biologically by a mechanism whereby a cell adjusts its threshold or gain so as to keep its average activity roughly constant. Stability was achieved in the model by introducing a mechanism whereby, once weights reach their maximum or minimum values, they are "frozen" and not allowed to change further (similarly to [Miller et al 1989]). Thus once all connection strengths have saturated the complete map is frozen.

6.4.2 Afferent normalization

Afferent normalization refers to a constraint on each retinal unit to have a fixed amount of weight. A normalization rule of this type was discussed in [Prestige & Willshaw 1975], and used in the Tea Trade model [von der Malsburg & Willshaw 1977, Willshaw & von der Malsburg 1979] and Cowan's model [Whitelaw & Cowan 1981] among others. A rough biological motivation for such a rule is the idea that afferent fibres are supported by a flow of nutrients, and each presynaptic cell generates a fixed quantity of these, which must be divided between the fibres emerging from that cell. A linear afferent normalization constraint, divisively imposed, was employed in the model, in addition to the aforementioned efferent normalization. A similar mechanism was used in Miller's model [Miller et al 1989, footnote 23], where it was found that this had the effect of constraining stripe width to be less than an arbor radius (i.e. stripe period less than an arbor diameter), as discussed in chapter 4. Without such afferent normalization, it is possible for all the weights for a retinal unit to go to zero.

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

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

6.5 Analysis of the efferent normalization mechanism

6.5.1 Introduction

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

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

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

We therefore consider an extremely simple competitive situation that is mathematically tractable, with the hope that a thorough analysis of this will provide at least hints as to the situation pertaining in reality in the model. We look at the case of two weight vectors evolving in the positive quadrant of a two-dimensional space under

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

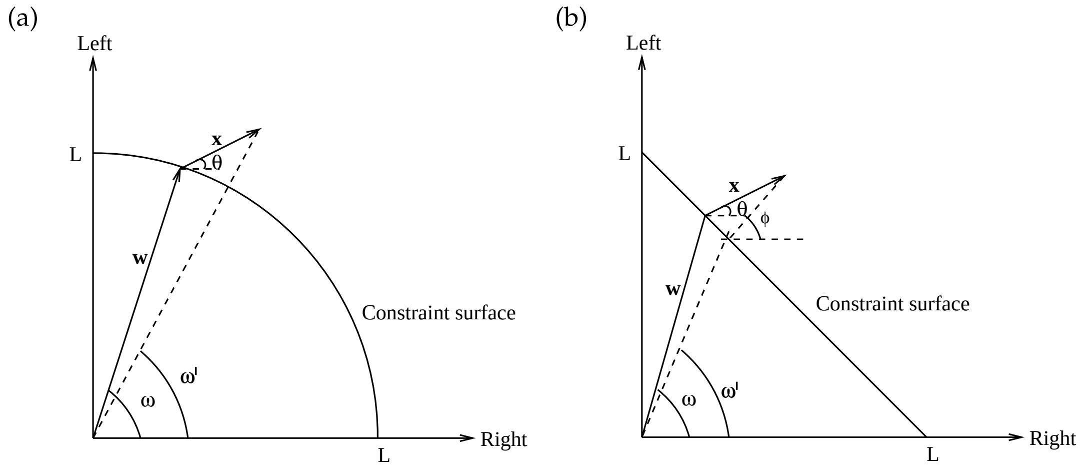

Figure 6.11: Geometry of an evolving weight for (a) squared constraints, divisive enforcement, and (b) linear constraints, subtractive enforcement. The squared subtractive and linear divisive cases are similar.

the influence of normalized input vectors uniformly distributed in direction. We consider a somewhat simpler competitive rule than that of the model for the evolution of weights w_i in response to input x:

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

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

The analysis proceeds in the following stages.

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

6.5.2 Step sizes

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

To calculate ω' in terms of ω, θ, ϕ, L and d we assume a small learning rate ϵ so that $\frac{\epsilon d}{L}$ is small to first order. We now derive expressions for the step in the squared constraint, divisive enforcement case shown in figure 6.11(a). The other cases follow similarly. Since $\frac{\epsilon d}{L}$ is small, we have that the distance from the tip of w to the tip

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

Table 6.4: Parameters for calculation of step sizes

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

Table 6.5: Value of ω' for winning unit.

of the new w is approximately equal to $\epsilon d\sin(\omega' - \theta)$. However, this distance is also (for $\omega - \omega'$ small) given by $L(\omega - \omega')$. Equating these two expressions gives

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

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

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

6.5.3 Rates of change

We now consider two competing weight vectors w_1 and w_2 with angles ω_1 and ω_2 as shown for the squared case in figure 6.13. The problem is to calculate the motion of each weight vector in response to the input patterns for which it wins, taking account of the fact that this set changes with time. We do this by assuming that the learning rate ϵ of the weight vectors is small enough so that they move infinitesimally in the time it takes to present all the input patterns. Pattern order is then not important, and we can average over the entire set of inputs in calculating the rates of change. Note that we assume $\omega_1 < \omega_2$ throughout: this is simply a matter of the labelling of the weights.

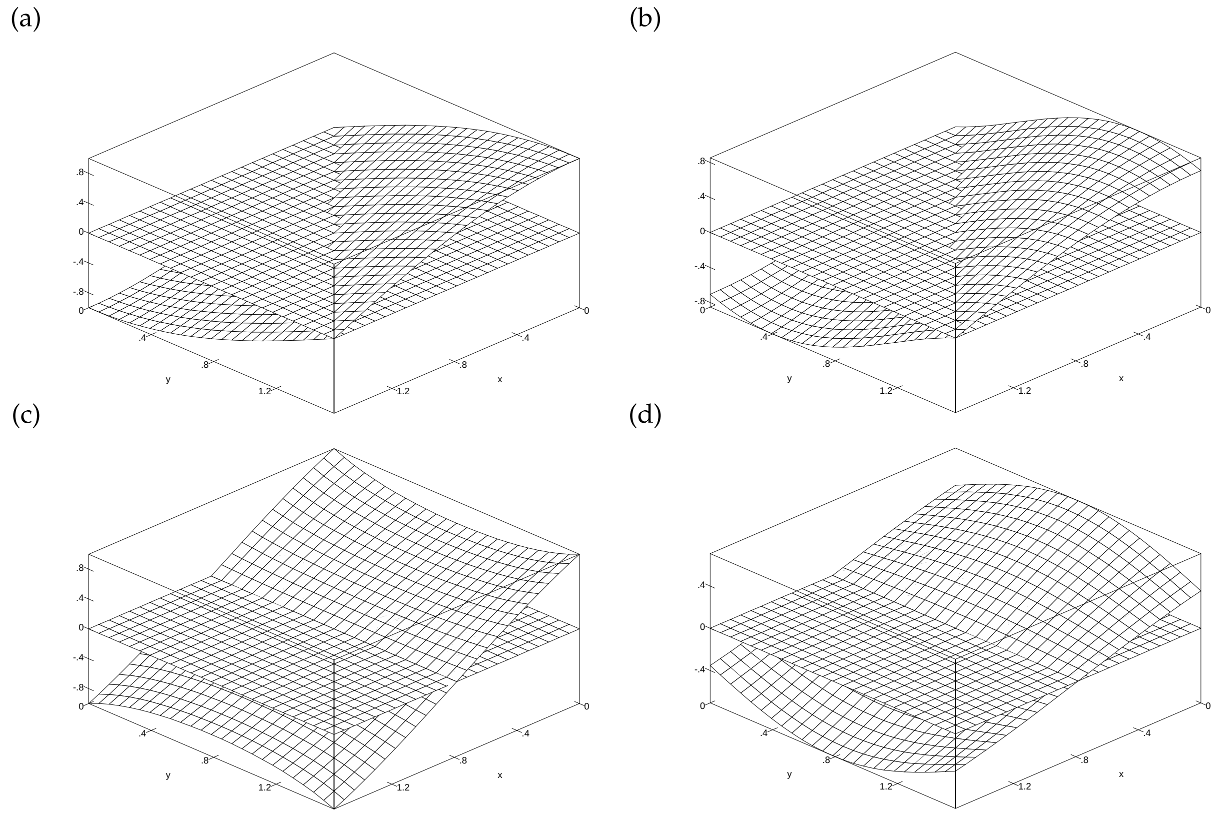

Figure 6.12: $-\dot{\omega}$ plotted as a function of θ (x axis) and ω (y axis) for each case. For the subtractive cases, ϕ was taken to be $\frac{\pi}{4}$. To make the sign of $\dot{\omega}$ clear the plane $\dot{\omega} = 0$ is also plotted. Vertical scale is arbitrary. (a) Squared constraints, divisive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

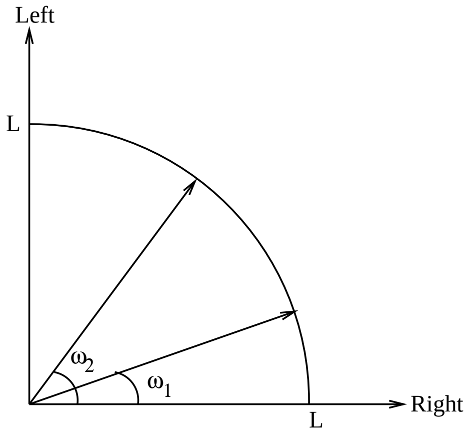

Figure 6.13: The geometry of two competing weight vectors in the squared divisive case.

Squared constraint, divisive enforcement case

Consider the evolution of ω_1. We have

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

Using the assumption that ϵ is small, we now average over all the patterns for which $\boldsymbol{w}_1$ wins the competition. In two dimensions this is straightforward: $\boldsymbol{w}_1$ wins for all $\theta < \frac{\omega_1+\omega_2}{2}$.

Thus for this case we have

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

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

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

We similarly obtain

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

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

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

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

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

That is

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

Thus $\boldsymbol{w}_1$ only tends towards the right axis when $\boldsymbol{w}_2$ is "close enough" to $\boldsymbol{w}_1$. By symmetry we have the following condition for which $\langle\dot{\omega}_2\rangle > 0$:

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

and thus the stable state is

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

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

Linear constraint, divisive enforcement case

For the linear constraint cases the weight vectors have variable length, and thus determining which unit wins the competition (in the sense of having the largest inner product with the input pattern) is slightly more complicated. In particular, the condition for $\boldsymbol{w}_1$ to win for input θ is

$$||\boldsymbol{w}_1||\cos(\theta - \omega_1)\ >\ ||\boldsymbol{w}_2||\cos(\theta - \omega_2)$$

where

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

This gives the condition

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

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

Similar reasoning for the linear divisive constraint to that above yields

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

and

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

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

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

and again the weights do not saturate at the axes.

Linear constraint, subtractive enforcement case

We now have

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

So

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

We similarly obtain

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

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

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

That is

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

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

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

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

Squared constraint, subtractive enforcement case

In the squared case, by similar reasoning we obtain

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

and

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

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

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