Chapter 4

An initial model of the retinogeniculate pathway

4.1 Introduction

This chapter introduces a model of retinogeniculate development proposed by Keesing et al. (1992). The first half of this chapter introduces the main components of the model and the visualisation techniques used to analyse network performance. We then present our replication of the results from the original publication. We also investigate the role of the different forms of the initial weight bias upon development.

In the second half of this chapter, the role of the different weight normalisation schemes in the network are investigated. The original model used both pre- and postsynaptic normalisation, although the results presented here indicate that only presynaptic normalisation is required for normal development. This is in contrast to the models reviewed in Chapter 3, where all of the models required postsynaptic normalisation for ocular dominance development. Given the nature of the projection columns within the LGN, we demonstrate that the presynaptic normalisation should be divisively implemented. However, results are presented showing that after certain modifications, subtractive presynaptic normalisation can also generate the required projection columns.

4.2 The initial model

The Keesing model consists of two one-dimensional retina fully connected to a two-dimensional LGN, as shown in Figure 4.1. The reduction in dimensionality of both the retina and LGN was made to keep the computational demands low.

The retinogeniculate model consists of only feedforward connections. There is no modelling of either the feedback pathway from the cortex or of the local inhibitory circuitry within the LGN (both described in Chapter 2). Both the inhibitory circuit and feedback pathway develop later than the feedforward connections, and so are likely to either be non-existent or have little effect in comparison with the feedforward circuitry. First, the lateral inhibition in the cat LGN is first measurable only at day E59, well after the start of segregation. Second, although corticogeniculate synapses are present in the newborn cat, they are not very well defined (Shatz & Kirkwood, 1984), and it is thought that in most mammals (excluding ferrets) the corticogeniculate pathway reaches its target several days later than the geniculocortical pathway (Johnson & Casagrande, 1993).

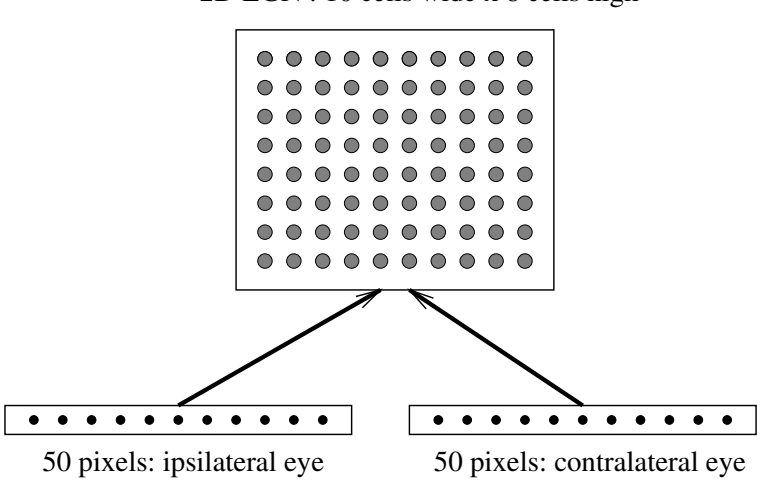


Figure 4.1: Architecture of the Keesing network. Each retina is one-dimensional and fully connects to the two-dimensional LGN. This represents a coronal slice of the LGN rather than the whole volume of the LGN. Each retina consists of X_{pre} units (normally 50); the LGN consists of $X_{\text{post}} \times Y_{\text{post}}$ units (normally 10×8).

4.2.1 Weight adaptation mechanisms

The aim of the model is to adjust the set of connection strengths in the network to replicate the pattern of connectivity found in the developing retinogeniculate pathway. These connection strengths (or weights) between presynaptic unit i and postsynaptic unit j are represented by the value w_{ij}. Postsynaptic units are assigned a unique index j as shown in Figure 4.2. Several mechanisms are used to adapt the strength of these weights during development:

1. Postsynaptic activations: $y_j = \sum_{i=1}^{N_{\text{pre}}} w_{ij} x_i$

   The activity of postsynaptic units is the dot product of the input vector and the unit's weight vector.

2. Covariance rule: $\Delta w_{ij} = \varepsilon(x_i - \alpha)(y_j - \beta)$

   The covariance rule uses constants (α, β) rather than long-term averages of unit activations as used in the original formulation of the rule (Sejnowski, 1977).

3. Normalisation:

$$w'_{ij} = \frac{w_{ij} \cdot T_{\text{pre}}}{\sum_{k=1}^{N_{\text{post}}} w_{ik}} \qquad \text{Presynaptic normalisation}$$

$$w'_{ij} = w_{ij} - \frac{r_s}{N_{\text{pre}}}\left(T_{\text{post}} - \sum_{k=1}^{N_{\text{pre}}} w_{kj}\right) \qquad \text{Postsynaptic normalisation}$$

In the Keesing model, the presynaptic normalisation is divisive and the postsynaptic normalisation subtractive. The values T_{pre} and T_{post} are the target values for presynaptic and postsynaptic normalisation respectively. For the subtractive normalisation, the parameter r_s controls the rate at which subtractive normalisation is enforced. Unless stated otherwise, this is set to 1.0 to ensure that the sum of weights after normalisation is exactly the target value. Smaller values of r_s can be used to push the weight vector towards (rather than directly onto) the constraint surface.

4. Growth rule: $\Delta w_{ij} = \gamma \sum_{k \in \text{neigh}(j)} w_{ik}$

   The growth rule encourages a retinal unit to increase the synaptic strength to neighbouring LGN units. The set of neighbours of a unit, neigh(j), consists of those units within a certain geometrical distance of the unit j. To avoid border effects, units at the left and right edges of the LGN are considered to be neighbours using wrap-around.

   The growth rule is used probabilistically after every iteration with a fixed probability p_g. The size of the neighbourhood decreases over time. It is assumed here that the radius of the neighbourhood, r, decreases by one unit every t_r epochs until the radius is zero, after which point the growth rule is not used for the remainder of the simulation.

5. Weight bounds: $w_{ij} \geq 0$

   All weights are bounded to be positive. If any of the rules modify the weights such that the weight will become negative, the weight is instead set to zero strength (and can be further modified). Unless explicitly stated otherwise, there is no upper bound placed on individual weights.

4.3 Retinal inputs

In the description of the original model, retinal inputs were assumed to be Gaussian functions of position, although no details were given on the temporal properties of the waves. In the earliest experiments (not reported here), it was assumed that retinal inputs were always anticorrelated (Eglen, 1995). Given that the retinal waves are independently generated within each eye, the anticorrelated waves were replaced with uncorrelated waves generated by the following procedure:

At the start of development, both eyes are quiet. At each time step, if an eye is quiet (no wave present), a new wave is initiated in the eye with probability p_w. Wavefronts are initiated at either the left- or right-hand side of the retina, and the wavefront moves by one pixel per time step to the other side of the retina. Once the wave reaches the other end of the retina (T time steps after the wave has started), the wave disappears and the retina goes into a refractory period for R time steps. After this refractory period, a new wave may then be initiated.

For these experiments, the wave is modelled as a Gaussian function of position (of standard deviation σ_w) from the wavefront centre c:

$$x_i = \exp(\frac{i-c}{2\sigma_w^2})$$

4.4 Results

In the following sections, we present the results of using the Keesing model under various different conditions. To provide a fair comparison between the different simulations, unless stated otherwise, all parameter values were kept the same. A list of the model parameters, along with their meaning and typical values, are given in Table 4.1.

Parameter	Meaning	Typical value
p_w	Probability of a new wave starting	0.02
σ_w	Standard deviation of the wave	0.85 – 1.00
R	Refractory period between waves	1
T	Number of time steps wave is present on retina	50 (X_{pre})
X_{pre}	Width of each retina	50
N_{pre}	Number of presynaptic units	100 ($X_{\text{pre}} \times 2$ eyes)
X_{post}	Width of the LGN	10
Y_{post}	Height of the LGN	8
N_{post}	Number of postsynaptic units	80 ($X_{\text{post}} \times Y_{\text{post}}$)
T_{pre}	Target sum for weights from each presynaptic unit	1.0
T_{post}	Target sum for weights to each postsynaptic unit	1.25 †
r_s	Enforcement rate of subtractive normalisation	1.0
ε	Rate of weight update in covariance rule	0.001
α	Presynaptic threshold in covariance rule	0.1
β	Postsynaptic threshold in covariance rule	0.0125
γ	Constant for growth term	0.1
r	Radius of growth rule	[2,1,0]
p_g	Probability of using growth rule	0.01
t_r	Number of epochs between radius decreasing by one	200

† There is a relationship between the target sums $T_{\text{pre}}$and T_{post}, since the total amount of weight strength in the network must be the same using either presynaptic or postsynaptic normalisation: $N_{\text{pre}} \times T_{\text{pre}} = N_{\text{post}} \times T_{\text{post}}$

Table 4.1: Parameters and typical values used in the Keesing network simulations.

4.4.1 Visualisation of results

Several different types of plots are used in the following sections to visualise network development. Here are brief descriptions of each type of plot:

- The raw weight matrix.

  The weight matrix is displayed as a raster image. One row of the image corresponds to all of the weights for a postsynaptic unit. One column of the image corresponds to all of the weights from a presynaptic unit. This is illustrated in Figure 4.2. The intensity of each pixel in the image is proportional to the magnitude of the corresponding weight. To enable the full range of different magnitudes of weights to be displayed in different matrix plots, the brightness of each pixel is scaled according to the minimum and maximum values of weights for each plot. Although this prevents a direct comparison of images from different

networks, in practice the range of weights in most experiments is roughly constant. (The two notable exceptions to this are shown in Figure 4.12, in the absence of any normalisation, and Figure 5.4, when the probability of waves being generated is very small.)

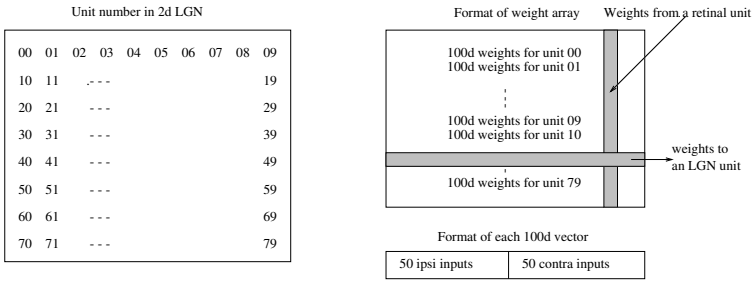


Figure 4.2: Structure of the weight matrix. Each LGN unit is numbered as shown, from 0 to 79. The right-hand figure shows how the weights are displayed: row n of the weight matrix corresponds to the 100-dimensional vector for LGN unit n. Each weight vector is subdivided into the 50 weights for the ipsilateral eye and 50 weights for the contralateral eye.

- Ocular dominance plots.

  Hinton diagrams are used to indicate the degree to which each postsynaptic unit receives input from the left and right eye. Each postsynaptic unit within the two-dimensional postsynaptic sheet is represented by a box. The size of the box represents the degree of monocularity, and the colour of the box represents the dominant eye.

$$t_j^{\text{left}} = \sum_{i=1}^{X_{\text{pre}}} w_{ij} \qquad t_j^{\text{right}} = \sum_{i=X_{\text{pre}}+1}^{2X_{\text{pre}}} w_{ij}$$

$$z_j = \frac{t_j^{\text{left}}}{t_j^{\text{left}} + t_j^{\text{right}}} - \frac{1}{2}$$

  If z_j is greater than zero, most of the input for postsynaptic unit j comes from the left eye, and so the box is coloured black. Otherwise, most of the input comes from the right eye, and the box is coloured white. The size of the box is proportional to the absolute value of z_j, scaled such that the maximum box corresponds to $z_j = \pm 0.5$, which indicates the unit receives all of its input from one eye. If the sum of weight strengths to a postsynaptic unit is less than a certain value (0.005 for the simulations presented here), the postsynaptic unit is defined as "dead", and is represented by a grey circle in the Hinton diagram (as shown for example in Figure 4.9).

- Topographic plots.

  The topographic plots show the centre of mass and the standard deviation of the weight vector for each postsynaptic unit. This assumes that the weight vector has a unimodal distribution. This is obviously not the case when the postsynaptic unit is binocular, but during development, as units become more dominant towards one eye, weight vectors become more unimodal. This progressive development towards a unimodal distribution is illustrated for a typical weight vector in Figure 4.3.

To reduce the effect of the non-dominant eye on the centre of mass measurements, we take the centre of mass and standard deviation for the subset of the weight vector from the eye providing the dominant input to the unit (given by z_j). The centre of mass, $\overline{x_j}$, and the standard deviation, s_j, for each postsynaptic unit are given by:

$$s_j = \left(\frac{1}{m_j} \sum_{i=1}^{X_{\text{pre}}} ((i - \overline{x_j})^2 w'_{ij}) \right)^{0.5}$$

$$\text{where} \qquad m_j = \sum_{i=1}^{X_{\text{pre}}} w'_{ij} \quad \text{and} \quad \overline{x_j} = \frac{1}{m_j} \sum_{i=1}^{X_{\text{pre}}} i \, w'_{ij}$$

$$w'_{ij} = \begin{cases} w_{ij} & \text{if } z_j \geq 0 \\ w_{(i+X_{\text{pre}})j} & \text{if } z_j < 0 \end{cases}$$

The dummy variable w'_{ij} for the weights allows us to sum over weights from either the left or the right eye, depending on the value of z_j.

Figure 4.3 shows the development of one postsynaptic unit which demonstrates the usefulness of the centre of mass and standard deviation measures. To allow easy comparison of the topographic plots with the raw weight matrices, the postsynaptic unit number is represented on the Y-axis (increasing in value down the page) and the centre of mass is represented along the X-axis (increasing to the right of the page). Solid and dashed lines are used in the topographic plot to indicate that the postsynaptic unit is dominant for the left or right eye respectively.

The topographic plots are not ideal for the Keesing network, due to the periodic boundary conditions imposed on the X dimension of the network. Postsynaptic units that respond to presynaptic units at the edges of the presynaptic sheets produce artificially high standard deviation values. Despite this limitation, the plots are still useful and are therefore included.

- Projection column plots.

  Projection column plots show how the centre of mass of the receptive field of postsynaptic units varies when moving vertically down through the layers of the LGN. (These plots are similar to those used by Sanderson (1971a) to show how the visual field maps onto the LGN.) The centre of mass of each postsynaptic unit is plotted for each unit. Lines are then drawn between points of neighbouring postsynaptic units that are within the same column of the LGN as long as the postsynaptic units receive most of their input from the same eye. These plots are drawn using the same axes as the topographic plots (postsynaptic unit along the Y-axis and centre of mass along the X-axis).

- Projective field plots.

  Most of the visualisation methods focus on showing the properties of postsynaptic units. Projective field plots focus on the presynaptic units instead by showing how one presynaptic unit connects to all of the postsynaptic units. The weight vector is plotted with the index of the postsynaptic unit on the X-axis and the magnitude of each weight on the Y-axis.

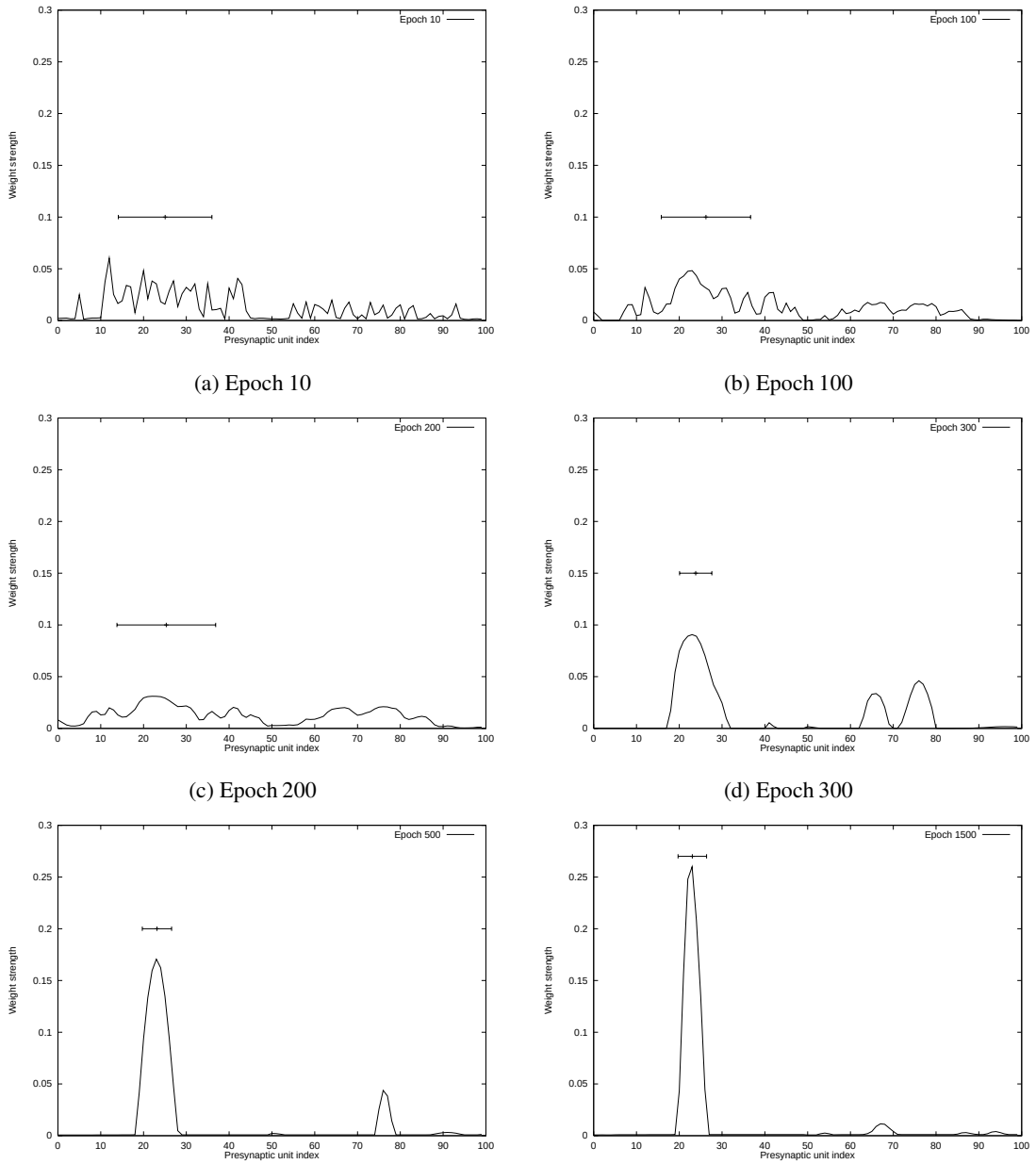
In addition to these plots, which show the state of the network at one particular time during development, two types of plot are used to summarise how the network develops over the time course of the simulation:

- Refinement of receptive field size.

  The standard deviation of the centre of mass for a postsynaptic unit's weight vector can be used as a rough indicator of the width of the unit's receptive field. These widths can be averaged over all postsynaptic units to produce a "mean receptive field width", which is plotted at various stages of development. Error bars for each point indicate ± 1.0 standard deviation of the receptive field width. An example is given in Figure 4.5(a).

- Development of monocularity.

  This plot shows the average value of the monocularity index z_j for all postsynaptic units at various points during development for both the left ($z_j > 0$) and right ($z_j < 0$) eyes. Error bars for each point indicate ± 1.0 standard deviation of the z_j values. An example is given in Figure 4.5(b).



(a) Epoch 10

(b) Epoch 100

(c) Epoch 200

(d) Epoch 300

(e) Epoch 500

(f) Epoch 1500

Figure 4.3: The refinement of receptive field width for a single postsynaptic unit. Plots (a) to (f) show the weights for one postsynaptic unit at different times during development. The error bar above the plot shows the centre of mass of the weight vector for the dominant eye ± 1.0 standard deviation. As time progresses, it can be seen that the receptive field refines to a small region of the ipsilateral eye.

4.5 The role of initial weight biases in development

In this section, we consider the effects of using various amounts of bias in the initial weights of the network to see how development is affected. Providing an initial bias in the network's weights is a common technique to prevent the network getting stuck in local minima and to provide a global orientation for retinotopic maps (Willshaw & von der Malsburg, 1976). In this section we present the results of three sets of experiments, gradually increasing the amount of initial weight bias in successive experiments:

1. No initial weight bias.
2. Earlier arrival of contralateral axons.
3. Earlier contralateral arrival plus retinotopic bias.

4.5.1 No initial weight bias

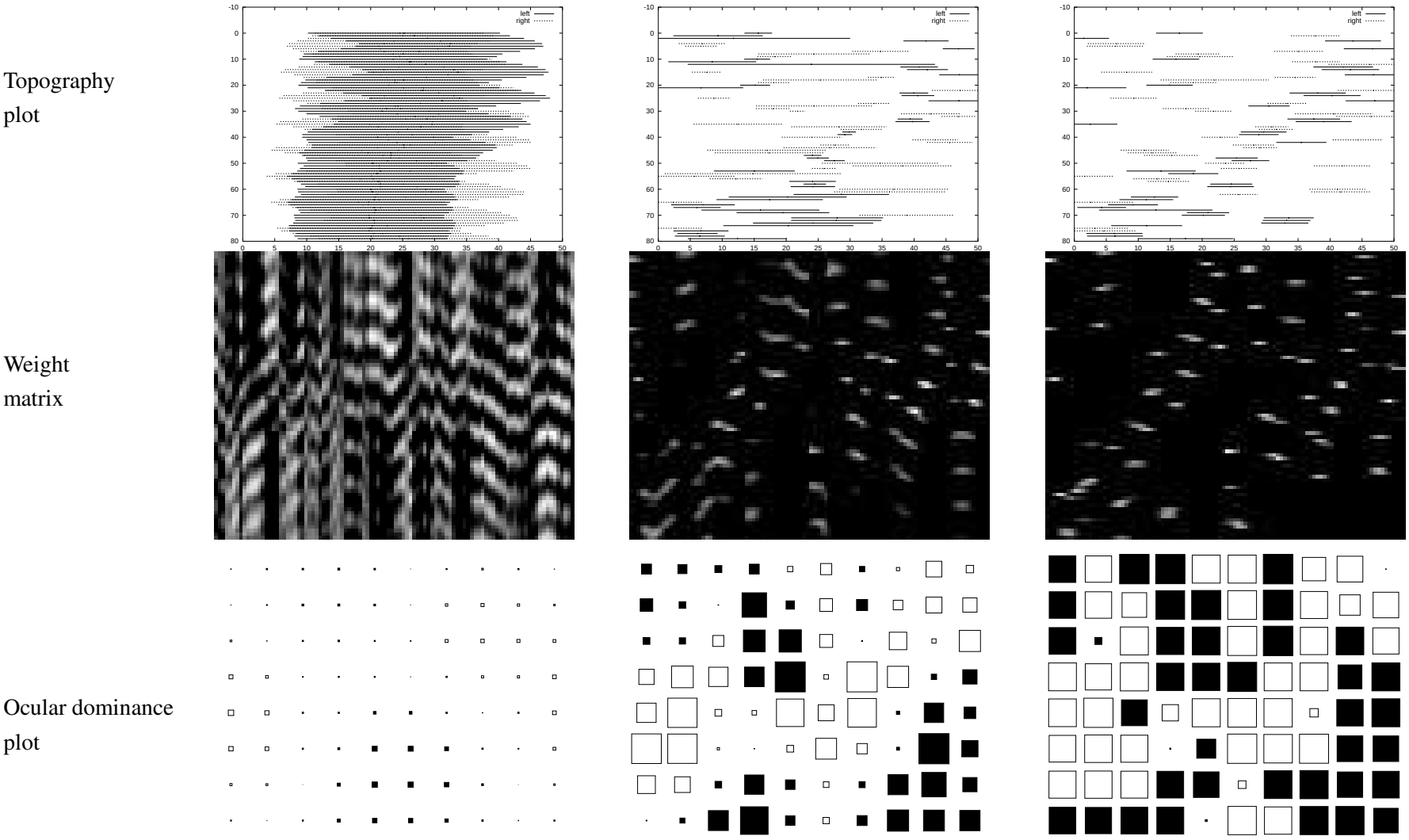
In the absence of any bias in the initial random weights, the modification rules have limited success in refining the weights to mirror the development of the retinogeniculate pathway. Figure 4.4 shows the network at various times during development. In the early stages, the postsynaptic units have wide receptive fields with little dominance of one eye over the other. As development continues, the receptive fields refine in two ways. First, as shown by the topographic plots, the unit receives input from a smaller part of the retina. This decrease in receptive field during development is summarised in Figure 4.5(a). Second, from the ocular dominance plots, the units also become more monocular, responding much stronger to one eye. The progressive development of monocularity is fairly smooth throughout the time course of the simulation, as shown in Figure 4.5(b)

By the end of development, most postsynaptic units have small, monocular receptive fields. However, there is little global organisation. Although nearest neighbour units have similar receptive field positions (two or three neighbouring units cluster together in the topographic plot), there is no global ordering of the receptive fields. This is shown by the many large jumps in receptive field position for neighbouring postsynaptic units in the topographic plots. It is also clearly seen in the projection columns plot shown in Figure 4.5(c): individual columns should be restricted to a small vertical region of the plot, but instead each column occupies a wide range of the input space.

As well as the absence of global topographic order, the final overall pattern of ocular dominance is very different to that found in the retinogeniculate pathway. The network has failed to fully segregate postsynaptic units into two large groups according to the eye of origin. Instead, the pattern of monocular dominance resembles the stripes found in the cortical areas, with a small group of neighbouring units responsive to one eye and then another group responsive to the other eye. Most of the units however have become strongly monocular, with only six units in the simulation presented in Figure 4.4 remaining binocular. All of these six units are surrounded by left- and right-eye dominant units. This is similar to other models of ocular dominance where binocular units are only found at the borders of stripes between alternating eyes (Goodhill, 1992).

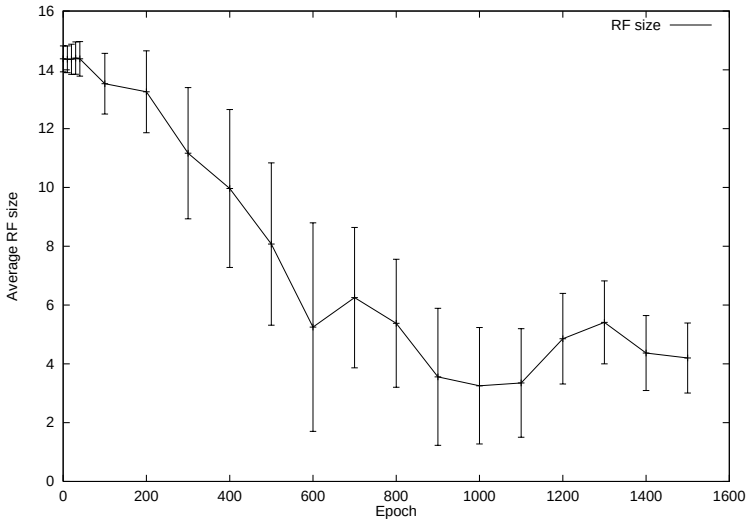
It is interesting to compare these results with the innervation pattern seen when ferret retinal inputs are rerouted into the medial geniculate nucleus (the thalamic region which normally receives auditory input) (see Figure 10b of Roe, Garraghty, Esguerra, & Sur, 1993). The rerouted

ferret retinal inputs cluster into patches rather than layers. This evidence can be interpreted as suggesting that extra information must be provided for the formation of layers, and it is likely that this information is specified by activity independent mechanisms (Sur, 1995; Angelucci, Clasca, Bricolo, Cramer, & Sur, 1997). This is investigated in the next set of experiments.

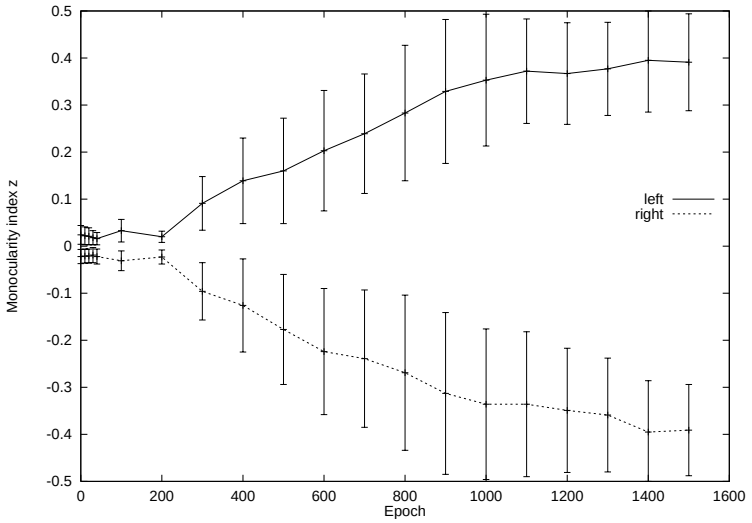


(a) Epoch 100 (b) Epoch 600 (c) Epoch 1500

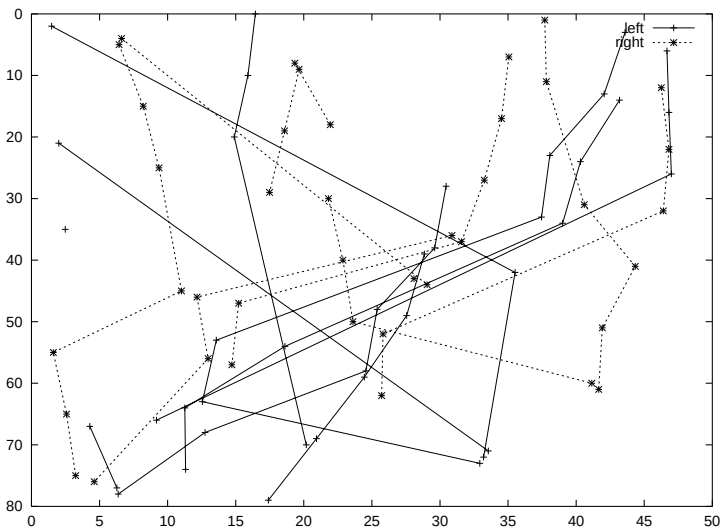
Figure 4.4: Network development with no initial weight bias – 1: topography and ocularity. The state of the network is shown after (a) 100 epochs, (b) 600 epochs and (c) 1500 epochs.



(a) Refinement of receptive field width



(b) Development of monocularity



(c) Projection columns

Figure 4.5: Network development with no initial weight bias – 2: refinement of receptive fields, monocularity and final projection columns.

4.5.2 Earlier arrival of contralateral axons

Contralateral axons innervate the cat LGN a few days earlier than the ipsilateral axons (Shatz, 1983). This gives the contralateral inputs a competitive advantage over the ipsilateral inputs for innervating further into the LGN. In these simulations, it is always (arbitrarily) assumed that the right eye corresponds to the contralateral eye.

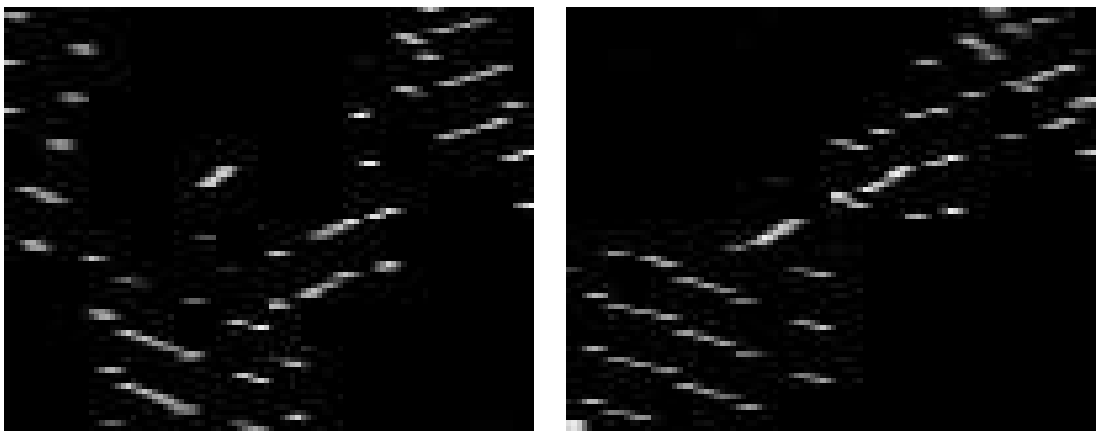
The competitive advantage for the contralateral eye is modelled by biasing the initial weights of the network so that the contralateral inputs initially innervates units higher in the LGN than the ipsilateral eye. For example, the initial weights between units of the left eye and postsynaptic units in the bottom two rows of the LGN are set to random values, with the remaining weights from the left eye initially set to zero. The weights from the right eye are initialised in a similar fashion, so that connections between the right eye and units in the bottom four layers of the LGN are set to random values, and the rest set to zero. With this initial setup, units in the bottom two rows (rows seven and eight) of the LGN initially receive input from both left and right eyes, and so are binocular units. Units in rows five and six of the LGN are initially monocular since they receive input from only the right eye. All of the units in the top four rows are initially dead since they receive no input. This is illustrated in Figure 4.9.

With this bias in favour of the contralateral eye, the network develops in a more ordered way in comparison to the networks without any initial weight bias. Figure 4.6 and Figure 4.7 show the development of topography and ocularity in this network. For this set of experiments, two slightly different networks were compared. The first network used both pre- and postsynaptic normalisation, whereas the second network used only presynaptic normalisation. This allowed us to examine network development under different normalisation conditions.

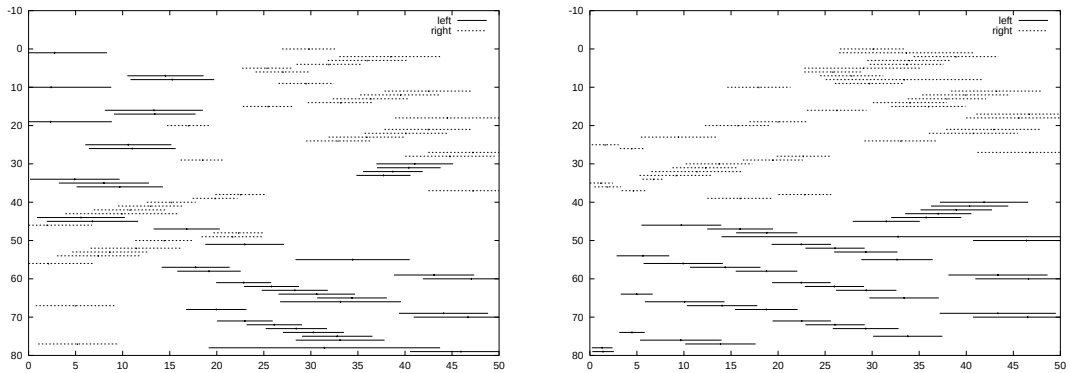
For both networks, the size of the receptive fields refined gradually throughout development. The final topographic mapping however was still only ordered on a local scale, rather than a global scale, as shown in the topography plots in Figure 4.6. In both networks, the topographic mapping is slightly more ordered for the left eye than for the right eye. The effect of the normalisation with respect to topography is most clearly seen in the projection column plots in Figure 4.7. Although some projection columns still vary over the majority of the input space, the projection columns in the network with just presynaptic normalisation tend to be more ordered than in the network with both pre- and postsynaptic normalisation.

The strongest effects of the normalisation in the network however are seen when looking at the development of monocularity in the network. First, the postsynaptic units reached a high level of monocularity much sooner during development when using just presynaptic normalisation than with normalisation at both sites. Second, the network with only presynaptic normalisation formed the desired pattern of ocular dominance: segregation into two eye-specific layers, each layer occupying four consecutive rows of the LGN. All units in this network are strongly monocular. In comparison, the network with both pre- and postsynaptic normalisation formed ocular dominance stripes. (Although the ocular dominance pattern is stripy, there is some overall layering in evidence however: most of the units responsive to the left eye are found in the bottom four rows of the LGN, and vice-versa for the right eye.) The network with both pre- and postsynaptic normalisation also has several binocular units, again at the borders of regions responding to different eyes.

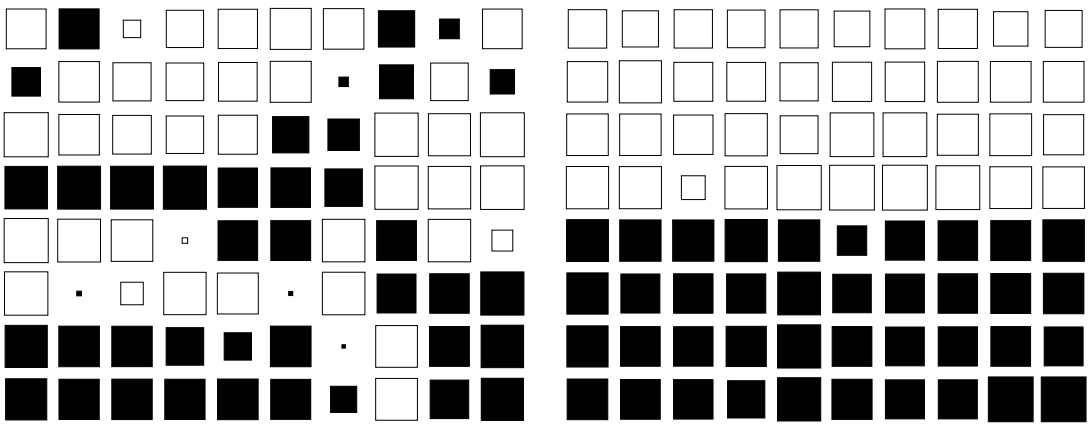
Pre+Post normalisation | Pre normalisation only



(a) Raw weight matrix



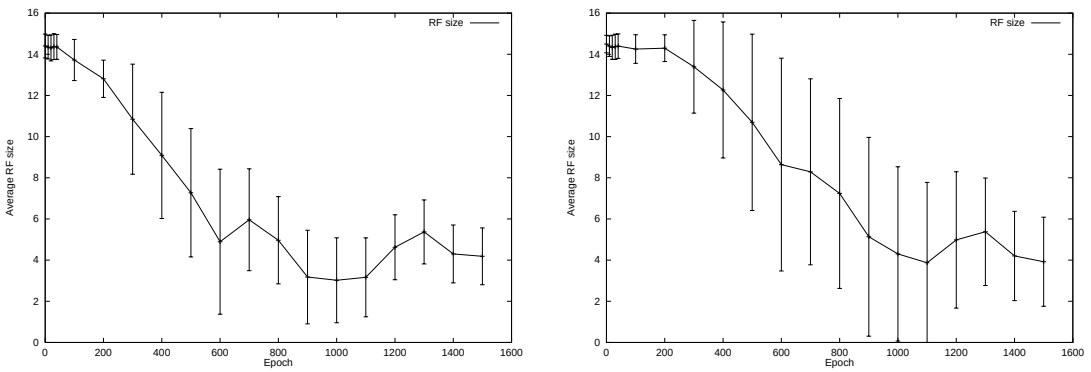
(b) Topography plots



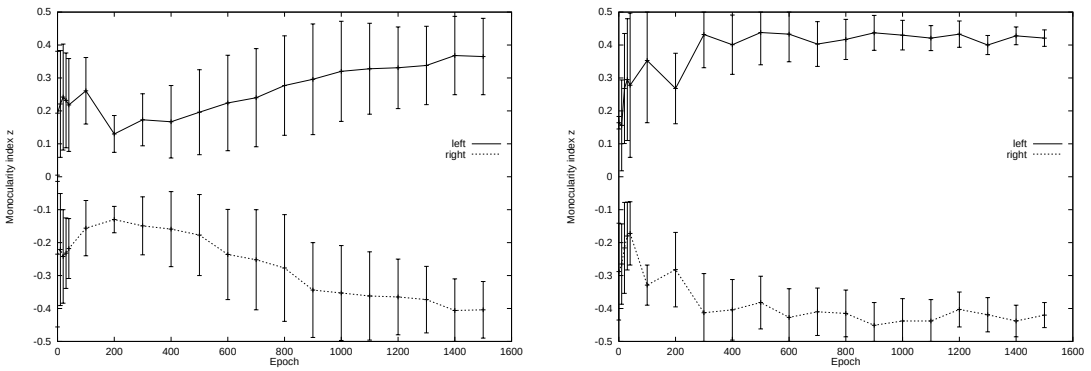
(c) Ocularity plots

Figure 4.6: The effect of contralateral bias in initial weights – 1: topography and ocularity. The left column shows the results of development using both presynaptic and postsynaptic normalisation; the right column shows the effect of using just presynaptic normalisation.

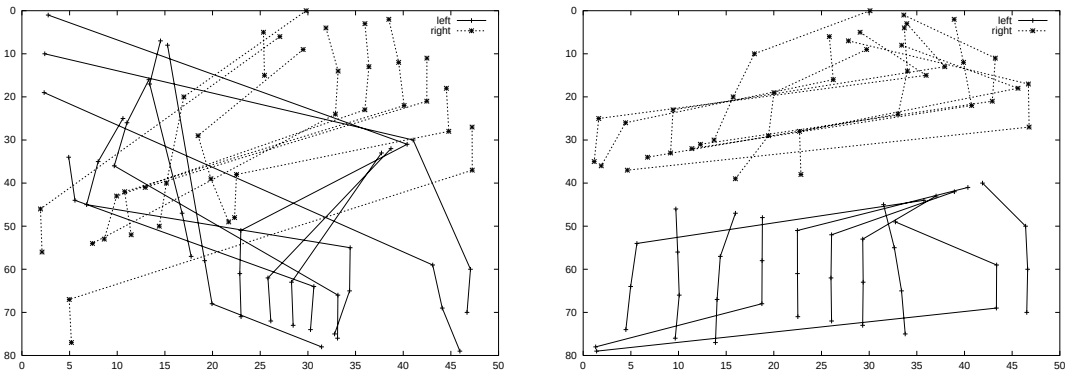
Pre+Post normalisation Pre normalisation only



(a) Refinement of receptive field width



(b) Development of monocularity



(c) Projection columns

Figure 4.7: The effect of contralateral bias in initial weights – 2: refinement of receptive fields, monocularity and final projection columns. The left column shows the results of development using both presynaptic and postsynaptic normalisation; the right column shows the effect of using just presynaptic normalisation.

4.5.3 Earlier arrival of contralateral axons and topographic bias

As well as an imposing a bias in the weights to favour the contralateral inputs going to the top rows of the LGN, we can also impose a topographic bias to ensure that the left end of each retina connects to the left end of the geniculate. This bias can be implemented by assuming that postsynaptic units initially receive input from some subset of all possible presynaptic units. This topographic bias is justified on two grounds. First, some form of topographic bias must be imposed in the retinogeniculate pathway since the mapping of visual space into the geniculate is consistent across animals. If there were no extra mechanism controlling topographic map formation, the mapping of elevation and azimuth into the LGN could we swapped in different ways (such as the superior visual field mapping into the anterior region of the LGN rather than the usual posterior region as shown in Figure 2.3). The second justification for topographic bias is that it is highly unlikely that an immature retinal axon could extend to branch throughout the entire LGN (or for the geniculate dendrite to extend enough to initially receive input from all of the incoming retinal axons). When retinal ganglion axons first innervate the geniculate, they are fairly simple in form and confined to a small region of the LGN, indicating that the axons only contact a few, rather than most of the LGN units (Sretavan & Shatz, 1986). Willshaw and von der Malsburg (1976) also note that rather than allowing each presynaptic unit initially to contact all postsynaptic units, in reality there is probably a more restricted arbor width.

Form of the initial retinotopic bias

Topographic bias is introduced into the initial weights for a set of inputs from one eye to a row of postsynaptic units by setting several consecutive weights to zero. The weight elements set to zero varies systematically across the units within one row of the LGN to provide an overall topographic bias on the initial weights, as shown in Figure 4.8. The degree of initial topographic bias is controlled by a parameter b. This parameter in turn defines two values, o and d (shown in Figure 4.8), which are used to calculate the weight bias:

$$o = \frac{X_{\text{pre}} - b}{2}$$
$$d = \frac{X_{\text{pre}}}{X_{\text{post}}}$$
$$l_x = (x \times d) + o$$

Using these equations, elements $[l_x], [l_{x+1}] \ldots [l_{x+b-1}]$ of the weight vector for the postsynaptic unit in column x (and some fixed row) are set to zero. The notation $[l] = l$ modulo X_{pre} is used to provide the periodic boundary conditions. All other weights are set to random values.

Results using retinotopic bias in initial weights

Figure 4.9 shows the development of topography and ocularity in a network whose initial weights were biased for both ocularity and topography. The initial weights are shown in Figure 4.9(a). The ocularity bias was imposed by the left eye innervating the bottom two rows of the LGN and the right eye innervating the bottom four rows (This experiment will be called L2R4, after the ocularity bias). The topographic bias was created using a value of $b = 20$ for weights from the left eye to units in row seven of the LGN and for weights from the right eye to units in row five of the LGN. Weights for all other rows were not topographically biased. Due to the topographic bias for

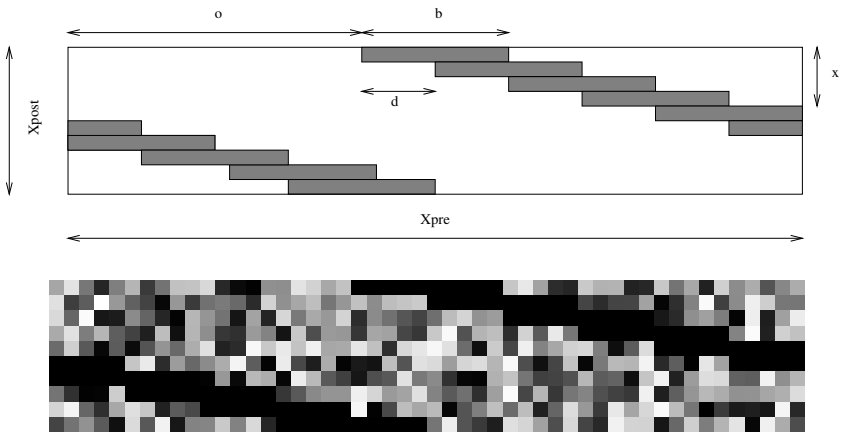


Figure 4.8: The nature of the topographic bias in the initial weights of one row of LGN units. The top figure shows how the initial bias is calculated. Row x of the diagram corresponds to the set of weights from one eye to the postsynaptic unit in column x (and some fixed row) of the LGN. The portion of the weights coloured grey are the elements of the weight vector which are set to zero; all other elements are set to random values. The bottom figure shows an example set of weights biased using this procedure (with $b = 10, X_{\text{pre}} = 50, X_{\text{post}} = 10$).

the left eye, units in row seven of the LGN are slightly dominant for the right eye rather than being equally responsive to both eyes.

Early in development, weights from the right eye tend to move up into the top rows of the LGN, which in turn gives the left eye an advantage in innervating the lower half of the LGN. Hence the layered patterning is in evidence by around 40 epochs, although many units are still binocular, as shown in Figure 4.9(b). The topographic bias has also spread from row seven of the left eye and row five of the right eye into all of the other rows of the LGN, as can be seen most clearly in the weight matrix.

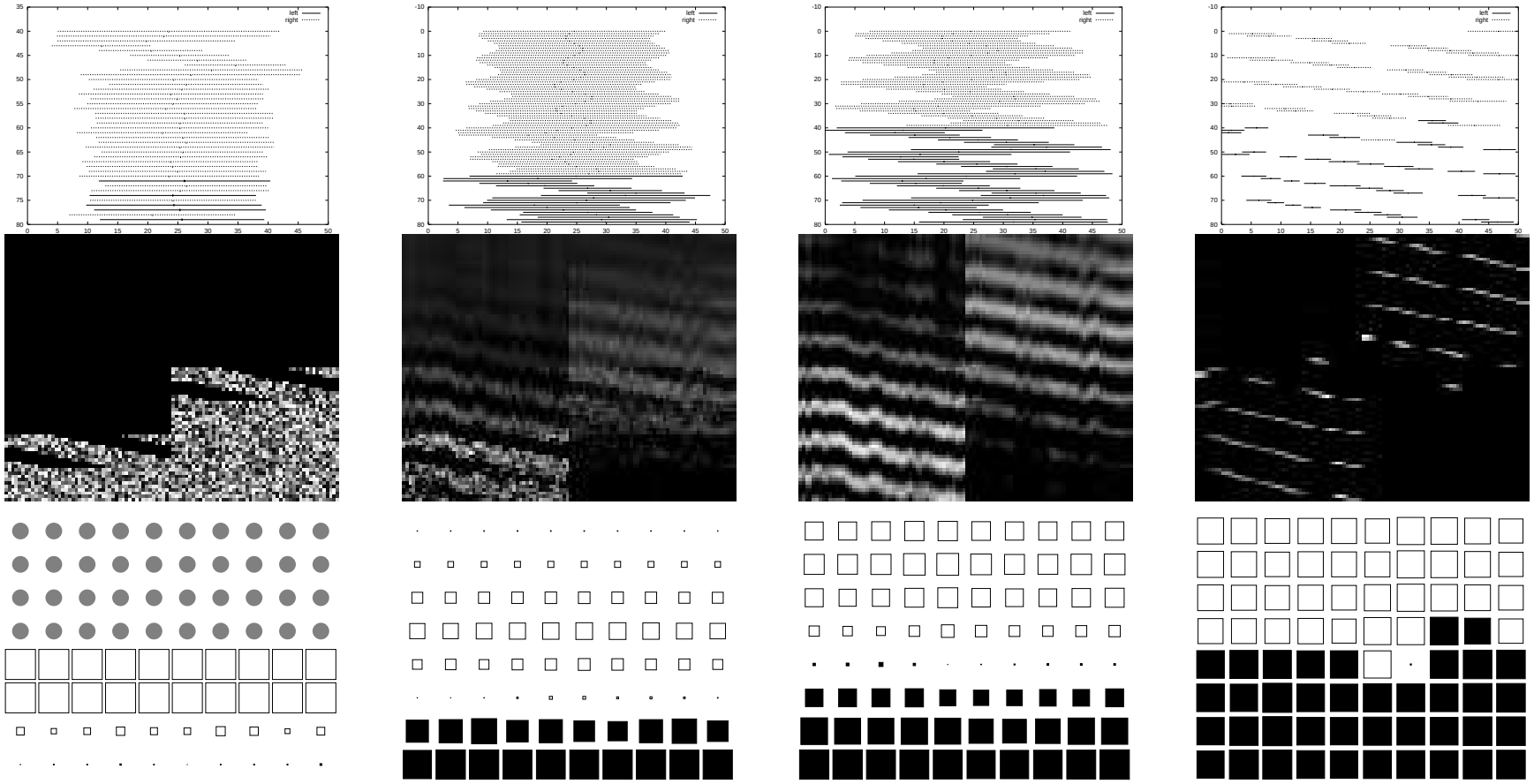
After 100 epochs, the network has settled into two eye-specific layers, each with the same number of postsynaptic units. The only binocular units left are found in the boundary between these two regions (Figure 4.9(c)). At this stage, the network has arranged its global layout, and for the rest of the simulation it then refines the monocularity and receptive field width. Only after the growth rule has finished (after 600 epochs when the radius of the growth rule reaches zero) does the network begin to convert its binocular units at the layer boundaries into monocular units. The receptive fields also gradually refine to receive input from only a small region of the retina.

At the end of development, the network has (almost) produced the desired patterns of ocular dominance and topography. The network has settled into two equal-sized layers of units responding to the two eyes (although at the border between the two layers, four units are wrongly positioned). Each row of the LGN has a smooth retinotopic mapping of visual space. These maps of visual space are aligned throughout the depth of the LGN, as shown by the projection columns of Figure 4.11. (The errant unit in the rightmost projection column for the right eye of Figure 4.11 is due to the periodic boundary conditions in the X dimension of the LGN.)

A similar experiment was performed with slightly different initial weights. This time, the left eye innervated the bottom four rows of the LGN, and the right eye innervated all eight rows of the LGN (This experiment will be called L4R8). To compensate for the increased ocularity

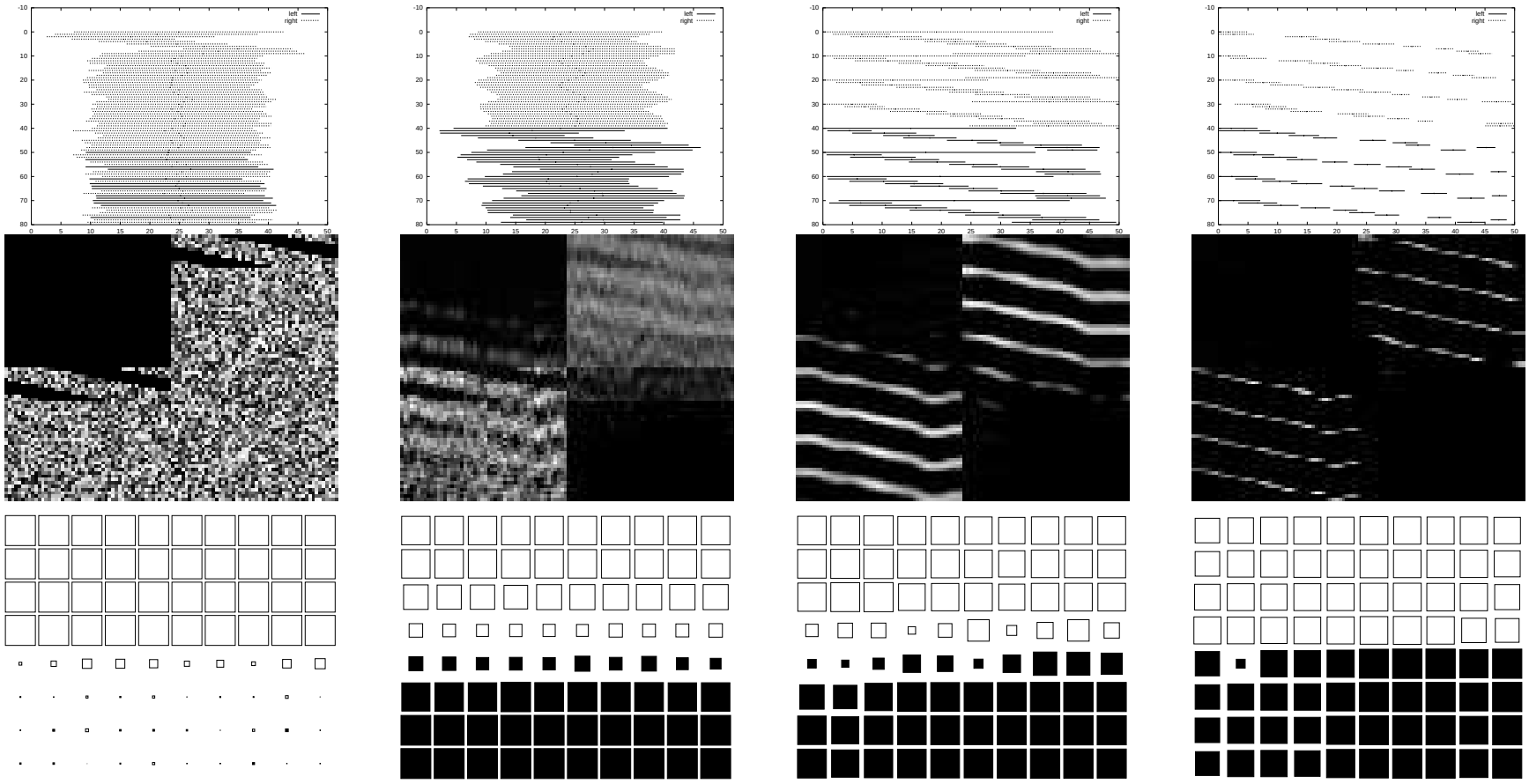
bias, the amount of topographic bias was reduced to $b = 10$. The initial weights, along with subsequent development of the network, are shown in Figure 4.10. Due to the increased ocularity bias, the network has already settled into the two eye-specific layers quite early in development, although the topography is not so well defined in comparison to the previous experiment. The final state of the network is slightly better than in the previous experiment, with a clean segregation of postsynaptic units into two regions of four rows each, smooth topographic mapping in each layer of the LGN, and ordered projection columns.

Figure 4.11 shows the refinement of monocularity and topography over time. Although the L4R8 experiment starts off with more monocular units, the growth rule ensures that in the early stages of development, units are less monocular (explaining the tendency for the average z to initially return towards zero rather than staying at extreme values). This decrease in z may also account for the sudden increase in the average receptive field width for the L4R8 experiment. The overall patterns of development for the two experiments however are pretty similar. The two sources of weight bias can therefore be balanced against each other: if there is less ocularity bias then the topographic bias needs to be increased slightly.



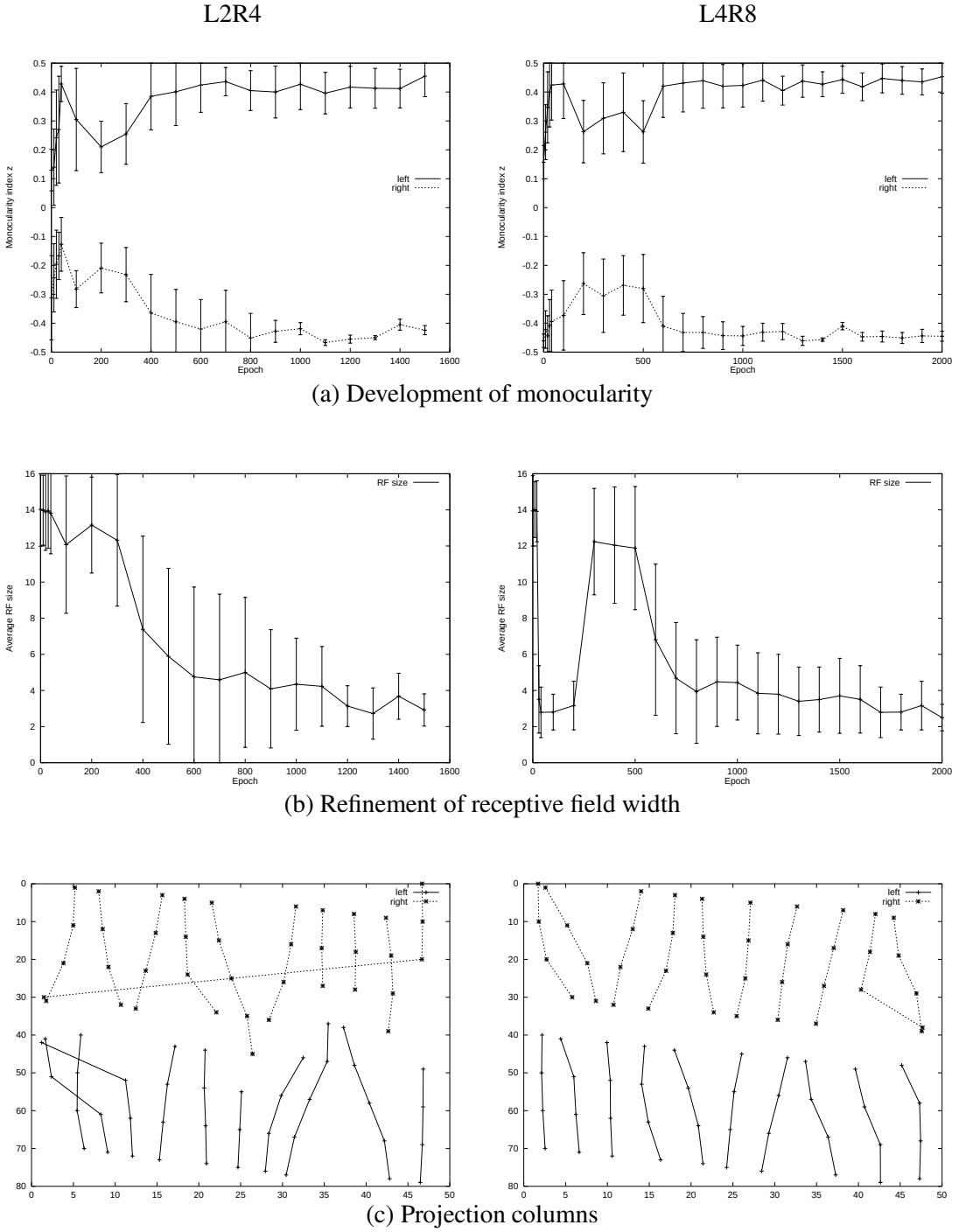
(a) Epoch 0:initial weights (b) Epoch 40 (c) Epoch 100 (d) Epoch 1500

Figure 4.9: The effect of ocular and topographic weight bias on development – 1: left eye innervates two rows, right eye innervates four rows (L2R4).



(a) Epoch 0:initial weights (b) Epoch 40 (c) Epoch 600 (d) Epoch 2000

Figure 4.10: The effect of ocular and topographic weight bias on development – 2: left eye innervates four rows, right eye innervates eight rows (L4R8).



(a) Development of monocularity

(b) Refinement of receptive field width

(c) Projection columns

Figure 4.11: Comparison of the L2R4 and L4R8 experiments with respect to the development of ocularity and receptive field width, and the final projection columns. Results for L2R4 are shown on the left and L4R8 on the right.

4.6 The importance of normalisation

As mentioned in Chapter 3, most previous models of ocular dominance require postsynaptic normalisation of weights to allow ocular dominance to develop. However, the results from Figure 4.6 and Figure 4.7 show that ocular dominance develops within this model either with or without any postsynaptic normalisation. These experiments only investigated network development with ocularity bias in the weights. To address the issue of normalisation more thoroughly, a set of experiments were therefore run to examine the effects of all possible combinations of normalisation techniques.

Normalisation can be applied either divisively or subtractively to both the presynaptic and postsynaptic units. In addition, there is the option to ignore normalisation of either pre- or postsynaptic units. This gives us nine combinations of normalisation techniques. The same network with weights initially biased for ocularity and topography (using the same initial weights as shown in Figure 4.10) was run nine times using each combination of normalisation technique. The results from these nine experiments are summarised in Figures 4.12, 4.13, 4.14 and 4.15.

From these figures, several points can be made concerning the role of the different normalisation techniques upon network development.

4.6.1 The effect of normalisation upon topography

First of all, for the normal pattern of topography to develop, the presynaptic normalisation must be divisive. As long as the presynaptic normalisation is divisive, the form of the postsynaptic normalisation is mostly redundant, as shown by Figures 4.12 and 4.14. Although visually the weight matrices look fairly similar for the three experiments with divisive presynaptic normalisation, there are two subtle differences. First, the projection columns for the left eye are marginally more organised when there is no postsynaptic normalisation than for the other two experiments. Second, in the absence of any postsynaptic normalisation, there is no constraint to ensure that all postsynaptic units should receive the same amount of synaptic strength. Figure 4.16 shows the sum of weights to each postsynaptic unit, which indicates that different postsynaptic units receive different amounts of weight strength. There is a systematic trend for postsynaptic units at the extremes of the network (those in rows one and eight of the LGN corresponding to units 0–9 and 70–79) to receive more weight strength than those units at the centre of the LGN (those in rows four and five, corresponding to units 30–39 and 40–49.)

The retinotopic mapping develops best if the presynaptic normalisation is divisive. When the presynaptic normalisation is subtractive, postsynaptic units still refine their receptive fields, but they fail to develop receptive fields similar to their neighbours (as shown in the middle column of plots of Figures 4.12 and 4.14). The subtractive presynaptic normalisation pushes weights to extreme values so that most weights go to zero, with all of the weight strength for the presynaptic unit allocated into one weight. This is in comparison to the divisive presynaptic normalisation, which can distribute its total weight strength amongst many postsynaptic units, as shown in Figure 4.17.

Finally, in the absence of presynaptic normalisation, most of the weight strength for each postsynaptic unit goes either to one or several neighbouring presynaptic units. This is due to the growth mechanism which encourages neighbouring postsynaptic units to develop similar receptive field properties. As shown in Figure 4.18(c), most of the presynaptic units have zero synaptic strength

and therefore the presynaptic units have disconnected from the postsynaptic sheet. (Any activity in these presynaptic units is therefore not propagated to the postsynaptic sheet.) If the growth rule is removed from the model, all presynaptic units can remain connected to the postsynaptic units in the absence of presynaptic normalisation (Figure 4.18(d)), although this prevents the normal development of topography and ocular dominance (Figures 4.18(a) and 4.18(b)). The size of the receptive field for the experiments without presynaptic normalisation is determined by the postsynaptic normalisation: divisive postsynaptic normalisation produces broader receptive fields than the sharper receptive fields produced by subtractive postsynaptic normalisation (as shown in Figure 4.14).

The experiment with neither presynaptic nor postsynaptic normalisation is included here for completeness, although it is difficult to compare with the other experiments. In this experiment, weights increased without bounds causing most weights to reach very large values. Some form of normalisation is therefore necessary to keep the total weight strength in the network within bounds.

4.6.2 The effect of normalisation upon ocular dominance

The form of normalisation also affects development of ocular dominance within the network. The results are summarised in Figure 4.13. The best results are achieved when the presynaptic normalisation is divisive, regardless of the form of postsynaptic normalisation. In these three cases, the network segregates perfectly into the two eye-specific layers.

When the presynaptic normalisation is subtractive, the results are similar to using divisive presynaptic normalisation, although some postsynaptic units failed to become totally monocular. The biggest difference was found for the experiment using subtractive presynaptic normalisation and no postsynaptic normalisation. In the absence of postsynaptic normalisation, there is no guarantee that every postsynaptic unit will receive input. Since the receptive fields produced by subtractive normalisation are much sharper than divisive normalisation, some postsynaptic units fail to receive any input. This produces a large number of "dead" postsynaptic units in the final network (28 out of 80 in the network shown in Figure 4.13).

In the absence of presynaptic normalisation, since all of the receptive fields have developed to receive input from the same part of the presynaptic sheet, all of the postsynaptic units respond to the same eye (in this case the right eye), with no units responding to the other eye. For these simulations, the initial weight strength for the right eye was stronger than for the left eye (due to the contralateral ocularity bias) which caused the network to develop a preference for the right eye. Changing the initial weight bias would affect the eye that dominates network development, but would not change the pattern of ocular dominance produced.

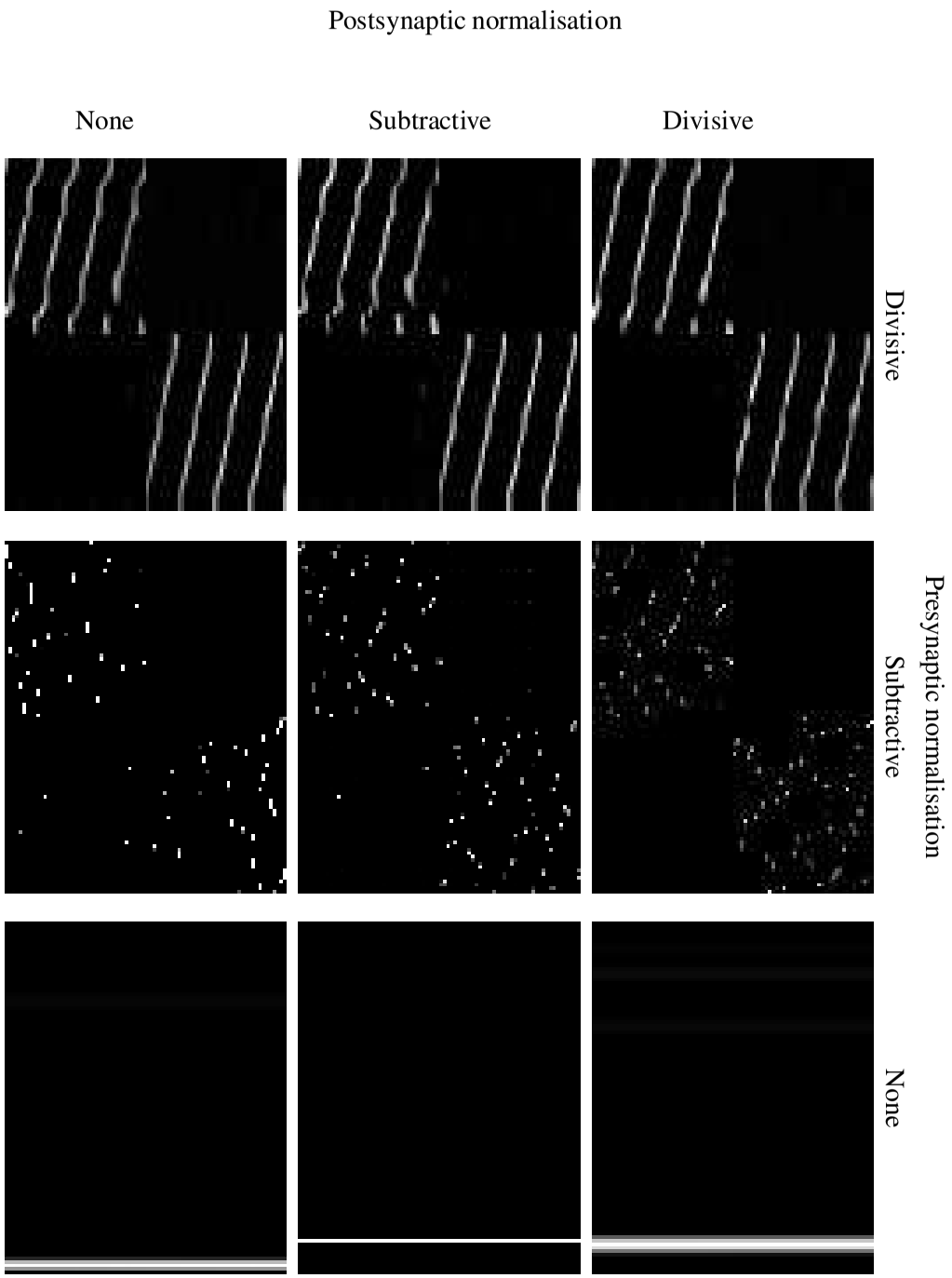


Figure 4.12: The effect of different normalisation methods – 1: weight matrices.

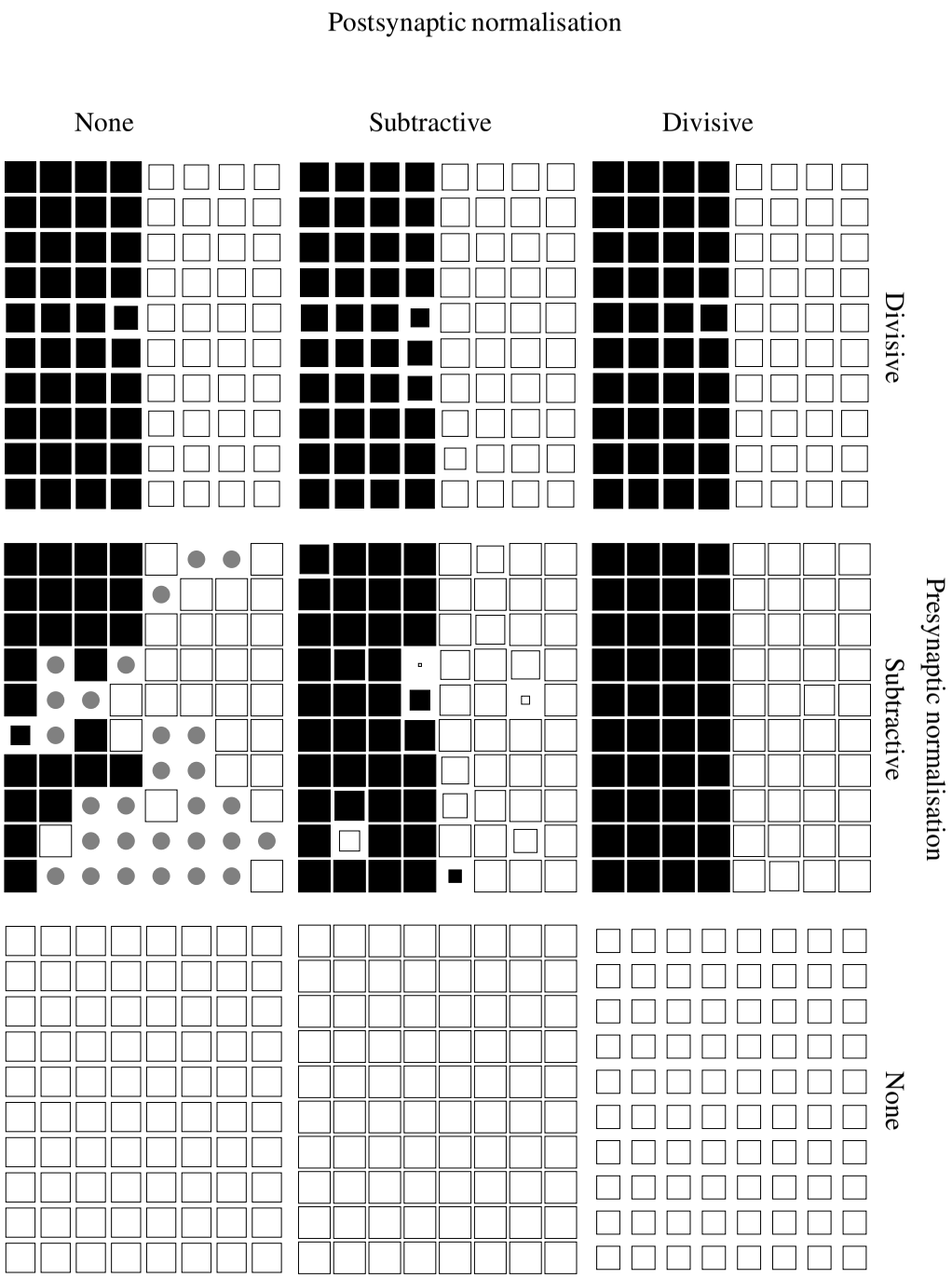


Figure 4.13: The effect of different normalisation methods – 2: ocular dominance plots.

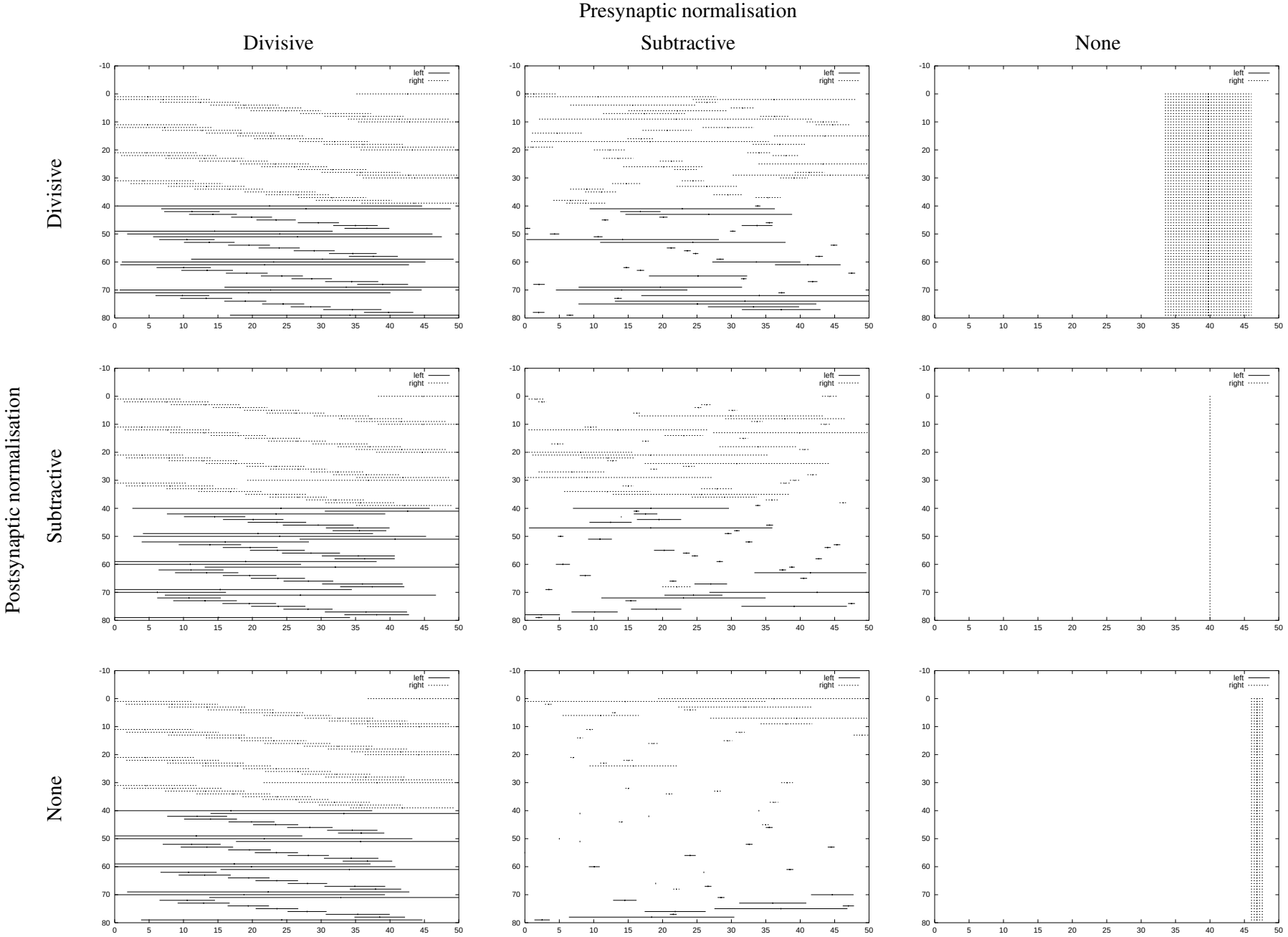


Figure 4.14: The effect of different normalisation methods – 3: topographic plots.

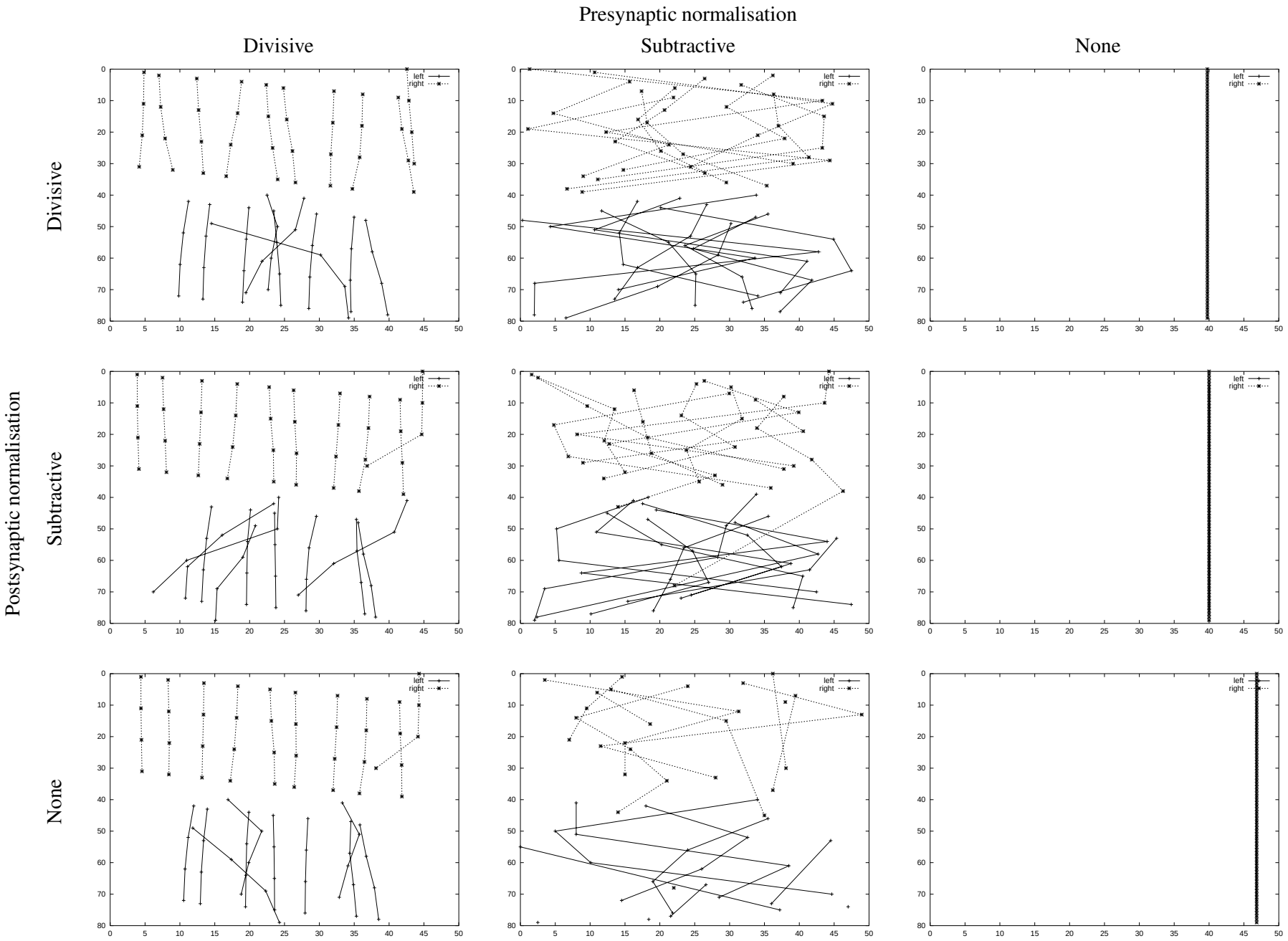


Figure 4.15: The effect of different normalisation methods – 4: projection columns.

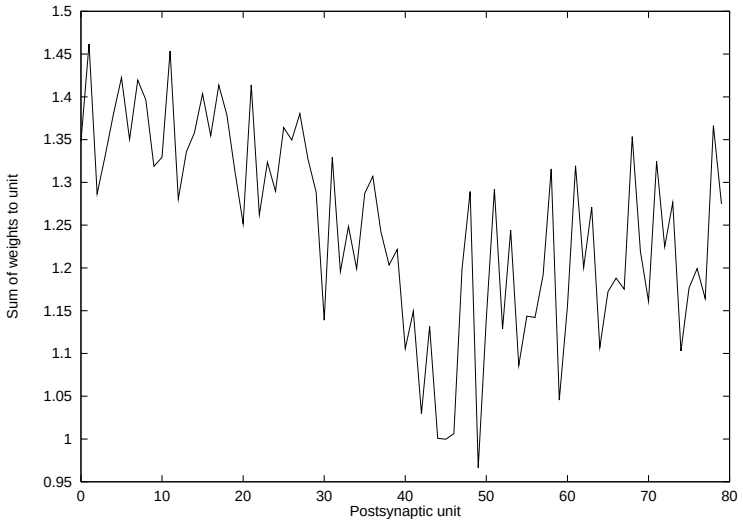


Figure 4.16: Sum of weights to each postsynaptic unit in a network with divisive presynaptic normalisation and no postsynaptic normalisation. For each postsynaptic unit j, the sum $\sum_i^{N_{\text{pre}}} w_{ij}$ is shown.

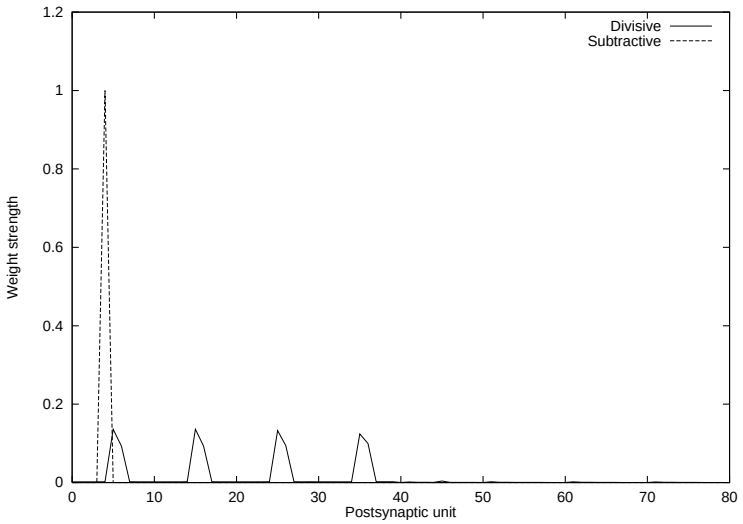
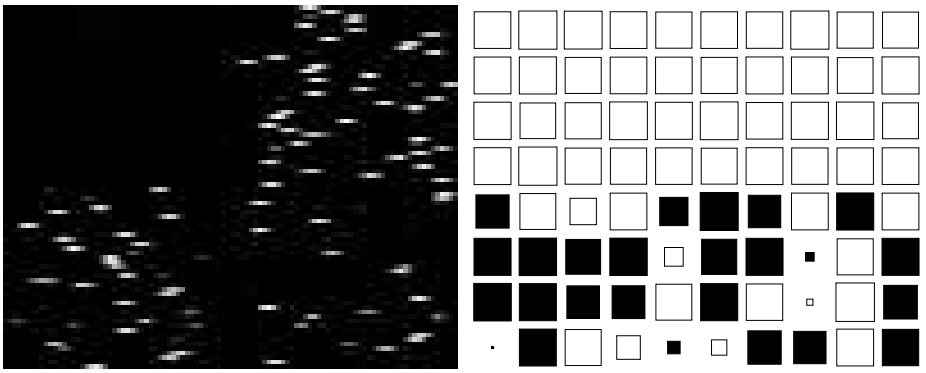
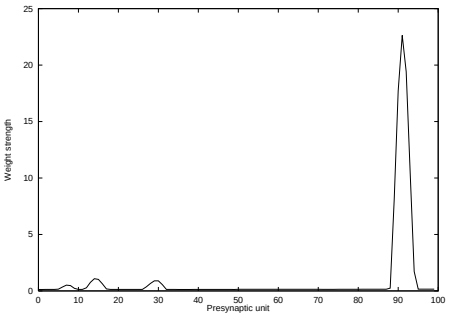


Figure 4.17: Projective field for one presynaptic unit from the right eye using either divisive (solid line) or subtractive (dashed line) presynaptic normalisation. These weights were taken from presynaptic unit 75 in the experiments with no postsynaptic normalisation.

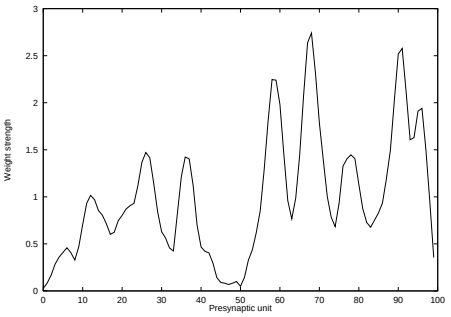


(a) Raw weight matrix

(b) Ocular dominance



(c) Sum of presynaptic connections with growth rule



(d) Sum of presynaptic connections without growth rule

Figure 4.18: The effect of the growth rule upon development in the absence of presynaptic normalisation. (a): Final weight matrix for network with divisive postsynaptic normalisation but no presynaptic normalisation or growth rule. (b): Ocular dominance plot for weights in (a). (c,d): The sum of weights $\sum_j^{N_{\text{post}}} w_{ij}$ for each presynaptic unit. Figure (c) corresponds to the experiment with no presynaptic normalisation and divisive postsynaptic normalisation presented in Figure 4.12. Figure (d) corresponds to the weight matrix shown here in (a).

4.6.3 The importance of the ordering of the normalisation techniques

The previous section has shown that presynaptic normalisation plays a crucial role in development, unlike postsynaptic normalisation. In these experiments, the presynaptic normalisation was always applied before the postsynaptic normalisation. Postsynaptic normalisation may be redundant simply because it is always performed after the presynaptic normalisation. To test for this possibility, a set of experiments were run varying the probability of whether presynaptic or postsynaptic normalisation was applied first.

A new parameter, p_n, was therefore introduced. This controlled the probability of postsynaptic normalisation occurring before presynaptic normalisation. For all of the experiments presented so far in this chapter, this parameter was implicitly set to 0.0 so that divisive presynaptic normalisation was applied first and then subtractive postsynaptic normalisation afterwards. Using the same set of initial conditions as for the previous normalisation experiments, network development was monitored using a range of values for p_n. The results of these experiments are shown in Figure 4.19.

For all values of p_n, the network developed the usual two eye-specific layers, although high values of p_n produced a small number of binocular units and monocular units in the wrong layer. The strongest effect of the p_n parameter was that it affected receptive field size: high values of p_n produced sharper and narrower postsynaptic receptive fields. This sharpening effect of the receptive fields is due to the subtractive normalisation which is sometimes applied first when $p_n > 0.0$. As a consequence of the sharper receptive fields, the topography within each row of the LGN is not so smooth as the value of p_n increases.

4.6.4 Satisfying both forms of normalisation

The previous section has showed that if the subtractive postsynaptic normalisation is applied first, the resulting topographic map is not very smooth, showing discontinuities in each row of the LGN. One possibility for this is simply that if the postsynaptic normalisation is applied first, the presynaptic normalisation can no longer be satisfied afterwards. To quantify how well the two forms of normalisation can be simultaneously satisfied, two error measures were produced:

$$\text{presynaptic error} \qquad E_{\text{pre}} = \left[\frac{1}{N_{\text{pre}}} \sum_{i=1}^{N_{\text{pre}}} \left(T_{\text{pre}} - \sum_{j=1}^{N_{\text{post}}} w_{ij} \right)^2 \right]^{\frac{1}{2}}$$

$$\text{postsynaptic error} \qquad E_{\text{post}} = \left[\frac{1}{N_{\text{post}}} \sum_{j=1}^{N_{\text{post}}} \left(T_{\text{post}} - \sum_{i=1}^{N_{\text{pre}}} w_{ij} \right)^2 \right]^{\frac{1}{2}}$$

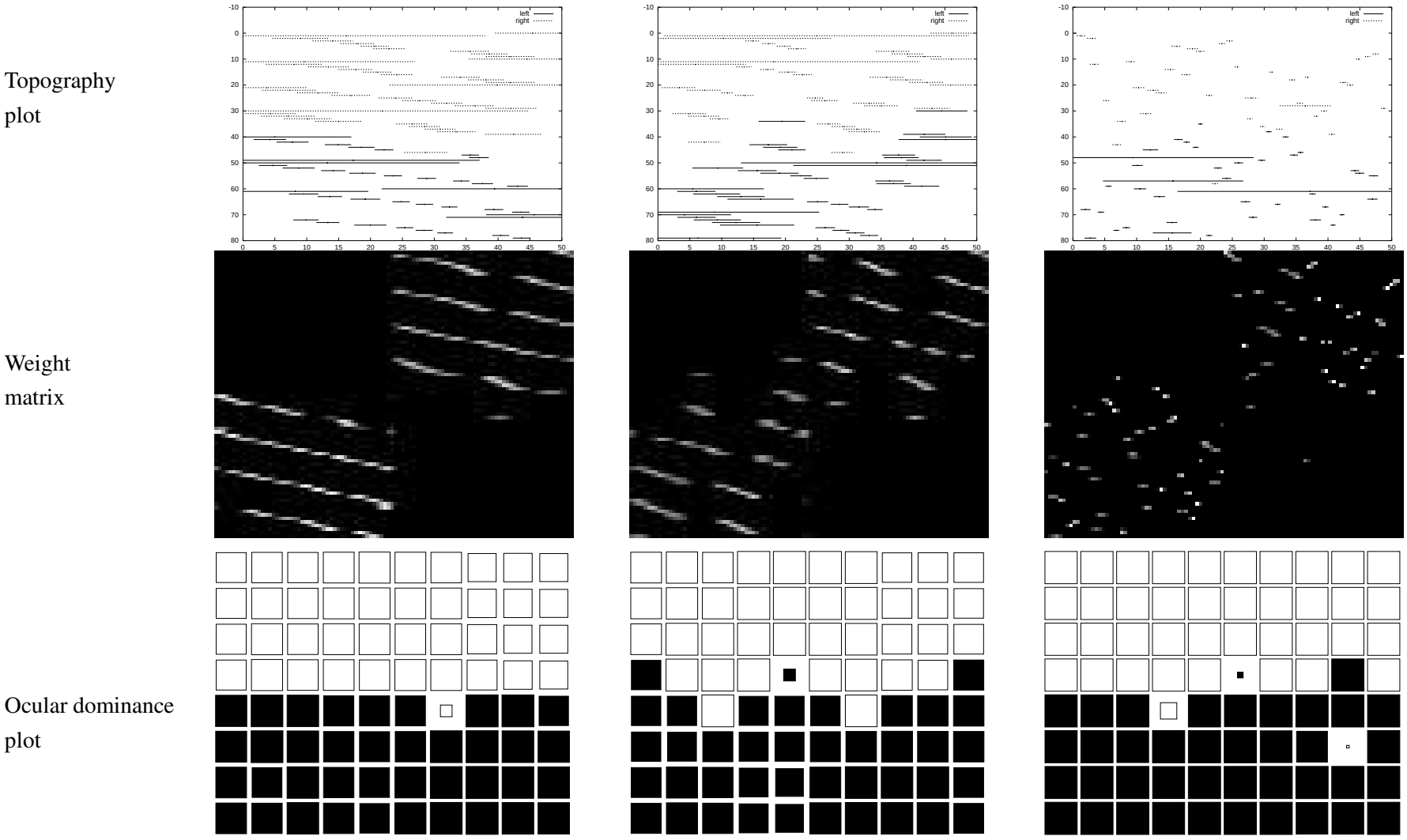
These measures give the root mean squared difference between the target sum and the actual sum of weights. Values smaller than 1e-2 for E_{pre} (or smaller than 1.25e-2 for E_{post}) indicate that the error is less than 1% of the desired target value.

Table 4.2 shows the error values for weights at different times during development for different values of p_n. From this table it can be seen that the difference between the actual normalised values and the target values are quite small for the normalisation applied first. (By definition, the

normalisation applied second will have a very low error.) The errors also tend to decrease during development, presumably because there are less weight changes made in the network. From these error measures, it is reasonable to conclude that both forms of normalisation can be approximately satisfied at the same time, regardless of whether pre- or postsynaptic normalisation is applied first. This result is in agreement with observations from other models using both pre- and postsynaptic normalisation (Miller & MacKay, 1992; Goodhill, 1992).

	Normalisation errors							
	$p_n = 0.00$		$p_n = 0.05$		$p_n = 0.20$		$p_n = 1.00$	
Epoch	E_{pre}	E_{post}	E_{pre}	E_{post}	E_{pre}	E_{post}	E_{pre}	E_{post}
1	1.09e-2	7.39e-6	1.09e-2	7.39e-6	1.09e-2	7.39e-6	1.02e-6	5.70e-1
350	6.03e-4	8.87e-6	1.61e-6	3.31e-3	1.45e-6	5.45e-3	6.00e-7	9.36e-3
700	6.76e-5	7.85e-6	9.91e-4	6.48e-6	1.74e-3	7.34e-6	4.80e-7	2.75e-2

Table 4.2: Normalisation error measures E_{pre} and E_{post} for different values of p_n at different epochs during development. See Figure 4.19 for the corresponding topography and ocular dominance plots. Error values smaller than 1e-6 are the errors remaining after the second normalisation has been applied. These errors ideally should be 0.0, but are slightly non-zero due to rounding errors.



(a) $p_n = 0.05$ (b) $p_n = 0.2$ (c) $p_n = 1.0$

Figure 4.19: Effect of varying the order of presynaptic and postsynaptic normalisation. The probabilities of postsynaptic normalisation occurring first are: (a) $p_n = 0.05$, (b) $p_n = 0.20$, (c) $p_n = 1.00$.

4.6.5 Capping subtractive presynaptic normalisation

The previous experiments have shown the importance of divisive presynaptic normalisation for the development of topography in the LGN. Networks using subtractive presynaptic normalisation are unable to replicate the topographic projections due to the way that the normalisation pushes individual weights to extreme values. Since there is no maximum weight value imposed on weights in the network, subtractive normalisation forces all of the synaptic weight strength into one of the elements of the weight vector, with all other elements pushed to the minimum weight value of zero, as shown in Figure 4.17. If, however, there is a maximum value imposed on each weight (which will be called $w_{\max}$), more than one weight for a unit is forced to take a non-zero value. (This assumes $w_{\max} < T_{\text{pre}}$ for presynaptic normalisation and $w_{\max} < T_{\text{post}}$ for postsynaptic normalisation. The case when $w_{\max}$ is equal or greater than the target value for normalisation (T_{pre}or T_{post}) is uninteresting.)

To test the effect of imposing upper limits on individual weights, an extra weight-capping mechanism was introduced to replace the existing weight bound mechanism mentioned in Section 4.2.1. This mechanism constrains each weight to lie in the range $[0, w_{\max}]$. This capping was performed after each of the mechanisms for changing the weights had been applied:

$$w'_{ij} = \begin{cases} 0 & \text{if } w_{ij} < 0 \\ w_{\max} & \text{if } w_{ij} > w_{\max} \\ w_{ij} & \text{otherwise} \end{cases}$$

The effect of imposing a maximum weight value onto individual weights for networks using subtractive presynaptic normalisation (without any postsynaptic normalisation) is shown in Figure 4.21. As can be seen from this Figure (especially when $w_{\max} = 0.2$), the imposition of a maximum weight value on subtractive normalisation does improve the nature of topography within the LGN, although there are several points to note:

1. The value of $w_{\max}$ must be carefully chosen. The larger the value of $w_{\max}$, the fewer postsynaptic units in different LGN rows are contacted by each presynaptic unit. This is clearly shown in the projective field plots shown in Figure 4.21. Since the target value of presynaptic normalisation is 1.0, values of $w_{\max}$ in the range [0.2,0.3] produce the best results since they allow presynaptic units to project to four rows of the LGN.

2. The subtractive normalisation must be applied gradually, using the parameter r_s which controls the rate of enforcement of subtractive normalisation. The experiments shown in Figure 4.21 were created with r_s set to 0.05; higher values of r_s causes the normalisation to be stricter, which produces results similar to the earlier results using just subtractive presynaptic normalisation (Figure 4.12) where r_s was set to 1.0. The effect of r_s on enforcing the normalisation constraint can be seen in Figure 4.20: the weights are still bounded, although the mean sum of weights from each presynaptic unit (1.560) is considerably above the desired target value of $T_{\text{pre}} = 1.0$.

3. The capping of weight values does not interfere with subtractive normalisation pushing weights to extreme values: at the end of the simulations presented in Figure 4.21, most

weights have strength 0 or $w_{\max}$. In contrast, simulations using divisive presynaptic normalisation produce weights distributed over a wide range of values.

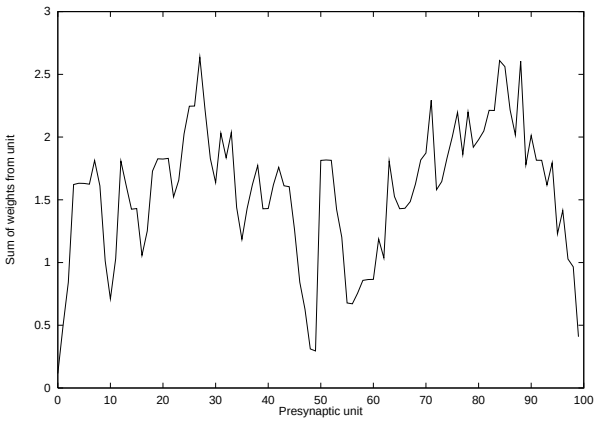
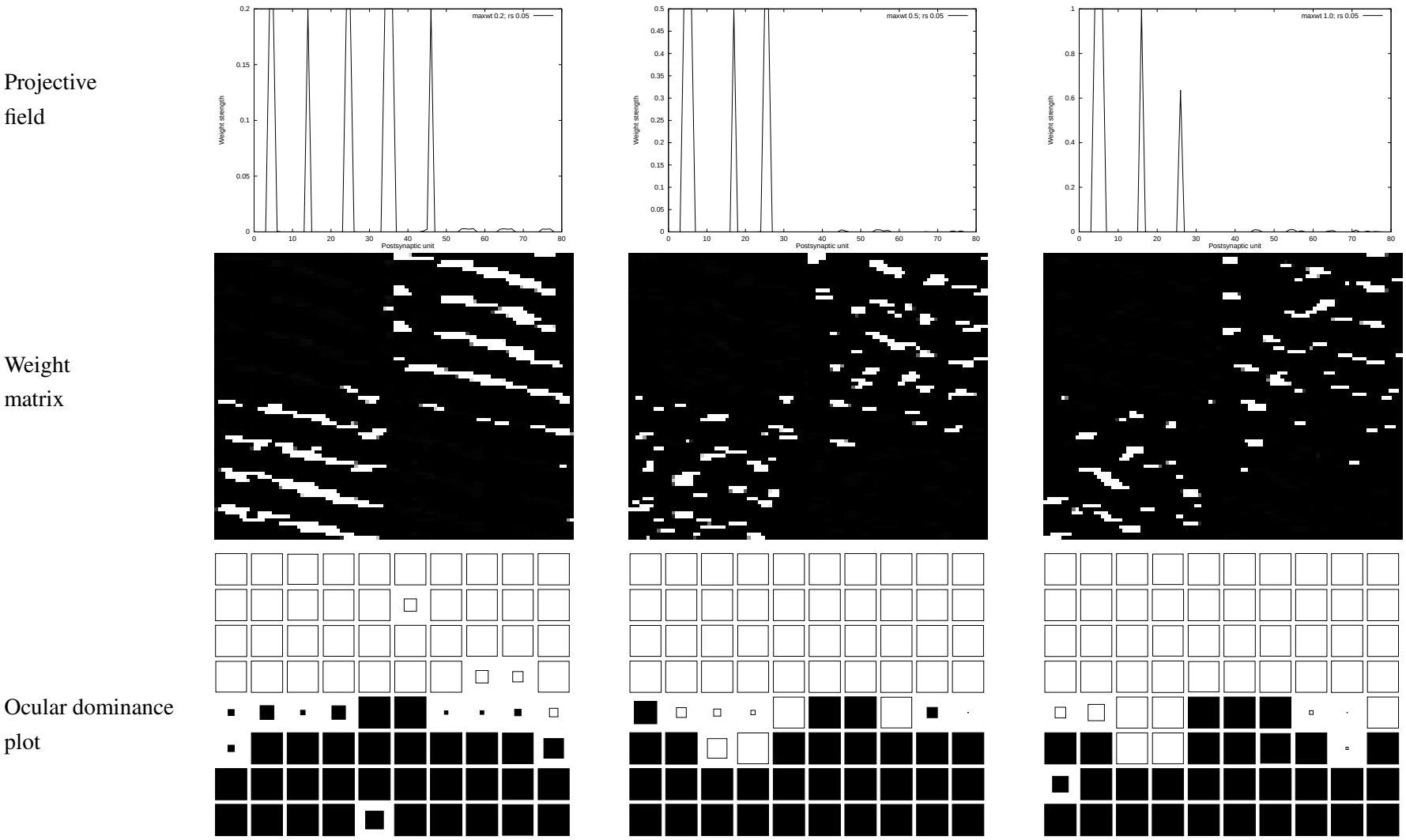


Figure 4.20: Sum of weights from each presynaptic unit under relaxed normalisation conditions. This figure shows the final sum of weights from each presynaptic unit for the network shown in Figure 4.21(a). The desired target value for the sum of weights from each units was $T_{\text{pre}} = 1.0$ with $r_s = 0.05$. (Average sum for all presynaptic units = 1.560; standard deviation = 0.532).



(a) Maximum weight = 0.2 (b) Maximum weight = 0.5 (c) Maximum weight = 1.0

Figure 4.21: The effect of imposing maximum values on individual weights upon development with subtractive presynaptic normalisation (and no postsynaptic normalisation). For these experiments, r_s was set to 0.05. The projective field plots show the final weight vector in each simulation for presynaptic unit 75 (located in the centre of the right eye).

4.7 Summary

This chapter has introduced the Keesing et al. (1992) model of retinogeniculate development. In the first half of this chapter, we have replicated the results of the Keesing model, showing in particular how the form of the initial weight bias influences network development. This replication stage was necessary since the only published information on the model omitted a number of crucial details of the model, including the form of the inputs and the nature of the initial weights (Keesing et al., 1992). We have been able to replicate these initial results, and also expand on the nature of the topographic map across all rows of the LGN, another aspect of the model that was not described in the original publication.

The second half of the chapter has analysed the role of the normalisation mechanisms in network development. In contrast to the original presentation of the model (Keesing et al., 1992), we have shown here that the form of the postsynaptic normalisation is redundant, as long as the presynaptic normalisation is implemented divisively, rather than subtractively. This is a new result, in contrast to previous models of ocular dominance which have used some form of postsynaptic normalisation (reviewed in Chapter 3) to ensure postsynaptic units become monocular. The only other model that relies on presynaptic normalisation of weights is the model by (von der Malsburg, 1979), although it also used a normalisation of the ocular marker concentrations induced into each postsynaptic unit. Postsynaptic normalisation (of either weights or markers) is very likely to produce monocular units, since any increase in weight strengths for some units from one eye is accompanied by a uniform decrease of all weights. Using subtractive enforcement of postsynaptic normalisation increases the tendency for monocular units due to the way that subtractive normalisation forces individual weights to extreme values. We have also shown that in some cases, the presynaptic normalisation can be implemented subtractively, as long as the normalisation is applied slowly and individual weights are constrained to lie between minimum and maximum values. Postsynaptic normalisation may not be required simply because no positive between-eye correlations are present in this model.

Normalisation techniques are often criticised for being too rigid, forcing the weight vectors to always lie exactly on a particular constraint surface, rather than within a wider region of weight space (Swindale, 1996). In this chapter, we have introduced an extra normalisation parameter, r_s, to address this criticism of normalisation techniques. If subtractive normalisation is gradually applied using the r_s parameter as shown in Section 4.6.5, the weights remain finite, but weight vectors no longer lie directly on the constraint surface (as shown in Figure 4.20), producing better topographic maps than when the vectors are forced to lie on the surface.