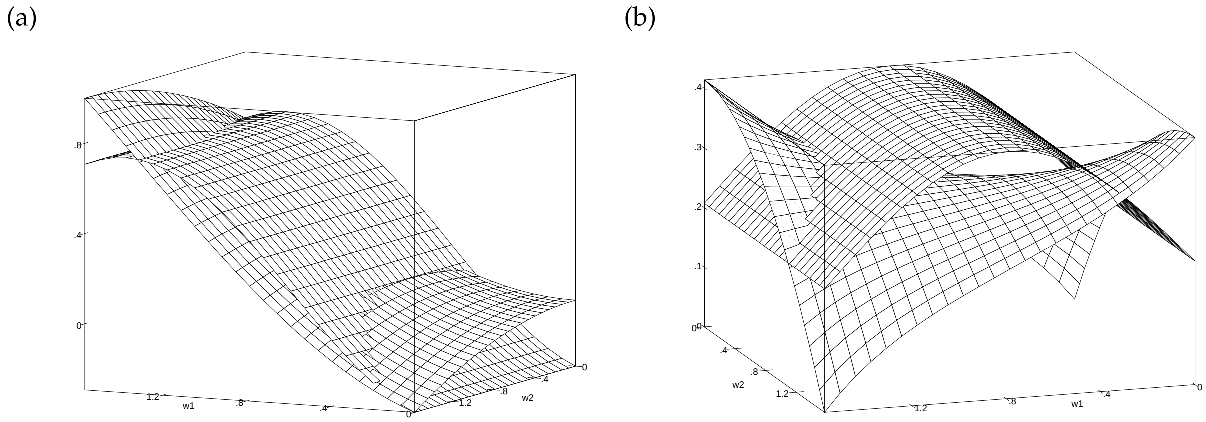

Figure 6.14: $\langle\dot{\omega}_1\rangle$ plotted as a function of ω_1 and ω_2. Vertical scale is arbitrary. (a) Divisive enforcement cases. (b) Subtractive enforcement cases for $\phi = \frac{\pi}{4}$. For the linear constraint cases, $\langle\dot{\omega}_1\rangle$ depends only on ω_1. Note that there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

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

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

which is satisfied for

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

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

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

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

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

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

6.5.4 Convergence times

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

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

Table 6.7: Convergence properties of the four cases

Divisive enforcement cases

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

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

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

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

and similarly

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

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

Subtractive enforcement cases

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

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

This yields the following relationship between ω_1 and ω_2:

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

where

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

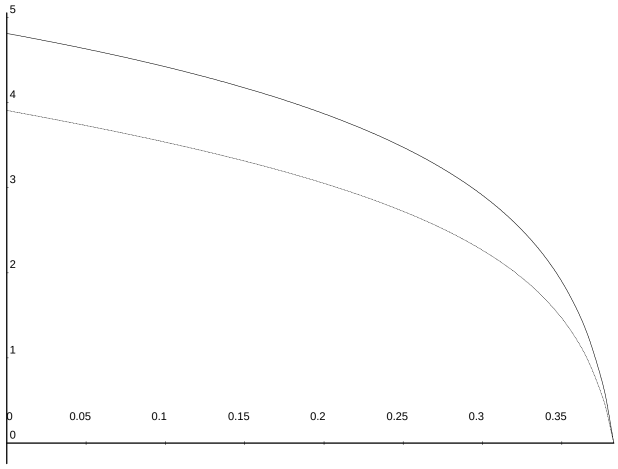

Figure 6.15: Time to convergence (arbitrary units) against initial ω_1 for divisive enforcement in the special case $\omega_1^0 + \omega_2^0 = \frac{\pi}{2}$, $\omega_1 < \frac{\pi}{8}$. Upper line is squared constraints, lower line is linear constraints.

We thus obtain

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

and

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

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

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

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

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

and

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

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

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

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

6.5.5 Summary

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

- The direction of movement of a weight vector is determined by (a) the side of the weight vector on which the input falls in the divisive enforcement cases, and (b) the side of ϕ on which the input falls in the subtractive enforcement cases.

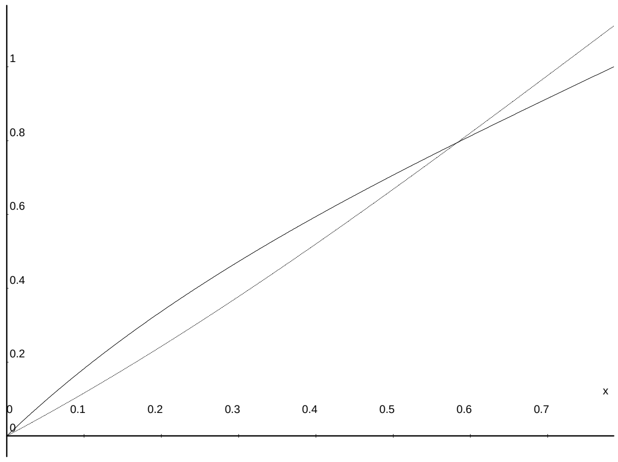

Figure 6.16: Time to saturation (arbitrary units) against initial ω_1 for subtractive constraints in the special case $\omega_1^0 + \omega_2^0 = \frac{\pi}{2}$, $\phi = \frac{\pi}{4}$. Squared constraints is the line that is higher for large x.

- The stable state for both linear and squared constraints in the divisive enforcement case is $\omega_1 = \frac{\pi}{8}$, $\omega_2 = \frac{3\pi}{8}$. For subtractive enforcement, the stable state depends on ϕ and whether the constraints are linear or squared. In both subtractive cases ϕ can be chosen so that $\boldsymbol{w}_1$ saturates at the right axis and $\boldsymbol{w}_2$ at the left axis, or so that both weights saturate at the same axis.
- The convergence times are different in each case. For both types of enforcement (with $\phi = \frac{\pi}{4}$), the time taken for the weights to saturate can be compared for the special case $\omega_1^0 + \omega_2^0 = \frac{\pi}{2}$. It is found that the squared case converges slower for divisive constraints but faster for subtractive constraints.

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

6.5.6 Extension to more units

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

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

For the linear constraint, divisive enforcement case the situation is different since it is the weight vector closest to the axis that the input vector is closest to that wins. First consider the case where $\frac{\pi}{8} < \omega_1 < \frac{3\pi}{8}$. Then $\boldsymbol{w}_1$ is the only vector that ever wins for $\theta < \frac{\pi}{4}$, and so is stable at $\omega_1 = \frac{\pi}{8}$ as before while all other vectors j such that $\omega_j < \frac{\pi}{4}$ remain in their initial positions. A similar situation holds in the upper octant. If $\omega_1 < \frac{\pi}{8}$ then it still eventually comes to rest at $\omega_1 = \frac{\pi}{8}$. However, if there are other vectors k such that $\omega_k < \frac{\pi}{8}$ then these will begin to win as ω_1 passes by on its way to $\frac{\pi}{8}$. Which unit wins changes as each progresses towards $\frac{\pi}{8}$, where they are all finally stable. Again, vectors with initial angles between $\frac{\pi}{8}$ and $\frac{3\pi}{8}$ remain in their initial states.

By similar arguments, the situation is even more straightforward in the linear constraint, subtractive enforcement case. $\boldsymbol{w}_1$ saturates at the right axis and $\boldsymbol{w}_N$ at the left axis as before (for appropriate ϕ), and all other weights remain unchanged. The squared constraint, subtractive enforcement case is however more complicated. In general all weights vectors $\boldsymbol{w}_i$ for which $\omega_i^0 < \phi$ saturate at the right axis: an analogous results holds for the left axis. However, this is not quite true of the two initial weight vectors $\boldsymbol{w}_j$ and $\boldsymbol{w}_{j+1}$ which are such that $\omega_j < \phi$ and $\omega_{j+1} > \phi$. Assume $\boldsymbol{w}_{j+1}$ is closer to ϕ than $\boldsymbol{w}_j$. Then $\boldsymbol{w}_{j+1}$ can win for inputs $\theta < \phi$, and thus $\boldsymbol{w}_{j+1}$ can eventually be pulled to the right axis.

We can thus clearly see the effect of a conscience mechanism that ensures that each unit wins roughly the same amount of the time. In the linear constraint, subtractive enforcement case, this would mean that eventually all weights would saturate at the axis they were initially closest too. For instance, for $\theta < \frac{\pi}{4}$, w_1 would win for the first pattern, but then w_2 would win for the next since w_1 is temporarily out of the running, and similarly for all weights.

6.5.7 Higher dimensions

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

6.5.8 The model

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

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

We return to address directly the model.

6.6 Discussion

6.6.1 Relation to Kohonen-type algorithms

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

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

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

6.6.2 Obermayer et al's algorithm

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

6.6.3 Miller's model

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

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

6.6.4 The elastic net

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

6.7 Conclusions

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

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