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.

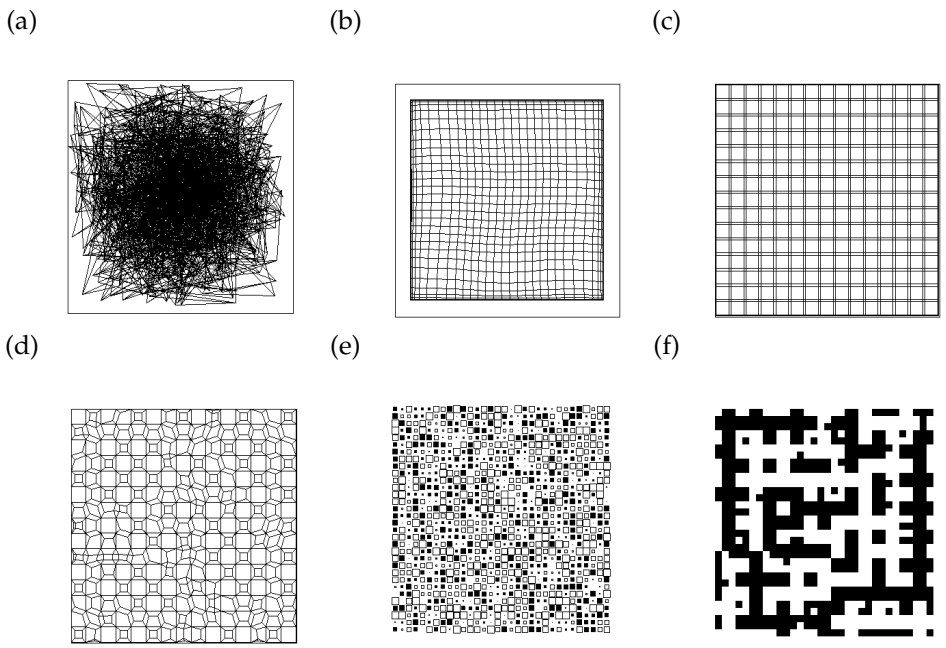

Figure 5.5: The time course of a typical elastic net simulation. There are two types of picture: (a-d) show the positions of the points in the elastic sheet as viewed from above, while (e-f) show the ocularity of each cortical unit. (a) Initially the points in the cortical sheet have large random scatter 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 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 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

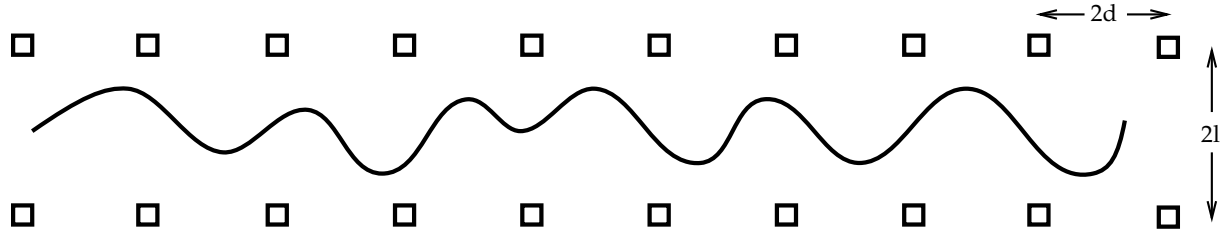

Figure 5.4: The representation of the topography and ocular dominance problem used in the one-dimensional case. The retinal units (squares) are arranged in two parallel rows. The rope, representing the mapping to the cortex, moves in between. The two dimensional case is analogous.

(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):

- 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

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

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

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

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

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 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.

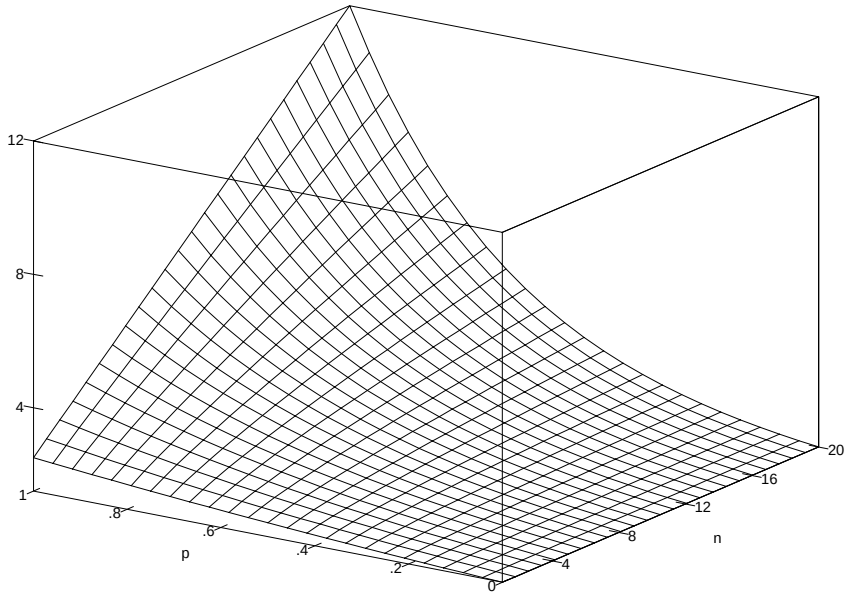

Figure 5.3: Neighbourhood and correspondence terms in minimal wiring argument plotted as a function of the norm parameter p and the stripe width n.

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

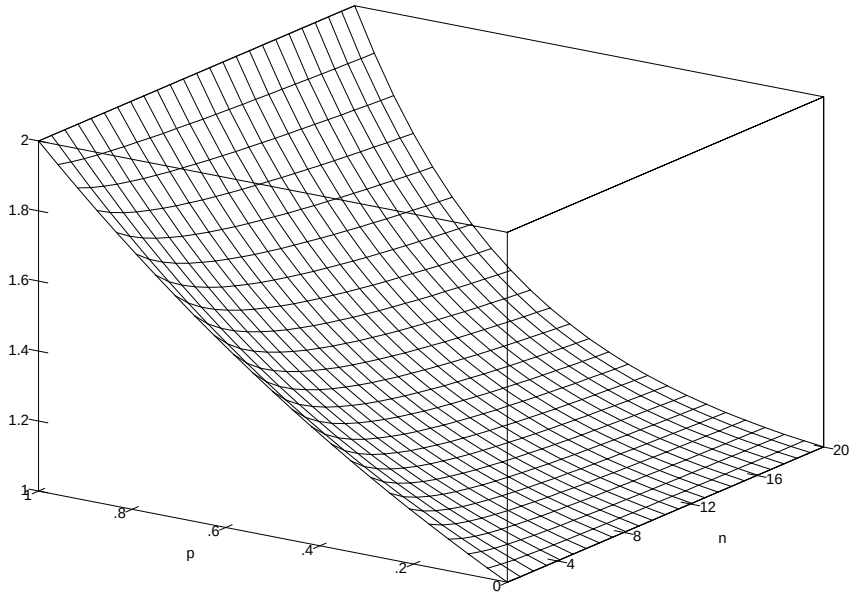

Figure 5.2: Neighbourhood term in minimal wiring argument plotted as a function of the norm parameter p and the stripe width n.

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 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.

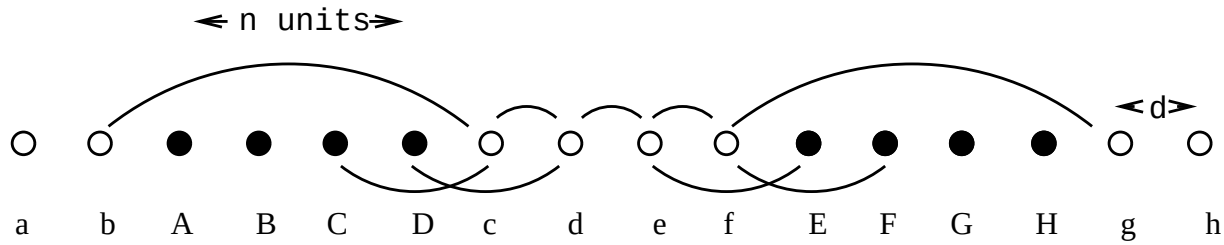

Figure 5.1: Application of the minimal wiring argument to the one dimensional topography and ocular dominance problem. White circles represent cortical units connected to the left eye at topographic position indexed by lower case letters. Black circles represent cortical units connected to the right 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.

(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.

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

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]:

Chapter 5

Elastic Models