and that information transfer must be optimized through this layer. However the computational role of the visual cortex has not been identified as clearly, and therefore optimality criteria in the cortex seem, to us, more speculative. Therefore it would come as no surprise to us, that the retinal receptive fields derived in these theories (Atick and Redlich, 1990; Atick and Redlich, 1992; Atick et al., 1992; van Hateren, 1992; van Hateren, 1993) seem to match experimental results closely, while those derived in the cortical theories (Li and Atick, 1994) show little resemblance to cortical receptive fields.

1.3 Hebb's Rule and Its Instability

The original Hebb rule, states how synapse efficacies are strengthened (Hebb, 1949):

> When an axon in cell A is near enough to excite cell B and repeatedly and persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency in firing B, is increased.

Later, it was suggested that synaptic strengths may decrease when the presynaptic cell fails to fire the postsynaptic cell (Stent, 1973). In a slightly more general context, this implies that synaptic strengths change as a function of the correlations between the pre- and post- synaptic neurons; they increase when the activities of the neurons are correlated, and decrease when they are anticorrelated. This generalized Hebbian rule is the cornerstone of many unsupervised learning schemes (von der Malsburg, 1973; Rochester et al., 1956; Linsker, 1986; Oja, 1982, eg:).

A mathematically simple form, which will nevertheless be general enough for understanding the basic properties of such learning rules, is the linear rule (Linsker, 1986)

$$\Delta m_i = -\eta(d_i - d_o)(c - c_o) \tag{1.1}$$

where d_i are the activities of presynaptic neurons, c is the activity of the postsynaptic neuron, m_i is the value of the synaptic efficacy between presynaptic neuron i and the postsynaptic neuron, and η, is the learning rate, and c_o and d_o are constants which may be related to spontaneous activity.

This concept is easiest to analyze in terms of single postsynaptic linear neurons, in which the neuron's activity is given by

$$c = \sum_i m_i d_i. \tag{1.2}$$

For this case, we can calculate the average of the synaptic efficacies over the input probability distribution. We denote this average by $E[\cdot]$. This average weight change is of interest to us since we define the fixed points as those values of m_i for which the average weight change is zero.

Inserting Equation 1.2 into Equation 1.1, averaging the result and assuming the same probability distribution for each input, i.e., assuming $E[d_j] = E[d_i] = \mu$, we obtain that

$$E[\Delta m_i] = -\eta \left(\sum_j m_j E[(d_i - \mu)(d_j - \mu)] - k_2 \sum_j m_j - k_1 \right) \tag{1.3}$$

$$= \quad -\eta \left(\sum_j (Q_{ij} - k_2) m_j + \hat{e} k_1 \right),$$

where $Q_{ij} = E[(d_i - \mu)(d_j - \mu)]$ is the correlation function of the input patterns, $k_1 = c_o(d_o - \mu)$, and $k_2 = \mu(d_o - \mu)$. It is important to notice that k_1 and k_2 are not free parameters which can be set arbitrarily. For instance, if $d_o = \mu$ then $k_1 = k_2 = 0$. This equation can also be put in the more convenient matrix form,

$$E[\Delta \mathbf{m}] = -\eta(Q - k_2 J)\mathbf{m} + \eta k_1 \hat{e} \tag{1.4}$$

where J is the matrix $J_{ij} = 1$.

A fundamental problem with this learning rule is that it has no stable fixed points. The vector $\mathbf{m}$ will asymptotically become parallel to the eigenvector with the largest eigenvalue of the matrix $Q - k_2 J$ but its magnitude will diverge to infinity. For example if $k_2 = 0$, the weight vector $\mathbf{m}$ will asymptotically become parallel to the principal component of the input (i.e. the largest eigenvector of the correlation matrix) but its magnitude will approach infinity.

This instability is not biologically plausible since it implies that the synaptic strengths will grow to be arbitrarily large. In order to avoid this unbiological behavior, one must go beyond the basic Hebb rule by imposing additional constraints. One approach is to assume the there exists a biologically imposed saturation limit which the synapses cannot exceed. This approach, although conceptually simple and easy to implement, does not preserve one of the appealing properties of the Hebb rule: the ability to identify the principal component of the data.

Before discussing how weight explosion problem can be corrected without appealing to saturation limits, we note that some researchers have avoided this problem by assuming that the neuron learns slowly enough for the receptive field to become approximately parallel to the principal component and that learning stops quickly enough for the synapses to avoid saturation (Linsker, 1986; Linsker, 88; MacKay and Miller, 1990; Miller et al., 1989).

1.3.1 Stabilized Hebb Rule

The absence of a bounded fixed point for the Hebb rule without saturation was first noted by Rochester (Rochester et al., 1956):

> If no additional rule were made the Hebb rule would cause synapse values to rise without bound.

Normalization procedures to bound synapse strengths by supplementing the Hebb rule with additional constraints have been proposed (von der Malsburg, 1973; Kohonen, 1982). The

method chosen by von der Malsburg was to renormalize the weight vector after each Hebbian update by dividing the strength of each synapse by the sum of the synapse strengths of the neuron. Kohonen chose to divide by the sum of the squares of synaptic strengths. Dividing by the sum of squares always keeps the weights bounded, whereas dividing by the sum of the weights can fail if the principal component has a sum of weights equal to zero. These stabilization schemes seem biologically implausible since they are nonlocal: Each synapse needs to "know" the weights of all the other synapses terminating on the same neuron. A mechanism for propagating this nonlocal information is not known; however this type of stabilization can be implemented in another way (Oja, 1982). Instead of dividing after each Hebbian update, an extra term can be added to the update rule. Setting for simplicity, $\mu = x_0$ which implies that, $k_1 = k_2 = 0$, this rule takes the form

$$\Delta m_i = d_i c - \beta(\mathbf{m}) m_i, \tag{1.5}$$

where $\beta(\mathbf{m})$ is a scalar function of of the vector $\mathbf{m}$. To see what is then the fixed point of the learning rule, it is convenient to go back to the correlational formulation of the learning rule. In this formulation the fixed points of Equation 1.4 (i.e. the points where we have $E[\Delta\mathbf{m}] = 0$) are given by

$$Q\mathbf{m} = \lambda\mathbf{m} \tag{1.6}$$

where $\lambda = \beta(\mathbf{m})$. Thus the fixed points are the eigenvectors of the correlation matrix Q, and the function $\beta(\mathbf{m})$ sets the normalization of these vectors. The choice Oja has made for $\beta(\mathbf{m})$ is $\beta(\mathbf{m}) = c^2$ and thus $E[\beta(\mathbf{m})] = m^\dagger \cdot Q\mathbf{m}$ and therefore the weights will be square normalized to one, i.e: $\mathbf{m} \cdot \mathbf{m} = 1$. In fact, the only stable fixed point is the principal component. Computationally, the importance of the principal component is that it is the projection which maximizes the variance, and, in a Gaussian channel, carries more information than any other projection.

When posed in this way the learning rule seems more biologically plausible, since the stabilization requires only that each synapse have information about its own efficacy, and of the square of the postsynaptic potential.

If instead of using Oja's choice of β we choose $\beta = \hat{e} \cdot dc$, this would when averaged result in $\hat{e}Q\mathbf{m}$. Thus an eigenvector with eigenvalue λ is normalized such that $\hat{e}\mathbf{m} = 1$, which is equivalent to von der Malsburgs choice of normalization (von der Malsburg, 1973). This type of normalization function β is also problematic in the cases where $\hat{e}\mathbf{m} = 0$ and normalization is not possible.

1.3.2 The Two Channel Problem

In some of the synaptic models of formation and organization of receptive fields in the cortex, and especially in those proposed by Miller (Miller et al., 1989; Miller, 1994), more then one input channel exists. In the two types of models proposed by Miller the two channels were inputs from two eyes (Miller et al., 1989) and inputs from the on-center and off-center ganglion cells. The full correlation therefore includes the correlations within channels Q_{ii} and the correlations between channels Q_{ij}, and therefore in the two channel case has the form

$$Q = \begin{pmatrix} Q_{11} & Q_{12} \\ \\ Q_{21} & Q_{22} \end{pmatrix}. \tag{1.7}$$

As in the single channel case the final state of the receptive fields is often determined by the largest eigen state of the complete correlation function.

In some cases the eigenstates of the two channel case are closely related to the eigen states of the one channel case.

Given a one channel correlation function Q, where the eigenstates ψ solve the equations $Q\psi = \lambda\psi$, we will denote the principal component as ψ_{max}.

Miller (Miller, 1990), had analyzed the case where $Q_{11} = Q_{22} = Q$ and $Q_{12} = Q_{21} = \eta Q$. Thus, for η positive, the principal component of the two channel problem has the form $(\psi_{max},\ \psi_{max})$ with an eigenvalue $(1+\eta)\lambda_{max}$. If η is negative then the eigenstate is $(\psi_{max},\ -\psi_{max})$ and the eigenvalue is $(1-\eta)\lambda_{max}$.

Another case which has been analyzed is when the cross correlation components Q_{12} and Q_{21} are not equal to Q, but are small, i.e: $Q_{12} = Q_{21} = \epsilon Q'$ for $\epsilon << 1$. Dayan and Goodhill (Dayan and Goodhill, 1992) have shown using perturbation theory, that up to $O(\epsilon)$ the same results apply, i.e a positive cross correlation yields the eigenstate $(\psi_{max},\ \psi_{max})$ with eigenvalue $(1+\eta)\lambda_{max}$.

In the two eye problem Miller has assumed an anti-correlation between the eyes which is difficult to justify biologically. In the On-center/Off-center case he has assumed a negative correlation as well, which is justified if the noise in the receptors drives the signal, and less justified if some of the noise originates in the ganglion cells themselves. By assuming that synapses can not take negative values, he managed to use this correlation structure in order to achieve receptive fields, or regions of receptive fields which disconnect from one of the channels.

1.3.3 Finding Multiple Principal Components

Learning rules which extract only the largest principal component generally capture only a fraction of the information in high dimensional data. In order to reduce the dimensionality of the data, and yet have a relatively small reconstruction error, it is useful to extract several principal components. This can be accomplished using a network with lateral inhibition. Different forms of lateral inhibition will pick out different features. If we assume that there are P output neurons, where P is smaller than the input dimensionality, then the resulting receptive fields will span the same subspace as the first P principal components but in general will be different from the principal components themselves. In some of the models (Rubner and Tavan, 1989; Foldia, 1989) the resulting receptive fields span the subspace, but are not the eigenvectors themselves, while others find ways of extracting exactly the first P principal components(Oja, 1992; Sanger, 1989). A heuristic way of understanding these methods is to note that the lateral inhibition between the first and second neurons acts effectively to subtract from the input to the second neuron those parts of the input which result from the first neuron's receptive field, and similarly for successive neurons.

1.4 The BCM learning rule

The BCM learning rule (Bienenstock et al., 1982) could also be viewed as a stabilized Hebbian rule. The method of achieving stabilization however, is completely different, and so are the resulting receptive fields. Mathematically the BCM rule has the form,

$$\dot{m}(t) = \eta \mathbf{d}(t) \Phi \left(c(t), \theta_m \right). \tag{1.8}$$

where c is the activity of the cell and η is the learning rate. The function Φ has two zero crossings, one at 0 and the other at the value θ_m, where θ_m is itself a non linear function of the history of the activity c. Examples of such functions Φ are shown in figure 1.7. As in the Hebbian case the neurons are often assumed to be linear, i.e: $c = \sum_i m_i d_i$. For this learning rule to be stable it has been shown that θ_m, termed the moving threshold, has to be a super-linear function of the running average of the activity (Bienenstock et al., 1982; Intrator and Cooper, 1992). The function of the moving threshold is to prevent the weights from escaping to infinity, i.e: it is a negative feedback mechanism, therefore heuristically it is clear that it has to depend on activity in a super-linear manner, because otherwise it will not catch up with the growth in activity.

If the inputs have a zero mean, then the average of the activity has a mean zero as well, therefore it is preferable to use a modified version of the threshold, thus in Intrator

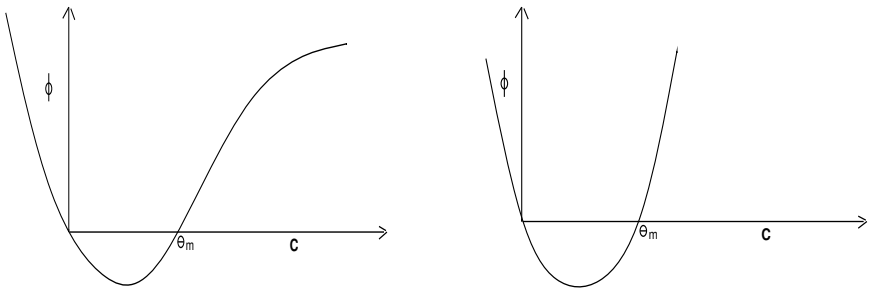

Figure 1.7: *An example of the Φ function used in Bienenstock et. al. [82] (left). It is sufficient for such a function to have two zero crossings, one at zero and the other at θ_m. The function used since Intrator et. al. [91] (right) this function is quadratic many of the properties have been proven explicitly for this function. In practice the results from both are similar.*

and Cooper (91) the definition of θ_m has been changed to the runing average of the square of the activity.

Here we adopt the same definition as Intrator and Cooper (91), and use,

$$\Phi(c, \theta_m) = c(c - \theta_m), \tag{1.9}$$

and

$$\theta_m(t) = < c^2 >_t = \frac{1}{\tau} \int_{-\infty}^{t} dt' \exp^{t'/\tau} C(t')^2. \tag{1.10}$$

With this definition of the learning rule, it has been proven (Intrator and Cooper, 1992) that when the input environment is composed of N linearly independent vectors $\mathbf{d}^i$ of dimensionality N, then the learning rule will have N stable fixed points, such that if $\mathbf{m}^j$ is fixed point weight vector, it will be orthogonal to all but one of the input vectors, i.e: $\mathbf{d}^i \cdot m^j \propto \delta_{ij}$. Furthermore, this result would not change if a small amount of Gaussian noise, with mean zero, is added to these vectors.

An interesting interpretation for this learning rule has been offered as well (Intrator and Cooper, 1992), which connects BCM with the statistical method of projection-pursuit. It has been shown that the fixed points of BCM, coincide with projection in directions of multimodality, i.e: with directions in the high dimensional space in which the projection is highly non Gaussian and preferably is concentrated in several clusters. It has been argued (Diaconis and Freedman, 1984) that such directions are interesting since they indicate structure whereas Gaussian directions are common indicators of noise.

Furthermore, it has been shown (Intrator et al., 1993) that when we average over the input distribution, the BCM learning rule takes the form,

$$\dot{m}_i(t) = \eta \left(\sum_{i,j} Q_{ijk} m_j(t) m_k(t) - \sum_{jkl} Q_{ij} Q_{kl} m_k(t) m_l(t) \right). \tag{1.11}$$

where Q_{ijk} is the three point correlation function, i.e: $Q_{ijk} = E[d_i d_j d_k]$, and where the average is taken over the input distribution.

This form stands in contrast to the averaged form of PCA learning since it depends on third order statistics, and not only on the covariance function as does the PCA rule.

1.4.1 BCM, and Natural Scenes

In order to obtain receptive fields which resemble cortical receptive fields, when the input environment is composed of natural scenes (Law and Cooper, 1994), assume an activation function σ which is non linear and non symmetric around zero (Law and Cooper, 1994), which is consistent with the properties of cortical neurons. Therefore for natural scenes we have abandoned the linear approximation for the activation function, and now use,

$$c = \sigma(\sum_i m_i d_i) \tag{1.12}$$

We have chosen to use the function σ as shown in figure 1.8. The essential feature of this function is that it has a low saturation value which is close to 0, whereas it has a much higher upper saturation value, therefore, a non symmetric sigmoid, with a positive saturation limit being much higher then the negative one, is also sensible, and would not alter the results.

This type of activation function is sensible for the cortex, since if spontaneous activity is defined as the difference between the minimal activity, and the average over the data set, it implies that the neurons have a low spontaneous activity, as do cortical neurons.

We have found, as discussed in appendix B, that if we assume that the input environment is symmetric around a point $\mathbf{d}_0$ [2] then the BCM rule with a symmetric neuron would lead us to trivial fixed points. These fixed points coincide with those found in simulations using preprocessed natural scenes. However, assuming a low spontaneous activity breaks this symmetry, and makes it possible to achieve receptive fields similar to those in the cortex.

1.5 Comparing Oja's rule and BCM

Since both the BCM learning rule, and some other versions of the Hebbian learning rule, such as Oja's rule, stabilize the Hebbian rule, it is sometimes assumed that they are similar.

[2] i.e: there is the same probability of having the vector $\mathbf{d} + \mathbf{d}_0$ as the vector $\mathbf{d} - \mathbf{d}_0$

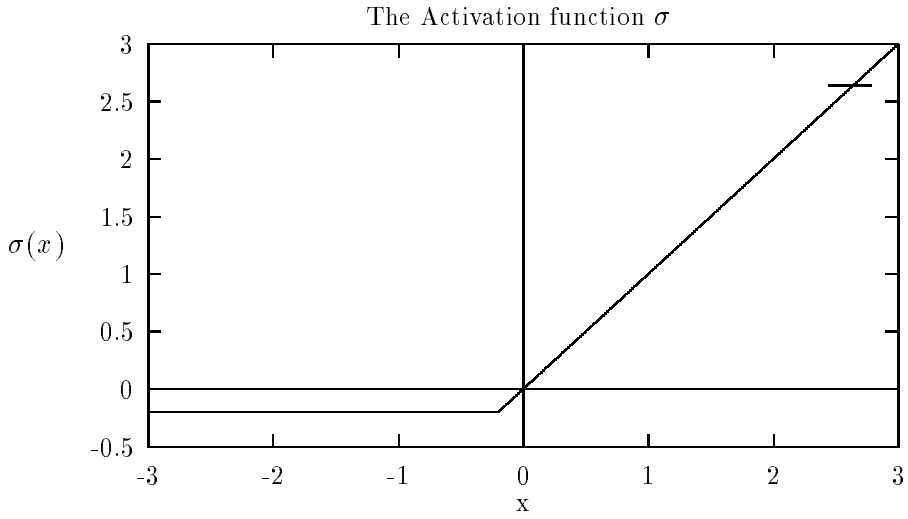

Figure 1.8: *The form of the activation function σ which has been used in our BCM simulations, in an environment composed of natural images. This form is consistent with properties of cortical neurons, which have a low spontaneous activity.*

We have already discussed some of the differences in previous sections, but we will point them out again in this section, and illustrate them with a simple example.

One difference, which relates to experimental findings, is that the decay term in Oja's rule means that a synapse for which there is no presynaptic input may decay; this is perhaps equivalent to what biologists have termed hetrosynaptic long term depression (LTD). In the BCM rule,, in contrast only homosynaptic LTD is needed, i.e only synapses which have had a presynaptic input can decay.

Another difference is that the receptive fields in the Oja rule depend only on the second order statistics, and in particular they are the largest eigenvector of the correlation matrix. The receptive fields in the BCM rule depend on third order statistics as well (1.11), and present direction in which there is a highly non Gaussian moltimodel projection.

It is easy to illustrate how different these learning rules are with an elementary two dimensional example. Suppose there are two 2D inputs $\mathbf{d}^1$ and $\mathbf{d}^2$, of unit length, such that the angle between them is α, as is shown in figure 1.9. For simplicity lets assume that $\alpha < \pi/2$. In this case it is easy to prove that for PCA the fixed points would be such that $\mathbf{m}^o \propto \pm(\mathbf{d}^1 + \mathbf{d}^2)/2$. For BCM however, the stable fixed points would be such that either $\mathbf{m}^b \cdot \mathbf{d}^1 = 0$ or $\mathbf{m}^b \cdot \mathbf{d}^2 = 0$, as is shown in figure 1.9.

Thus it can be seen that using the Oja rule, the receptive field responds equally to

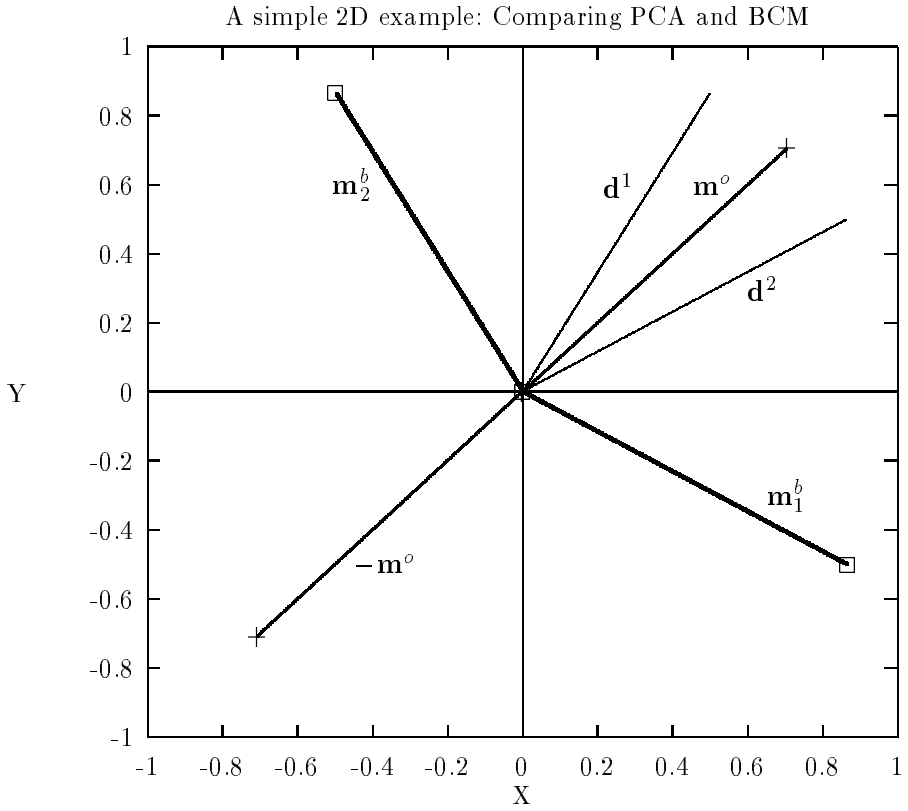

Figure 1.9: *Comparing PCA and BCM, in a simple 2D example. The thin lines represent the inputs* $\mathbf{d}^1$ *and* $\mathbf{d}^2$*, the thicker are the Oja rule stable fixed points,* $\pm\mathbf{m}^o$*, and the thickest are the BCM stable fixed points* $\mathbf{m}_1^b$ *and* $\mathbf{m}_2^b$*.*

both inputs, in contrast to the BCM receptive field which responds to only one input, and does not at all to the other. This difference generalizes, under similar conditions, to higher dimensional examples as well.

It might be thought that BCM cells are trying to become totally selective i.e: cells which respond only to one input. This is impossible in realistic situations in which the number of possible inputs greatly exceeds the dimensionality of the input. However, BCM cells do respond to a small fraction of the inputs, and in that sense they tend to form a sparse representation. PCA neurons in contrast point to the direction of maximum variance, and thus tend to maximize the projection of as many as possible input vectors onto the receptive field. Therefore they tend to form a relatively denser representation.

1.6 An Environment Composed of Natural Images

As explained in the first part of this chapter the focus of this work is plasticity in the presence of a realistic input environment. Thus when dealing with post eye-opening plasticity the relevant environment is the type of environment the animal is likely to encounter namely Natural Scenes. In order to perform our simulations, and our analysis we have used a set of 24 scenes, some of which are displayed in figure 1.10. These pictures were taken at Lincoln Woods State Park, and later scanned and digitized. For different uses we have used a resolution of 256 X 256 pixels or of 512 X 512 pixels, we have not used any corrections for the optical distortions of the instruments used. We have avoided man-made objects, because they have many sharp edges, and straight lines, which would make it easier to achieve oriented receptive fields.

The nature of the Correlation function of Natural Scenes was first investigated by Field (1987), who found that the correlation function of natural scenes is scale invariant [3] , i.e: that the power-spectrum of natural scenes has the form $\tilde{Q}(\mathbf{k}) = 1/k^2$. The scale invariance can be put into the form of the differential equation

$$\nabla^2 Q(\mathbf{r}) = -c\delta(\mathbf{r}). \tag{1.13}$$

The two dimensional radially symmetric function which solves this equation is

$$Q(\mathbf{r}) = c(-log(r/a) + b). \tag{1.14}$$

Although b can not be obtained from equation 1.13, it can be obtained directly from the data. In order to examine this, and find the value of b, we have extracted the correlation function from a set of natural images. The results are displayed in figure 1.11, the different points, in the graph, correspond to all different correlation lengthes at different angles.

However the correlation function has a non-radially symmetric component as well, as can be seen on the right hand side of figure 1.12 where the complete spectrum is presented (Field, 1987, see also). We will see later that in certain cases this non symmetric component is important. In order to examine quantitatively the non symmetric part of the correlation function we will assume that the Fourier transform of the correlation function, i.e: the power spectrum of the images, can be decomposed into angular modes:

$$\begin{aligned} \tilde{Q}(\mathbf{k})^2 &= \tilde{Q}_0(k) + \epsilon \sum_{l\neq 0} |\tilde{Q}_l(k) e^{il\theta} \\ &= \tilde{Q}_0(k) + \epsilon \left[\sum_{l=1} \tilde{D}_l(k)\cos(l\theta) + \tilde{E}_l(k)\sin(l\theta) \right]. \end{aligned} \tag{1.15}$$

[3] A later extensive study (Ruderman and Bialek., 1994), in which not pixel intensity, but the *log* of pixel intensity is the variable measured, concluded that for this variable this is only approximately true.

Figure 1.10: *Eight of the natural images used to represent the visual environment.*

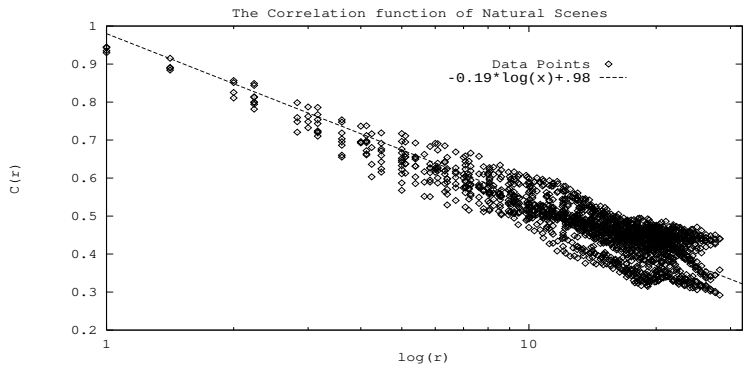

Figure 1.11: *The correlation function of natural scenes in coordinate space. The different points correspond to correlation lengthes at all different angels, and as will be seen below, the radial asymmetry of the environment contributes to the spread of points.*

Where $\tilde{Q}$ is the power spectrum of the images. This form can then be used to find the different angular modes of spectrum i.e find the different $Q_l(k)$ functions.

On the right side of figure 1.12 we display results of this analysis when performed on 20 images from our data set of natural scenes, for the first two even modes. The odd modes vanish due to the mirror symmetry.

This decomposition would be used in the next chapter, where it will be shown that under certain conditions the non-symmetric portion is essential for understanding receptive field structures.

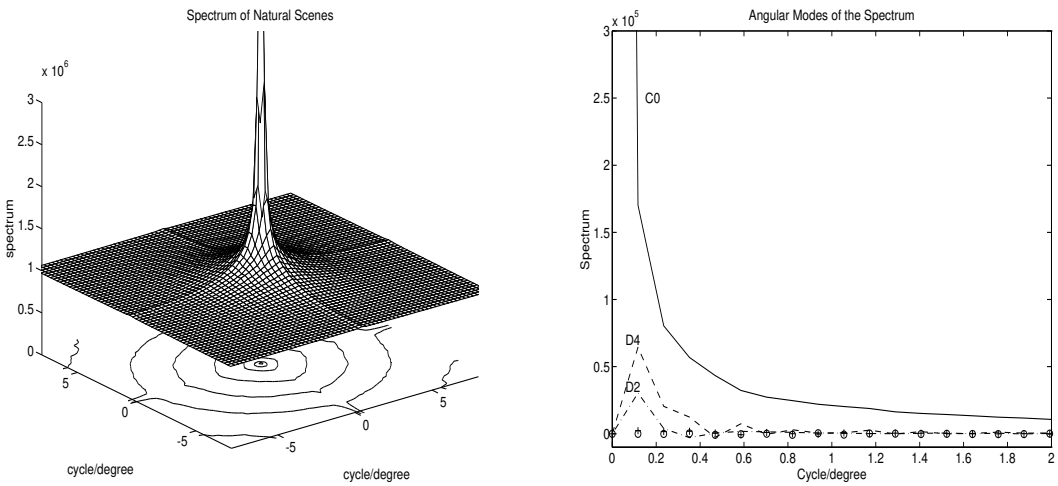

Figure 1.12: *On the left the spectrum of our 20 raw images scanned at 512 X 512 pixels, the spectrum has been shifted up, and below we can see a contour plot of the* **log** *of the spectrum. On the right the decomposition of the correlation function into angular modes is displayed, the odd modes are not displayed because they vanish, due to the reflection symmetry of the spectrum. The solid line represents* $\tilde{\mathbf{Q}}_{\mathbf{0}}(\mathbf{k})$*, the "-."=*$\tilde{\mathbf{D}}_{\mathbf{2}}(\mathbf{k})$ *, the "+"=*$\tilde{\mathbf{E}}_{\mathbf{2}}(\mathbf{k})$*, the "- -"=*$\tilde{\mathbf{D}}_{\mathbf{4}}(\mathbf{k})$ *and "o"=*$\tilde{\mathbf{E}}_{\mathbf{4}}(\mathbf{k})$*.*

Chapter 2

A single Cell theory of PCA Neurons Trained with Natural Images

2.1 Introduction

In this Chapter we analyze the receptive fields of neurons trained with PCA type learning rules (Oja, 1982; Oja, 1992; Sanger, 1989), when their inputs are composed of natural scenes. In section 2.2 we analyze the case in which the inputs are small circular patches extracted from the natural images. Since inputs coming into the cortex are preprocessed by the retina and LGN as mentioned in the introduction, this assumption about the inputs is incomplete. Therefore, in section 2.3 we examine how preprocessing similar to that found in the retina affects the receptive fields of such neurons.

Our approach in this chapter is to use the knowledge about the spectrum of natural scenes in order to perform the analysis. We will see that some knowledge about the non radial parts of the correlation function are needed as well, in order to match simulation results. It is possible to perform this sort of analysis for the Oja rule, since it's outcome depends only on the second-order statistics; and therefore, knowledge of the spectrum, and assumptions about the symmetries of the images are sufficient for deriving the receptive fields.

The principal components are the eigen-vectors of the covariance matrix, therefore the equation we try to solve is the eigenvalue problem, i.e., the eigen-equation, which has

the form

$$\sum_j Q_{ij} m_j = \lambda m_i, \tag{2.1}$$

where m_i are eigen-vectors, λ is the eigenvalue, and Q_{ij} is the covariance matrix which is defined as $Q_{ij} = \mathrm{E}[(d_i - \mathrm{E}[Idi])(d_j - \mathrm{E}[d_j])]$ where the average is taken over the input pattern distribution. Since we are dealing with a two-dimensional space, the index i really denotes a point in the plane, so it is more convenient to rewrite the covariance matrix in the form $Q(\mathbf{r}_i, \mathbf{r}'_j)$. Due to translational invariance, $Q(\mathbf{r}_i, \mathbf{r}'_j) = Q(\mathbf{r}_i - \mathbf{r}'_j)$. In the continuous limit, the summation becomes an integral over $\mathbf{r}'$, thus the eigen-equation becomes

$$\int_a Q(\mathbf{r} - \mathbf{r}')\psi(\mathbf{r}')d^2\mathbf{r}' = \lambda\psi(\mathbf{r}). \tag{2.2}$$

where $\psi(\mathbf{r})$ is the continuous limit of the eigen-vectors m_i, and a is the circular region in the input space in which synapses are allowed to form. It is important to note that the solutions to this equation 2.2, depends on the boundary condition imposed on them. Imposing a square boundary results in a different set of solutions from those we derive for a circular boundary.

The correlation function Q depends on the nature of the inputs, in section 2.2, Q will be the correlation function of the natural scenes, whereas in section 2.3, it will be denoted as Q^p and represents the correlation function of the preprocessed natural scenes.

2.1.1 The Method of Solving the Rotationally Symmetric problem

When the correlation function is radially symmetric, i.e when $Q(\mathbf{r} - \mathbf{r}') = Q(|\mathbf{r} - \mathbf{r}'|)$, all solutions to the eigen-equation 2.2 have the form, $\psi_l(\mathbf{r}) = f_l(r)\exp[il(\theta + \phi)]$.

We have chosen to represent the correlation function and receptive fields in the normalized Bessel-Fourier set $W_n(k_i^n\mathbf{r})$ which is zero at the boundary $r = a$ (Jackson, 1975, etc.). The receptive fields are, therefore, a superposition of these functions, i.e., $\psi(\mathbf{r}) = \sum_i A_{ni} W_n(k_i^n\mathbf{r})$.

When the correlation function is rotationally symmetric, then it is possible to represent the correlation function, in the Fourier Bessel basis, in the following simplified form.

$$Q(\mathbf{r} - \mathbf{r}') = \sum_{ljj'} [Q(l)]_{jj'} W_l(k_j^l\mathbf{r}/a)W_l^*(k_{j'}^l\mathbf{r}/a). \tag{2.3}$$

where W_l are the normalized Fourier Bessel functions as defined in appendix A, and k_j^l are defined so that the Bessel functions are zero at the boundary, i.e $W_l(k_j^l) = 0$. The important factor in this transformation is that we converted a function of $\mathbf{r} - \mathbf{r}'$ to a sum of functions of $\mathbf{r}$ times functions of $\mathbf{r}'$. This transformation is possible due to the addition theorem of Fourier-Bessel functions (Abramowitz and Stegun., 1972).

When this transformation is carried out, the integral equation 2.2 is transformed into the matrix equation

$$\sum_j [Q(l)]_{ij} A_{lj} = \lambda A_{li}. \tag{2.4}$$

In order to solve the rotationally symmetric problems, we will therefore have to find the values of $[Q(l)]_{ij}$ for the raw and preprocessed images, and solve the matrix equations for those values.

2.2 Principal Components of Natural Scenes

In this section we analyze the receptive fields when the inputs are small circular regions of the raw natural scenes. In figure 2.1 we have schematically displayed the architecture of this single cell model.

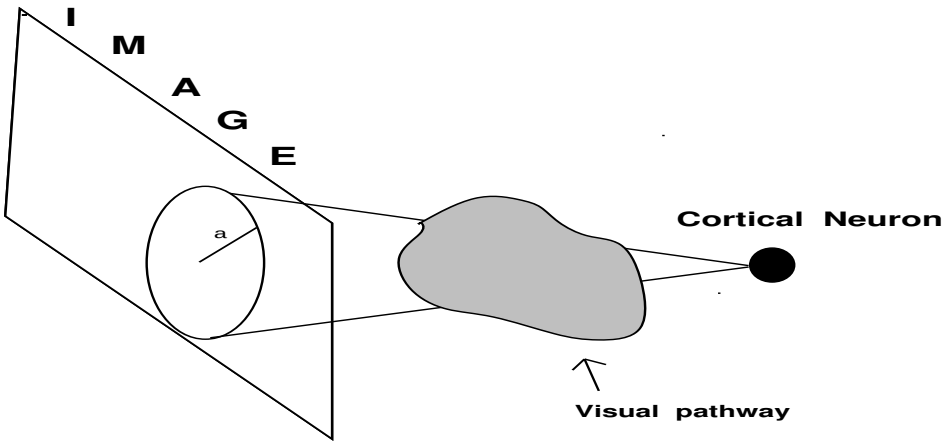

Figure 2.1: *Small circular regions in the input layer serve as inputs to a single neuron. This neuron uses the Oja (82) learning rule in order to form the receptive fields.*

The analysis performed is valid for the Oja rule as displayed in the image, but since we extract several of the major principal components, it is also valid for methods which extract several principal components simultaneously (Oja, 1982; Sanger, 1989; Oja, 1992, etc.).

2.2.1 The Rotationally Symmetric Solution

In this section we have to find the values of $[Q(l)]_{ij}$, when the inputs are the raw natural images. This will be done using the properties of the spectrum of natural images, as well as the assumption of radial symmetry.

As discused in the introduction, the power spectrum of natural images, which is equivalent to the Fourier transform of the correlation function, has the form $Q(\mathbf{k}) = c/\mathbf{k}^2$ where c is a constant (Field, 1987). Hereafter we shall set $c = 1$ for convenience, thus $Q(\mathbf{r})$ satisfies

$$\nabla^2 Q(\mathbf{r}) = -\delta(\mathbf{r}), \tag{2.5}$$

which can be readily proven by taking Fourier transformation of both sides of this equation. In principle, a gauge $g(\mathbf{r})$ which satisfies $\nabla^2 g(\mathbf{r}) = 0$, can be added to $Q(\mathbf{r})$. But the only such gauge, which is radially symmetric, is a constant (denoted as b). This constant will only affect the nature of the radially symmetric, or $l = 0$, solutions.

As is shown in detail in appendix A.1 these properties of the correlation function are sufficient for finding it's representation in this basis, which turns out to be

$$[Q(l)]_{jj'} = \sum_i \frac{\varepsilon_{lij}\varepsilon_{lij'}}{2\pi N_0(k_i^0)\ (k_i^0)^2}. + b\delta_{l0}V_jV_{j'} \tag{2.6}$$

where

$$\begin{aligned}
\varepsilon_{lnj} &= \int J_l(k_n^0 r)e^{il\theta}m_l(k_j^l\mathbf{r})d^2r \\
&= \begin{cases} \sqrt{2\pi N_0(k_i^0)}\delta_{jn} & \text{for } l = 0 \\ \sqrt{\frac{2\pi}{N_l(k_j^l)}}\left(\frac{k_j^l}{(k_n^0)^2-(k_j^l)^2}\right)J_l(k_n^0)J_{l+1}(k_j^l) & \text{for } l > 0 \end{cases}
\end{aligned}$$

and

$$V_j = \frac{2\sqrt{\pi}}{k_j^0}.$$

and N_m is the normalization constant, which takes the value $N_m(k) = (\frac{1}{2})[J_{m+1}(k)]^2$.

This matrix $[Q(l)]$ is infinite and in general non-diagonal. However, if $b = 0$, then for the case $l = 0$, it takes the particularly simple diagonal form, $[Q(0)]_{ij} = \delta_{ij}/(k_i^0)^2$, putting this into the matrix equation 2.4, we obtain the radially symmetric solutions to the eigen-equation. Hence in this case the solutions are simply the Bessel-Fourier functions $W_0(k_l^0\mathbf{r})$ with corresponding eigenvalues $\lambda_{0l} = 1/(k_l^0)^2$.

For $l \neq 0$, it is still possible to obtain a good approximation for the first eigen-vectors. Due to the diminishing nature of the eigenvalues of the Bessel equation, the infinite

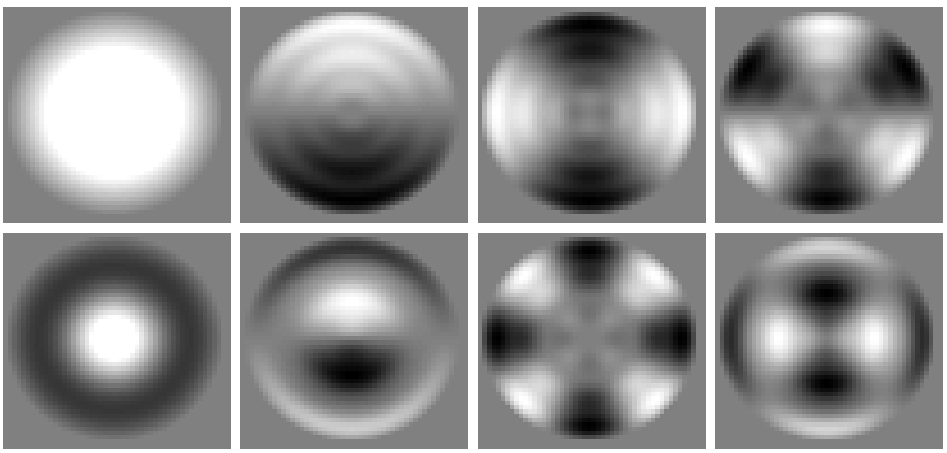

Figure 2.2: *The shapes of the first eight principal components ($b = 0$ $a = 1$ and $\phi_{mi} = 0$), ordered from left to right, top to bottom.*

summation involved in computing $Q(l)$ can be well approximated by a finite summation, furthermore, for small values of l, this matrix will typically have a near-diagonal form. Thus the effective number of Bessel functions of which each eigen-function is composed, will not be large at least for small l. We use this approximation and solve equation 2.4 numerically.

Some of these eigen-functions, with the arbitrary choice $b = 0$, are shown in figure 2.2, in which only the real part ($\propto cos(m\theta)$) has been displayed. In general, the complete set of solutions should include both parts, i.e., $\psi^1_m(\mathbf{r}k^m_i) \propto cos(m(\theta + \phi))$ and $\psi^2_m(\mathbf{r}k^m_i) \propto sin(m(\theta + \phi))$.

2.2.2 Retrieving the Phase

The solutions above, $\psi^1_{mi}(\mathbf{r}) = \psi^1_m(\mathbf{r}k^m_i) \propto cos(m(\theta + \phi))$ and $\psi^2_{mi}(\mathbf{r}) = \psi^2_m(\mathbf{r}k^m_i) \propto sin(m(\theta + \phi))$, not only have undetermined phases ϕ, but are degenerate as well, i.e., both have the same eigenvalues. In this section we l show that if the correlation function is not completely radially symmetric, this degeneracy will be broken, and that the value of the phase ϕ will be determined. Furthermore the preferred directions will be the horizontal and vertical, due to the longer correlation lengthes in these directions which exist in the natural environment. The assumption, that the correlation function is not radially symmetric, has been discussed in section 1.6. These results explain the feature of simulation results (Hancock et al., 1992), in which the phases seem to always take the value zero, and the ψ^1_{mi} solution has a different eigenvalue from the ψ^2_{mi} solution.

Hereafter we assume, for simplicity, that this perturbation term has, in $\mathbf{k}$ space, the

form

$$Q'(\mathbf{k}) = U(k)T(\theta_{\mathbf{k}}). \tag{2.7}$$

In order to calculate this perturbation, the representation of this perturbation in the two degenerate eigen-functions $\psi^1_{mi}(\mathbf{r})$ and $\psi^2_{mi}(\mathbf{r})$ has to be calculated. It is easier to perform this in $\mathbf{k}$ space in which the eigen-functions $\psi^1_{mi}(\mathbf{r})$ and $\psi^2_{mi}(\mathbf{r})$ are replaced by their Fourier transformations,

$$\begin{aligned} \psi^1_{mi}(\mathbf{k}) &= f_{mi}(k)cos(m(\theta_{\mathbf{k}} + \phi_{mi})) \\ \psi^2_{mi}(\mathbf{k}) &= f_{mi}(k)sin(m(\theta_{\mathbf{k}} + \phi_{mi})) \end{aligned} \tag{2.8}$$

in which $f_{mi}(k)$, is the transformation of the radial component of the solutions. If we denote

$$T(\theta_{\mathbf{k}}) = \sum_l t_l cos(l(\theta_{\mathbf{k}} - \alpha_l)) \tag{2.9}$$

which is the Fourier expansion of $T(\theta_{\mathbf{k}})$. The representation of the perturbation matrix with respect to the two degenerate eigen-functions has the form

$$\begin{aligned} (Q'_{(\mu,m,i|\gamma,m,i)})_{(\mu=1,2|\gamma=1,2)} &= (\int \psi^{\mu}_{mi}(\mathbf{k})^* Q'(\mathbf{k}) \psi^{\gamma}_{mi}(\mathbf{k}) d^2\mathbf{k})_{(\mu=1,2|\gamma=1,2)} \\ &= g_{mi} \begin{pmatrix} cos(\delta) & sin(\delta) \\ sin(\delta) & -cos(\delta) \end{pmatrix} \end{aligned} \tag{2.10}$$

in which $\delta = 2\phi_{mi} + 2m\alpha_{2m}$ and $g_{mi} = \frac{\pi}{2} t_{2m} \int U(k)|f_{mi}(k)|^2 k dk$. Since the two eigen-functions are degenerate, any linear combination of these two eigen-functions is an eigen-function of Q. Therefore, all we have to do is to find a linear combination of them which diagonalizes the perturbation matrix, i.e., to find the eigenvalues and eigen-vectors of the matrix in equation 2.10, which are

$$\begin{pmatrix} cos(\delta/2) \\ sin(\delta/2) \end{pmatrix} \qquad \begin{pmatrix} -sin(\delta/2) \\ cos(\delta/2) \end{pmatrix} \tag{2.11}$$

with eigenvalues g_{mi} and $-g_{mi}$, respectively.

Thus the eigen-functions and eigenvalues after the perturbation can be readily written as

$$\begin{aligned} \psi^+_{mi}(\mathbf{k}) &= f_{mi}(r)cos(m(\theta - \alpha_{2m})) \\ \psi^-_{mi}(\mathbf{k}) &= f_{mi}(r)sin(m(\theta - \alpha_{2m})) \end{aligned} \tag{2.12}$$

with eigenvalues $\lambda^+_{mi} = \lambda_{mi} + g_{mi}$, and $\lambda^-_{mi} = \lambda_{mi} - g_{mi}$, respectively, thus the degeneracy is broken. This result is in agreement with Hancock's simulations. These solutions have an important feature: their phases are determined by the properties of the real world correlation function. If the correlation function has a definite symmetry with an inclination angle α, then the solutions will have the same symmetry angle, since in this case $\alpha_{2m} = \alpha$ for all m. The spectra of the correlation function, shown in the introduction, indeed indicates a symmetry axis along $\alpha = 0$. Thus equation 2.12 predicts the zero phase result found in Hancock's simulation. When Hancock used images which were tilted by 45 degrees before being scanned, the preferred axis of the receptive fields was found to be 45 degrees. Again this is predicted by equation 2.12, because the symmetry axis of the correlation function spectrum also gets rotated by 45 degrees due to the rotated images, i.e., $\alpha = 45°$, and thus the solutions also get rotated by 45 degrees.

2.2.3 Discussion

We have derived analytically the forms of the localized principal components of natural images, based on results about the correlation function of natural scenes (Field, 1987). These receptive fields are shown in figure 2.2. We have also shown that a non-rotationally symmetric perturbation can break the degeneracy and give a definite phase which only depends on the properties of the real world correlation function.

When comparing the receptive fields predicted here, with the receptive fields of neurons in the visual cortex, two major differences stand out. The first is that some of the receptive fields predicted, such as the radially symmetric ones, do not exist in the visual cortex. This difference may be due, at least in part, to the absence of a retinal preprocessing from the model we have analyzed. Another difference, is that for each of the non-radially symmetric receptive fields predicted, only two orientations are necessary to form a complete set. This feature is a property of any method which extracts complete orthogonal sets, such as the Sanger (89) rule, and clearly contradicts the near-continuum of orientations found in the visual cortex.

2.3 Including Retinal Preprocessing

In the previous section we solved equation 2.2 for an environment composed of natural scenes, In this section we take into account the effects of the preprocessing which the natural images undergo in the retina and LGN. This preprocessing has a complex spatio-temporal form. We will neglect the temporal aspects of this preprocessing, and model the spatial preprocessing as a center surround filter of a finite extent. In figure 2.3 we schematically

display the architecture of this single cell model.

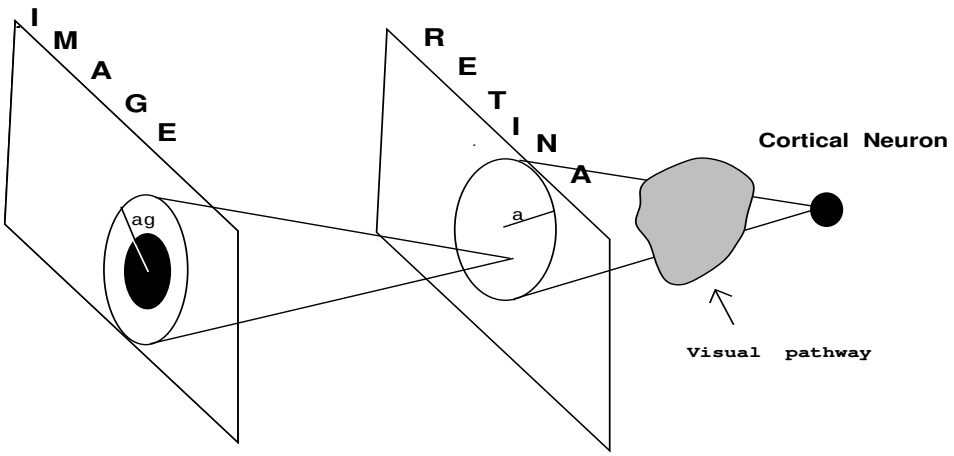

Figure 2.3: *The input images are preprocessed in the retina by a center surround filter of radius ag. Small circular patches of radius a are then chosen from the preprocessed images, as input vectors to the single neuron. This neuron uses the Oja (82) learning rule in order to form the receptive fields.*

We have chosen this filter to have a balance between the negative and positive parts. The filter $K(r)$ has a center surround shape, and is typically assumed to be a DOG (difference of Gaussians) function; for mathematical convenience have chosen a different center surround function,

$$K(\mathbf{r}) = \frac{1}{g^2} \begin{cases} J_0(q_2^0 r/a) + q J_0(q_1^0 r/a) & r < ag \\ 0 & r \geq ag \end{cases} \tag{2.13}$$

where $J_0(qr)$ are Bessel functions of order zero and $q_i^0 = g k_i^0$ are set so that the zeroes of the 0 order Bessel functions will be on the boundary with radius $r = ag$. The coefficient $q = 0.289$ so that $\int K(\mathbf{r}) d^2 r = 0$.

In general, any filter with a spatial extent smaller than the size of the receptive fields, can be represented in terms of the complete set of functions. Transforming to the set used for representing the correlation function we get that

$$K(\mathbf{r}) = \frac{1}{g^2} \sum_i \alpha_i J_0(k_i^0 r/a)$$

where $\alpha_i = \eta_{2i} + q\eta_{1i}$, and

$$\eta_{ij} = \frac{\int_0^g J_0(q_i^0 r/a) J_o(k_i^0 r/a) d^2 r}{\int_0^a J_0^2(k_i^0 r/a) d^2 r}$$
$$= \frac{g q_1^0}{N_0(k_j^0)[(q_i^0)^2 - (k_j^0)^2]} J_0(k_j^0 g/a) J_1(q_j^0 g/a).$$

In which k_j^0 is the j'th zero of the 0 order Bessel function, and $N_0(k_j^0) = (\frac{1}{2})[J_1(k_j^0/a)]^2$ is the normalization constant of $J_0(k_j^0 r/a)$.

It is important to notice that the number of terms sufficient for a good approximation depends the value of g. As can be seen in figure 2.4.

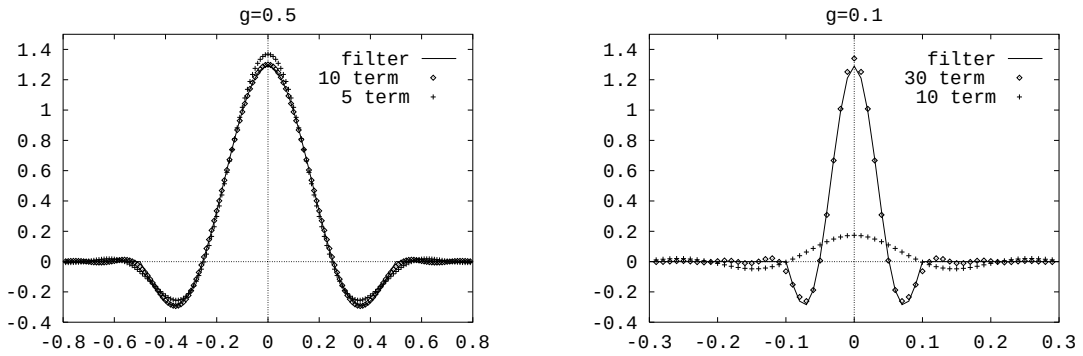

Figure 2.4: *Approximations of two different filter sizes, on the left $g = 0.5$ (left) and on the right $g = .1$ (right). When $g = 0.5$ ten terms are sufficiently good approximation, for $g = 0.1$ an equivalent approximation requires thirty terms.*

2.3.1 The Radially Symmetric Solution

Mathematically the aim of this paper is to solve equation 2.2. In the radially symmetric case we know that the eigenstates have the form $\psi(\mathbf{r}) \propto e^{il\theta} f(r)$. Solving the eigen-equation will be done as in the previous section, by representing both the receptive fields, and the correlation function in terms of an orthonormal Fourier-Bessel Basis W_{lj} as defined in Appendix A. Thus the eigen-states take the form $\psi(\mathbf{r}) = \sum_j A_{lj} W_{lj}(k_j^l \mathbf{r}/a)$. And the eigen-equation is transformed into the matrix equation

$$\sum_j [Q^p(l)]_{ij} A_{lj} = \lambda A_{li}. \tag{2.14}$$

In the previous section we solved this type of equation for an environment composed of natural scenes, appendix A.1 outlines how to obtain the representation for the correlation function in such a case. In this case Q^p is the correlation function of the preprocessed images, i.e of the images convolved with the retinal preprocessing filter.

When the appropriate approximation to the filter is convolved with the images, as detailed in appendix A.2, the resulting correlation function in this representation, takes the form,

$$[Q^p(l)]_{nm} = \sum_{jj'ip} [Q(l)]_{jj'} \alpha_i \alpha_p \varepsilon_{lij} \varepsilon_{lpj'} \varepsilon_{lin} \varepsilon_{lpm}. \tag{2.15}$$

Where $[Q(l)]_{jj'}$ the the components of the correlation function of natural scenes, and ε_{lpm} are defined in Appendix A.1.

For the case $l = 0$ this has the form,

$$[Q^p(0)]_{mn} = \frac{\pi^2}{g^4} \left[J_1^4(k_m^0) \alpha_m^2 \frac{1}{(k_m^0)^2} \delta_{mn} + b \frac{J_1^2(k_m^0) J_1^2(k_n^0)}{k_m^0 k_n^0} \alpha_m \alpha_n \right].$$

Only the radially symmetric matrix depends on b. which dependent on the size of the receptive field. The image to neuron receptive field, can be obtained, by replacing A_{lj} the components of the eigenvectors of this matrix, with $B_{ln} = \sum_{ij} A_{lj} \alpha_i \varepsilon_{lij} \varepsilon_{lin}$. Thus $\tilde{\psi}_l(\mathbf{r}) = \sum_n f_{ln} m_l(k_n^l \mathbf{r}/a)$.

2.3.2 Results and comparison to simulations

In this section the results of the theory outlined in the first section are displayed, and compared to simulation results obtained under similar conditions.

The radially symmetric solutions depend on the constant b. The value of b depends also on the measurements performed in the first chapter, as shown in figure 1.11, as well as on the size of the receptive field. We have chosen the size $a = 10$ for our receptive fields.

The theoretical analysis in the previous section assumes a radially symmetric environment, this assumption is only an approximation. The non radially symmetric correlations can be taken into account, in order to give a more exact and complete theory, however in this section we try to test the radially symmetric theory, and therefore we have forced a larger degree of radial symmetry on our images by rotating each of our images by 90 degrees and including the rotated images in the data set.

In Figure 2.5 we have displayed the theoretical prediction for the eigenvalues and eigenvectors as a function of g which is the ratio between the receptive field size, and the size of the center surround preprocessing filter. Changing the receptive field size would affect only the radially symmetric solutions.

In order to test the validity of the theory we have compared the theoretical predictions at $g = 1$ shown in figure 2.6 with simulation results, for the first five principal components shown in figure 2.7. In these simulations we used the symmetrized data, and the algorithm proposed by Sanger (89) in order to extract several principal components.

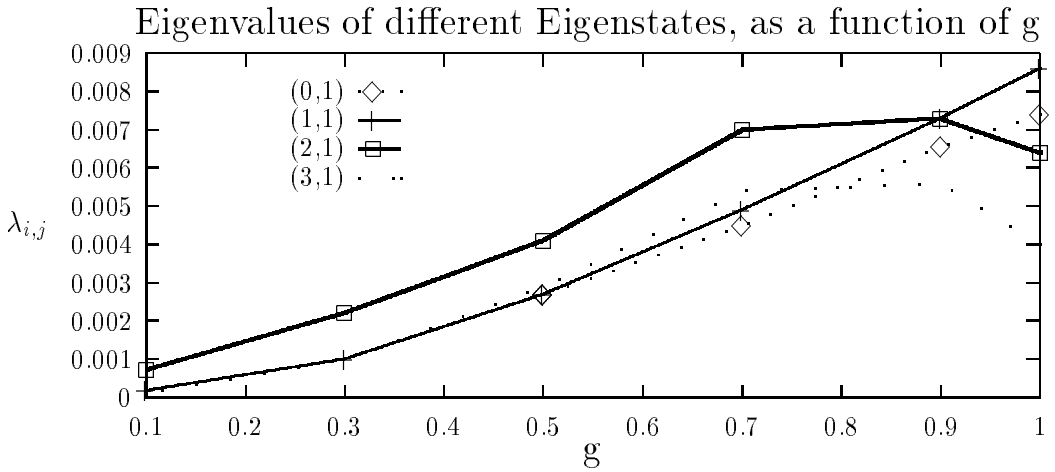

Figure 2.5: *The change in eigen-values as a function of g, for receptive fields with radius a=10. The notation* (m,n) *denotes the n'th solution with angular component m, eg:* $(0,1)$ *denotes the radially symmetric solution with highest eigenvalue. It can be seen the near* $g=1$ *the principal component is proportional to* $\cos(\theta)$ *whereas near* $g=0.7$ *the dominant solution becomes the one proportional to* $\cos(2\theta)$*. For small values of g i.e:* $g \leq .3$ *the results displayed here are not exact since larger matrices should be used in order to achieve a good approximation.*

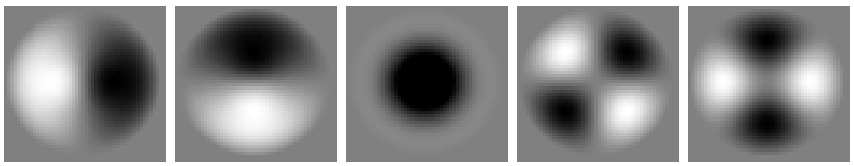

Figure 2.6: The first 5 receptive fields, from LGN to cortex with $g=1$ according to the theory, with $a=10$.

It should also be noted that theoretically, for a radially symmetric environment, the two solutions identical apart from a rotation, have the same eigenvalue. We have displayed here the LGN to cortex receptive fields which which have a different radial component then the retina to cortex receptive fields, but an identical angular form.

2.3.3 An angular correction

In order to achieve an agreement between the theory and simulations, we artificially symmetrized the correlation function of natural scenes by rotating each image by 90 degrees, and including it in the data base. In figure 2.8 we have contrasted the first principal component as extracted by simulation from a non symmetrized data set with the Principal

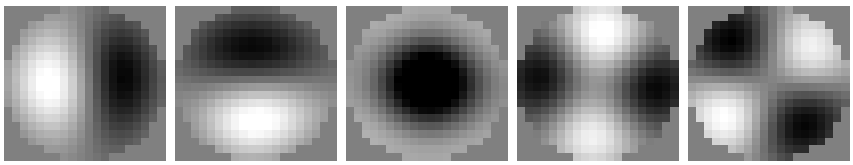

Figure 2.7: *Simulations results. The inputs are* 12 *natural scenes which have been preprocessed with a center surround filter of radius* 10*, and in addition each image has been rotated by 90 degrees in order to increase the radial symmetry. The receptive filed size used had a radius* $a = 10$ *thus* $g = 1$*, and the conditions are identical to those in the theoretical derivation above. There is a considerable similarity between the theory and the simulations.*

component as extracted from the symmetrized inputs, at the value $g = 0.7$.

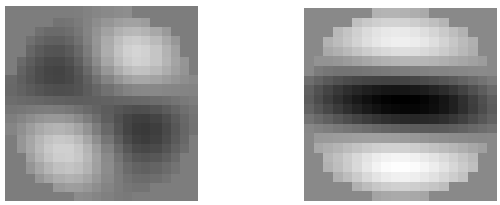

Figure 2.8: *Comparison between simulations with symmetrized Data set (left) and non symmetrized (right). The results of the symmetrized simulations, match the theoretical prediction.*

By inspecting the form of the receptive field on the right hand side of figure 2.8 we conclude, that the non symmetric portion of the correlation function of the natural scenes is responsible for the difference between the theoretical predictions and the simulation results using non rotated natural scenes.

In section 1.6 we analyzed the non symmetric component of spectrum of natural scenes. Since the non symmetric parts of the correlation function are relatively small compared to the symmetric part, we can use perturbation theory in order to correct the predictions of the symmetric theory. The solutions of the symmetric theory will be denoted by ψ^0_{li} where the l represents the angular mode, and i is the solution with the i'th eigenvalue in that mode. When performing first order perturbation theory in Fourier space, we find that the first order correction ψ^1_{li} takes the form,

$$\psi^1_{li}(\mathbf{k}) = \psi^0_{li}(\mathbf{k}) + \epsilon \sum_{i \neq j,m} \mathrm{a(l,i;m,j)}\psi^0_{mj}(\mathbf{k}) \tag{2.16}$$

where

$$\mathrm{a(l,i;m,j)} = \sum_n \delta(\mathrm{l+n,m}) \frac{\int d^2k Q^p_n(k) e^{-in\theta} \psi^0_{li}(k) \psi^0_{mj}(k)}{\lambda^0_{li} - \lambda^0_{mj}}, \tag{2.17}$$

where Q_n^p are the angular components of the correlation function of the preprocessed natural images. In Fourier space the correlation function takes the form $\tilde{Q}^p(k) = Q(k)K^2(k)$, where $K(k)$ is the Fourier transform of the retinal filter, and $Q(k)$ is the Fourier transform of the correlation function of natural scenes.

The major qualitative difference which results from this perturbation is therefore that the receptive fields become combinations of different angular states. For the case of $g = 0.7$ the major contribution comes from the component

$$\mathrm{a}(2,1;0,1) = \frac{\int d^2k D_2(k)K^2(k)cos(2\theta)\psi^0_{2,1}(k)\psi^0_{0,1}(k)}{\lambda^0_{2,1} - \lambda^0_{0,1}}. \tag{2.18}$$

In order to perform this calculation we need to have a measurement of or $D_2(k)$. We measured this component of the power spectrum, as shown in figure 1.12, but as can be seen it is a crude measurement. We used this measurement in order to approximate this integral, in this approximation we only considered contributions from the lowest frequencies, thus it is an estimated lower bound. The coupling constant we obtained is $a(2,1;0,1) = 0.47$. In figure 2.9 we have displayed the receptive field for this coupling constant, as well as for $a(2,1;0,1) = 1$ which gives a more reasonable result, and is a sensible number due to the nature of our approximation.

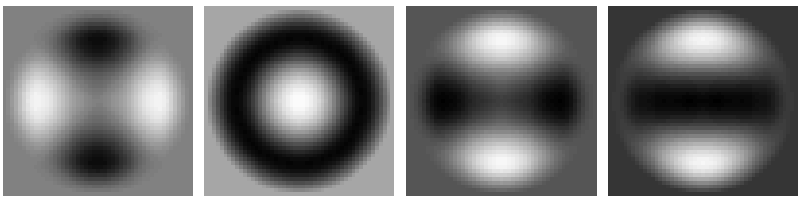

Figure 2.9: *The two images on the left are the first and fourth eigenfunctions of the radially symmetric theory at $g = .7$. The third from left is the result of mixing them with the calculated coupling constant .47. On the right is the result when the constant is assumed to be* 1.

As can be seen even before carrying out the exact calculations, this approach may indeed explain exactly the receptive field structures obtained from unsymmetrized natural images, using the Oja (82) rule.

2.3.4 Discussion

We examined the form of the receptive field structures resulting from a class of unsupervised learning rules. We assumed that the environment to which these neurons are exposed is composed of natural scenes, which are preprocessed by a center-surround filter.

At first we analyzed the radially symmetric environment, and have shown that the receptive field structure is critically dependent on the ratio between the size of the

preprocessing filter, and the size of the receptive field. We also investigated the nature of the non radially symmetric components of the correlation function, and have demonstrated how perturbation theory can be used to explain the effect of the non symmetric components on the receptive fields of neurons in non radially symmetric environments.

In the previous section, we found that the principal components of raw natural scenes are quite different from receptive fields in the visual cortex. In this section we have examined the effect of retinal preprocessing and have found that, for certain sizes of receptive fields and retinal filters, the resulting receptive field is more similar to cortical receptive fields. This similarity we have shown, results from the nonradially symmetric portion of the correlation function. Receptive fields obtained with this environment have a horizontal oreiented receptive fields, due to the non rotationally symmetric components of the correlation function, which originate from the predominance of horizontal components such as horizons, and vertical components such as trees.

One difference which still stands out between this single cell theory and receptive fields in the cortex, is that the cortex neurons with all orientations exist whereas the single cell theory predicts only a single orientation.

This stands in contrast to what happens in the BCM theory, in which already the single cell model exhibits many different orientations, due to the many local minima of the BCM cost function. This difference could already be seen in the simple 2D example discussed in the introduction, and displayed in figure 1.9.

Chapter 3

Synaptic Plasticity and the structure of the Visual Cortex

3.1 Introduction

In the previous chapter we analyzed receptive fields for a single cortical neuron, which does not interact with other cortical neurons. In the cortex, however, many of the synapses onto a cortical cell do not come from sensory projections, but rather from other cortical neurons, which makes it clear that the interaction term is very significant. Furthermore, in order to understand the regular relatively continuous organization of different receptive fields across the cortex, networks of interacting neurons must be assumed.

In this chapter we analyze a simple cortical network of learning neurons, each of the neurons in this network forms plastic bonds with the sensory projections, and fixed bonds with other cortical neurons. The learning rule assumed herein will typically be of a Hebbian type, and we will, as in the previous chapter, put an emphasis on the Oja rule. The type of network architecture assumed here is similar to that used in many other synaptic models of cortical plasticity (von der Malsburg, 1973; Nass and Cooper, 1975; Linsker, 1986; Miller et al., 1989; Miller, 1994, etc.). Rather then presenting a new model we have attempted, in this chapter, to analyze the behavior of these existing models, and map them onto much simpler models which we might be able to understand better[1]. When this analysis is be performed, it becomes clear that in order to understand the behavior of such networks, it is first necessary to understand how single neurons using the same learning rule behave.

The purpose of this chapter is to start giving answers to two related questions about such networks.

[1] This analysis is similar to the one performed by (Rubinstien, 1994) although it was derived separately

(a) How do the cortical interactions affect the receptive fields of single cortical neurons, and in particular can cortical interactions produce oriented receptive fields, where the single cell model produces radially symmetric receptive fields.

(b) How do the cortical interactions affect the organization of different receptive fields across the cortex.

Apart form the synaptic network models analyzed herein, there have been phenomenological models proposed in order to explain the organization of receptive fields across the cortex.

The experience of the last couple of decades in physics of complex systems has shown that phenomenological models are extremely important in explaining properties of complex systems. For complex systems of many interacting particles it is typically hopeless to try and solve the exact microscopic theory of these interacting systems, however in many cases coarse grained phenomenological models, which fail to take into account most of the microscopic detail, do manage to predict the properties of such systems. It seems that if this is the case for the relatively simple complex systems of the physical sciences, then it must also be the case for the much more complex systems of neurobiology.

In this chapter we show how some of the detailed synaptic theories (microscopic) are related to some of the proposed phenomenological theories. This give further justification to probing the properties of the network, with the much simpler phenomenological analogue, as we do in this chapter.

It has been claimed that some of the properties of some phenomenological models differ to a large degree from those of the synaptic models (Erwin et al., 1994). In particular when the models of Miller (94) and Swindale (82) are compared, it is found that the Swindale model fits the experimental results much better. In this chapter we will see that these families of models are actually related, and that some of the phenomenological models (Cowan and Friedman, 1991; Swindale, 1982) are under certain conditions approximations of the full synaptic models. We hope that the analytic tools provided herein can help in understanding what the origin of such differences is.

3.2 A self organizing network model

In this section we examine a network of interconnected neurons, which are connected with modifiable synapses to the input layer. The learning rule examined is a stabilized Hebbian rule which is sensitive only to the second order statistics.

The output $c(\mathbf{r})$, of each neuron, in this model is defined as

$$c(\mathbf{r}) = \sum_{x,\mu} I(\mathbf{r}-\mathbf{x})A(\mu-\beta(\mathbf{x}))d(\mu)m(\mathbf{x},\mu).$$

where β is the center of the symmetric arbor function A, and is a function of $\mathbf{x}$. I is the cortical interaction function, m are the weights, d the input vectors, μ denotes points in input space, $\mathbf{r}$ and $\mathbf{x}$ in output space.

The energy function H, we will try to minimize is now simply

$$\begin{aligned} H &= -\sum_{r} c^2(\mathbf{r}) - \gamma(\mathbf{m}) \qquad (3.1)\\ &= -\sum_{r,x,x'} I(\mathbf{r}-\mathbf{x})I(\mathbf{r}-\mathbf{x}') \\ &\quad \sum_{\mu,\mu'} A(\mu-\beta(\mathbf{x}))A(\mu'-\beta(\mathbf{x}'))m(\mathbf{x},\mu)m(\mathbf{x}',\mu')d(\mu)d(\mu') - \gamma(\mathbf{m}) \end{aligned}$$

where γ is an abstract constraint on the allowed values of $\mathbf{m}$.

We now perform the standard procedure, of turning this into a correlational model, by taking the ensemble average over the input environment, and also taking the sum over $\mathbf{r}$. Thus

$$\begin{aligned} H &= -\sum_{x,x'} \tilde{I}(\mathbf{x}-\mathbf{x}') \qquad (3.2)\\ &\quad \sum_{\mu,\mu'} A(\mu-\beta(\mathbf{x}))A(\mu'-\beta(\mathbf{x}'))m(\mathbf{x},\mu)m(\mathbf{x}',\mu')Q(\mu-\mu') - \gamma(\mathbf{m}). \end{aligned}$$

Where $\tilde{I}(\mathbf{x}-\mathbf{x}') = \sum_r I(\mathbf{r}-\mathbf{x})I(\mathbf{r}-\mathbf{x}')$ is the new effective interaction term, and $Q(\mu-\mu') =< d(\mu)d(\mu') >$ is the correlation function.

From now on, in order to simplify matters, we assume a step arbor function of the form,

$$A(\mathbf{x}) = \begin{cases} 1 & \text{for } |\mathbf{x}| \le \mu_{max} \\ 0 & \text{for } |\mathbf{x}| > \mu_{max} \end{cases} \qquad (3.3)$$

Using this function is justified since it can produce oriented receptive fields in both network models (Miller, 1994) and single cell models (Liu, 1994).

The gradient descent dynamics implied by this Hamiltonian are now

$$\frac{dm(\mathbf{x},\mu)}{dt} = \begin{cases} \sum_{\mathbf{x}'} \sum_{|\mu'-\beta(x')|<\mu_{max}} \tilde{I}(\mathbf{x}-\mathbf{x}')m(\mathbf{x}',\mu')Q(\mu-\mu') & \text{for } |\mathbf{x}-\beta(\mathbf{x})| < \mu_{max} \\ 0 & \text{for } |\mathbf{x}-\beta(\mathbf{x})| > \mu_{max} \end{cases} \qquad (3.4)$$

This equation is the same to Millers (1994) synaptic plasticity equations for the step arbor function.

The method we use now, is to represent the weights m in terms of the complete set, i.e $m(\mu, \mathbf{x}) = \sum_{ln} a_{ln}(\mathbf{x}) m_{ln}(\mu, \mathbf{x})$. We have chosen this complete set to be a solution of the non interacting eigen-value problem.

$$\sum_{|\mu'-\alpha|} m_{ln}(\mu', \mathbf{x}) Q(|\mu - \mu'|) = \lambda_{ln} m_{ln}(\mu - \alpha, \mathbf{x}), \tag{3.5}$$

In a radially symmetric environment the eigenfunctions will take the form

$$m_{ln}(\mu, \mathbf{x}) = m_{ln}(\mu, \phi_{ln}(\mathbf{x})) = e^{il\phi_{ln}(\mathbf{x})} U_{ln}(\mu)$$

where $\phi_{ln}(\mathbf{x})$ is an arbitrary phase which is undetermined, due to the radially symmetric nature of the correlation function.

Using this expansion, and taking sum over μ the Hamiltonian takes the form,

$$\begin{aligned} H &= -\sum_{x,x'} \sum_{lnl'n'} \tilde{I}(\mathbf{x} - \mathbf{x}') \lambda_{ln} a_{ln}(\mathbf{x}) a_{l'n'}(\mathbf{x}') \\ &\qquad O\left[l, n, l', n', \phi_{ln}(\mathbf{x}), \phi_{ln}(\mathbf{x}'), \beta(\mathbf{x}), \beta(\mathbf{x}')\right] - \gamma(a) \end{aligned} \tag{3.6}$$

Where O the overlap between two receptive fields, is

$$\begin{aligned} & O\left[l, n, l', n', \phi_{ln}(\mathbf{x}), \phi_{ln}(\mathbf{x}'), \beta(\mathbf{x}), \beta(\mathbf{x}')\right] \\ = & \sum_{\mu'} m_{ln}\left(\mu' - \beta(\mathbf{x}), \phi_{ln}(\mathbf{x})\right) m_{l'n'}\left(\mu' - \beta(\mathbf{x}'), \phi_{l'n'}(\mathbf{x}')\right). \end{aligned} \tag{3.7}$$

It is important to notice that in general, O depends on the distance between the two arbor functions, as well as on the shapes and phases of the two eigen-states.

3.3 Analysis of the Common Input model:

We will now make the simplification that $\beta = \alpha = 0$ i.e all neurons have an arbor function with the same center point, this simplification makes analysis simpler, and is a good approximation when the number of output neurons is much greater than the dimensionality of the input.

It is important to notice though that the conclusions which we deduce from this Common input model, may not be applicable to more complex models. In particular it is important to notice that in other models, O may change signs as a function of $|\mathbf{x} - \mathbf{x}'|$ and thus turn an excitatory model, i.e: one in which $I \geq 0$ to one in which there is effectively an inhibitory portion to the interaction.

For this choice of arbor function the overlap O takes a very simple form,

$$O\left[l,n,l',n',\phi_{ln}(\mathbf{x}),\phi_{ln}(\mathbf{x}')\right] = \delta_{ll'}\delta_{nn'}\exp^{il[\phi_{ln}(\mathbf{x})-\phi_{ln}(\mathbf{x}')]}.$$

Thus for the common input model the Hamiltonian takes the form

$$H = -\sum_{xx'lm}\tilde{I}(\mathbf{x}-\mathbf{x}')\lambda_{ln}a_{ln}(\mathbf{x})a_{ln}(\mathbf{x}')e^{il[\phi_{ln}(\mathbf{x})-\phi_{ln}(\mathbf{x}')]} - \gamma(a). \tag{3.8}$$

When observing equation: 3.8 it is easy to see that the eigenvalues of the correlation function, play a major role in the resulting Hamiltonian. They could serve as an importance parameter by which approximations can be made. Hence the approximation proposed here is to neglect all terms except those multiplying the largest eigenvalues, these approximations will be termed herein toy models.

In order to make things concrete, we assume that $\lambda_{01} = \lambda_0$, and $\lambda_{11} = \lambda_1$ are the largest eigenvalues and considerably larger then the other eigen-values. This is not an artificial choice, since in many circumstances this indeed seems to be the case (MacKay and Miller, 1990; Liu and Shouval, 1994; Liu et al., 1994), although in other regimes other eigen-states will dominate.

Under this approximation the Hamiltonian will take the form.

$$H = -\sum_{\mathbf{r}\mathbf{r}'}\tilde{I}(\mathbf{r}-\mathbf{r}')\left\{\lambda_0 a(\mathbf{r})a(\ r') + \lambda_1 b(\mathbf{r})b(\mathbf{r}')cos[\phi(\mathbf{r})-\phi(\mathbf{r}')]\right\} - \gamma(a,b)$$

Different types of constraints can be considered. Given the constraint that the weights are square normalized, i.e that $\sum_\mu m(\mu,\mathbf{x})^2 = 1$,[2] which for the approximate model takes the form $a^2(\mathbf{r}) + b^2(\mathbf{r}) = 1$, the model is equivalent to the non isotropic Heisenberg spin model embedded in two dimensions. This can be made more evident when this is rewritten in two equivalent forms. If we set $a(\mathbf{r}) = S_z(\mathbf{r}) = cos(\theta(\mathbf{r}))$, $b(\mathbf{r}) = sin(\theta(\mathbf{r}))$ and $S_x(\mathbf{r}) = b(\mathbf{r})cos(\phi(\mathbf{r}))$ and $S_y(\mathbf{r}) = b(\mathbf{r})sin(\phi(\mathbf{r}))$. then the Hamiltonian takes the form,

$$H = -\sum_{\mathbf{r}\mathbf{r}'}\tilde{I}(\mathbf{r}-\mathbf{r}')\left\{\lambda_0 S_z(\mathbf{r})S_z(\mathbf{r}') + \lambda_1[S_x(\mathbf{r})S_x(\mathbf{r}') + S_y(\mathbf{r})S_y(\mathbf{r}')]\right\} \tag{3.9}$$

Subject to the constraint that $S_x^2 + S_y^2 + S_z^2 = 1$.

This can also be put in the closed form,

$$\begin{aligned} H &= \sum_{\mathbf{r},\mathbf{r}'} I(\mathbf{r}-\mathbf{r}')\{\lambda_0\cos(\theta(\mathbf{r}))\cos(\theta((\mathbf{r}')) \\ &+ \lambda_1\sin(\theta(\mathbf{r}))\sin(\theta(\mathbf{r}'))[\cos(\phi(\mathbf{r}))\cos(\phi(\mathbf{r}')) + \sin(\phi(\mathbf{r}))\sin(\phi(\mathbf{r}'))]\}. \end{aligned} \tag{3.10}$$

[2]This is the constraint to which the Oja (82) rule converges

Which is exactly the nonisotropic Heisenberg model of Statistical Mechanics.

Thus it can be seen that each neuron is associated with only two dynamical variables ϕ and θ, thus this approximation has resulted in a major simplification in the complexity of the model.

If $\lambda_1 >> \lambda_0$ and we can discard of the terms multiplying the λ_0 this becomes an XY model and is equivalent to the phenomenological model proposed by Cowan and Friedmann (Cowan and Friedman, 1991). Thus in this case the Hamiltonian takes the simple form

$$H = -\sum_{\mathbf{r}\mathbf{r}'} I(\mathbf{r}-\mathbf{r}')\cos\left(\phi(\mathbf{r})-\phi(\mathbf{r}')\right), \tag{3.11}$$

and the dynamics are

$$\dot{\phi}(\mathbf{r}) = \eta\sum_{\mathbf{r}'} I(\mathbf{r}-\mathbf{r}')\sin\left(\phi(\mathbf{r})-\phi(\mathbf{r}')\right). \tag{3.12}$$

Another, well known, phenomenological models is the Siwindale model (82) for the development of orientation selectivity, we will show that it is similar to the Cowan Friedmann model. Swindale defines the complex variable $Z(\mathbf{r}) = S_x(\mathbf{r}) + iS_y(\mathbf{r})$ or $Z(r) = b(\mathbf{r})\exp(i\phi(\mathbf{r}))$ [3].

For which he defines the dynamics

$$\dot{Z}(\mathbf{r}) = \sum_{\mathbf{r}'} I(\mathbf{r}-\mathbf{r}')Z(\mathbf{r}')(1-|Z(\mathbf{r})|). \tag{3.13}$$

Thus for the variables $\phi(\mathbf{r}) = tg^{-1}(S_x/S_y)$ and $b(\mathbf{r}) = |Z(\mathbf{r})|$, the dynamics are

$$\dot{\phi}(\mathbf{r}) = \eta(1-b(r))\sum_{\mathbf{r}'} I(\mathbf{r}-\mathbf{r}')\sin\left(\phi(\mathbf{r})-\phi(\mathbf{r}')\right) \tag{3.14}$$

and

$$\dot{b}(\mathbf{r}) = \eta\left(1-b(\mathbf{r})\right)b(\mathbf{r})\sum_{\mathbf{r}'} I(\mathbf{r}-\mathbf{r}')\cos\left(\phi(\mathbf{r})-\phi(\mathbf{r}')\right) \tag{3.15}$$

Comparing equations 3.12 and 3.14 we see that for $b(\mathbf{r}) < 1$ the dynamics of the angle ϕ in the Swindale model are parallel to those in the Cowan model. However, when $b(\mathbf{r}) = 1$ the Swindale dynamics terminate, and when $b(\mathbf{r}) > 1$ they are anti parallel to the Cowan and Friedmann dynamics. The dynamics for $b(\mathbf{r})$ when $b(\mathbf{r}) < 1$ are such that if $h(\mathbf{r}) = \sum_{\mathbf{r}'} I(\mathbf{r}-\mathbf{r}')\sin(\phi(\mathbf{r})-\phi(\mathbf{r}')) > 0$ then $b(\mathbf{r}) \to 1$, this would be the common case since the dynamics for $\phi(\mathbf{r})$ try to maximize $h(\mathbf{r})$ however in some cases, such as near orientation singularities this may not be possible and in these cases $b(\mathbf{r}) \to 0$.

[3] This angle is different than the one Swindale chooses for the orientation angle, He defines the orientation angle ϕ_s as $\phi_s = \frac{1}{2}\phi$.

Thus it can be expected that if the Swindale simulations are started at small values of $b(\mathbf{r})$ and the learning rate is slow enough, then the fixed points of the orientation maps would be similar to the Cowan and Friedmann fixed points.

In synaptic space the constraint imposed in the Swindale model is equivalent to stopping the update of each neuron once it's synaptic weight becomes normalized, this seems a non local constraint, although it could possibly be made local by a procedure similar to the one Oja has used to attain square normalization. The final state of a Swindale type model can depend on the initial conditions, and on the learning rate η, thus the Swindale model has extra (spurious) fixed points, which the Cowan and Friedmann model does not possess.

Another possible choice for the normalization, is to set a+b=1 , then the approximate Hamiltonian takes the form

$$H = -\sum_{\mathbf{r}\mathbf{r}'} I(\mathbf{r}-\mathbf{r}')\left\{\lambda_0 a(\mathbf{r})a(\ r') + \lambda_1[1-a(\mathbf{r})][1-a(\mathbf{r}')]cos[\phi(\mathbf{r})-\phi(\mathbf{r}')]\right\}$$

It is important to note that this Hamiltonian is not bounded, as may be the case when the van der Malsburg (73) type of normalization is used. This choice of normalization is therefore not valid for our purpose.

The questions which we would like to answer using these toy models are

a) Can the Lateral interaction break the symmetry, i.e if $\lambda_0 > \lambda_1$ can the network settle in a state where most of the neurons are oriented nevertheless.
b) What would be the form of the ordering, if the network is in an oriented state. We will analyze these questions numerically for the square normalized case, In the next section.

3.3.1 The Excitatory case $I \geq 0$

In the Excitatory case, we know that the global minimum of the common input energy function (eq:3.8) as well as in the approximation (eq:3.10), occurs when all receptive fields are equivalent to the principal component, and have the same orientation.

Since we know what the global minim is, it is interesting to see how simulations of the simplified model manage to reach this minimum. Doing this is a test for our procedures, but it also has interesting implications on simulations of detailed synaptic models of the Miller, Linsker type.

In Figures:3.1 and 3.2 we show the time evolution of the excitatory toy model, when $\lambda_1/\lambda_0 = 1.3$, i.e: the oriented solution has a higher eigenvalue. In these simulations we used a Gaussian interaction function, with a standard deviation $\sigma = 2.5$. We have defined

one iteration in this network as the number of random updates, required for updating each neuron on average once.

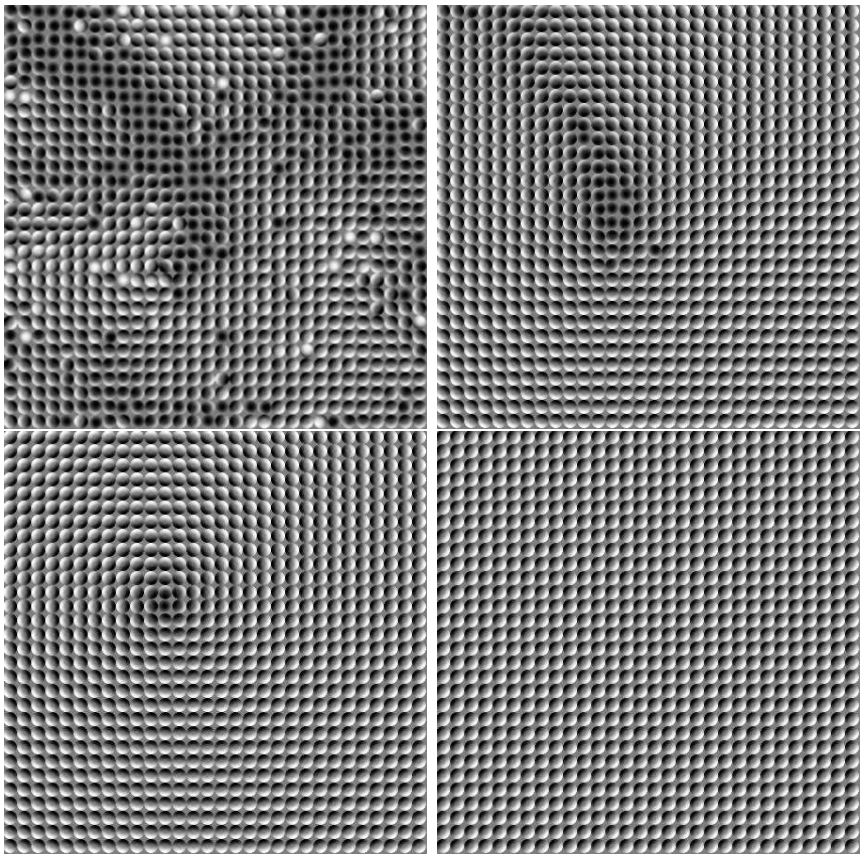

Figure 3.1: *Time evolution of receptive fields in an excitatory network, in each iteration each neuron is updated on average one time. The images show the receptive fields in a 30 by 30 network of neurons, which were run for 5,30,100 and 600 iterations, displayed from left to right, top to bottom. The learning rate η was reduced during learning from .1 to .02, in order to speed up the convergence.*

We can see that the global minimum is indeed reached, however it takes several hundred iterations to reach the global minimum for the parameters we have used, which we have tried to optimize. The neurons become selective quickly, most of them after 30 iterations, however it takes a very long time in comparison for the neurons to become parallel.

Simulations of the detailed synaptic models which are much more complex, were typically run with a smaller number of iteration (Miller, 1994). In the Miller model the learning rule used was unstable, therefore saturation limits were set on the synapses. When synapses reached these limits they were frozen, these type of bounds we term sticky bounds. When a high fraction of the synapses become saturated, the simulations are terminated.

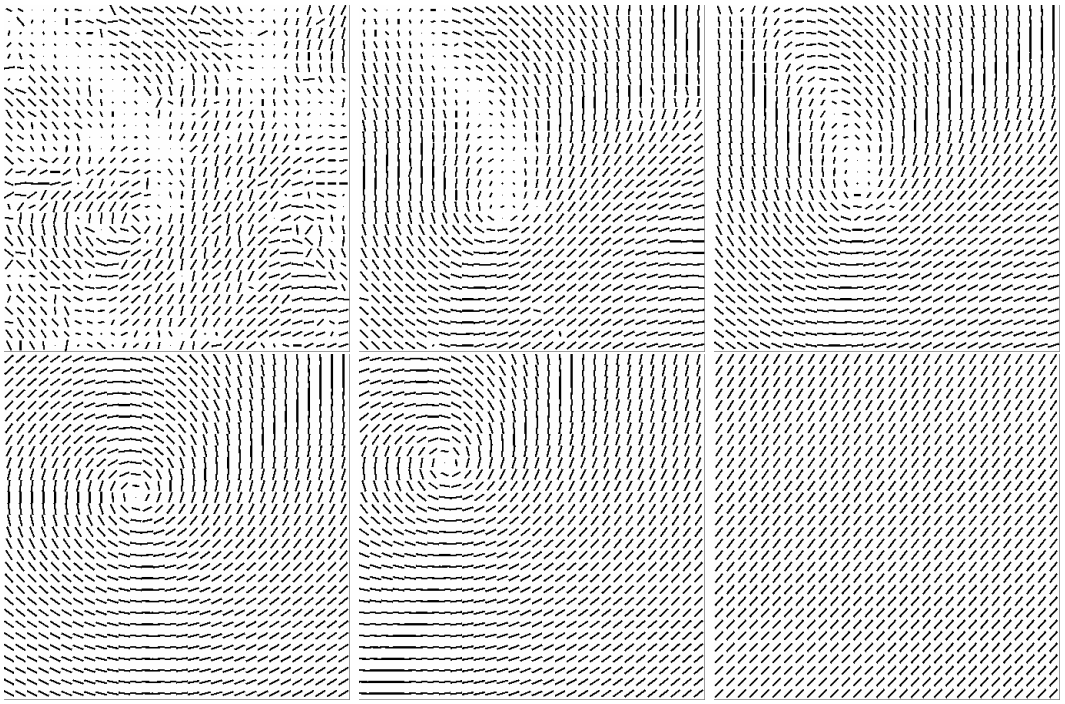

Figure 3.2: *Time evolution of orientation in an excitatory network. These simulations are the same as in the previous figure but here we have only displayed the optimal orientation and added some intermediate time steps. The images show the orientations in a 30 by 30 network of neurons, which were run for 5,15,30,100,400 and 600 iterations, displayed from left to right, top to bottom. The direction of the line codes orientation, and it's intensity codes for the degree of orientation selectivity. A great majority of the cells have already attained a high degree of selectivity, at 100 iterations, but it takes several hundred more iterations for them to become completely parallel.*

Since in our simulations neurons attained a high degree of orientation selectivity long before they became parallel, it implies that if the sticky bounds were not used by Miller (Miller, 1994) and if simulations were run for more iterations, then the resulting orientation maps may be quite different. Thus sticky boundaries are more than a computational convenience and they should either have biological motivation, or be omited from simulations.

Another difference between the way these simulations, and the detailed synaptic simulations have been performed is that we used random sequential dynamics, whereas parallel dynamics have been used in the simulations of the complete synaptic models (Miller, 1994; Linsker, 1986). This implies that all synapses in the network are updated synchronously, i.e: at once. This assumption does not seem biologically plausible, furthermore although parallel dynamics are computationally convenient, they may alter the final state the network will reach and in particular they can prevent the model from obtaining the real energy minima.

These types of continuous spin models, even with only excitatory connections, can attain vortex states as well, if a finite amount of noise is introduced into the dynamics (Kosterlitz and Thouless, 73). These vortex states however are fluctuations in the statistical mechanics sense, therefore they will move in time and in this sense they are quite different than the vortex states we will show can be attained with models which include inhibition.

3.3.2 Interactions with Inhibition

In the case where there is some inhibition in the network, the situation is more complex, and more interesting. In this case we do not know what the global minima is, we will therefore use simulations of the toy models in order to examine numerically what these minima are.

Using the toy model has the great advantage of being much less complex, and therefore requiring much less computation, furthermore definitions such as orientation selectivity, and orientation angle have a transparent definition. On the other hand it runs the risk of not being a valid approximation. This would happen if when running the full model, the final receptive fields would have strong components of the less prominent eigenfunctions which are not represented in the approximation.

In order to examine the amount of single cell symmetry breaking, we define an order parameter, $rms(a)$, which is defined as $rms(a) = \sqrt{< a^2 >}$ where the average is taken over all network sites, this order parameter shows how much the single cell receptive fields in the network deviate from non interacting single cells. Since in the all excitatory case the values of $rms(a)$ should be like those in the non interacting case, we can asses the effect of the interaction by comparing the values of $rms(a)$ for interactions with inhibition, to those which have only excitation. Strong effects of the interaction may also imply that the approximation may break down.

In our simulations, we used an interaction function of the form,

$$\tilde{I}(\mathbf{r}) = \exp[-(r^2/2\sigma^2)] - IN * exp[-(r^2/2(2\sigma)^2)],$$

where we have fixed $\sigma = 2.5$. Hence for simplicity we have varied only the value of IN. A cross sections of the different interaction terms used are shown in figure: 3.3.

The cases we have examined are **Excitatory** $IN = 0.0$, **Balanced** $IN = .25$, where the integral of the interaction function is zero, **Net Inhibitory** $IN = .5$ in which the is more inhibition then excitation, and **Net Excitatory** $IN = .125$ where there is more excitation then inhibition. We have run all of our simulations for at least 1500 iterations and tested the stability of the final state, in order to make sure that we have indeed reached a stable fixed point.

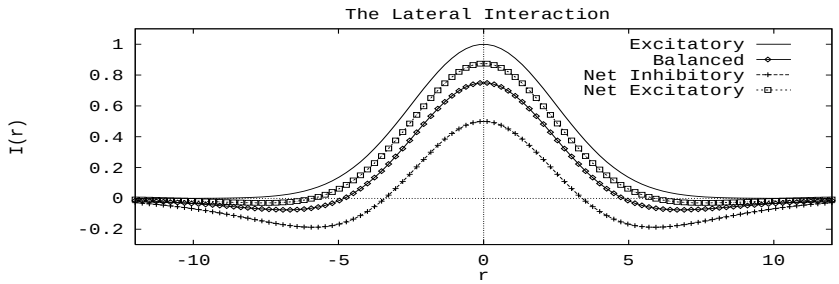

Figure 3.3: *This image presents the different types of interaction functions of the form* $I(r) = \exp(-r^2/\sigma^2) - IN * \exp(-r^2/(2\sigma)^2)$. **Excitatory** $IN = 0.0$*, i.e there is only excitation,* **Balanced** $IN = .25$ *i.e the inhibition and excitation are balanced,* **Net Inhibitory** $IN = .5$ *more excitation then inhibition, and* **Net Excitatory** $IN = .125$ *more excitation then inhibition.*

As can be seen figure 3.4 the interaction can indeed break the symmetry, although only mildly since under the most conditions investigated most receptive fields remain dominated by the first principal component. In figure 3.5, we chose to display the network, in an extreme case of symmetry breaking, even here most receptive fields have a strong, radially symmetric component.

The effect of the interaction on the global ordering is more pronounced. In figure 3.6, we can see that when the radially symmetric principal component is dominant, the effect of the interaction would be a staggered type of final state.

When the oriented solution dominates, the emerging structure has bands with similar orientations, and some orientation singularities as can be seen in figure: 3.7.

These singularities however, are of the $S = 1$ type which means that, around a singularity, the orientation angle changes by 360 deg. It has been claimed (Erwin et al., 1994) that in the cortex and in many models such as in the Swindale (92) model and in the Miller (94) model, the singularities are of the $S = \frac{1}{2}$ type. i.e the orientation angle changes by 180 deg, examples of such singularities are shown in figure 3.8.

This stands in contradiction to the results in the toy model, which therefore raises the question of what is the source of these differences. In the Swindale model the orientations angles are defined to be half of the angle of the complex variable z, i.e $\phi_s = \frac{1}{2}\phi$, and thus when ϕ has a $S = 1$ singularity ϕ_s has a $S = \frac{1}{2}$ singularity.

The differences between the toy model and the Miller model actually stem from a similar reason. We claim that when the dominant single cell solution has the angular component $\cos(m\theta)$, or in general when a rotation of $(360/n)$ deg leaves the receptive field as it was, then the resulting singularities will be of the type $S = \frac{1}{m}$. The heuristic reasoning for this claim is the following; suppose that we divide the interaction function into two terms,

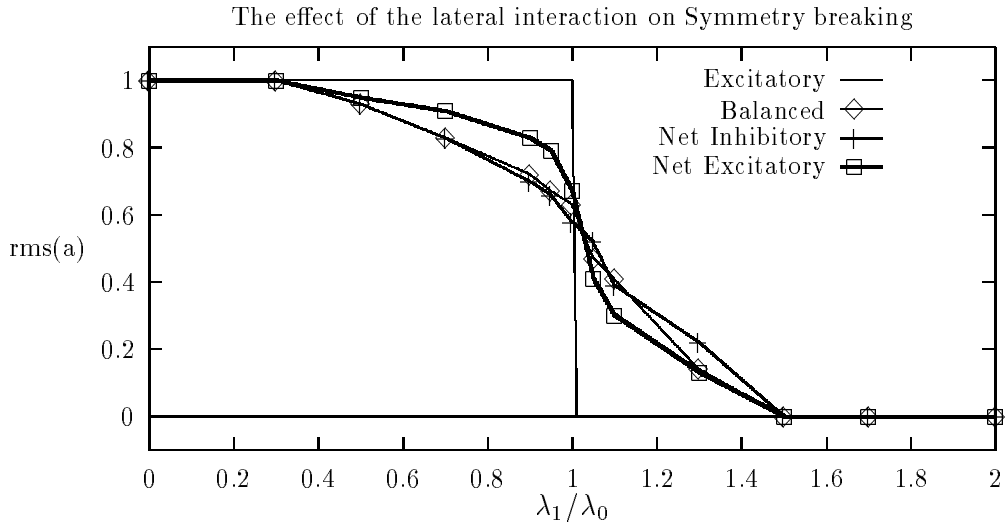

Figure 3.4: *This image presents the rms value of a as a function of* λ_1/λ_0 *for the different types of interactions as defined above. Inhibitory interactions can create orientation selective cells in regions in which single cells are radially symmetric. The different types of interaction we have used have the form* $I(r) = \exp(-r^2/\sigma^2) - IN * \exp(-r^2/(2\sigma)^2)$. *and differ in the value of* IN, **Excitatory** $IN = 0.0$, *i.e there is only excitation,* **Balanced** $IN = .25$ *i.e the inhibition and excitation are balanced,* **Net Inhibitory** $IN = .5$ *more excitation then inhibition, and* **Net Excitatory** $IN = .125$ *more excitation then inhibition.*

an inhibitory term and an excitatory term, $I = I_E + I_I$ such that $I_E > 0$ and $I_I < 0$. For the functions I we have chosen, the excitatory term dominates for short range interactions. The excitatory term attempts to enforce continuity, and low curvature, i.e: a small as possible change between neighboring orientations. Thus a closed path around a singularity should exhibit the minimal curvature which retains continuity which for receptive fields with an angular dependence $\cos(m\theta)$, is a rotation around the curve by $360/m$ degrees, i.e a singularity of type $S = \frac{1}{m}$. More explicitly we predict that when the single cell solution has a $\cos(2\theta)$ angular dependence, then the singularities exhibited will be of the type $S = \frac{1}{2}$, in figure 3.9 we have displayed what the approximate model produces when we make this assumption.

This is actually similar to Swindale's original argument for why he chose $\phi_s = \frac{1}{2}\phi$, which originated from the type of symmetry the receptive fields exhibit. Performing the Swindale transformation in the toy model with $\cos(\phi)$ angular dependence will transform it into a $\cos(2\phi)$ model, which is exactly the model simulated in figure 3.9.

The receptive fields obtained in the $\cos(2\theta)$ model look far from plausible biologically,

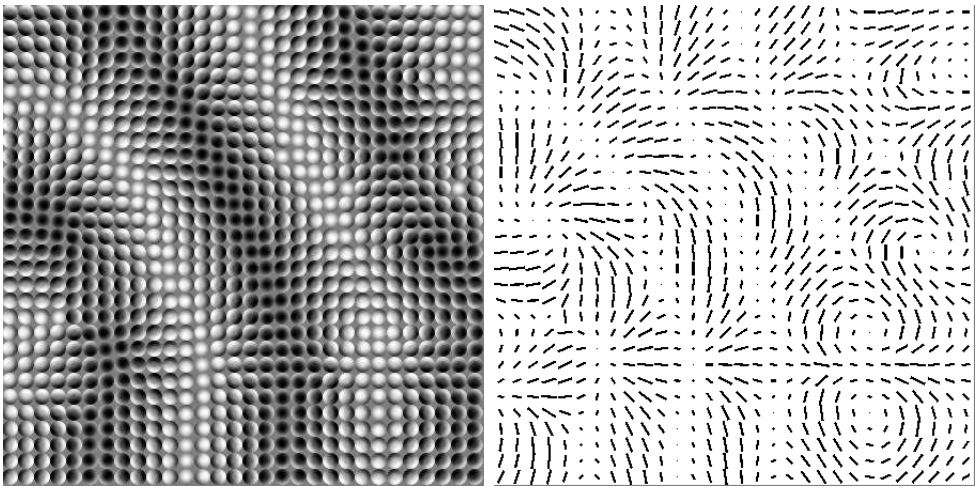

Figure 3.5: *Simulations results for $\lambda_1/\lambda_0 = 0.95$ and a balanced interaction. The receptive fields (left) and the orientation directions (right). The single cell solution is radially symmetric, but here some symmetry breaking does occur although many of the cells are still dominated by the radially symmetric component (shown by the short length of many of bars representing orientation). In order to represent the single receptive fields we have chosen an arbitrary normalized functions, one radially symmetric and one with an angular dependence of $\propto \cos(\theta)$*

in the previous chapter we saw how this may be corrected by the asymmetric portion of the correlation function, which we did not take into account here, however, if we simulate the degenerate case $\lambda_2 = \lambda_0$ as shown in figure 3.10 we obtain more plausible receptive field even in the absence of the non symmetric portion of the correlation function. This case also maintains the $S = \frac{1}{2}$ singularities.

Hence we predict that the type of singularities exhibited in synaptic types of networks, depend critically on the dominant eigen function of the single cell in the same parameter regime. Therefore claims about synaptic models (Erwin et al., 1994) should be restricted to the parameters examined. This prediction results from the simplifications obtained by the analysis performed in this chapter, and is but one example of the power and usefulness of such analysis.

3.4 Related Models

When we do not make the common input assumption, the overlap function O takes on a different form.

The other extreme from the common input case, is the case where inputs to neighboring cells have no spatial overlap. In this case the overlap function takes the simple form

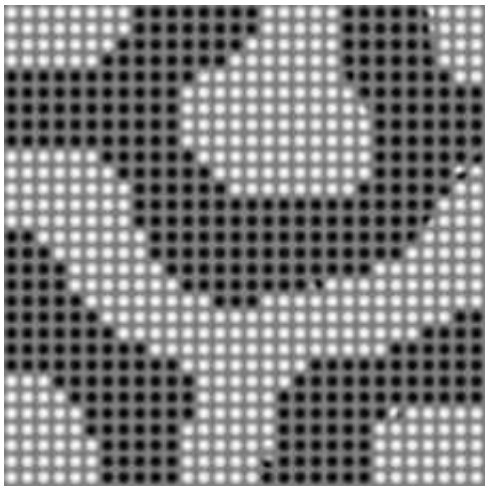

Figure 3.6: *Simulations results for $\lambda_1/\lambda_0 = 0.3$ Balanced interaction. The single cell solution is radially symmetric, and so are the network solutions. The effect of the inhibitory term in the lateral interaction has been to form staggered ordering.*

$O\left[1,1,1,1,\mathbf{x},(\mathbf{x}'),\beta(\mathbf{x}),\beta(\mathbf{x}')\right] = \delta(\mathbf{x}-\mathbf{x}')\delta_{ll'}\delta_{mm'}$ In which case, the Hamiltonian takes the form

$$H = \sum_x \sum_{ln} \tilde{I}(0)\lambda_{ln}a_{ln}^2 - \gamma(a).$$

Thus the Hamiltonian becomes a a non interacting Hamiltonian, and no effective interaction exists between neighboring neurons. Therefore the final state will be such that each neuron will settle in the single cell receptive fields, and the orientations will not be coupled. For this model no continuity of the orientation field should be expected.

A more complex case is the one which is examined in the Linsker and Miller models, in which $\beta(\mathbf{x}) = \mathbf{x}$, in this case the overlap function is complex, and depends on the exact details of the receptive fields, however we can obtain some understanding by looking qualitatively at the form of the overlap function. When $\beta(\mathbf{x}) = const$ an overlap exists only between same type receptive fields, when the inputs are shifted, interactions exists between different types of receptive fields as well, e.g between $\cos(\theta)$ and $\cos(2\theta)$ receptive fields. The interaction between same type receptive fields is also a function of the distance of their centers in the input plane, and in particular it can change signs as a function of $|\mathbf{x}-\mathbf{x}'|$. Some such abstract examples are shown in figure 3.11.

When this happens, a model which is purely excitatory can behave like a model with inhibition. To understand this lets assume for simplicity that $\lambda_{11} > \lambda_{ij}$ for every i and j, and that the cross term overlaps are very small as well. In such a case the Hamiltonian

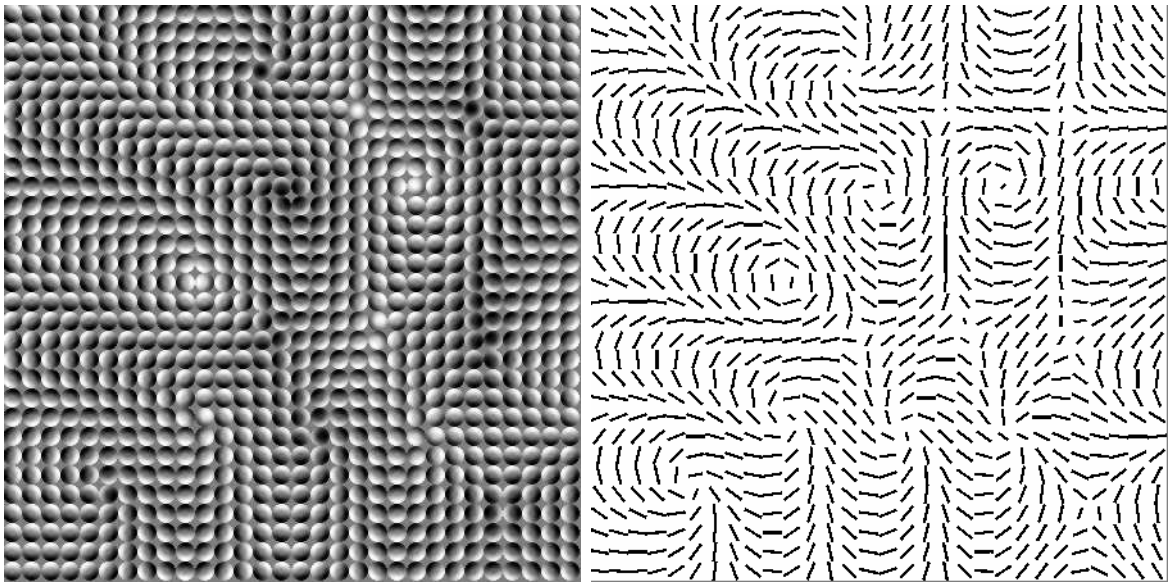

Figure 3.7: *When $\lambda_1/\lambda_0 = 1.3$ the final state shows a banded structure, and structures sometimes described in the biological literature as pinwheels, The map displayed here shows singularities of the $S = 1$ type.*

will take the form,

$$H = \sum_{\mathbf{x}\mathbf{x}'} \tilde{I}(\mathbf{x}-\mathbf{x}')O\left[l,n,l',n',\phi_{ln}(\mathbf{x}),\phi_{ln}(\mathbf{x}'),\beta(\mathbf{x}),\beta(\mathbf{x}')\right]. \tag{3.16}$$

For a purely excitatory $\tilde{I}$ the aim will be to minimize the overlap functions O, as a function of the angles ϕ. When observing figure 3.11, it is obvious that for $Y = \beta(\mathbf{x}) - \beta(\mathbf{x}')$ small the two receptive fields would preferentially have the same phase angle ϕ, where as at larger distances it would be minimized when $\phi(\mathbf{x}) = \phi(\mathbf{x}') + \pi$. This observation enables us to understand that the common input approximation is sensible, If the size of the interaction term $\tilde{I}$ is smaller that the first zero crossing of the overlap term O.

3.5 Discussion

We investigated a model of interacting Hebbian neurons and have shown that the learning in such a network is equivalent to minimizing an energy function. We have further shown that under certain conditions this network can be approximated by continuous spin models, and is similar to some of the proposed phenomenological models as well. In the common input case we have shown that for excitatory interactions there is no symmetry breaking, and that all receptive fields should have the same orientation. We have also used this case to make the point that sticky boundary assumptions can drastically change the organization of receptive fields in the model.

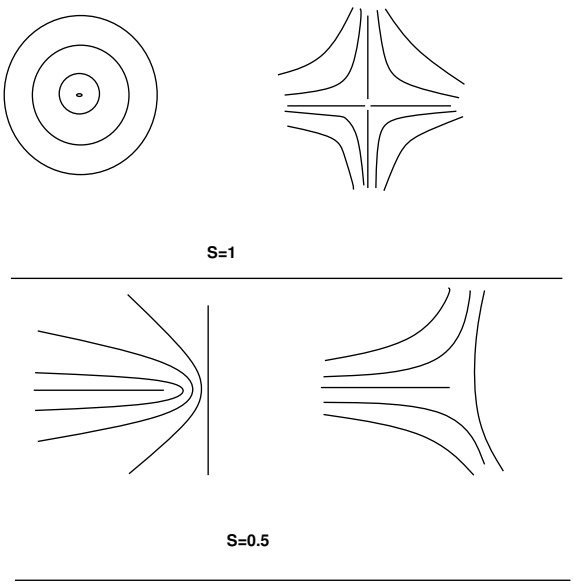

Figure 3.8: *A schematic drawing of different types of singularities , $S = 1$ are drawn above, $S = \frac{1}{2}$ below .*

Inhibition we have shown can effect both the structure of the receptive fields and the organization of different receptive fields across the cortex. In extreme cases we have shown that inhibition can create oriented receptive fields, where the single cell receptive fields are radially symmetric.

The global organization produced by models with inhibition, shows Iso-orientation patches, and pinwheel type of singularities. We demonstrated that the type of singularities depend critically on the on the symmetry of the receptive fields.

The major qualitative difference between the common input model, and shifted input models, is that even in the absence of inhibition, shifted input models behave like models with inhibition. We have qualitatively demonstrated why this is the case.

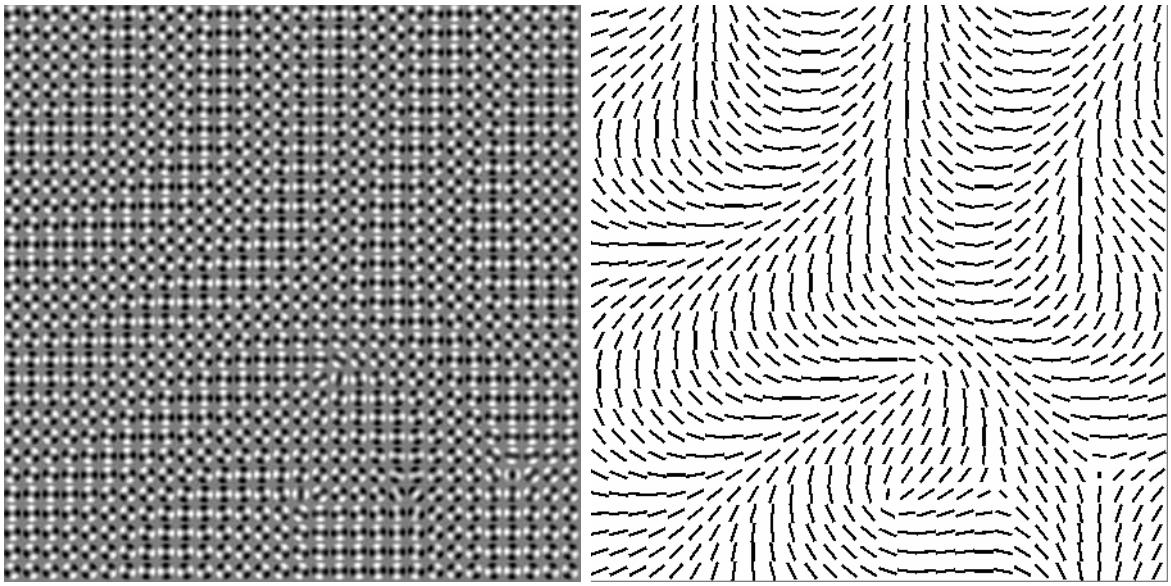

Figure 3.9: *When the dominant eigenfunction has a* $\cos(2\theta)$ *angular form, and respectively* $\lambda_2/\lambda_0 = 1.3$ *the final state shows* $S = \frac{1}{2}$ *type singularities.*

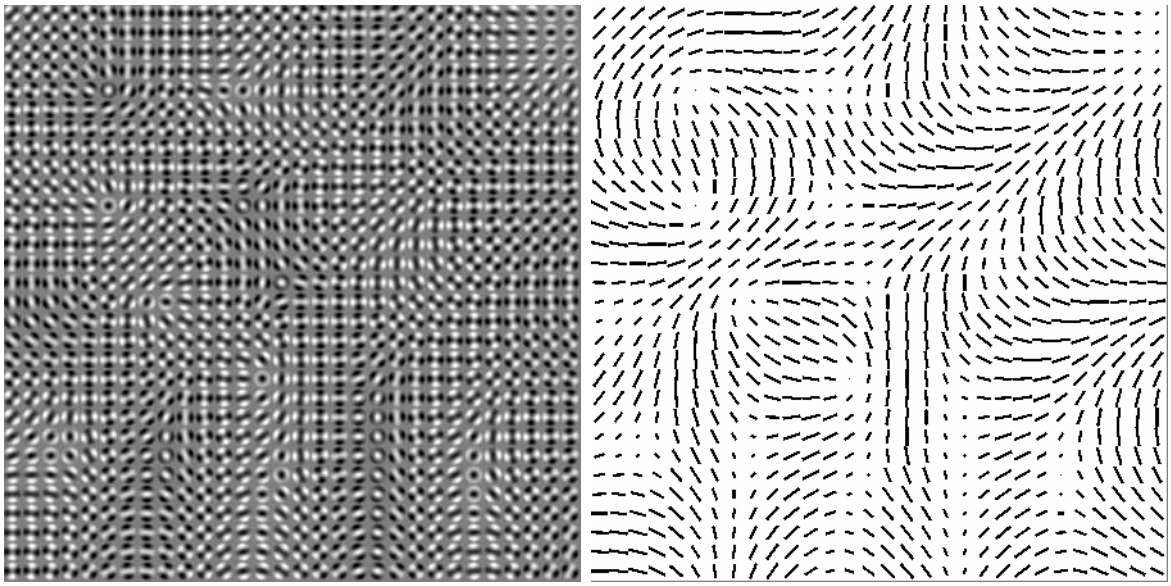

Figure 3.10: *In the degenerate case, when the eigen value of the radially symmetric, center surround, solution* λ_0 *is equal to that of the* $\cos(2\theta)$ *solution, the resulting receptive field structure seems biologically plausible, and the final state shows* $S = \frac{1}{2}$ *type singularities.*

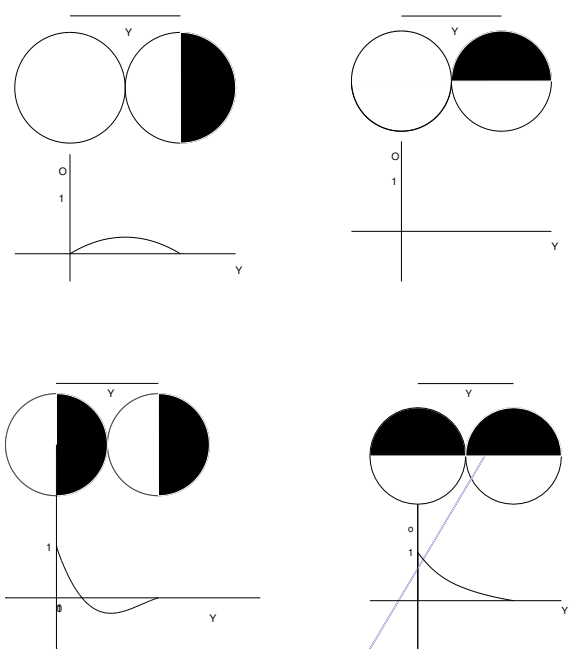

Figure 3.11: *A schematic drawing of some aspects of the overlap functions for shifted inputs. We have defined* $Y = \beta(\mathbf{x}) - \beta(\mathbf{x}')$. *Above cases of overlaps between different types of RF's are displayed, below between same kinds of RF's*

Chapter 4

Simultaneous formation of Binocularity and Orientation Selectivity

4.1 Introduction

The importance of visual experience for the formation of orientation selectivity, has been reviewed in the introduction to this thesis, however the best known model for plasticity in the visual cortex is the formation of binocularly driven receptive fields in visual cortex neurons. Furthermore it seems that whereas some degree of orientation selectivity can appear before eye opening, the formation of binocular cells which have closely matched orientation preference in both eyes, requires binocular visual experience (Blakemore and Van-Sluyters, 1975). Our aim is to understand how plasticity can achieve oriented receptive fields, which are closely matched in both eyes. Since this happens only after eye opening, the use of a realistic patterned visual environment is essential.

In this chapter we adopt the approach used throughout the thesis, we try to understand the simple cases before delving into the more complex models. Thus we will only examine a single-cell model herein, as a first step to an understanding of network properties. Unlike the other chapters of this thesis, in this chapter we compare how two different learning rules PCA (Oja, 1982) and BCM (Bienenstock et al., 1982) perform this task, under otherwise identical conditions.

At a first glance the task seems simple, if both eyes receive the same input, and if the input vectors from both eyes are identical, it is most likely that identical receptive fields would evolve in both eyes. However this approach in the single cell model, can only create

binocular cells, furthermore it assumes that the developmental process is precise enough to exactly align the receptive fields in both eyes. It was suggested by Blakemore and van Sluyters (75) that plasticity may be needed in animals with binocular vision in order to overcome an imprecise developmental alignment. It may also be needed in order to adjust the receptive field as the head size grows. We will examine what happens when the receptive fields from both eyes are not exactly aligned, and see that for the BCM learning rule this misalignment is sufficient for producing varying degrees of ocular dominance, already in the single cell model, whereas for the PCA learning rule binocular neurons will be produced in every case.

4.2 Model of Binocular inputs

In figure 4.1 we have illustrated how the inputs from both eyes are chosen. First the natural scenes are convolved with a center-surround filter, which is analogous to the retinal preprocessing. Here we have chosen a DOG filter with a center radius of one pixel ($\sigma_1 = 1.0$) and a surround radius of three ($\sigma_2 = 3$). [1] The input vectors from both eyes are chosen as small, partially overlapping, circular regions of these preprocessed images.

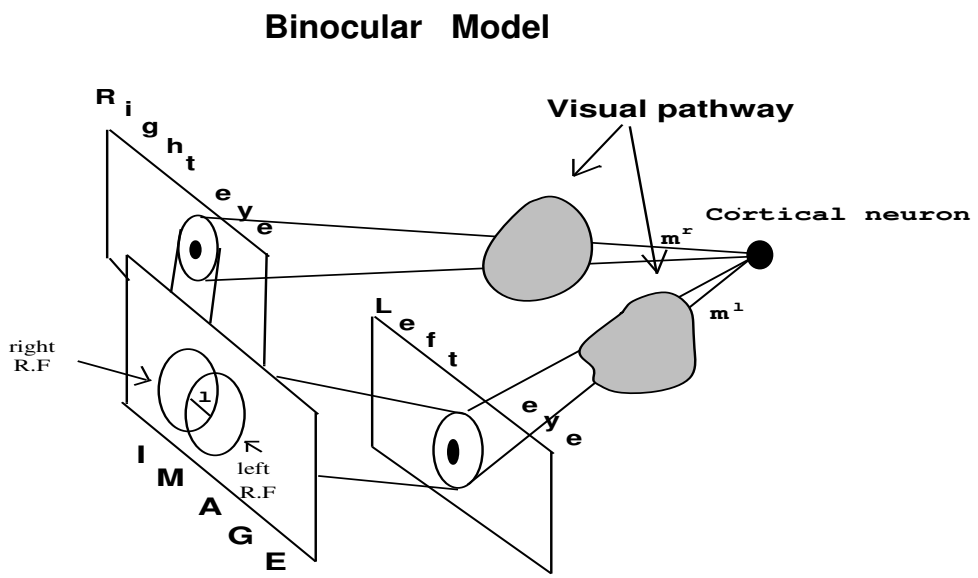

Figure 4.1: *The natural scenes which are encoded by receptors in the retina, are convolved with a center surround filter. Then two partially overlapping circular regions are chosen as the input vectors.*

The input from the right and left eye respectively are denoted by $\mathbf{d}^R$ and $\mathbf{d}^L$, and

[1] This ratio between the center and surround in biologically plausible, and enables both learning rules to produce oriented receptive fields.

the output of each neuron then becomes

$$c = \sigma(\mathbf{d}^r \cdot \mathbf{m}^r + \mathbf{d}^l \cdot \mathbf{m}^l), \tag{4.1}$$

where σ is the, possibly non linear, activation function of each neuron.

In order to get sensible receptive fields with the BCM learning rule it is essential to use a non symmetric activation function σ of the type shown in figure 1.8, as discussed in section 1.4.1. In the case of the PCA learning we have examined both the linear case, and the asymmetric activation function of figure 1.8

Choosing this type of function is a sensible choice for the cortex, because cortical cells have low spontaneous activity. Since we equate the spontaneous level with the zero of the activation function, then a low spontaneous firing rate implies that the activation function should not be able to go much below zero. A non symmetric sigmoid, with a positive saturation limit being much higher then the negative one, is also sensible, and would not alter the results.

In order to examine the effect of varying the overlap between the receptive fields we define an overlap parameter O, such that $O = l/2a$, where a is the receptive field radius in pixels, and l is the linear overlap in pixels, as shown in figure 4.2. When the left and right receptive fields are completely overlapping $O = 1$, and when they are completely separate $O \leq 0$.

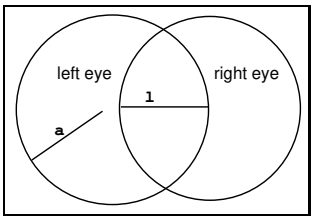

Figure 4.2: *The overlap parameter O is defined as $O = l/2a$. When $O = 1$ the receptive fields are completely overlapping, when $O = 0$ they are non overlapping but touching $O < 0$ means that they are non overlapping and not touching.*

In order to asses the degree to which a cell is binocular, and in order to make ocular dominance histograms we devised an ocular dominance measure B,

$$B = \frac{L - R}{L + R}. \tag{4.2}$$

To calculate B, we have to first find the orientation at which the cell has the greatest binocular response to a sinusoidal grating, and at this orientation we measure L and R, the left and right eye responses respectively.

We have been motivated by the measure used by Albus (Albus, 1975) in defining the bin boundaries, for a seven bin ocular dominance histogram, thus we have defined

$$\begin{cases} -1 \leq B \leq -.85 & \text{Bin 1} \\ -0.85 \leq B \leq -0.5 & \text{Bin 2} \\ -0.5 \leq B \leq -0.15 & \text{Bin 3} \\ -0.15 \leq B \leq 0.15 & \text{Bin 4} \\ 0.15 \leq B \leq 0.5 & \text{Bin 5} \\ 0.5 \leq B \leq 0.85 & \text{Bin 6} \\ 0.85 \leq B \leq 1 & \text{Bin 7} \end{cases}$$

This measure was motivated by the one used by Albus since it is one of the few quantitative measures used. We altered it in the following manner, Albus measure defines cells to be monocular when they respond to one eye, and not to the other, this never happens in simulations since there is always some small response to both eyes. In our measure a cell falls into the binocular group when the peak response to the weak eye is less then 8% of the strong eye response, even this measure is probably too extreme since random fluctuations are usually sufficient to drive eyes out of this range, and as will be seen receptive fields may easily look monocular, and not fall into the monocular bins. The other bin boundaries have been slightly shifted as well in order to get a simpler evenly spaced measure.

4.3 BCM rule: results and robustness

We performed simulations of binocular receptive field formation using the BCM learning rule as defined in equations 1.8,1.9,1.10 and with an activation function f as shown in figure 1.8. At first we will present the effect of fixed horizontal shifts between the two receptive fields, at seven different ocular overlap values O = 1.0, 0.8, 0.6, 0.4, 0.2, 0.0, -0.2. Receptive fields had a radius $r = 10$. At each overlap value we ran 100 single cell simulations. In figure: 4.3 we display the different ocular dominance histograms for each different ocular overlap value. As O decreases so does the degree of binocularity.

In the following set of figures, figures 4.4 to 4.13, we have shown receptive fields and tuning curves of neurons with different degrees of overlap O. The general trend is that for a complete overlap ($O = 1$) highly binocular cells are also highly orientation selective as can be seen in figure 4.4 , when the overlap decreases however, the highly selective cells tend to be monocular, as can be seen for example in figure 4.10.

This happens since for low overlap binocular cells are those which have an oriented receptive field which is strong only in the overlapping region, and therefore such receptive fields are typically short, as seen in figures 4.6 and 4.9. Less binocular cells however, as

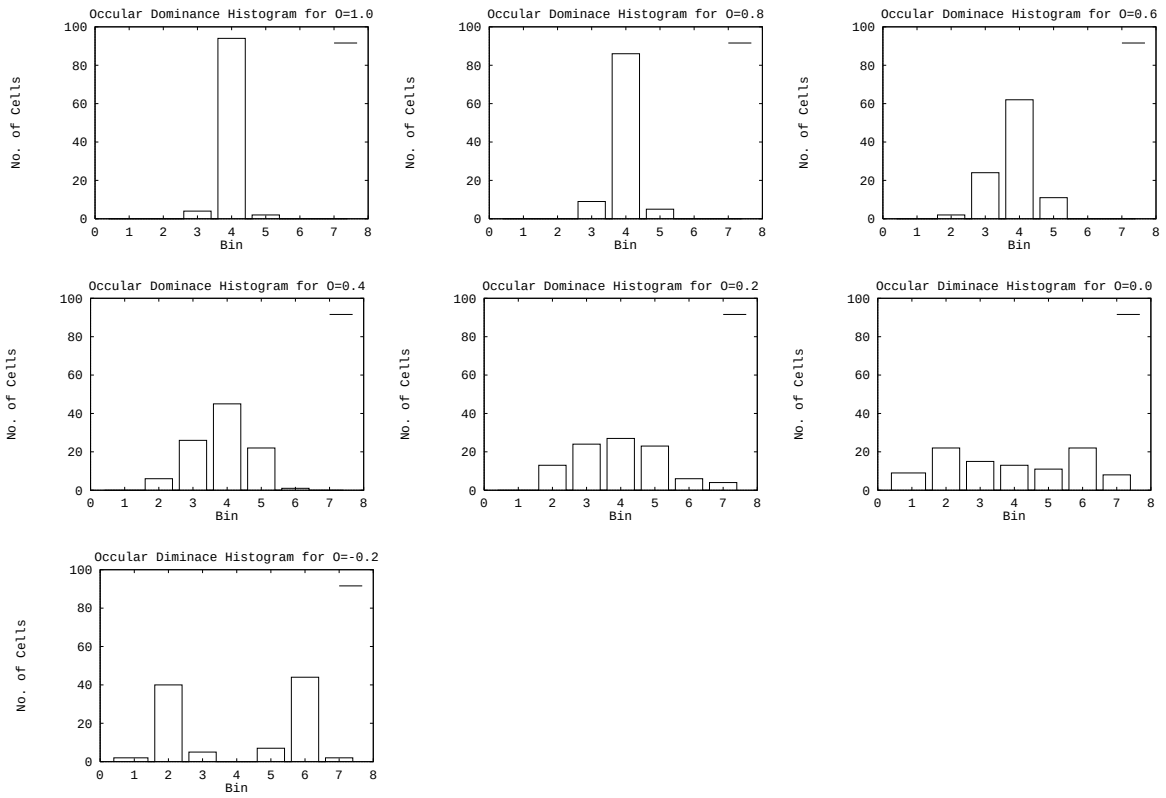

Figure 4.3: *Ocular dominance histograms for seven different values of the ocular overlap O. Obtained using the BCM rule. At each ocular overlap value* 100 *different single cell simulations have been run, using the BCM learning rule. It is seen that as O decreases so does the degree of binocularity.*

seen in figures 4.7, 4.10, 4.11 typically have an elongated receptive field in one of the eyes, and a short, flat receptive field in the other, thus becoming less binocular.

Even when there is no overlap between the cells, correlations between the receptive fields exist, due to elongated patterns which cross both receptive fields. In figures 4.12 4.13, two examples of receptive fields which developed in a non overlapping case, are presented. In figure 4.12 a cell of weak binocularity is shown, it can be seen that such cells have typically only weak orientation tuning, whereas in figure 4.13 a completely monocular cell is presented, which is highly orientation selective in the responsive eye.

In order to examine whether shifting in non horizontal directions would alter the results, we compared the histograms for an overlap parameter of $O = 0.5$, at three different angles horizontal, vertical, and at 36 deg. We have run 60 different single cell simulations at each different angle. As can be seen in figure 4.14, these different histograms are qualitatively similar, there seems to be a tendency for more binocular cells when the shift is horizontal, but the size of the sample we tested does is too small for these results to be

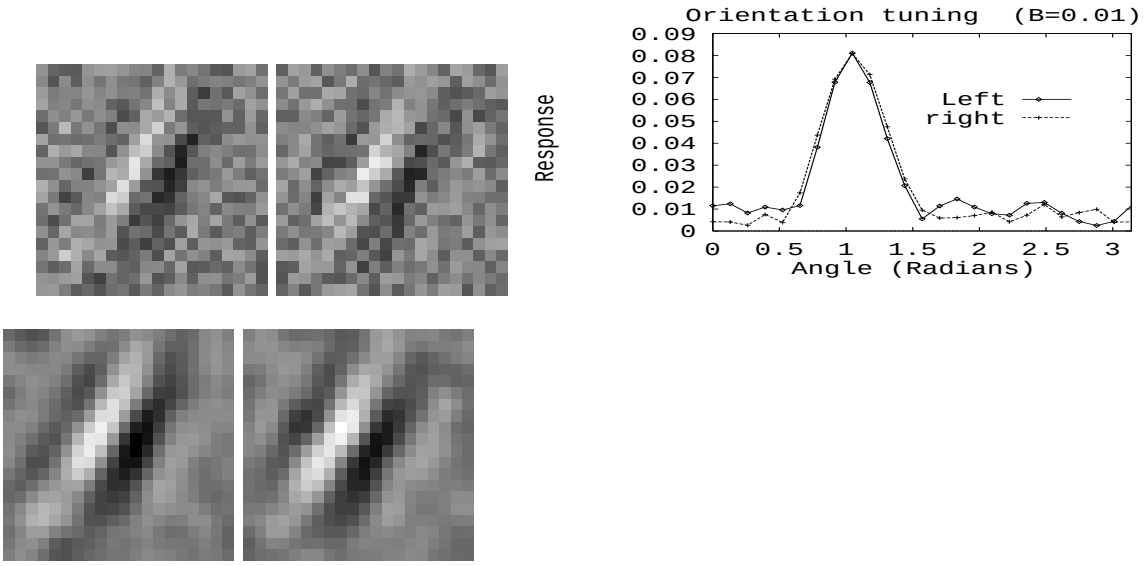

Figure 4.4: *Receptive fields when $O = 1.0$; using the BCM rule. Both eyes have nearly identical receptive fields, and tuning curves. Above the LGN-to-cortex receptive fileds are presented, below the retina-to-cortex receptive fields. All subsequently receptive fields displayed retina-to-cortex receptive fields.*

significant.

We examined the robustness of these results, by changing the receptive field sizes to 10 and 30 pixels, and found no qualitative differences. Furthermore we changed slightly the learning rule to the one used in (Law and Cooper, 1994), in which

$$\phi(c) = \frac{1}{\theta_m} c(c - \theta_m). \tag{4.3}$$

This change speeds up the simulations, but does not alter the final results.

From these results we can conclude that the BCM learning rule is sufficient for matching partially overlapping receptive fields from both eyes, furthermore it can possibly

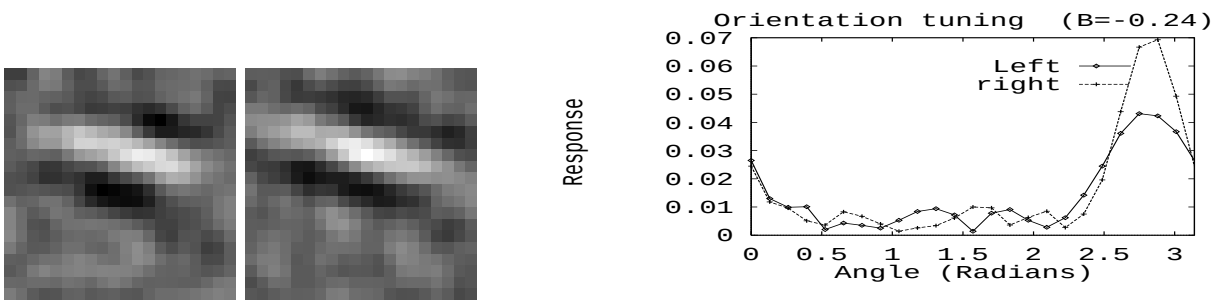

Figure 4.5: *Receptive fields when $O = 1.0$, using the BCM rule. A less typical case, the receptive field is less binocular, the cells are not as sharply tuned as well.*

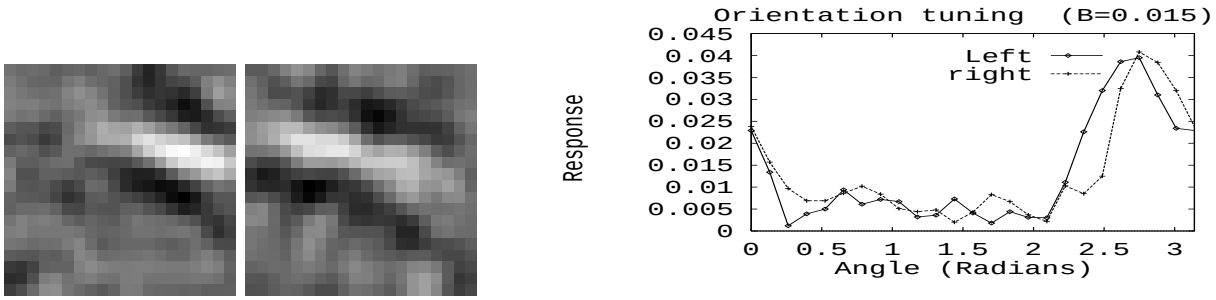

Figure 4.6: *Receptive fields when $O = 0.6$, Using the BCM rule. Here a highly binocular neuron is displayed, typically receptive fields evolve only in the overlapping region.*

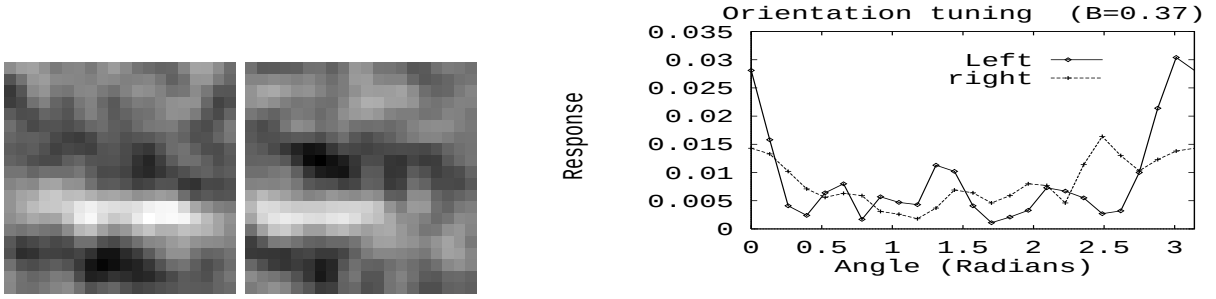

Figure 4.7: *Receptive fields when $O = 0.6$, Using the BCM rule. Less binocular cells may have a receptive field which is more extended in one of the eyes then in the other.*

explain the origin of varying degrees of binocularity in the cortex. We do not attempt here to produce exactly the shape of the ocular dominance histograms in the cat, let alone of different species. It is clear, however, that by having different initial distributions of overlaps i.e: by adding together with different weights histograms of the type displayed in figure: 4.3, we can get almost any ocular dominance histogram we desire.

Since this is a single cell study, it is clear that we do not explain here the origin of ocular dominance columns, and their interactions with orientation columns.

4.4 PCA rule: results and robustness

In order to examine the effects of varying degrees of overlap on neurons which use the PCA learning rule, and compare them to BCM neurons, we performed single cell simulation using the PCA rule, with otherwise identical conditions. We used both the linear, and the asymmetric activation function f. As can be seen in figure 4.15, the different choices of activation functions, do not seem to alter the shapes of the resulting receptive fields. [2]

[2]For the PCA learning rule with linear neurons, there is a symmetry with respect to a sign change of all of the weights, this is not the case for non symmetric activation functions

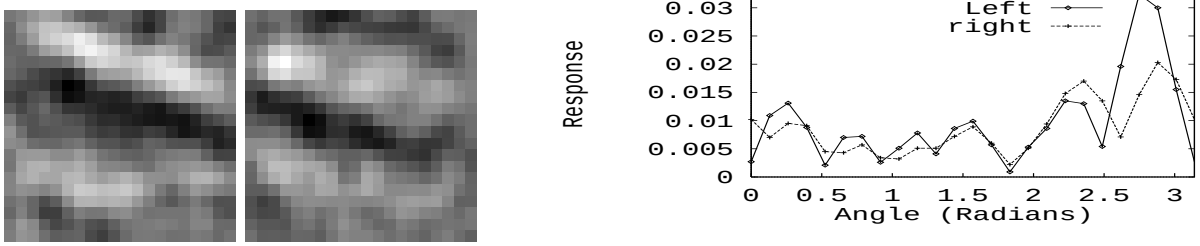

Figure 4.8: *Receptive fields when $O = 0.6$, using the BCM rule. A typical case with intermediate binocularity.*

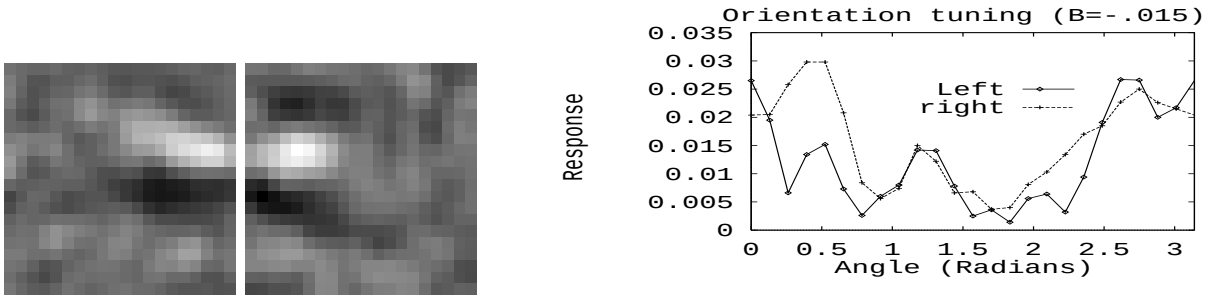

Figure 4.9: *Receptive fields when $O = 0.2$, using the BCM rule. In cases when the overlap between the receptive fields is small, binocular cells are typically only weakly orientation selective.*

For PCA learning, it seems that for a given overlap, the final shapes of the receptive fields are almost identical, and independent of the initial conditions. Whenever the receptive fields are completely, partially, or non overlapping, the PCA learning rule always yields binocular receptive fields.

For the partial overlap case the receptive fields seem to concentrate in the overlap region between the eyes. As seen in figures 4.16, and 4.17, the receptive fields for a low overlap are very short in both eyes, thus resulting in a broad tuning curve.

The receptive field in the completely non overlapping case, is similar to the completely overlapping case. This result is actually simple to explain, since for the non overlapping case there is still a small positive correlation between the eyes, thus this case falls into the perturbative regime examined by Dayan and Goodhill (Dayan and Goodhill, 1992). In this regime, the receptive fields in both eyes should be identical, and similar to the single channel receptive field, obtained under similar conditions. Another way of looking at this result, is that it stems from the fact that receptive fields for a PCA cost function has a single, non degenerate minima, due to the horizontal bias in the images. The receptive fields for both eyes, even if they do not interact will converge to the same horizontal form,

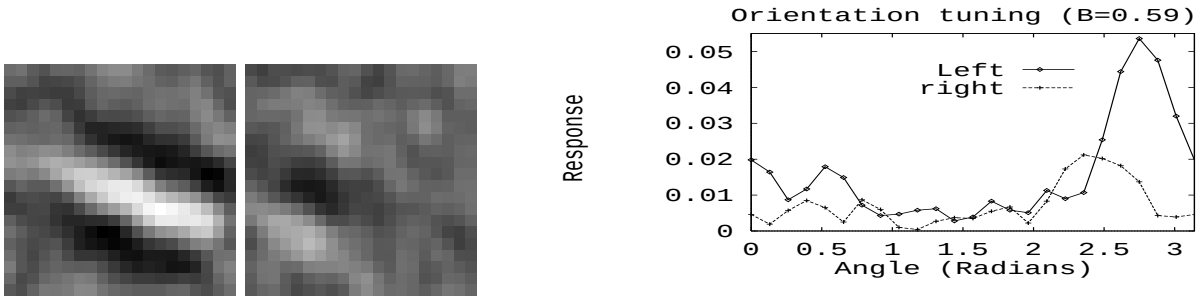

Figure 4.10: *Receptive fields when $O = .2$, using the BCM rule. Less binocular cells have well defined receptive field in one eye, and are highly orientation selective*

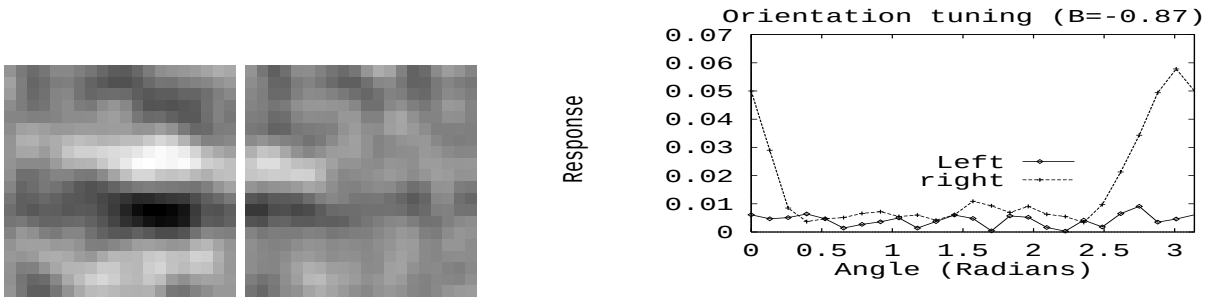

Figure 4.11: *Receptive fields when $O = .2$, using the BCM rule. An example of a monocular cell, with a elongated receptive field, and a sharp monotonic tuning curve.*

independently, and thus the elongated receptive field is highly selective.

As in the BCM case we have examined the effect of shifting in non horizontal directions, several such shifts are displayed in figure 4.19, where it can be seen that the cells remain binocular horizontal, and develop mainly in the overlap region. The overlap region though changes when the direction of the shift changes.

Another difference from receptive fields obtained with the BCM rule, is that the cells trained with the PCA rule, are more broadly tuned than BCM cells.

It is obvious that in the PCA case the single cell model fails to account both for the varying degrees of binocularity, and for the different orientations encountered in the visual cortex.

4.5 Discussion

In animals which exhibit binocular vision, both the binocularity, and the orientation cells are modified after eye opening, and are controlled by the input environment (Blakemore and Van-Sluyters, 1975). Thus the animal attains orientation selective cells, which are matched in both eyes. These cells have varying degrees of binocularity, and different orientation

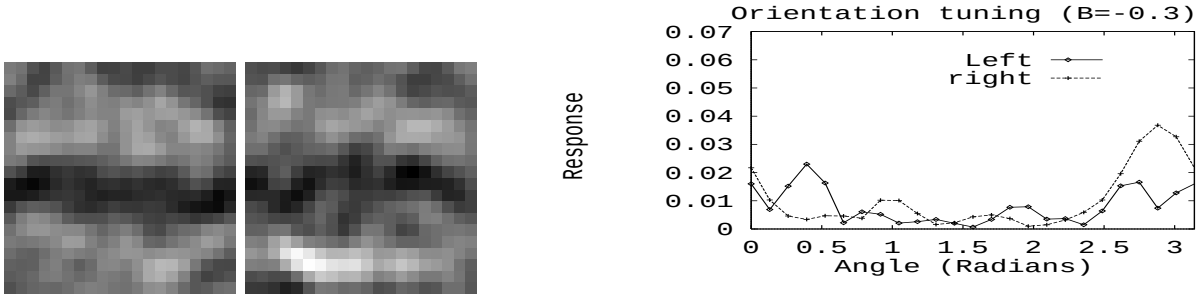

Figure 4.12: *Receptive fields when $O = -0.2$, using the BCM rule. An example of a relatively binocular cell.*

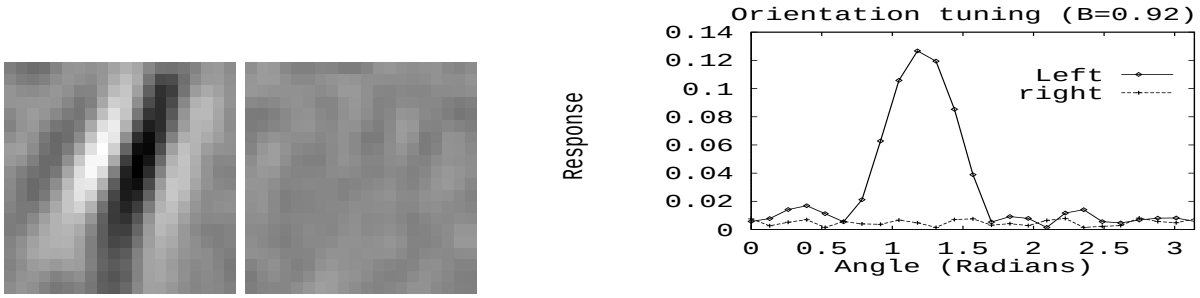

Figure 4.13: *Receptive fields when $O = -0.2$, using the BCM rule. An example of a monocular cell, with a well defined elongated receptive field, and a sharp monotonic tuning curve. Monocular cells are more highly tuned in this no overlap case.*

preferences.

We chose to examine how two different learning rules compare in achieving these two tasks simultaneously. In order to do this we examined a simple single cell model, in which we assume that receptive fields from both eyes are not prefectly aligned.

For both learning rules the receptive fields obtained, have the same orientation preference in both eyes. However here the similarity ends. For the BCM learning rule the assumptions we made are sufficient for attaining different degrees of binocularity and different orientations. This is achieved since for the BCM rule the two eye receptive fields are not symmetric.

When the PCA learning rule is used, the resulting receptive fields are always quite symmetric, thus they are always binocular, and the preferred orientation is always horizontal. Thus it is clear that to obtain sensible receptive fields with the PCA rule, extra assumptions must be made.

An obvious step, which may improve the results of the PCA model, is to go beyond the single cell model to a network model. A network of Hebbian neurons, with binocular receptive fields trained with an abstract correlated input environment has been examined,

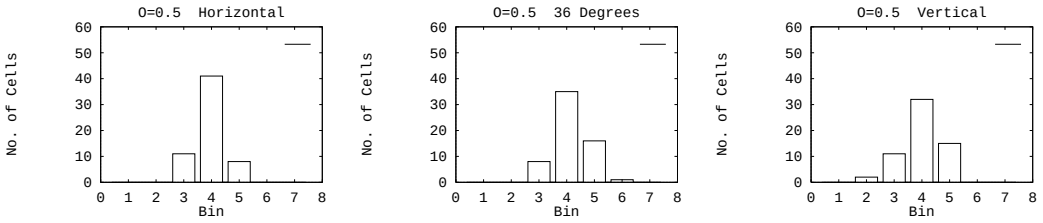

Figure 4.14: *The ocular dominance histograms, for three shift angles (Horizontal , 36 deg and Vertical), are compared, for BCM learning. No qualitative difference is seen. A slight inclination for more cells to be binocular when the shift is in the horizontal direction can be see, although the size of the sample is too small for this small tendency to be significant.*

Figure 4.15: *Receptive fields for completely overlapping input* $O = 1.0$ *using the PCA rule. Comparison of linear neuron (right) and a non linear neuron (left). These examples are almost identical up to a sign (see footnote 2)*

and has failed to produce oriented receptive fields which are not completely binocular (Erwin et al., 1994). For this type of inputs these results are expected in light of the two channels properties of the PCA rule. It remains to be seen if these negative results would change for more realistic inputs , or if the network interactions are non linear.

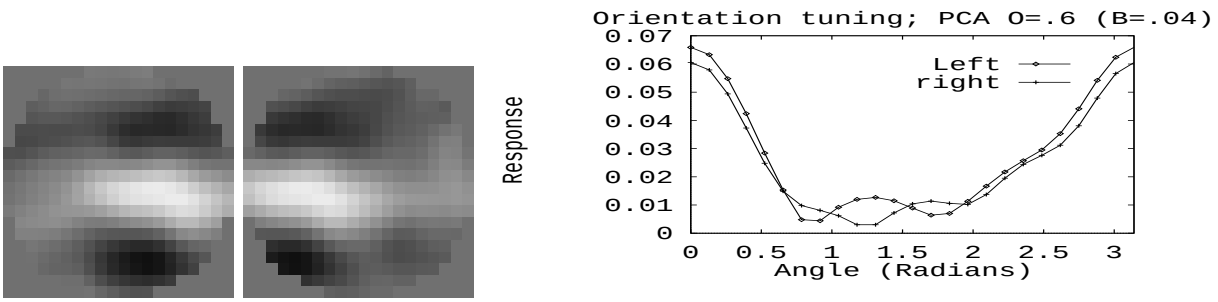

Figure 4.16: *Receptive fields for partially overlapping input,* $O = .6$*, using the PCA rule. The cell is binocular, and the receptive fields are negligible outside of the overlap region.*

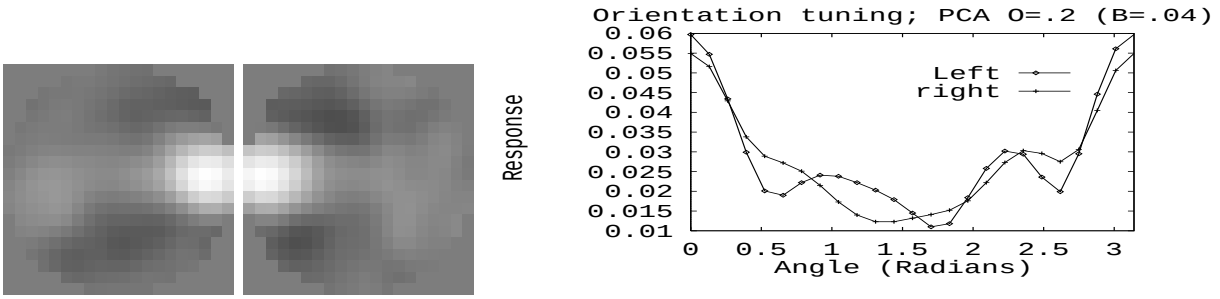

Figure 4.17: *Receptive fields for a small partial overlap, $O = .2$, using the PCA rule. The cell is binocular, and the receptive fields are negligible outside of the overlap region, producing overlapping small almost circular receptive fields in each eye.*

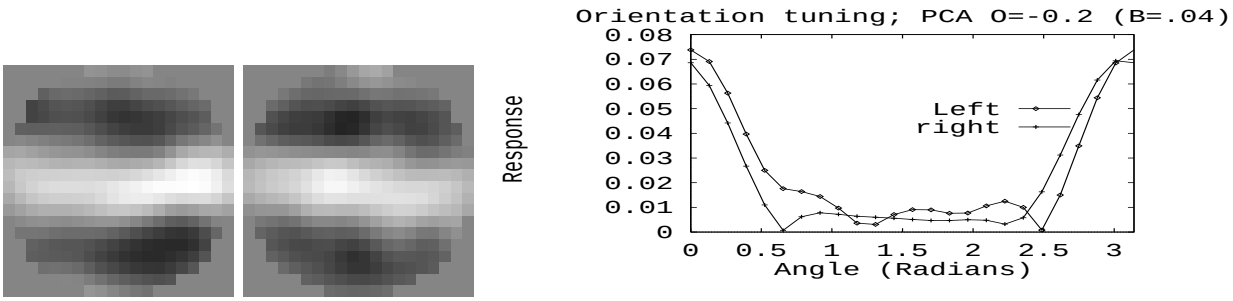

Figure 4.18: *Receptive fields for non overlapping receptive fields, $O = -.2$, using the PCA rule. A small correlation exists between both eyes nevertheless. The cell is binocular, and the receptive fields are similar to the completely overlapping case, and do not concentrate in a small region as in the partial overlap case.*

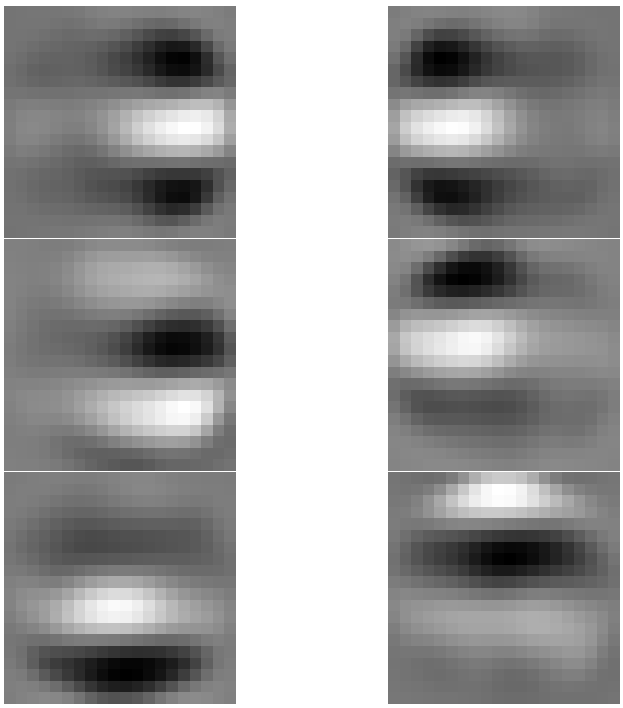

Figure 4.19: *Receptive fields have been shifted in three different directions Horizontally (center), at* 36 deg *(Center) and Vertically (Bottom). The results shown from top to bottom, were obtained for receptive fields of radius* $r = 10$ *and overlap* $O = .5$.

Chapter 5

Conclusions

The underlying assumption of the work in this thesis, is that receptive field formation and organization in the visual cortex is influenced by the input environment. In the introduction I have tried to present evidence supporting this view. We have tried to use an input environment which is as realistic as possible. The major novel feature of this work is the choice we made to use an environment composed of natural images, to model the environment after eye opening.

We tried to keep the models as simple as possible in order to gain real understanding, and avoided complex models, whose behavior may resemble that of the cortex, but which we understand no better than the cortex. Extra complications do have to be added to the simplest models for them to resemble known experimental results. We however tried to introduce these complications gradually, so we will be able to use our knowledge of simpler models in order to understand the more complex ones. We also believe it is more informative, to use the same model to explain the different phenomena, rather then different models which explain different phenomena, but which it is not necessarily possible to combine.

We have not only tried to understand the behavior of the models, but also compare their predictions to those found experimentally. However understanding a model can sometimes help in understanding it's predictions better as well, as seen for example in chapter 3 when we have demonstrated that sticky boundaries can alter the final states of synaptic model simulations.

In chapter 2 we derived the forms of the localized principal components of natural images, based on the result about the correlation function of natural scenes. We have also shown that a non-rotationally symmetric perturbation can break the degeneracy and give a definite phase which only depends on the properties of the real world correlation function.

When comparing the receptive fields predicted here, with the receptive fields of

neurons in the visual cortex, two major differences stand out. The first is that some of the receptive fields predicted, such as the radially symmetric ones, do not exist in the visual cortex. Another difference, is that for each of the non-radially symmetric receptive fields predicted, only two orientations are necessary to form a complete set. This feature is a property of any method which extracts complete orthogonal sets, such as the sanger (89) rule, and is clearly in contradiction to the near-continuum of orientations found biologically.

We then considered the effect that the preprocessing which the inputs undergo, in the retina and the LGN on the receptive field structure. We first analyzed the radially symmetric environment, and have shown that the receptive field structure is critically dependent on the ratio between the size of the preprocessing filter, and the size of the receptive field. We then used perturbation theory in order to approximate the receptive field in the unsymmetrized environment. If the non symmetric components are included, then there exists a regime in which the receptive fields obtained, resemble those found in the visual cortex.

One difference which still stands out between this single cell theory and receptive fields in the cortex, is that the cortex neurons with all orientation exist whereas the single cell theory predicts only a single orientation, the horizontal one.

In order to examine the effects of the lateral interactions on both receptive field formation, and organization we examine, in Chapter 3, a simple model network of interacting neurons, which exhibit hebbian type learning.

We have shown that learning in such a network is equivalent to minimizing an energy function. We have further shown that under certain conditions this network can be approximated by continuous spin models, and is similar to some of the proposed phenomenological models as well. In the common input case we have shown that for excitatory interactions there is no symmetry breaking, and that all receptive fields should have the same orientation. We also used this case to make the point that sticky boundary assumptions can drastically change the organization of receptive fields in the model.

Inhibition, we have shown, can effect both the structure of the receptive fields and the organization of different receptive fields across the cortex. In extreme cases inhibition can create oriented receptive fields, where the single cell receptive fields are radially symmetric.

The global organization produced by models with inhibition, shows Iso-orientation patches, and pinwheel type of singularities. We demonstrated that the type of singularities depend critically on the type of symmetries in the receptive fields.

The major qualitative difference between the common input model, and shifted input models, is that even in the absence of inhibition, shifted input models behave like models with inhibition. We have qualitatively demonstrated why this is the case.

Finally we addressed the problem of the formation of binocular receptive fields, here we compared the results for PCA type learning rules, with results for the BCM learning rule. The formation of binocular receptive fields is experimentally the best studied model of plasticity in the cortex, and therefore it is useful to examine how different learning rules compare in this respect. We assume that the developmental process which occurs before eye opening is incapable of exactly aligning the anatomical receptive fields from both eyes, and that plasticity is responsible for matching the receptive field orientations in both eyes.

Since we have attempted to model the simultaneous formation of orientation selectivity and binocularity, we have designed these models so that they would indeed produce oriented receptive fields for both learning rules, as we have learned from our previous studies (Law and Cooper, 1994; Shouval and Liu, 1994).

For both learning rules the receptive fields obtained, have the same orientation preference from both eyes, however here the similarity ends. For the BCM learning rule the assumptions we made were sufficient for attaining different degrees of binocularity and different orientations. This is achieved since for the BCM rule the two eye receptive fields are not symmetric.

When the PCA learning rule is used, the resulting receptive fields are always quite symmetric, thus they are binocular. The preferred orientation is always horizontal. Thus it is clear that to obtain sensible receptive fields with the PCA rule, extra assumptions must be made.

An obvious step, which may which may improve the results of the PCA model, is to go beyond the single cell model to a network model. A network of Hebbian neurons, with binocular receptive fields trained with an abstract correlated input environment has been examined, and has failed to produce oriented receptive fields which are not completely binocular (Erwin et al., 1994). For the type of inputs used these results are expected in light of the two channels properties of the PCA rule. It remains to be seen if these negative results would change for more realistic inputs, or if the network interactions are non linear.

Appendix A

Appendix

A.1 Representing the correlation matrix in a Bessel function Base

Since the Spectrum of the correlation function has the form, $C(\mathbf{k}) = c/\mathbf{k}^2$ where c is a constant (Field, 1987), $C(\mathbf{r})$ satisfies

$$\nabla^2 C(\mathbf{r}) = -c\delta(\mathbf{r}), \tag{A.1}$$

which can be readily proven by taking Fourier transformation on both side of this equation. In principle, a gauge $g(\mathbf{r})$ which satisfies $\nabla^2 g(\mathbf{r}) = 0$, can be added to $C(\mathbf{r})$. But the only such gauge, which is radially symmetric, is a constant (denoted as b).

A representation for the radially symmetric form of the $\delta(r)$ function in terms of Bessel functions, is

$$\delta(\mathbf{r}) = \frac{1}{2\pi a^2} \sum_i \frac{J_0(k_i^0 r/a)}{N_0(k_i^0)}$$

where where J_l is the lth order Bessel function, k_i^0 is the ith zero of J_0, r and θ are the polar coordinates of $\mathbf{r}$, and $N_0(k_j^0) = \int_0^1 r dr [J_0(k_j^0 \mathbf{r})]^2$ is the normalization constant of $J_0(k_j^0 \mathbf{r})$. Therefore it can be seen that

$$C(\mathbf{r}) = -c \left[\frac{1}{2\pi} \sum_i \frac{J_0(k_i^0 r/a)}{(k_i^0)^2 N_0(k_i^0)} + b \right],$$

since it solves eq:A.1. Using the addition theorem for Bessel functions(Abramowitz and Stegun., 1972), we obtain a representation for the correlation function $C(\mathbf{r} - \mathbf{r}')$

$$C(\mathbf{r} - \mathbf{r}') = c \left[\sum_{jl} \frac{J_l^*(k_j^0 r/a) J_l(k_j^0 r'/a)}{2\pi (k_j^0)^2 \ N_0(k_j^0)} e^{il(\theta - \theta')} + b \right].$$

We shall rewrite this correlation function in terms of the normalized Bessel-Fourier basis W_{mi}. These functions are zero on the boundary and take the form (Jackson, 1975, etc.)

$$W_m(k\mathbf{r}/a) = \begin{cases} \frac{J_m(k\mathbf{r}/a)}{a\sqrt{2\pi N_m(k)}} e^{im\theta} & \text{for } r \leq a \\ 0 & \text{for } r > a \end{cases}$$

In which $N_m(k)$ is the normalization constant, and is equal to

$$N_m(k) = \frac{1}{2}[J_{m+1}(k)]^2.$$

From now on we will set for simplicity $a = 1$.

These W functions solve the differential equation,

$$\nabla^2 W_m = -k^2 W_m. \tag{A.2}$$

Since these functions are set to be zero on a circular boundary of radius 1, thus the solutions for k are quantized, which take discrete values k_i^m where i denotes the ith zero of the function, i.e $W_m(k_i^m) = 0$. Since we can rewrite $J_l(k_n^0 r)e^{il\theta} = \sum_j \varepsilon_{lnj} W_l(k_j^l \mathbf{r})$, where the value of the coefficients is the value of the coefficients is

$$\begin{aligned} \varepsilon_{lnj} &= \int J_l(k_n^0 r)e^{il\theta} W_l(k_j^l \mathbf{r}) d^2 r \\ &= \begin{cases} \sqrt{2\pi N_0(k_i^0)}\delta_{jn} & \text{for } l = 0 \\ \sqrt{\frac{2\pi}{N_l(k_j^l)}} \left(\frac{k_j^l}{(k_n^0)^2 - (k_j^l)^2} \right) J_l(k_n^0) J_{l+1}(k_j^l) & \text{for } l > 0 \end{cases} \end{aligned}$$

we can obtain a representation of the correlation function in terms of the Bessel-Fourier set,

$$C(\mathbf{r} - \mathbf{r}') = c \left[\sum_{ljj'} [C(l)]_{jj'} W_l(k_j^l \mathbf{r}) W_l^*(k_{j'}^l \mathbf{r}') \right],$$

where

$$[C(l)]_{jj'} = \sum_i \frac{\varepsilon_{lij}\varepsilon_{lij'}}{2\pi N_0(k_i^0)\ (k_i^0)^2}.$$

This representation of $C(\mathbf{r} - \mathbf{r}')$ is unique only up to a constant b, which adds the term $\delta_{l0} V_j V_{j'}$ to the representation of the correlation matrix, where $V_j = \int W_0(k_j^0 \mathbf{r}) d^2 r = 2\sqrt{\pi}/k_j^0$.

Thus plugging this form of the correlation function into the eigen-equation 2.2, representing the eigen-functions as a sum of this complete set, or, $\psi(\mathbf{r}) = \sum_{mn} A_{mn} W_m(k_n^m \mathbf{r})$, and using the orthogonality relations, we obtain the matrix equation,

$$\sum_j [C(l)]_{ij} A_{lj} = \lambda A_{li}.$$

In this appendix we have set the boundary $a = 1$, in order to take a into account the simple transformation $\mathbf{r} \rightarrow \mathbf{r}/a$ needs to be performed to the representation of the correlation function. It is important to note that the value of the constant b is dependent on the value of a.

Including a non zero b results in

$$[C(l)]_{jj'} = \sum_i \frac{\varepsilon_{lij}\varepsilon_{lij'}}{2\pi N_0(k_i^0)\ (k_i^0)^2}. + b\delta_{l0}V_jV_{j'} \tag{A.3}$$

where

$$\begin{aligned} \varepsilon_{lnj} &= \int J_l(k_n^0 r)e^{il\theta}W_l(k_j^l\mathbf{r})d^2r \\ &= \begin{cases} \sqrt{2\pi N_0(k_i^0)}\delta_{jn} & \text{for } l = 0 \\ \sqrt{\frac{2\pi}{N_l(k_j^l)}}\left(\frac{k_j^l}{(k_n^0)^2-(k_j^l)^2}\right)J_l(k_n^0)J_{l+1}(k_j^l) & \text{for } l > 0 \end{cases} \end{aligned}$$

and

$$V_j = \frac{2\sqrt{\pi}}{k_j^0}.$$

and N_m is the normalization constant, which takes the value $N_m(k) = (\frac{1}{2})[J_{m+1}(k)]^2$.

A.2 Finding the correlation function of the preprocessed images.

The New correlation function

$$C^p(\mathbf{r}-\mathbf{r}') = \int d^2x \int d^2y K(\mathbf{r}-x)C(x-y)K(y-\mathbf{r}')$$

Using the addition formula for Bessel functions we get that

$$K(\mathbf{r}-\mathbf{r}') = \frac{1}{g^2}\sum_i \alpha_i J_0(k_i^0(\mathbf{r}-\mathbf{r}')) = \frac{1}{g^2}\sum_i \alpha_i \sum_n J_n(k_i^0\mathbf{r})J_n(k_i^0\mathbf{r}')e^{in(\theta-\theta')}$$

Thus

$$\int W_l(k_j^l\mathbf{r}')K(\mathbf{r}-\mathbf{r}')d^2r = \frac{1}{g^2}\sum_i \alpha_i\varepsilon_{lij}J_l(k_i^0 r) = \frac{1}{g^2}\sum_{in}\alpha_i\varepsilon_{lij}\varepsilon_{lin}W_l(k_n^l\mathbf{r})$$

and therefore

$$C^p(\mathbf{r}-\mathbf{r}') = \sum_{lmn}[C^p(l)]_{mn}\,W_l(k_m^l\mathbf{r})W_l(k_n^l\mathbf{r}')$$

where

$$[C^p(l)]_{nm} = \frac{1}{g^4}\sum_{jj'ip}[C(l)]_{jj'}\alpha_i\alpha_p\varepsilon_{lij}\varepsilon_{lpj'}\varepsilon_{lin}\varepsilon_{lpm}. \tag{A.4}$$

Appendix B

Appendix

B.1 BCM : Fixed Points in a Symmetric Distributions

Input vectors $\mathbf{x}$ with bias bit $x_0 = 1$, weight vectors $\mathbf{m}$ with bias bit m_0. If the distribution is symmetric then for each input vector $\mathbf{x}$ there exists another $-\mathbf{x}$ with the same probability. I will now look at the dynamics averaged over all input vectors, assuming that θ_m is a constant over this time scale. The key to this method, is splitting up the distribution of input vectors into the two symmetric sub populations each with P input vectors.

$$\begin{aligned} < \dot{m}_0 > &= \frac{1}{2P}\sum_{i=1}^{P}[\Phi(m_0 + \mathbf{x}^i \cdot \mathbf{m}, \theta_m) + \Phi(m_0 - \mathbf{x}^i \cdot \mathbf{m}, \theta_m)] \\ < \dot{\mathbf{m}}_j > &= \frac{1}{2P}\sum_{i=1}^{P}\mathbf{x}_j^i[\Phi(m_0 + \mathbf{x}^i \cdot \mathbf{m}, \theta_m) - \Phi(m_0 - \mathbf{x}^i \cdot \mathbf{m}, \theta_m)] \end{aligned}$$

Using the standard quadratic Φ function this takes the form,

$$\begin{aligned} < \dot{m}_0 > &= \frac{1}{2P}\sum_{i=1}^{P}[(m_0 + \mathbf{x}^i \cdot \mathbf{m})(m_0 + \mathbf{x}^i \cdot \mathbf{m} - \theta_m) + (m_0 - \mathbf{x}^i \cdot \mathbf{m})(m_0 - \mathbf{x}^i \cdot \mathbf{m} - \theta_m)] \\ < \dot{\mathbf{m}}_j > &= \frac{1}{2P}\sum_{i=1}^{P}\mathbf{x}_j^i[(m_0 + \mathbf{x}^i \cdot \mathbf{m})(\mathbf{m}_0 + \mathbf{x}^i \cdot \mathbf{m} - \theta_m) - (m_0 - \mathbf{x}^i \cdot \mathbf{m})(m_0 - \mathbf{x}^i \cdot \mathbf{m} - \theta_m)] \end{aligned}$$

$$\begin{aligned} < \dot{m}_0 > &= m_0^2 - m_0\theta_m + < (x \cdot \mathbf{m})^2 > = \theta_m(1 - m_0) \\ < \dot{\mathbf{m}}_j > &= (2m_0 - \theta_m)m_k C_{kj} \end{aligned} \tag{B.1}$$

Where $C_{kj} = (1/P)\sum_i x_k^i x_j^i$ is the correlation matrix, and the $< >$ denote an average over the input vectors. These equations have two types of fixed points.

$$< \dot{m}_j >= 0 \quad If \quad \begin{cases} \theta_m = 2m_0 \\ Cm = 0 \end{cases} \tag{B.2}$$

In the simplest case when there is no bias term, $\mathbf{m}_0 = 0$ and then the only fixed point is the trivial fixed point $Cm = 0$, i.e: the synaptic vector is orthogonal to all inputs or is zero.

We will now only look at the more complex situation in which there is a bias bit. The case where $\theta_m = 2m_0$ is the equilibrium point of interest, since it is the one found in simulations, since the trivial point is found to be unstable. For this case eq. 1 takes the form,

$$m_0^2 = \frac{1}{P}\sum_{i=1}^{P}(\mathbf{x}^i \cdot \mathbf{m})^2 = \frac{1}{P}\sum_{i=1}^{P}(S^i - m_0)^2$$

Where $S^i = m_0 + \mathbf{x}^i \cdot \mathbf{m}$. Using the identities $\theta_m =< {S^i}^2 >$, and $\theta_m = 2m_0$. we get that $1/P \sum S^i =< S >= 1$, but since the distribution is symmetric, $< S >= m_0$. Therefore

$$m_0 = 1 \qquad \theta_m = 2 \tag{B.3}$$

The other condition $C\mathbf{m} = 0$ generally implies that $\mathbf{m}$ is the Null vector, unless $det(C) = 0$ which is the non generic case.

When $\mathbf{m} = 0$ using the same method as before, we obtain that

$$m_0 = 1 \qquad \theta_m = 1 \tag{B.4}$$

B.1.1 Stability of the fixed points

By linearizing the dynamic equations around the fixed points, an equation of the form $\dot{M} = A \cdot M$, is produced, where M is the vector $(m_0, \mathbf{m})$. A fixed points is stable if all the eigen valuesλ_i of the matrix A are such that $|\lambda_i| < 0$. The form of the matrix A is

$$A = \begin{pmatrix} R - 2{m_0}^2 & 2(1-m_0)m_j C_{1j} & \dots & 2(1-m_0)m_j C_{Nj} \\ 2(1-m_0)m_j C_{j1} & RC_{11} - 2m_k C_{1k} m_l C_{1l} & \dots & RC_{1N} - 2m_k C_{1k} m_l C_{Nl} \\ \vdots & \vdots & \ddots & \vdots \\ 2(1-m_0)m_j C_{jN} & RC_{N1} - 2m_k C_{Nk} m_l C_{1l} & \dots & RC_{NN} - 2m_k C_{Nk} m_l C_{Nl} \end{pmatrix} . \tag{B.5}$$

where $R = (2m_0 - \theta_m)$

For the trivial case $m = 0$, the stability matrix has the form,

$$A = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & C_{11} & \dots & C_{1N} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & C_{N1} & \dots & C_{NN} \end{pmatrix} . \tag{B.6}$$

Thus for this case stable points exist only if all eigen-values of the correlation matrix are negative. For a correlation type matrix this can not happen since $TR(C) = \sum C_{ii} = \sum \lambda_i > 0$, therefore at least some of the eigen-values are positive.

For the other case,

$$A = \begin{pmatrix} -1 & 0 & \dots & 0 \\ 0 & -{g_1}^2 & \dots & -g_1 g_N \\ \vdots & \vdots & \ddots & \vdots \\ 0 & -g_N g_1 & \dots & -{g_N}^2 \end{pmatrix}. \tag{B.7}$$

where $g_i = -m_k C_{ik}$. The matrix D is a matrix whose elements are $D_{ij} = -g_i g_j$. The stable fixed points are such that D has only negative eigen-values.

A two dimensional example: This will be illustrated with the case of two symmetric clusters of variance σ located at $x_0,\ y_0$ and at $-x_0,\ -y_0$. The correlation matrix thus has the form

$$C = \begin{pmatrix} {x_0}^2 + \sigma^2 & x_0 y_0 \\ x_0 y_0 & {y_0}^2 + \sigma^2 \end{pmatrix} \tag{B.8}$$

It will be useful to note that the eigen-vectors and eigen-values are $d_1 \propto (y_0, -x_0),\ \lambda_1 = \sigma^2$ and $d_2 \propto (x_0, y_0),\ \lambda_2 = \sigma^2 + {x_0}^2 + {y_0}^2$

When written in the eigen-vector of C base i.e: when $\mathbf{m} = \sum a_j \mathbf{d}_j$, the matrix elements of D take the form, $D_{ij} = -g_i g_j$ where $g_i = \lambda_l a_l d_{li}$. The determinant of such a matrix is 0, therefore the stability matrix has two eigen-values $\mu_1 = 0$ and $\mu_2 = TR(D) = -({g_1}^2 + {g_2}^2) < 0$. Thus all fixed points will be stable in the one direction, and marginally stable in the other. This implies that the is a closed curve connecting all sable points.

In the general case a similar point can be made using the objective function formulation, Since

$$R_m = -\mu\{\frac{1}{3}E(X \cdot M)^3 - \frac{1}{4}E^2(X \cdot M)^2\},$$

Where X and M are the input and weight vectors which include the bias bit m_0 and the $x_0 = 1$ dummy input bit. Since we know that for fixed points $m_0 = 1$ and $E(x \cdot m)^2 = 1$ then

$$\begin{aligned} R_m &= -\frac{1}{3}E[1 + 3x \cdot m + 3(x \cdot m)^2 + (x \cdot m)^3] - \frac{1}{4}E^2[x \cdot m]^2 \\ &= -\frac{1}{3}E[1 + 3(x \cdot m)^2] - \frac{1}{4}E^2[x \cdot m]^2 = -\frac{13}{12}. \end{aligned} \tag{B.9}$$

i.e: for all non trivial fixed points in symmetric distributions, with linear neurons, the objective function has the same value.

B.1.2 Distributions Symmetric around $\mathbf{x}^0$

We will write $\mathbf{x} = \mathbf{x}^0 + \mathbf{x}'$, symmetry implies that $< x_i >=< x_i x_j x_k >= 0$ The dynamic eqatins for the bias bit take the form,

$$\begin{aligned} < \dot{m}_0 > &= < C(C-\theta) >=< C^2 >= \theta < C >= \theta(1- < C >) \\ &= \theta(1 - m_0 - m \cdot \mathbf{x}^0) \end{aligned} \tag{B.10}$$

Thus at the fixed point $m_0 + m \cdot \mathbf{x}^0 = 1$.

For the other weights,

$$\begin{aligned} < \dot{m}_j > &= < C(C-\theta)(x_j^0 + \mathbf{x}'_j) >= \underbrace{< C(C-\theta) >}_{0} x_j^0 + < C(C-\theta)x'_j > \\ &= < (m_0 + m \cdot \mathbf{x}^0 + m \cdot \mathbf{x}')(m_0 + m \cdot \mathbf{x}^0 + m \cdot \mathbf{x}' - \theta)x'_j > \\ &= \underbrace{m_0^2 < x'_j >}_{0} + \underbrace{(m \cdot \mathbf{x}^0)^2 < x'_j >}_{0} + \underbrace{< \mathbf{x}'_j (m \cdot \mathbf{x}')^2 >}_{0} \\ &\quad + \underbrace{2m_o m \cdot \mathbf{x}^0 < x'_j >}_{0} + 2m_0 < x'_j m \cdot \mathbf{x}' > + 2m \cdot \mathbf{x}^0 < x'_j m \cdot \mathbf{x}' > \\ &\quad -\theta \left(\underbrace{< x'_j m_0 >}_{0} + \underbrace{1 < x' m \cdot \mathbf{x}^0 >}_{0} + < x'_j m \cdot \mathbf{x}^p rime > \right) \end{aligned}$$

Denoting $< x'_i x'_j >= C_{ij}$, where C_{ij} is the correlations matrix of the shifted inputs, we obtain that,

$$< \dot{m}_j >= m_i C_{ij} \left[2(m_o + m \cdot \mathbf{x}^0) - \theta \right] = m_i C_{ij}(2 - \theta) \tag{B.11}$$

Thus at equilibrium,

$$\begin{aligned} & m_0 + m \cdot \mathbf{x}^0 = 1 \\ \textit{and} & \\ & \theta = 2 \end{aligned} \tag{B.12}$$

Bibliography

Abramowitz, M. and Stegun., I. A. (1972). *Handbook of mathematical functions.* Dover.

Albus, K. (1975). Predominance of monocularly driven cells in the projection area of the central visual field in cat's striate cortex. *Brain Research*, 89:341–347.

Atick, J. J., Li, Z., and Redlich, A. (1992). Understanding retinal color coding from first principlas. *Neural Computation*, 4:559–572.

Atick, J. J. and Redlich, A. (1990). Towards a theory of early visual processing. *Neural Computation*, 2:308–320.

Atick, J. J. and Redlich, A. (1992). What does the retina know about natural scenes. *Neural Computation*, 4:196–211.

Barlow, H. B. (1961). Possible principles undelying the transformation of sensory messeges. In Rosenblith, W., editor, *Sensory Comunication.*, pages 217–234. MIT Press, Cambridge MA.

Barlow, H. B. (1989). Unsupervised learning. *Neural Computation.*, 1:295–311.

Barlow, H. B. and Pettingrew, J. D. (1971). Lack of specificity in the visual cortex of yong kittens. *J. Physiol.*, 218:98–100.

Bialek, W., Ruderman, D. L., and Zee, A. (1991). Optimal sampeling of natural images: A design principle for the visual system? In R. Lippman, J. M. and Touretzkey, D., editors, *Neural Information Processing 3*, pages 363–369. Morgan Kaufmann.

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. *Journal Neuroscience*, 2:32–48.

Blakemore, C. and Van-Sluyters, R. R. (1975). Innate and environmental factors in the development of the kittens visual cortex. *J. Physiol.*, 248:663–716.

Blasdel, G. G. (1992). Orientation selectivity,preference, and continuity in monkey striate cortex. *The Journal of Neuroscience*, 12(8):3139–3161.

Bonhoeffer, T. and Gringald, A. (1991). Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, 353:429–431.

Braastad, B. O. and Heggelund, P. (1985). Development of spatial receptive-field organization and orientation selectivity in kitten's visual cortex. *J. Neurophysiol.*, 53:1158–1178.

Campbell, F. W. and Robson, J. G. (1968). Application of fourier analysis to the visibility of gratings. *J. Physiol (Lond)*, 197:551–566.

Chow, K. and Spear, P. (1974). Morphological and functional effects of deprivation on the rabbit visual cortex. *Expl Neurol*, 42:429–447.

Clothiaux, E. E., Cooper, L. N., and Bear, M. F. (1991). Synaptic plasticity in visual cortex: Comparison of theory with experiment. *Journal of Neurophysiology*, 66:1785–1804.

Cooper, L. N., Liberman, F., and Oja, E. (1979). A theory for the acquisition and loss of neurons specificity in visual cortex. *Biol. Cyb.*, 33:9–28.

Coren, S., Porac, C., and Ward, L. M. (1984). *Sensation and Perception.* Harnacort Brace Janovich.

Cowan, J. D. and Friedman, A. E. (1991). Simple spin models for the development of ocular dominance columns and iso-orientation patches. In *Advances in Neural Information Processing Systems 3.*

Dayan, P. and Goodhill, G. (1992). Perturbing hebbian rules. In *Advances in Neural Information Processing Systems 4.*

Diaconis, P. and Freedman, D. (1984). Asymptotics of graphical projection pursuit. *Annals of Statistics*, 82:249:266.

Erwin, E., Obermayer, K., and Schulten, K. (1994). A simple network model of unsupervised learning. Technical report, University of Illinois.

Field, D. J. (1987). Ralations between the statistics of natural images and the response properities of cortical cells. *J. Opt. Soc. Am. A*, 4(12):2379–2394.

Foldia, P. (1989). Adaptive network for optimal linear feature extraction. *In Proc. IJCNN, Washington D.C.*, pages 401–406.

Hancock, P. J., Baddeley, R. J., and Smith, L. S. (1992). The principal components of natural images. *Network*, 3:61–70.

Hebb, D. O. (1949). *The Organization of Behavior*. Wiley.

Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurons in the cat striate cortex. *J. Physiol. (London)*, 148:509–591.

Hubel, D. H. and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol*, 160:106–154.

Hubel, D. H. and Wiesel, T. N. (1963a). Binocular interactions in striate cortex of kittens reared with artificial squint. *J. Neurophysiol.*, 28:1041–1059.

Hubel, D. H. and Wiesel, T. N. (1963b). Receptive fields of cells in stiate cortex of very yong, visually inexperienced kittens. *J. Neurophysiol.*, 28:1041–1059.

Imbert, M. and Buisseret, P. (1975). Receptive field charectaristics and plastic properties of visual cortical cells in kittens reared with or without visual experiance. *Exp. Brain Res.*, 22:25–36.

Intrator, N., Bear, M. F., Cooper, L. N., and Paradiso, M. A. (1993). Theory of synaptic plasticity in visual cortex. In Thomppson, R., Baudry, M., and Davis, J., editors, *Synaptic Plasticity: Molecular, Cellular and Functional Aspects*. MIT Press, Cambridge, MA.

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.

Jackson, J. D. (1975). *Classical Electrodynamics*. Wiley.

Jones, J. P. and Palmer, L. A. (1987). The two-dimensional spatial structure of simple receptive fields in cat striate cortex. *jnp*, 58(6):1187–1258.

Kohonen, T. (1982). Self-organization of topologically correct feature maps. *Biological Cybernetics*, 43:59–69.

Kosterlitz, M. J. and Thouless, D. J. (73). Ordering, metastability and phase transitions in two-dimensional systerms. *J. Phys.*, C6:1181.

Law, C. C. and Cooper, L. N. (1994). Formation of receptive fields according to the bcm theory of synaptic plasticity in realistic visual environment. *PNAS*, 91:7797–7801.

Levay, S., Hubel, D., and Weisel, T. (78). *Journal of Comparative Neurology*, 159:559–576.

Li, Z. and Atick, J. (1994). Towards a theory of the striate cortex. *Neural Computation*, 6:127–146.

Linsker, R. (1986). From basic network principles to neural architecture. *Proc. Natl. Acad. Sci. USA*, 83:7508–12, 8390–4, 8779–83.

Linsker, R. (88). Self-organization in a perceptual network. *Computer*, March 1988:105–117.

Liu, Y. (1994). *Synaptic Plasticity: From Single Cell to Cortical Network*. PhD thesis, Brown University.

Liu, Y. and Shouval, H. (1994). Localized principal components of natural images - an analytic solution. Network., 5.2.

Liu, Y., Shouval, H., and Cooper, L. N. (1994). Formation of receptive fields of single neurons with hebbian type learning in a correlated noise environment. *submitted to* NIPS *94*, 7.

MacKay, D. J. and Miller, K. D. (1990). Analysis of linsker's simulations of hebbian rules to linear networks. *Network*, 1:257–297.

Miller, K. D. (1990). Derivation of linear hebbian equations from non linear models of synaptic plasticity. *Neural Computation*, 2:321–333.

Miller, K. D. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between on- and off-center inputs. *J. of Neurosci.*, 14.

Miller, K. D., Keller, J. B., and Striker, M. P. (1989). Ocular dominance column development: Analysis and simulation. *Science*, 245:605–615.

Nass, M. N. and Cooper, L. N. (1975). A theory for the development of feature detecting cells in visual cortex. *Biol. Cyb.*, 19:1–18.

Neibur, E. and Wögötter (1994). Design principles of columnar organization in visual cortex. *Neural Computation*, 6(4):602–615.

Obermayer, K., Schulten, K., and Blasdel, G. (1992). Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps. *Phys. Rev.*, 45:7568–7589.

Oja, E. (1982). A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15:267–273.

Oja, E. (1992). Principal components, minor components, and linear neural networks. *Neural Networks*, 5:927–935.

Orban., G. A. (1984). *Neuronal Operations in the Visual Cortex*. Springer Verlag.

Rochester, N., Holland, J., Haibt, L., and Duda, W. (1956). Tests on a cell assembly theory of the action of the brain, using a large scale digital computer. *IRE Transactions of Information Theory*, IT-2:80–93.

Rojer, A. S. and Schwartz, E. L. (1990). Cat and monkey cortical columnar patterns modeled by bandpass-filtered white noise. *Biological Cybernetics*, 62:381–391.

Rubinstien, J. (1994). Pattern formation in associative neural networks with weak lateral interaction. *Biological Cybernetics*, 70.

Rubner, J. and Tavan, P. (1989). A self-organizing network for principal component analysis. *Europhysics Letters*, 10(7):693–698.

Ruderman, D. L. and Bialek., W. (1994). Statistics of natural images: Scaling in the woods. In J. D. Cowan, G. T. and Alspector, J., editors, *Advances in Neural Information Processing Systems 6*.

Sanger, T. D. (1989). Optimal unsupervised learning in a single-layer linear feedforward neural network. *Neural Networks*, 2:459–473.

Scofield, C. L. and Cooper, L. N. (1985). Developmental properties of neural networks. *Contemp. Phys.*, 26(2):125–145.

Shouval, H. and Liu, Y. (1994). The effect of retinal preprocessing on the receptive field of a stabilized hebbian neuron. In *Proceedings of the CNS*94 confrence*. Kluwer. [To Appear].

Sirosh, J. and Miikkulainen, R. (1994). Cooperative self-organization of afferent and lateral connections in cortical maps. *Biological Cybernetics*, 71:66–78.

Stent, G. (1973). A physiological mechnism for hebb's postulate of learning. *Proceedings of the National Academy of Sciences, U.S.A*, 70:997.

Swindale, N. (1982). A model for the formation of orientation columns. *Proc. R. Soc. London*, B208:243.

Swindale, N. (1991). Covarage, and the design of striate cortex. *Biological Cybernetics*, 65:415:424.

van Hateren, J. H. (1992). Theoretical predictions of spatiotemporal receptive fields of fly [lmcs], and experimantal validation. *J Comp Physiol A*, 171:157–170.

van Hateren, J. H. (1993). Three modes of spationtemporal processing in eyes. *J Comp Physiol A*, 172:583–591.

von der Malsburg, C. (1973). Self-organization of orientation sensitive cells in striata cortex. *Kybernetik*, 14:85–100.