Correlations, Competition, and Optimality: Modelling the Development of Topography and Ocular Dominance

Geoffrey Goodhill

April 1992

Cognitive Science Research Paper
Serial No. CSRP 226
The University of Sussex
School of Cognitive and Computing Sciences
Falmer
Brighton BN1 9QH
GREAT BRITAIN

Abstract

There is strong biological evidence that the same mechanisms underly the formation of both topography and ocular dominance in the visual system. However, previous computational models of visual development do not satisfactorily address both of these phenomena simultaneously. In this thesis we discuss in detail several models of visual development, focussing particularly on the form of correlations within and between eyes.

Firstly, we analyse the "correlational" model for ocular dominance development recently proposed in [Miller et al 1989]. This model was originally presented for the case of identical correlations within each eye and zero correlations between the eyes. We relax these assumptions by introducing perturbative correlations within and between eyes, and show that (a) the system is unstable to non-identical perturbations in each eye, and (b) the addition of small positive correlations between the eyes, or small negative correlations within an eye, can cause binocular solutions to be favoured over monocular solutions.

Secondly, we extend the elastic net model of [Goodhill 1988, Goodhill & Willshaw 1990] for the development of topography and ocular dominance, in particular considering its behaviour in the two-dimensional case. We give both qualitative and quantitative comparisons with the performance of an algorithm based on the self-organizing feature map of Kohonen, and show that in general the elastic net performs better. In addition we show that (a) both algorithms can reproduce the effects of monocular deprivation, and (b) that a global orientation for ocular dominance stripes in the elastic net case can be produced by anisotropic boundary conditions in the cortex.

Thirdly, we introduce a new model that accounts for the development of topography and ocular dominance when distributed patterns of activity are presented simultaneously in both eyes, with significant correlations both within and between eyes. We show that stripe width in this model can be influenced by two factors: the extent of lateral interactions in the postsynaptic sheet, and the degree to which the two eyes are correlated. An important aspect of this model is the form of the normalization rule to limit synaptic strengths: we analyse this for a simple case.

The principal conclusions of this work are as follows:

- It is possible to formulate computational models that account for (a) both topography and stripe formation, and (b) ocular dominance segregation in the presence of *positive* correlations between the two eyes.
- Correlations can be used as a "currency" with which to compare locality within an eye with correspondence between eyes. This leads to the novel prediction that stripe width can be influenced by the degree of correlation between the two eyes.

Acknowledgments

I would firstly like to thank my supervisor Harry Barrow, for his help, guidance, and the large number of ideas he contributed which helped greatly to shape and direct the work described here.

I am also deeply indebted to both David Willshaw and Peter Dayan. David was a great source of inspiration and encouragement to me throughout the thesis, and I am particularly grateful to him for his hospitality during my occasional visits to Edinburgh. Peter is a very special person who combines both vast technical expertise and deep insight with great generosity and good-naturedness. His contribution to seeing the project through, especially in the closing stages, is immesurable. Chapter 4 of this thesis describes work that was carried out jointly with Peter.

I was also extremely fortunate to have occasional but important feedback on some specific aspects of the work from Geoff Hinton and Christopher Longuet-Higgins. Geoff has helped me in many ways both before and during this project, for which I am deeply grateful. I greatly appreciate both the trouble Christopher has taken to try to instil in me academic standards as rigourous as his own, and our very enjoyable music-making. In addition, I am grateful to him for pointing me in the direction of perturbation analysis in Chapter 4.

Ken Miller has provided help in many ways to this project, for which I am grateful. He willingly answered my many questions about his model, provided kind hospitality in San Francisco, and in particular emphasised to me the properties of subtractive normalization. I also much appreciated occasional conversations with Jack Cowan and Klaus Obermayer, and extremely useful feedback on the thesis from my examiners, Graeme Mitchison and David Young.

Many people at Sussex have helped me in many ways: in particular I am grateful for the useful conversations I've had with Julian Budd, Dave Cliff, Antonio Da Silva Costa, Inman Harvey, Jim Stone, and particularly Alistair Bray, whose patience in helping me with various matters went far beyond what could be reasonably expected. I also thank Richard Dallaway for much practical assistance with LaTeX. More recently in Edinburgh I have gained much from discussions of biological matters with David Price, and also Richard Ribchester.

Deep thanks are due to all those who provided useful comments on earlier drafts of this thesis: David Willshaw, Peter Dayan and Harry Barrow, who greatly improved several versions, and also Inman Harvey and Antonio Da Silva Costa. Any errors that remain are, of course, my own.

I would also like to mention David Griffel, both for his excellent teaching and kind words while I was a mathematics undergraduate, and more recently for some insightful feedback on particular aspects of this work.

I wish to express my very great thanks to Isabelle Desjeux, both for various practical assistance, but more importantly for providing much-appreciated emotional support in some of my weakest moments.

I am grateful to the SERC and SD-Scicon for their financial support, and to the Nuffield Foundation, from whom a Science Travel Grant supported a useful period of study in Edinburgh.

Contents

Chapter 1

Introduction

1.1 Development of brain connections

Brains are mostly wire. The human brain contains about 10^{11} neurons, but roughly 1000-10,000 times as many connections between them. These connections vary in length between very short range (less than a millimetre) linking small patches of brain, to quite long range (several centimetres or more) connecting different regions of the brain, or connecting sensory structures to more central brain structures.

There is strong evidence that to a large extent these connections are "specified" during ontogenesis: fibres growing from a cell are very specific for which other cells they connect to, both in terms of general regions and particular areas within a region. It is clear that, if the connection matrix of the brain were random, the amount of information available in DNA would be grossly insufficient to specify each of these connections explicitly.

However, the overwhelming biological evidence is that connection patterns in the brain are *not* random. A number of particular types of pattern occur frequently throughout the brain, the general properties of which are remarkably similar over a wide range of animal species. Patterns, even if made up of millions of connections, can be specified using a much smaller amount of information than if these connections were random. It thus becomes feasible to imagine that connection strengths are specified in DNA, but in the form of rules that generate patterns of connections rather than an explicit connection matrix (for discussion see [Gierer 1988]). One part of understanding the development of brain connections is therefore to try to discover these rules. Note that we are not claiming that these rules necessarily exist for the purpose of generating patterns: Pattern-formation is simply one of their outcomes in the biological system.

We must also consider the following. Although in general outline connection patterns in the brains of two individuals of the same species are very similar, on a fine scale there can be large variations between them. This is true even for genetically identical individuals (see e.g. [Changeux & Danchin 1976]). Thus besides finding the rules that generate overall patterns, we also have to account for the individual variation in a way that does not imply large amounts of genetic specification. Part of this variation may be due to entirely random processes. However, the experimental data shows that the detailed patterns of connections between sensory structures and more central brain structures can be systematically influenced by the *environment* in which that particular pattern develops. Different environments, both in terms of sense data from the outside world and the biochemical environment inside the brain itself, can lead to differences in the detailed pattern of connections (see e.g. [Blakemore 1978]). It is clear that the outcome of the rules that generate patterns is bound up with the context in which those rules operate (see e.g. [Changeux & Danchin 1976]).

In this thesis we examine a number of different models for the development of connections in the visual system, and in particular study the way in which the outcome of these models is influenced by the form of their inputs.

1.1.1 The importance of development

Understanding of the development of the brain may also provide insight into how *learning* takes place. Rules that account for developmental processes may also have relevance to learning: for instance, issues of normalization may play a similar role in both (for further discussion see [Gierer 1988]). One of the advantages of studying development is that it involves substantial changes in connectivity and the response properties of individual cells that can be observed using standard neuroanatomical and neurophysiological techniques. This is in contrast to the changes that take place during learning: these can be hard to measure, especially given that the overall properties of a network of neurons can be altered significantly by the cumulative effects of just very small changes at a large number of synapses. Thus there is the possibility of more biological information being available to constrain models of development than of learning.

1.2 Topographic maps and striped projections

We now focus on one particular type of connection in the brain: that between individual structures such as lateral geniculate nucleus and V1, or V1 and V2. One of the most prevalent features of these connection patterns is that they are very often arranged in the form of a *topographic map*. We will return later to the issue of a precise mathematical definition of this term: for the present, we take it to mean that neighbouring points in the first structure are connected to neighbouring points in the second. For instance, for mappings from sensory structures this means that a small movement on the sensory surface corresponds to a small movement in the region of brain representing that sensory surface. There are a very large number of such maps in the brain: for example over 200 in the macaque monkey, where in some cases the entire visual field is represented in an ordered manner across a surface 1mm square (M. Sereno, Personal Communication). In general we will refer

to the first structure (e.g. the eye) as the input structure, or input space if we are considering more general mappings (see below), and the second structure as the output or target structure.

Many regions of brain that receive such connections have a columnar structure: moving in a direction orthogonal to the surface, cells with similar response properties are encountered. These regions can almost always be regarded as two-dimensional from the point of view of understanding the abstract properties of the map. This is relatively unproblematic for "spatial" maps, where the mapping is from some sensory surface or other brain surface that is also two-dimensional. However, it can be the case that variables other than, or in addition to, position are represented on the target surface: so called "computational maps" [Knudsen et al 1987]. It is a general property of such maps in the brain that small movements in the input space again lead to small movements on the target surface, but now with some discontinuous jumps that are inevitably produced by mapping to a space of lower dimensionality. The simplest case of this is where one variable additional to position is represented, and this variable can take only two discrete values: in other words, a mapping from two and (exactly) a bit dimensions to two dimensions. In cases such as these in the brain, the map is often laid out in a series of alternating regions, each representing one of the two values of the binary variable. These maps are commonly "striped": that is, the alternating regions are a series of bands which are much longer in one direction than the other. This is well described for the case of sensory inputs from the monkey hand in [Constantine-Paton & Law 1982]:

> The bands separate sensory inputs from the hand according to the kinds of information the input carries. In some bands the neurons respond only to the onset of a touch; in the intervening bands the neurons have a more prolonged response. It is as if the map of the hand in the somatic sensory cortex of the monkey has been constructed by alternating stripes cut out of two distinct maps of the hand.

The representation of position in such maps commonly has two important properties. Firstly, the map of each value of the binary variable (i.e. the map that would be formed if all the stripes representing the other value were removed), is topographic in the above sense. Secondly, these maps are "aligned": the two values of the binary variable at each position on the input surface are mapped to nearby positions in the target structure, given the constraint that they are represented in different stripes.

1.3 Topography and ocular dominance in the visual system

In the course of vertebrate ontogenesis, the ganglion cells of the retina send axons through the eye stalk to the brain, where they project to cells of central structures. In mammals the retinae project to the lateral geniculate nucleus (LGN), which then project to primary visual cortex. In lower vertebrates such as amphibians and fish there are direct projections from retina to optic tectum. In both cases, the mapping is topographic, or as sometimes referred to in this context, "retinotopic". However, in mammals the primary visual cortex is naturally binocularly innervated, i.e. the same part of brain receives inputs from both eyes. In this case, in the early stages of development, the retinotopic projections from the two eyes to the common target structure are uniformly intermingled. As development proceeds, each part of the target gradually becomes more densely innervated by one eye and less densely innervated by the other. Eventually, a striped pattern of innervation can be observed that is reminiscent of the pattern of zebra stripes: so-called "ocular dominance stripes" (see e.g. [Hubel & Wiesel 1977]). The extent to which this happens varies between species. In animals such as cats and monkeys segregation is almost total, and the thickness of these stripes is approximately 400μm. Ocular dominance stripes are also present in humans, with width approximately 500μm to 1000μm [Horton et al 1990]. In lower vertebrates, similar striped projections do not occur naturally but can be induced experimentally (see e.g. [Constantine-Paton & Law 1978]). One very important feature of all these mappings is that they are *plastic*: their detailed form is influenced by the visual experience of the animal, and other procedures which alter the internal environment of the brain.

In one sense there is difference between the case of the striped mapping from the hand described earlier and ocular dominance stripes. In the former case, both values of the binary variable exist at the same spatial location in the input space, whereas in the latter the two values exist at different spatial locations. However, as we argue further later, the crucial point is that, even though cells at similar locations in the two eyes are spatially separated, they will be *correlated*, and it is this fact which defines the correspondance between the two eyes, rather than spatial relationships.

1.3.1 Why study this phenomenon?

There are a number of reasons why topography and ocular dominance are useful developmental phenomena to study for both neuroscientists and theoreticians. Firstly, the visual system is perhaps the most extensively researched and well-understood part of the brain. The form of the map from the eye (via the LGN) to the brain is a macroscopic property which develops as a result of microscopic changes in synaptic strengths. It therefore provides a way of magnifying alterations to various parameters of the system so that their effects can be clearly seen. In addition, ocular dominance is a relatively straightforward property of neurons in the visual system to measure: it is comparatively easy to determine by anatomical and physiological techniques to what extent cells have become monocular. Alterations in the parameters can be seen to have a clear effect on the monocularity of cells and on the global properties of stripe morphology: for instance, stripe width. For these reasons (among others - see chapter 2) there exists a substantial amount of experimental data.

The motivation for studying both phenomena together comes largely from experiments on frogs and goldfish performed by Constantine-Paton and others. Here, stripes were produced in normally monocular tectum by procedures such as diversion of both optic nerves into the same tectum, and implant of a third eye (reviewed in chapter 2). The importance of these experiments is summed up in [Constantine-Paton & Law 1982]:

> Questions about the functional or evolutionary significance of stripes in the brain of three-eyed frogs are clearly irrelevant. The third eye is abnormal, and in the absence of substantial input from both of the eyes to a single optic tectum the normal frog would get no benefit from a mechanism that evolved specifically to segregate tectal inputs into stripes. On the other hand, the survival of free-living frogs, and in particular their ability to catch the insects on which they feed, depends critically on a robust mechanism to ensure that a precise map of the contralateral retinal surface develops in each tectum.[1] We began, therefore, to consider the possibility that stripes might arise from the same developmental mechanism that generates maps.

The hypothesis that the same developmental mechanisms underly both topography and ocular dominance stripes is the main motivation that drives this thesis. As we will see, there are very many models that address either the formation of topographic maps *or* of ocular dominance stripes, but very few that satisfactorily address both. Considering both simultaneously allows a wider range of biological knowledge to be applied to constrain models.

1.4 Levels, models and mathematics

1.4.1 Levels

[Marr 1982] identified three levels of analysis for understanding "....any machine carrying out an information-processing task": in our case, the brain. The highest of these he called the computational level, which is the abstract definition of the problem: the goal of the computation, and why it is appropriate. The second is the algorithmic level, at which an abstract procedure is defined that can solve the problem as formulated at the computational level. Lastly, there is the implementational level, where the algorithm is actually carried out in some piece of hardware: at this level we might have for instance a computer program, or a demonstration of how the algorithm could be realised in a certain piece of brain tissue. Working down through these levels has sometimes been taken as a methodology for research: *first* formulate the problem, *then* find an algorithm, *then* show how it can be implemented.

This thesis does not strictly follow this "top-down" approach. We present algorithms without necessarily defining a computational problem, and use certain implementation issues to constrain theorizing at the algorithmic level. Our reasons for this are as follows [Churchland & Sejnowski 1988, Sejnowski et al 1988, Sejnowski & Tesauro 1989]:

- It is often difficult, and sometimes misleading, to attempt to define "the computational problem". Brains evolved rather than being designed by an engineer. It may be the case that a certain mechanism in the brain solves a number of computational problems, or that the phenomenon under investigation is a byproduct of mechanisms for solving a quite different computational problem from that which our intuitions might first suggest.

[1] The contralateral eye is on the opposite side of the brain from the tectum under consideration: the ipsilateral eye is on the same side.

- Even if we can accurately define a particular computational problem, to paraphrase [Churchland & Sejnowski 1988], "the space of algorithmic possibilities is consumately vast": there may be many possible algorithms that solve the problem. However, by considering the implementational level in conjunction with the algorithmic level, that is restricting attention to those algorithms which satisfy certain constraints arising from consideration of the hardware in which they are implemented (i.e. nervous tissue), one can proceed much more quickly to discovering the algorithm the brain actually employs. Furthermore, the implementational level can often provide inspiration to the algorithmic level.

- There may not be just one computational level and one algorithmic level. The brain contains many levels of structure, and "....it is likely that there is a corresponding multiplicity of algorithmic and computational levels as well" [Sejnowski & Tesauro 1989].

We now discuss the *types* of models presented in this thesis.

1.4.2 Models

[Sejnowski et al 1988] identify two types of brain models which they refer to as "realistic" and "simplifying": these can be seen as extremes on a continuum. A realistic model of a particular brain system attempts to take into account all that is known about the system, down to the level of the biophysics of individual cells. A simplifying model on the other hand tries to extract and work with only those properties of the system that are relevant to the phenomenon under investigation. [Sejnowski et al 1988] draw attention to a number of general problems with investigating realistic models, which also apply directly to the topography and ocular dominance problem. Firstly there is the issue that to formulate a realistic model with confidence requires the support of a huge amount of experimental data, which may not be available in sufficient detail to adequately constrain the parameters of the model. Secondly, the danger arises with realistic models that they may contain so many parameters and low-level details that very little higher-level understanding of the system can be obtained from the model. Thirdly, there is the practical problem that the computational requirements for simulating the vast detail of a realistic model may restrict consideration to an uninterestingly small fragment of the system.

The models discussed in this thesis tend, to differing degrees, towards the simplifying end of the spectrum. That is, they are formulated in terms of just a few key parameters that are assumed to be important in map formation. For instance, we study the effects of correlations in activity between cells in the retina or LGN, and the lateral connectivity in the cortex/tectum. These are simply not known at a detailed enough level in the natural system to be investigated in the context of a realistic model of map formation. We prefer therefore to examine the influence of general correlation and interaction parameters, abstracted away from cellular details. Our motivation is the hypothesis that these are the key parameters that determine the properties of the map, in particular the stripe width. Lastly, regarding the issue of computational resources, a general problem with studying map formation is that it is a property of many rather than just a few cells. Therefore it is not practically possible to devote large computational resources to simulating each individual cell.

1.4.3 Mathematics

What mathematical statements can be made about the topography and ocular dominance problem, and to what extent are these useful?

We note firstly that the term "topographic" does not have a well-defined mathematical meaning. It is used in the biological and theoretical literature pertaining to this problem to mean loosely that, in general, either points close together in the target structure are mapped to points close together in the input structure, or that points close together in the input structure map to points that are close together in the target structure (we return to issues of the difference between these two statements in chapter 5). Here "close together" is usually a purely qualitative measure, perhaps based on the visual appearance of the map as realised in some biological experiment or as a picture on a computer screen.

Although topography is not well-defined, unfortunately some authors appear to use this word interchangeably with "topology", which does have a precise meaning. A topological map, or homeomorphism, is a map that is one-to-one and bicontinuous (i.e. continuous in both directions) [Alexandroff 1961]. We argue that it may not be useful to apply such a definition to brain maps. Firstly, biological maps are virtually never one-to-one: each fibre from an input cell may contact thousands of cells in the target structure. Secondly, given a mapping between many thousands or even millions of cells in the brain, it would be most unlikely that there was not even one deviation from perfect ordering, and one deviation is all that is required for a map to fail to be

topological. Another (ab)use of terminology is to drop the word "topographic" altogether, and simply use the word "map" to mean topographic map (as in the quote from [Constantine-Paton & Law 1982] above), Here "map" is referring to a much more restricted class of entities than is defined by the mathematical meaning of the word.

It seems clear that what is needed is a measure of the *degree* of topography of a particular map, that is, the degree to which it approaches some ideal of perfect continuity (or bicontinuity). Indeed, this is the way in which the word is often used informally in the literature. However, these distinctions are again qualitative. A number of quantitative measures of topography have been proposed (see e.g. [Rankin & Cook 1986, Roe et al, 1990, Cottrell & Fort 1986, Érdi & Barna, 1984, Kohonen 1988]). Unfortunately each makes particular assumptions with respect to an individual system, and cannot be applied in all situations.

Another important reason for not using the mathematical concept of a topological map here is that the ocular dominance case involves mapping between spaces of different dimensionalities, and such maps (if they are assumed 1-1) cannot preserve all neighbourhoods of the higher dimensional space in the lower dimensional space. However, the topography and ocular dominance map still has some property in common with a "perfectly" topographic map as could exist between spaces of the same dimensionality, and we will tend, in common with most others, to refer to this case as being topographic with occasional discontinuities.

1.5 Mechanisms of map formation

1.5.1 Cooperation and competition

Two important concepts for understanding pattern formation in physical (and more particularly biological) systems are cooperation and competition. To explain these we will refer to an abstract system, initially disordered, which organizes itself due to the growth of various "modes" (patterns) of the system. *Cooperation* is the process whereby different modes can enhance each other's strength (ability to dominate in the final system). Balanced against this are often processes of *competition*. This has been defined by [Keddy 1989] as

> the negative effects which one mode[2] has upon another by consuming, or controlling access to, a resource that is limited in availability.

Here a resource is something required for the continued growth or maintenance of modes. Thus cooperation encourages mutual support between modes, while competition selects between modes. Both can be important for the development of interesting structure. Without cooperation, modes other than the very strongest have no influence on development, while without competition all modes can grow unrestrained, possibly obscuring the interesting structure contained within each.

Classic examples of physical systems which can be described by these concepts are the Rayleigh-Bénard instability, and the organization of crystal lattices and magnetic domains (for further discussion see [von der Malsburg & Singer 1988]). One of the first applications to biological systems was made by [Turing 1952], who formulated mathematically mechanisms by which an array of cells can become chemically differentiated from an almost uniform initial state. For other examples of biological interest see [Meinhardt 1982]. We now focus on the relevance of the concepts of cooperation and competition to the formation of brain maps, within the context of an extremely influential theory of the mechanism underlying changes of synaptic strength in the nervous system.

1.5.2 Hebbian rules

[Hebb 1949] made the following hypothesis regarding the circumstances under which the connection between two cells in the brain is increased in strength:

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

This has become known as the Hebb rule, and the term "Hebbian" is generally applied to any postulated synaptic learning rule that implements some version of this idea (see [Stent 1973] for discussion of a possible physiological implementation). Direct evidence for the existence of such a rule comes from, among others, various experiments in the visual system (reviewed in [Rauschecker 1991]).

[2] The word appearing here was originally "organism", since [Keddy 1989] is primarily concerned with Population Biology.

1.5.3 Some consequences of Hebbian rules

We now relate this to the concepts of cooperation and competition described above. There are two ways in which the Hebb rule can be seen as implementing a form of cooperation. Firstly, the postsynaptic cell cooperates with the presynaptic cells with which it simultaneously fires: these connection strengths are increased, making it yet more likely that the postsynaptic cell will fire when these presynaptic cells fire, thus further increasing the connection strength, and so on. Secondly, presynaptic cells whose activities are *correlated* will tend to cooperate in forming strong connections with the same postsynaptic cell. Their joint firing makes it more likely that the postsynaptic cell will fire, thus increasing all their connection strengths, and so on. This form of cooperation between correlated presynaptic cells is crucially important in the models of map formation discussed in this thesis.

The Hebb rule as expressed above prescribes only increases in synaptic strengths (which we will also refer to as "weights"). Applied just in this form, all weights would grow without bounds, however small the correlation between the postsynaptic cell and each presynaptic cell. One way of limiting synaptic strengths is to specify that weights are reduced under certain pairings of pre- and postsynaptic activity. Recent (though controversial) evidence that such a process may be operating in the brain comes from the discovery of long term depression in the hippocampus [Stanton & Sejnowski 1989]. Here, the pairing of presynaptic activity with postsynaptic inactivity produces a decrease in synaptic strength (see [Willshaw & Dayan 1990] for a discussion of some of the computational issues this raises). An alternative way of preventing unbounded growth is to introduce some form of competition, where the resource is some substance required for the maintenance of synaptic strengths. This is often implemented in theoretical models by means of weight normalization: each pre- or postsynaptic cell (or sometimes both) is assumed to have a fixed amount of "weight substance" that it can share out between the connections to or from that cell. This normalization can take various forms, and the form can greatly influence the final pattern of weights, as we will see later.

1.6 Introduction to the thesis

In this thesis we investigate three different models of map formation in the visual system, and compare them along a number of dimensions that speak to their adequacy as models of the biological system. The most general considerations are as follows:

- Their ability to form maps and stripes from the same mechanism.
- The degree to which their mechanisms and parameters can be interpreted biologically.
- Their ability to form appropriate mappings in the presence of positive correlations between the two eyes (the significance of this will become apparent in chapters 2 and 3).
- The degree to which the morphology of stripes produced by the model matches that of natural stripes, and the effect on model results of monocular deprivation.
- The factors that determine the width of the stripes.

We now briefly summarize the content and main results of the thesis.

1.6.1 Chapter 2: Review of biological data

In this chapter we outline what is known experimentally about the development of topography and ocular dominance, and draw attention to the results most relevant to the thesis. We review the literature regarding both amphibians and fish (in particular frogs and goldfish), and higher mammals (in particular cats and monkeys), and argue that the development of topography and ocular dominance is similar enough in both systems for both to be modelled by the same abstract mechanisms.

1.6.2 Chapter 3: Previous theoretical work

Here we review previous theoretical work on the formation of maps and stripes in the visual system. We consider two levels: firstly, qualitative hypotheses as have commonly been used to explain various aspects of map formation, and secondly specific computational theories that have been implemented as computer programs. We conclude that most models of topography do not naturally extend to also account for ocular dominance segregation.

1.6.3 Chapter 4: Correlational models

In this chapter we present theoretical analysis of a recent and influential model of stripe formation [Miller et al 1989], related to the model of visual receptive field development of [Linsker 1986]. Both assume a pre-existing topography. Cortical unit receptive field development is driven by the correlations within and between two populations of presynaptic (LGN) units, and stripe width is determined mainly by the extent of lateral connections in the cortex. Due to the simple form of the learning rule, the behaviour of the system can be largely predicted from the eigendecomposition of the matrix of presynaptic correlations. In Miller's original analysis (see e.g. [Miller 1990(a)]) he assumed that the correlational structure of activity in both eyes is exactly the same, and that there are no correlations between the eyes. We present analysis of the effect of relaxing these two assumptions by introducing perturbating correlations, which reveals that in the first case there is an instability in the model, and in both cases monocularity can begin to break down. Generally speaking, in the perturbed case eigenvalues of binocular solutions are more likely to be larger than eigenvalues of monocular solutions, and thus there is a greater tendency for cortical units to evolve to a binocular, as opposed to monocular, final state. Some examples of numerically calculated eigenvectors and eigenvalues are presented to support these conclusions.

1.6.4 Chapter 5: Elastic and Kohonen models

This chapter investigates the properties of the elastic net model of topography and ocular dominance, and also those of a related algorithm based on Kohonen's self-organizing feature map algorithm. The elastic net is a developmental algorithm which can be interpreted as performing an optimization. It attempts to form a map, from an input space of points to a target sheet, that minimizes the separation between the representations in the target sheet of points that are neighbouring in the input space. It was first applied to the topography and ocular dominance problem in [Goodhill 1988, Goodhill & Willshaw 1990]: here we significantly extend this work by (among other things) an investigation of the effect of certain parameters on the pattern of stripes in the two dimensional case. In addition, we also investigate a version of Kohonen's algorithm for the same problem, and compare its performance with that of the elastic net. In both these models stripe width is determined by parameters which we interpret in terms of the ratio of correlations within and between eyes. As a final point we show that a global orientation for ocular dominance stripes in the elastic net case can be produced by anisotropic boundary conditions in the cortex.

1.6.5 Chapter 6: A competitive model

The final model presented considers a sheet of interacting cortical cells that compete for the right to respond to input patterns in the two eyes (similar to [Barrow 1987]). Unlike many earlier models, the mechanism we propose forms appropriate maps in the presence of distributed, simultaneous activity in the two eyes. An important difference with earlier competitive models is the form of the normalization rule used to limit synaptic strengths: we analyse this in a simple case. We show that stripe width in this model is influenced by two factors: the extent of lateral interactions in the cortex (as in Miller's model), and the degree of correlation between the two eyes (as in the elastic and Kohonen models).

1.6.6 Chapter 7: Conclusions

Here we sum up the main conclusions and contributions of the thesis. These are as follows:

- Addition of perturbing correlations in Miller's model of ocular dominance leads to an increased tendency to produce binocular rather than monocular units.
- Elastic and Kohonen models naturally produce topography and ocular dominance from the same developmental mechanisms, and can account for the effects of monocular deprivation.
- A new model is presented that for the first time accounts for the development of topography and ocular dominance in the presence of distributed activity simultaneously in the two eyes. Stripe width in this model is related to both the degree of correlation between the two retinae and the width of interactions in the cortex.

Finally we suggest possible directions for future work, and ways in which the predictions regarding the effect of input correlations on stripe width could be tested in the natural system.

Chapter 2

Review of Biological Data

2.1 Introduction

One of the advantages of studying the vertebrate visual system is that, over the past several decades, a large amount of experimental data has been accumulated regarding how this system develops. This data is a vital resource for both constraining models of visual development and inspiring theoretical analysis.

In this chapter we introduce some aspects of the visual system of two different kinds of vertebrate: firstly amphibians and fish (in particular frogs and goldfish), and then higher mammals (in particular cats and monkeys). We note immediately that, although data from a number of different species will be referred to within each group, for our purposes the differences between species within each group are less important than the similarities. We also argue that, despite the many obvious differences between the two groups themselves, in certain key respects they are similar enough for a unified theoretical treatment of visual map formation.

The key points we emphasise are as follows:

- Topographic map development is an active process, rather than relying purely on the preservation of order from retina to cortex/tectum.
- There is evidence of a role for both chemically and activity based processes in topographic map formation, and correlations in activity are crucially important.
- Stripes can form when two eyes innervate the same target structure.

2.2 The retinotectal system

In animals such as frogs and goldfish the main visual centre is the optic tectum. During development, fibres grow from each retina to the opposite tectum,[1] crossing at the optic chiasm, to form a topographic map. This is oriented such that the nasal-temporal and dorsal-ventral axes of each retina map to the caudal-rostral and lateral-medial axes of each tectum respectively (see figure 2.1 for a diagram of the system and an explanation of this terminology). For many purposes, the system can be considered as a sheet of retinal ganglion cells connected to a sheet of tectal cells by a bundle of ganglion cell fibres [Meyer 1982(a)]. The density of ganglion cells over the retina is roughly uniform [Schmidt 1985], and most of these animals have a field of view of 180 degrees, which can be well approximated by a hemisphere centred on the eye. This mapping has the virtue of being a central brain system easy to access [Schmidt 1985] and manipulate [Edelman et al 1985] for a long period of time during development, and it has provided neuroscientists with a model system for studying a variety of phenomena [Fraser & Hunt 1980]. We now focus on the process of map formation.

There is clear evidence that topographic map formation in this system is an active process, rather than simply being due to fibres from the eye retaining their order while growing from retina to tectum. Some evidence relating to this is as follows:

- Although there is wide variation between species in the degree of order existing in the optic nerve, it is almost always the case that the final map in the tectum is ordered to a greater extent than is the optic nerve [Udin & Fawcett 1988].
- The map refines from crude to more precise order during development [Meyer 1983].
- In the frog, retina and tectum grow asymmetrically: the retina adds cells to its periphery in a series of annuli, whereas the tectum adds cells to its caudomedial edge. Despite this, at each stage of development an ordered retinotectal map exists (reviewed in [Gaze & Keating 1972]).
- If the optic nerve of an adult animal is cut, it can regrow to form a normal map, even if the original order of the fibres is artificially scrambled after optic nerve section (reviewed in [Sperry 1963]).

2.2.1 Effect of various surgical manipulations

A large amount of data exist regarding the effect of various surgical manipulations on the retinotectal system. Studying the behaviour under abnormal conditions helps to illuminate the mechanisms of map formation under ordinary conditions,[2] and provides additional constraints for computational models. We first survey

[1] For a recent review of how retinal axons are guided to the tectum see [Hankin & Lund 1991].

[2] Though in general it should be noted that regeneration results cannot always be generalized to the case of development.

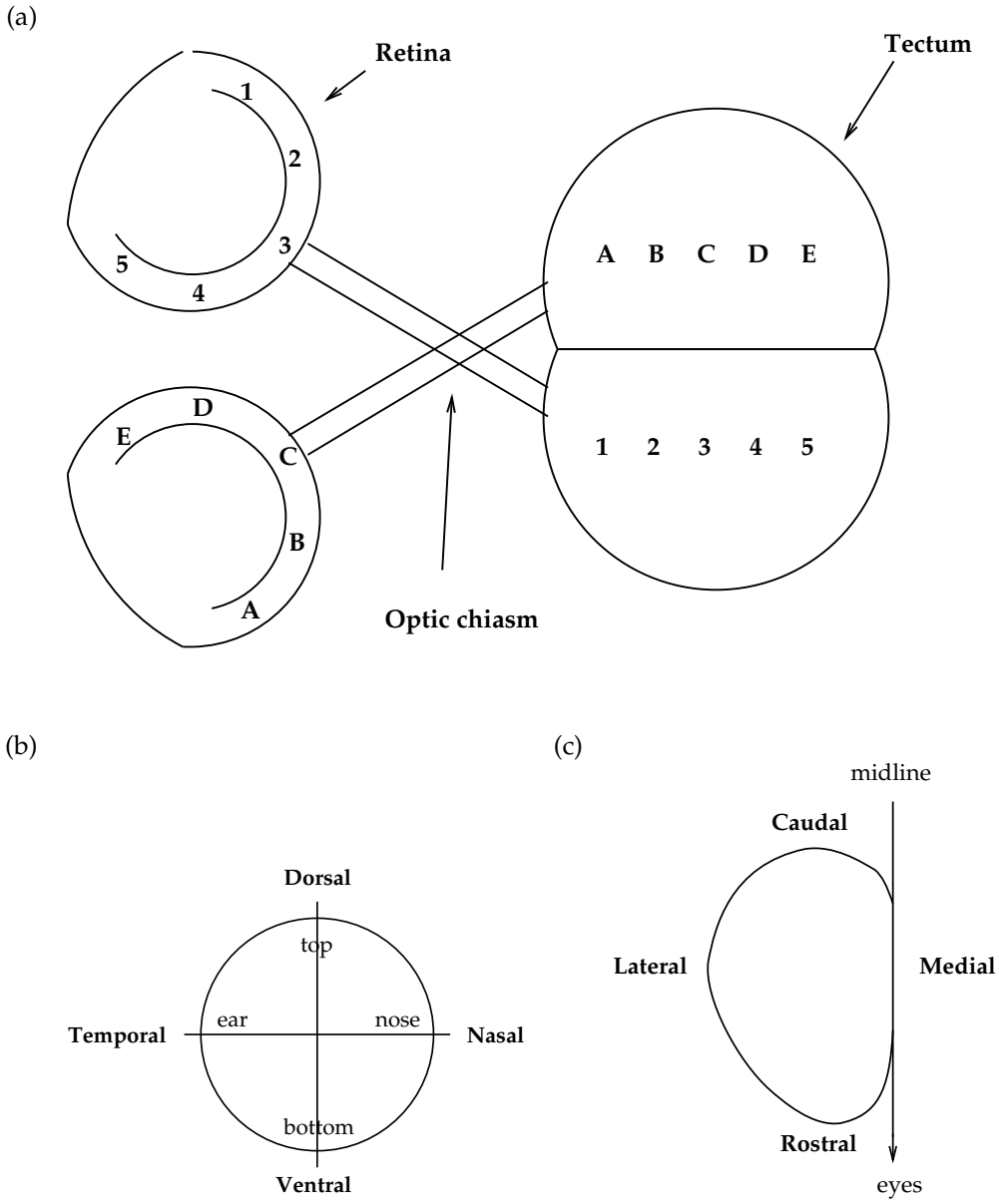

Figure 2.1: (a) A simplified picture of the retinotectal system (after [Meyer 1979]). (b) Terminology for referring to retinal positions. (c) Terminology for referring to tectal positions.

very briefly some of the data regarding manipulations to the single-eye projection: there are a large number of review papers that cover this material in greater detail (e.g. [Fraser & Hunt 1980, Meyer 1982(a), Trisler 1982, Schmidt 1985, Schmidt 1988, Udin & Fawcett 1988, Fraser & Perkel 1990, Hankin & Lund 1991]). In particular, where specific references for the results below are not shown, see [Udin & Fawcett 1988, Fraser & Perkel 1990].

- If a *Xenopus*[3] eye is rotated in situ during development, a normal map results if the rotation is carried out before a certain age, whereas an appropriately rotated map occurs if the rotation is carried out after that age [Hunt & Jacobson 1972].

- If the tectum is rotated during regeneration of the optic nerve, in some cases a normal map results whereas in other cases a rotated map results. Similar results follow if parts of the tectum are translocated: in some cases the cut retinal axons regrow to innervate their normal piece of tectum, and in other cases a map is formed that ignores the translocation.

- Compound eye experiments: When a whole eye is created by fusing together two half eyes, the two halves being of the same type (e.g. nasal, ventral or temporal), they each map across the whole tectum.

- Ablating half the tectum leads to a gradual compression of the regenerating map to fit into the remaining available space [Yoon 1971].

- The map formed after removal of half the retina initially covers half the tectum, but then gradually expands to fill the whole tectum [Schmidt 1978]. Axon terminal density remains the same [Schmidt et al 1978]. If the optic nerve is then made to regenerate again an expanded map is immediately formed [Schmidt 1978].

- Fibres innervating transplanted tissue show a strong tendency to align with fibres in the surrounding tectum, regardless of the orientation of the transplant [Jacobson & Levine 1975]. However, even in the absence of other optic nerve fibres, fibres can find their correct position in the tectum (reviewed in [Fraser & Perkel 1990]).

A number of models based on the matching of chemical markers between retina and tectum have been proposed to account for the above observations. These will be discussed in the next chapter.

2.2.2 Role of activity

There is also a body of data addressing the role of neural activity in map formation. The main results are as follows (for more detailed reviews see [Fawcett & O'Leary 1985, Constantine-Paton et al 1990])

- Spontaneous activity of neighbouring retinal ganglion cells is correlated [Arnett 1978, Ginsburg et al 1984].

- Topographic refinement of the map can be prevented by blocking impulse activity in retinal ganglion cells with Tetrodotoxin (TTX)[4] [Schmidt & Edwards 1983, Meyer 1983].

- Topographic refinement is also prevented by rearing in stroboscopic light [Schmidt & Eisele 1985, Cook & Rankin 1986]. More recent work has revealed some subtleties in these strobe experiments. Firstly, the effect of strobe light varies at different periods during optic nerve regeneration [Cook 1988]. Secondly, the strobe experiments described above had the lens of the goldfish *ablated*, thus creating a diffuse retinal image (with diurnal light the map forms normally under these conditions, i.e. a lack of patterned stimuli). However, if the lens is left intact, the map refines as normal in strobe light [Cook 1987]. Lastly, goldfish reared in the dark (i.e. with only spontaneous retinal activity) have maps as refined as goldfish reared under ordinary lighting conditions [Cook 1988, Cook & Becker 1990].

We return to discuss particular theoretical proposals regarding activity-based mechanisms in the next chapter.

[3] This is a species of amphibian commonly used in retinotectal mapping experiments.

[4] Tetrodotoxin is a potent neurotoxin that blocks sodium channels.

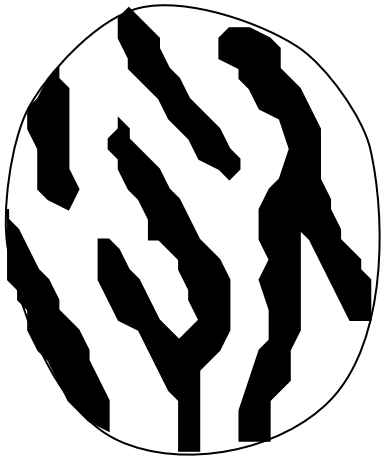

Figure 2.2: A typical pattern of stripes created in frog tectum by implant of a third eye (after [Constantine-Paton & Law 1982]).

2.2.3 Ocular dominance stripes

One of the most remarkable features of the retinotectal system is its behaviour in the unnatural situation of two retinae innervating the same tectum. Here a process of competition appears to take place, and from an initially binocular pattern of innervation a pattern of monocular stripes is formed. The basic results are as follows:

- Early experiments: [Cronly-Dillon & Glaizner 1974] stripped away optic fibres from the left tectum in goldfish, redirected them into the right tectum, and simultaneously removed the left tectum. After leaving the fish for one year they found that terminals from the two eyes lay in two segregated zones, with an overlapping zone in between. In a similar experiment [Levine & Jacobson 1975] found several segregated regions arranged in irregular patches.
- [Constantine-Paton & Law 1978, Law & Constantine-Paton 1981] created a binocularly innervated tectum by implanting an extra eye into embryonic frogs. Here, the fibres from the supernumerary eye grew into one or both tecta, creating regions of binocular innervation. Autoradiographic tracing revealed initially continuous labelling over the entire tectal surface, followed by segregation into a set of eye-specific stripes running rostro-caudally.
- [Law & Constantine-Paton 1980] found that when one tectum was totally or partially ablated, fibres which had lost their normal termination sites crossed to the other tectum and made connections in the appropriate region of tectum. Here, stripes were formed in those parts of the remaining tectum that were innervated by both eyes: invading ipsilateral fibres managed to displace the established connections of the normal contralateral fibres. All bands were highly uniform in width and oriented rostro-caudally.
- [Fawcett & Willshaw 1982] reported stripes from a single surgically reconstructed compound (e.g. double-nasal) eye, and [Straznicky & Tay 1982] found stripes when a compound eye competed for the same tectal space as a normal eye.
- In the so-called "isogenic eye experiment", [Ide et al 1983] took a nasal eye fragment and let it grow into a complete eye. They found that this initially formed a double-nasal map across the whole tectum, which gradually segregated into a pattern of stripes.
- Role of activity: Various experiments have shown that blocking retinal ganglion cell activity with TTX leads to a failure of double projections to segregate into stripes, for instance [Meyer 1982(b)] and [Boss & Schmidt 1984] in goldfish, and [Reh & Constantine-Paton 1985] in frogs.

A typical picture of stripes resulting from one of these experiments is shown in figure 2.2.

2.2.4 Summary

To summarize, the above results point to the existence of the following basic types of mechanisms of map formation in the retinotectal system:

- A polarity mechanism that crudely matches regions of the retina to regions of the tectum [Hunt & Jacobson 1972, Willshaw et al 1983, Fraser & Perkel 1990]. This is not activity dependent [Meyer 1983], and is enabled only after a certain stage of development [Hunt & Jacobson 1972].
- A mechanism that enables the map to expand or compress so as to form smooth order between a mismatched retina and tectum [Yoon 1971, Schmidt 1978, Udin & Fawcett 1988]. Under some circumstances the form of this map can be influenced by the history of the system [Schmidt 1978].
- An interaction between fibres causing them to align with their neighbours, which can sometimes outweigh the polarity mechanism [Jacobson & Levine 1975, Udin & Fawcett 1988].
- An activity-based mechanism that causes the crude map formed by the polarity mechanism to refine [Meyer 1983]: this is dependent on the activity of retinal cells being correlated, but not perfectly correlated [Cook & Rankin 1986].

Further conclusions for the case of ocular dominance stripes are:

- The mechanisms that form a topographic map for a single eye also form ocular dominance stripes (from an initially unsegregated map) for a double projection from two eyes [Constantine-Paton & Law 1978, Law & Constantine-Paton 1980, Law & Constantine-Paton 1981], or a compound eye where both halves attempt to innervate the same part of tectum [Fawcett & Willshaw 1982].
- Stripe formation is activity dependent [Meyer 1982(b), Boss & Schmidt 1984, Reh & Constantine-Paton 1985].
- It is unlikely that a global chemical difference between the two projections is the cause of ocular dominance segregation [Ide et al 1983].

We return to these points in the next chapter.

2.3 The retinocortical system

In mammals such as cats and monkeys the central brain structure receiving the first inputs from the two eyes is area V1 of visual cortex, also known as area 17, or primary or striate visual cortex. Connections from the retina are not direct: retinal ganglion cells synapse at the lateral geniculate nucleus (LGN) in an ordered map, and then geniculate cells project to V1, again forming an ordered map. Thus, visual space is mapped topographically across the cortex. An important difference with the retinotectal system is that V1 naturally receives inputs from both eyes. Fibres from the two eyes are only partially crossed at the optic chiasm: those representing the left visual field from each eye go to the left hemisphere and those representing the right visual field to the right hemisphere. Although terminations from the two eyes are initially evenly distributed in V1, segregation of the two projections occurs during development into an alternating pattern of ocular dominance stripes. This is summarized in figure 2.3. We now discuss this system in more detail (for a fuller account see for instance [Kuffler et al 1984, Lund 1988, Miller & Stryker 1990, Price 1991]). Note that this is a rather simplified description: for instance, we do not consider the large number of fibres carrying information back from cortex to LGN, the function of which is largely unknown.

2.3.1 Precortical stages

The retina is made up of three layers of cells: rods and cones, which convert incoming light into electrical signals, a middle layer of bipolar cells, and then finally retinal ganglion cells, the axons of which form the optic nerve. There are also horizontal and amacrine cells, which make predominantly lateral connections. These layers perform some preliminary processing of the visual image. Retinal ganglion cells have centre-surround receptive fields,[5] and these axons project to the LGN. Although still controversial, many people believe that in cats the optic nerve is again quite disordered (e.g. [Horton et al 1979], also Glen Jeffreys, Personal Communication).

The LGN is a structure lying between the optic chiasm and V1: thus it receives inputs from both eyes. Inputs from the two eyes are initially intermixed during development: however they gradually segregate into a

[5]Cells with centre-surround receptive fields respond best either to an appropriately located dark spot on a light background or an appropriately located light spot on a dark background.

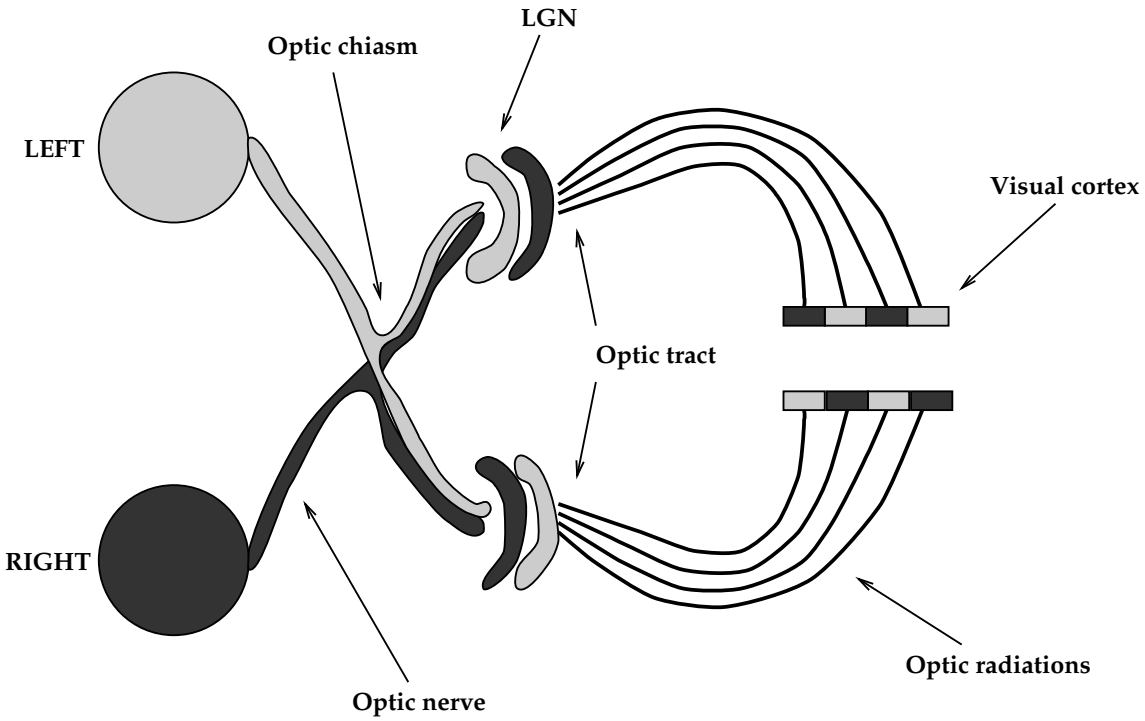

Figure 2.3: A simplified picture of the retinocortical system (after [Kuffler et al 1984]). Only two of the six layers of the LGN are shown.

series of eye-specific layers, each of which receives input directly from the retina. This segregation is activity-dependent [Shatz & Stryker 1988]. In the cat there are three layers for each eye: each receives input across the whole visual field in an orderly map. These maps are aligned, so that an electrode penetrating orthogonal to the layers would encounter cells responding to a similar position in the visual field in each layer (even if the layers correspond to different eyes). The area over which each ganglion cell axon makes contacts varies depending on where in the retina that fibre originated from. Axons representing cells in the fovea of the retina arborize over a smaller area than cells in the periphery. There are also many more LGN cells representing the fovea than the periphery. In general LGN cells have centre-surround receptive fields. Each layer sends a projection to V1.

2.3.2 V1

Primary visual cortex lies at the back of the brain, slightly folded in towards the opposite hemisphere. It is also made up of several layers, some of which receive mostly direct inputs from the LGN, some mostly inputs from other layers, and some a mixture. However, the primary site of termination for LGN fibres is layer IV. In the monkey, the receptive fields of cells in this layer are primarily of centre-surround type, whereas in the cat they are mostly simple (oriented) cells.[6] Initially during development the projection from LGN to layer IV is diffuse with only crude topographic order, and cortical cells are binocular. The formation of this initial map is not activity dependent (reviewed in [Meister et al 1991]). Subsequently, topographic refinement and segregation into ocular dominance stripes occurs (see e.g. [Shatz & Stryker 1978], that is dependent on neural activity [Stryker & Harris 1986]. Recent work has shown that such stripes are also present in humans [Horton et al 1990]. These have a similar qualitative form, though are approximately twice as wide as in cats and monkeys.

The topography of the final map is illustrated in figure 2.4(a). In monkeys, cortical cells are almost entirely monocular in layer IV, so that an electrode traversing this region is driven by one eye and then the other in an alternating fashion (although there are some differences between species: for instance, New World monkeys do not possess ocular dominance stripes, and layer IV is entirely binocular [Hendrickson 1985]). A typical pattern of stripes is shown in 2.4(b): each stripe is approximately 400μm in width. In cats, there is a greater range of ocularities, and the stripes are less sharply defined [Shatz & Stryker 1978]. Hubel and Wiesel identified 7 degrees of dominance, so that a cell of degree 1 or 7 is completely dominant for one eye or the other respectively, and a cell of degree 4 can be driven equally by both eyes (see e.g. [Hubel & Wiesel 1965, Hubel 1990]). After recording from a number of cells, an ocular dominance histogram can be drawn, showing the number of cells falling into each category. For cats this is convex, i.e. the most common ocularity is degree 4 (binocular)

[6] Simple cells respond optimally to an oriented edge or bar at some particular position and angle.

(a) (b)

Figure 2.4: **(a)** Result of an experiment by Roger Tootell that demonstrates the topography of the map from retina via LGN to cortex. The stimulus shown was presented to an anesthetized macaque monkey for 45 minutes. V1 had previously been injected with a radioactively labelled chemical that concentrates where cells are most active. The resulting audioradiograph is of a section of V1 parallel to the surface. As can be seen, neighbourhoods in the retina are preserved in the cortex. The discontinuities in the image lines on the cortex represent shifts in ocular dominance: the pattern was shown to only one eye. **(b)** Ocular dominance stripes in V1 in the macaque as revealed by staining techniques. Both pictures taken from [Hubel 1990].

dropping off on each side.[7] For monkeys it is concave, i.e. rising towards the ends so that the most common ocularity is complete dominance for one or other eye. Lower mammals such as mice have a very slight variation in dominance across the cortex: most cells can be driven equally by both eyes. For cats and monkeys in the layers above and below layer IV similar variations in dominance are present but to a lesser extent, so that an electrode penetrating orthogonal to the surface encounters cells which are dominant for the same eye as in layer IV. Thus the stripes seen on the surface are in fact columns extending through the thickness of the cortex. Although qualitatively similar, the detailed layout of the stripe pattern is different for each individual.

In addition to ocularity, other variables besides space are also mapped across V1 (for review see [Knudsen et al 1987]). One of these is orientation. The overall topography of the map ensures that position in the visual field progresses smoothly across V1. However, on a smaller scale there is smoothness of the orientation map: neighbouring cells respond to similar orientations in a smooth progression. Cells of the same orientation are also arranged in a pattern of stripes, though the thickness of these stripes is several times less than that of ocular dominance stripes [Hubel & Wiesel 1977, Hubel 1990]. Again these extend through the whole thickness of the cortex.

From V1, fibres project mainly to V2, and then from V2 to V4 and V5. These projections are broadly topographic, but with a detailed fine structure (see e.g. [Thompson 1985, Ferrer et al 1988]). We do not consider these "association projections" further.

2.3.3 Detailed structure of ocular dominance stripes

A detailed account of the structure of ocular dominance stripes in macaque monkey is given in [LeVay et al 1985], who used computer-aided 3-D surface modelling techniques to reconstruct the cortical surface in V1 from a series of slices cut through the brain. Below we quote part of the abstract.

[7] Although this is the picture as presented in for instance [Kuffler et al 1984, page 537], it has been suggested [Shatz & Stryker 1978] that using greater care in determining whether cells lie in layer IV rather than deeper or more superficial layers, in which cells have a greater tendency towards binocularity, shows that in fact the ocular dominance histogram for layer IV of the normal cat is roughly flat.

> As described in earlier studies, the stripes formed a system of parallel bands, with numerous branches and islands. They were roughly orthogonal to the V1/V2 border throughout the binocular segment of the cortex. In the lateral part of the operculum, where the fovea is represented, the stripes were less orderly than elsewhere. In the calcarine fissure the stripes ran directly across the striate cortex from its dorsal to its ventral margin. In the far periphery the stripes for the ipsilateral eye became progressively narrower, eventually fragmenting into small islands at the edge of the monocular segment. The overall periodicity (width of left- plus right-eye pair of stripes) averaged 0.88 mm but decreased by a factor of about 2 from center to periphery. This decrease was not accounted for solely by shrinkage of the ipsilateral eye stripes.

The fact that stripes have a tendency to run into the boundaries of V1 at right angles has been commented on by many authors. Also noted has been the tendency for stripe endings to be rounded rather than flat (see e.g. [Swindale 1980]).

An important step forward was taken in [Blasdel & Salama 1986], who used voltage-sensitive dyes to examine patterns of both ocular dominance and orientation selectivity in living macaque monkeys. They found ocular dominance stripes of the same thickness as those found by earlier methods, again intersecting the borders of V1 at right angles. Their technique allowed them to study ocularity and orientation domains in the same animal, and thus see how the discontinuities in the two maps compare. It was found that ocular dominance and orientation stripes do not intersect at a consistent angle, as was first suggested by Hubel and Wiesel (see e.g. [Hubel & Wiesel 1977]). They concluded that "...disruptions in the gradient of preferred orientation tend to coincide with reversals in the gradient of ocular dominance...". That is, orientation changes most rapidly when ocular dominance changes least rapidly. However, their conclusions regarding the structure of the orientation map have recently been contradicted by [Bonhoeffer & Grinvald 1991], who suggest that regions of iso-orientation are more patch-like than band-like.

2.3.4 Development of stripes and the critical period

Ocular dominance stripes develop from an initially unsegregated binocular projection. The stripes appear to form by retraction of exuberant axons, although in the final state it can still be the case that a single axon arborizes to contact two stripes. However, there are differences in the time course of this segregation between cats and monkeys. In cats, the stripes first begin to appear about one month after birth [Kuffler et al 1984, page 536], but in monkeys segregation begins before birth [Rakic 1976]. It is important to note that this prenatal segregation could be driven by spontaneous activity in the visual system (for evidence of spontaneous activity before birth see [Galli & Maffei 1988, Meister et al 1991]), and thus may not require a substantially different explanation from the case of segregation driven by postnatal visual experience, except for the difference in input correlations (see chapter 3).

Both cats and monkeys have a "critical period" for visual development: roughly the first six weeks after birth in monkeys, and from week 3 to week 6 in cats (see e.g. [Kuffler et al 1984, pages 543–544]). This is the period for which the system is plastic: procedures such as monocular deprivation (covering up one eye, as described below), even for a very short time, can have a dramatic effect on the structure of V1. After this period the structure is fixed, and for instance monocular deprivation (even for an extended period of time) has no effect. However, we note later that under some circumstances the critical period can be extended. There is evidence that certain chemicals play a role in determining when the system is plastic [Kuffler et al 1984, page 546].

2.3.5 Experimental manipulations to dominance and the pattern of stripes

As in the retinotectal system, a large number of experiments can be performed in the retinocortical system to create abnormal circumstances during development. In the retinocortical system many of these experiments are concerned with the effect on ocular dominance stripes of disrupting the normal patterns of activity in one or both eyes. These results can shed light on the mechanisms underlying stripe formation in the normal system, and also provide extra constraints for computational models. Some of these results are now described.

- **Monocular suture:** One eye is occluded or sewn shut during the critical period. After a few weeks it is found that (a) a higher proportion of cells than normal are completely monocular, (b) substantially more of the cells in layer IV can be driven by the normal eye as compared to the deprived eye, and (c) ocular dominance stripes are now of different thicknesses for the two eyes. The stripes receiving input from the normal eye expand at the expense of the stripes from the deprived eye: however, the stripe periodicity remains the same (see e.g. [Shatz & Stryker 1978, Hubel et al 1977, LeVay et al 1980]). Even occlusion by

the nictating membrane (which produces a much weaker decrease in light levels than suturing but has the effect of blurring the image) causes the same shifts in ocularity [Kuffler et al 1984, page 541].

- **Binocular suture / dark rearing:** Both eyes are sewn shut, with the result that ocular dominance columns form roughly as normal [Kuffler et al 1984, page 547]. In the paradigm of dark-rearing (i.e. lack of normal visual experience), it was originally found that stripes also formed as normal [Frégnac & Imbert 1978], although later [Swindale 1981] (also Swindale, personal communication) found no evidence of ocular dominance stripes in dark reared cats. However, Swindale's results are controversial (for instance [Swindale & Cynader 1986] reported differences between physiological and anatomical data in a similar experiment). Evidence from other experiments is that dark-rearing can extend the critical period almost indefinitely ([Cynader 1979, Mower et al 1983], for review see [Munro 1986]). Periods of visual experience as little as 6 hours long in dark-reared cats can cause the onset of the critical period, which then appears to run its course even if the animals are immediately returned to the dark [Kuffler et al 1984, pages 543–544].

- **Impulse blockade:** [Stryker & Harris 1986] removed even spontaneous activity by blocking impulses in the two eyes with TTX, and found that when both eyes were thus treated, stripes failed to form. A related experiment was performed in [Chapman et al 1986], who examined the consequences of competition between an occluded eye and one treated with TTX. The result was a shift in ocularity to the occluded eye.

- **Reverse suture:** Here one eye is shut for a time, then opened again and the other eye is shut. If this reversal takes place during the critical period, the impoverished ocular dominance stripes from the eye first deprived are able to expand and eventually (given sufficient time within the critical period) become wider than the stripes from the eye that was initially open [Movshon & Blakemore 1974].

- **Artificial strabismus:** Here the correlational structure of the inputs from the two eyes is disrupted by cutting the extra-ocular muscles, which prevents image registration in the two retinae [Hubel & Wiesel 1965]. This has the effect that almost all cells become entirely monocular, and the pattern of stripes becomes correspondingly sharper. However, these changes occur more slowly than the changes following monocular deprivation [Freeman et al 1982].

- **Alternating occlusion:** Various experiments have been performed to assess to exactly what degree activity in corresponding positions in the two eyes needs to be simultaneous to maintain binocularity. [Hubel & Wiesel 1965] investigated a regime of alternating occlusion on a time scale of one day per eye, from normal eye opening until 10 weeks after birth, and found that the shift towards monocularity was even greater than for the strabismatic cats. The effects of more rapid alternations in deprivation were investigated in [Blasdel & Pettigrew 1979] and [Altmann et al 1987]. Using mechanical and solid-state shutters respectively over the eyes, they found that alternation rates of a few seconds or even fractions of a second reduced cortical binocularity, and also disrupted normal depth discrimination in behavioural experiments [Altmann et al 1987].

- **Effect of eye movements:** Various experiments suggest that eye movements are essential for the plasticity effects described above to operate. [Freeman & Bonds 1979] found that monocular deprivation did not lead to a breakdown of binocularity in 4-week old kittens if the animals were paralysed. However, if either eye was moved mechanically, there was a shift to monocularity. A similar result has been found for plasticity with regard to orientation selectivity in [Buisseret et al 1978].

2.3.6 Mechanisms

The experiments described above have usually been interpreted to mean that there is some competitive interaction in V1 between fibres from the two eyes. This is activity dependent [Stryker & Harris 1986], and depends on both the strength and the correlational structure of the incoming activity. If one eye is given a competitive disadvantage, there is a shift to the other eye. Relative rather than absolute strengths seem to be the important factor [Chapman et al 1986], and a competitive imbalance increases the degree of monocularity. If correlations between the two eyes are disrupted, the degree of monocularity is also increased. However, stripes can form naturally even when the two eyes are likely to be highly correlated (e.g. in monkeys).

Direct measurements have been made of the retinal correlations produced by spontaneous activity. [Mastronade 1989] found positive correlations between the firing of neighbouring cells of similar receptive field type in the cat. More recently [Meister et al 1991] studied the correlations in fetal cat retina *in vitro*, and found that spontaneous activity had the form of a wave of excitation sweeping across the retina in an arbitrary direction, lasting a few seconds, followed by one to two minutes silence. Both these experiments show that spontaneous

activity is indeed locally correlated. However, there are no experimental results regarding the form of retinal correlations produced by normal patterned input. It is likely that these will be significantly different from the spontaneous case. In particular, one would not expect the two eyes to be correlated during spontaneous activity, whereas one would for the evoked activity case.

Two experiments which provide further evidence about mechanisms are as follows. [Reiter & Stryker 1988] used the easily-measurable effects of monocular occlusion to test the Hebbian hypothesis that both pre- and postsynaptic activity are required for changes in synaptic strengths. They blocked only postsynaptic activity in the cortex during the critical period while one eye was occluded, and found that there was a consistent shift in the responsiveness of the visual cortex in favour of the less-active, *closed eye*. This indicates that repeated pairing of presynaptic activity with zero postsynaptic activity leads to a competitive disadvantage for that eye compared to an occluded eye. Another experiment designed to test the effect of pre- and postsynaptic activity pairing was [Frégnac et al 1988]. Here, it was found that a cell dominant for eye 1 could be converted to a cell dominant for eye 2 by repeated artificial (i.e. direct current injection from an electrode) increase in the cell's response to input from eye 2, thus confirming the importance of temporal correlations in pre- and postsynaptic activity.

2.4 Conclusions

A large proportion of what is known experimentally about the mappings from retina to tectum and from LGN to V1 can be understood in terms of a sheet of presynaptic cells connected to a sheet of postsynaptic cells by fibres with variable connection strengths. In frogs and goldfish, experimental manipulations to this system have included removing or rotating parts of the pre- and/or postsynaptic sheet, fusing together parts of two or more presynaptic sheets, and introducing fibres from an additional presynaptic sheet to compete with the normal fibres for space on the postsynaptic sheet. In cats and monkeys, experiments have focused on the effects of monocular deprivation on patterns of ocular dominance as a way of examining the behaviour of the system under abnormal conditions.

The retinotectal and retinocortical systems are rather different in detail. However, they have the following important properties in common:

- There appears to be a polarity mechanism that provides an initial orientation for the map.
- An initially crude map on the postsynaptic sheet gradually refines during development.
- Ocular dominance stripes are formed when two presynaptic sheets compete for the same postsynaptic space.

Phenomena two and three are our primary interest in this thesis: in particular, how they can arise from the same mechanisms.

Chapter 3

Previous Models

3.1 Introduction

In the previous chapter we reviewed a range of experimental data, and also outlined the main organizing principles that have been proposed to account for that data. In this chapter, we first discuss informal models for how these organizing principles are realised, and we then review a number of computational models. Each of the latter specifies a mechanism or a combination of mechanisms for map formation in a form precise enough for implementation as a computer program, and thus helps to make clear the capabilities and limitations of each mechanism.

Some important questions to be borne in mind when considering these models are as follows:

- How are presynaptic locations specified?
- How is information about presynaptic positions communicated to the postsynaptic sheet?
- How are synaptic strengths updated?
- What cooperative and competitive interactions take place to ensure a smooth map across the postsynaptic sheet?

Although there are several models, they are all based on quite a small number of principal mechanisms, usually some form of the Hebb rule. The defining characteristics of a particular model are therefore the precise mathematical forms adopted for each of these mechanisms. One such characteristic that will be important later in this thesis is the type of normalization used to limit synaptic strengths.

We judge each model according to the criteria outlined in the introduction: their ability to develop topography and ocular dominance from the same mechanisms, their biological plausibility, their success in the presence of positive correlations between the eyes, the properties of the ocular dominance stripes produced, and the predictions they make regarding stripe width.

3.2 Informal models for topography

Over the past several decades there has been much discussion and controversy over the mechanisms of retinotectal map formation in amphibians and fish. We take as our starting point the hypothesis of chemical matching between pre- and postsynaptic cells, discussed first by Langley (e.g. [Langley 1895], see also [Brown et al 1991]) and then in much greater detail by Sperry.

3.2.1 Chemospecificity

To account for the experimental facts that retinal axons can find their normal sites of termination in the tectum even when optic nerve fibres are artificially scrambled, Sperry postulated a scheme of chemical matching between retina and tectum (see [Sperry 1963] for review). He describes it thus:

> The establishment and maintenance of synaptic associations [is] conceived to be regulated by highly specific cytochemical affinities that arise systematically among the different types of neurons involved via self-differentiation, induction through terminal contacts, and embryonic gradient effects [Sperry 1963].

This "chemospecificity" hypothesis has been extremely fruitful in both providing an informal account of various experimental results, and inspiring theoretical models. A distinction can be made between the three related chemical mechanisms mentioned in the above quotation. (1) The idea of self-differentiation is that some process takes place whereby each retinal and tectal cell assumes a unique chemical identity, and there is a process of matching between appropriately located retinal and tectal cells. One possibility is that there might be no relation between the labels associated with neighbouring retinal or tectal cells: this is sometimes referred to as a "mosaic"of labels. (2) An alternative to mosaics is explicitely mentioned by Sperry: that there are *gradients* of chemical differences in both retina and tectum, and it is the gradients rather than individual cells that are matched. At least two gradients are needed to match two-dimensional structures. (3) The notion of induction through terminal contacts is that the chemical identities of retinal cells are passed via retinal axons into the tectum. It is not clear from the above whether Sperry intended this to be simply a way for the chemical identities of retinal and tectal cells to be brought together for matching purposes, or a map-forming mechanism

in itself. We will discuss computational models of map formation based on each of these three mechanisms later.

Besides the unscrambling of abnormal projections in Sperry's own experiments, the chemospecificity hypothesis fits well with the eye rotation results of [Hunt & Jacobson 1972] and others (for a fuller review of data supporting this hypothesis see [Gaze & Keating 1972]). However, subsequent experimental data began to cause problems for the hypothesis. Firstly, the expansion and compression results of [Yoon 1971] and others (see for example [Gaze & Keating 1972, Udin & Fawcett 1988]) showed that the matching scheme cannot be "rigid": even though there are a mismatched number of termination sites, a smooth map still forms. An interpretation of these results favoured by Yoon was that "respecification" of retinal or tectal cells occurs after surgical intervention, so that each cell develops a new label appropriate for the new context. However, as pointed out by [Gaze & Keating 1972] this is a post-hoc explanation, and has no predictive power (see also [Prestige & Willshaw 1975]). In addition, the respecification hypothesis requires that the pre- or postsynaptic sheet has direct knowledge about how much the other sheet was altered by the surgery. The expansion and compression results mentioned above have been referred to by [Gaze & Keating 1972] as "systems-matching". They put forward the important hypothesis that this could occur by a competitive mechanism: retinal fibres compete for space in the tectum, and since each fibre has equal competitive strength they spread out to fill the entire space available, be that smaller or larger than the normal case.

We have presented these issues regarding the role of chemical specificity in map formation here to aid understanding of the computational models reviewed in this chapter that are based on these ideas. More recently debate has continued, fueled by the results of ever more exotic surgical manipulations (for reviews see [Trisler 1982, Meyer 1982(a), Schmidt 1985, Udin & Fawcett 1988, Fraser & Perkel 1990]), and also direct evidence for the existence of gradients of chemical markers in retina and tectum (reviewed in [Gottlieb & Glaser 1981]). We do not discuss this further here, since our primary concern in the following chapters is the effects of correlated retinal activity on map formation.

3.2.2 Correlated activity

The idea that correlations in presynaptic activity could drive the formation of an ordered map was discussed by several experimentalists before being formulated as a computational model in [Willshaw & von der Malsburg 1976]. The informal notion behind the model is as follows.

Consider a postsynaptic cell p receiving connections from two presynaptic cells, a and b. The following assumptions are made:

- The activity of p is related (perhaps by a threshold function) to the sum of the activities of a and b weighted by their connection strengths.
- Connection strengths are updated according to a Hebbian rule such that at each step the strength of each connection is increased in proportion to the product of pre- and postsynaptic activity.
- The activities of a and b are highly correlated only if they are close together in the presynaptic sheet.

There are now two possibilities for the development of the two connection strengths in this system. If a is near to b, they will tend to be active at the same time, and thus be likely to make p active at this time. This will lead to both connection strengths being increased. In other words, a and b *cooperate*. However, if a and b are widely separated then their activities are not correlated, and the two connection strengths evolve independently. If a normalization rule is introduced to limit connection strengths (e.g. the sum of connection strengths for cell p is maintained at a constant), then in the latter case a and b actively *compete*. Thus we see how locally correlated activity in the presynaptic sheet has a tendency to cause neighbouring presynaptic cells to be connected to the same postsynaptic cell.

However, two other components now need to be added to explain map formation. So far there is no pressure for neighbouring postsynaptic cells to be connected to (1) different but (2) neighbouring cells in the presynaptic sheet. The first can be achieved by limiting the total synaptic strength available to each postsynaptic cell (this will become clearer when we discuss [Prestige & Willshaw 1975] below). The second can be achieved by introducing lateral interactions in the postsynaptic sheet. The principal requirement is that neighbouring postsynaptic cells excite one another: this is equivalent to specifying that their activities be correlated. Then, by a similar argument to that for correlations between presynaptic cells, neighbouring postsynaptic cells will have a tendency to receive connections from the same presynaptic cell, or more generally, presynaptic cells with correlated activities, i.e. those that are close together. In this way an activity-based learning mechanism can lead to the formation of a locally-ordered map.

By itself, such a mechanism can generate only local order. One additional assumption that assists the establishment of global order is that there exists some other mechanism to specify an overall bias or polarity for the map. The experimental evidence supports the notion that map formation is activity-dependent but that there is also some other mechanism that establishes a crude map in the absence of activity [Meyer 1983]. Many theoreticians have postulated a two-stage process whereby a crude initial map is formed by matching of chemical gradients, followed by refinement of this map due to activity-based mechanisms (first proposed in [Willshaw & von der Malsburg 1976]). The mathematical effects of this initial symmetry breaking are discussed in [Häussler & von der Malsburg 1983].

3.2.3 Summary

Cytochemical matching, normalization, and a combination of correlated presynaptic activity and lateral excitation of neighbouring postsynaptic cells, are the main mechanisms that have been put forward to explain topographic map formation in the retinotectal and retinocortical systems. We will see shortly various ways in which these qualitative notions have been formalized into computational models, and how different implementations of these ideas can lead to different results.

3.3 Informal models for ocular dominance

Two issues involved in the development of ocular dominance are: (1) Why should tectal or cortical cells become monocular despite being initially binocularly innervated, and (2) Given that half the cells are dominant for one eye and half for the other, why should these cells be laid out in a periodic pattern of stripes?

The first issue was discussed by [Hubel & Wiesel 1965] (see also [Stent 1973]). In their "artificial strabismus" experiments they made kittens unable to focus both eyes on the same point, in effect severely reducing the normal correlations between cells at corresponding positions in the two eyes. This led to an increased tendency for cortical cells to be monocular. [Hubel & Wiesel 1965] suggested that this could be explained by similar arguments to those made above for the topography case: uncorrelated afferents converging on the same postsynaptic cell compete. They refer to a "lack of synergy between the two afferent paths": following from our discussion above it is clear that Hubel and Wiesel were assuming some form of competition (e.g. normalization) between afferents. However, this argument does not account for how monocularity arises in normal cats and monkeys, where there is a high degree of correlation between the two eyes. Also, we note that both [Hubel & Wiesel 1965] and [Stent 1973] refer to "asynchronous" rather than uncorrelated activities in the two afferent pathways. These are not equivalent: asynchronous means "occuring at different times" (Collins English Dictionary, third edition), which implies anticorrelation between the two pathways. Uncorrelated afferents may still sometimes be synchronous. If the two eyes move effectively independently, it seems more reasonable to suppose that activity is uncorrelated than anticorrelated. We emphasise this point here because the precise nature of the correlations between the two eyes will be important in later chapters.

Moving now to the second issue, an idea which we will refer to as the "balancing" model was first introduced in [LeVay et al 1975]:

> How is the columnar pattern arrived at during development? The segregation of the two sets of afferents into alternating bands might be understood as the consequence of two conflicting tendencies, the specification of the two sets to occupy the same cortical space according to a single topographic map, and an opposing tendency for grouping within each set, as if caused by mutual repulsion between members of the two sets. The first principle would tend to increase the degree of interpenetration of the two sets, that is, to reduce the maximum distance from any cortical point to the nearest column border. If this distance....is kept constant, then a repulsion between the two sets would tend to produce an arrangement with the smallest total length of interface. Alternating bands give a shorter interface than any other arrangement such as a checkerboard or circular islands in a matrix [LeVay et al 1975].

The idea is that the final pattern is a balance between two contradictory constraints: that corresponding cells in the two eyes are mapped close together, and that afferents from neighbouring positions in each eye are kept together. This notion was also proposed more specifically to explain ocular dominance stripe formation in the retinotectal system in [Constantine-Paton 1983]. She suggests that the first tendency is caused by chemospecific matching, and states "Only the striped pattern optimizes each retina's ability to contact all regions of the tectal lobe while simultaneously minimizing the borders where ganglion cells must synapse adjacent to the non-neighbouring cells of the second retina".

The notion of minimal boundary length was also mentioned by von der Malsburg in [von der Malsburg & Willshaw 1976], and explored more quantitatively in [von der Malsburg 1979]. He imagined a postsynaptic sheet with periodic boundary conditions (i.e. a torus), and considered the topologically different ways in which this can be divided into two regions by closed lines. He showed that the topologies with the shortest boundary lengths are circles or stripes, and went on to calculate the minimal boundary lengths for both cases as a function of α, the fraction of the plane taken up by one ocularity. For $1/\pi < \alpha < 1 - 1/\pi$, stripes have shorter boundary length than circles: the normal case of $\alpha = \frac{1}{2}$ is in this range. [von der Malsburg 1979] also predicted that the determiner of stripe width in the balancing model is the relative strength of these two tendencies.

Neither [LeVay et al 1975, Constantine-Paton 1983, von der Malsburg 1979] put forward any convincing justification for why the system should attempt to minimize the length of stripe boundaries. [LeVay et al 1975] refer to a repulsive force between opposite-eye fibres, but without saying how this force arises. [Constantine-Paton 1983] implies that there is a "cost" incurred if neighbouring tectal cells represent different eyes. The isogenic-eye experiments [Ide et al 1983] seem to rule out a global chemical difference between the two eyes as being a factor in stripe formation. Whether or not a mechanism based on correlated activity could implement a suitable cost function depends on the precise form of correlations between the two eyes and the nature of the competition in the tectum/cortex.

3.4 Computational models - topography

The qualitative models of topography described above were systematically developed as computational models in [Prestige & Willshaw 1975, Willshaw & von der Malsburg 1976, von der Malsburg & Willshaw 1977, Willshaw & von der Malsburg 1979]. We describe these in some detail, and then discuss other models that have been presented.

Some of the models are phrased in terms of mappings in lower vertebrates from retina to tectum, some in terms of mappings in higher vertebrates from LGN to cortex, and some in terms of a general mapping between a pre- and postsynaptic sheet. However, most of the models described here do not rely on the details of either the retino-tectal or geniculo-cortical system beyond the general idea of a mapping between a pre- and postsynaptic sheet. The difference lies in which set of data the model attempts to explain. If the emphasis is on higher vertebrates, the data usually considered is that regarding ocular dominance and deprivation results, whereas for lower vertebrates it is expansion, compression, translocation and rotation of maps. For completeness, in the latter case we briefly indicate which set of data each model can account for, though these results are not part of the criteria for model evaluation listed in the introduction. At the end of the chapter we summarize in tabular form the important properites of each model.

3.4.1 Prestige and Willshaw

[Prestige & Willshaw 1975] formalized notions of chemospecific matching suggested by Sperry's hypothesis. They drew a clear distinction between two forms of chemical matching, which they termed "group I" and "group II" (often subsequently referred to as type I and type II). In group I (which they also refer to as *direct*) matching, each presynaptic cell j has an affinity for just a small neighbourhood of postsynaptic cells, with peak affinity for the cell topographically matching to j in the postsynaptic sheet: a mosaic model. While such a rigid scheme can form a map under normal conditions, it cannot account for certain experimental data without invoking the unsatisfactory notion of respecification (see earlier). In group II matching, "...*all* axons have maximum affinity for making and retaining contacts at one end of the postsynaptic sheet of cells, and progressively less for the cells at greater distances from that end. Similarly, all postsynaptic cells have maximum affinity for axons from one end of the postsynaptic set, and axons remote from this end have correspondingly less likelihood of retaining *any* contact. We may thus talk of *graded affinity* within both pre- and postsynaptic sets..." [Prestige & Willshaw 1975]. Simulations showed that this mechanism is capable of forming an ordered map if competition is introduced, in the form of normalization. They specified that each presynaptic cell could make only a fixed number of contacts N_{pre} among the postsynaptic cells, and similarly each postsynaptic cell could only support a fixed number of contacts N_{post} from presynaptic cells. This then ensures an even spread of connections: without competition for postsynaptic sites every presynaptic cell would connect only to the highest affinity end of the postsynaptic set, while without competition for presynaptic sites every postsynaptic cell would receive connections only from cells at the highest affinity end of the presynaptic set. However, to explain compression and expansion of retinotectal maps, they found that it was necessary to make the additional assumption ("regulation") that N_{pre} and N_{post} were also altered. For instance, if half the postsynaptic set is removed then, unless N_{post} is increased, connections will only be made from the highest-affinity end of the presynaptic set.

3.4.2 The Neural Activity model

[Willshaw & von der Malsburg 1976] proposed a theory in which topographic map formation is driven by correlated activity in the presynaptic sheet. A two-dimensional sheet of presynaptic cells is totally connected, with synapses of variable strengths, to a two-dimensional sheet of postsynaptic cells. Initial connectivity is random, with a small initial bias to model the effect of an initial crude polarity-matching mechanism: the initial weights between a few topographically matching pairs of pre- and postsynaptic cells were artificially strengthened. There also exist fixed lateral connections in the postsynaptic sheet, excitatory at short range and inhibitory at long range (similar to [von der Malsburg 1973]). At each time step, a pattern of "dipole" activity is imposed on the presynaptic sheet. This pattern is generated by choosing a cell at random and giving it and one of its neighbours unit activity, while all other cells in the presynaptic sheet have zero activity. Presynaptic cells are thus anticorrelated with all other presynaptic cells apart from their immediate neighbours. The activity of cells in the postsynaptic sheet is calculated by adding together the feedforward and lateral interactions and iterating until a stable state is reached. All postsynaptic cells with activity above a certain threshold then have their connection strengths updated according to a Hebbian rule. This is a form of competition between units, similar to [von der Malsburg 1973]. In order to prevent the weights (which are always positive) growing without bounds, the total sum of connection strengths for each postsynaptic unit is normalized to a constant value by dividing each connection strength by the total sum. They showed that such a mechanism is capable of generating topographic maps, given a small degree of initial order.

Although providing an elegant account of how activity can drive map formation under normal conditions, the Neural Activity model fails to account for translocated maps in graft transplantation experiments, and also some of the compound eye experiments (for review see [Willshaw & von der Malsburg 1979]).

3.4.3 The Tea Trade model

The one remaining idea of chemospecific matching referred to by Sperry that had not been investigated theoretically was that of map formation being somehow dependent on induction of chemicals from the retina into the tectum. [von der Malsburg & Willshaw 1977, Willshaw & von der Malsburg 1979] proposed a model based on this mechanism (for a summary of the model see [von der Malsburg & Willshaw 1981(a)]; [von der Malsburg & Willshaw 1981(b)] presents a more formal statement of the equations, and [von der Malsburg 1989] is a recent review). The model is expressed in terms of chemical markers, but markers intrinsic only to the presynaptic sheet. There are no pre-existing markers in the postsynaptic sheet: it is assumed that presynaptic markers are transported to the postsynaptic sheet via induction along retinal axons (evidence for such a mechanism in the natural system comes from the experiments of [Schmidt 1978, Schmidt et al 1978]). Several markers exist in the presynaptic sheet, the sources of which are spaced out in the presynaptic sheet at fairly regular intervals. An analogy presented to explain the working of the model was the import of tea from plantations in India to British towns [von der Malsburg & Willshaw 1977]: hence the name "Tea Trade model".

Initially, markers diffuse in the presynaptic sheet until a stable distribution is set up. It is assumed that each presynaptic axon then induces the vector of markers existing at that point in the presynaptic sheet into the postsynaptic sheet, where the markers diffuse according to the same rule as in the presynaptic sheet. The rate of induction at each synapse is proportional to the strength of that synapse. Synaptic strengths are updated periodically according to the degree of similarity between the vector of markers each fibre carries and the vector of markers already existing at those points it contacts in the postsynaptic sheet. An interesting aspect of this model is that synaptic updating occurs with a molecular analogue of the Hebb rule: the strength of connection between presynaptic cell i and postsynaptic cell j is increased in proportion to the similarity of their vectors of chemical markers, rather than on the basis of the correlations in their activities as Hebb originally formulated the rule. A slight overall bias in the connection strengths is specified initially in order to provide a global orientation for the map. Synaptic strengths are normalized by division so that each *presynaptic* cell can only support a certain total strength of connections. This ensures that every presynaptic axon makes contacts in the postsynaptic sheet. Willshaw and von der Malsburg showed that such a process reaches a stable state and can form appropriate maps under normal conditions, and also correctly predict the outcome of a range of results under abnormal conditions. In particular, some regeneration experiments (e.g. [Schmidt 1978]) can be accounted for under this scheme by assuming that when regeneration occurs, the postsynaptic sheet holds a "memory" of the previous pattern of innervation in terms of the previous stable distribution of markers.

An algorithm similar to the Tea Trade model has been analysed mathematically for the one-dimensional case in [Häussler & von der Malsburg 1983]. The technique of analysis used is synergetics: "Linear analysis near the initial state, classification of the linear modes into principal and ancillary, adiabatic elimination of the latter from the original system and discussion of the resulting equations". Their analysis makes the assumption that the two modes corresponding to an ordered map (one in each direction) are the only modes with positive

eigenvalue. A linear analysis shows that eventually these modes dominate. By considering higher-order terms, they were also able to show that asymmetries in the initial configuration can select between the two modes (cf [Turing 1952]).

3.4.4 Cowan's model

The three models described so far each address one of the basic mechanisms that could underlie map formation, and each can account for a different subset of the experimental data. In order to explain a more complete range of data, some experimentalists (e.g. [Meyer 1982(a), Fraser & Perkel 1990]) have argued that it is important to consider the simultaneous operation of several mechanisms. These are hypothesised to act in concert in normal map formation, but under abnormal conditions different mechanisms can dominate to produce varying outcomes.

One such model is that of [Whitelaw & Cowan 1981]. This was an attempt to integrate both marker- and activity-based mechanisms in map formation, by combining a gradient of adhesive specificity (i.e. a group II mechanism) with synaptic updating as in the Neural Activity model [Willshaw & von der Malsburg 1976]. They employed an activity regime and updating rule similar to the Neural Activity model, but with changes in synaptic strengths multiplied by the degree of adhesion between the corresponding pre- and postsynaptic cells (it is unclear how such a mechanism could be implemented in the natural system). Both pre- and postsynaptic normalization are employed. They predict a range of the experimental literature (including expansion and compression, mismatch, rotation and compound eye experiments), and also draw attention to experimental evidence contradictory to the induction hypothesis of the Tea Trade model.

Recent additions to this model [Cowan & Friedman 1990(a), Cowan & Friedman 1991] have increased the range of data it can account for. An extra term in the updating equations has been added to represent ".... a postulated random depolarization which occurs at synapses due to the quantal release of neurotransmitter...." [Cowan & Friedman 1990(a)]. This means that a crude map will now form in the absence of electrical activity, in agreement with the experimental data. In addition, motivated by experimental evidence that fibre-fibre interactions play a part in map formation (see e.g. [Fraser & Perkel 1990]), a tendency for fibres to stick to their retinal neighbours has been added. With these and other modifications the model reproduces the results of [Schmidt et al 1978], among others.

3.4.5 Fraser's model

Another example of the consideration of multiple influences on map formation is the model of [Fraser 1980, Fraser 1985, Fraser & Perkel 1990]. Fraser introduced the notion that the state of the system could be described by an "adhesive free energy" G, which depends on how successfully a number of constraints are satisfied. In his model the form of the final mapping is found from minimising G by simulated annealing [Kirkpatrick et al 1983]. In [Fraser & Perkel 1990], the constraints employed (in order of decreasing weighting in G) are (1) a position independent adhesion between retinal and tectal cells, (2) a general competition among retinal axons for tectal space, (3) a tendency for neighbouring axon terminations in the tectum to stabilize if those axons come from neighbouring positions in the retina (implemented perhaps by correlated activity), (4) a dorsoventral gradient of adhesive specificity in retina and tectum, and (5) an anteroposterior gradient in retina and tectum. In [Fraser & Perkel 1990] an impressive array of results is presented that qualitatively accounts for a large proportion of the experimental literature. However, Fraser does not provide a developmental mechanism which could perform such a minimization. Also, the way he proposes constraints (4) and (5) to be implemented is somewhat unclear. In fact a group I type of specificity is employed, with the suggestion that this could be implemented by gradients of different molecules running in opposite directions [Fraser & Perkel 1990]. However, he does not properly explain how the addition of two linear gradients results in a local minimum of G for each retinal cell.

3.4.6 The arrow model

Shortly after [Prestige & Willshaw 1975], an alternative way of looking at map formation was proposed [Hope et al 1976], the so-called "arrow" model. The assumptions of this model are that (a) retinal fibres that terminate next to each other in the tectum are able to compare their relative positions of origin, and (b) retinal fibres can identify both rostral-caudal and medial-lateral axes of the tectum (this could be implemented in terms of a polarity gradient of marker along each of these axes). The procedure is basically the bubblesort algorithm, though they do not give a biological implementation for this. Starting from initially random connections where

each retinal fibre contacts one tectal cell, two retinal fibres are chosen at random that terminate next to each other in the tectum. Their retinal positions are compared, and if appropriate the sites of termination of the fibres are exchanged. This simple algorithm is capable of forming topographic maps under normal conditions, and also rotated maps when a piece of the tectal array is rotated. However, it fails to account for the translocated map experiments. This algorithm has since been extended in a number of ways [Overton & Arbib 1982]. The tectum is now considered to be continuous, retinal fibres are assumed to branch and make several contacts in the tectum, and "....the movement of a retinal fibre is determined by the overall trend of the movements of its individual branches" [Overton & Arbib 1982]. They also explicitly introduced group II position markers for each axis in retina and tectum. This "Extended Branch Arrow model" is able to account for more of the experimental data than the original arrow model.

3.4.7 The work of Amari

[Amari 1980, Takeuchi & Amari 1979] performed a mathematical analysis of a one-dimensional version of the Neural Activity model, but assuming that the postsynaptic sheet is a continuous neural field, and that signals are chosen at random in a continuous rather than discrete presynaptic sheet. In [Amari 1980] inputs were assumed to be single points, but in [Takeuchi & Amari 1979] the case of inputs with finite spatial extent was analysed. This latter case produces more interesting results, since now there is an interaction between the two length scales involved: the width of the input stimuli L_{in}, and the width of lateral interactions in the postsynaptic sheet L_{lat}. It was shown that if $L_{in} < L_{lat}$, a smooth topographic map forms as normal, but if $L_{in} > L_{lat}$ the map forms a series of *microstructures*: although topographic on a large scale, on a fine scale the map is broken up into small blocks, which Amari identified with columns in visual cortex [Hubel & Wiesel 1977]. Later [Amari 1983, Amari 1988] further extended the analysis to show that the area in the postsynaptic sheet devoted to the representation of a particular part of the presynaptic sheet is proportional to the probability of occurrence of stimuli in that part of the presynaptic sheet. That is, more frequently occurring stimuli are represented in more detail in the postsynaptic sheet (in agreement with experimental data on somatosensory cortex - see for example [Sur et al 1980]). This analysis has recently been extended to the case of a dual population of postsynaptic cells, some inhibitory and the rest excitatory in accordance with Dale's law, and similar results are obtained [Da Silva Filho 1991].

3.4.8 Kohonen's algorithm

The models of topographic map formation so far discussed have generally been addressed towards the retinotectal or retinocortical systems. We now mention a *general* map formation technique that sacrifices biological plausibility for algorithmic expedience, since it will be important later in the thesis. This is the algorithm of [Kohonen 1982(a)], which is basically the Neural Activity model [Willshaw & von der Malsburg 1976], abstracted away from biological details and simplified to make it much faster computationally [Kohonen 1982(a), von der Malsburg 1989]. Kohonen's insight was to see that the time-consuming process of iterating activity in the postsynaptic sheet to reach a stable state in response to input both from the presynaptic sheet and from the lateral excitatory and inhibitory connections in the postsynaptic sheet could be replaced by a much simpler process with approximately the same behaviour. In Kohonen's algorithm, it is *assumed* that the result of iteration in the postsynaptic sheet is that a small "bubble" of activity forms, centred on the postsynaptic cell with the largest activity in response to purely feedforward input, such that these cells are the only postsynaptic cells with non-zero activity: they are assumed to inhibit all other cells. Only the synapses for this small bubble of cells are updated in response to that pattern. The process of iteration is thus now simplified to the step of just finding the postsynaptic cell with the largest activity (or equivalently the largest input, if the activation function of cells is assumed to be monotonic) for the input pattern presented. This winning cell and its geometric neighbours then have their connection strengths updated. This implements a form of competition between cells, and is similar to competitive learning [Rumelhart & Zipser 1986, Carpenter & Grossberg 1988] with a topology on the output units. It has been shown analytically that this process converges to an ordered map in the case of a discrete one-dimensional presynaptic sheet mapping to a discrete one-dimensional postsynaptic sheet [Kohonen 1982(b)]. Additionally simulations show that this algorithm leads to maps with interesting geometric structure in a number of more complex cases, including mapping between spaces of different dimensionalities [Kohonen 1982(a), Kohonen 1988]. The final results seem to be reasonably insensitive to the precise form of the neighbourhood function, although different functions can lead to quite different convergence times [Erwin et al 1991]. One important aspect of Kohonen's simulations is that usually the width of the neighbourhood function is changed over the course of a simulation. Best results are obtained if this width is initially large and then gradually reduced: a process often referred to in this context as "annealing", though it is rather different from the technique of simulated annealing [Kirkpatrick et al 1983]. The most obvious correlate for such a

process in the biological system is the critical period: however, there is no evidence for such changes in lateral connectivity during natural development.

[Cottrell & Fort 1986] applied Kohonen's algorithm specifically to the case of retino-cortical mappings, although still in an algorithmic rather than biological form. They considered a discrete array of retinal cells making contact with a continuous cortical sheet. A small part of the retina was stimulated at each time, and the relative positions of the centres of mass of retinal weights in the cortex was used to assess the topography of the map. Unusually, the initial bias was in the form of fixing the boundaries of the map. They showed theoretically that this variant of Kohonen's algorithm leads to a topographic map in the one dimensional case (where as stated above topography is measured in the retina-cortex direction), and presented simulation results suggesting that the algorithm also produces a topographic map in the two dimensional case.

3.5 Computational models - ocular dominance

Attempts have been made to extend a number of the models of topography discussed in the previous section to account for ocular dominance stripes with the same mechanisms. We first review these, and then some other models which have been proposed specifically for ocular dominance.

3.5.1 The Neural Activity model

A model similar to the Neural Activity model was applied to the formation of ocular dominance stripes by von der Malsburg in [von der Malsburg & Willshaw 1976]. He assumed that activity in the two presynaptic sheets is anticorrelated between the two eyes: the activity regime in the model consists of activity in one eye *or* the other, but never both eyes simultaneously. He showed that this model is capable of producing ocular dominance stripes, and he argued that the anticorrelation between the two eyes could be produced by inhibition within the LGN. However, there is no physiological evidence such inhibition. This model is somewhat similar to that explored in chapter 6 of this thesis, except that in the model of chapter 6 ocular dominance segregation can occur with distributed patterns of activity presented in both eyes simultaneously.

3.5.2 The Tea Trade model

[von der Malsburg 1979] also proposed a version of the Tea Trade model which produces ocular dominance segregation. This assumes the same chemical markers in the two sheets, plus one extra unique marker for each sheet. These markers become active once the regular markers have set up a smooth map, and segregation occurs. Stripe width is determined by the balance between two factors: a tendency for cells to map to their correct retinotopic locations as in the monocular model, and a tendency for cells from one eye to map close to one another due to the attractive force of the chemical unique to that eye. However, the existence of such a global chemical difference between the eyes appears unlikely in light of the experiments where stripes are formed between fibres from compound eyes [Fawcett & Willshaw 1982], and the isogenic eye experiments [Ide et al 1983]. In addition, there is no evidence for the proposed time course of marker activation.

3.5.3 The models of Cowan and Fraser

The most recent versions of Cowan's model also generate ocular dominance stripes [Cowan & Friedman 1990(b)]. However, segregation only occurs if the correlation between the two eyes is less than or equal to zero (J. Cowan, personal communication).

Ocular dominance stripes can be formed in Fraser's model if it is assumed that constraint (3), the tendency for neighbouring axon terminations in the tectum to stabilize if those axons come from neighbouring positions in the retina, is stronger for terminals from the same eye than for terminals from different eyes [Fraser 1985, Fraser & Perkel 1990]. This could be so if activity within an eye is generally more correlated than activity between the two eyes. While this is reasonable to a degree (though fibres from the two eyes can still be highly correlated if they are from corresponding positions), since this model does not propose a learning rule that can identify such differences in correlation, it is simply equivalent to the assumption of a global label marking one eye as different from the other. Also, there is no consideration of the fact that in the natural system cells are initially binocularly innervated. [Fraser & Perkel 1990] state that stripe width is determined purely by the width of terminal arbors (a fixed parameter of the model), rather than on the balance of the constraints.

3.5.4 Kohonen's algorithm

An application of Kohonen's algorithm to the problem of the simultaneous formation of topography, ocular dominance, and also orientation selectivity, has recently been presented in [Obermayer et al 1990, Ritter et al 1990, Obermayer, Blasdel & Schulten 1991, Obermayer, Ritter & Schulten 1991]. This can be looked at in two ways: either as a biological model based on the presentation of a sequence of images to the retina and the subsequent evolution of synaptic strengths between the retinae and the cortex, or as an abstract optimization model where the aim is to find a "good" mapping between an appropriate feature space and a cortical sheet, without placing emphasis on the biological interpretation of the mechanism used to find the mapping. The first point of view (stressed in [Obermayer et al 1990]) suggests analogies with the model presented in chapter 6 of this thesis, while the second point of view (pervading more recent articles) suggests analogies with the elastic models of chapter 5. We will discuss these analogies in the appropriate chapters: here we describe the basic framework.

The model consists of an array of cortical units receiving input from an array of retinal units onto which patterns of activity are imposed. These patterns are localised, oriented patches, where only one patch in the retina is active at a time. In the binocular case, activity patterns in the two eyes are generally of different strengths but otherwise identical. The cortical units compete for each pattern and are updated according to a neighbourhood function, as in Kohonen's algorithm. The degree to which the stimulus is oriented and ocularly dominant is determined by two parameters, q and z. $q = 0$ corresponds to round stimuli, and as q increases the pattern becomes more elongated (i.e. more sharply tuned to a particular orientation). $z = 0$ corresponds to the pattern being of equal strengths in the two eyes: $z > 0$ represents the pattern stronger in one eye, and $z < 0$ represents the pattern stronger in the other eye. In the simulations presented in [Obermayer, Ritter & Schulten 1991] the inputs were chosen at random positions, but with fixed orientation tuning $q = q_{pat}$ and fixed dominance $|z| = z_{pat}$.

One of the most interesting aspects of this model is that Obermayer et al have been able to calculate critical values for their parameters, which determine when orientation and ocular dominance stripes will be formed. This is related to the issue of *automatic selection of feature dimensions*, which was introduced by Kohonen (see e.g. [Kohonen 1988]) and has been analysed for Kohonen's algorithm in [Ritter & Schulten 1988]. The basic question is to predict, when mapping from a high-dimensional feature space to (for instance) a two-dimensional sheet, which variables from the feature space will be "primary" in the sense of being mapped smoothly, and which variables will be secondary: that is, changing rapidly on a local scale. For instance, in the case of mapping two eyes to the cortex, if space is the primary variable then a pattern of ocular dominance stripes will be formed. However, if space is secondary to ocularity, projections from the two eyes completely segregate such that each eye is mapped over one half of the cortex. In the Obermayer et al model, there are 3 possibilities for ocular dominance which depend on their parameter z_{pat}. If z_{pat} is small, cells do not become dominant (i.e. ocularity is not represented), and if z_{pat} is large the projections become completely segregated into two regions. It is only between these two critical values that stripes form [Obermayer, Ritter & Schulten 1991]. Similar results hold in the case where the dominance of the patterns is chosen uniformly in the range $z \in [-z_{pat}, z_{pat}]$ [Obermayer, Blasdel & Schulten 1991][1]. They also show that stripe width in their model is determined by the width of the cortical neighbourhood function [Obermayer, Blasdel & Schulten 1991].

3.5.5 The elastic net algorithm

The elastic net is a developmental algorithm that has the virtue that there exists an energy function which the algorithm optimizes. It was originally formulated as a simplification of the Tea Trade model specifically for application to combinatorial optimization algorithms which have a geometric interpretation: in particular the Travelling Salesman problem [Durbin & Willshaw 1987]. In [Goodhill 1988, Goodhill & Willshaw 1990] it was applied to the topography and ocular dominance problem and shown to be capable of producing topography and ocular dominance from the same mechanisms. The retinae are represented as two parallel sheets of units lying in an abstract "feature" space. The directions parallel to the two planes represent the spatial dimensions: the third dimension is ocularity. The cortex is represented by a sheet of points in this feature space. Each cortical point is attracted to each retinal unit, and also to its neighbouring cortical points: this latter force attempts to minimize the distances between neighbouring cortical points. The trade-off between these two components serves to establish a map that matches cortical points to retinal units so that the distances between

[1] It has been pointed out by Peter Dayan (personal communication) that there is an inconsistency in the way this model is formulated. Under the abstract optimization interpretation, all feature dimensions (e.g. space and ocular dominance) should be regarded equally. However, when patterns are presented, corresponding positions within the two eyes are allowed to be active at the same time, but not widely separated positions in the two eyes.

neighbouring cortical points as measured in the feature space are kept small. We discuss the elastic net in much greater detail in chapter 5.

We now turn to models designed exclusively for ocularity.

3.5.6 Swindale's model

One of the first mathematical theories specifically addressing the formation of ocular dominance stripes was presented by Swindale [Swindale 1979, Swindale 1980]. His model is expressed in terms of stimulating and inhibiting interactions between neighbouring synapses in the cortex, for which he suggests interpretations both in terms of correlated activity and diffusing chemical substances. In the second case, the model bears similarities with that of [Turing 1952], and could also account for pattern formation in other systems such as zebra or mackerel skin.

The model consists of a cortical array filled with synapses from both eyes that are initially of random strength and uniformly intermingled. A fixed topography is assumed, and only the distribution of synapses rather than individual cells is considered. There are four interaction functions between synapses, relating to the four possible left-eye, right-eye combinations, although in practice Swindale usually reduced this to just "same-eye" and "opposite-eye". Both of these interaction functions are radially symmetric with two regions: a narrow inner one and a broader outer one. For synapses from the same eye, the inner region is excitatory and the outer region inhibitory, and vice-versa for synapses from opposite eyes. It is assumed therefore that it is possible to distinguish fibres from the two eyes (for instance, on the basis of correlated activity). At each time step, the effects on each synapse are summed and the strength of that synapse increased or decreased according to whether the sum is positive or negative. Synaptic strengths are bounded between upper and lower limits. Various forms for the interaction function were investigated, and (providing they were of centre-surround type) the results were found to be reasonably insensitive to this.

This mechanism produces results that share a number of the morphological features of ocular dominance stripes as described in chapter 2: narrowing, branching, and perpendicular intersection with the boundaries. Although the stripes twist and turn more than in the natural system, a more regular global orientation can be produced by assuming growth of the cortex during development [Swindale 1980]. Narrowing of stripes in response to monocular deprivation can be produced by the model if it is assumed that deprivation corresponds to a reduction in the strength of the interaction functions for one eye. Qualitative arguments are presented based on the form of the interaction functions which suggest that the stable patterns should indeed be stripes. A similar model has also been formulated to account for the pattern of orientation columns seen in V1 [Swindale 1982], and more recently for the interaction between orientation columns and ocular dominance columns [Swindale 1992].

3.5.7 The model of Bienenstock

An influential model of cortical receptive field development is that of [Bienenstock et al 1982, Bienenstock 1980] (for a more recent discussion, including a possible physiological implementation of this algorithm, see [Bear & Cooper 1990]). Although primarily a theory for the development of orientation selectivity of isolated cortical cells, it is relevant here since the development of binocularly innervated cells is also considered. Weights to a cortical cell evolve according to a Hebbian rule, driven by input patterns randomly chosen from an "environment" of patterns. The unusual feature of the particular rule they use is the way that decreases in the weights are specified, which is a different method from that of normalization. Firstly, the learning rule says that in the absence of input, weights decay exponentially to zero. More importantly, they introduce a threshold for the output of the cell such that below this threshold all weights decrease. This threshold is chosen to be a nonlinear function of the response of the cell. Bienenstock at al showed that for certain environments of oriented patterns, weights converge to a stable state that is orientationally selective. With respect to ocular dominance, they showed the following results. (1) If inputs to the two eyes are perfectly correlated (which they referred to as "normal rearing"), cells are binocular. (2) Monocular cells develop under conditions of "binocular deprivation" (constant vector presented to both eyes, but with random variation in relative strength between the eyes), "monocular deprivation", and "uncorrelated rearing" (inputs in the two eyes completely uncorrelated). Again we have the problem of explaining why cells become monocular even when the two sets of inputs are positively correlated.

Model	Principal mechanisms	Topography		Ocular dominance		
		Refinement	Map	Monocularity	Stripes	Principal determiner of stripe width
Prestige and Willshaw	Group 1 and group 2 chemospecificity	-	Yes	-	-	-
Neural Activity	Activity	Yes	Yes	Yes if activity nonsimultaneous in two eyes		Width of cortical interaction function
Tea Trade	Induction of retinal markers	Yes	Yes	Yes if extra marker is introduced for one eye		Balance of chemical forces
Cowan	Group 2 chemospecificity and activity	Yes	Yes	Yes if no correlations between eyes		?
Fraser	Multiple	-	Yes	-	Yes	Width of terminal arbor
Amari	Activity	Yes	Yes	-	-	-
Swindale	Activity	Fixed		-	Yes	Width of cortical interaction function
Bienenstock	Activity	-	-	Yes	-	-
Miller	Activity	Fixed		See chapter 4		Width of cortical interaction function
Elastic net	See Chapter 5	(Yes)	Yes	(Yes)	Yes	Ratio of correlations within and between eyes
Obermayer et al	Activity	Yes	Yes	Yes under certain conditions		Width of cortical neighbourhood function

Table 3.1: Summary of models. We indicate for each of the models discussed (1) the main mechanisms underlying the model, (2) whether the model accounts for the formation of topography, and also provides a mechanism whereby an initially crude topographic map can refine, (3) whether the model explains how individual cells can become monocular, and how these monocular cells then come to form a pattern of stripes, (4) The main determiner of stripe width in the model. Dashes indicate that this is outside the scope of the model as originally presented. The entries for the elastic net should be read in the light of the discussion in chapter 5 of this thesis. There is a question mark for stripe width in Cowan's model since this has not been discussed by Cowan.

3.5.8 Miller's model

[Miller et al 1989] proposed a model of ocular dominance stripe formation based on a "correlational" learning rule similar to that of [Linsker 1986], though the two models were developed independently (see also [Miller 1989, Miller 1990(a), Miller 1990(b), Miller & Stryker 1990]). Given certain assumptions about the input-output function of postsynaptic units and a simple mathematical implementation of the Hebb rule, Miller derived an expression for the rate of change of weights that depends only on the correlations between presynaptic units rather than between pre- and postsynaptic units. In particular, the behaviour of the system is determined largely by the eigenvalues and eigenvectors of the matrix of time-averaged correlations between presynaptic units. Miller assumed that the strength of these correlations drop off as a gaussian function of distance in each presynaptic sheet (representing the LGN), but that there are zero correlations between units in the two presynaptic sheets. Fixed lateral connections exist in the postsynaptic sheet (representing the cortex), generally of a short-range excitatory and long-range inhibitory form. At each time step, activity is iterated to a stable state in the postsynaptic sheet and then synaptic strengths are updated according to this pattern of activity and the presynaptic correlation matrix. Stripe width is determined largely by the width of the cortical interaction function. We discuss this model in much greater detail in chapter 4 of this thesis.

3.6 Conclusions

We have briefly reviewed the most important models of the development of topography, and ocular dominance, in the retinotectal and retinocortical systems. The properties of each model are summarized in table 3.1. It is clear that there are a number of different types of model, which depend on different types of mechanisms. The Tea Trade model, Cowan's model and Fraser's model are essentially in competition, since they tackle roughly the same set of experimental data. The Tea Trade model is the most elegant, Cowan's is the most detailed, and Fraser's, being based on the direct optimization of an objective function, has the clearest computational level description. By contrast, [Prestige & Willshaw 1975] and the Neural Activity model are complementary:

instead of attempting to account for a wide range of data, they each take a single mechanism and explore its capabilities. The same can be said of Bienenstock's and Swindale's models, except that the former is concerned only with the ocular dominance of single cells, and the latter only with the laying out of two projections in a pattern of stripes. Miller's is a more detailed model, that nonetheless has the virtue of yielding to extensive mathematical analysis. The models deriving from Kohonen are more algorithmic in nature, and are thus not directly competing with the more directly biological models such as the Neural Activity model and the Tea Trade model.

Strong analogies can be made between models based on chemical labelling and those based on activity labelling. Both specify that axons of presynaptic units are "similar" (in the sense of expressing similar chemical identities or having correlated activity) if they originate from nearby cells in the presynaptic sheet. The Tea Trade model is related to activity-based models in that they each specify some mechanism for information to be transmitted between postsynaptic units during map formation about which presynaptic axons connect to them. This is by diffusion of presynaptic markers in the postsynaptic sheet in the Tea Trade model, and lateral connections between neighbouring postsynaptic units in for instance the Neural Activity or Miller's model. Such models require at least some initial polarity to generate global order. Models such as Prestige and Willshaw, Cowan, and Fraser are somewhat different in that they assume fixed chemical gradients in the postsynaptic sheet, and communication between postsynaptic units of the connections they receive is not required. Such fixed gradients can produce global order.

To what extent do the models we have described fit into the criteria listed in the introduction? (We do not include Miller's and the elastic net model here, since they will be discussed in much greater detail later.) The first point is the development of topography and ocular dominance from the same mechanisms. This is difficult to judge for models such as Swindale's and Bienenstock's, since these were formulated exclusively for ocularity. It has not been shown that similar mechanisms could *not* also produce a topographic map: the issue has not been addressed. Regarding biological relevance, for extensions of models of topography we have indicated certain assumptions with relation to ocular dominance in each of the models which appear to be biologically implausible. Concerning the issue of correlations, the binocular versions of the Neural Activity model and Cowan's model require non-positive correlations between the eyes. The stripe morphology produced by the models is again hard to judge, since there are no generally accepted ways of assessing this.

It is apparent from our discussion that up to now, ocular dominance has often been regarded either as a separate phenomenon from the development of topography, or as an adjunct to theories of map formation. This has led to the development of models for ocular dominance that are strongly wedded to a pre-existing topography without showing how this could arise from the same mechanisms, and models for topography which consider ocular dominance only as an afterthought. Although a great amount of progress has been made by making clear distinctions between various phenomena to be explained, and between the mechanisms employed to explain them, we prefer a different perspective motivated by the experimental evidence reviewed in chapter 2. We propose that topography and ocular dominance should be regarded as equally important aspects of the functioning of the same basic mechanisms. This leads us to develop new models where from the outset topography and ocular dominance are considered equal partners, with a model considered inadequate unless it can naturally encompass both phenomena.

Chapter 4

Correlational Models

4.1 Introduction

Two of the most influential recent models of the development of the visual system are those of [Linsker 1986] (see also [Linsker 1988]) and [Miller et al 1989]. Linsker considered the maturation of a number of layers of units with weights between successive layers, modelling very roughly the layers in the visual pathway from one eye to V1. Miller considered a two-layer binocular system of two presynaptic sheets (which he identified with the lateral geniculate nuclei) innervating a single postsynaptic sheet (the cortex). Linsker's primary concern was the development of orientation selectivity given both excitatory and inhibitory weights, whereas Miller's was the development of ocular dominance given two purely excitatory projections. However, though developed independently, there is a close relationship between these models. In fact, Miller's model (in its simplest form) is the natural way of expressing Linsker's model in the binocular case. Each is based on a type of Hebbian learning that Miller refers to as a *correlational* rule: the form of receptive field development is determined by the correlations between presynaptic units. These correlational models have generated much interest, partly because the form of the learning rule makes them susceptible to extensive mathematical analysis. It is thus important for us to attempt to make clear the behaviour of these models (particularly Miller's binocular model) with regard to their position along the dimensions of interest identified in the introduction.

Miller argues that the important parameters of his model are both biologically meaningful and experimentally measurable in the natural system. His model produces good stripe morphology, and can account for the effects of monocular deprivation (although we will return to this point later). Thus he states "In summary, a simple correlation-based plasticity rule, combined with a very simple model of geniculate correlations, geniculocortical connectivity, and cortical connectivity or interactions, is sufficient to account for a large part of the observed phenomenology of ocular dominance segregation" [Miller 1990(a)].

Miller and Linsker both assume a fixed topography in their models. We leave further discussion of whether they could produce topography to the Conclusions chapter. The purpose of this chapter is to investigate to what extent the form of correlations within and between eyes influences the development of monocularity in Miller's model. In his analysis and simulations, Miller assumed that activity in the two eyes is described by identical correlation functions. In addition, in his simulations he usually assumed that there was zero or negative correlation between the eyes. We start by reviewing Linsker's model of receptive field development, the recent eigenvector analysis of this model [MacKay & Miller 1990(a), MacKay & Miller 1990(b)], and its relationship to Miller's model for the development of ocular dominance. We then analyse a more general situation: the addition of small correlations that are different in the two eyes, and of small correlations between the two eyes. We do this by means of a perturbation analysis of the general two-eye correlation matrix. A preliminary account of this work is in [Dayan & Goodhill 1992].

4.1.1 The assumptions behind correlational models

The most important assumptions underlying the models of Linsker and Miller are as follows (for a derivation of rules in this form by the simplification of general Hebbian rules, see [Miller 1990(b)]):

- The output of a post-synaptic unit is a linear weighted sum of the activity on each input line multiplied by the weight on that line.
- During learning, each weight is updated in proportion to the product of pre- and post-synaptic activity on that line.
- Learning is sufficiently slow that we can assume changes in the weights are insignificant over the time taken to present a subset of the input patterns large enough to accurately represent their internal statistics.

These three assumptions allow us to express the learning rule in terms of the covariance matrix of the input activities, which can then be analysed in terms of its eigenvectors. These are the weight structures that develop independently in the system, and govern the ultimately stable weight pattern. Since the behaviour of correlational models is to a large extent governed by the eigenvectors of the covariance matrix, they bear a close relationship with Principal Components Analysis. However, it can be the case in correlational models that finite saturation limits lead to subprincipal eigenvectors determining the final weight pattern [MacKay & Miller 1990(a), MacKay & Miller 1990(b)].

It is instructive to compare these assumptions with those of the Neural Activity model [Willshaw & von der Malsburg 1976] as discussed in chapter 3. Although in the Neural Activity model the *input* to a postsynaptic unit is a linear weighted sum of input activities, there is a threshold non-linearity in the response of units.

Learning also occurs in the Neural Activity model by a Hebbian mechanism, but weights are updated every time an input pattern is presented: it is not assumed that learning is much slower than the rate of pattern presentation. These differences mean that the Neural Activity model does not succumb to the same type of analysis as the correlational models of Miller and Linsker.

4.2 Mathematics of correlational models

4.2.1 Linsker's model

[Linsker 1986] considered a system of several layers of units that mature sequentially, with feed-forward connections only between adjacent layers. There are no connections between units in the same layer (except for the last layer which we will not consider), so that each unit develops independently. A predefined topography is assumed: each unit in layer $\mathcal{M}$ is connected to units in the previous layer $\mathcal{L}$ with a probability given by a gaussian function of the distance in $\mathcal{L}$ from the topographically matching position. To speed computation, on occasion Linsker used a simpler version of this gaussian spread of randomly located synapses: he assumed a regular grid of 'representative' synapses, the strength of each scaled by a gaussian arbor function. This is equivalent to the first case when there are enough synapses so that random fluctuations are not significant. However, it complicates the analysis slightly, since the covariance matrix Q is now multiplied by the arbor matrix A, and the product QA is not in general symmetric.

Linsker's equation is now derived (following [MacKay & Miller 1990(b)]). Consider a single unit in layer $\mathcal{M}$ which has connections of strength w_i to a group of units in layer $\mathcal{L}$ with activities $x_i^{\mathcal{L}}$. The activity of this unit in $\mathcal{M}$, $x^{\mathcal{M}}$, is calculated as

$$x^{\mathcal{M}} = \sum_{i \in \mathcal{L}} w_i x_i^{\mathcal{L}} + a_1 \tag{4.1}$$

where a_1 is a constant. Linsker assumed that a_1, and all other constants he introduced, are the same for all units within a particular layer. He used the following Hebbian rule for weight-updating (see the second assumption above):

$$\dot{w}_i = (x^{\mathcal{M}} - c_2)(x_i^{\mathcal{L}} - c_1) + c_3 \tag{4.2}$$

where c_1, c_2 and c_3 are constants. We can now substitute for $x^{\mathcal{M}}$ from (4.1) into (4.2):

$$\dot{w}_i = \left(\sum_{j \in \mathcal{L}} w_j x_j^{\mathcal{L}} + a_1 - c_2 \right) (x_i^{\mathcal{L}} - c_1) + c_3$$

From the third assumption (see page 36), we can average over the input patterns $\mathbf{x}^{\mathcal{L}}$ to obtain an equation for the average rate of change of w_i, in terms of the mean activities $\overline{x}_i^{\mathcal{L}}$ in layer $\mathcal{L}$ and the covariance matrix $Q_{ij}^{\mathcal{L}} \equiv \langle (x_i^{\mathcal{L}} - \overline{x}_i^{\mathcal{L}})(x_j^{\mathcal{L}} - \overline{x}_j^{\mathcal{L}}) \rangle$, where angle brackets denote a time average. If we also assume that the first-order statistics of the patterns in layer $\mathcal{L}$ are uniform, i.e. $\overline{x}_i^{\mathcal{L}} = \overline{x}^{\mathcal{L}}$ $\forall i$, we can combine constants to obtain the simple equation

$$\dot{w}_i = \sum_{j \in \mathcal{L}} (Q_{ij}^{\mathcal{L}} + k_2) w_j + k_1$$

or in vector form

$$\dot{\mathbf{w}} = (Q + k_2 J)\mathbf{w} + k_1 \mathbf{n} \tag{4.3}$$

where J is the matrix $J_{ij} = 1$, $\forall\, i, j$, and $\mathbf{n}$ is the 'DC' vector $n_i = 1$, $\forall\, i$. Note that this is slightly different from Linsker's original equation, due to a rescaling of the constants k_1 and k_2 to avoid implicit dependence on the number of synapses appearing in (4.3).

The stability of this equation depends on the eigenvalues of the matrix $Q + k_2 J$. The fixed point $\mathbf{w}_{FP}$ of the equation is given by

$$\mathbf{w}_{FP} = -k_1 (Q + k_2 J)^{-1} \mathbf{n},$$

and it can be shown (e.g. [MacKay & Miller 1990(b)]) that each eigenvector of the matrix $Q + k_2 J$ is a weight structure that grows exponentially relative to the fixed point, with rate given by the appropriate eigenvalue. Since $Q + k_2 J$ is symmetric, it has a complete orthonormal set of eigenvectors with real eigenvalues (see e.g. [Horn & Johnson 1985]). The fixed point is unstable if there exist eigenvectors with positive eigenvalues, and in this case the eigenvector with the largest eigenvalue will dominate as $t \to \infty$. In order to keep the weights within bounds, Linsker used saturation: once weights reach upper or lower limits they are 'frozen' and no further change is allowed. He used two forms of saturation rule. The first assumed that there exist

two statistically indivisible populations of units in layer $\mathcal{L}$, one excitatory and one inhibitory. The saturation limits for each unit were then taken to be 0 and 1. This was later replaced with a single type of unit, and weights that could take on both positive and negative values. This is formally equivalent, though somewhat less biologically plausible, given Dale's law that cells in the natural system are either excitatory *or* inhibitory. Usually he considered weights in the range $[-0.5, 0.5]$. It can be shown [Linsker 1986] that all, or all but one, of the weights saturate at the upper or lower limits.

The regime of activity considered was uncorrelated random firing in the first layer $\mathcal{A}$. This leads to gaussian correlations in layer $\mathcal{B}$ (due to the assumption of a gaussian synaptic density function) and for each subsequent layer the correlations can be calculated for that layer once the weights have matured.

Linsker reported (for carefully chosen parameter values that varied from layer to layer) all-excitatory or all-inhibitory receptive fields in layer $\mathcal{B}$ (the second layer), centre-surround units in layers $\mathcal{C}$ to $\mathcal{F}$, and eventually oriented units in layer $\mathcal{G}$. Oriented units can in fact be obtained in layer $\mathcal{C}$ for appropriate parameter values, but these units are less robust than those that appear in layer $\mathcal{G}$.

4.2.2 The eigenvector analysis

In a Hebbian rule such as (4.3), without any constraints, the eigenvector with the largest positive eigenvalue will eventually dominate. It is thus clear that an eigendecomposition of the covariance matrix can reveal much about the development of the system. However, matters are complicated by the introduction of saturation constraints. Firstly, non-linear phenomena may now occur once the weights start to reach the saturation limits, and secondly the initial conditions may give some eigenvectors a 'head start' such that, despite not growing as fast as other eigenvectors, they are still the first to reach the saturation limits. [MacKay & Miller 1990(a), MacKay & Miller 1990(b)] performed the eigendecomposition of Linsker's covariance matrix $Q + k_2 J$, and analysed under what circumstances the principal eigenvector determined the final weight structure. This is somewhat similar to the analysis of the Tea Trade model performed by [Häussler & von der Malsburg 1983], except that Häussler and von der Malsburg also consider the effect of higher-order terms, which Mackay and Miller do not (see also [Turing 1952]). Mackay and Miller made the following important assumption:

> The principal features of the dynamics are established before the hard limits are reached. When the hypercube is reached, it captures and preserves the existing weight structure with little subsequent change. [MacKay & Miller 1990(b)].

Thus they ignored the possibility of non-linear effects after weight saturation.

They used the standard wave-vector notation of Quantum Mechanics (see e.g. [Schiff 1968]) to describe the eigenvectors. Under most circumstances, they found that the following receptive field structures are the principal eigenvectors:

- 1s: An all-excitatory or all-inhibitory receptive field, which has the largest possible DC component of all the eigenvectors.
- 2s: A centre-surround receptive field: a core of positive (negative) components surrounded by a ring of negative (positive) components. The sum of the components is not generally equal to zero, and thus this structure has a medium-sized DC component.
- 2p: An oriented receptive field: a region of positive components separated from a region of negative components by a straight line. The two regions are equally balanced, and thus this structure has zero DC component.

In this chapter we label 2p structures as $\mathbf{x}_i$, and for the moment do not distinguish between the eigenvectors with non-zero DC, labelling them as $\mathbf{y}_i$. Mackay and Miller found that the three varieties of eigenvector listed above are usually the three principal components of Linsker's Q matrix. However, whether the eigenvector with the highest eigenvalue wins depends on the parameters k_1 and k_2:

- k_1 does not affect the eigenvalues, but biases the competition between zero DC and non-zero DC eigenvectors. A positive (negative) value of k_1 gives a head start to eigenvectors with positive (negative) DC component.
- k_2 determines which is the principal eigenvector. Modes $\mathbf{x}$ with zero DC (i.e. $\mathbf{x}.\mathbf{n} = 0$) have eigenvalues that are constant with k_2, while modes $\mathbf{y}$ with non-zero DC (i.e. $\mathbf{y}.\mathbf{n} \neq 0$) have eigenvalues that increase monotonically with k_2.

The effect of this for various combinations of k_1 and k_2 is shown in [MacKay & Miller 1990(a), MacKay & Miller 1990(b)].

4.3 Extension to two eyes

We now consider the natural extension of Linsker's equations to the case of two eyes, or more generally two populations (for instance on-off and off-on centre-surround cells: for concreteness we will express results in terms of two eyes). This leads to a system more general than that studied by Miller: it is shown that Miller's model can be regarded as a special case of this. We consider a two-layer system, which in general will correspond to Linsker's $\mathcal{B}$ to $\mathcal{C}$ connections, with gaussian correlations in layer $\mathcal{B}$ (the "input" layer).

A correlational rule for two populations can now be found in the same way as the monocular derivation given above:

$$\begin{pmatrix} \dot{w}_L \\ \dot{w}_R \end{pmatrix} = \begin{pmatrix} Q_L + k_2 J & Q_C + k_2 J \\ Q_C + k_2 J & Q_R + k_2 J \end{pmatrix} \begin{pmatrix} w_L \\ w_R \end{pmatrix} + k_1 \begin{pmatrix} n \\ n \end{pmatrix} \tag{4.4}$$

where Q_L gives the covariance between units within the left eye, Q_R gives that between units within the right eye, and Q_C the covariance between units in the two eyes. We assume Q_C is symmetric: this is equivalent to saying that the left and right eyes see the same set of patterns. Define Q_* as the full two eye iteration matrix. We constrain all weights to be positive, for reasons that will be apparent later.

4.3.1 Miller's special case

The above equation with $k_1 = k_2 = 0$ is the basis of Miller's model, as we now show. His equation for the development of the left-eye weights is:

$$\begin{aligned} \dot{w}^L(x,\alpha,t) &= \lambda A(x-\alpha) \sum_{y,\beta} I(x-y)[C^{LL}(\alpha-\beta)w^L(y,\beta,t) + C^{LR}(\alpha-\beta)w^R(y,\beta,t)] \\ &\quad - \gamma w^L(x,\alpha,t) - \epsilon^L A(x-\alpha) \end{aligned}$$

where α, β index over retinal units and x, y index over cortical units, A is the arbor function, I is related to the cortical interaction function, C^{LL} and C^{LR} are the correlation functions within the left eye and between the left and right eyes respectively, and the last two terms are decay terms (which Miller usually took to be zero). There is an analogous equation for the right eye weights.

If we consider just the development of monocularity for an individual cortical unit, we can ignore the cortical interaction function. Ignoring also the arbor function (which in practice appears to have little effect) and setting the decay terms to zero yields

$$\begin{aligned} \dot{w}^L(\alpha,t) &= \lambda \sum_{\beta} [C^{LL}(\alpha-\beta)w^L(\beta,t) + C^{LR}(\alpha-\beta)w^R(\beta,t)] \\ \dot{w}^R(\alpha,t) &= \lambda \sum_{\beta} [C^{RR}(\alpha-\beta)w^R(\beta,t) + C^{RL}(\alpha-\beta)w^L(\beta,t)] \end{aligned}$$

Or, in matrix notation (taking $\lambda = 1$),

$$\begin{aligned} \dot{w}^L &= C^{LL}w^L + C^{LR}w^R \\ \dot{w}^R &= C^{RR}w^R + C^{RL}w^L \end{aligned}$$

These are the same equations as expressed in (4.4), with k_1 and k_2 equal to zero. The absence of k_2 is equivalent to assuming that the average activity of each input unit is zero [Miller 1990(a)].

These equations are coupled. However, Miller assumed that $C^{LL} = C^{RR}$ and $C^{LR} = C^{RL}$ (similarly to [Swindale 1980]), and was thus able to separate the equations into sum ($S^S = w^L + w^R$) and difference ($S^D = w^L - w^R$) variables, which evolve independently. S^D describes the development of ocular dominance, and Miller showed that with saturation, bottom-limited at zero, and a subtractive normalization rule to limit synaptic strengths (an issue to which we return later), the patterns of dominance between the two eyes are determined by the initial weight values and the fastest-growing components of S^D. The starting condition Miller adopts has $w^L - w^R = \epsilon' p$ where ϵ' is small, and $w^L + w^R = q$, which is $\mathcal{O}(1)$.

4.3.2 The eigendecomposition of Miller's matrix

In his simulations, Miller usually took $Q_C = 0$, i.e. the case of no correlations between the two eyes. Using this and the assumption that $Q_R = Q_L = Q$, we can calculate the eigendecomposition of the resulting matrix Q_*^m in

E-vector	E-value	Conditions	
(x_i, x_i)	λ_i	$Qx_i = \lambda_i x_i$	$x_i.n = 0$
$(x_i, -x_i)$	λ_i	$Qx_i = \lambda_i x_i$	$x_i.n = 0$
$(y_i, -y_i)$	μ_i	$Qy_i = \mu_i y_i$	$y_i.n \neq 0$
(z_i, z_i)	ν_i	$(Q + 2k_2 J)z_i = \nu_i z_i$	$z_i.n \neq 0$

Table 4.1: The eigendecomposition of Q_*^m.

the same way that Mackay and Miller analysed the eigendecomposition of the one-eye matrix $Q + k_2 J$ (see table 4.1).[1] There are two basic types of eigenvectors: $e_1 = (u, u)$ and $e_2 = (u, -u)$. These can be further subdivided depending on the form of u. If u has zero DC, then both e_1 and e_2 correspond to "striped" binocular patterns (oscillation between eyes occurring within a receptive field): we refer to these as x. In this case e_1 and e_2 are degenerate, and thus any linear combination of them is also an eigenvector of Q_*^m. We see below how the introduction of perturbative correlations can break this degeneracy. An eigenvector u with non-zero DC (y and z) is also binocular if u still has both positive and negative components (e.g. a 2s centre-surround pattern). The only eigenvector that corresponds to a monocular pattern is e_2 when u is entirely DC (i.e. 1s). Note that since weights are positive in Miller's system, the dominance of e_2 for u entirely DC leads to a final weight pattern of $(u, 0)$, i.e. monocular. Monocularity could not develop in the context of Linsker's saturation limits even if eigenvectors of the form $(u, 0)$ existed, because growth of sub-principal eigenvectors always ensures that all weights eventually saturate at their limiting values. Thus referring to table 4.1 we see that the only monocular solution is $(y_i, -y_i)$ for i such that y_i is 1s. [Miller 1990(a), pages 296–300] compares empirically the development of various eigenvectors, including circumstances under which the monocular eigenvector does and does not dominate.

4.4 Perturbation analysis

We now return to the complete two eye equation (4.4), and analyse a more general case than that considered by Miller. In particular we do not assume $Q_L = Q_R$: Miller's device of decomposing equation (4.4) into sum and difference variables is no longer valid, and more general methods are required. We also do not assume that $k_1 = k_2 = 0$. We perform the eigendecomposition of the general matrix Q_* with the assumption that $Q_L - Q_R$ and Q_C are small. This amounts to a perturbation analysis of Miller's special case, and serves two functions:

- It provides a step in the direction of understanding the general two eye system.
- It explains some of the empirical results found by Miller, reveals problems of stability in his model, and predicts the effects of positive correlations between eyes.

Let $Q_L = Q + \epsilon Q_1$, $Q_R = Q + \epsilon Q_2$ and $Q_C = \epsilon Q_c$, and call the resulting matrix Q_*^ϵ:

$$Q_*^\epsilon = \begin{pmatrix} Q + \epsilon Q_1 + k_2 J & \epsilon Q_c + k_2 J \\ \epsilon Q_c + k_2 J & Q + \epsilon Q_2 + k_2 J \end{pmatrix} \tag{4.5}$$

It is now of interest to calculate the effects of these perturbations on the eigenvectors and eigenvalues, particularly with regard to the following points:

- Different eigenvectors may be perturbed by different amounts, possibly changing the nature of the competition between them.
- Some eigenvectors may not be stable to the perturbation, *i.e.* if (u_1, u_2) is an eigenvector of Q_* it may not be the case that $(u_1 + \epsilon a_1, u_2 + \epsilon a_2)$ is an eigenvector of Q_*^ϵ for some vectors a_1 and a_2.
- Previously degenerate eigenvectors may split, implying one is more favoured in the race to saturation. This phenomenon is well known in Quantum Mechanics (see e.g. [Schiff 1968]).

[1] We do not describe the case of degeneracy between modes of different index i. This is for notational convenience, and does not affect the results. Also for convenience we write (x_i, x_i) for $(x_i, x_i)^T$, *etc.*

To calculate the eigenvalues and eigenvectors of the perturbed system, we consider the following equation:

$$\begin{pmatrix} Q + \epsilon Q_1 + k_2 J & \epsilon Q_c + k_2 J \\ \epsilon Q_c + k_2 J & Q + \epsilon Q_2 + k_2 J \end{pmatrix} \begin{pmatrix} \alpha_i u_i + \epsilon a_i \\ \beta_i u_i + \epsilon b_i \end{pmatrix} = (\lambda + \epsilon \mu_i) \begin{pmatrix} \alpha_i u_i + \epsilon a_i \\ \beta_i u_i + \epsilon b_i \end{pmatrix} \tag{4.6}$$

where u_i is a general eigenvector of Q with eigenvector λ. That is, we look for solutions of the form $(\alpha_i u_i + \epsilon a_i, \beta_i u_i + \epsilon b_i)$. Consideration of zero-order terms gives the results in table 4.1: thus u_i is an eigenvector of Q. Calculating terms of $\mathcal{O}(\epsilon)$ gives the following two equations (where for clarity we drop the subscript i):

$$(Q - \lambda I)a + k_2 J(a + b) = -(\alpha Q_1 + \beta Q_c - \alpha \mu I)u \tag{4.7}$$

$$(Q - \lambda I)b + k_2 J(a + b) = -(\beta Q_2 + \alpha Q_c - \beta \mu I)u \tag{4.8}$$

We now use the fact that since u is an eigenvector of Q, $(Q - \lambda I)p$ can have no components in the direction of u for any p.(This becomes clear if one imagines p expressed in the eigenvector basis: $p = \sum_i a_i u_i$). There are two cases to consider: zero DC and non-zero DC.

Zero DC: $u.n = 0$

For vectors x such that $x.n = 0$, we have $x^T J = 0$ and so multiplying equations (4.7) and (4.8) by x^T makes both left-hand sides vanish, enabling us to write down two expressions for μ. Combining these gives an equation for $\rho = \beta/\alpha$:

$$x^T Q_1 x + \rho x^T Q_c x = x^T Q_2 x + \frac{1}{\rho} x^T Q_c x \tag{4.9}$$

For $Q_c \neq 0$ we define

$$\gamma = \frac{x^T (Q_1 - Q_2) x}{2 x^T Q_c x}$$

and thus obtain from (4.9)

$$\rho = -\gamma \pm \sqrt{1 + \gamma^2}$$

Therefore in the perturbed system with $Q_c \neq 0$, the degeneracy is broken: only particular combinations of zero-DC vectors are allowed, namely $(\alpha x, \beta x)$ and $(-\beta x, \alpha x)$ where the ratio of β/α is as given above. The new eigenvalues are

$$\lambda + \frac{\epsilon}{2} \left\{ x^T (Q_1 + Q_2) x \pm \sqrt{[x^T (Q_1 - Q_2) x]^2 + 4 [x^T Q_c x]^2} \right\}$$

To explain the meaning of this equation we will return to consider some special cases shortly. For $Q_c = 0$ we refer back to equations (4.7) and (4.8) using $\alpha = 0$, $\beta = 1$, and $\alpha = 1$, $\beta = 0$. This yields the same expression for the perturbation as above, except with $Q_c = 0$.

Non-zero DC: $u.n \neq 0$

For non-zero DC vectors, the $k_2 J(a + b)$ terms in (4.7) and (4.8) do not disappear. However, we can find a value for μ for vectors $(y, -y)$ in table 1 by subtracting (4.7) from (4.8) and substituting $\alpha = -\beta$, and likewise for (z, z) by adding (4.7) and (4.8) and substituting $\alpha = \beta$. The results are summarized in table 4.2, where for clarity we have assumed that x_i, y_i and z_i are normalised (and the subscript i has been replaced).

4.5 Special cases

In order to highlight the effects of these perturbations, we now consider some special cases. This also allows contact to be made with some of the results found empirically by Miller.[2]

4.5.1 Cross-eye perturbation only: $Q_1 = Q_2 = 0, Q_c \neq 0$

[2]There is a technical point that we have not so far discussed. The point of our analysis is to reveal under what circumstances the eigenvalues for different eigenvectors cross. However, it is well known in Quantum Mechanics that perturbation theory breaks down when eigenvalues cross, since the second order terms in the perturbation expansion have differences between the eigenvalues on the bottom, and thus blow up when these become equal (see e.g. [Schiff 1968]). It should be emphasised that this does *not* mean that the eigenvalues do not cross. For the Gaussian correlation and perturbation functions, the first order analysis is correct, and simulation results show that the predictions about crossing are also emprically accurate for random perturbations.

E-vector	E-value of Q_*^M	Eigenvalue of Q_*^e
$(\alpha_i x_i, \beta_i x_i)$	λ_i	$\lambda_i + \frac{\epsilon}{2}\left\{x_i^T(Q_1+Q_2)x_i + \sqrt{[x_i^T(Q_1-Q_2)x_i]^2 + 4[x_i^T Q_c x_i]^2}\right\}$
$(-\beta_i x_i, \alpha_i x_i)$	λ_i	$\lambda_i + \frac{\epsilon}{2}\left\{x^T(Q_1+Q_2)x_i - \sqrt{[x_i^T(Q_1-Q_2)x_i]^2 + 4[x_i^T Q_c x_i]^2}\right\}$
$(y_i, -y_i)$	μ_i	$\mu_i + \frac{\epsilon}{2} y_i^T(Q_1+Q_2-2Q_c)y_i$
(z_i, z_i)	ν_i	$\nu_i + \frac{\epsilon}{2} z_i^T(Q_1+Q_2+2Q_c)z_i$

Table 4.2: Effect of perturbations on eigendecomposition of Q_*^e.

E-vector	E-value of Q_*^M	Eigenvalue of $Q_*^e(0,0,Q_c)$
(x_i, x_i)	λ_i	$\lambda_i + \epsilon x_i^T Q_c x_i$
$(x_i, -x_i)$	λ_i	$\lambda_i - \epsilon x_i^T Q_c x_i$
$(y_i, -y_i)$	μ_i	$\mu_i - \epsilon y_i^T Q_c y_i$
(z_i, z_i)	ν_i	$\nu_i + \epsilon z_i^T Q_c z_i$

Table 4.3: Case of $Q_1 = Q_2 = 0$: perturbations only between the eyes.

We will refer to this case of only perturbative correlations between the two eyes as $Q_*^e(0,0,Q_c)$. Results are shown in table 4.3. Firstly, we see that all eigenvectors are stable to this perturbation. Secondly, it should be noted that $x^T Q_c x$ is less for zero DC vectors than non-zero DC vectors if we consider all components of Q_c to be of the same sign. We see from table 4.3 that there is a hierarchy of effects on different eigenvectors. For $Q_c < 0$, the eigenvalue of the monocular $(y_i, -y_i)$ for $y_i = 1$s is increased the most, followed by a smaller increase in the binocular $(x_i, -x_i)$ and $(y_i, -y_i)$ for $y_i \neq 1$s. (x_i, x_i) is decreased slightly, with a bigger decrease in (z_i, z_i). These predictions can be compared with the empirical results of [Miller 1990(a), pages 296–300]. Here, the effect of small-magnitude long-range gaussian anticorrelations between the two eyes on the eigenvalues of (1s,-1s), and (2p,2p) and (2p,-2p), is examined, and it is shown that an increase in eigenvalue for $(y_i, -y_i)$ occurs, while the eigenvalues for $(x_i, x_i), (x_i, -x_i)$ are changed less, as we predict.

The effects of these changes depend critically on the relative sizes of the different eigenvectors: it could be that they are so different that perturbations do not affect the competition, or it could be that they are close enough so that perturbations can select a different winning eigenvector from the unperturbed case. We first show the application of this analysis to some numerical results provided by Miller for the case of anticorrelations between the eyes. This also allows us to predict the effects of *positive* correlations between the eyes on Miller's results. Firstly we see from [Miller 1990(a), page 300] that addition of anticorrelations between the eyes leads to an increase in the eigenvalue for the monocular (1s,-1s) solution, and a lesser increase in the eigenvalue for the (2p,2p) and (2p,-2p) solutions, as we predict. Secondly we see that in his third case (relatively short-range correlations within an eye) the addition of positive correlations between the eyes would be enough to cause binocular rather than monocular development of a unit (since reversing the sign of Q_c is equivalent to subtracting the same amount from the eigenvalue of (1s,-1s) that was added before).

We now present simulation results for the case of gradually increasing positive correlations between the two eyes. Parameters are as follows: number of units in LGN = 12×12, within-eye correlations C_w given by e^{-d^2/σ^2} where d is the Euclidean distance between LGN units and $\sigma = 2.0$, between-eye correlations C_b as for within-eye except with $\sigma = 6.0$. Total correlations are given by $C_w + \epsilon C_b$, where ϵ is the perturbation. Figure 4.1 shows the dependence of the eigenvalues on the perturbation for some of the principal eigenvectors. Figure 4.2 shows the principal eigenvector at each magnitude of perturbation. The following effects can be seen.

- The qualitative forms of the eigenvectors are not affected by the perturbation.
- The eigenvalues are affected by the perturbation, in the direction predicted by our analysis.
- At a certain point the eigenvalues of monocular and binocular solutions cross, implying the development of binocular rather than monocular cells.

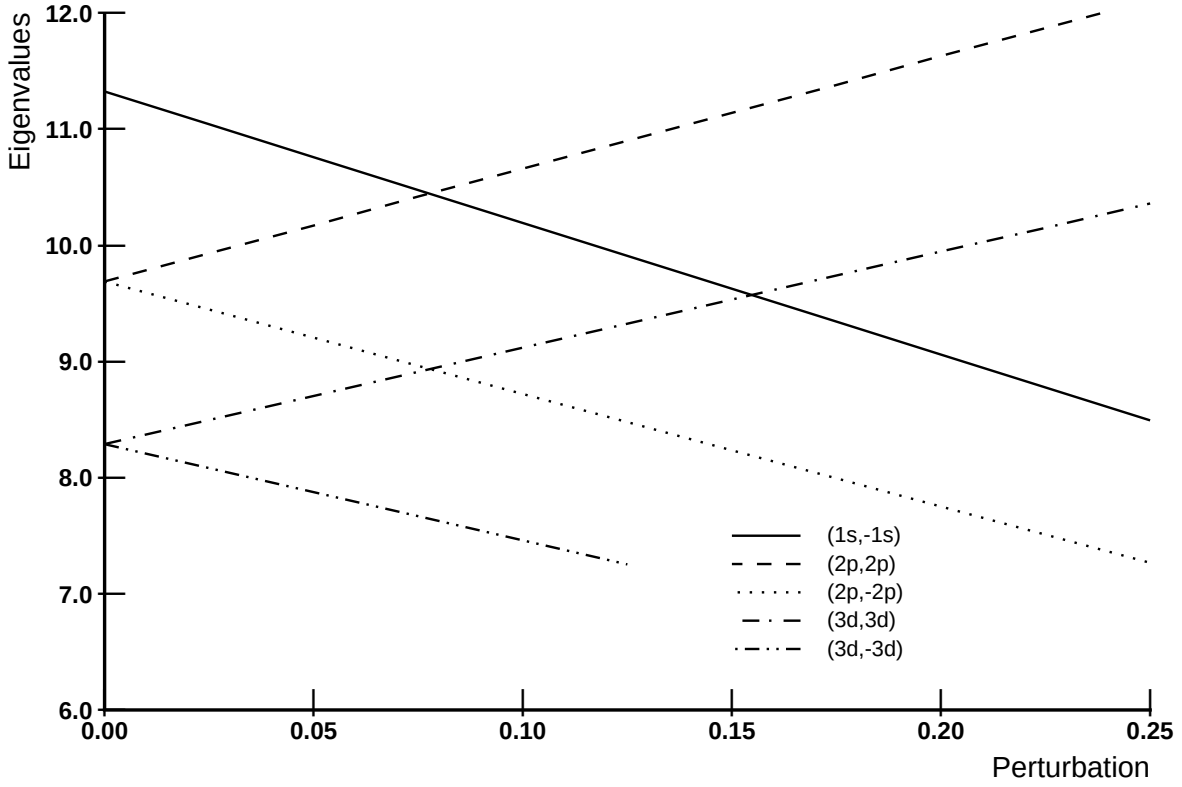

Figure 4.1: Effect of positive cross-eye correlations on the eigenvalues of the principal eigenvectors (not including the (1s,1s) eigenvector, which has largest eigenvalue but is removed by the normalization). Note that the eigenvalues change linearly with the perturbation, and that the eigenvalues cross.

E-vector	E-value of Q_*^M	Eigenvalue of $Q_*^e(0, 0, Q_c)$
(x_i, x_i)	λ_i	$\lambda_i + \epsilon x_i^T Q_s x_i$
$(x_i, -x_i)$	λ_i	$\lambda_i + \epsilon x_i^T Q_s x_i$
$(y_i, -y_i)$	μ_i	$\mu_i + \epsilon y_i^T Q_s y_i$
(z_i, z_i)	ν_i	$\nu_i + \epsilon z_i^T Q_s z_i$

Table 4.4: Case of $Q_1 = Q_2 = Q_s$, $Q_c = 0$: equal perturbations in each eye.

Thus we see that there are conditions under which perturbations can affect the competition, and that increasing positive correlations between the eyes leads to an increasing tendency to binocularity. These findings suggest an analogy with the results of 'artificial strabismus' experiments [Hubel & Wiesel 1965], where it is found that reducing the correlations between the two eyes leads to increased monocularity. However, the analysis above describes only the case of adding small correlations between two uncorrelated eyes, rather than the case investigated by Hubel and Wiesel of reducing the correlations between strongly correlated eyes.

4.5.2 Within-eye perturbation only: $Q_c = 0, Q_1 = Q_2 = Q_s$

Here all eigenvalues are changed by the same factor, $u^T Q_s u$ (see table 4.4). Note that ρ is in this case a free parameter. However, this can still change the competition, since the eigenvalues of the non-zero DC eigenvectors $(y_i, -y_i)$ and (z_i, z_i) will be perturbed to a greater extent. Miller shows empirically [Miller 1990(a), pages 296–300] that addition of anticorrelations within an eye ($Q_s < 0$) can lead to a 'striped' binocular receptive field. This is as we predict, since the eigenvalue of $(y_i, -y_i)$ is decreased more than those of (x_i, x_i) and $(x_i, -x_i)$.

Some simulation results for the case of gradually increasing long range anticorrelations identical within each eye are shown in figure 4.4. Parameters are as above, except that total correlations are now of the form

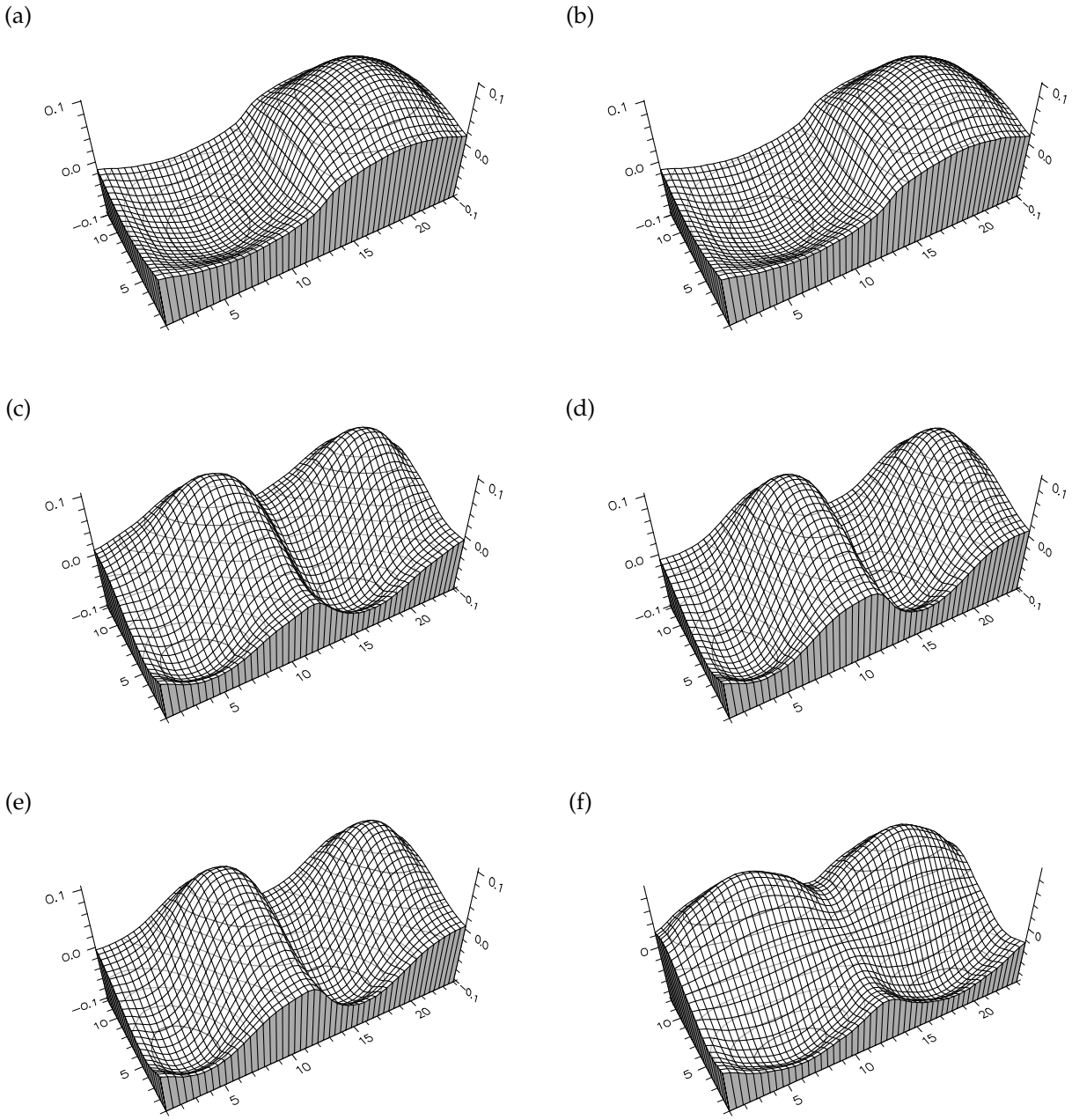

Figure 4.2: Effect of positive cross-eye correlations on the principal eigenvector (not including the (1s,1s) eigenvector, which has largest eigenvalue but is removed by the normalization). (a) No perturbation: the monocular (1s,-1s) dominates. (b) $\epsilon = 0.05$: (1s,-1s) is unchanged and remains dominant. (c) $\epsilon = 0.10$: Now a binocular (2p,2p) structure has taken the lead. (d) $\epsilon = 0.15$: (2p,2p) remains in the lead. (e) $\epsilon = 0.20$: (2p,2p) remains in the lead. (f) $\epsilon = 0.25$: A different binocular (2p,2p) structure takes the lead.

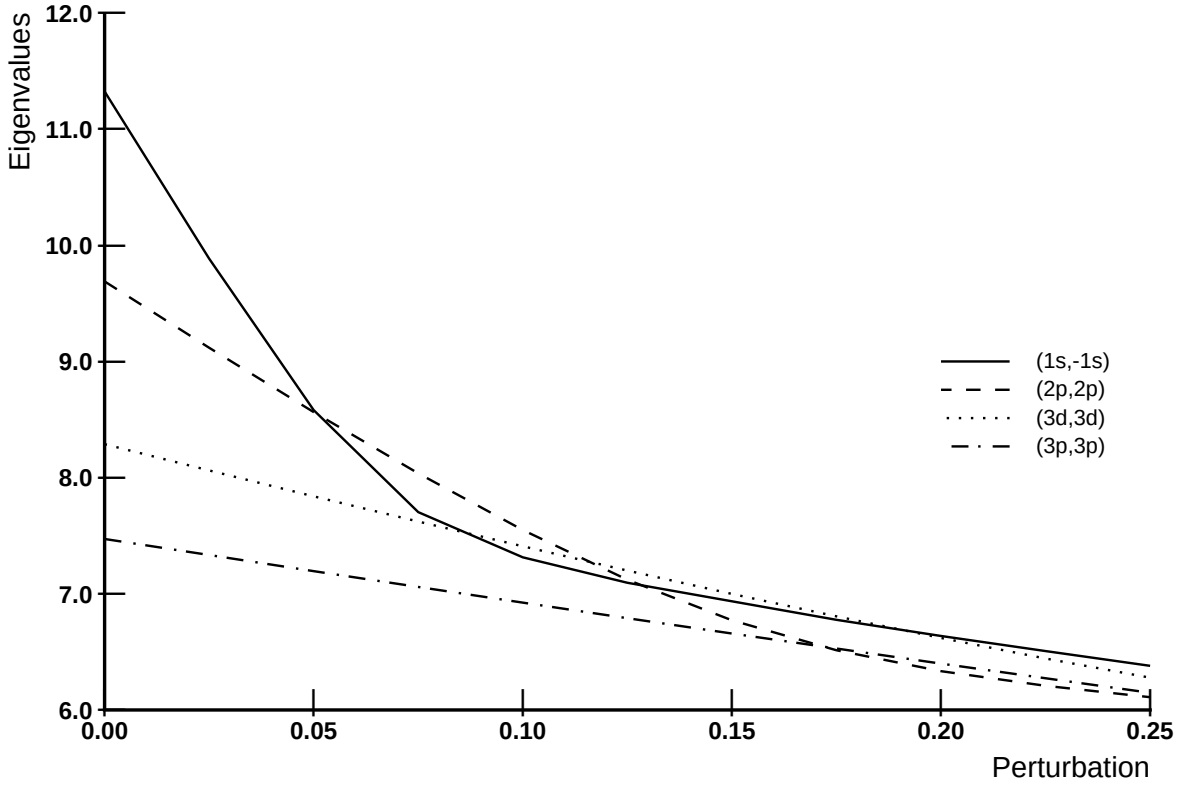

Figure 4.3: Effect of negative correlations within an eye on the eigenvalues of the principal eigenvectors (not including the (1s,1s) eigenvector, which has largest eigenvalue but is removed by the normalization). Note that the eigenvalues change nonlinearly with the perturbation, since the perturbation gradually changes the qualitative form of the eigenvectors, but that eigenvalues still cross.

$C_w + \epsilon C_a$, where C_a represents within-eye anticorrelations with $\sigma = 6.0$. We see that the qualitative forms of the eigenvectors are now continuously distorted by the perturbation. This means that our prediction for the size of the perturbative changes in the eigenvalues will become increasingly inaccurate as the perturbation increases. However, our analysis still correctly predicts the directions and relative sizes of the changes. Again the principal eigenvector becomes binocular for perturbations beyond a certain point. An interesting point to note is that although the eigenvector that was initially of monocular form (1s,-1s) retakes the lead for large perturbations, it has now deformed into roughly (2s,-2s), and is thus no longer monocular.

4.5.3 Different perturbations in each eye: $Q_c = 0, Q_1 \neq Q_2$

The eigendecomposition of $Q_*^e(Q_1, Q_2, 0)$ is given in table 4.5. The most important point to note is that the zero-DC eigenvectors are not stable to the perturbation: they have changed their form from (x_i, x_i) and $(x_i, -x_i)$ to $(x_i, 0)$ and $(0, x_i)$. As explained above for Linsker's system, even if an eigenvector of the form $(x_i, 0)$ was dominant, this would not lead to left-eye dominant monocular units since right-eye weights would grow according to other eigenvectors. Thus just considering $(y_i, -y_i)$ we see that as before within-eye anticorrelations

E-vector	E-value of Q_*^M	Eigenvalue of $Q_*^e(Q_1, Q_2, 0)$
$(x_i, 0)$	λ_i	$\lambda_i + \epsilon x_i^T Q_1 x_i$
$(0, x_i)$	λ_i	$\lambda_i + \epsilon x_i^T Q_2 x_i$
$(y_i, -y_i)$	μ_i	$\mu_i + \epsilon y_i^T (Q_1 + Q_2) y_i / 2$
(z_i, z_i)	ν_i	$\nu_i + \epsilon z_i^T (Q_1 + Q_2) z_i / 2$

Table 4.5: Case of $Q_1 \neq Q_2, Q_c = 0$: different perturbations in each eye.

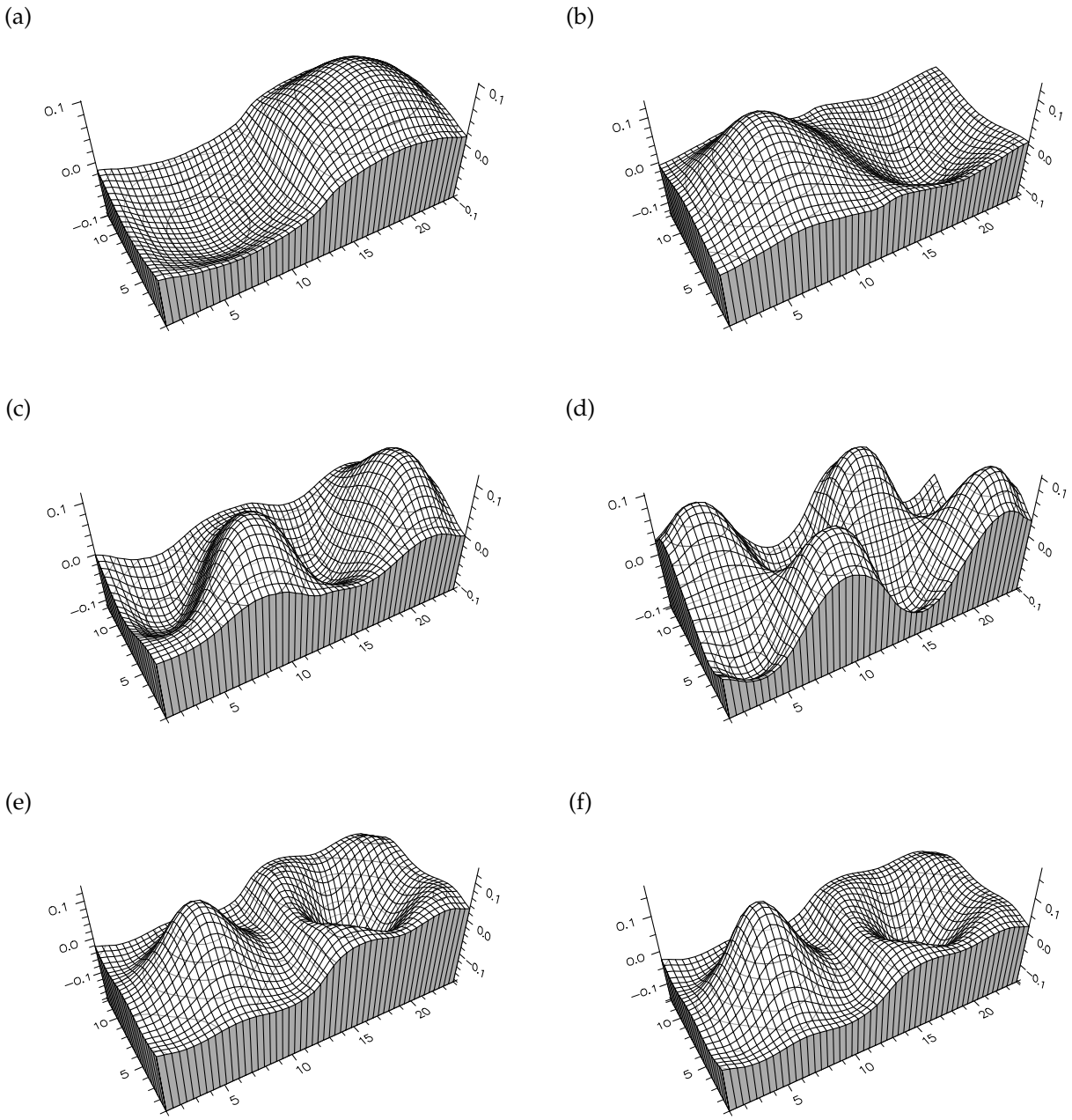

Figure 4.4: Effect of within-eye anticorrelations on the principal eigenvector (not including the (1s,1s) eigenvector, which has largest eigenvalue but is removed by the normalization). (a) No perturbation: the monocular (1s,-1s) dominates. (b) $\epsilon = 0.05$: (1s,-1s) remains dominant, but has deformed slightly (monocularity has now flipped to the other eye, as can happen since (1s,-1s) is degenerate with (-1s,1s)). (c) $\epsilon = 0.10$: Now a binocular (2p,2p) structure has taken the lead. (d) $\epsilon = 0.15$: A (3d,3d) structure now dominates. (e) $\epsilon = 0.20$: (1s,-1s) dominates once more, but now it is no longer monocular since it has deformed significantly towards becoming a (2s,-2s) structure. (f) $\epsilon = 0.25$: (1s,-1s) again in the lead, but its deformation is yet more severe.

favour binocularity.

4.5.4 Perturbations in growth rates

We calculated above the change in the eigenvalues caused by perturbing correlations. However, perturbations can also affect the growth rates of the different eigenvectors in the context of the evolution of a general weight vector. We briefly demonstrate that, for a general weight vector satisfying Miller's initial condition that the difference between w_L and w_R is small, the effect of the perturbation is to bias the competition towards the zero DC (binocular) solutions.

Consider the expansion of (w_L, w_R) in terms of the eigenvectors of Q^{ϵ}_{*} after time t:

$$\begin{aligned}
\begin{pmatrix} w_L \\ w_R \end{pmatrix} &= \sum_{x_i} (\lambda_i + \epsilon\xi_i)^t \begin{pmatrix} \alpha_i x_i + \epsilon a_i \\ \beta_i x_i + \epsilon b_i \end{pmatrix} [(\alpha_i x_i + \epsilon a_i).w_L + (\beta_i x_i + \epsilon b_i).w_R] \\
&+ \sum_{x_i} (\lambda_i + \epsilon\zeta_i)^t \begin{pmatrix} -\beta_i x_i + \epsilon c_i \\ \alpha_i x_i + \epsilon d_i \end{pmatrix} [(-\beta_i x_i + \epsilon c_i).w_L + (\alpha_i x_i + \epsilon d_i).w_R] \\
&+ \sum_{y_i} (\mu_i + \epsilon\gamma_i)^t \begin{pmatrix} y_i + \epsilon e_i \\ -y_i + \epsilon f_i \end{pmatrix} [(y_i + \epsilon e_i).w_L + (-y_i + \epsilon f_i).w_R] \\
&+ \sum_{z_i} (\nu_i + \epsilon\delta_i)^t \begin{pmatrix} z_i + \epsilon g_i \\ z_i + \epsilon h_i \end{pmatrix} [(z_i + \epsilon g_i).w_L + (z_i + \epsilon h_i).w_R]
\end{aligned}$$

If we assume that the difference between w_L and w_R is of $\mathcal{O}(\epsilon)$, we can calculate terms of $\mathcal{O}(\epsilon)$ in the expansion, using $(\lambda_i + \epsilon\xi_i)^t \approx \lambda_i(1 + \epsilon t\xi_i/\lambda_i)$, *etc.* We find that $w_L - w_R$ grows as $\lambda_i^t(1 + \xi_i t)$ in the direction of x_i and μ_i^t in the direction of y_i. This extra t term could bias the competition towards the zero DC solutions, depending on the sign of ξ_i.

4.6 Summary and significance of the results

Linsker's correlational learning rule can be generalised to the binocular case. A particular version of this was studied independently by Miller: he looked in detail only at the case of identical correlations in the two eyes and zero correlations between the two eyes. We have presented a more general analysis, and shown what happens to the eigenvalues and eigenvectors when small perturbations to Miller's correlations are considered. The main results are as follows:

- When positive correlations are added between the eyes while correlations within each eye remain identical, the eigenvalues of the monocular eigenvectors decrease. The other non-zero DC eigenvectors have increased or decreased eigenvalues depending on whether they are of the form $(+,+)$ or $(+,-)$ respectively, but to a lesser extent. The previously degenerate zero DC eigenvectors split, one eigenvalue being slightly greater than before and the other being slightly less.

- If there are identical negative correlations within each eye (a case investigated empirically by Miller), all eigenvalues are reduced, but the monocular eigenvectors are reduced the most. In addition, eigenvectors begin to change their form.

- When the correlations within the two eyes are made slightly different, the zero DC eigenvectors change their character: they are not stable to the perturbation. This also breaks the degeneracy. If the perturbing correlations are negative in each eye, then the same conclusions hold as for the case of identical negative correlations as above.

- General perturbations can lead to a slightly increased *growth rate* for components of the evolving weight vector in the direction of binocular solutions as compared to an unchanged growth rate for monocular solutions.

Thus binocular solutions can be favoured over monocular solutions. However, the following two points need to be taken into consideration before these results can be used to make definite predictions about the weight structure that will finally dominate:

- The absolute size of the eigenvalues needs to be known. Even though monocular solutions are relatively punished, it might be that this effect is not enough for any particular perturbation to make a qualitative difference to the final outcome.
- The form of the perturbing correlation matrices needs to be known. Although our analysis is general, the concrete cases we concentrated on assumed that all components of each perturbing matrix have the same sign. More complicated perturbation matrices will lead to results that are impossible to predict without performing the precise calculations.

These results are significant because perturbations are inevitable in the natural system. It is implausible to think that the correlation functions are identical in the two eyes, or that correlations between the eyes are exactly zero. In the former case there is a particular form of instability in Miller's model, and in the latter case monocularity becomes less likely as cross-eye correlations increase.

It is important to make a distinction between different types of result. Firstly there is the question, are the eigenvectors stable to perturbations? We have shown that for some perturbations they are not. The other type of result is using perturbation theory to take a small (size ϵ) step in analysing the effect of more realistic correlations on model results. One step of this type is the analysis of cross-eye correlations. For animals such as cats and monkeys it seems highly likely (although not yet shown experimentally) that there should be substantial correlations between activities in the two eyes for the case of evoked activity. Perturbation theory gives a hint as to what would happen in Miller's model in this case. Another result of this type is the analysis of adding negative correlations within an eye: again, perturbation theory goes some way to analysing this case.

4.7 Other aspects of Miller's model

We now briefly discuss two other aspects of Miller's model that have relevance later in the thesis.

4.7.1 Constraints

An important point about Miller's system is the type of constraints used to limit synaptic strengths. Besides saturation limits he also used a type of normalization which we refer to as *subtractive*. This is to contrast it with the more common *divisive* type of normalization (as used for instance in the Neural Activity model [Willshaw & von der Malsburg 1976]). These are two different ways of keeping the sum of components or sum of squares of components of the weight vector constant for a particular unit, depending on the way these constraints are enforced. In divisive normalization, after each weight update, every component is divided by an appropriate number to enforce the constraint. In subtractive normalization, a constant amount is subtracted from each component, the constant vector being chosen so as to enforce the constraint. Usually we consider the case where all components of the constraint vector are the same. A detailed analysis of the difference between these cases in this context is presented in [Miller & MacKay unpublished]. For our purposes, there are two points of significance:

- Normalizing in either case by the sum of components is equivalent to normalizing according to the (z_{1S}, z_{1S}) eigenvector. Thus the amount of (z_{1S}, z_{1S}) in the evolving weight structure is kept constant, and perturbations to the eigenvalue of this eigenvector will not affect the nature of the competition.
- If the constraint is enforced divisively, the fastest-growing $(y_i, -y_i)$ solutions have to grow faster than (z_i, z_i) in order for monocularity to develop. However, if the constraint is enforced subtractively, the rate of growth of $(y_i, -y_i)$ only has to be positive to have a chance of winning the race to saturation.

For Miller's model, this means that monocularity can develop under *divisive* enforcement of constraints only if the two eyes are *anticorrelated*, whereas under *subtractive* enforcement of constraints there is the potential for monocularity to develop even when the two eyes are *positively correlated* (although we have shown above that this cannot be guaranteed). We discuss the difference between these two types of enforcement of the constraints in the context of a simple competitive model in chapter 6.

4.7.2 Stripe width in Miller's model

An important aspect of any model of ocular dominance stripe formation is the dependence of stripe width on the parameters of the model. It is argued in [Miller 1990(a)] that, if we now consider a complete layer of output

units in the model, there are several different cases with respect to stripe width. Firstly, stripes only form if the fastest-growing eigenvector of S^D is monocular, and this depends on the parameters. If we assume this is true, then the width depends on the form of the cortical interaction function I, which we have not so far considered. If I is of Mexican hat form (Miller usually used a difference of gaussians), then the periodicity of the resulting ocular dominance stripes is related to the position of the peak of the fourier transform of the cortical interaction function. However, if I is all-excitatory, then Miller argues that the addition of constraints on input units is required to force segregation:

> In the case of a purely excitatory cortical interaction function, the peak of the fourier transform of I corresponds to an infinite wavelength, that is dominance by a single eye everywhere. Addition of constraints on the total synaptic strength of input cells modifies the equation, forcing the left- and right-eye inputs from each retinotopic location to divide between them the cortical surface over which they initially arborize. The result is to suppress the growth of monocular patterns with wavelength longer than an arbor diameter, so that the fastest-growing monocular pattern has a wavelength of about an arbor diameter [Miller 1990(a)].

We will discuss further the issue of including normalization constraints for presynaptic as well as postsynaptic units in chapter 6. In the meantime, we make the following observation with regard to the implications of this for Miller's monocular deprivation results. Miller presents simulations (e.g. [Miller et al 1989]) where monocular deprivation is modelled by reducing the strength of the correlations in one eye. Stripes from the deprived eye are narrower, as in the natural system. However, this only occurs because the normalization constraints on presynaptic units are now relaxed: in particular Miller allowed the sum of synaptic strengths for each presynaptic unit to vary between 0.5 and 1.5 times its initial value. Without this, presynaptic normalization compels each presynaptic unit to take over the same amount of cortex regardless of the strength of the correlations.

4.8 Conclusions

Miller's model provides an explanation for the formation of ocular dominance stripes that has the virtues of both rich possibilities for mathematical analysis, and a clear biological interpretation for some of its parameters. However the model has limitations. Firstly, it is unstable to various types of perturbative correlations, and monocularity breaks down in the presence of certain perturbative correlations either within or between eyes. Secondly, it has not yet been shown that the same mechanism can satisfactorily account for the development of topography.

Chapter 5

Elastic Models

5.1 Introduction

The elastic net is a developmental algorithm originally presented in [Durbin & Willshaw 1987]. It was first applied to the topography and ocular dominance problem in [Goodhill 1988, Goodhill & Willshaw 1990], where it was shown to be capable of producing both topography and ocular dominance from the same mechanisms. An important feature of the algorithm is that there exists an "energy" function which the algorithm minimizes. The elastic net can also therefore be regarded as an optimization algorithm: indeed, it was first presented in [Durbin & Willshaw 1987] as a method for solving combinatorial optimization problems that have a geometric interpretation.

In this chapter we consider the properties of the elastic net as an algorithm for finding a mapping between a set of presynaptic units and a set of postsynaptic units. Presynaptic units are represented as laying at fixed positions in an abstract space, and postsynaptic units as an elastic sheet of points moving in this abstract space. Gradually the elastic sheet deforms to "visit" each presynaptic unit, defining a map. Distances between presynaptic units in the space represent the degree to which they are "similar". There is a tension in the elastic sheet that attempts to ensure that the sheet deforms as little as possible in visiting all the presynaptic units: the aim is to map neighbouring points in the elastic sheet to presynaptic units that are similar (i.e. are nearby in the abstract space).

In this chapter we extend earlier work on this algorithm [Goodhill 1988, Goodhill & Willshaw 1990] in the following ways:

- Here we concentrate on the two-dimensional case, previous work being mostly restricted to the one-dimensional case.
- We present a more detailed discussion of the assumptions and biological meaning of the representation and the algorithms used.
- We compare the performance of the elastic net to a new algorithm related to Kohonen's self-organizing feature map algorithm.
- We show that the elastic net and the Kohonen-type algorithm can account for the experimental results of monocular deprivation.
- We use the elastic net to test the hypothesis of [LeVay et al 1985], that a global orientation of the stripes can be produced purely by the boundary conditions, without invoking any other mechanisms to produce anisotropy.

Firstly we discuss possible motivations for formulating the problem of the development of brain maps in terms of optimization.

5.2 Optimization principles for map formation

5.2.1 The minimal wiring argument

[Cowey 1979] put forward the following argument for why ordered maps should exist in the visual system:

> [Why is there a map?]....the answer rests on an assumption, which is that interactions between cortical neurones are much more important for cells representing points close together in visual space than for widely separated points. We are all familiar with the area of lateral inhibition in the retina whereby the simultaneous luminous contrast between adjacent portions of a scene is highlighted. If that process of selection and relative exaggeration of some feature of the visual image is continued at the cortex we require cortical interneurones....By having a retinotopic map in striate cortex the length of their axons can be minimized.

These arguments can be used generally for many types of mapping in the brain: points in the target sheet that should be connected are those representing points that are close together in the input (parameter) space. Two biological motivations for the minimization of wire length have been outlined in [Nelson & Bower 1990]: (a) wire takes up volume, and there is only a very limited space available in the cranium, and (b) short interconnection lengths minimize the delays in propagating information from one processor to another.

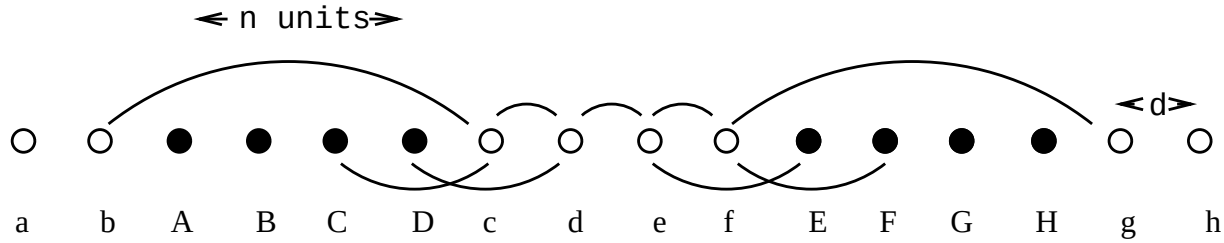

Figure 5.1: Application of the minimal wiring argument to the one dimensional topography and ocular dominance problem. White circles represent cortical units connected to the left eye at topographic position indexed by lower case letters. Black circles represent cortical units connected to the right eye at topographic position indexed by upper case letters. Lateral connections (shown only for one stripe) are assumed to exist between units representing neighbouring points in the same eye, and between units representing corresponding points in the two eyes. Periodic boundary conditions are assumed.

This mapping problem is straightforward in the case where the input and target spaces have the same dimensionality: however, if they have different dimensionalities matters are more complicated. Refer to the input space as I with dimension D_I, and the target space as T with dimension D_T. Minimizing the length of wire as described above by Cowey says that we should try to map neighbouring points in the input space close together in the target space. For $D_I = D_T$ it is straightforward to preserve all neighbourhoods of I in T. However, if $D_I > D_T$ this is clearly not possible, and a compromise solution is needed. For instance, in the discrete case with $D_I = 2$, $D_T = 1$ and a square topology in D_I, each point in I has 4 (orthogonal) neighbours whereas points in T have only two neighbours. The consequences of the minimal wiring principle for this case have been investigated in detail by [Mitchison & Durbin 1986]: they studied the mapping from a two dimensional parameter space consisting of a square array of points of size $N \times N$ to a one dimensional rope of N^2 points. They assumed the cost of a length of wire l required to connect points in the target space representing neighbouring parameter values to be l^p, and showed that the optimal map has a qualitatively different form depending on whether p is greater than or less than unity.

5.2.2 Application to ocular dominance stripes

We now consider the predictions made by this "minimal wiring principle" for a simple case of the topography and ocular dominance problem. The input space is the two retinae, and the target space is the cortex.

Consider two one-dimensional "retinae" each of N units mapping in a 1-1 fashion to a one-dimensional "cortex" of $2N$ units, each pair of cortical units distance d apart. Assume now that the cortex desires to process a unit in one retina together with its neighbours in that retina, and the corresponding unit in the other retina (see figure 5.1). Following [Mitchison & Durbin 1986, Durbin & Mitchison 1990] we investigate the case where the cost of a length of wire l joining two units is l^p for some norm parameter p. Then the total cost $L_p(n)$ of connecting appropriate units in the cortex for a particular stripe width n is

$$L_p(n) = Nd^p \left[1 - \frac{1}{n} + \frac{(n+1)^p}{n} + \frac{1}{2}\left(\frac{n}{2}\right)^p\right].$$

First consider the case $p = 1$. We then have

$$L_1(n) = Nd\left[2 + \frac{n}{4}\right].$$

Note firstly that the cost of connecting neighbourhoods does not depend on n: the wider jumps required for large n are exactly compensated by the fact there are fewer of them. In order to minimise $L(n)$ it is therefore only required to minimise the correspondence term. This specifies stripes as narrow as possible i.e. $n = 1$.

For general p the minimum cannot be found analytically, so instead we resort to graphical means. The neighbourhood component of the expression for $L_p(n)$ (i.e. the contribution made by connecting neighbouring units for same-eye stripes) is plotted as a function of n and p for $0 < p < 1$ in figure 5.2. We see that the neighbourhood cost decreases gradually as n increases, as expected. However, figure 5.3 shows that when the correspondence term (i.e. the contribution made by connecting corresponding units in opposite-eye stripes) is added, $L_p(n)$ increases with n for all p in this range. Thus again the lowest cost is achieved for $n = 1$.

One other case investigated was that of including a greater range of connectivity than simply nearest-neighbour. We calculated the cost ($p = 1$) of the neighbourhood term for the case where each unit makes connections with

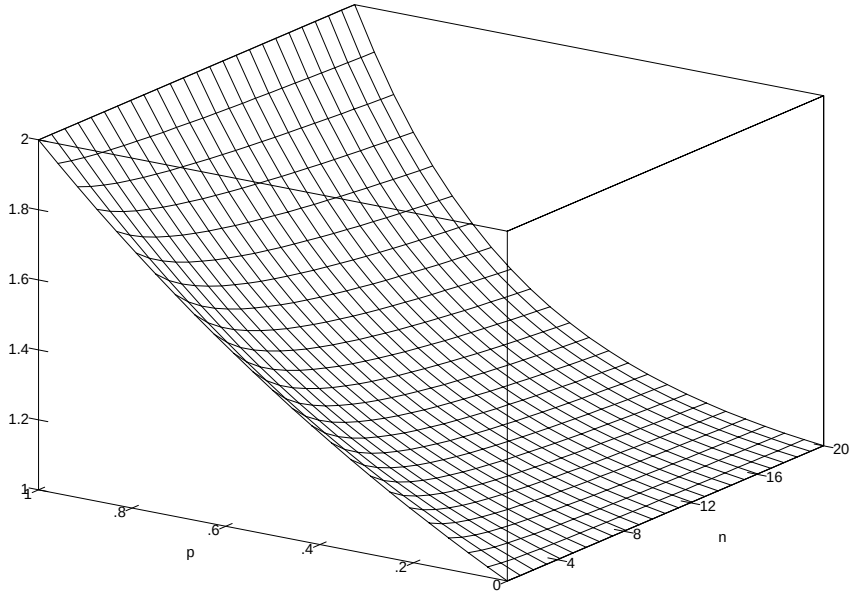

Figure 5.2: Neighbourhood term in minimal wiring argument plotted as a function of the norm parameter p and the stripe width n.

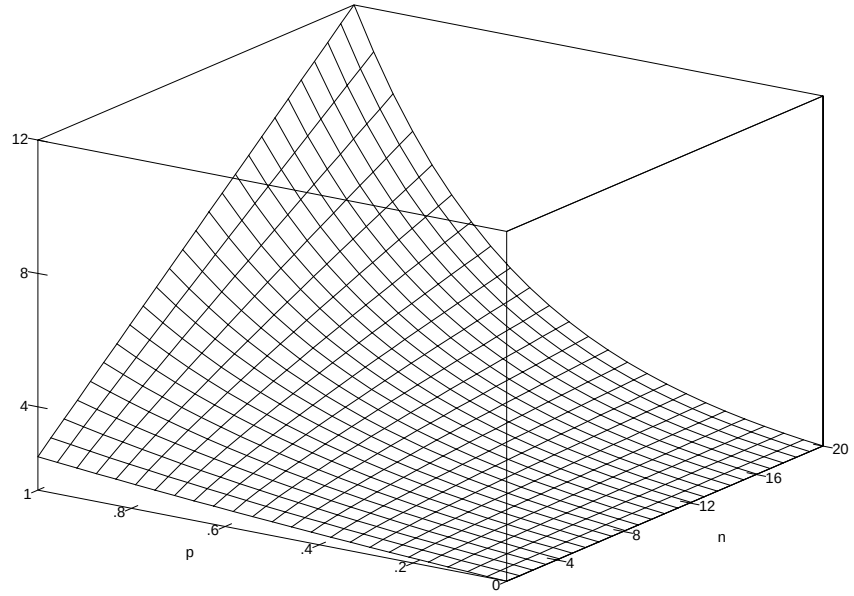

Figure 5.3: Neighbourhood and correspondence terms in minimal wiring argument plotted as a function of the norm parameter p and the stripe width n.

its m neighbours on each side. This is $Nm(m+1)$, which is independent of stripe width. Thus $n = 1$ is still the optimum width, due to the correspondence term.

Thus we conclude that a simple case of the minimal wiring principle in one dimension does not lead to wide stripes, as are observed in the natural system. One possibility to try to repair this situation is to specify that the cost of connecting corresponding units in the two eyes is different from the cost of connecting neighbouring units within an eye. The simplest possibilities are making the cost for corresponding units differ from the cost for units within an eye by a fixed amount or by a scaling factor. The first case does not change the position of the minimum of $L_p(n)$. In the second case, it would be necessary for the scaling factor to be such that the cost of joining corresponding units is substantially *less* than the cost of joining neighbouring units, for stripes wider than $n = 1$ to form.

It is difficult to draw firm conclusions from these results, since it would certainly be possible to construct cost functions which do have a minimum for $n > 1$. A crucial point is the weighting given to the cost of neighbourhoods along different dimensions (in our case, space and ocularity): there is currently little biological data available to guide the choice of these weights.

A substantially more detailed application of minimal wiring arguments to the formation of stripes, and also cortical areas, has recently been undertaken in [Mitchison 1991, Mitchison 1992]. Here, rules for arborization of axons are considered, along with factors such as the decrease in axonal width at branching. It is shown that, given certain assumptions, such considerations can predict division of the cortex into areas, and also areas into stripes.

Up to now, we have just considered the formation of maps that keep neighbouring points in the input space I close together in the target space T. However, there is an alternative "direction" for optimizing the map: we could instead specify that neighbouring points in T represent points that are close together in I. If $D_I = D_T$, this specifies the same map as the minimal wiring principle. However, when $D_I > D_T$ a mapping that preserves neighbourhoods in this new sense is still possible. [Durbin & Mitchison 1990] investigated the maps formed by the elastic net algorithm (which optimizes in this direction) for the $D_I = 2$, $D_T = 1$ case described above. They found that these had a somewhat similar form to the minimal wiring maps found with $p < 1$, but produced using much less computational effort than performing a direct minimization in the $I \rightarrow T$ direction. It was argued that the brain uses algorithms similar to the elastic net to optimise in the $T \rightarrow I$ direction as a computationally efficient approximation to the desired $I \rightarrow T$ optimization.

This lengthy preamble has served to establish some motivation for applying the type of optimization implemented by the elastic net to the problem of brain maps. The rest of this chapter concerns the application of the elastic net to the topography and ocular dominance problem.

5.3 The elastic net algorithm

5.3.1 Application to the Travelling Salesman Problem

The elastic net algorithm [Durbin & Willshaw 1987] was originally developed as an approximate method for the Travelling Salesperson Problem (TSP), a well-known NP-complete combinatorial optimization problem. The objective is to find the shortest distance a salesperson can travel to visit a set of N cities in a plane and return to where they started. In general the matrix of city distances can be arbitrary, but here we consider the simplest version (the "geometric" case), where the distances between cities are just the Euclidean distances in the plane (for a recent application of retinotopic mapping algorithms to the non-geometric TSP, see [Frean 1990]). This problem has been extensively discussed in the combinatorial optimization literature, being both easy to state and hard to solve. For N cities there are $(N-1)!/2$ possible routes: for large N it is impossible to search them all to find the optimal tour. Therefore many heuristic algorithms have been investigated, which aim to provide good solutions in reasonable time (for review see [Lawler et al 1986]).

The set of valid tours for a TSP of size N can be represented as the vertices of an N-dimensional hypercube. Most techniques aim to provide good ways of stepping from one vertex to another to gradually improve the quality of the solution [Lawler et al 1986]. However, an alternative method is for the search to proceed through the continuous space inside the body of the hypercube, only converging to a valid solution in the final state. The first neurally-inspired algorithm employing this strategy was introduced in [Hopfield & Tank 1985], and shown to be capable of producing reasonable solutions for small scale ($N \leq 30$) problems.

The elastic net also operates in a continuous space. The inspiration for the development of the algorithm was that the TSP is analogous to the development of topographic mappings in the nervous system: the problem is to find a mapping between an array of presynaptic units (cities) and an array of postsynaptic units (tour points)

such that neighbouring postsynaptic units are mapped to by presynaptic units that are close to one another. Durbin and Willshaw simplified the equations of the Tea Trade model [Willshaw & von der Malsburg 1979] to a form suitable for the TSP case. They considered a string of "beads" j with positions y_j moving on an elastic rope, each bead being pulled towards the cities i with positions x_i. The change in the position Δy_j of each bead at each time step is given by

$$\Delta y_j = \alpha \sum_i w_{ij}(x_i - y_j) + \beta k(y_{j+1} - 2y_j + y_{j+1}) \tag{5.1}$$

The first term is a matching term that represents the "pull" of cities for beads, which is traded off with ratio $\alpha/\beta k$ against a regularization term representing a "tension" in the rope, i.e. a tendency for neighbouring beads to remain close together. The "weights" w_{ij} are defined as follows:

$$w_{ij} = \frac{\Phi(|x_i - y_j|, k)}{\sum_p \Phi(|x_i - y_p|, k)} \tag{5.2}$$

where

$$\Phi(|x_i - y_j|, k) = \exp(\frac{-|x_i - y_j|^2}{2k^2}) \tag{5.3}$$

Over the course of a simulation, the scale parameter k is gradually reduced, so that the matching term dominates the regularization term. It was shown by Yuille that this equation can be integrated to produce an "energy" function E [Durbin & Willshaw 1987]:

$$E = -\alpha k \sum_i \log \sum_j \Phi(|x_i - y_j|, k) + \frac{\beta}{2} \sum_j |y_{j+1} - y_j|^2 \tag{5.4}$$

This is such that

$$\Delta y_j = -k \frac{\partial E}{\partial y_j}$$

Thus the length of the path in the L_2 norm and the distance of beads to cities is being simultaneously minimised. For k small enough the rope passes arbitrarily close to the cities for appropriate α and β. The reduction of k over the course of a simulation has a similar effect to that of simulated annealing [Kirkpatrick et al 1983], although the way it operates is somewhat different. Since E is a function of k, reducing k changes the energy landscape. It has been shown [Durbin et al 1989] that for k large enough, E is a bowl with a unique minimum, but as k is reduced a critical value is reached at which the energy bifurcates. The algorithm picks up a particular minimum and tracks it as the energy surface bifurcates further: this has been referred to as "deterministic annealing" [Yuille 1990]. Application of this algorithm to a standard 30 city problem reproduced the best known solution: application to larger problems gave tours within reasonable time with lengths only a few percent greater than those produced by simulated annealing [Durbin & Willshaw 1987].

5.3.2 Relation to Hopfield and Tank's algorithm

Recently [Simic 1990, Simic 1991] and [Yuille 1990] have demonstrated a mathematical relationship between the elastic net and the algorithm of [Hopfield & Tank 1985]. They showed that both can be derived from a general objective function for the TSP under a statistical mechanical interpretation. The two algorithms represent different ways in which the "hard" constraints can be enforced[1]: weakly in the Hopfield and Tank algorithm (i.e. bias terms are added to the energy function) or strongly in the elastic net (i.e. constraints explicitly imposed in calculating the "partition function"). In deriving the Hopfield and Tank algorithm, an approximation is used that keeps only nearest-neighbour interactions between tour points. [Yuille 1990] argues that this, along with the addition of extra terms in the energy function, increases the number of local minima of the energy function and this helps to account for the poor performance of the Hopfield and Tank algorithm compared to the elastic net [Wilson & Pawley 1988, Peterson 1990]. [Simic 1990] goes further into the statistical mechanics analogy, and suggests that the elastic net approach amounts to evaluating statistical averages of the system, such as mean path, at given levels of noise, and then tracking the averages as the noise decreases to zero (a continuation method). He argues that this is equivalent to minimising the thermodynamic free energy F of the system, and this is useful since F "takes account of" the noise (controlled by the temperature) in the system. F is much smoother than the original cost function for non-zero noise, but approaches it as the noise goes to zero.

[1] A hard constraint is one that *must* be satisfied: for instance for the TSP that each city must be visited by a path point. That the tour length be short is a soft constraint.

5.3.3 A statistical interpretation

A statistical interpretation for the elastic net algorithm has recently been found that sheds further light on what is being optimized. If it is imagined that the feature points (cities) represent the set of data points it is desired to model with a set of gaussian functions with variance k^2 centred on the beads, the matching term in the elastic net equations move the beads to the positions where they are most likely to have generated that data at each k [Durbin et al 1989]. This suggests an analogy with the Expectation and Maximisation (EM) algorithm [Dempster et al 1977], and inspires a faster way of optimizing E than steepest descent. This was used in [Durbin & Mitchison 1990], and involves explicitly calculating the optimum positions for the beads at each k in terms of the city forces, and then iterating to relax the tension (a form of Gauss-Seidel relaxation). The amount of computation for each k is slightly greater than steepest descent, but k can now be reduced much faster and thus the computationally intensive step of calculating the weights is performed less frequently.

5.3.4 A "biological" interpretation

We follow the discussion of [Durbin & Mitchison 1990]. Equation 5.1 can be interpreted as saying that each postsynaptic unit has a gaussian receptive field at position y_j in the presynaptic sheet. The amount postsynaptic unit j responds to input i at position x_i is given by w_{ij}. The normalization of w_{ij} by the response of all other postsynaptic units to input i implements a form of soft competition between postsynaptic units. Although all postsynaptic units are adapted towards input i, those that respond most strongly are adapted the most. The first term can therefore be seen as Hebbian [Durbin & Mitchison 1990]. The second term says that postsynaptic units are also adapted towards inputs that their neighbours respond to.

5.4 Kohonen-type algorithms

Shortly after [Durbin & Willshaw 1987], two similar algorithms for the TSP appeared [Angeniol et al 1988, Fort 1988], based on Kohonen's self-organizing feature map algorithm [Kohonen 1982(a)]. These work by finding the nearest bead to each city at each iteration, and then moving that bead and its neighbours on the ring closer to that city. The system is "annealed" over the course of a simulation by gradually reducing the extent of neighbourhood updating around the ring. These algorithms can be compared with the elastic net in statistical terms. The elastic net can be seen as an algorithmic version of the Tea Trade model, and implements a gaussian mixture model plus regularization (a form of soft competitive learning). Kohonen's algorithm can be seen as an algorithmic version of the Neural Activity model, and implements the K-means algorithm [MacQueen 1967] plus regularization (a form of hard competitive learning). The regularizer in both cases attempts to keep tour lengths short. [Nowlan 1990] has shown that the K-means algorithm implemented by Kohonen is an approximation to the statistically "correct" way of placing a set of gaussian probability generators implemented by the elastic net. The elastic net is thus more sophisticated than the Kohonen-type algorithms, since it both takes into account the effect of all cities on each bead, and allows "slippery" matching (that is, particular beads take longer to become committed to particular cities). Unsurprisingly, the elastic net performs better than these algorithms on average if one run at a time is compared [Angeniol et al 1988]. However, an original advantage of the Kohonen-type algorithms was that they were computationally faster. Also they are non-deterministic: the order of consideration of the cities is randomly chosen at each time step. It was argued in [Angeniol et al 1988] that several Kohonen-type simulations can be run in the time taken for one (deterministic) run of the elastic net, and although on the average these tours will be longer, the *best* tour is likely to be shorter than the elastic net tour. However, the efficiency of the elastic net can be improved by techniques such as Gauss-Seidel relaxation, and this begins to make the computation times comparable with that of Kohonen-type algorithms (R. Durbin, personal communication).

An algorithm proposed in [Burr 1988] can be viewed as a combination of the elastic net and Kohonen-type approaches. The equations are similar to the elastic net equations, with the addition of an extra city-force term expressing nearest-neighbour attractions, as described in the algorithms above. In addition, Burr proposed a more sophisticated annealing schedule than that used in [Durbin & Willshaw 1987]. At each step, k is reduced to about the same size as the root-mean-square city to nearest-bead distance. This algorithm can produce good tours much faster than the elastic net on a small set of problems [Burr 1988], but so far a more complete comparison has not been presented.

A comparison of several TSP algorithms was recently undertaken in [Peterson 1990]. The performance of the elastic net, an algorithm similar to Hopfield and Tank, a genetic algorithm, simulated annealing with a "2-opt" strategy [Lawler et al 1986], and a hybrid of several approaches, was compared on a standard set of 50, 100 and 200 city problems. The best algorithm was the genetic algorithm, followed by the hybrid approach and then

the elastic net. However, these results are difficult to assess since no consideration was given to the extremely important practical issue of the time taken by each algorithm. A more detailed comparison can be found in [Simmen 1992].

5.4.1 A new Kohonen-type algorithm

There has been much interest in how the performance of the elastic net compares with that of Kohonen-type algorithms for the TSP. One of the aims of this chapter is to compare these two approaches for the topography and ocular dominance problem. This helps to shed light on the relation between the quality of the solution to the optimization problem and the morphology of the stripes thus generated.

We now describe a slightly modified form of the Kohonen algorithm, that was first introduced in [Goodhill 1990].[2] The equations are as follows (notation as for the elastic net case):

$$\Delta y_j = \alpha \sum_i w_{ij}(x_i - y_j) \tag{5.5}$$

where

$$w_{ij} = \frac{\Phi(|j - j_i^*|, k)}{\sum_p \Phi(|j - j_i^*|, k)} \tag{5.6}$$

where j_i^* is the unit in the cortical sheet that is closest to x_i, and $\Phi(|j - j_i^*|, k)$ is given by

$$\Phi(|j - j_i^*|, k) = \exp(\frac{-|j - j_i^*|^2}{2k^2}) \tag{5.7}$$

The most important difference between this algorithm and those of [Angeniol et al 1988] and [Fort 1988] is that here updating occurs only after all feature points have been presented and their effects on each point normalized, rather than inputs being presented one at a time. The reason for this difference is that it was found that the highly regular and symmetric distribution of cities/presynaptic units (see below) caused unwanted effects when the cities were presented one at a time. Again k is started large and reduced over time. A form of hard competition is implemented: rather than each cortical unit responding to each input according to its distance from that input in feature space as in the elastic net, only the closest unit responds. The updating of the neighbours of the closest unit that also occurs takes the place of an explicit regularization term, and serves to keep cortical wiring short. Some trial simulations showed that the above algorithm could produce good solutions to the TSP, for instance reproducing the shortest known tour for the standard 30-city problem mentioned above.

5.4.2 A "biological" interpretation

This is similar to that of the elastic net as discussed above, except that now only the postsynaptic unit with the biggest response to each input is considered. Only the winning unit and its topological neighbours are adapted towards each input.

5.5 Application to the topography and ocular dominance problem

For the topography and ocular dominance problem, the input space (which we will also refer to as the feature space) to be represented consists of all positions in both eyes. We refer to feature points as being retinal units. Distances between retinal units in the space encode the "similarity" of units to each other. In the original Tea Trade model, the similarity between two units is the degree to which their chemical marker vectors are similar, and this is a possible biological interpretation of the distances in the feature space under consideration here (later we will identify a problem with this interpretation). An alternative implementation of similarity is in terms of the activity of retinal units: distances in the feature space represent the degree to which two units are *correlated* (a mathematical justification for this interpretation of distances comes from [Yuille et al 1991(a)], which is discussed in chapter 7). Retinal units that have highly correlated activities lie close together in the feature space.

We represent each eye as being a two-dimensional sheet of points, making the following assumptions (to which we return later):

[2] This algorithm was developed in conjunction with Harry Barrow.

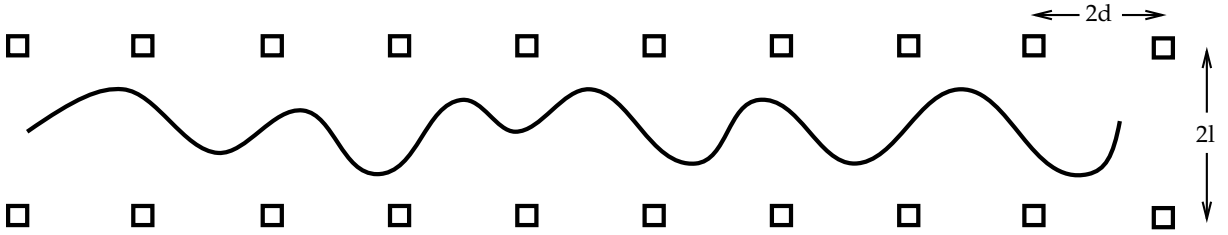

Figure 5.4: The representation of the topography and ocular dominance problem used in the one-dimensional case. The retinal units (squares) are arranged in two parallel rows. The rope, representing the mapping to the cortex, moves in between. The two dimensional case is analogous.

- There is a direct relationship between the correlation of retinal units and their relative physical positions in the eye, and units are ordered in feature space in the same way as in the retina.
- The correlation between two units is a function only of their separation in feature space, and not their absolute positions.

The third dimension of the space is taken to represent ocularity, but *distances in the ocularity dimension represent correlations according to the same metric as in the two "spatial" dimensions*. Thus our representation of the problem is that of two square arrays of units representing each eye laying on top of one another in horizontal planes, separated by a constant vertical distance (illustrated in the one-dimensional case in figure 5.4). To reiterate, the distance between any two retinal units, either within or between retinae, represents the similarity of the two units, or as we have interpreted it above, the correlation between those two units. Considering now the mapping to the cortex, we envisage this as an elastic sheet of points moving in this feature space: we refer to these points as cortical units.

This abstract representation affords a powerful tool for investigating the way in which correlations between retinal units determine the properties of the mapping. There is in effect just one dimensionless parameter to manipulate: the ratio of the separation of corresponding units between retinae to the separation of neighbouring units within a retina. We will refer to these two separations as $2l$ and $2d$ respectively. The ratio l/d determines the way correlations vary across the two eyes. If it is less than one, it says that a unit in eye 1 is more correlated with its corresponding unit in eye 2 than with its neighbouring units in eye 1. In the case we more usually consider below, the ratio is greater than one, saying that a unit is most correlated with its same-eye neighbours, but less correlated with same-eye units more than a short distance away than with its corresponding unit in the other eye. A problem with interpreting similarities as chemical should now be apparent. Under this interpretation, the ocularity dimension of the space represents a chemical difference between the two eyes. Such a global difference is hard to justify in biological terms (as discussed in chapters 2 and 3).

5.6 Results

5.6.1 Development

We now review the course of development over an elastic net simulation (see also [Goodhill 1988, Goodhill & Willshaw 1990]). The main stages are shown for a representative simulation in figure 5.5. For any arbitrary initial configuration of the cortical sheet at large enough initial k, the sheet flattens, becomes untangled, and moves to a position equidistant between the two retinae. Points are quite closely spaced in a square lattice, so that the total area of the sheet is somewhat less than the area of the retinae. A polarity can be introduced in the form of a bias in the initial random distribution of points. The state so far reached represents an ordered but only crudely refined map, where each cortical unit is binocular. As k is reduced further, gradually the sheet expands until its edges reach the edges of the retinae. The sheet remains flat. Points are equally spaced, up to a particular value of k at which point they form clusters (for discussion see [Goodhill 1988, Goodhill & Willshaw 1990]). At a certain critical value of k which will be referred to as k_s, the vertical symmetry is lost and some points move quickly upwards towards one retina while other points move quickly downwards to the other retina. Once reaching the vicinity of the retinal units, points move slowly for very small values of k to become roughly coincident with retinal units. The resulting pattern is one of stripes: the sheet alternates periodically between the two retinae, representing alternation of units connected only to one eye with units connected only to the other. This mapping is "topographic": apart from inevitable discontinuities at stripe boundaries, neighbouring retinal units map to neighbouring points in the cortex. To test the possibility that

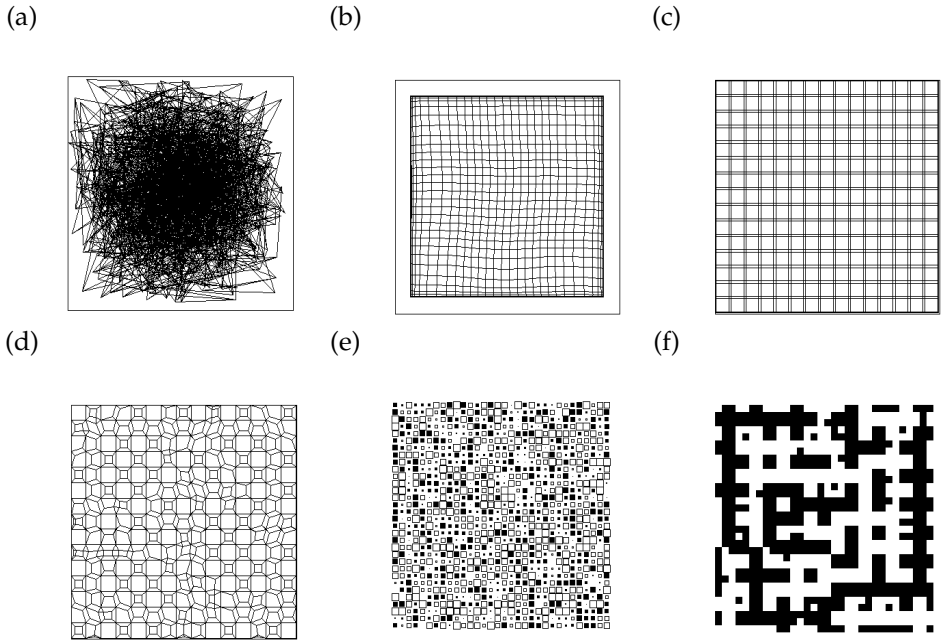

Figure 5.5: The time course of a typical elastic net simulation. There are two types of picture: (a-d) show the positions of the points in the elastic sheet as viewed from above, while (e-f) show the ocularity of each cortical unit. (a) Initially the points in the cortical sheet have large random scatter in spatial position. (b) The sheet soon develops a topographic order and lies flat, equidistant from the two retinal sheets. (c) As k is reduced further neighbouring groups of cortical points may start to cluster. (d) The final position of the sheet. (e) Initially the sheet has random ocularity component. (f) At a certain critical value of k stripes form.

the striping behaviour observed was an artifact of the effects of the rope having ends, some one-dimensional experiments were also performed using periodic boundary conditions in [Goodhill 1988]. The two rows of units were wrapped around into a cylinder, and the rope (with ends joined) constrained to move on the surface of the cylinder. No qualitative difference in development or final patterns could be detected. The topography pictures in figure 5.5 show the horizontal positions of the cortical units: the sheet is flat except in the initial state and when stripes have formed. This last picture is slightly hard to interpret as many of the lines joining neighbouring cortical units are coincident. However, it can be seen that the basic grid structure of the sheet remains, indicating overall topography.

Dependence of k_s on parameters

In [Goodhill 1988, Goodhill & Willshaw 1990] some approximations were used in order to estimate k_s, the value of k at which stripes form. The simplest approximation assumes that a group of beads lie together, equidistant between a pair of corresponding units in the two retinae, and k is sufficiently small that these are the only two units that affect the beads. An equation is derived specifying when a small vertical perturbation in the position of one bead will be magnified rather than reduced. Other approximations are considered, each taking into account the effect of a bigger neighbourhood of retinal units. All the approximations match fairly closely the values of k_s observed empirically for the one-dimensional case [Goodhill & Willshaw 1990], and all are almost entirely dependent only on the parameter l. This agrees with intuitive notion that stripe formation is a "vertical" phenomenon and should therefore be controlled by the parameter measuring vertical scale.

This completes our review of [Goodhill 1988, Goodhill & Willshaw 1990], and we now consider detailed results in the two-dimensional case for the elastic net and Kohonen-type algorithms.

5.6.2 Comparison of the two algorithms

The time course of development of the elastic net and Kohonen-type algorithms is similar. There is a slight difference with regard to the onset of stripe formation: the vertical deformation of the cortical sheet is more gradual for the Kohonen-type algorithm than for the elastic net. Results for a range of separations of the two retinae and rate of reduction of k for the two algorithms are compared in figures 5.6 to 5.9.

Figures

For each simulation the following pictures are presented (see figures 5.6 to 5.9):

- The spatial positions of points in the cortical sheet, ignoring their positions in the ocularity dimension.
- The ocular dominance map, thresholded according to whether points in the cortical sheet are closer to the left eye or the right eye.
- The fourier transform of the (unthresholded) pattern of ocularity.
- The power spectrum of the fourier transform.

Parameters

Parameters used for the elastic net simulations were $\alpha = 0.2$, with β set so that the ratio β/α remained close to $1/2\mu$, where μ is the maximum expected distance between neighbouring points in the final state (taken to be roughly $2l$). This was to ensure proper matching of cortical points to retinal points, as discussed in [Simmen 1991]. k_{init}, the initial value of k, was taken to be 0.2. For the Kohonen-type algorithm simulations we used $\alpha = 1.0$, $k_{init} = 20.0$. In both cases there were 16×16 units in each retinal sheet, and 32×32 units in the cortical sheet: a ratio of two cortical units to every retinal unit. The retinal units were arranged in the unit square: thus the separation $2d$ between units within a retina was 0.0625. The "annealing rate" R is the factor by which k was multiplied at each iteration. Two annealing rates were investigated for each algorithm: $R = 0.95$ and $R = 0.98$, which for convenience we will refer to as "fast" and "slow" annealing respectively. In each case the computation time required for the two algorithms was roughly comparable: approximately 1.5 hours for fast annealing, and 3.0 hours for slow annealing (implemented in C on a Sparcstation1).

Initial conditions

Initial conditions were as shown in figure 5.5. These were defined by assigning each cortical unit an arbitrary vertical position between the two sheets, and a horizontal position randomly chosen with a uniform distribution within 0.5 (ie half the width of the retinal sheets) of its "ideal" topographic location. Results are shown after 200 iterations for fast annealing and 400 iterations for slow annealing: these are stable. Boundary conditions for the elastic net were used to ensure that, despite the lower number of neighbours for edge and corner points, the magnitude of the regularization term was roughly the same as that for interior points.

Stripe width and morphology

A general trend for both algorithms is that stripes tend to increase in thickness as the separation between the two layers increases. This is indicated quantitatively by the fact that the peak of the power spectrum shifts to lower frequencies as l increases. This is to be expected: as the retinae become more correlated, it is more desirable to keep the representations of corresponding positions in the two eyes close to one another.

Regarding stripe morphology, we see that there are general differences between the two algorithms. The elastic net produces patterns that are more regular, with longer stripes of a more uniform width. The Kohonen-type algorithm has a greater tendency to form blobs, especially for fast annealing. These qualitative impressions are confirmed by the fourier transforms of the stripe patterns and their power spectra: the spectra tend to show sharper peaks for the elastic net. We also note that, for a given separation, stripes produced by the Kohonen algorithm tend to be wider than those produced by the elastic net.

Topography

Qualitatively, it appears that the elastic net produces a more ordered topography for the map than the Kohonen-type algorithm. We now attempt to measure this difference quantitatively in two different ways. Firstly, the total distance D as measured in feature space between neighbouring cortical units was calculated:

$$D = \sum_i \sum_{j \in N(i)} ||\mathbf{y}_j - \mathbf{y}_i||,$$

where $N(i)$ refers to the four topological neighbours of cortical unit i. Values of D for all separations and both annealing rates for the two algorithms are shown in table 5.1. D inevitably increases as the separation

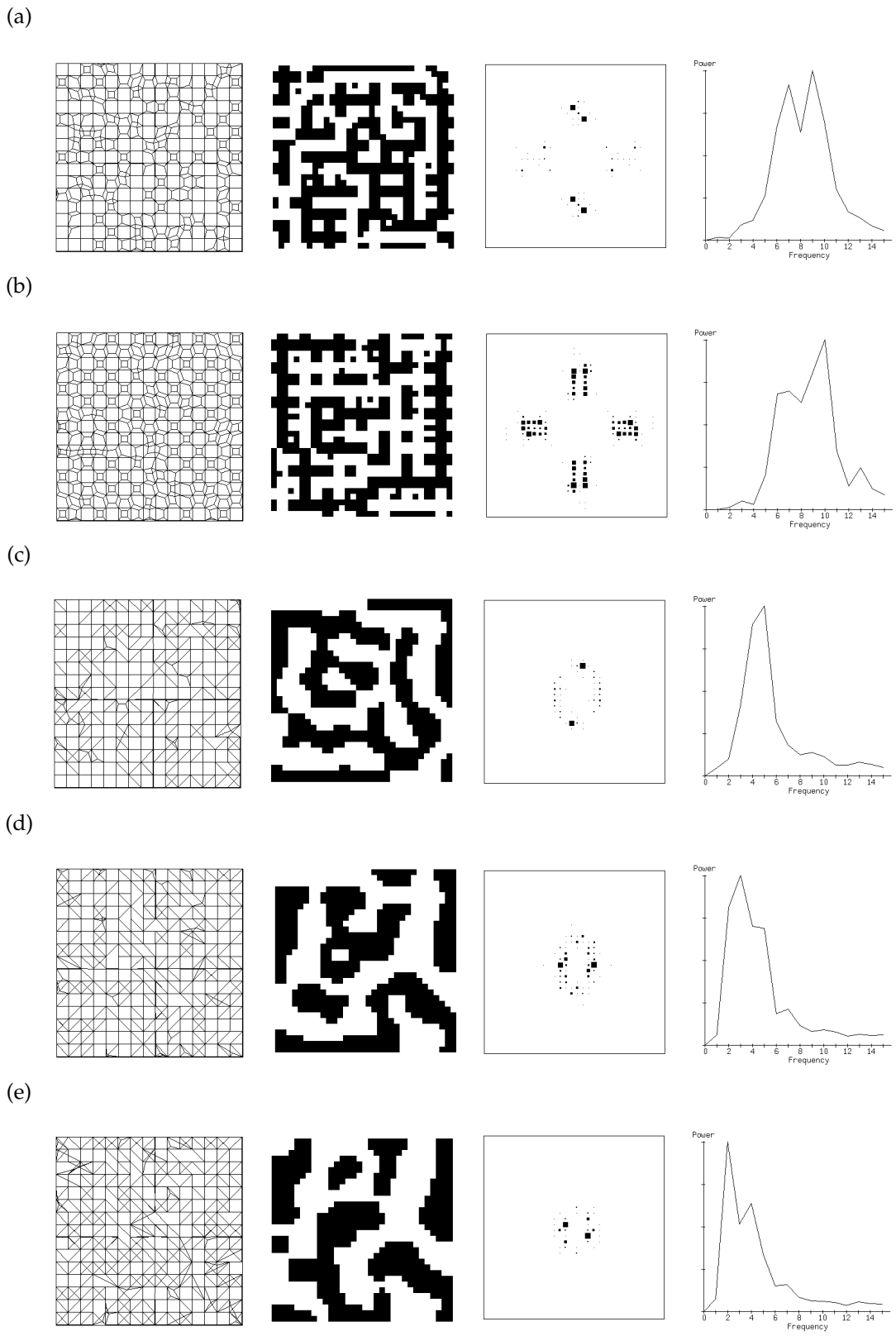

Figure 5.6: Topography, stripes, fourier transforms and power spectra for elastic net: fast annealing. A progression of separations $2l$ between the two layers of retinal units are shown: (a) $2l = 0.10$. (b) $2l = 0.15$. (c) $2l = 0.20$. (d) $2l = 0.25$. (e) $2l = 0.30$.

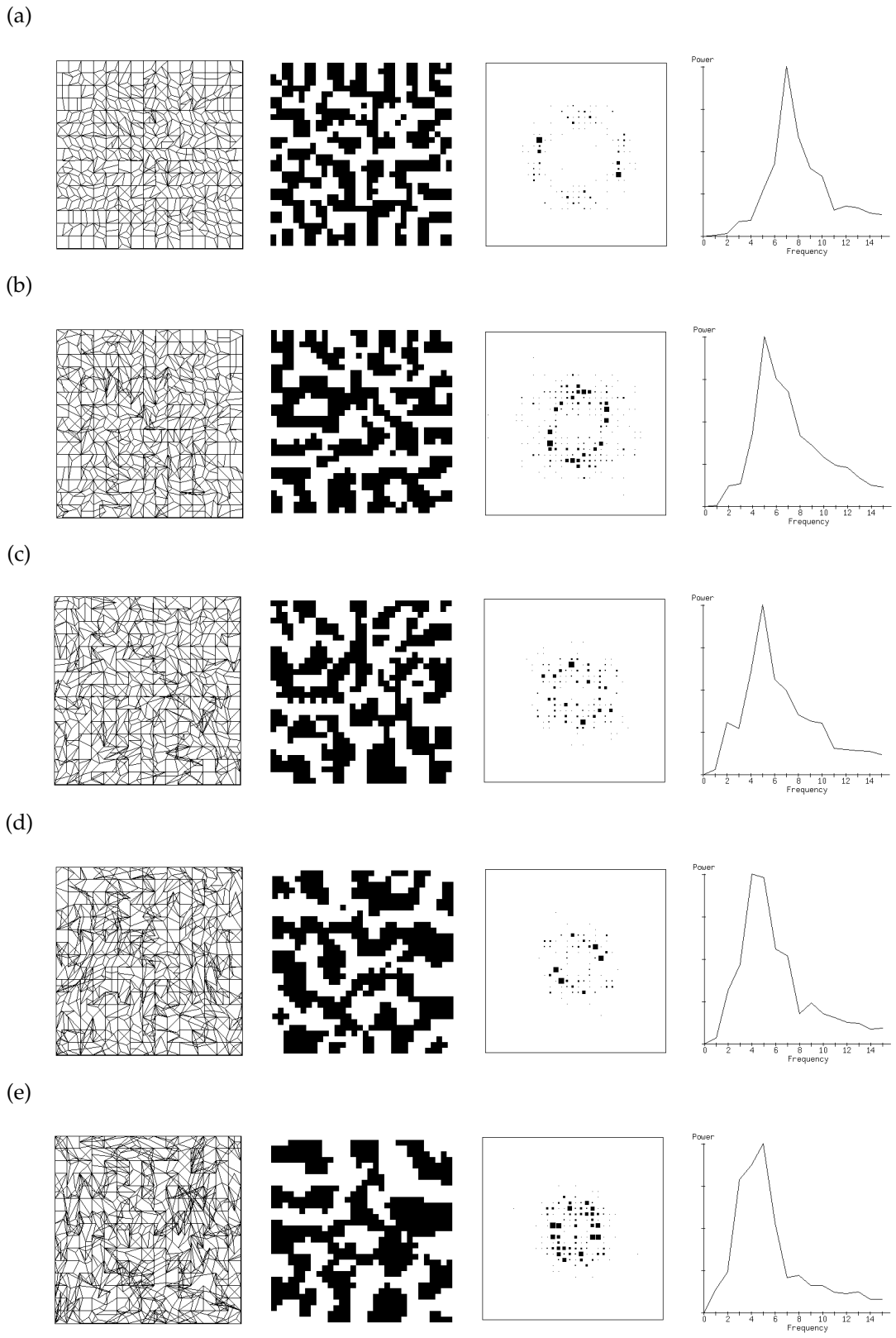

Figure 5.7: Topography, stripes, fourier transforms and power spectra for Kohonen-type algorithm: fast annealing. (a) $2l = 0.10$. (b) $2l = 0.15$. (c) $2l = 0.20$. (d) $2l = 0.25$. (e) $2l = 0.30$.

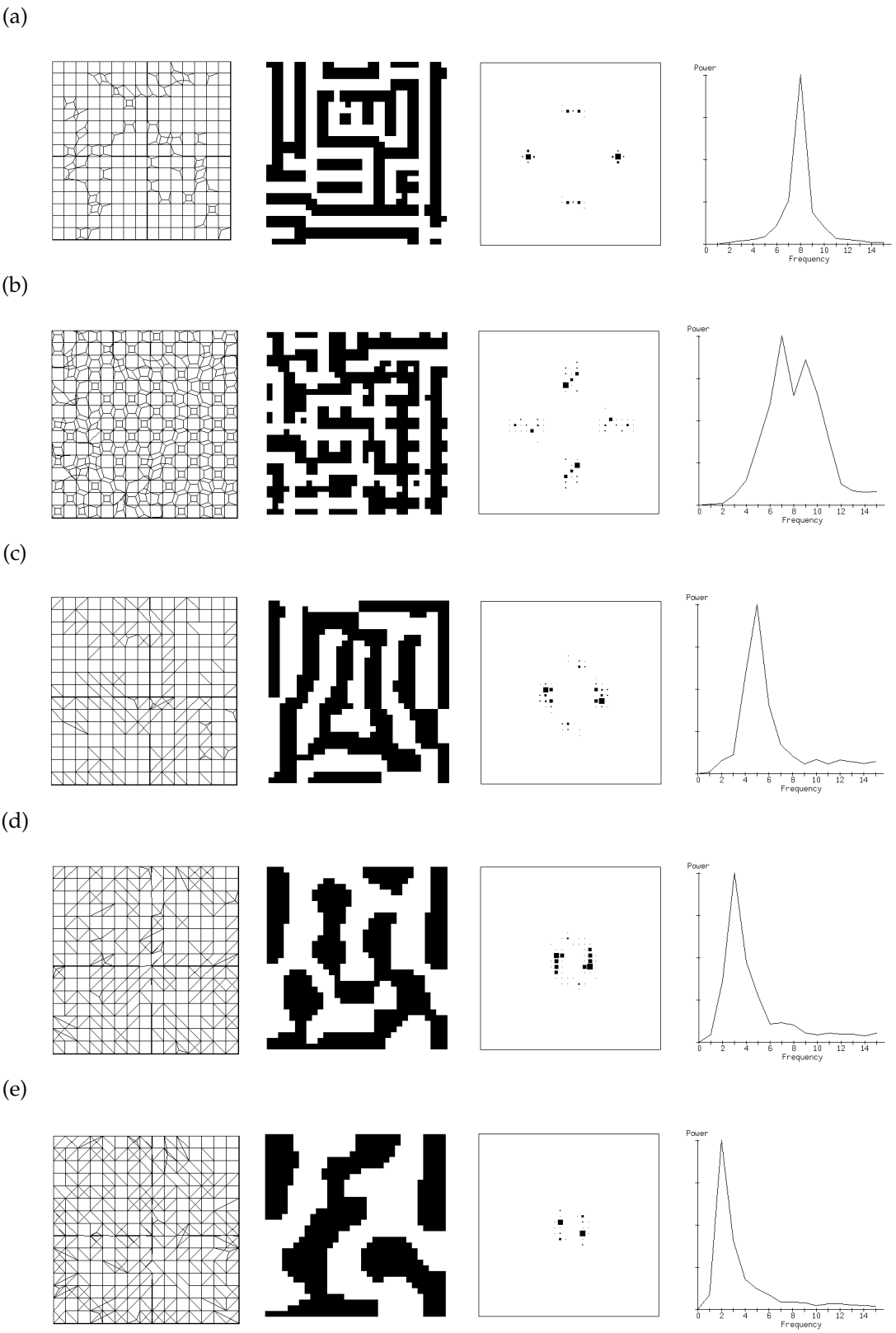

Figure 5.8: Topography, stripes, fourier transforms and power spectra for elastic net: slow annealing. A progression of separations $2l$ between the two layers of retinal units are shown: (a) $2l = 0.10$. (b) $2l = 0.15$. (c) $2l = 0.20$. (d) $2l = 0.25$. (e) $2l = 0.30$.

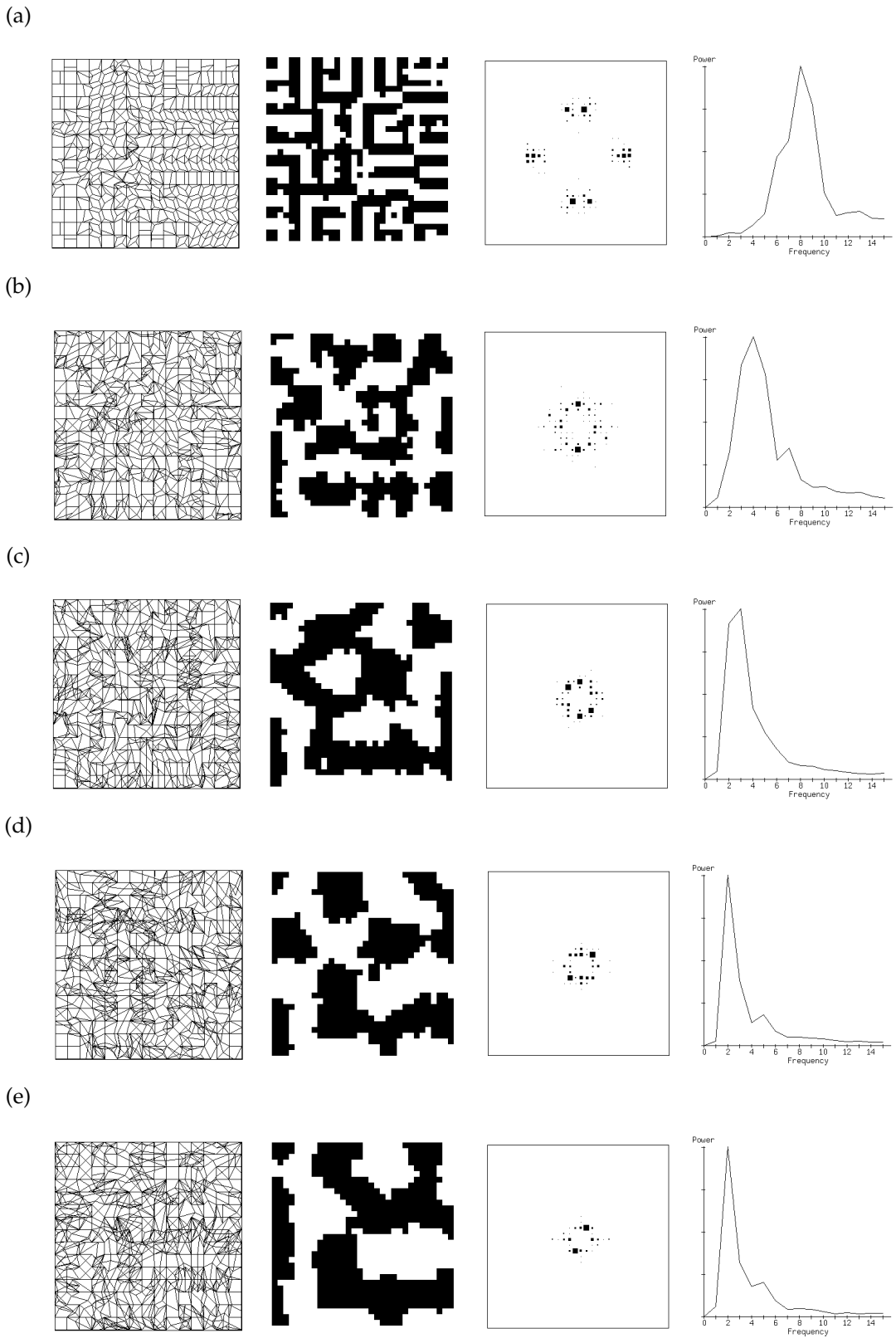

Figure 5.9: Topography, stripes, fourier transforms and power spectra for Kohonen-type algorithm: slow annealing. (a) $2l = 0.10$. (b) $2l = 0.15$. (c) $2l = 0.20$. (d) $2l = 0.25$. (e) $2l = 0.30$.

Separation of layers	Elastic net		Kohonen-type	
	Fast	Slow	Fast	Slow
0.10	111.2	114.3	121.4	119.7
0.15	135.6	133.7	151.7	137.5
0.20	155.7	155.8	176.9	148.5
0.25	165.3	161.4	198.6	160.3
0.30	180.1	165.3	215.0	168.1

Table 5.1: Distance D for the results shown in figures 5.6 to 5.9.

Separation of layers	Elastic net			Kohonen-type algorithm		
	L_N	L_C	Total	L_N	L_C	Total
0.10	4236.9	340.9	4577.8	4624.3	376.2	5000.5
0.15	4211.1	350.4	4561.6	4866.8	495.7	5362.5
0.20	4512.3	619.0	5131.3	4986.4	593.2	5579.6
0.25	4622.1	805.4	5427.5	5225.0	699.7	5924.7
0.30	4788.1	914.8	5702.9	5528.0	782.3	6310.3

Table 5.2: Length of cortical wiring required for the fast annealing cases.

of the retinae increases: however, for each separation, D is usually smaller for the elastic net than for the Kohonen-type algorithm. We also see that, as expected, a smaller D is usually produced by annealing more slowly.

Secondly, we implemented a simple cost function to measure length of wiring as in the minimal wiring principle: the total length L ($p = 1$ in the terminology of page 53) of wire required to join up all the cortical units that represent neighbouring or corresponding retinal units was calculated. The results are shown for fast annealing in table 5.2 and slow annealing in table 5.3, where besides the total length we also give the length of the neighbourhood, which we now refer to as L_N, and correspondence, L_C, terms separately. An inevitable effect of increasing stripe width is that the correspondence term increases. Although as discussed earlier it is not clear how the neighbourhood and correspondence terms should be weighted, they are each usually bigger for the Kohonen algorithm than the elastic net for each separation of the two retinae. Just the total wiring length is compared for both algorithms and both annealing rates in 5.4. Again it is clear that both quantitatively and qualitatively slower annealing usually gives better solutions, as expected.

Thus it appears that in general the elastic net is optimizing more effectively than the Kohonen-type algorithm. This would appear to account for the qualitative difference between the two sets of results, thus indicating that better solutions to the optimization problem can indeed imply improved stripe morphology.

It should of course be noted that the comparisons we have made between the two algorithms have only been for certain sets of parameters. Indeed, we have shown that the Kohonen-type algorithm annealed slowly can produce better results than the elastic net annealed quickly. Our criteria of equal run-time is only a crude comparison of the two algorithms, being somewhat implementation-dependent. However, it is encouraging that our quantitative results are in general agreement with the comparisons that have been performed in the TSP case [Angeniol et al 1988, Peterson 1990, Simmen 1992].

Separation of layers	Elastic net			Kohonen-type algorithm		
	L_N	L_C	Total	L_N	L_C	Total
0.10	4215.7	336.4	4552.1	4510.7	355.5	4866.2
0.15	4248.4	362.9	4611.3	4670.6	612.9	5283.5
0.20	4378.5	546.7	4925.2	4847.0	791.6	5638.6
0.25	4654.5	823.4	5477.9	4846.6	927.1	5773.7
0.30	4887.1	1094.2	5981.3	4839.5	1010.7	5852.2

Table 5.3: Length of cortical wiring required for the slow annealing cases.

Separation	Elastic net		Kohonen-type	
	Fast	Slow	Fast	Slow
0.10	4577.8	4552.1	5000.5	4866.2
0.15	4561.6	4611.3	5362.5	5283.5
0.20	5131.3	4925.2	5579.6	5638.6
0.25	5427.5	5477.9	5924.7	5773.7
0.30	5702.9	5981.3	6310.3	5852.2

Table 5.4: Comparison of total cortical wiring for fast and slow results for both algorithms.

(a) (b)

Figure 5.10: Results of one run of the (a) elastic net and (b) Kohonen-type algorithms for $2 \times 32 \times 32$ retinal units mapping to 64×64 cortical units, $2l = 0.15$, slow annealing.

Larger scale

One larger-scale simulation was performed for each algorithm (two 32×32 retinal sheets mapping to one 64×64 cortical sheet), and the results shown in figure 5.10. Similar conclusions apply to this case as to the smaller-scale cases: the stripes produced by the elastic algorithm are more uniform. Note the marked difference in the stripe width produced by the two algorithms.

5.6.3 Monocular deprivation

Monocular deprivation (reducing the competitive strength of one eye) was modelled in both algorithms by using a different value of α for each eye (this was first studied for these algorithms in [Goodhill 1990]). The amount of deprivation, or "deprivation parameter", was defined to be the ratio of α for the deprived eye to α for the normal eye. The variation of stripe pattern with the size of this deprivation parameter is shown for the elastic net algorithm in figure 5.11, and for the Kohonen-type algorithm in figure 5.12.

Firstly, we see the desired result that for both algorithms the deprived eye takes over less of the cortex than the normal eye: this effect is more marked for the elastic net than the Kohonen-type algorithm. However, there appears to be a significant difference in the way that this occurs for the two algorithms. For the elastic net, the stripes become thinner but remain continuous. For the Kohonen-type algorithm, the mapping from the deprived eye appears rather to break up into a series of blobs, each blob being of roughly the same size as the original stripe thickness. The degree of deprivation appears to exert a stronger effect on the elastic net than on the Kohonen-type algorithm. Both sets of results also reflect the biological data that for the deprived case stripe periodicity (i.e. sum of widths of neighbouring right- and left-eye stripes) remains the same as in the

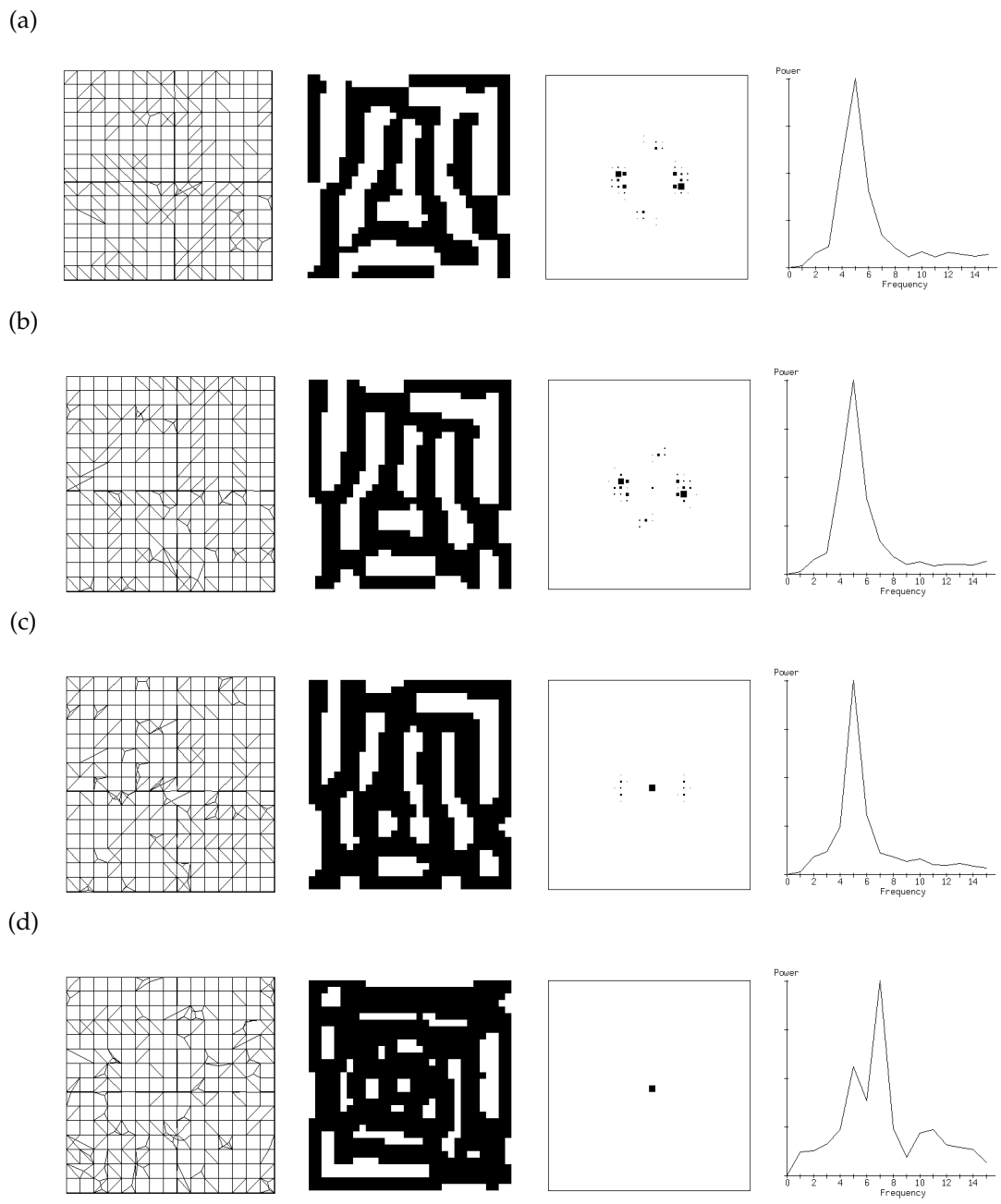

Figure 5.11: Effects of monocular deprivation - elastic net algorithm. In all cases slow annealing and a separation of retinae of 0.2 was used. (a) Normal case. (b) Deprivation parameter = 0.8. (c) Deprivation parameter = 0.6. (d) Deprivation parameter = 0.4.

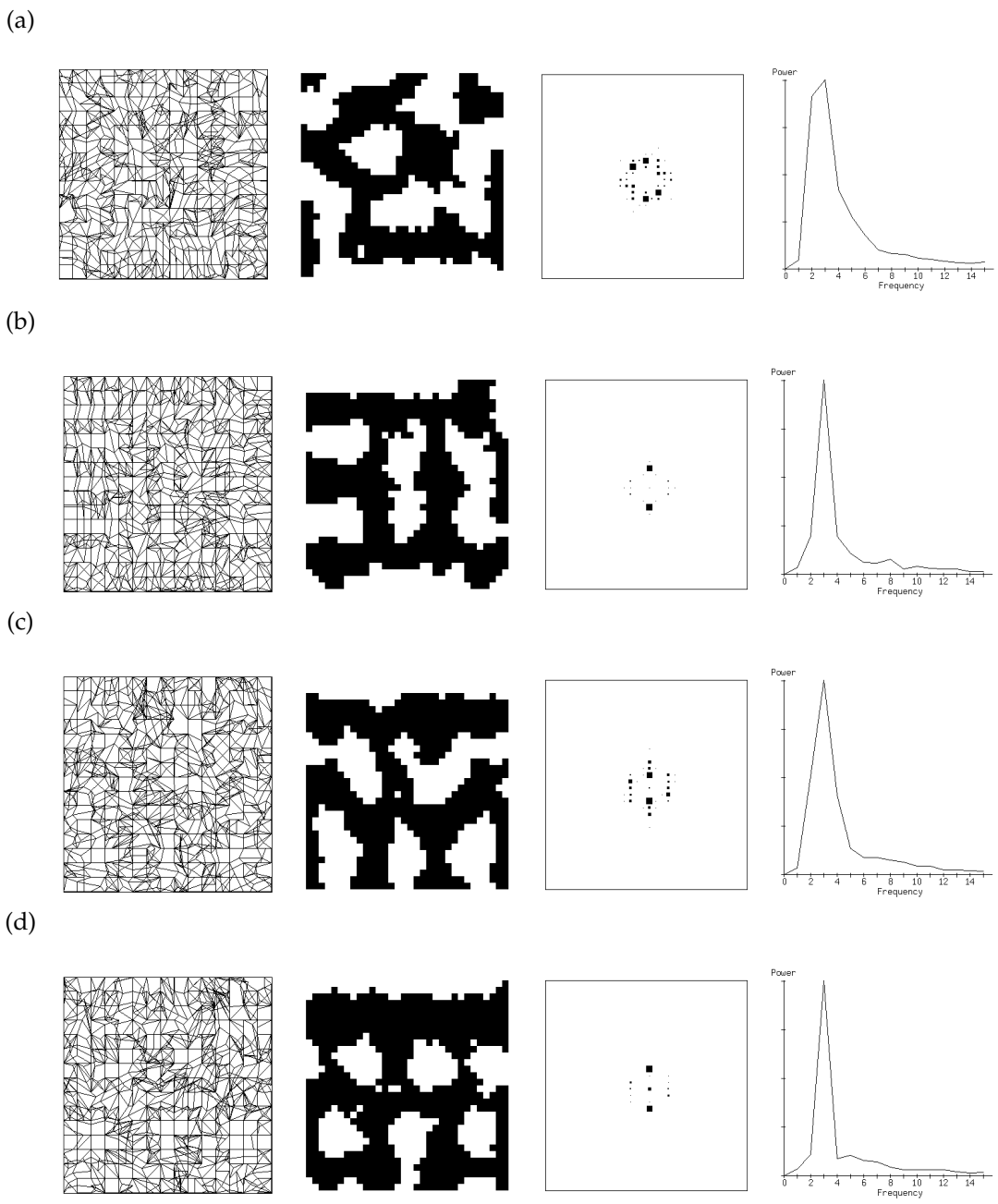

Figure 5.12: Effects of monocular deprivation - Kohonen-type algorithm. In all cases slow annealing and a separation of retinae of 0.2 was used. (a) Normal case. (b) Deprivation parameter = 0.8. (c) Deprivation parameter = 0.6. (d) Deprivation parameter = 0.4.

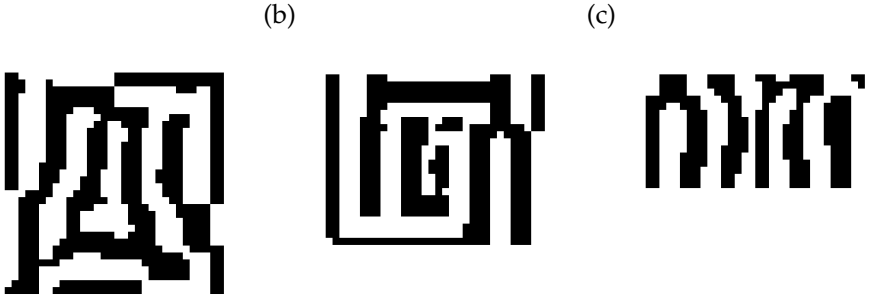

Figure 5.13: Testing the prediction of [LeVay et al 1985] regarding stripe orientation. (a) Normal case: 32×32 points in cortical sheet. (b) 32×24 points in cortical sheet. (c) 32×16 points in cortical sheet. It can be seen there is an increasingly strong tendency towards stripes oriented parallel to the short axis as the cortical sheet becomes more anisotropic. Slow annealing was used.

normal case, as shown by the constant position of the peak of the power spectrum.

5.6.4 Anisotropic boundary conditions

One interesting phenomenon of natural stripes is that they exhibit global order: stripes remain parallel over large distances, which might not be expected from the operation of purely local mechanisms as are present in most models. [LeVay et al 1985] provides a review of theoretical ideas regarding possible sources of an anistropic ordering influence, and hence how such global order might arise. Such ideas include elongation of geniculocortical arborizations [von der Malsburg 1979], anisotropic growth of the cortex [Swindale 1980], and differences in the strength of two orthogonal gradients of adhesiveness [Fraser 1985]. In [LeVay et al 1985] however it is suggested that the effect could be due to anisotropic boundary conditions:

> "The largest of the LGN layers are roughly circular in outline....whereas the cortex is approximately twice as long as it is wide, with the long axis oriented horizontally in our standard format. Thus, to a first approximation, it is reasonable to regard the geniculocortical mapping problem as one of mapping two circular discs, one for each eye, onto a horizontally elongated ellipse. From a global standpoint, the simplest way of achieving this is to slice each disc into thin vertical strips and then to interdigitate the left and right eye strips, thereby forming an appropriately elongated ellipse. If, instead, one started with discs cut into horizontal stripes that were stacked vertically, the strips would have to be stretched horizontally or shrunk vertically by 4-fold in order to obtain the requisite elongation. More generally, the only way to obtain elongation along one axis without anisotropic stretching of the inputs is to interdigitate stripes cut orthogonal to that axis." [LeVay et al 1985]

We tested this general hypothesis by performing experiments with the elastic net algorithm as above except that the cortical sheet was now rectangular. Two cases were considered: 32×16 and 32×24 cortical units, in order to test the amount of anisotropy required to produce global ordering. In both cases the separation between retinae was 0.2. Results are shown in figure 5.13.

In the first case almost entirely vertical stripes are formed, while in the second there is a strong bias towards vertical over horizontal, agreeing with LeVay et al's prediction (though for the squares-rectangles rather than circles-ellipses case). We suggest that this effect could perhaps provide another test of the adequacy of models of ocular dominance in general. It would not be expected to occur in models where there is no consideration of topography (such as [Swindale 1980]), since the effect relies on the undesirability of "stretching", i.e. deformations in topography in certain directions.[3]

5.7 Discussion

We address the following three issues:

[3]The hypothesis of [LeVay et al 1985] has also recently been tested in [Jones et al 1991]. They exactly minimized a measure of topographic distortion of mapping from two LGNs, and showed that boundary conditions had the predicted effect. However, they did not discuss how such an optimization could be carried out in the biological system, or what determines the width of the stripes in this framework.

- **Representation:** What is the meaning of formulating the problem in these geometric terms, and what is its relationship to biology?
- **Algorithms:** How important is the choice of algorithm? How do the results (e.g. stripe width) depend on the parameters of the model?
- **Relation to other work:** What relation do both this representation and these algorithms bear to other models of topography and ocular dominance?

5.7.1 Representation and algorithms

The notion of a feature space containing the inputs it is desired to represent has been used extensively (see e.g. [Kohonen 1988]). However, a problem that always arises when dimensions having different meanings (such as orientation and position, or ocularity and position) are regarded as the axes of a feature space is that of how "distances" between these features are to be measured, and what such distances actually mean. It is not immediately clear in our case why distances in the ocularity dimension should contribute in exactly the same way as distances in the spatial dimensions. Two arguments can however be made to support this interpretation. Firstly, in the model we directly control the ratio of the separation of units within and between retinae: we assume that they can both be measured in the Euclidean metric, but not what their weightings should be. Secondly, due to our interpretation of distances as correlations, we are claiming that the two types of distance really are measuring the same kind of thing: the correlation between units, whether these units be in the same eye or in different eyes (the same argument also applies to an interpretation in terms of chemical concentrations).

We turn now to the issue of how it was chosen to situate points (retinal units) in the feature space. This brings us directly to the two assumptions mentioned in section 5.3 above. It should be emphasised that, given the correlational interpretation, these are empirical questions: their validity can (at least in principle) be determined by biological experiment. Assuming that there is a relationship between the position of units in feature space and in physical space, and that the same ordering of units applies to both, is essential for the approach to ensure that a mapping that is topographic in feature space is also topographic in physical space. Whether this assumption is valid in reality depends on both the structure of the visual world and on the nature of spontaneous activity in the retina: as we saw in chapter 2, both play a part in map development. The statistics of natural scenes have been investigated in [Field 1987], who found the existence of correlations at all length scales (see also [Hancock et al 1992]. These correlations decrease with distance, as in the representation here. Similarly, [Mastronade 1989, Meister et al 1991] showed that spontaneous firing of retinal ganglion cells was correlated between neighbours and this correlation decreased with distance. It is clear the assumption that the correlation between two units is a function only of their separation in feature space (and hence physical space) is not true on a large scale, due to non-uniform sampling of the visual scene by the retina and non-uniform receptor density over large distances. However, on a local scale, which is the scale addressed in the simulation results presented above, the assumption may be approximately valid. Lastly we observe that the fact that the arrangement of points in feature space is highly regular (in contrast to for instance [Durbin & Mitchison 1990] and [Obermayer et al 1990] (discussed later), algorithms which choose points randomly in their respective feature space according to particular distributions) may be influencing the results. One elastic net experiment that was tried to investigate this was to randomly displace each retinal unit a small distance ($<< \mathtt{d}$) from its initial location so as to break the symmetry. Although qualitatively the final stripe pattern was similar, there was a slight difference in the course of development. In the randomly displaced case, the cortical sheet did not become entirely flat, and vertical movement of cortical units during stripe formation was more gradual than in the undisplaced case.

Optimum stripe widths

The optimum stripe width (i.e. the pattern of visitation of retinal units by the cortical sheet that minimizes the sum of distances between neighbouring cortical units in feature space) can in theory be calculated directly for this representation for any combination of the parameters $\mathtt{l}$ and $\mathtt{d}$. This was investigated for the one-dimensional case in [Goodhill & Willshaw 1990], where the following results were proved:

- The $\sqsupset$ path (i.e. the 1-D rope runs the entire length of one retina, then changes to the other retina and runs back in the opposite direction: the mapping segregates into non-overlapping regions) is the shortest path for $\mathtt{l} > \mathtt{d}$.
- Stripes of width 2 units gives the shortest path for $\mathtt{l} < \mathtt{d}$.

- Stripes of width n units have minimum length when

$$n = 1 + \frac{l^2}{d^2} \tag{5.8}$$

From this last result it follows that the Z path is always longer than the best striped path (the Z path is the counterpart of the $\sqsupset$ path, where now the topographies are laid out side by side rather than being reflected: the rope travels completely from one end of the left eye to the other, and then doubles back and travels the complete length of the right eye in the same direction).

We also note that the elastic net actually minimises distance in the L_2 norm. In this case the optimal stripe width is

$$n = \frac{2l}{d} \tag{5.9}$$

Comparing this with equation 5.8, we see that optimum stripe width increases more slowly with l/d in the L_2 norm than the L_1 norm.

It is difficult to generalize these calculations regarding stripe width to the two dimensional case of interest here, since there is now a much richer space of possibilities for the pattern of stripes, even assuming uniformity. However, it is clear that a similar argument for the generalization of the $\sqsupset$ path in 2-D also holds. Thus we have the significant result that for $l > d$, the case for all the simulation results presented in this chapter, the mapping corresponding to two non-overlapping regions always has a shorter path length than any striped mapping. Why do these algorithms not find these solutions? A possible explanation is that, for the range of l/d we have considered, the two retinae are cooperating over topography but competing over ocularity. Each pair of corresponding retinal units attempts to attract the cortical point(s) in the same topographic location.

5.7.2 Related work

The approach described in this chapter is related to work applying Kohonen's algorithm to a somewhat similar feature space. In [Kohonen 1988, Ritter & Schulten 1988] pictures are briefly presented of an elastic sheet deforming within a three-dimensional block of feature points regularly spaced through the volume of the block, with the suggestion that this could be related to ocular dominance stripe formation. However this is not pursued: this example is just given to illustrate their main point of "automatic selection of feature dimensions". Kohonen-type algorithms cause an elastic rope/sheet to align itself along the main axis/axes of variation in the feature space, with (small) deformations to take account of feature units not on this main axis/axes.

A detailed application of Kohonen's algorithm to a feature space similar to the one investigated here, and the orientation-position case of [Durbin & Mitchison 1990], has been undertaken in [Obermayer, Ritter & Schulten 1991, Obermayer, Blasdel & Schulten 1991]. Although the model was originally presented as a pattern-based competitive learning model [Obermayer et al 1990] (as in the model to be presented in chapter 6), recently these authors have placed greater emphasis on a more abstract, feature-mapping interpretation. This model has already been described in chapter 3. Two important differences between this model and the elastic model are that [Obermayer, Ritter & Schulten 1991] use a set of feature points located at random in the space, and also sometimes consider arbitrary degrees of dominance between the two eyes, as described in chapter 3, whereas we use only feature points which are entirely left-eye or right-eye.

5.8 Conclusions

We have shown that the elastic net and Kohonen-type algorithms can both form topography and ocular dominance from the same mechanisms, and we have investigated their behaviour over a range of parameters. Various measures of the quality of the maps produced have been defined, and these show that in general the elastic net performs slightly better than the Kohonen-type algorithm for parameters chosen such that the time taken by both algorithms is comparable. In addition we have demonstrated that both algorithms can successfully model the effects of monocular deprivation. The behaviour of the elastic net in the case of anisotropic boundary conditions has been shown to agree with the prediction of [LeVay et al 1985].

Chapter 6

A New Competitive Model

6.1 Introduction

So far in this thesis we have studied in some detail two different models of map formation. To recapitulate:

- Miller's model is based on a correlational rule, and has been presented by Miller as a reasonably realistic model: he links some model parameters with biological parameters. Although accounting for ocular dominance stripe formation when correlations within each eye are identical and there are non-positive correlations between the eyes, we have shown that (a) the model is unstable to the addition of perturbative correlations making the two eyes different, and (b) addition of small positive correlations between the eys or small negative correlations within each eye can favour binocular over monocular development. In addition, it has not been shown that the same mechanism can generate topography.
- The elastic net model is based on a more abstract developmental rule, and its parameters are harder to interpret biologically. However, it does form topography and ocular dominance from the same mechanism, and has the virtue that its behaviour can be understood in terms of optimizing an objective function.

Neither of these models, or any of the models we reviewed in chapter 3, has been shown to satisfy *all* of the criteria presented in the introduction. As a reminder, these are as follows: (a) forms topography and ocular dominance from the same mechanisms, (b) is expressed in terms of mechanisms and parameters that can be naturally expressed biologically, (c) develops appropriate mappings in the presence of positive correlations between the two eyes, and (d) produces stripe patterns with similar characteristics to natural stripes.

In this chapter we introduce a model which goes further to simultaneously satisfying these criteria than any model yet presented. This new model consists (as usual) of two presynaptic sheets of units innervating a single postsynaptic sheet of units, and is based on the following principles. At each time step, *distributed* patterns of activity are presented simultaneously in both presynaptic sheets. The postsynaptic units compete for the right to respond to the input pattern. The strongest-responding unit and its neighbours have the weights on their input lines updated. Weights are then normalized by a *subtractive* normalization rule.

We first explain the basic mechanisms of the model, and then present some results, drawing particular attention to the predictions that the model makes regarding stripe width. We then discuss in some detail the important issue of the type of normalization used in the model, and present some analysis of this for a simple case. Finally, we explain how this model relates to some of the other models we have discussed. A preliminary account of this work appeared in [Goodhill 1991], and an earlier version of the model was discussed in [Goodhill 1990].

6.2 Formulation of the model

6.2.1 The framework of the model

The model consists of two two dimensional sheets of input cells (indexed here by r) connected to one two-dimensional sheet of output cells (indexed by c) by fibres with variable synaptic weights w_{cr}. It is assumed that the topography of the retina is essentially unchanged by the LGN on its way to the cortex. In addition, the effects of retinal and LGN processing are taken together, and thus for simplicity we refer to the input layers of the model as being "retinae". It should be understood that the model is formulated at a general enough level to be applicable to both the retinocortical and the retinotectal systems.

Both retina and cortex are arranged in square arrays. All weights and cell activities are positive. Lateral interactions exist in the cortical sheet of a circular center-surround excitation/inhibition form [von der Malsburg 1973], although these are not modeled explicitly. The initial pattern of weights is random apart from a small bias that specifies an orientation for the map (similar to the initial conditions in the Tea Trade Model [Willshaw & von der Malsburg 1979]). At each time step, a pattern of activity is presented by setting the activities a_r of retinal cells. Each cortical cell c calculates its total input x_c according to a linear summation rule:

$$x_c = \sum_r w_{cr} a_r$$

We make the following assumptions to speed calculation at each time step (the significance of these assumptions is discussed later):

- Lateral inhibition in the cortex is sufficient to suppress the activity of all but the cell g with maximum response, and cells in a small neighbourhood of g. Cells close to g have an output that decreases as a function of distance from g.

- The "winning" cell g is the cell that has the largest activity when driven purely by the retina, before taking into account the effect of lateral feedback.
- The input-output function of the cortical cells is monotonic, so that g is simply the cortical cell with the greatest input.

This winning cell and its neighbours have their weights updated at each time-step in a Hebbian manner by adding in a small fraction of the input pattern:

$$w_{cr} = w_{cr} + \alpha a_r s(c, g) \tag{6.1}$$

α is a small positive constant, and s is the function that specifies how the activities of cells c near to g decrease with distance from g. We assume s to be a gaussian function of the Euclidean distance between cells in the cortical sheet. So far these assumptions are similar to those made by Kohonen (see e.g. [Kohonen 1988]). However, an important difference is in the form of the inputs to the model, and the normalization rule used to maintain weights within bounds.

6.2.2 The inputs to the model

Input to the model are random dot patterns with short range spatial correlation introduced by convolution with a blurring function. Patterns were generated by assigning the value 0 or 1 to each pixel in the input array with a fixed probability (usually 50%), and then convolving this array with a gaussian function of standard deviation σ_r. This produces patterns of activity with a range of correlation determined by σ_r. It is important to note that these patterns are *distributed*: all cells in the input layer may be active simultaneously.

6.2.3 Efferent normalization

To prevent weights evolving in accordance with the equations above from growing without bounds, some form of constraint on weight values is needed. Here we assume this constraint to be in the form of a normalization rule: each cortical cell has a fixed amount of "weight substance" to share out amongst the efferent fibres that synapse on it. [Miller 1990(a)] and [Miller & MacKay unpublished] have drawn attention within the framework of a correlational model to the fact that there is a sensitive interaction between the statistics of the inputs, the form of the normalization of efferent fibres contacting a postsynaptic cell, and the receptive fields developed by units in such models. They highlight two ways in which either a linear or squared constraint on synaptic strengths may be enforced: *divisively* (as in e.g. [von der Malsburg 1973]) and *subtractively* (as in e.g. [Miller et al 1989]). In the former (which is more common), the normalization is enforced by dividing each weight by the same amount. In the latter, the normalization is enforced by subtracting the same amount from each weight. As one example of the difference in the behaviour of these two rules, [Miller & MacKay unpublished] focus on ocular dominance segregation. They show that a postsynaptic cell evolving under a correlational learning rule as in [Miller et al 1989] with divisive enforcement will only become dominant for one or the other eye if there exist *anti*correlations between the two eyes. They further argue that this constraint does not exist for a subtractive enforcement rule (but see chapter 4).

Although the correlational framework analysed by Miller and Mackay is somewhat different to the competitive model presented in this chapter, the simulation results we present suggest that a similar conclusion applies to the current model with respect to ocular dominance segregation. The same argument applies to receptive field refinement. With divisive enforcement, cortical unit receptive fields remain distributed over the entire retina for the inputs used here. They do not refine to a small localized region unless retinal cells are anticorrelated with each other (an example is shown below). A form of subtractive enforcement was therefore employed in the current model, as follows.

For each cortical unit, the quantity t is calculated:

$$t = \sum_r w_{cr} - N_c$$

where N_c is the total weight available to each cortical cell (a constant). Define t' to be t divided by the total number of retinal units for which $w_{cr} \neq 0$. Then

$$w_{cr} = \begin{cases} w_{cr} - t' & \text{if } w_{cr} - t' > 0 \\ 0 & \text{otherwise} \end{cases}$$

Parameter	Value
Number of retinae	2
Number of retinal units in each dimension	16
Number of cortical units in each dimension	32
Width of bounding box for initial weights as proportion of total width	0.8
Width of uniform noise distribution added to each weight	0.1
Learning rate = α	0.01
Constant for efferent normalization = N_c	10.0
Width of cortical interactions σ_c	1.5
Range of positive part of retinal convolution σ_r	1.5
Parameter controlling correlation between two retinae = h	0.15
Number of input patterns presented	350,000

Table 6.1: Parameters for initial results.

t is now recalculated: if $t \neq 0$ (i.e. some weights have become zero), divisive normalization is applied:

$$w_{cr} = \frac{N_c w_{cr}}{\sum_r w_{cr}}$$

We now present results of the model. This is followed by a more detailed exposition of some aspects of the algorithm. We then discuss the issue of efferent normalization again in more detail, and present some analysis of this in a simple competitive setting.

6.3 Results

We first study a typical case of the model in action. Later we investigate the effect of various parameters on model results.

6.3.1 Parameter values

For our typical case (figures 6.1, 6.2, 6.3, 6.7(c) and 6.8(c)), the parameter values of table 6.1 were used. Further parameters will be given below as they are discussed. Initial bias in the weights, representing a crude polarity matching machanism, was defined similarly to the Tea Trade model [Willshaw & von der Malsburg 1979]. Correlations within an eye were produced as explained above by convolving with a gaussian: all within-eye correlations were positive, and the two eyes were correlated to a certain degree (the meaning of the parameter h will be explained later). Figure 6.1 shows the development of the map, and figures 6.2, 6.3, 6.7(c) and 6.8(c) show the topography and ocular dominance properties of the final state. The meaning of these pictures is now explained: we start with figure 6.2 and return to figure 6.1 later.

6.3.2 Ocularity

In figure 6.2 cortical units are coloured white or black if they are primarily dominant for the left or right eyes respectively, and grey otherwise. The threshold for a unit to be classified as monocular was taken to be that the sum of a unit's weights for one eye is at least 4 times bigger than the sum of its weights for the other. The development of monocularity of each unit is shown on a continuous scale in figure 6.1(d-f). Here the size of the square for each unit is proportional to the monocularity of the unit, and the colour of the square represents for which eye is is dominant. All units are initially binocular, and become gradually more dominant for one eye or the other over time. Longer runs showed that eventually all units become monocular: however the last few units take an extremely long time to saturate and thus runs presented here were terminated when the final properties of the map were clearly visible. Note that all remaining binocular units are on the borders between stripes. Once saturated, weights are stable (see below for further discussion). It can be seen that a pattern of stripes is formed. That is, regions of same-eye dominance are connected into regions that tend to be longer than they are wide, as opposed to blobs.

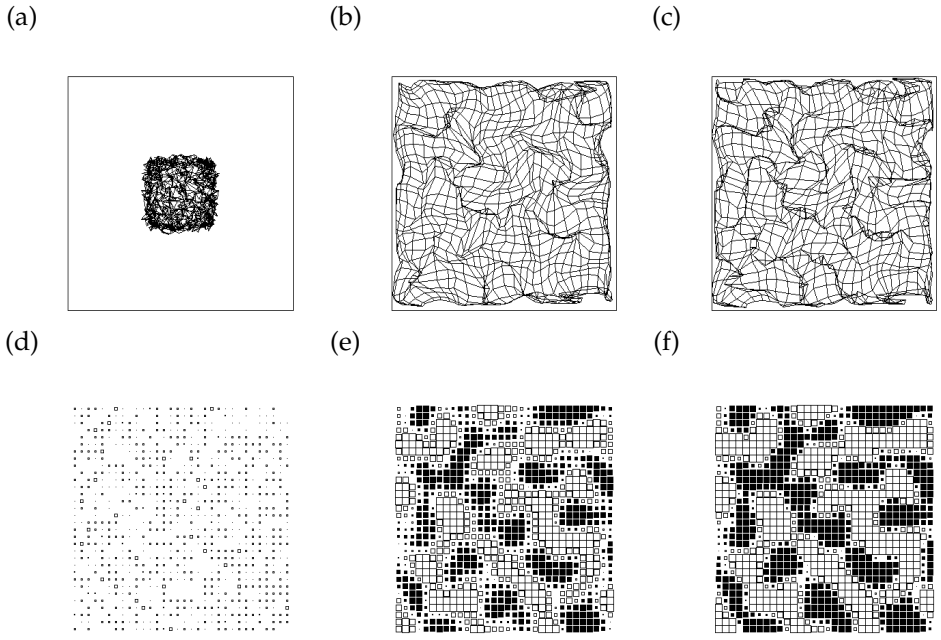

Figure 6.1: Time course of development of the model. (a-c) show the cortical topography averaged over both eyes (see text for further explanation) at 0, 50,000 and 100,000 iterations (i.e. different input patterns presented) respectively. (d-f) show the ocularity of each unit in the cortical array at 0, 50,000 and 100,000 iterations respectively.

6.3.3 Topography and its interaction with ocular dominance

In figure 6.2, each monocular cortical unit contains a Hinton diagram of that unit's weights (i.e. efferent connection strengths) for the eye for which it is dominant. This means that there is a 16 by 16 grid within each cortical unit with each grid point representing a retinal unit, and the size of the box at each grid point encodes the strength of the connection between each retinal unit and the cortical unit. For binocular (grey) units, the larger of the two corresponding weights in the two eyes is drawn at each position, coloured white for the left eye and black for the right eye.

It can be seen that neighbouring positions in each eye tend to be represented by neighbouring cortical units, apart from discontinuities across stripe boundaries. For instance, the bottom right corner of the right retina is represented by the bottom right cortical unit, but the bottom left corner of the right retina is represented by cortical unit (3,3) (counting along and up from the bottom left corner of the cortex), since unit (1,1) is representing the left eye.

The topography can be seen more clearly in figures 6.3, 6.7(c) and 6.8(c). In figure 6.3 the receptive fields in the cortex of retinal units are represented. For both eyes, each retinal unit is shown as a square containing a Hinton diagram of its weights to the cortex. Firstly, note that each retinal unit is represented in the cortex. Secondly, we see that the map is broadly topographic: moving across each retina, one observes a continuous shift in the cortical position at which that point in the retina is represented, apart from discontinuities at stripe boundaries. Thirdly, retinal units occasionally arborize to contact two stripes. Fourthly, we see that corresponding units in the two eyes are usually represented close to one another in the cortex.

In figures 6.1(a-c), 6.7(c) and 6.8(c), the centre of mass of the weights of each cortical or retinal unit is represented as a point in retinal or cortical space respectively, and neighbouring units are connected by lines to form a grid (similar to [Willshaw & von der Malsburg 1976]), except that now we have two sets of pictures, one for each eye. There are three types of picture:

- Retinal topography for each eye (figure 6.8(c)). For each retinal unit, the centre of mass of the weights for each retinal unit is plotted as a point in cortical unit space. It can be seen that for both eyes this map is locally continuous, with gaps (complementary between the two eyes) left for regions of the cortex where the other eye is dominant. However, since we have seen from the retinal weights pictures (figure 6.3) that retinal units can occasionally arborize to contact two stripes, retinal topography pictures are slightly unreliable indicators of topography, since the centre of mass of weight for a retinal unit with contacts evenly balanced between two stripes will be in the middle of the stripe for the other eye.

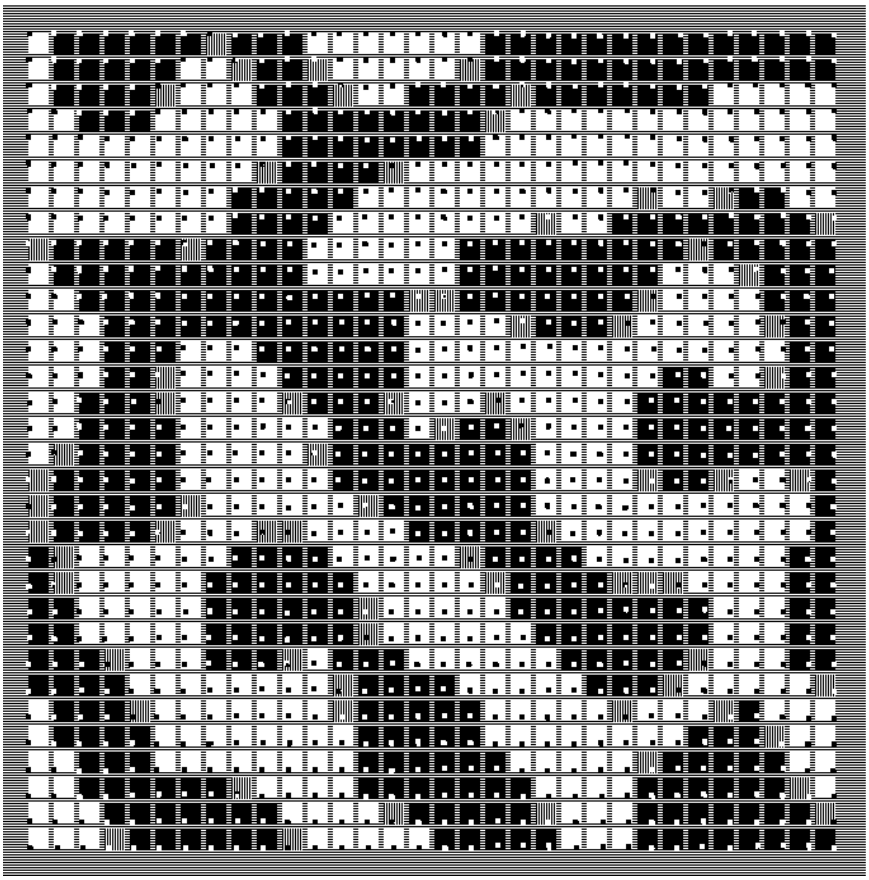

Figure 6.2: The receptive fields of cortical units. Units are coloured white if they are strongly dominant for the left eye, black if they are strongly dominant for the right eye, and grey if they are primarily binocular. Within each unit is a representation of its receptive field (see text for further details).

(a) (b)

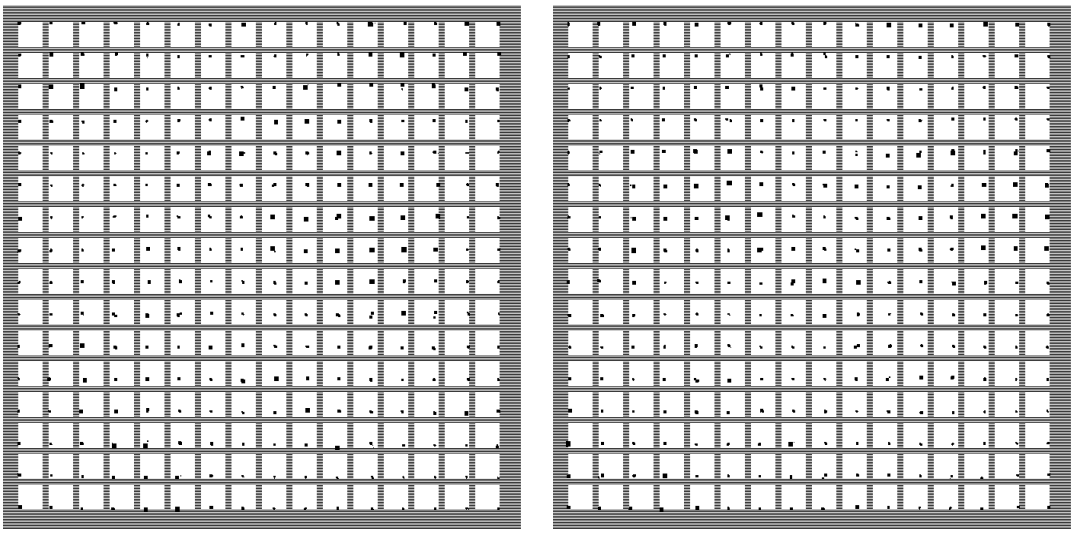

Figure 6.3: The receptive fields in the cortex of retinal units. (a) Left eye. (b) Right eye. Each white box represents a retinal unit, and inside each box is a representation of that unit's connections to the cortex.

- Cortical topography for each eye (figure 6.8(c)). For each cortical unit, the centre of mass of the weights for that unit is plotted as a point in retinal unit space. The "holes" in the map arise where cortical units are dominant for the other eye, and are largely complementary between the two eyes. Comparing these picture with the ocularity picture of figure 6.7(c), we see that for each picture topography is continuous within a stripe.
- Cortical topography for both eyes (figures 6.1(a-c), 6.7(c)). Here the centre of mass of weights for each cortical unit is averaged over *both* eyes, imagining the retinae to be lying atop one another. This type of picture reveals where the map is folded to take into account that the cortex is trying to represent both eyes. We see here that discontinuities in terms of folds tend to follow stripe boundaries: first particular positions in one eye are represented, and then the cortex "doubles back" as its ocularity changes in order to represent corresponding positions in the other eye.

Thus the map appears to proceed smoothly for one eye within a stripe, and then double back and "pick up where it left off" to a certain degree for the other eye. This agrees with the experimental data for the natural system (see e.g. [Hubel & Wiesel 1977]).

6.3.4 Development

Various stages of the development of the map are shown in figure 6.1. We see that receptive field refinement and the appearance of topography and the general properties of the stripes occur quite quickly, followed by a more gradual process of further refinement.

6.3.5 Divisive normalization

The effect of using divisive rather than subtractive normalization is shown in figure 6.4, for the parameters in table 6.1. It can be seen that receptive fields fail to refine and cortical units remain primarily binocular. This behaviour will be analysed later.

We now examine the effect of changing some of the parameters of the model.

6.3.6 Width of interaction in the cortex

We examined the effect on stripe width of varying the width of cortical interactions. Results are compared in figure 6.5 for a progression of values of σ_c, the width of the gaussian function s in equation 6.1. Also shown in

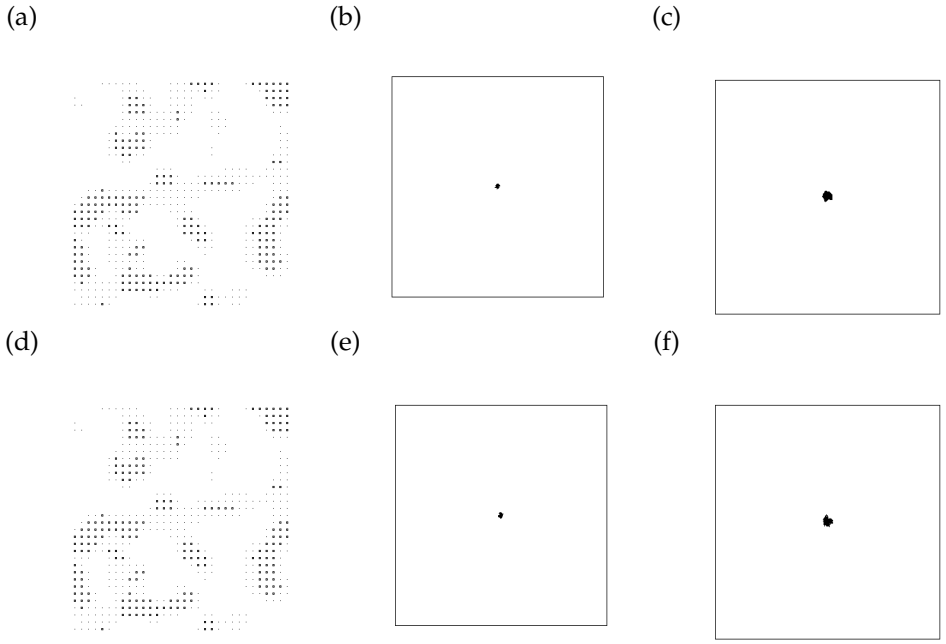

Figure 6.4: Results using a divisive normalization rule: parameters as in table 6.1. (a) Ocularity of cortical units. (b) Retinal topography for left eye. (c) Cortical topography for left eye. (d) Cortical topography averaged over both eyes. (e) Retinal topography for right eye. (f) Cortical topography for right eye.

σ_c	L_N	L_C	Total
1.0	5281.7	958.4	6240.1
2.0	4787.4	913.3	5700.7
3.0	4639.0	1064.8	5703.8

Table 6.2: Length of cortical wiring required for various σ_c.

figure 6.5 are the cortical topographies averaged over both eyes, the fourier transforms of the stripe pictures, and the power spectra calculated as described in chapter 5. It can be seen that stripe width increases with the width of the cortical interactions. The retinal and cortical topographies for these simulations are shown in figure 6.6. The parameter value $h = 0.05$ (i.e. a slight amount of correlation between the eyes) was used in these cases.

We also calculated the total length of wire L needed to connect all cortical units representing neighbouring positions within an eye (L_N) or corresponding positions in the two eyes (L_C) as described in chapter 5. Here the cortical unit representing a given retinal unit was taken to be that with the largest weight from that unit. Results are shown in table 6.2. It can be seen that the neighbourhood term tends to decrease as the stripes get wider.

6.3.7 Cross-eye correlations

As we have already argued, it seems highly likely that for animals such as cats and monkeys there exist strong positive correlations between the two eyes. We investigated the behaviour of the competitive model when the two eyes are correlated to varying degrees. Cross-eye correlations were produced in the following way. Once each retina had been convolved individually with a gaussian function to produce local correlations within each eye, the activity a_j of each unit j in each retina was replaced with $ha_j + (1-h)a'_j$, where a'_j is the activity of the corresponding unit to j in the other eye, and h specifies the degree of correlation between the two eyes. Thus by varying h we can vary the degree of correlation between the eyes: if $h = 0$ they are uncorrelated, and if $h = 0.5$ they are perfectly correlated (i.e. the pattern of activity is identical in the two eyes). Results for a progression of values of h are compared in figure 6.7, with the retinal and cortical topographies for each case in figure 6.8. The following effects can be seen as h increases:

- Cortical units take longer to saturate at complete dominance for one or the other eye.

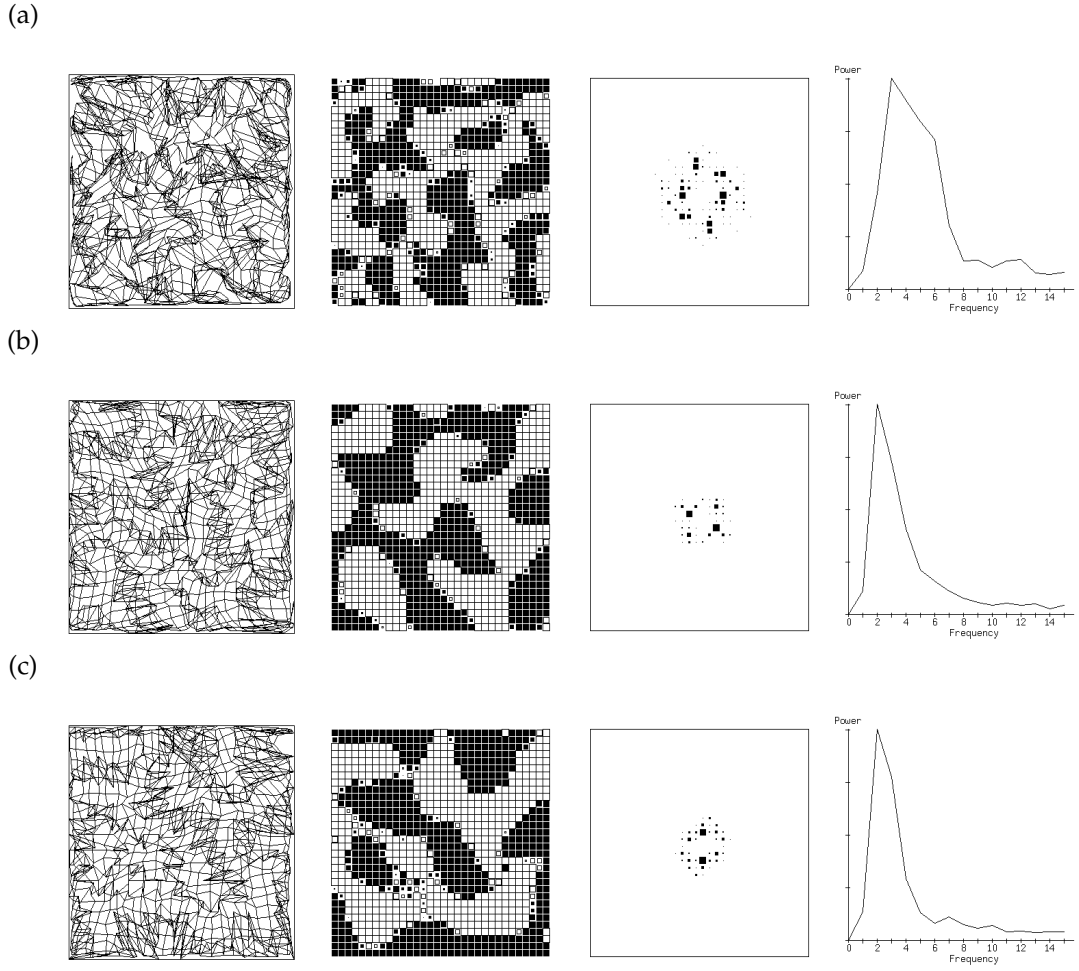

Figure 6.5: Effect of varying the width of cortical interactions. We show the cortical topography averaged over both eyes, the stripe pattern, the fourier transform of the stripes, and the power spectrum of the fourier transform for each case. (a) $\sigma_c = 1.0$. (b) $\sigma_c = 2.0$. (c) $\sigma_c = 3.0$. Note that stripe width tends to increase with σ_c.

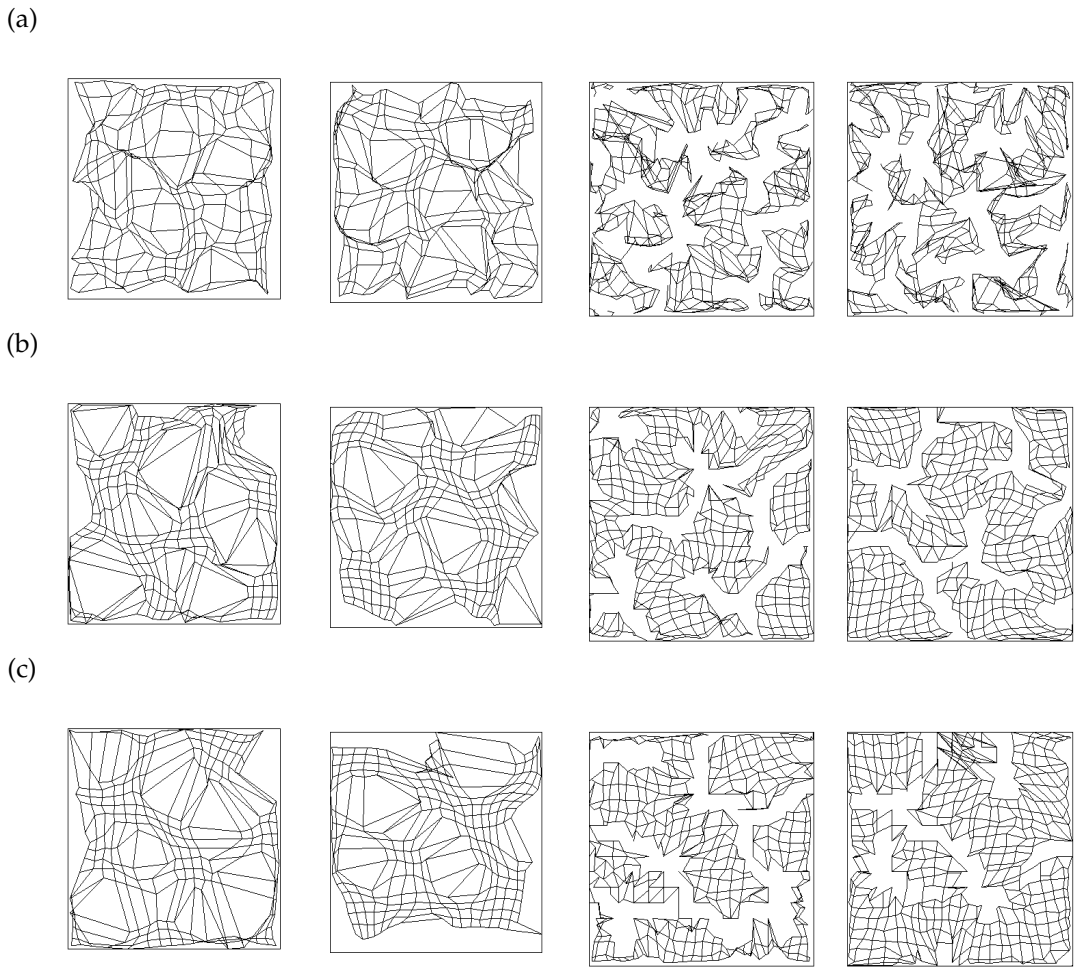

Figure 6.6: Retinal and cortical topography for varying extent of cortical interactions. For each σ we show the retinal topographies for the left and right eyes, followed by the cortical topographies for the left and right eyes respectively. (a) $\sigma = 1.0$. (b) $\sigma = 2.0$. (c) $\sigma = 3.0$.

h	L_N	L_C	Total
-0.1	5296.2	1529.0	6825.2
-0.05	5164.4	1258.4	6422.8
0.0	5108.2	973.3	6081.6
0.05	5002.5	1048.2	6050.7
0.10	4857.4	809.0	5666.4
0.15	4654.7	662.0	5316.7
0.20	4622.4	579.5	5201.8

Table 6.3: Length of cortical wiring required for various values of h.

- The topography of the map becomes smoother.
- Stripe width *decreases*.

Improvement of the topography is expected: as the two eyes become more correlated, more information becomes available to cortical units regarding the relative positions of units in the two retinae, and thus the two maps are aligned to a greater extent. From intuitive arguments similar to those made in chapter 5 we also expect the observed reduction in stripe width: greater correlation between the two eyes increases the desire to keep cortical units representing corresponding positions in the two eyes together.

The effects of zero and negative correlations between the eyes were also studied. These results are compared in figure 6.9, and the retinal and cortical topographies in figure 6.10. The degree of *negative* correlation appears to have little effect on stripe width.

We calculated the lengths of cortical wire required for results of each value of h, as shown in table 6.3. It can be seen that the correspondence term L_C decreases very rapidly as the correlation between the two eyes increases: this reflects the fact that the maps from the two eyes become increasingly accurately aligned as the correlations increase (see figures).

It is difficult to directly compare these wire lengths with those calculated for the results of the elastic net and Kohonen-type algorithms in chapter 5, due to varying stripe widths. However, we can see that for roughly comparable stripe widths, the present model gives wire lengths of approximately the same magnitude as those found in chapter 5.

6.4 Further characterization of the model

6.4.1 Dead units, stability, and saturation

The basic mechanism of the model is that of competitive learning [Rumelhart & Zipser 1986]. The model thus shares the same problem of "dead" units: that is, units which do not capture any patterns and hence whose weights do not progress beyond their initial values. Various mechanisms have been proposed to avoid this problem (for review see [Hertz et al 1991]), and in the present model we adopt a form of "conscience" mechanism [Hertz et al 1991]. Here the activity of each cortical unit in response to a pattern is divided by the number of times it has won the competition so far. This serves to roughly equalize the number of times each cortical unit wins the competition, and thus ensures that dead units do not occur. This could be implemented biologically by a mechanism whereby a cell adjusts its threshold or gain so as to keep its average activity roughly constant. Stability was achieved in the model by introducing a mechanism whereby, once weights reach their maximum or minimum values, they are "frozen" and not allowed to change further (similarly to [Miller et al 1989]). Thus once all connection strengths have saturated the complete map is frozen.

6.4.2 Afferent normalization

Afferent normalization refers to a constraint on each retinal unit to have a fixed amount of weight. A normalization rule of this type was discussed in [Prestige & Willshaw 1975], and used in the Tea Trade model [von der Malsburg & Willshaw 1977, Willshaw & von der Malsburg 1979] and Cowan's model [Whitelaw & Cowan 1981] among others. A rough biological motivation for such a rule is the idea that afferent fibres are supported by a flow of nutrients, and each presynaptic cell generates a fixed quantity of these, which must be divided

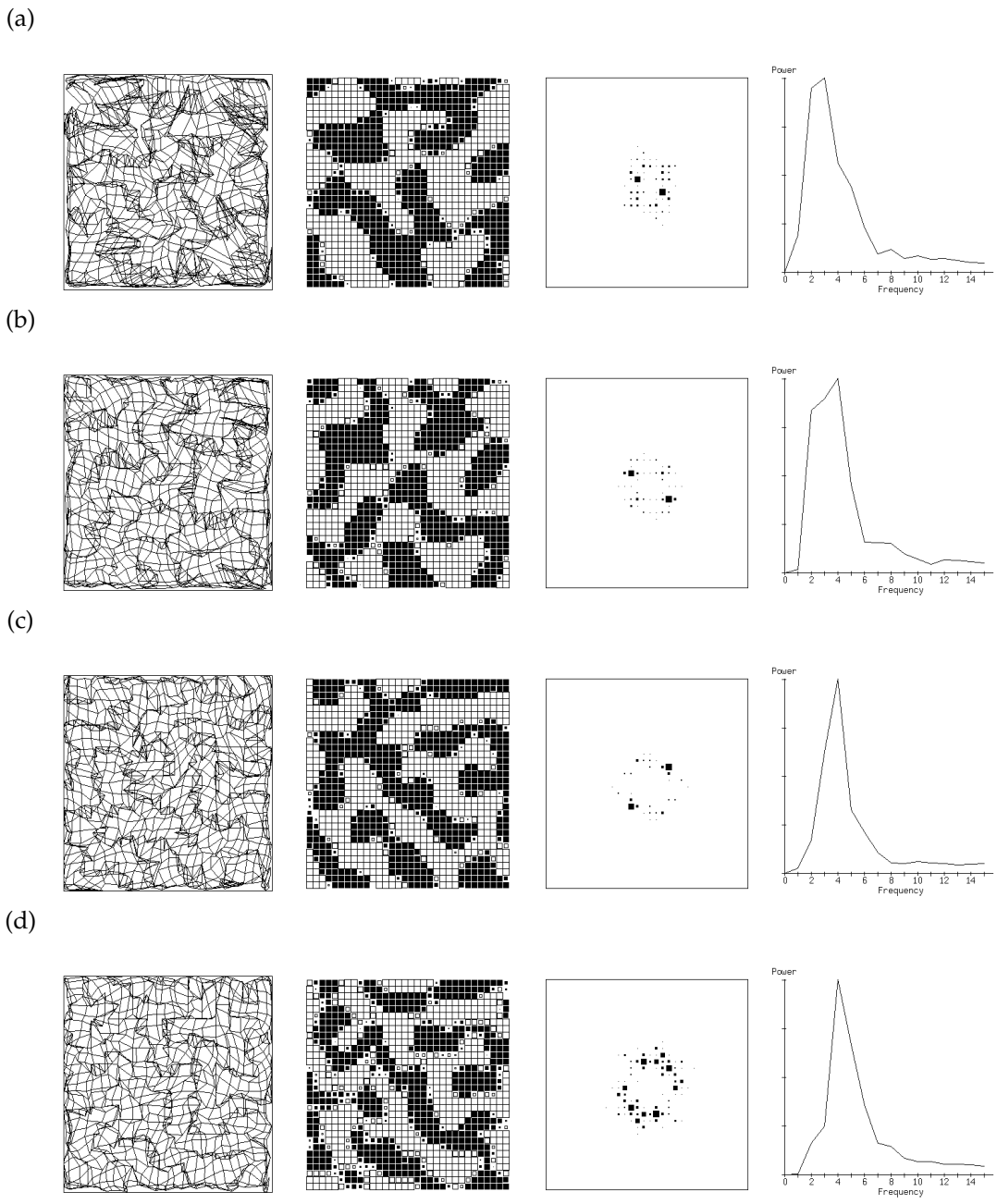

Figure 6.7: Effect of varying the extent of correlations between the eyes. We show the cortical topography averaged over both eyes, the stripe pattern, the fourier transform of the stripes, and the power spectrum of the fourier transform for each case. (a) $h = 0.05$, 250,000 iterations. (b) $h = 0.10$, 280,000 iterations. (c) $h = 0.15$, 350,000 iterations. (d) $h = 0.20$, 440,000 iterations. Note that stripe width tends to decrease as h increases, and that more iterations are required for convergence with increasing h.

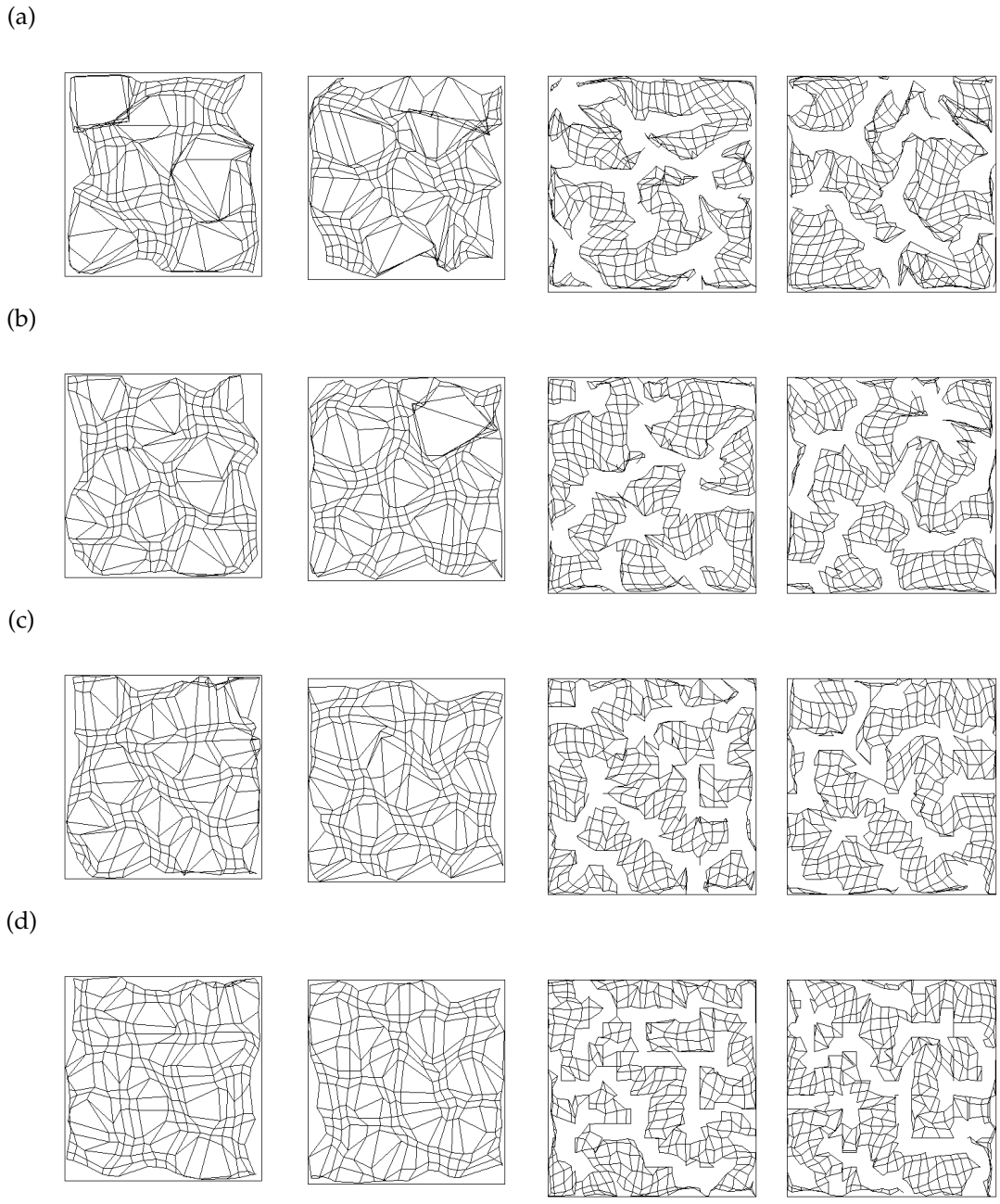

Figure 6.8: Retinal and cortical topography for varying extent of degree of correlations between the two eyes. For each h we show the retinal topographies for the left and right eyes, followed by the cortical topographies for the left and right eyes respectively. (a) $h = 0.05$. (b) $h = 0.10$. (c) $h = 0.15$. (d) $h = 0.20$.

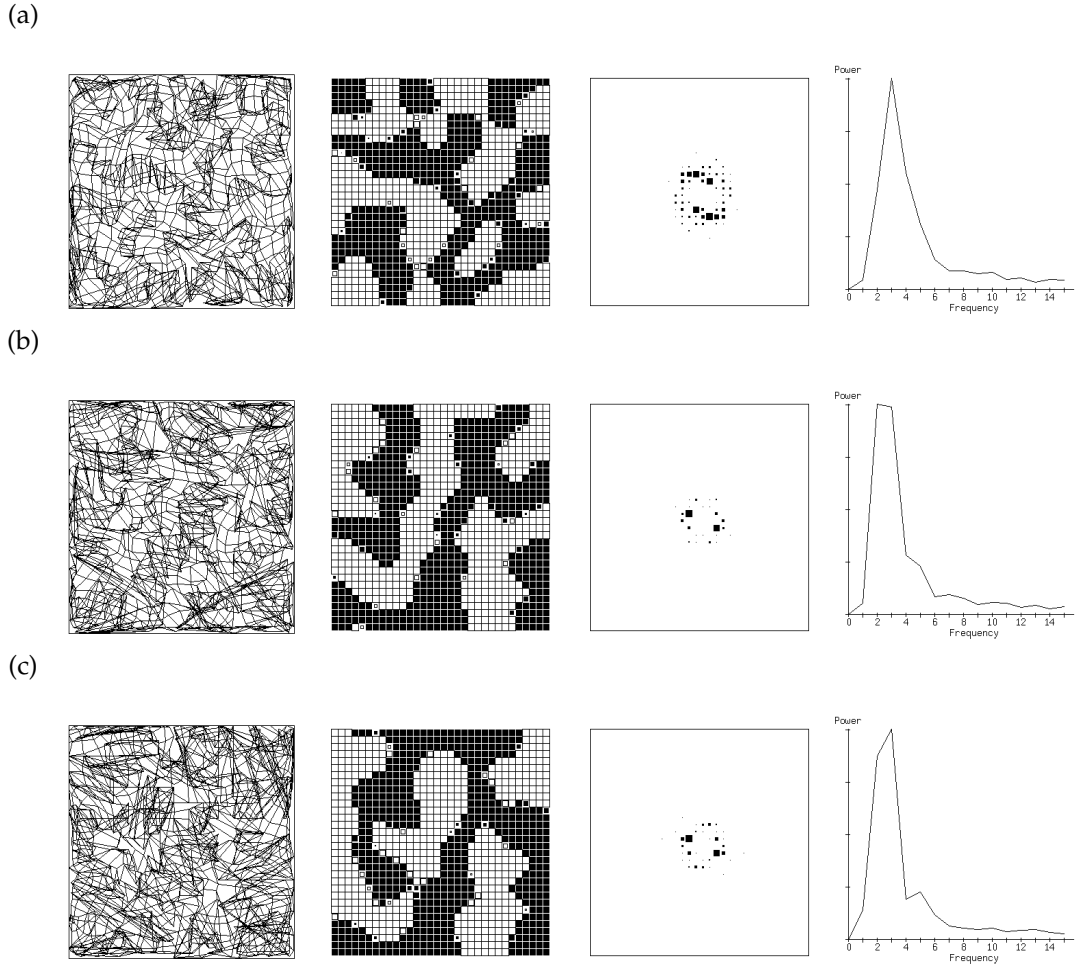

Figure 6.9: Effect of zero and negative correlations between the eyes. We show the cortical topography averaged over both eyes, the stripe pattern, the fourier transform of the stripes, and the power spectrum of the fourier transform for each case, after 250,000 iterations. (a) $h = 0.0$. (b) $h = -0.05$. (c) $h = -0.10$.

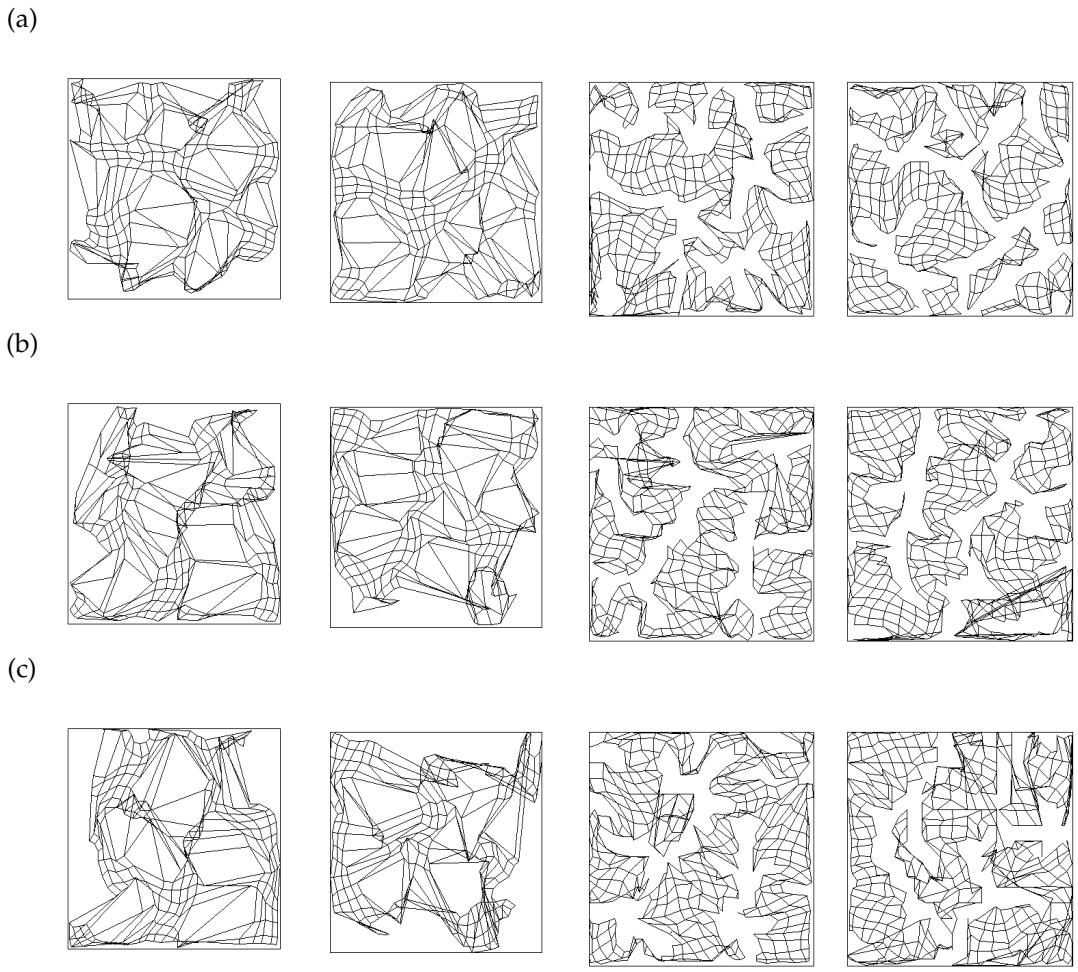

Figure 6.10: Retinal and cortical topography for varying extent of degree of negative correlation between the two eyes. For each h we show the retinal topographies for the left and right eyes, followed by the cortical topographies for the left and right eyes respectively. (a) $h = 0.0$. (b) $h = -0.05$. (c) $h = -0.10$.

between the fibres emerging from that cell. A linear afferent normalization constraint, divisively imposed, was employed in the model, in addition to the aforementioned efferent normalization. A similar mechanism was used in Miller's model [Miller et al 1989, footnote 23], where it was found that this had the effect of constraining stripe width to be less than an arbor radius (i.e. stripe period less than an arbor diameter), as discussed in chapter 4. Without such afferent normalization, it is possible for all the weights for a retinal unit to go to zero.

However, it is generally not possible to satisfy both afferent and efferent constraints simultaneously. Thus in models employing both types it is common to apply one and then the other, ignoring the fact that the first will now only be approximately satisfied.[1] In this model we apply (divisive) afferent normalization after the efferent normalization step. Efferent normalization is thus only approximate.

This introduces a new parameter for the model: the normalization constant for the afferent constraint, or more particularly the ratio of the normalization constants N_r and N_c for afferent and efferent constraints respectively. For all the results presented earlier $N_c = 10$ and $N_r = 20$. The ratio was motivated by the desire that the two constraints should "balance", in the sense that the one latterly applied should not cause a substantial rescaling of all weight values. Since in the simulations presented here there were twice as many postsynaptic cells as presynaptic cells, a ratio of $N_r = 2N_c$ was used. A brief investigation of the effect of varying N_r suggested that this ratio did not greatly affect simulation results.

6.5 Analysis of the efferent normalization mechanism

6.5.1 Introduction

As we have seen empirically, the type of normalization employed can make important differences to the outcome of a competitive model. We now analyse the effects of different types of normalization for a simple case. The following issues are addressed:

- Whether it is the linear sum or the sum-of-squares of the weights that is normalized.
- Whether this normalization is enforced divisively or subtractively.
- In the case of subtractive enforcement, the direction of the constraint vector.

Competitive systems such as the present model do not in general yield to the same kind of analysis as can be used for correlational models (see chapter 4). Even though each individual unit evolves according to a simple rule, the statistics of the set of inputs with which it is being presented is constantly changing as units specialise. Each unit sees only a subset of the patterns, and that subset is a function of the behaviour of all the other units. The form of the normalization changes the nature of the competition, and hence the analysis of [Miller & MacKay unpublished] is not directly applicable.

We therefore consider an extremely simple competitive situation that is mathematically tractable, with the hope that a thorough analysis of this will provide at least hints as to the situation pertaining in reality in the model. We look at the case of two weight vectors evolving in the positive quadrant of a two-dimensional space under the influence of normalized input vectors uniformly distributed in direction. We consider a somewhat simpler competitive rule than that of the model for the evolution of weights w_i in response to input $\mathbf{x}$:

$$\dot{w}_{winner} = \epsilon \mathbf{x} \tag{6.2}$$

where the winner is the unit with the largest inner product with the input pattern, and is the only unit that is updated for each input pattern. The competition is "hard": we do not consider any Kohonen-type interactions between units, such as those that exist in the model. It is shown that divisive enforcement leads to weight vectors becoming evenly distributed through the space, while subtractive enforcement leads to weight vectors tending to the axes of the space. The practical question of the speed of convergence is also addressed, and we show that there are differences between the two types of enforcement.

The analysis proceeds in the following stages.

- Calculate the size of the step taken by the winning unit in response to an input pattern.
- Calculate the average rate of change of the direction of the weight vectors, by averaging over all patterns for which that unit wins.

[1] An alternative, used in [Häussler & von der Malsburg 1983] is to impose a mathematical constraint that satisfies both simultaneously, but does not have a direct interpretation in terms of normalization.

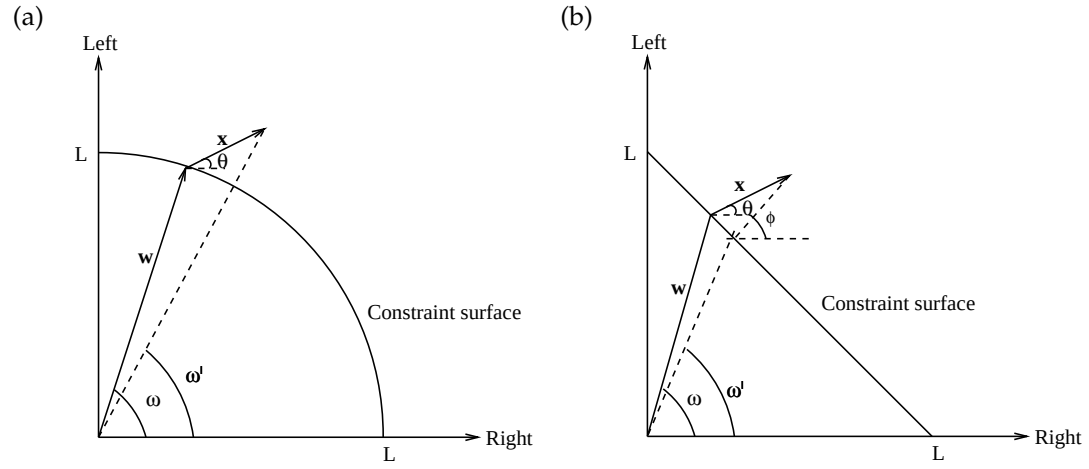

Figure 6.11: Geometry of an evolving weight for (a) squared constraints, divisive enforcement, and (b) linear constraints, subtractive enforcement. The squared subtractive and linear divisive cases are similar.

- Find the stable states.
- Calculate the convergence time.

6.5.2 Step sizes

We derive the step sizes by considering the geometric effect of updating weights and then applying the various normalization constraints. The geometry of an evolving weight for the squared divisive and linear subtractive cases is shown in figure 6.11 (see table 6.4 for notation). For convenience we refer to the axes as "left" and "right". Note that it is assumed throughout that $\omega_1, \omega_2, \phi \in [0, \frac{\pi}{2}]$, and only solutions to the equations in this range are considered. Sometimes for plotting purposes we consider the special case $\phi = \frac{\pi}{4}$, which is the case of relevance to the model. The diagrams show the effect of a weight being updated according to equation 6.2. A small fraction of x is added to w, and then the constraint is imposed by projecting back to the constraint surface, thus defining the new weight with angle ω'. In the case of divisive enforcement this regression takes place along a line towards the origin, while for subtractive enforcement the regression is along a line at angle ϕ with the right axis.

To calculate ω' in terms of ω, θ, ϕ, L and d we assume a small learning rate ϵ so that $\frac{\epsilon d}{L}$ is small to first order. We now derive expressions for the step in the squared constraint, divisive enforcement case shown in figure 6.11(a). The other cases follow similarly. Since $\frac{\epsilon d}{L}$ is small, we have that the distance from the tip of w to the tip of the new w is approximately equal to $\epsilon d \sin(\omega' - \theta)$. However, this distance is also (for $\omega - \omega'$ small) given by $L(\omega - \omega')$. Equating these two expressions gives

$$\omega' = \omega - \frac{\epsilon d}{L} \sin(\omega - \theta)$$

where we have used the fact that $\sin(\omega' - \theta) - \sin(\omega - \theta)$ is small to first order, and thus $\frac{\epsilon d}{L}\sin(\omega' - \theta) = \frac{\epsilon d}{L}\sin(\omega - \theta)$ to first order. Results for all cases are summarized in table 6.5.

An important difference between divisive and subtractive enforcement is immediately apparent: for divisive enforcement the sign of the change is dependent on $\text{sign}(\omega - \theta)$, while for subtractive enforcement it is dependent on $\text{sign}(\phi - \theta)$ (note that $\cos(\phi - \omega)$, $\cos(\phi - \frac{\pi}{4})$ and $\cos(\omega - \frac{\pi}{4})$ are always positive for $\omega, \phi \in [0, \frac{\pi}{2}]$). Thus in the divisive case a weight vector only moves towards the right axis (say) if the input pattern is more inclined to the right axis than the weight is already, whereas in the subtractive case the vector moves towards the right axis whenever the input pattern is inclined further to the right axis than the constraint vector. Graphs of the step sizes as a function of ω and θ are shown in figure 6.12, for the special case $\phi = \frac{\pi}{4}$.

6.5.3 Rates of change

We now consider two competing weight vectors w_1 and w_2 with angles ω_1 and ω_2 as shown for the squared case in figure 6.13. The problem is to calculate the motion of each weight vector in response to the input patterns for which it wins, taking account of the fact that this set changes with time. We do this by assuming

Name	Parameter
w	Weight vector
x	Input vector
ω	Angle of old weight vector to right axis
ω'	Angle of new weight vector to right axis
θ	Angle of input pattern vector to right axis
ϕ	Direction of subtractive constraint vector
L	Magnitude of the normalization constraint
d	$\|\|x\|\|$ (constant)
ϵ	Learning rate

Table 6.4: Parameters for calculation of step sizes

Constraint	Enforcement	ω'
Squared	Divisive	$\omega - \frac{\epsilon d}{L}\sin(\omega - \theta)$
Linear	Divisive	$\omega - \frac{\epsilon d\sqrt{2}}{L}\cos(\omega - \frac{\pi}{4})\sin(\omega - \theta)$
Squared	Subtractive	$\omega - \frac{\epsilon d}{L}\mathrm{cosec}(\phi - \omega)\sin(\phi - \theta)$
Linear	Subtractive	$\omega - \frac{\epsilon d\sqrt{2}}{L}\mathrm{cosec}(\phi - \frac{\pi}{4})\cos^2(\omega - \frac{\pi}{4})\sin(\phi - \theta)$

Table 6.5: Value of ω' for winning unit.

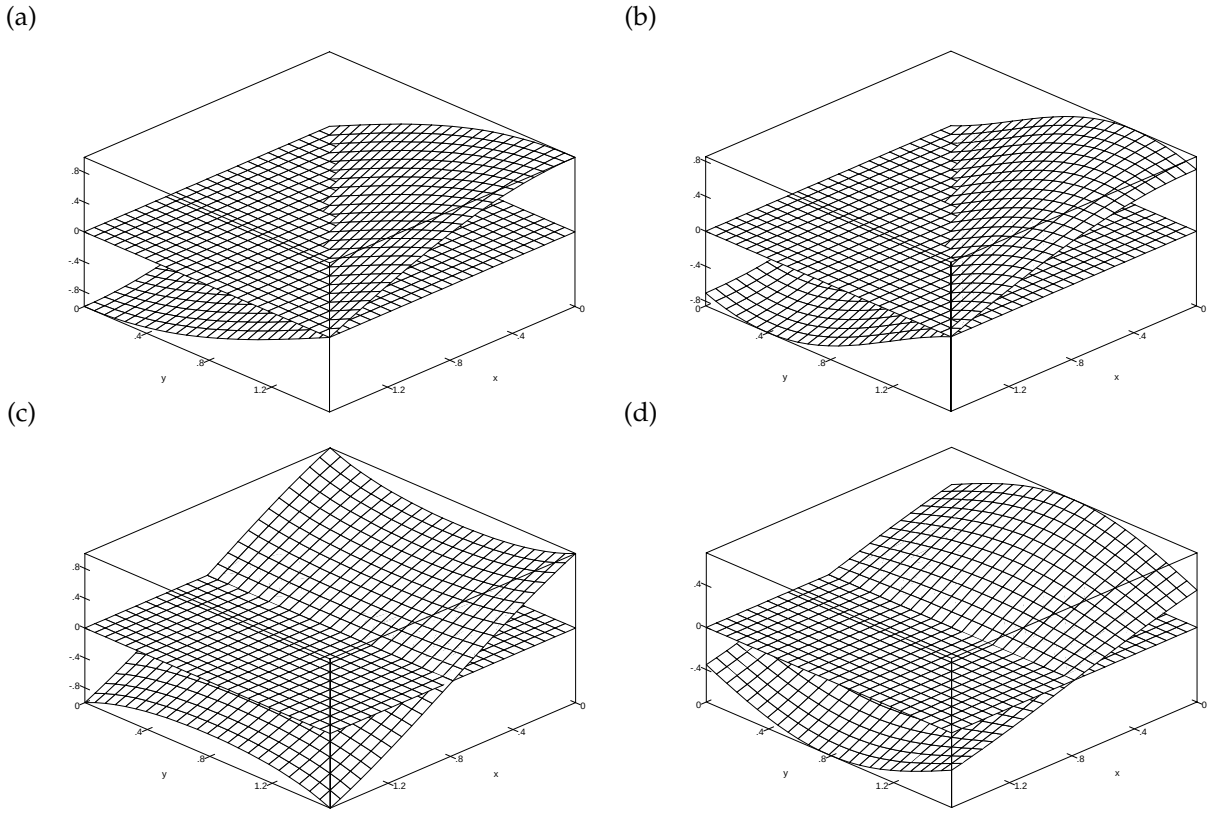

Figure 6.12: $-\dot{\omega}$ plotted as a function of θ (x axis) and ω (y axis) for each case. For the subtractive cases, ϕ was taken to be $\frac{\pi}{4}$. To make the sign of $\dot{\omega}$ clear the plane $\dot{\omega} = 0$ is also plotted. Vertical scale is arbitrary. (a) Squared constraints, divisive enforcement. (b) Linear constraints, divisive enforcement. (c) Squared constraints, subtractive enforcement. (d) Linear constraints, subtractive enforcement. Note that for the divisive cases $\dot{\omega}$ changes sign on the line $\omega = \theta$, while for the subtractive cases $\dot{\omega}$ changes sign on the line $\theta = \frac{\pi}{4}$.

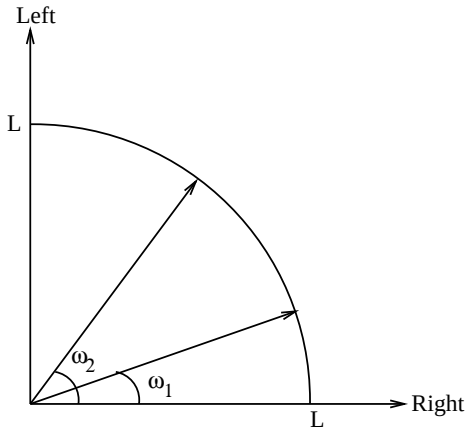

Figure 6.13: The geometry of two competing weight vectors in the squared divisive case.

that the learning rate ϵ of the weight vectors is small enough so that they move infinitesimally in the time it takes to present all the input patterns. Pattern order is then not important, and we can average over the entire set of inputs in calculating the rates of change. Note that we assume $\omega_1 < \omega_2$ throughout: this is simply a matter of the labelling of the weights.

Squared constraint, divisive enforcement case

Consider the evolution of ω_1. We have

$$\dot{\omega}_1 = -\frac{\epsilon d}{L}\sin(\omega_1 - \theta)$$

Using the assumption that ϵ is small, we now average over all the patterns for which w_1 wins the competition. In two dimensions this is straightforward: w_1 wins for all $\theta < \frac{\omega_1+\omega_2}{2}$.

Thus for this case we have

$$\langle\dot{\omega}_1\rangle_{0<\theta<\frac{\omega_1+\omega_2}{2}} = -\frac{\epsilon d}{L}\int_0^{\frac{\omega_1+\omega_2}{2}} \sin(\omega_1 - \theta)P(\theta)d\theta$$

where the angle brackets denote averaging over the specified range of θ, and $P(\theta)$ is the probability of input θ. We assume this is uniform: we take $P(\theta) = p$, a constant. It follows that

$$\langle\dot{\omega}_1\rangle = -\frac{\epsilon dp}{L}[\cos(\frac{\omega_1 - \omega_2}{2}) - \cos(\omega_1)] \tag{6.3}$$

We similarly obtain

$$\langle\dot{\omega}_2\rangle = -\frac{\epsilon dp}{L}[\sin(\omega_2) - \cos(\frac{\omega_2 - \omega_1}{2})], \tag{6.4}$$

where the average has been calculated over $\frac{\pi}{2} > \theta > \frac{\omega_1+\omega_2}{2}$. To determine the behaviour of the system we examine under what conditions $\langle\dot{\omega}_1\rangle$ and $\langle\dot{\omega}_2\rangle$ are positive, negative, and zero. For $\langle\dot{\omega}_1\rangle < 0$, i.e. w_1 tending towards the right axis, we require

$$\cos(\frac{\omega_1 - \omega_2}{2}) > \cos(\omega_1)$$

which (for $\omega_1, \omega_2 \in [0, \frac{\pi}{2}]$ and $\omega_2 > \omega_1$ as already assumed) implies

$$\frac{\omega_2 - \omega_1}{2} < \omega_1$$

That is

$$\omega_1 > \frac{\omega_2}{3}$$

Thus w_1 only tends towards the right axis when w_2 is "close enough" to w_1. By symmetry we have the following condition for which $\langle\dot{\omega}_2\rangle > 0$:

$$\omega_2 < \frac{\omega_1 + \pi}{3}$$

and thus the stable state is

$$\omega_1 = \frac{\pi}{8}\ ,\ \omega_2 = \frac{3\pi}{8}$$

Each weight captures half the patterns, and comes to rest balanced by inputs on either side of it. Weights do not saturate at the axes.

Linear constraint, divisive enforcement case

For the linear constraint cases the weight vectors have variable length, and thus determining which unit wins the competition (in the sense of having the largest inner product with the input pattern) is slightly more complicated. In particular, the condition for w_1 to win for input θ is

$$||w_1||\cos(\theta - \omega_1) \ > \ ||w_2||\cos(\theta - \omega_2)$$

where

$$||w_i|| = \frac{L}{\sqrt{2}\sin(\omega_i + \frac{\pi}{4})} \ , \ i = 1, 2$$

This gives the condition

$$\theta < \frac{\pi}{4}$$

Thus in the linear cases it is the unit closest to the axis which the input is closest to that wins, and the weights evolve effectively independently of each other.

Similar reasoning for the linear divisive constraint to that above yields

$$\langle \dot{\omega}_1 \rangle = -\frac{\epsilon d p \sqrt{2}}{L}\cos(\omega_1 - \frac{\pi}{4})[\cos(\omega_1 - \frac{\pi}{4}) - \cos(\omega_1)]$$

and

$$\langle \dot{\omega}_2 \rangle = -\frac{\epsilon d p \sqrt{2}}{L}\cos(\omega_2 - \frac{\pi}{4})[\cos(\omega_2 - \frac{\pi}{2}) - \cos(\omega_2 - \frac{\pi}{4})]$$

It follows that the stable state is the same as in the squared constraint, divisive enforcement case:

$$\omega_1 = \frac{\pi}{8} \ , \ \omega_2 = \frac{3\pi}{8},$$

and again the weights do not saturate at the axes.

Linear constraint, subtractive enforcement case

We now have

$$\langle \dot{\omega}_1 \rangle_{0<\theta<\frac{\pi}{4}} = -\frac{\epsilon d p \sqrt{2}}{L}\frac{\sin^2(\omega_1 + \frac{\pi}{4})}{\sin(\phi + \frac{\pi}{4})}\int_0^{\frac{\pi}{4}} \sin(\phi - \theta) d\theta$$

So

$$\langle \dot{\omega}_1 \rangle = -\frac{\epsilon d p \sqrt{2}}{L}\frac{\sin^2(\omega_1 + \frac{\pi}{4})}{\sin(\phi + \frac{\pi}{4})}[\cos(\phi - \frac{\pi}{4}) - \cos(\phi)] \tag{6.5}$$

We similarly obtain

$$\langle \dot{\omega}_2 \rangle = -\frac{\epsilon d p \sqrt{2}}{L}\frac{\sin^2(\omega_1 + \frac{\pi}{4})}{\sin(\phi + \frac{\pi}{4})}[\cos(\phi - \frac{\pi}{2}) - \cos(\phi - \frac{\pi}{4})] \tag{6.6}$$

Now consider the circumstances under which $\langle \dot{\omega}_1 \rangle < 0$, that is w_1 heading for the right axis. This depends only on ϕ:

$$\cos(\phi - \frac{\pi}{4}) > \cos(\phi)$$

That is

$$\phi > \frac{\pi}{8}$$

Similarly for w_2 to be heading for the left axis we have

$$\phi < \frac{3\pi}{8}$$

Thus in this case the weights saturate, one at each axis, if $\frac{\pi}{8} < \phi < \frac{3\pi}{8}$. They both saturate at the right axis for $\phi < \frac{\pi}{8}$, and both at the left axis for $\phi > \frac{3\pi}{8}$.

(a) (b)

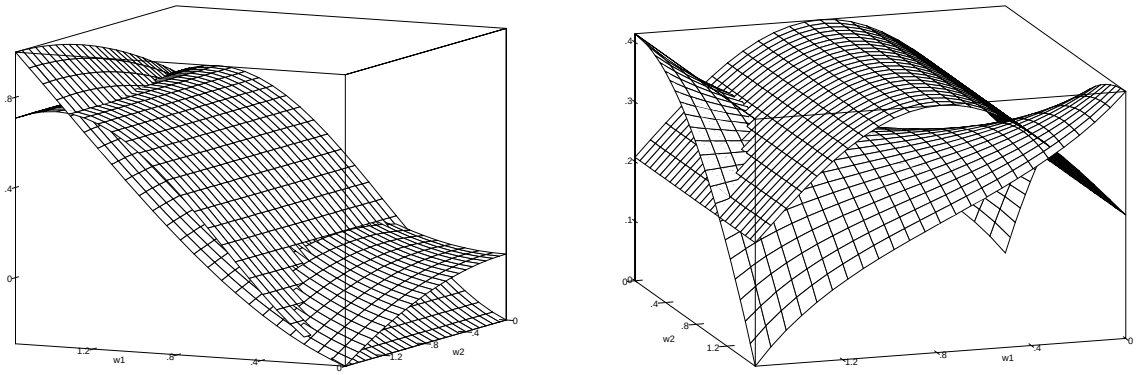

Figure 6.14: $\langle \dot{\omega}_1 \rangle$ plotted as a function of ω_1 and ω_2. Vertical scale is arbitrary. (a) Divisive enforcement cases. (b) Subtractive enforcement cases for $\phi = \frac{\pi}{4}$. For the linear constraint cases, $\langle \dot{\omega}_1 \rangle$ depends only on ω_1. Note that there can be a large difference in the size of the steps for linear and squared constraints in each case.

Squared constraint, subtractive enforcement case

In the squared case, by similar reasoning we obtain

$$\langle \dot{\omega}_1 \rangle = -\frac{\epsilon\, dp}{L} \frac{1}{\cos(\phi - \omega_1)} [\cos(\phi - \frac{\omega_1 + \omega_2}{2}) - \cos(\phi)] \tag{6.7}$$

and

$$\langle \dot{\omega}_2 \rangle = -\frac{\epsilon\, dp}{L} \frac{1}{\cos(\phi - \omega_2)} [\cos(\phi - \frac{\pi}{2}) - \cos(\phi - \frac{\omega_1 + \omega_2}{2})] \tag{6.8}$$

For $\langle \dot{\omega}_1 \rangle < 0$ we require

$$\cos(\phi - \frac{\omega_1 + \omega_2}{2}) > \cos(\phi)$$

which is satisfied for

$$\phi > \frac{\omega_1 + \omega_2}{4}$$

Similarly for $\langle \dot{\omega}_2 \rangle > 0$ we require

$$\phi < \frac{\omega_1 + \omega_2}{4} + \frac{\pi}{4}$$

That is, weights saturate one at each axis for $\frac{\omega_1+\omega_2}{4} < \phi < \frac{\omega_1+\omega_2}{4} + \frac{\pi}{4}$.

Thus in both subtractive enforcement cases there is a "window" of size $\frac{\pi}{4}$ within which ϕ must lie in order for weights to saturate at different axes: this window is in a slightly different place for the linear and squared constraint cases. In particular in the squared constraint case the suitable range of ϕ changes as the weights evolve. This case reduces to the linear constraint case when $\omega_1 + \omega_2 = \frac{\pi}{2}$, ie when the weights are symmetrically balanced about $\frac{\pi}{4}$. Thus if for instance w_1 is much closer to saturating than w_2, ϕ is required to be slightly smaller to compensate: e.g. in the case $\omega_1 \simeq 0$, $\omega_2 = \frac{\pi}{4}$ we require $\frac{\pi}{16} < \phi < \frac{5\pi}{16}$. Note however that the choice $\phi = \frac{\pi}{4}$ ensures that each weight vector is always moving towards the appropriate axis (except in the special case that both weights are already saturated at the same axis). For ϕ outside the range above, both weights saturate at the same axis as in the linear constraint case.

Formulae for $\langle \dot{\omega}_1 \rangle$ obtained above are summarized in table 6.6. The results are compared in figure 6.14 for the special case $\phi = \frac{\pi}{4}$. Convergence properties for each of the 4 cases of constraints and enforcement are summarized in table 6.7.

6.5.4 Convergence times

From a practical point of view, it is of interest to compare the time taken for each of these cases to reach a stable state. The convergence time for each case can be calculated by integrating the equations for the evolution of the weights from the initial state to the converged state. Refer to the initial state as (ω_1^0, ω_2^0). We note immediately that for the divisive cases, since $\langle \dot{\omega}_1 \rangle$ goes to zero as the system converges, we expect these times to be infinite.

Constraint	Enforcement	$\langle \dot{\omega}_1 \rangle$
Squared	Divisive	$-\frac{\epsilon d p}{L}[\cos(\frac{\omega_1-\omega_2}{2}) - \cos(\omega_1)]$
Linear	Divisive	$-\frac{\epsilon d p\sqrt{2}}{L}\cos(\omega_1 - \frac{\pi}{4})[\cos(\omega_1 - \frac{\pi}{4}) - \cos(\omega_1)]$
Squared	Subtractive	$-\frac{\epsilon d p}{L}\mathrm{cosec}(\phi - \omega_1)[\cos(\phi - \frac{\omega_1+\omega_2}{2}) - \cos(\phi)])$
Linear	Subtractive	$-\frac{\epsilon d p\sqrt{2}}{L}sec(\phi + \frac{\pi}{4})[\cos(\phi - \frac{\pi}{4}) - \cos(\phi)]\cos^2(\omega_1 - \frac{\pi}{4})$

Table 6.6: Formulae for $\langle \dot{\omega}_1 \rangle$.

Constraints	Divisive	Subtractive
Linear	Weights stable at $\omega_1 = \frac{\pi}{8}$, $\omega_2 = \frac{3\pi}{8}$	Weights saturate at different axes for $\frac{\pi}{8} < \phi < \frac{3\pi}{8}$
Squared	Weights stable at $\omega_1 = \frac{\pi}{8}$, $\omega_2 = \frac{3\pi}{8}$	Weights saturate at different axes for $\frac{\omega_1+\omega_2}{4} < \phi < \frac{\omega_1+\omega_2}{4} + \frac{\pi}{4}$

Table 6.7: Convergence properties of the four cases

Divisive enforcement cases

We have shown above that for the divisive enforcement cases the stable state is $\omega_1 = \frac{\pi}{8}, \omega_2 = \frac{3\pi}{8}$. Unfortunately for the squared constraint case the system of differential equations 6.3 and 6.4 cannot be solved in closed form. However, a straightforward solution is possible in the special case $\omega_1^0 + \omega_2^0 = \frac{\pi}{2}$. By symmetry, if this holds initially it must be true for all time, that is $\omega_1 + \omega_2 = \frac{\pi}{2}$ (we will use this special case again later). The equations now simplify so that $\langle \dot{\omega}_1 \rangle$ is a function only of ω_1, and similarly $\langle \dot{\omega}_2 \rangle$ is a function only of ω_2. We thus obtain T_{sd}^{sp}, the time to convergence of both w_1 and w_2 (by the assumed symmetry they are the same) in the squared constraint, divisive enforcement, special case, as

$$T_{sd}^{sp} = -\frac{L}{\epsilon d p}\int_{\omega_1^0}^{\frac{\pi}{8}} \frac{1}{\cos(\omega_1 - \frac{\pi}{4}) - \cos(\omega_1)} d\omega_1$$

For the linear case $\langle \dot{\omega}_1 \rangle$ and $\langle \dot{\omega}_2 \rangle$ evolve independently, and thus without making any assumptions about ω_1^0 and ω_2^0 we obtain

$$T_{ld}^{1} = -\frac{L}{\epsilon d p\sqrt{2}}\int_{\omega_1^0}^{\frac{\pi}{8}} \frac{1}{\cos(\omega_1 - \frac{\pi}{4})[\cos(\omega_1 - \frac{\pi}{4}) - \cos(\omega_1)]} d\omega_1$$

and similarly

$$T_{ld}^{2} = -\frac{L}{\epsilon d p\sqrt{2}}\int_{\omega_2^0}^{\frac{3\pi}{8}} \frac{1}{\cos(\omega_2 - \frac{\pi}{4})[\cos(\omega_2 - \frac{\pi}{2}) - \cos(\omega_2 - \frac{\pi}{4})]} d\omega_2$$

We note that all these integrands diverge at the upper limits. As pointed out above, this is expected since the convergence times are infinite. To compare the convergence times we therefore consider the time to reach a value very close to the limit. These results are shown in figure 6.15 for T_{sd}^{sp} and T_{ld}^{sp} (i.e. the special case $\omega_1^0 + \omega_2^0 = \frac{\pi}{2}$). It can be seen that squared constraints take longer to converge than linear constraints.

Subtractive enforcement cases

The stable state is now saturation of w_1 at the right axis and w_2 at the left axis (providing ϕ is chosen appropriately). Again the equations for the evolution of the weights in the squared constraint case (equations 6.7 and 6.8) are coupled. However, for the special case $\phi = \frac{\pi}{4}$ the equations can be separated since they are now of the form

$$\dot{\omega}_1 = \frac{f(\omega_1, \omega_2)}{g(\omega_1)} \ , \ \dot{\omega}_2 = -\frac{f(\omega_1, \omega_2)}{g(\omega_2)}$$

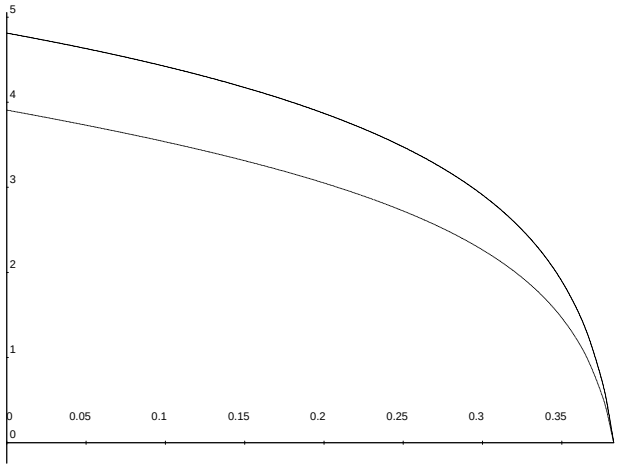

Figure 6.15: Time to convergence (arbitrary units) against initial ω_1 for divisive enforcement in the special case $\omega_1^0 + \omega_2^0 = \frac{\pi}{2}$, $\omega_1 < \frac{\pi}{8}$. Upper line is squared constraints, lower line is linear constraints.

This yields the following relationship between ω_1 and ω_2:

$$\sin(\phi - \omega_1) + \sin(\phi - \omega_2) = c(\omega_1^0, \omega_2^0)$$

where

$$c(\omega_1^0, \omega_2^0) = \sin(\phi - \omega_1^0) + \sin(\phi - \omega_2^0)$$

We thus obtain

$$T_{ss}^1 = -\frac{L}{\epsilon\, dp} \int_{\omega_1^0}^{0} \frac{\cos(\phi - \omega_1)}{[\cos(\frac{\phi}{2} - \frac{1}{2}(\omega_1 - \arcsin(c(\omega_1^0, \omega_2^0) + \sin(\omega_1 - \phi)))) - \cos(\phi)]} d\omega_1$$

and

$$T_{ss}^2 = \frac{L}{\epsilon\, dp} \int_{\omega_2^0}^{\frac{\pi}{2}} \frac{\cos(\phi - \omega_2)}{[\cos(\frac{\phi}{2} - \frac{1}{2}(\omega_2 - \arcsin(c(\omega_1^0, \omega_2^0) + \sin(\omega_2 - \phi)))) - \cos(\phi - \frac{\pi}{2})]} d\omega_2$$

These cannot be integrated in closed form. However, if we again consider the special case $\omega_1 + \omega_2 = \frac{\pi}{2}$, we then obtain

$$T_{ss}^{sp} = -\frac{L}{\epsilon\, dp} \frac{1}{(1 - \frac{1}{\sqrt{2}})} [\frac{1}{\sqrt{2}} - \sin(\frac{\pi}{4} - \omega_1^0)]$$

Moving now to the linear case, the integrals are straightforward:

$$T_{ls}^1 = -\frac{L}{\epsilon\, dp\sqrt{2}} \frac{\sin(\phi + \frac{\pi}{4})}{\cos(\phi - \frac{\pi}{4}) - \cos(\phi)} [1 - \cot(\omega_1^0 + \frac{\pi}{4})]$$

and

$$T_{ls}^2 = -\frac{L}{\epsilon\, dp\sqrt{2}} \frac{\sin(\phi + \frac{\pi}{4})}{\cos(\phi - \frac{\pi}{2}) - \cos(\phi - \frac{\pi}{4})} [1 + \cot(\omega_2^0 + \frac{\pi}{4})]$$

Note that these blow up for $\phi = \frac{\pi}{8}$, $\phi = \frac{3\pi}{8}$ respectively as expected.

We now compare the time to saturation for squared and linear constraints for the special case $\omega_1 + \omega_2 = \frac{\pi}{2}$, $\phi = \frac{\pi}{4}$ (which implies that $T_{ls}^1 = T_{ls}^2$). This is shown in figure 6.16.

It can be seen from these graphs that squared constraints generally saturate faster than linear constraints in this case.

6.5.5 Summary

In this section we have analysed a very simple case of competitive learning under various constraints and constraint enforcement, and have found the following:

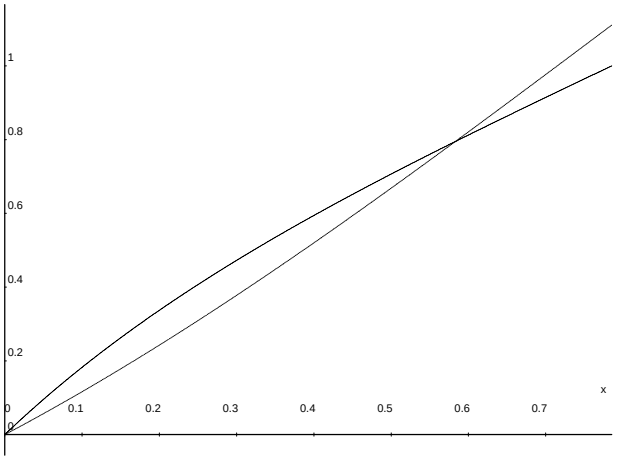

Figure 6.16: Time to saturation (arbitrary units) against initial ω_1 for subtractive constraints in the special case $\omega_1^0 + \omega_2^0 = \frac{\pi}{2}$, $\phi = \frac{\pi}{4}$. Squared constraints is the line that is higher for large x.

- The direction of movement of a weight vector is determined by (a) the side of the weight vector on which the input falls in the divisive enforcement cases, and (b) the side of ϕ on which the input falls in the subtractive enforcement cases.
- The stable state for both linear and squared constraints in the divisive enforcement case is $\omega_1 = \frac{\pi}{8}$, $\omega_2 = \frac{3\pi}{8}$. For subtractive enforcement, the stable state depends on ϕ and whether the constraints are linear or squared. In both subtractive cases ϕ can be chosen so that w_1 saturates at the right axis and w_2 at the left axis, or so that both weights saturate at the same axis.
- The convergence times are different in each case. For both types of enforcement (with $\phi = \frac{\pi}{4}$), the time taken for the weights to saturate can be compared for the special case $\omega_1^0 + \omega_2^0 = \frac{\pi}{2}$. It is found that the squared case converges slower for divisive constraints but faster for subtractive constraints.

The question now naturally arises of how applicable these results are to more weight vectors and higher dimensions.

6.5.6 Extension to more units

Extending the above analysis to the case of more than two units evolving in two-dimensional space is straightforward. Consider N units with weight vectors w_i, indexed according to their angle with the right axis, so that the smallest angle with the right axis is ω_1 and so on.

For the squared constraint, divisive enforcement case we have the stable states $\omega_1 = \frac{\omega_2}{3}$, $\omega_N = (\frac{\omega_{N-1}+\pi}{3})$. The weight vectors in between are stable when they are equidistant from their neighbours. The angle between each pair is thus $\alpha = \frac{\pi}{2N}$, and the angle between ω_1, ω_N and the right and left axes respectively is $\frac{\alpha}{2}$.

For the linear constraint, divisive enforcement case the situation is different since it is the weight vector closest to the axis that the input vector is closest to that wins. First consider the case where $\frac{\pi}{8} < \omega_1 < \frac{3\pi}{8}$. Then w_1 is the only vector that ever wins for $\theta < \frac{\pi}{4}$, and so is stable at $\omega_1 = \frac{\pi}{8}$ as before while all other vectors j such that $\omega_j < \frac{\pi}{4}$ remain in their initial positions. A similar situation holds in the upper octant. If $\omega_1 < \frac{\pi}{8}$ then it still eventually comes to rest at $\omega_1 = \frac{\pi}{8}$. However, if there are other vectors k such that $\omega_k < \frac{\pi}{8}$ then these will begin to win as ω_1 passes by on its way to $\frac{\pi}{8}$. Which unit wins changes as each progresses towards $\frac{\pi}{8}$, where they are all finally stable. Again, vectors with initial angles between $\frac{\pi}{8}$ and $\frac{3\pi}{8}$ remain in their initial states.

By similar arguments, the situation is even more straightforward in the linear constraint, subtractive enforcement case. w_1 saturates at the right axis and w_N at the left axis as before (for appropriate ϕ), and all other weights remain unchanged. The squared constraint, subtractive enforcement case is however more complicated. In general all weights vectors w_i for which $\omega_i^0 < \phi$ saturate at the right axis: an analogous results holds for the left axis. However, this is not quite true of the two initial weight vectors w_j and w_{j+1} which are such that $\omega_j < \phi$ and $\omega_{j+1} > \phi$. Assume w_{j+1} is closer to ϕ than w_j. Then w_{j+1} can win for inputs $\theta < \phi$, and thus w_{j+1} can eventually be pulled to the right axis.

We can thus clearly see the effect of a conscience mechanism that ensures that each unit wins roughly the same amount of the time. In the linear constraint, subtractive enforcement case, this would mean that eventually all

weights would saturate at the axis they were initially closest too. For instance, for $\theta < \frac{\pi}{4}$, w_1 would win for the first pattern, but then w_2 would win for the next since w_1 is temporarily out of the running, and similarly for all weights.

6.5.7 Higher dimensions

In higher dimensional spaces, the situation immediately becomes much more complicated, for two main reasons. Firstly, it becomes much harder to calculate the new weight vector in terms of the old for the winning unit by the geometric methods we have used here. Secondly, the dividing lines between which unit wins for the set of inputs is a Voronoi Tesselation [Hertz et al 1991] of the constraint surface. Thus the limits of the integral required to calculate $\langle \dot{\omega} \rangle$ are hard to determine, and the integrand is more complicated. We do not pursue this case further here.

6.5.8 The model

The model is somewhat different from the situation analysed here, both because of the much higher number of dimensions, and the fact there are interactions between units in the model. We have already discussed the problems of analysis in higher dimensions. For the case of interactions between units, as we have seen empirically (and was shown analytically for a simple case [Kohonen 1982(b)]), this in general causes neighbouring units to represent neighbouring inputs. Analysing this case using the techniques of this chapter is difficult since now the average change for unit i needs to be calculated over not just all patterns for which i wins, but in addition over all patterns for which any unit within the range of the postsynaptic updating function of i wins.

The simple analysis given in this chapter at least provides a hint about the behaviour of competitive systems with varying types of constraints. The conclusions regarding the stable state of each system appear also to apply empirically to the model. However, higher dimensional analysis is required to assess to what extent explanations in the two-dimensional case can be generalized.

We return to address directly the model.

6.6 Discussion

6.6.1 Relation to Kohonen-type algorithms

The model presented here employs a version of Kohonen's algorithm as a computationally efficient approximation to specifying with an explicit cortical interaction function how much the activity of each cortical cell is affected by its neighbours. However, an important difference is in the presentation of distributed rather than localized patterns of activity. This affects the degree to which Kohonen's algorithm really is a good approximation to explicit interactions, as we now explain.

Consider four neighbouring units in a 1-D cortex. Kohonen's algorithm assumes that the unit with the greatest feed-forward activation (say unit 1) still has the greatest activation after the effects of lateral interactions have been taken into account. However, consider the case where unit 4 has an activation almost as big as that of unit 1, unit 3 has an activation only slightly smaller than unit 4, and unit 2 has very low activity. Assume that the lateral connectivity is strongly excitatory but short range: just 1 cell wide. Then units 3 and 4 will mutually excite each other to a much greater extent than units 1 and 2, with the result that after settling unit 4 will have greater activity than unit 1, the unit that would originally have won the competition. This pertains until receptive fields have refined sufficiently so that there is a very big difference in the activity of the winning unit (or at least a small cluster of units in a local neighbourhood of the winner) and the rest.

That Kohonen's algorithm is only a rough approximation to full centre-surround interactions is well known. The important point here is that this situation is much more likely to arise using distributed input patterns than those that are localized. If all input units are active to roughly the same degree, postsynaptic units with similar (high) activations can occur in widely-distributed positions: receptive field refinement does not now guarantee large differentials in activity. However, in [von der Malsburg & Singer 1988] it is suggested that in the biological system differences of activity as low as 1% are sufficient to lead to complete inhibition of all but a few cells in the neighbourhood of the winner, after taking account of cortical interactions.

6.6.2 Obermayer et al's algorithm

Parallels can be drawn between the model presented here and that of [Obermayer et al 1990]. Firstly, the models are similar in that both use a sequence of retinal inputs to drive reorganization of a layer of weights to a cortical sheet via a Kohonen-type algorithm. The main difference is that we use a subtractive normalization rule, which allows us to adopt distributed patterns of activity in the retinae, and thus investigate the role of correlations within and between retinae. Ocular dominance segregation is driven by the form of the correlations between the two eyes, rather than systematic differences in strength between the two eyes as in Obermayer et al's model. In addition, Obermayer et al consider the formation of orientation stripes besides ocular dominance stripes, which we do not. They are therefore able to study the interaction between these two stripe systems in their model.

6.6.3 Miller's model

Miller did not employ the same kind of competition between cortical units as we have used in the present model: he included lateral inhibition explicitly. There is little pressure for Miller's cortical units to learn different things: however, since the topography of the map is fixed this is not so important. Each cortical unit learns the same eigenvector of the matrix of presynaptic correlations: when this is monocular some units to develop dominance for the left eye and the others dominance for the right eye, determined by the lateral inhibition in the cortex and the initial weights.

There is a non-trivial relationship between full centre-surround excitation and inhibition as used by Miller and the computationally efficient approximation used here. In particular, it is hard to say precisely how the widths compare, which makes it difficult to compare stripe width in the two models. However, it is clear that in both cases stripes get wider as the width of the cortical interaction function increases.

6.6.4 The elastic net

Like the elastic net, the present model predicts that stripe width decreases as the two eyes become more correlated. Both models involve competition between postsynaptic units (the relation between the elastic net and Kohonen-type algorithms has been discussed in chapter 5). Stripe morphology is different in the two models: elastic net stripes tend to be more regular with more global order than in the present model.

6.7 Conclusions

In this chapter we have presented a new model that accounts for the development of topography and ocular dominance by the same mechanism. It is based on a type of competitive learning, with a subtractive normalization rule to limit synaptic strengths, and a cortical interaction function simplified in a way introduced by Kohonen. The most important properties of the model are:

- Localized and ocularly dominant receptive fields form in the cortex when distributed patterns of activity are presented simultaneously in both retinae.
- Stripe width is a function of both the range of the cortical interaction function (as in Miller's model), and the degree of correlation between the two retinae (as in the elastic net).

Chapter 7

Conclusions

7.1 Introduction

We first summarize the main points made in the thesis. We then present a more detailed comparison of the models we have addressed, paying particular attention to the terms of reference set out in the introduction. Directions for future work are suggested.

7.2 Summary of the main points

- Motivations: Experimental data from the natural system suggests that topography and ocular dominance can arise from the same developmental mechanisms. Despite this, previous models for topography or ocular dominance have tended not to simultaneously address both issues in a satisfactory way.
- We therefore discussed several models of visual development from the point of view of (a) whether they could form topography and ocular dominance from the same mechanisms, (b) to what extent they could be interpreted biologically, (c) whether they could form appropriate maps in the presence of positive correlations between the two eyes, and (d) the extent to which the morphology of model stripes matched that of natural stripes. These were a correlational model [Miller et al 1989], an elastic net model [Goodhill & Willshaw 1990], a model based on a particular version of Kohonen's self-organizing feature map algorithm, and a competitive model which explicitly considered the presentation of a sequence of input patterns.
- A perturbation analysis of Miller's model of ocular dominance showed that binocular solutions can be favoured over monocular solutions if small correlations are introduced between the eyes or if there are negative correlations within each eye. The model is also not stable to perturbative differences in the correlations within the two eyes.
- We demonstrated that both the elastic net and a new Kohonen-type algorithm can produce topography and ocular dominance from the same developmental mechanisms. Considering the development of topography and ocular dominance as an optimization problem, we formulated various different ways of assessing the quality of the maps produced. With respect to these metrics it was shown that the elastic net generally produces better results than the Kohonen-type algorithm for comparable amounts of computation. The numerical methods of evaluating the map generally agreed with qualitative assessments of stripe morphology. Both algorithms were shown to correctly predict the results of monocular deprivation. Stripe width in both models is governed by a parameter we interpret to be the ratio of the strength of correlation within and between retinae.
- A particular type of competitive model in conjunction with a suitable weight normalization rule can simultaneously produce topography and ocular dominance from biologically plausible mechanisms. In particular, distributed patterns of activity can be presented simultaneously in the two retinae, which allows investigation of the influence of positive correlations between the eyes on stripe width. It was found that stripe width is *decreased* by increasing the strength of correlations between the eyes, and *increased* by increasing the spread of lateral interactions in the cortex. This behaviour combines aspects of both Miller's and the elastic net and Kohonen-type models.

7.3 Comparison of models

7.3.1 Topography and ocular dominance

The elastic net, Kohonen-type and competitive model each produce topography and ocular dominance from the same mechanisms, without any additional assumptions. It is not clear whether Miller's model is capable of this: we discuss this further below. One interesting difference is that in Miller's model refinement does not guarantee the development of dominance, whereas in the others receptive field refinement and the development of ocular dominance proceed hand in hand, driven by the same mechanisms.

7.3.2 Biological plausibility

Miller has discussed in some detail how the mechanisms and parameters of his model can be interpreted biologically [Miller 1989, Miller et al 1989, Miller 1990(a), Miller 1990(b), Miller & Stryker 1990]. However, it is

interesting to compare Miller's presentation with that of Linsker's for a similar model [Linsker 1986, Linsker 1988]. Linsker lays greater emphasis on the *simplicity* of the model, and the fact that it attempts to reproduce the behaviour of the natural system using a minimal number of rules. He refers to these rules as "biologically plausible", without attempting a more detailed justification. The elastic net and Kohonen-type models are more algorithmic in nature, although they can be interpreted as implementing Hebbian-type rules. Although we have identified distances as correlations in the representation of the problem used by these models, this has not been rigourously justified. The competitive model has the advantage that it is expressed in terms of a succession of inputs rather than averaged correlations: furthermore, these patterns are distributed and simultaneous in both retinae. However, certain assumptions are made regarding the mechanisms of competition and cooperation in the cortex, the plausibility of which has not been addressed.

7.3.3 Input Correlations

We have shown that monocularity in Miller's model begins to break down when there are positive correlations between the eyes. Our interpretation of distances in feature space as representing correlations suggests that the elastic net and Kohonen-type algorithms can form appropriate maps for positive cross-eye correlations. The competitive model gives appropriate maps for a range of positive correlations between the two eyes. Since it is driven by a succession of input patterns rather than the regular correlations of the elastic and Kohonen-type models, a much wider range of correlational structure can be investigated in this model. In the elastic net, Kohonen-type and competitive models stripes become narrower as the two eyes become more correlated, which is expected from intuitive arguments.

7.3.4 Stripe morphology

It is difficult to compare stripe morphology for the different models beyond the general statement that the segregated regions in each of the models are long, thin, and of roughly uniform width. There are a number of problems with making more detailed comparisons between the models. Firstly, Miller used periodic boundary conditions, which we have not used for the elastic and competitive models. Secondly, due to various technical features of the models, the stripes presented tend to be of different thicknesses. Thirdly, all these simulations are on a small scale.[1] However, we have seen that of the models for which simulation results are presented in this thesis, the elastic net tends to produce the most regular stripes (i.e. the most sharply-peaked power spectrum of the fourier transform), especially for small separations of the two retinae.

It is also difficult to make detailed comparisons with the natural system for reasons of scale. Whereas in all the models presented stripe width is generally just a few units across, in the natural system stripes can be many tens or hundreds of cells wide, and so comparisons with, for instance, the way the stripes terminate are not particularly meaningful. However, one area of stripe morphology where progress has been made in this thesis is of showing, as suggested by [LeVay et al 1985], that anisotropic boundary conditions can lead to an overall orientation for the stripes. A related issue to this is the degree to which the elastic and competitive models succeed in producing a "good" map. We have investigated this for the elastic net compared to a Kohonen-type algorithm, and also for the competitive model.

The elastic net and Kohonen-type algorithm reproduce the results of monocular deprivation without requiring any adjustments in the parameters apart from a reduction in strength for one eye. Monocular deprivation can be modelled in Miller's framework, but it is necessary to alter the form of the afferent normalization mechanism. Deprivation has not yet been investigated in the competitive model. However, it should be pointed out that, since it uses a similar afferent normalization mechanism to Miller's model, it is likely that similar adjustments to this will need to be made.

7.4 Further work

7.4.1 Miller's model

We have been able to point to some of Miller's experimental results that support certain aspects of the perturbation analysis presented in this thesis. However, a more thorough empirical investigation is needed, in

[1] Note that all the 40×40 arrays of cortical units presented by Miller have a border of 15×15 units which are repeated: this was done to demonstrate continuity across the edges of his array.

particular of the effect of positive correlations between the eyes, which Miller does not discuss. A more direct empirical comparison with the elastic, Kohonen-type and competitive models would also be useful.

Our analysis only considers the addition of very small correlations. An analytic investigation of more general cases would be of interest, although it is not immediately clear how to go about this. As a starting point larger perturbations could be investigated empirically.

So far the analysis of Miller (e.g. [Miller 1989]) has only considered first-order terms - a linear approximation. It would be interesting to analyse the role of higher-order terms, as was done for different but related systems in [Turing 1952] and [Häussler & von der Malsburg 1983].

The largest outstanding question regarding Miller's model is its ability or otherwise to form a topographic map. A fixed topography was assumed in all the simulation results Miller presented for his model: each cortical unit saw a 7×7 window of those presynaptic units directly below it, with connection strengths multiplied by the arbor function. We now consider what would happen if each cortical unit initially had non-zero connections to *all* units in the presynaptic sheet, and there was no arbor function making each unit in the postsynaptic sheet specific for the presynaptic units directly below it. There are two issues. Firstly, would receptive fields refine: that is, would the majority of connection strengths for each cortical unit go to zero leaving attention focussed on a small region of the presynaptic sheet? Secondly, would the regions that each cortical unit focussed on progress in an orderly way across the LGN in moving across the cortical sheet?

Some of Miller's simulation results are relevant to the first question. In [Miller et al 1989] the development of receptive fields for a number of individual cortical units is shown. Refinement occurs over time: eventually non-zero weights are restricted to a significantly smaller region than the original 7×7 window. Miller explains this [Miller 1990(a), pages 295–296] as being due to inputs in the centre of the receptive field having more correlated neighbours than inputs in the periphery (assuming the correlations are a decreasing function of distance). However, this seems to suggest that without the fixed topography, all cortical units would refine attention on the centre of the presynaptic sheet. In particular, there is no competitive pressure to learn different features. This makes the second question irrelevant.

To make these arguments more precise it might be possible to analyse this extended system in terms of its eigenvectors. The correlation matrix to be considered is now for the whole presynaptic sheet, rather than just the presynaptic region defined by the arbor function. Assuming gaussian correlations of range significantly less than the width of the presynaptic sheet, this correlation matrix would take the form of a gaussian "blur" down the leading diagonal (for the one-dimensional case), with zeros away from the neighbourhood of the leading diagonal. It might also be possible to investigate the spatial layout of these refined receptive fields by including the cortical iteration matrix in the eigenvector calculation, thus calculating both receptive field form and location for the whole set of cortical units simultaneously. We review below some very recent work that is relevant to this question.

7.4.2 The elastic net and Kohonen-type models

In the one-dimensional case of the elastic net model investigated in [Goodhill 1988, Goodhill & Willshaw 1990], it is possible to calculate explicitly which are the shortest paths, and thus to compare simulation results with these optimal results. It would be interesting to attempt similar calculations for the two-dimensional case. Although the general case is hard, it may be possible to compare analytically the sum D of distances between neighbouring points in the cortical sheet of certain general forms, such as parallel stripes of fixed width, a checker-board pattern, and the two-dimensional equivalent of the $\sqsupset$ path. This would enable us to see whether there are significantly shorter paths that the elastic net and Kohonen-type algorithm avoids, and whether these shorter paths represent patterns that are closer or otherwise to biological reality. It might also reveal something about the D-value of stripes with and without twists and turns. On the same theme it would be interesting to see the result of applying a more standard TSP algorithm to this optimization problem, to address again to what extent the elastic net and Kohonen-type algorithm are restricted to finding only a class of possible solutions.

There is an outstanding question as to why the elastic net but not the Tea Trade model is capable of forming stripes. A detailed comparison of the mathematics underlying both models might answer this, and in addition shed light on the correct biological interpretation of elastic net parameters.

A difficulty in the biological interpretation of the elastic net is that it is hard to see how the regularization term relates to a more realistic cortical interaction function as used in Miller's model. A similar problem applies to the version of neighbourhood updating used in the Kohonen-type algorithm. Recent analysis has begun to make this clearer [Dayan 1992].

A more accurate expression for the dependence of k_s on the parameters of the model could be found by following the techniques employed in [Durbin et al 1989]. k_s corresponds to the first bifurcation of the energy

function, i.e. the first time E has more than one minimum. Due to the highly regular arrangement of retinal units, this occurs for a much lower value of k than in the TSP case where cities are usually arranged more randomly, and so E develops a very large number of minima very quickly. In [Durbin et al 1989] the value of k at the first bifurcation was found by calculating the smallest eigenvalue of the Hessian matrix: when this is zero, the system bifurcates. This method could be applied to the ocular dominance case by writing down the Hessian matrix for the feature space of retinal units used here.

There are problems with knowing precisely what is being compared in chapter 5 between the elastic net and Kohonen-type algorithms. Firstly, it is not known whether the energy functions each optimizes have the same minima, so that the difference in results is only to do with differing abilities of the algorithms to optimize, or whether the energy functions actually have different minima. Secondly, our comparison of equal run times is obviously rather crude, being highly implementation-dependent. Further theoretical work might begin to answer these questions. Also, it would be interesting to compare the algorithms for even slower annealing.

7.4.3 Competitive model

An important feature of the competitive model is that in general, the effect of arbitrary input patterns can be investigated. In particular, it would be possible to use fragments of (suitably preprocessed) real images. Following [Barrow 1987], retinal and LGN processing could be simply modelled as a convolution with a difference of gaussians. This pre-processing would introduce negative components in the inputs, which are not allowed in the model as it is presently formulated, but this could perhaps be dealt with by separating on and off channels as in for example [Barrow & Bray 1992].

A mathematical analysis of the model would be enlightening from several points of view. Firstly, it would be interesting to know the precise analytical relationship between stripe width and the parameters of degree of correlation between the two eyes and extent of lateral interactions in the cortical sheet. Secondly, analysis might make it clearer how the model relates to Miller's and the elastic net, and also to the model of [Obermayer, Ritter & Schulten 1991]. It might be possible to apply some of the mathematical techniques of [Obermayer, Blasdel & Schulten 1991, Obermayer, Ritter & Schulten 1991] to this model, which would help to reveal the mathematical differences between presenting localized and distributed patterns of activity. One length parameter that could be expected to have an effect on stripe width is σ_r, the range of correlations within each retina. For the simulations presented here this was fixed at $\sigma_r = 1.5$: generally less than or approximately equal to the values used for σ_c, the range of the cortical interaction function. We might expect the case $\sigma_r > \sigma_c$ to produce interesting results, since now the correlations in the retinae "overflow" the range of lateral interactions in the cortex. For instance, in Amari's (monocular) model it was found that this regime led to a break-up of the previously smooth map into columnar microstructures [Takeuchi & Amari 1979]. In addition, the ability of the model to account for the results of monocular deprivation, i.e. a reduction in the strength of the input from one eye, needs to be investigated.

The biological plausibility of the model would be enhanced by adopting an explicit interaction function in the cortex as in for instance [Willshaw & von der Malsburg 1976] (and Miller's model), rather than the Kohonen approximation used up to now. However, a problem with this would be the large additional computational burden this would impose on an already computationally-intensive algorithm.

7.4.4 Unification of elastic and correlational models

We now review some very recent work [Yuille et al 1991(a), Yuille et al 1991(b)] that points to a mathematical link between the elastic net model and Miller's correlational model. The suggestion is that both can be derived from a more general objective function for the topography and ocular dominance problem, in a very similar way to that in which the elastic net and Hopfield and Tank's algorithm can be derived from a more general objective function for the TSP. [Yuille et al 1991(a)] suggest that an appropriate "generalized deformable model" for this problem has cost function

$$\begin{aligned} E[V^L, V^R, y, B] \;=\; & \sum_{i,j,a,b} V^L_{(ij)(ab)} \{ |x^L_{ij} - y_{ab}|^2 + (B_{ab} - h)^2 \} \\ & + \sum_{i,j,a,b} V^R_{(ij)(ab)} \{ |x^R_{ij} - y_{ab}|^2 + (B_{ab} + h)^2 \} \\ & + \nu \sum_{a,b} \{ (B_{a+1,b} - B_{ab})^2 + (B_{a,b+1} - B_{ab})^2 \} \\ & + \tau \sum_{a,b} \{ |y_{a+1,b} - y_{ab}|^2 + |y_{a,b+1} - y_{ab}|^2 \}. \end{aligned} \tag{7.1}$$

The x's represent retinal points: these are laid out in a grid for each retina. The y's represent points in a cortical sheet, and the B's represent the ocularity of each cortical point. y and B are both continuous variables. The V's are matching variables that specify which cortical points are matched to which retinal points. These are initially conceived to be binary variables, but become continuous under a statistical mechanics interpretation for finite temperature. The first two terms impose the matching constraints and the last two terms impose smoothness of the B and y fields. In [Yuille et al 1991(a)] it is shown that eliminating the V fields leads to an elastic net type energy function. It is also suggested that eliminating the B and y fields leads to a model closely related to Miller's.

The most interesting aspects of this work from our point of view are the following:

- It addresses the question as to the true nature of the relationship between Miller's model and the elastic net (cf [Goodhill & Willshaw 1990]).
- In order to identify the model derived by eliminating the B's and the y's with Miller's model, it is necessary for [Yuille et al 1991(a)] to explicitly identify distances with correlations: in particular their analysis suggests that the correlation between x_i and x_j should be $K - |x_i - x_j|^2$, where K is a constant. This agrees with the interpretation of distances as correlations we have adopted so far.
- Since the elastic model can produce topography by the same mechanism as ocular dominance, it may be possible to use this analysis to address theoretically the question of whether a correlational mechanism can produce topography. This is not considered in [Yuille et al 1991(a)].
- Extensions of this work provide insight into the relationship between the regularization term of the elastic model and cortical interaction functions, and show how more general neighbourhood terms may be included in the elastic net [Dayan 1992].

However, there are a number of technical errors and ambiguities in [Yuille et al 1991(a)] that need to be addressed. In addition, it would be interesting to see if the Kohonen-type algorithm and the competitive model can also be subsumed under this framework. Examples are given in [Yuille 1990] of generalized deformable models for competitive-type models of various phenomena.

7.4.5 Experimental issues

This thesis highlights a number of questions about the natural system which have not been answered experimentally. One of the most important of these here is the form of correlations between the two eyes. Various experiments have investigated the form of correlations resulting from spontaneous correlations *within* an eye (e.g. [Arnett 1978, Ginsburg et al 1984, Mastronade 1989, Meister et al 1991]). However, there are a number of problems with exploring the correlations *between* eyes. Firstly, such correlations are not likely to be produced by spontaneous firing: only in response to real visual input, and are thus much harder to measure. Secondly, it would be necessary to very closely match the positions of cells in the two eyes/lateral geniculate nuclei that were recorded from.

We have predicted that stripe width could be affected by the strength of the correlations between the two eyes. Thus it would be interesting to try to detect changes in stripe width in response to artificially induced strabismus. Since this severly reduces the extent of correlations between the two eyes, we predict an increase in stripe width. Although the effects of strabismus on the degree of dominance of individual cells has been investigated extensively, the effect on stripe width has not been examined (D. Hubel, Personal Communication). It would also be interesting to try and investigate the effect on stripe width of the cortical interaction function. One possibility might be to attempt to pharmacologically alter the form of lateral interactions in the cortex/tectum.

Two interesting questions which (as far as the author is aware) have not been answered experimentally are whether (a) there is any detailed similarity between the patterns of stripes in the two hemispheres, and (b) there is any similarity between the patterns in closely related individuals. Answers to both these would help to illuminate the extent to which the detailed form of the pattern is determined by genetic as opposed to environmental factors.

7.4.6 Measures and morphology

As we have already suggested, in order for work in this area to progress it is necessary to have better ways of evaluating the output of models, in terms of the degree of topography of the map produced, and the degree to which stripe morphology matches that of the natural system. The latter requires a more careful characterization

of natural stripes than has hitherto been performed. It will also be necessary to decide, out of the many possible features of this morphology that could be considered, which are the *key* features that should be used to constrain theoretical models.

7.4.7 Balancing models

A set of proposals regarding stripe formation that has not been properly addressed theoretically are the balancing models as discussed in chapter 3. Although these address only the interaction between topography and ocular dominance rather than the reason individual cells become ocularly dominant, a proper theoretical investigation could help to clarify a number of important issues. For instance, which kinds of patterns are produced by what kinds of mechanisms or interactions between mechanisms, and which classes of objective functions give patterns of stripes when minimized.

7.5 Summary: principal contributions

To conclude, we list the principal contributions made by this thesis.

- We have shown that computational models can be formulated in which topography and ocular dominance naturally arise from the same developmental mechanisms.
- We have shown analytically that the addition of perturbing correlations in Miller's model of ocular dominance leads to an instability, and an increased tendency to produce binocular rather than monocular units, thus highlighting a weakness of correlational models in general.
- We have addressed a number of issues related to elastic net and Kohonen-type models of ocular dominance, including the extent to which various measures of optimality imply good stripe morphology, and have shown that these models can naturally account for the results of monocular deprivation.
- We have introduced a new model based on a competitive mechanism, in which for the first time topography and ocular dominance are able to develop in the presence of distributed activity simultaneously in the two eyes. Stripe width in this model can be influenced by both the degree of correlation between the eyes and the extent of lateral interactions in the cortex. The former influence represents a novel prediction regarding the determiners of stripe width in the natural system.

Bibliography

Alexandroff, P. (1961). Elementary concepts of topology. Dover.

Altmann, L., Luhmann, H.J., Gruel, J.M. & Singer, W. (1987). Functional and neuronal binocularity in kittens raised with rapidly alternating monocular occlusion. *Jou. Neurophys.*, **58**, 965-980.

Amari, S. (1980). Topographic organization of nerve fields. *Bull. Math. Biol.*, **42**, 339-364.

Amari, S. (1983). Field theory of self-organizing nets. *IEEE Trans. Syst., Man, Cybernetics*, SMC-**13**, 741-748.

Amari, S. (1988). Dynamical stability of formation of cortical maps. In *Dynamic interactions in neural networks: models and data*, eds. M.A.Arbib & Amari, S, Springer Research Notes in Neural Computation, **1**, 15-34.

Angeniol, B., de la Croix Vaubois, G. & Le Texier, J. (1988). Self-organizing feature maps and the travelling salesman problem. *Neural Networks*, **1**, 289-293.

Arnett, D.W. (1978). Statistical dependence between neighbouring retinal ganglion cells in goldfish. *Exp. Brain. Res.*, **32**, 49-53.

Barrow, H.G. (1987). Learning receptive fields. *Proc. IEEE First Annual Conference on Neural Networks*, **IV**, 115-121.

Barrow, H.G. & Bray, A. (1992). A Model of Adaptive Development of Complex Cortical Cells. Accepted for ICANN92, Brighton, UK.

Bear, M.F. & Cooper, L.N. (1990). Molecular mechanisms of synaptic modification in visual cortex: interaction between theory and experiment. In M.A. Gluck & D.E. Rumelhart, editors, *Neuroscience and Connectionist Theory.* Hillsborough, NJ: Lawrence Erlbaum.

Bienenstock, E.L. (1980). A theory of development of neuronal selectivity. PhD Thesis, Brown University, Providence, RI.

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

Blakemore, C. (1978). Maturation and modification in the developing visual system. In *Handbook of Sensory Physiology. Perception*, eds. R. Held, H.W. Leibowitz, & H.L. Teuber **VIII**, 377-426. Springer-Verlag, Berlin.

Blasdel, G.G. & Pettigrew, J.D. (1979). Degree of interocular synchrony required for maintenance of binocularity in kitten's visual cortex. *Jou. Neurophsiol.*, **42**, 1692-1710.

Blasdel, G.G. & Salama, G. (1986). Voltage-sensitive dyes reveal a modular organization in monkey striate cortex. *Nature*, **321**, 579-585.

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

Boss, V.C. & Schmidt, J.T. (1984). Activity and the formation of ocular dominance patches in dually innervated tectum of goldfish. *Jou. Neurosci.*, **4**, 2891-2905.

Brown, M.C., Hopkins, W.G. & Keynes, R.J. (1991). Essentials of neural development (revised edition). Cambridge University Press, Cambridge.

Buisseret, P., Gary-Bobo, E. & Imbert, M. (1978). Ocular motility and recovery of orientational properties of visual cortical neurones in dark-reared kittens. *Nature*, **272**, 816-817.

Burr, D. J. (1988). An improved Elastic Net method for the travelling salesman problem. *Proc. IEEE International Conference on Neural Networks*, San Diego, California, July 24-27 1988.

Carpenter, G. & Grossberg, S. (1988). The ART of adaptive pattern recognition by a self-organizing neural network. *IEEE Computer*, March.

Changeux, J.P. & Danchin A. (1976). Selective stabilization of developing synapses as a mechanism for the specification of neural networks. *Nature*, **264**, 705-712.

Chapman, B., Jacobson, M.D., Reiter, H.O. & Stryker, M.P. (1986). Ocular dominance shift in kitten visual cortex caused by imbalance in retinal electrical activity. *Nature*, **324**, 154-156.

Churchland, P.S. & Sejnowski, T.J. (1988). Perspectives on Cognitive Neuroscience. *Science*, **242**, 741-745.

Constantine-Paton, M. (1983). Position and proximity in the development of maps and stripes. *Trends Neurosci.* **6**, 32-36.

Constantine-Paton, M., Cline, H.T. & Debski, E. (1990). Patterned activity, synaptic convergence, and the NMDA receptor in developing visual pathways. *Ann. Rev. Neurosci.*, **13**, 129-154.

Constantine-Paton, M. & Law, M.I. (1978). Eye-specific termination bands in tecta of three-eyed frogs. *Science*, **202**, 639-641.

Constantine-Paton, M. & Law, M.I. (1982). The development of maps and stripes in the brain. *Sci. Am.*, **247**, 54-62.

Cook, J.E. (1987). A sharp retinal image increases the topographic precision of the goldfish retinotectal projection during optic nerve regeneration in stroboscopic light. *Exp. Brain Res.*, **68**, 319-328.

Cook, J.E. (1988). Topographic refinement of the goldfish retinotectal projection: sensitivity to stroboscopic light at different periods during optic nerve regeneration. *Exp. Brain Res.*, **70**, 109-116.

Cook, J.E. & Becker, D.L. (1990). Spontaneous activity as a determinant of axonal connections: darkness and diurnal light are equally effective for activity-dependent refinement of the regenerating retinotectal projection in goldfish. *European Journal of Neuroscience*, **2**, 162-169.

Cook, J.E. & Rankin, E.C.C. (1986). Impaired refinement of the regenerated retinotectal projection of the goldfish in stroboscopic light: a quantitative WGA-HRP study. *Exp. Brain. Res.*, **63**, 421-430.

Cottrell, M. & Fort, J.C. (1986). A stochastic model of retinotopy: a self organizing process. *Biol. Cybern.*, **53**, 405-411.

Cowan, J.D. & Friedman, A.E. (1990). Development and regeneration of eye-brain maps: A computational model. In D.S. Touretzky, ed, *Advances in Neural Information Processing Systems*, **II**, 92-99.

Cowan, J.D. & Friedman, A.E. (1990). Development and regeneration of eye-brain maps: A computational model. Unpublished manuscript.

Cowan, J.D. & Friedman, A.E. (1991). Studies of a model for the development and regeneration of eye-brain maps. In D.S. Touretzky, ed, *Advances in Neural Information Processing Systems*, **III**, 3-10.

Cowey, A. (1979). Cortical maps and visual perception. *Qua. Jou. Exper. Psychol.*, **31**, 1-17.

Cronly-Dillon, J.R. & Glaizner, B. (1974). Specificity of regenerating optic fibres for left and right optic tecta in goldfish. *Nature*, **251**, 505-507.

Cynader, M. (1979). Coopetitive interactions in postnatal development of the kitten's visual system. In *Developmental Neurobiology of Vision. NATO Advanced Study Institute* ed. Freeman, R.D., Plenum Press.

Da Silva Filho, A.C.R. (1991). An analytical study of the resolution and amplification of properties of neural maps. Manuscript submitted for publication.

Dayan, P.S. (1992). Arbitrary elastic topologies. Manuscript submitted for publication.

Dayan, P.S. & Goodhill, G.J. (1992) Perturbing Hebbian rules. To appear in *Advances in Neural Information Processing Systems*, **4**, eds. J.E. Moody, S.J. Hanson and R.P. Lippman, Morgan Kaufman, San Mateo, CA.

Dempster, A.P., Laird, N.M. & Rubin, D.B. (1977). Maximum likelihood from incomplete data using the EM algorithm. *Jou. Roy. Statistical. Soc. B*, **39**, 1-37.

Durbin, R. & Mitchison, G. (1990). A dimension reduction framework for understanding cortical maps. *Nature*, **343**, 644-647.

Durbin, R., Szeliski, R. & Yuille, A. (1989). An analysis of the elastic net approach to the traveling salesman problem. *Neural Computation*, **1**, 348-358.

Durbin, R. & Willshaw, D.J. (1987). An analogue approach to the travelling salesman problem using an elastic net method. *Nature*, **326**, 689-691.

Edelman, G.M., Gall, W.E. & Cowan, W.M. (1985). Map formation in the retinotectal system. *Molecular Bases of Neural Development*, eds. G.M. Edelman, W.E. Gall & W.M. Cowan, Wiley, New York.

Érdi, P. & Barna, Gy. (1984). Self-organizing mechanism for the formation of ordered neural mappings. *Biol. Cybern.*, **51**, 93-101.

Erwin, E., Obermayer, K. & Schulten, K. (1991). Convergence properties of self-organizing maps. *Proc. Int. Conf. Artificial Neural Networks, Helsinki*.

Fawcett, J.W. & O'Leary, D.D.M. (1985). The role of electrical activity in the formation of topographic maps in the nervous system. *Trends Neurosci.*, **8**, 201-206

Fawcett, J.W. & Willshaw, D.J. (1982). Compound eyes project stripes on the optic tectum in Xenopus. *Nature*, **296**, 350-352.

Ferrer, J.M.R., Price, D.J. & Blakemore, C. (1988). The organization of corticocortical projections from area 17 to area 18 of the cat's visual cortex. *Proc. R. Soc. Lond. B.*, **233**, 77-98.

Field, D. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Jou. Opt. Soc. Am.*, **4**, 2379-2394.

Fort, J.C. (1988). Solving a combinatorial problem via self-organizing process: an application of the Kohonen algorithm to the travelling salesman problem. *Biol. Cybern.*, **59**, 33-40.

Fraser, S.E. (1980). A differential adhesion approach to the patterning of neural connections. *Dev. Biol.*, **79**, 453-464.

Fraser, S.E. (1985). Cell interactions involved in neural patterning. In *Molecular Bases Of Neural Development*, eds. G.M. Edelman, W.E. Gall & W.M. Cowan, 481-507, Wiley, New York.

Fraser, S.E. & Hunt, R.K. (1980). Retinotectal specificity: models and experiments in search of a mapping function. *Ann. Rev. Neurosci.*, **3**, 319-352.

Fraser, S.E. & Perkel, D.H. (1990). Competitive and positional cues in the patterning of nerve connections. *Jou. of Neurobiol.*, **21**, 51-72.

Frean, M.R. (1990). Small nets and short paths: optimising neural computation. PhD thesis, University of Edinburgh.

Freeman, R.D. & Bonds, A.B. (1979). Cortical plasticity in monocularly deprived immobilized kittens depends on eye movement. *Science*, **206**, 1093-1095.

Freeman, R.D., Sclar, G. & Ohzawa, I. (1982). Cortical binocularity is disrupted by strabismus more slowly than by monocular deprivation. *Dev. Br. Res.*, **3**, 311-316.

Frégnac, Y. & Imbert, M. (1978). Early development of visual cortical cells in normal and dark reared kittens: relationship between orientation selectivity and ocular dominance. *Jou. Physiol.*, **278**, 27-44.

Frégnac, Y., Shulz, D., Thorp, S. & Bienenstock, E. (1988). A cellular analogue of visual cortical plasticity. *Nature*, **333**, 367-370.

Galli, L. & Maffei, L. (1988). Spontaneous impulse activity of rat retinal ganglion cells in prenatal life. *Science*, **242**, 90-91.

Gaze, R.M. & Keating, M.J. (1972). The visual system and "neuronal specificity". *Nature*, **237**, 375-378.

Gierer, A. (1988). Spatial organization and genetic information in brain development. *Biol. Cybern.*, **59**, 13-21.

Ginsburg, K.S., Johnsen, J.A. & Levine, M.W. (1984). Common noise in the firing of neighbouring ganglion cells in goldfish retina. *Jou. Physiol.*, **351**, 433-450.

Goodhill, G.J. (1988). Application of the elastic net algorithm to the formation of ocular dominance stripes. Unpublished M.Sc. thesis, University of Edinburgh.

Goodhill, G.J. (1990). The development of topography and ocular dominance. In Touretzky, D.S., Elman, J.L., Sejnowski, T.J. & Hinton, G.E. (eds.) *Proceedings of the 1990 Connectionist Models Summer School*, San Mateo, CA: Morgan Kaufman.

Goodhill, G.J. (1991). Topography and ocular dominance can arise from distributed patterns of activity. *International Joint Conference on Neural Networks, Seattle, July 1991*, **II**, 623-627.

Goodhill, G.J. & Willshaw, D.J. (1990). Application of the elastic net algorithm to the formation of ocular dominance stripes. *Network*, **1**, 41-59.

Gottlieb, D.I. & Glaser, L. (1981). Cellular recognition during neural development. *Studies in Developmental Neurobiology*, ed. W.M. Cowan, 243-260, Oxford University Press, New York.

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

Hankin, M. & Lund, R. (1991). How do retinal axons find their targets in the developing brain? *Trends Neurosci.*, **14**, 224-228.

Häussler, A.F. & Malsburg, C. von der (1983). Development of retinotopic projections: an analytical treatment. *Jou. Theoret. Neurobiol.*, **2**, 47-73.

Hebb, D.O. (1949). The Organization of Behaviour. New York: Wiley.

Hendrickson, A.E. (1985). Dots, stripes and columns in monkey visual cortex. *Trends Neurosci.*, Sept, 406-410.

Hertz, J., Krogh A. & Palmer, R.G. (1991). Introduction to the Theory of Neural Computation. Lecture notes in the Santa Fe Institute Studies in the sciences of complexity: Adison Wesley.

Hope, R.A., Hammond, B.J. & Gaze, R.M. (1976). The arrow model: retinotectal specificity and map formation in the goldfish visual system. *Proc. Roy. Soc. Lond. B.*, **194**, 447-466.

Hopfield, J.J. & Tank, D.W. (1985). Neural computation of decisions in optimization problems. *Biol. Cybern.*, **52**, 141-152.

Horn, R.A. & Johnson, C.R. (1985). Matrix Analysis. Cambridge University Press.

Horton, J.C., Dagi, L.R., McCrane, E.P. and de Monasterio, F.M. (1990). Arrangement of ocular dominance columns in human visual cortex. *Arch-Ophthalmol.*, **108**, 1025-1031.

Horton, J.C., Greenwood, M.M. & Hubel, D.H. (1979). Non-retinotopic arrangement of fibres in cat optic nerve. *Nature*, **282**, 720-722.

Hubel, D.H. (1990). Eye, Brain, and Vision. Scientific American Library, W.H. Freeman and Company, New York.

Hubel, D.H. & Wiesel, T.N. (1965). Binocular interaction in striate cortex of kittens reared with artificial squint. *Journal of Neurophysiology*, **28**, 1041-1059.

Hubel, D.H. & Wiesel, T.N. (1977). Functional architecture of the macaque monkey visual cortex. *Proc. R. Soc. Lond. B*, **198**, 1-59.

Hubel, D.H., Wiesel, T.N. & LeVay, S. (1977). Plasticity of ocular dominance columns in monkey striate cortex. *Phil. Trans. R. Soc. Lond. B*, **278**, 377-409.

Hunt, R.K. & Jacobson, M. (1972). Specification of positional information in retinal ganglion cells of *Xenopus*: stability of the specified state. *Proc. Nat. Acad. Sci., U.S.A.*, **69**, 2860-2864.

Ide, C.F., Fraser, S.E. & Meyer, R.L. (1983). Eye dominance columns formed by an isogenic double-nasal frog eye. *Science*, **221**, 293-295.

Jacobson, M. & Levine, R.L. (1975). Plasticity in the adult frog brain: filling in the visual scotoma after excision or translocation of parts of the optic tectum. *Brain Res.*, **88**, 339-345.

Jones, D.G., Van Sluyters, R.C. & Murphy, K.M. (1991). A computational model for the overall pattern of ocular dominance. *J. Neurosci.*, **11**, 3794-3808.

Keddy, P.A. (1989). Competition. Chapman and Hall, New York.

Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P. (1983). Optimization by simulated annealing. *Science*, **220**, 671-680.

Knudsen, E.I., du Lac, S. & Esterly, S.D. (1987). Computational maps in the brain. *Ann. Rev. Neurosci.*, **10**, 41-65.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biol. Cybern.*, **43**, 59-69.

Kohonen, T. (1982). Analysis of a simple self-organizing process. *Biol. Cybern.*, **44**, 135-140.

Kohonen, T. (1988). Self-organization and associative memory. Springer, Berlin.

Kolodny, J.A. (1990). Generalized deformable models for early vision development. B.A. Thesis, Harvard University.

Kuffler, S.W., Nicholls, J.G. & Martin, A.R. (1984). From neuron to brain. A cellular approach to the function of the nervous system. 2nd Edition. Sinauer, Massachusetts.

Langley, J.N. (1895). Note on regeneration of præ-ganglionic fibres of the sympathetic. *J. Physiol.*, **18**, 280-284.

Law, M.I. & Constantine-Paton, M. (1980). Right and left eye bands in frogs with unilateral tectal ablations. *Proc. Natl. Acad. Sci., U.S.A.*, **77**, 2314-2318.

Law, M.I. & Constantine-Paton, M. (1981). Anatomy and physiology of experimentally produced striped tecta. *Jou. Neurosci.*, **1**, 741-759.

Lawler, E.L., Lenstra, J.K., Rinnooy Kan, A.H.G. & Shmoys, D.B. (1986). The Travelling Salesman Problem. John Wiley and Sons, Chichester, England.

LeVay, S., Connolly, M., Houde, J. & Van Essen, D.C. (1985). The complete pattern of ocular dominance stripes in the striate cortex and visual field of the macaque monkey. *Jou. Neurosci.*, **5**, 486-501.

LeVay, S., Hubel, D.H. & Wiesel, T.N. (1975). The pattern of ocular dominance columns in macaque monkey visual cortex revealed by a reduced silver stain. *Jou. Comp. Neurol.*, **159**, 559-576.

LeVay, S., Wiesal, T.N. & Hubel, D.H. (1980). The development of ocular dominance columns in normal and visually deprived monkeys. *Jou. Comp. Neurol.*, **191**, 1-51.

Levine, R.L. & Jacobson, M. (1975) Discontinuous mapping of retina onto tectum innervated by both eyes. *Brain Res.*, **98**, 172-176.

Linsker, R. (1986). From basic network principles to neural architecture (series). *Proc. Nat. Acad. Sci., USA*, **83**, 7508-7512, 8390-8394, 8779-8783.

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

Lund, J.S. (1988). Anatomical organization of macaque monkey striate visual cortex. *Ann. Rev. Neurosci.*, **11**, 253-288.

MacKay, D.J.C. & Miller, K.D. (1990). Analysis of Linsker's simulations of Hebbian rules. *Neural Computation*, **2**, 169-182.

MacKay, D.J.C. & Miller, K.D. (1990). Analysis of Linsker's application of Hebbian rules to linear networks. *Network*, **1**, 257-297.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. *Fifth Berkeley Symposium on Mathematical Statistics and Probability 1965-66*, eds. J. Neyman & L.M. Le Cam. University of California Press.

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

Malsburg, C. von der (1979). Development of ocularity domains and growth behaviour of axon terminals. *Biol. Cybern.*, **32**, 49-62.

Malsburg, C. von der (1989). Network self-organization. In *An Introduction to Neural and Electronic Networks*, S.F. Zornetzer, J. Davis & C. Lau, eds., Academic Press, 421-432.

Malsburg, C. von der & Singer, W. (1988). Principles of cortical network organization. In *Neurobiology of Neocortex*, eds. P. Rakic & W. Singer, John Wiley and Sons, 69-99.

Malsburg, C. von der & Willshaw, D.J. (1976). A mechanism for producing continuous neural mappings: ocularity dominance stripes and ordered retino-tectal projections. *Exp. Brain. Res. Supplementum 1*, 463-469.

Malsburg, C. von der & Willshaw, D.J. (1977). How to label nerve cells so that they can interconnect in an ordered fashion. *Proc. Nat. Acad. Sci., U.S.A.*, **74**, 5176-5178.

Malsburg, C. von der & Willshaw, D.J. (1981). Co-operitivity and brain organization. *Trends Neurosci.*, April 80-83.

Malsburg, C. von der & Willshaw, D.J. (1981). Differential equations for the development of topological nerve connections. *SIAM-AMS Proceedings*, **13**, 39-47.

Marr, D. (1982). Vision. W.H. Freeman and Company, New York.

Mastronade, D.N. (1989). Correlated firing of retinal ganglion cells. *Trends Neurosci.*, **12**, 2, 75-80.

Meinhardt, H. (1982). Models of biological pattern formation. Academic Press, London.

Meister, M., Wong, R.O.L., Baylor, D.A. & Shatz, C.J. (1991). Synchronous bursts of action potentials in ganglion cells of the developing mammalian retina. *Science*, **252**, 939-943.

Meyer, R.L. (1979). Retinotectal projection in goldfish to an inappropriate region with a reversal in polarity. *Science*, **205**, 819-821.

Meyer, R.L. (1982). Ordering of retinotectal connections: a multivariate operational analysis. *Current Topics in Developmental Biology*, **17**, 101-145.

Meyer, R.L. (1982). Tetrodotoxin blocks the formation of ocular dominance columns in goldfish. *Science*, **218**, 589-591.

Meyer, R.L. (1983). Tetrodotoxin inhibits the formation of refined retinotopography in goldfish. *Dev. Brain. Res.*, **6**, 293-298.

Miller, K.D. (1989). Correlation-based Mechanisms in Visual Cortex: Theoretical and Empirical Studies. PhD Thesis, Stanford University Medical School.

Miller, KD (1990). Correlation-based mechanisms of neural development. In M.A. Gluck & D.E. Rumelhart, editors, *Neuroscience and Connectionist Theory.* Hillsborough, NJ: Lawrence Erlbaum.

Miller, K.D. (1990). Derivation of linear Hebbian equations from a nonlinear Hebbian model of synaptic plasticity. *Neural Computation,* **2**, 321-333.

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

Miller, K.D. & MacKay, D.J.C. (unpbublished). The role of constraints in Hebbian Learning.

Miller, K.D. & Stryker, M.P. (1990). The development of ocular dominance columns: mechanisms and models. In *Connectionist modeling and brain function: the developing interface*, eds. S.J. Hanson and C.R. Olsen, MIT Press, Cambridge, MA.

Mitchison, G. (1991). Neuronal branching patterns and the economy of cortical wiring. *Proc. Roy. Soc. B.*, **245**, 151-158.

Mitchison, G. (1992). Axonal trees and cortical architecture. *Trends Neurosci.*, **15**, No. 4, 122-126.

Mitchison, G. & Durbin, R. (1986). Optimal numberings of an $N \times N$ array. *SIAM J. Alg. Disc. Meth.*, **7**, 571-581.

Movshon, J.A. & Blakemore, C. (1974). Functional reinnervation in kitten visual cortex. *Nature*, **251**, 504-505.

Mower, G.D., Christen, W.G. & Caplan, C.J. (1983). Very brief visual experience eliminates plasticity in the cat visual cortex. *Science*, **221**, 178-180.

Munro, P.W. (1986). State-dependant factors affecting neural plasticity: a partial account of the critical period. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, eds. J.L. McClelland & D.E. Rumelhart, 471-503, MIT Press, Cambridge, M.A.

Nelson, M.E. & Bower, J.M. (1990). Brain maps and parallel computers. *Trends Neurosci.*, **13**, 403-408.

Nowlan, S.J. (1990). Max likelihood competition in RBF networks. *Technical report CRG-TR-90-2, University of Toronto*

Obermayer, K., Blasdel, G.G. & Schulten, K. (1991). A neural network model for the formation and for the spatial structure of retinotopic maps, orientation- and ocular dominance columns. *Proc. Int. Conf. Artificial Neural Networks, Helsinki*.

Obermayer, K., Ritter, H. & Schulten, K. (1990). A principle for the formation of the spatial structure of cortical feature maps. *Proc. Nat. Acad. Sci., U.S.A.*, **87**, 8345-8349.

Obermayer, K., Ritter, H. & Schulten, K. (1991). Development and spatial structure of cortical feature maps: a model study. In *Neural information Processing Systems 3*, eds. R.P. Lippmann, J. Moody & D.S. Touretzky, Morgan Kaufmann, CA.

Overton, K.J. & Arbib, M.A. (1982). The extended branch-arrow model of the formation of retino-tectal connections. *Biol. Cybern.*, **45**, 157-175.

Peterson, C. (1990). Parallel distributed approaches to combinatorial optimization: benchmark studies on traveling salesman problem. *Neural Computation*, **2**, 261-269.

Press, W.H., Flannery, B.P., Teukolsky, S.A. & Vetterling, W.T. (1988). Numerical recipes in C: the art of scientific computing. Cambridge University Press: Cambridge.

Prestige, M.C. & Willshaw, D.J. (1975). On a role for competition in the formation of patterned neural connexions. *Proc. R. Soc. Lond. B*, **190**, 77-98.

Price, D.J. (1991). The development of visual cortical afferents. In: *Development and plasticity of the visual system*, ed. J.R. Cronly-Dillon, Macmillan, London.

Rakic, P. (1976). Prenatal genesis of connections subserving ocular dominance in the rhesus monkey. *Nature*, **261**, 467-471.

Rankin, E.C.C & Cook, J.E. (1986). Topographic refinement of the regenerating retinotectal projection of the goldfish in standard laboratory conditions: a quantitative WGA-HRP study. *Exp. Brain Res.*, **63**, 409-420.

Rauschecker, J.P. (1991). Mechanisms of visual plasticity: Hebb synapses, NMDA receptors, and beyond. *Physiol. Rev.*, **71**, 587-615.

Reh, T.A. & Constantine-Paton, M. (1985). Eye-specific segregation requires neural activity in three-eyed Rana pipiens. *Jou. Neurosci.*, **5**, 1132-1143.

Reiter, H.O. & Stryker, M.P. (1988). Neural plasticity without postsynaptic action potentials: less active inputs become dominant when kitten visual cortical cells are pharmacologically inhibited. *Proc. Nat. Acad. Sci. USA*, **85**, 3623-3627.

Ritter, H., Obermayer, K., Schulten, K. & Rubner (1990). Self-organizing maps and adaptive filters. In *Physics of neural networks*, Springer, New York.

Ritter, H. & Schulten, K. (1988). Convergence properties of Kohonen's topology conserving maps: fluctuations, stability and dimension selection. *Biol. Cybern.*, **60**, 59-71.

Roe, A.W., Pallas, S.L., Hahm, J. & Sur, M. (1990). A map of visual space induced in primary auditory cortex. *Science*, **250**, 818-820.

Rumelhart, D.E. & Zipser, D. (1986). Feature discovery by competitive learning. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, eds. J.L. McClelland & D.E. Rumelhart, 151-193, MIT Press, Cambridge, M.A.

Schmidt, J.T. (1978). Retinal fibres alter tectal position markers during the expansion of the half retinal projection in goldfish. *Jou. Comp. Neurol.*, **177**, 279-300.

Schmidt, J.T. (1985). Factors involved in retinotopic map formation: complementary roles for membrane recognition and activity-dependent synaptic stabilization. *Molecular Bases of Neural Development*, eds. G.M. Edelman, W.E. Gall & W.M. Cowan, Wiley, New York.

Schmidt, J.T. (1988). Activity-dependent sharpening of the retinotopic proection on the tectum of goldfish in regeneration and development. *The making of the nervous system*, eds. J.G. Parnavelas, C.D. Stern and R.V. Stirling, Oxford 1988.

Schmidt, J.T., Cicerone, C.M. & Easter, S.S. (1978). Expansion of the half retinal projection to the tectum in goldfish: an electrophysiological and anatomical study. *Jou. Comp. Neurol.*, **177**, 257-278.

Schmidt, J.T. & Edwards, D.L. (1983). Activity sharpens the map during the regeneration of the retinotectal projection in goldfish. *Brain Research*, **269**, 29-39.

Schmidt, J.T. & Eisele, L.E. (1985). Stroboscopic illumination and dark-rearing block the sharpening of the regenerated retinotectal map in goldfish. *Neuroscience*, **14**, 535-546.

Sejnowski, T.J., Koch, C. & Churchland, P.S. (1988). Computational Neuroscience. *Science*, **241**, 1299-1306.

Sejnowski, T.J. & Tesauro, G. (1989). The Hebb rule for synaptic plasticity: algorithms and implementations. In *Neural Models of Plasticity: experimental and theoretical approaches*, eds. J.H.Byrne & W.O. Berry. Academic Press, San Diego.

Shatz, C.J. & Stryker, M.P. (1978). Ocular dominance in layer IV of the cat's visual cortex and the effects of monocular deprivation. *Jou. Physiol.*, **281**, 267-283.

Shatz, C.J. & Stryker, M.P. (1988). Prenatal tetrodotoxin infusion blocks segregation of retinogeniculate afferents. *Science*, **242**, 87-89.

Schiff, L.I. (1968). Quantum Mechanics (Third Edition). McGraw-Hill, Simgapore.

Simic, P.D. (1990). Statistical mechanics as the underlying theory of 'elastic' and 'neural' optimisations. *Network*, **1**, 89-103.

Simic, P.D. (1991). Constrained nets for graph matching and other quadratic assignment problems. *Neural Computation*, **2**, 268-281.

Simmen, M.W. (1991). Parameter sensitivity of the elastic net approach to the travelling salesman problem. *Neural Computation*, **3**, 363-374.

Simmen, M.W. (1992). PhD Thesis, forthcoming.

Sperry, R.W. (1963). Chemoaffinity in the orderly growth of nerve fiber patterns and connections. *Proc. Nat. Acad. Sci., U.S.A.,* **50**, 703-710.

Stanton, P.K. & Sejnowski, T. (1989). Associative long-term depression in the hippocampus induced by hebbian covariance. *Nature*, **339**, 215.

Stent, G.S. (1973). A physiological mechanism for Hebb's postulate of learning. *Proc. Nat. Acad. Sci. USA*, **70**, 997-1001.

Straznicky, C. & Tay, D. (1982). Retinotectal map formation in dually innervated tecta: a regeneration study in *Xenopus* with one compound eye following bilateral optic nerve section. *Jou. Comp. Neurol.*, **206**, 119-130.

Stryker, M.P. & Harris, W. (1986). Binocular impulse blockade prevents the formation of ocular dominance columns in cat visual cortex. *Jou. Neurosci.*, **6**, 2117-2133.

Sur, M., Merzenich, M.M. & Kaas, J.H. (1980). Magnification, receptive-field area and "hypercolumn" size in areas 3b and 1 of somatosensory cortex in owl monkeys. *Jou. Nerophysiol.*, **44**, 295-311.

Swindale, N.V. (1979). How ocular dominance stripes may be formed. In *Developmental neurobiology of vision*, ed. R.D. Freeman. NATO advanced study institute series. Series A, Life sciences, **27**, 267-273. New York: Plenum Press.

Swindale, N.V. (1980). A model for the formation of ocular dominance stripes. *Proc. R. Soc. Lond. B*, **208**, 243-264.

Swindale, N.V. (1981). Absence of ocular dominance patches in dark-reared cats. *Nature*, **290**, 332-333.

Swindale, N.V. (1982). A model for the formation of orientation columns. *Proc. R. Soc. Lond. B*, **215**, 211-230.

Swindale, N.V. (1992). A model for the coordinated development of columnar systems in primate striate cortex. *Biol. Cybern.*, **66**, 217-230.

Swindale, N.V. & Cynader, M.S. (1986). Physiological segregation of geniculo-cortical afferents in the visual cortex of dark-reared cats. *Brain Res.*, **362**, 281-286.

Takeuchi, A. & Amari, S. (1979). Formation of topographic maps and columnar microstructures in nerve fields. *Biol. Cybern.*, **35**, 63-72.

Thompson, I. (1985). Stars and stripes present the colours. *Nature*, **315**, 277-278.

Trisler, G.D. (1982). Are molecular markers of cell position involved in the formation of neural circuits? *Trends Neurosci.*, **5**, 306-310.

Turing, A.M. (1952). The chemical basis of morphogenesis. *Phil. Trans. Roy. Soc. Lond. B*, **237**, 37-72.

Udin, S.B. & Fawcett, J.W. (1988). Formation of topographic maps. *Ann. Rev. Neurosci.*, **11**, 289-327.

Whitelaw, V.A. & Cowan, J.D. (1981). Specificity and plasticity of retinotectal connections: a computational model. *Jou. Neurosci.*, **1**, 1369-1387.

Willshaw, D.J. & Dayan, P.S. (1990). Optimal plasticity from matrix memories: what goes up must come down. *Neural Computation*, **2**, 85-93.

Willshaw, D.J., Fawcett, J.W. & Gaze, R.M. (1983). The visuotectal projections made by *Xenopus* 'pie slice' compound eyes. *J. Embryol. Exp. Morphol.*, **94**, 121-137.

Willshaw, D.J. & Malsburg, C. von der (1976). How patterned neural connections can be set up by self-organization. *Proc. R. Soc. Lond. B*, **194**, 431-445.

Willshaw, D.J. & Malsburg, C. von der (1979). A marker induction mechanism for the establishment of ordered neural mappings: its application to the retinotectal problem. *Phil. Trans. Roy. Soc. B*, **287**, 203-243.

Wilson, G.V. & Pawley, G.S. (1988). On the stability of the travelling salesman problem algorithm of Hopfield and Tank. *Biol. Cybern.*, **58**, 63-70.

Yoon, M. (1971). Reorganization of retinotectal projection following surgical operations on the optic tectum in goldfish. *Exp. Neurol.*, **33**, 395-411.

Yuille, A.L. (1990). Generalized deformable models, statistical physics, and matching problems. *Neural Computation*, **2**, 1-24.

Yuille, A.L., Kolodny, J.A. & Lee, C.W. (1991). Dimension reduction, generalized deformable models and the development of ocularity and orientation. *Harvard Robotics Laboratory Technical Report no. 91-3.*

Yuille, A.L., Kolodny, J.A. & Lee, C.W. (1991). Dimension reduction, generalized deformable models and the development of ocularity and orientation. *International Joint Conference on Neural Networks, Seattle, July 1991*, **II**, 597-602.