To appear in NEURAL COMPUTATION 6:2

The Role of Weight Normalization in Competitive Learning

Geoffrey J. Goodhill
University of Edinburgh
Centre for Cognitive Science
2 Buccleuch Place
Edinburgh EH8 9LW
UNITED KINGDOM

Harry G. Barrow
University of Sussex
School of Cognitive & Computing Sciences
Falmer
Brighton BN1 9QH
UNITED KINGDOM

Abstract

The effect of different kinds of weight normalization on the outcome of a simple competitive learning rule is analysed. It is shown that there are important differences in the representation formed depending on whether the constraint is enforced by dividing each weight by the same amount ("divisive enforcement"), or subtracting a fixed amount from each weight ("subtractive enforcement"). For the divisive cases weight vectors spread out over the space so as to evenly represent "typical" inputs, whereas for the subtractive cases the weight vectors tend to the axes of the space, so as to represent "extreme" inputs. The consequences of these differences are examined.

1 Introduction

Competitive learning [Rumelhart & Zipser 1986] has been shown to produce interesting solutions to many unsupervised learning problems (see e.g. [Becker 1991, Hertz et al 1991]). However, an issue that has not been greatly discussed is the effect of the type of weight normalization used. In common with other learning procedures that employ a simple Hebbian-type rule, it is necessary in competitive learning to introduce some form of constraint on the weights to prevent them from growing without bounds. This is often done by specifying that the sum (e.g. [von der Malsburg 1973]) or the sum-of-squares (e.g. [Barrow 1987]) of the weights for each unit should be maintained at a constant value.

Weight adaptation in competitive learning is usually performed only for the "winning" unit w, which we take to be the unit whose weight vector has the largest inner product with the input pattern x. Adaptation usually consists of taking a linear combination of the current weight vector and the input vector. The two most common rules are

$$w' = w + \epsilon x, \tag{1}$$

and

$$w' = w + \epsilon(x - w). \tag{2}$$

Consider the general case

$$w' = aw + \epsilon x, \tag{3}$$

where $a = 1$ for rule 1 and $a = 1 - \epsilon$ for rule 2. For a particular normalization constraint, e.g. $\|w\| = L$, there are various ways in which that constraint may be enforced. The two main approaches are:

$$w = w'/\alpha, \tag{4}$$

and

$$w = w' - \beta c, \tag{5}$$

where c is a fixed vector, and α and β are calculated to enforce the constraint. For instance, if the constraint is $||w|| = L$ then $\alpha = ||w'||/L$. The simplest case for c is $c_i = 1\ \forall i$. We refer to rule 4 as "divisive" enforcement, since each weight is *divided* by the same amount so as to enforce the constraint, and rule 5 as "subtractive" enforcement, since here an amount is *subtracted* from each weight so as to enforce the constraint. It should be noted that the qualitative behaviour of each rule does not depend on the value of a. It is straightforward to show that any case in which $a \neq 1$ is equivalent to a case in which $a = 1$ and the parameters ϵ and L have different values. In this paper, therefore, we will consider only the case $a = 1$.

The effect of these two types of enforcement on a model for ocular dominance segregation, where development is driven by the time-averaged correlation matrix of the inputs, was mentioned by [Miller 1990(a), footnote 24]. Divisive and subtractive enforcement have been thoroughly analysed for the case of general linear learning rules in [Miller & MacKay 1992]. They show that in this case divisive enforcement causes the weight pattern to tend to the principal eigenvector of the synaptic development operator, whereas subtractive enforcement causes almost all weights to reach either their minimum or maximum values.

Competitive learning however involves choosing a winner, and thus does not succumb to the analysis employed by [Miller & MacKay 1992], since account needs to be taken of the changing subset of inputs for which each output unit wins. In this paper we analyse a special case of competitive learning which, although simple, highlights the differences between divisive and subtractive enforcement. We also consider both normalization constraints $\sum_i w_i =$ constant and $\sum_i w_i^2 =$ constant, and thus compare four cases in all.

The analysis focuses on the case of two units (i.e. two weight vectors) evolving in the positive quadrant of a two-dimensional space under the influence of normalized input vectors uniformly distributed in direction. Later it is suggested how the conclusions can be extended to various more complex situations. It is shown that, for uniformly distributed inputs, 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.

2 Analysis

The analysis proceeds in the following stages: (1) Calculate the weight change for the winning unit in response to an input pattern. (2) Calculate the average rate of change of a weight vector, by averaging over all patterns for which that unit wins. (3) Calculate the phase plane dynamics, in particular the stable states.

2.1 Weight changes

The change in direction of the weight vector for the winning unit is derived by considering the geometric effect of updating weights and then enforcing the normalization constraint. A formula for the change in the weight in the general case is derived, and then instantiated to each of the four cases under consideration. For convenience the axes are referred to as "left" (y axis) and "right" (x axis). Figure 1 shows the effect of updating a weight vector w with angle ω to the right axis, and then enforcing a normalization constraint. Notation is summarized in table 1. A small fraction of x is added to w, and then the constraint is enforced by projecting back to the normalization surface (the surface in which all normalized vectors lie) at angle ϕ, thus defining the new weight. For the squared constraint case, the surface is a circle centred on the origin with radius L. For the linear constraint case, the surface is a line normal to the vector $(1, 1)$ which cuts the right axis at $(L, 0)$. When ϵ is very small, we may consider the normalization surface to be a plane, even in the squared constraint case. For this case the normalization surface is normal to the weight vector, a tangent

Parameter	Description
w	Weight vector
x	Input vector
δw	Change in weight vector
ω	Angle of weight vector to right axis
$\delta\omega$	Change in angle of weight vector to right axis
θ	Angle of input pattern vector to right axis
ϕ	Angle of enforcement vector to right axis
σ	Angle of normal to constraint surface to right axis
L	Magnitude of the normalization constraint
d	$\|\|x\|\|$ (constant)
ϵ	Learning rate

Table 1: Notation for calculation of weight vectors

of the circle. For divisive enforcement, the projection direction is back along w', directly towards the origin. For subtractive enforcement, the projection direction is back along a fixed vector c, typically $(1, 1)$.

Referring to figure 1(a), consider

$$\delta w = \epsilon x - \beta c.$$

Resolving horizontally and vertically and then eliminating $\beta||c||$ yields

$$||\delta w|| = -\epsilon ||x|| \frac{\sin(\theta - \phi)}{\cos(\sigma - \phi)}. \tag{6}$$

Now referring to figure 1(b), consider the change in angle ω, $\delta\omega$:

$$||w||\delta\omega = -||\delta w|| \cos(\sigma - \omega),$$

which in conjunction with equation 6 gives

$$\delta\omega = \frac{\epsilon||x||}{||w||} \frac{\sin(\theta - \phi)}{\cos(\sigma - \phi)} \cos(\sigma - \omega). \tag{7}$$

For the squared constraint case $||w|| = L$, whereas in the linear constraint case

$$||w|| = \frac{L}{\sqrt{2}\cos(\sigma - \omega)}$$

For divisive enforcement $\phi = \omega$, whereas for subtractive enforcement ϕ is constant. From now on we assume $||x|| = d$, a constant. Table 2 shows the instantiation of equation 7 in the four particular cases studied below.

An important difference between divisive and subtractive enforcement is immediately apparent: for divisive enforcement the sign of the change is dependent on $\text{sign}(\theta - \omega)$, while for subtractive enforcement it is dependent on $\text{sign}(\theta - \phi)$. (Note that $\cos(\omega - \phi)$, $\cos(\frac{\pi}{4} - \phi)$ and $\cos(\frac{\pi}{4} - \omega)$ are always positive for $\omega, \phi \in [0, \frac{\pi}{2}]$.) Thus in the divisive case a weight vector only moves towards (say) the right axis 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.

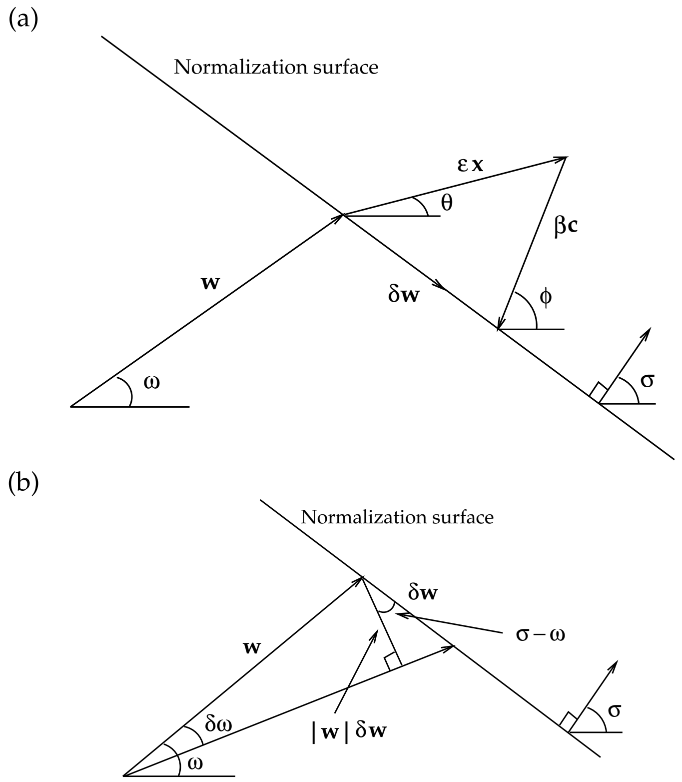

Figure 1: (a) The general case of updating a weight vector w by adding a small fraction of the input vector x and then projecting at angle ϕ back to the normalization surface. (b) The change in angle ω, $\delta\omega$, produced by the weight update.

2.2 Averaged weight changes

The case of two competing weight vectors w_1 and w_2 with angles ω_1 and ω_2 respectively to the right axis is now considered. It is assumed that $\omega_1 < \omega_2$: this is simply a matter of the labelling of the weights. 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. This is done by assuming that the learning rate ϵ is small enough so that the weight vectors move infinitesimally in the time it takes to present all the input patterns. Pattern order is then not important, and it is possible to average over the entire set of inputs in calculating the rates of change.

Consider the evolution of ω_i, $i = 1, 2$. In the continuous time limit, from equation 7 we have

$$\dot{\omega}_i = \frac{\epsilon d}{||w_i||} \frac{\cos(\sigma - \omega_i)}{\cos(\sigma - \phi)} \sin(\theta - \omega_i).$$

Using the assumption that ϵ is small, an average is now taken over all the patterns for which w_i wins the competition. In two dimensions this is straightforward. For instance consider w_1: in the squared constraint cases w_1 wins for all $\theta < \frac{\omega_1 + \omega_2}{2}$. In the linear constraint cases the weight vectors have variable length, and the condition for w_1 to win for input θ is now

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

Constraint	Enforcement	Equivalences	$\delta\omega$
Squared	Divisive	$\sigma = \omega \quad \phi = \omega \quad \|\|w\|\| = L$	$\frac{\epsilon d}{L}\sin(\theta - \omega)$
Linear	Divisive	$\sigma = \frac{\pi}{4} \quad \phi = \omega \quad \|\|w\|\| = \frac{L}{\sqrt{2}\cos(\sigma - \omega)}$	$\frac{\epsilon d\sqrt{2}}{L}\cos(\frac{\pi}{4} - \omega)\sin(\theta - \omega)$
Squared	Subtractive	$\sigma = \omega \quad \|\|w\|\| = L$	$\frac{\epsilon d}{L}\frac{1}{\cos(\omega - \phi)}\sin(\theta - \phi)$
Linear	Subtractive	$\sigma = \frac{\pi}{4} \quad \|\|w\|\| = \frac{L}{\sqrt{2}\cos(\sigma - \omega)}$	$\frac{\epsilon d\sqrt{2}}{L}\frac{\cos^2(\frac{\pi}{4} - \omega)}{\cos(\frac{\pi}{4} - \phi)}\sin(\theta - \phi)$

Table 2: Value of $\delta\omega$ for winning unit

where

$$||w_i|| = \frac{L}{\sqrt{2}\cos(\frac{\pi}{4} - \omega_i)}\ ,\ i = 1, 2.$$

This yields the condition $\theta < \frac{\pi}{4}$ for w_1 to win for input θ. That is, in the linear cases the unit that wins is the unit closest to the axis to which the input is closest, and the weights evolve effectively independently of each other. (Note that we have only assumed $\omega_1 < \omega_2$, not $\omega_1 < \frac{\pi}{4}$.)

First equation 7 is integrated for general limits θ_1 and θ_2, and then the particular values of θ_1 and θ_2 for each of the cases are substituted. We have

$$\langle\dot{\omega}_i\rangle_{\theta_1<\theta<\theta_2} = \frac{\epsilon d}{||w_i||}\frac{\cos(\sigma - \omega_i)}{\cos(\sigma - \phi)}\int_{\theta_1}^{\theta_2}\sin(\theta - \omega_i)P(\theta)d\theta \tag{8}$$

where the angle brackets denote averaging over the specified range of θ, and $P(\theta)$ is the probability of input θ. The outcome under any continuous distribution can be determined by appropriate choice of $P(\theta)$. Here we just consider the simplest case of the uniform distribution $P(\theta) = p$, a constant. With some trigonometrical manipulation it follows that

$$\langle\dot{\omega}_i\rangle = \frac{2\epsilon dp}{||w_i||}\frac{\cos(\sigma - \omega_i)}{\cos(\sigma - \phi)}\sin\left(\phi - \frac{\theta_1 + \theta_2}{2}\right)\sin\left(\frac{\theta_1 - \theta_2}{2}\right). \tag{9}$$

2.3 Stable states

2.3.1 Linear constraints

Substituting the limits derived above for linear constraints into equation 9 yields for the divisive enforcement case

$$\langle\dot{\omega}_1\rangle = -C\sqrt{2}\cos\left(\frac{\pi}{4} - \omega_1\right)\sin\left(\omega_1 - \frac{\pi}{8}\right)\sin\left(\frac{\pi}{8}\right)$$
$$\langle\dot{\omega}_2\rangle = -C\sqrt{2}\cos\left(\frac{\pi}{4} - \omega_2\right)\sin\left(\omega_2 - \frac{3\pi}{8}\right)\sin\left(\frac{\pi}{8}\right)$$

where for conciseness we have defined $C = 2\epsilon dp/L$. To determine the behaviour of the system the conditions for which $\langle\dot{\omega}_1\rangle$ and $\langle\dot{\omega}_2\rangle$ are positive, negative, and zero are examined. It is clear that w_1 moves towards the right axis for $\omega_1 > \pi/8$, w_2 moves towards the left axis for $\omega_2 < 3\pi/8$, and the stable state is

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

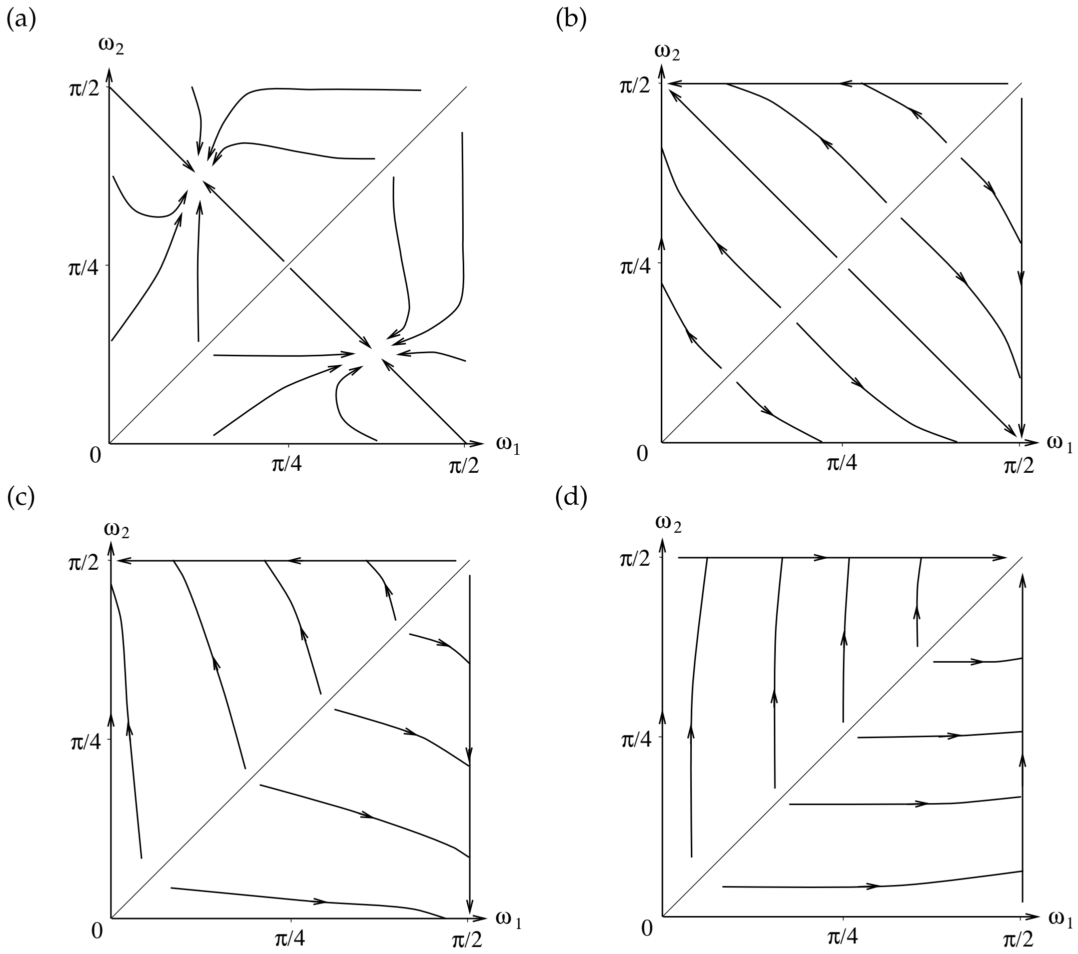

Figure 2: Phase plane portraits of the dynamics for linear constraint cases. (a) Divisive enforcement: weights tend to $(\pi/8, 3\pi/8)$. (b,c,d) Subtractive enforcement for $\phi = \pi/4$, $\phi = \pi/6$, and $\phi = \pi/16$ respectively. For $\pi/8 < \phi < 3\pi/8$ weights saturate one at each axis, otherwise both saturate at the same axis.

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. This behaviour can be clearly visualised in the phase plane portrait (figure 2(a)).

For the subtractive enforcement case

$$\langle \dot{\omega}_1 \rangle = -C\sqrt{2}\frac{\cos^2(\frac{\pi}{4} - \omega_1)}{\cos(\frac{\pi}{4} - \phi)} \sin\left(\phi - \frac{\pi}{8}\right) \sin\left(\frac{\pi}{8}\right)$$

$$\langle \dot{\omega}_2 \rangle = -C\sqrt{2}\frac{\cos^2(\frac{\pi}{4} - \omega_2)}{\cos(\frac{\pi}{4} - \phi)} \sin\left(\phi - \frac{3\pi}{8}\right) \sin\left(\frac{\pi}{8}\right)$$

For $\langle \dot{\omega}_1 \rangle < 0$, that is w_1 heading for the right axis, it is required that $\phi > \pi/8$. Similarly for w_2 to be heading for the left axis it is required that $\phi < 3\pi/8$. Thus the weights saturate, one at each axis, if $\frac{\pi}{8} < \phi < \frac{3\pi}{8}$. They both saturate at the left axis for $\phi < \frac{\pi}{8}$, and both at the right axis for $\phi > \frac{3\pi}{8}$. Phase plane portraits for some illustrative values of ϕ are shown in figure 2(b-d).

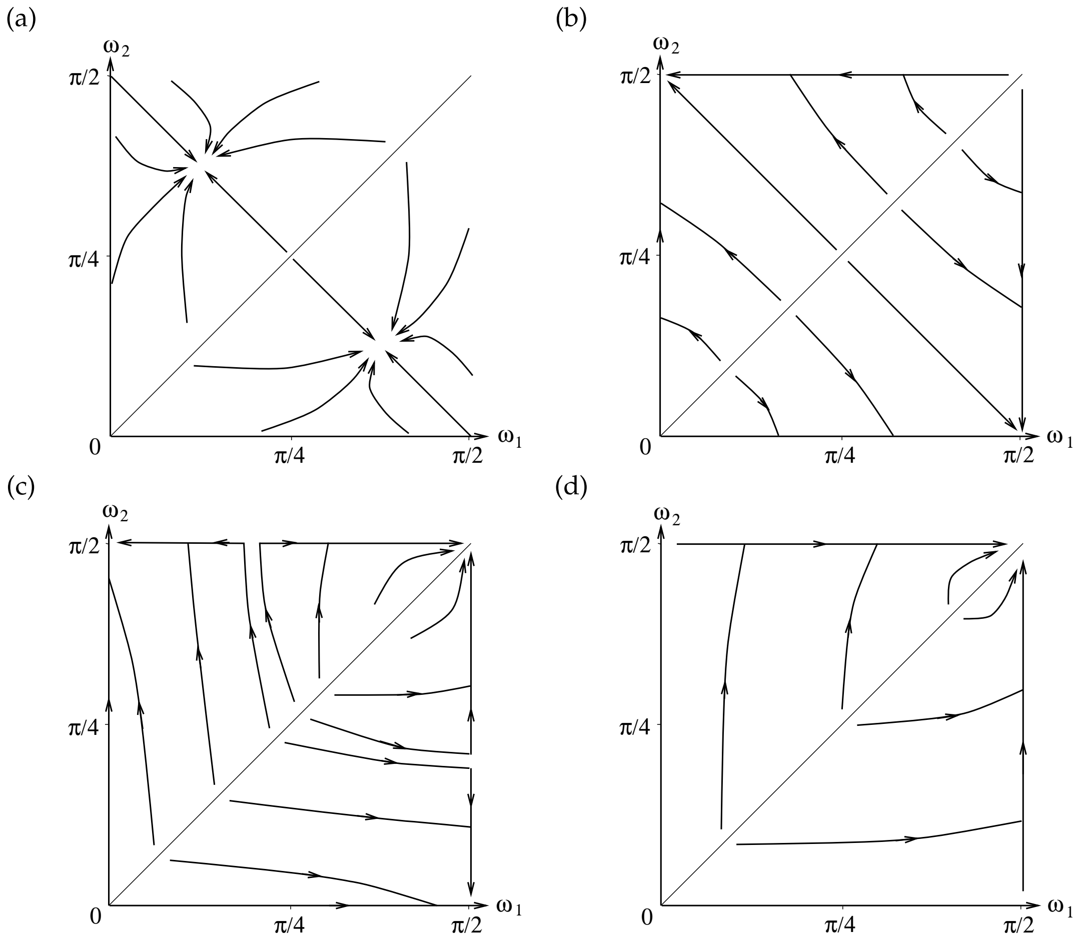

Figure 3: Phase plane portraits of the dynamics for squared constraint cases. (a) Divisive enforcement: weights tend to $(\pi/8, 3\pi/8)$. (b,c,d) Subtractive enforcement for $\phi = \pi/4$, $\phi = \pi/6$, and $\phi = \pi/16$ respectively. For $\phi = \pi/4$ weights saturate at different axes. As ϕ moves from $\pi/4$, there is an increasing region of the (ω_1, ω_2) plane for which the final outcome is saturation at the same axis.

2.3.2 Squared constraints

Instantiating equation 9 in the divisive enforcement case yields

$$\langle \dot{\omega}_1 \rangle = C \sin\left(\frac{\omega_2 - 3\omega_1}{4}\right) \sin\left(\frac{\omega_1 + \omega_2}{4}\right)$$

$$\langle \dot{\omega}_2 \rangle = C \sin\left(\frac{\pi + \omega_1 - 3\omega_2}{4}\right) \sin\left(\frac{\pi - \omega_1 - \omega_2}{4}\right)$$

For $\langle \dot{\omega}_1 \rangle < 0$ we require $3\omega_1 > \omega_2$, for $\langle \dot{\omega}_2 \rangle > 0$ we require $3\omega_2 < \omega_1 + \pi$, and the stable state is the same as in the linear constraint, divisive enforcement case:

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

The phase plane portrait is shown in figure 3(a).

Constraints	Divisive	Subtractive
Linear	Weights stable at $\omega_1 = \frac{\pi}{8}$, $\omega_2 = \frac{3\pi}{8}$	Weights saturate at different axes for $\frac{\pi}{8} < \phi < \frac{3\pi}{8}$. For $\phi < \frac{\pi}{8}$ both weights saturate at left axis. For $\phi > \frac{3\pi}{8}$ both weights saturate at right axis.
Squared	Weights stable at $\omega_1 = \frac{\pi}{8}$, $\omega_2 = \frac{3\pi}{8}$	Weights saturate at different axes for $\phi = \frac{\pi}{4}$. For $\frac{\pi}{8} < \phi < \frac{3\pi}{8}$ weights may saturate at the same or different axes (see text). For other ϕ both weights saturate at same axis as in linear case.

Table 3: Convergence properties of the four cases

In the subtractive enforcement case we have

$$\langle \dot{\omega}_1 \rangle = -C \frac{1}{\cos(\phi - \omega_1)} \sin\left(\phi - \frac{\omega_1 + \omega_2}{4}\right) \sin\left(\frac{\omega_1 + \omega_2}{4}\right)$$

$$\langle \dot{\omega}_2 \rangle = C \frac{1}{\cos(\phi - \omega_2)} \sin\left(\phi - \frac{\omega_1 + \omega_2 + \pi}{4}\right) \sin\left(\frac{\omega_1 + \omega_2 - \pi}{4}\right)$$

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

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

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

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

If $\phi = \pi/4$ both these conditions are satisfied for all but two initial states and weights saturate one at each axis, i.e. the only stable attractor is $(0, \pi/2)$. The two points for which this is not true are the critical points $(0, 0)$ and $(\pi/2, \pi/2)$. Here $\dot{\omega}_1 = 0$, $\dot{\omega}_2 = 0$, and these are unstable equilibria.

If $\pi/8 < \phi < \pi/4$, both $(\pi/2, \pi/2)$ and $(0, \pi/2)$ are stable attractors. Both weights can saturate at the left axis if they start sufficiently close to it. The size of the basin of attraction around $(\pi/2, \pi/2)$ gradually increases as ϕ decreases, until for $\phi < \pi/8$ the point $(0, \pi/2)$ is no longer an attractor and all initial conditions lead to saturation of both weights at the left axis. Analogous results hold for $\pi/4 < \phi < 3\pi/8$, and $\phi > 3\pi/8$. Phase plane portraits for some illustrative values of ϕ are shown in figure 3(b-d). We have not been able to find an analytic expression for the boundary between the different basins of attraction in this case.

Convergence properties for each of the 4 cases of constraints and enforcement are summarized in table 3.

3 Discussion

3.1 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 the stable state is $\omega_1 = \frac{\omega_2}{3}$, $\omega_N = (\frac{\omega_{N-1}+\pi}{3})$. 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 to which the input vector is closest that wins. First consider the case where $\frac{\pi}{8} < \omega_1 < \frac{3\pi}{8}$. Then w_1 is the only vector that ever wins for $\theta < \frac{\pi}{4}$, and so is stable at $\omega_1 = \frac{\pi}{8}$ as before while all other vectors j such that $\omega_j < \frac{\pi}{4}$ remain in their initial positions. A similar situation holds in the upper octant. If $\omega_1 < \frac{\pi}{8}$ then it still eventually comes to rest at $\omega_1 = \frac{\pi}{8}$. However, if there are other vectors k such that $\omega_k < \frac{\pi}{8}$ then these will begin to win as ω_1 passes by on its way to $\frac{\pi}{8}$. Which unit wins changes as each progresses towards $\frac{\pi}{8}$, where they are all finally stable. Again, vectors with initial angles between $\frac{\pi}{8}$ and $\frac{3\pi}{8}$ remain in their initial states.

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

The effect of a conscience mechanism [Hertz et al 1991] that ensures that each unit wins roughly the same amount of the time can thus be clearly seen. In the linear constraint, subtractive enforcement case, this would mean that eventually all weights would saturate at the axis to which they were initially closest. 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.

3.2 Higher dimensions

In higher dimensional spaces, the situation immediately becomes much more complicated, for two main reasons. Firstly, it is harder to calculate the new weight vector for the winning unit in terms of the old by the geometric methods we have used here. Secondly, the dividing lines between which unit wins for the set of inputs forms a Voronoi Tesselation [Hertz et al 1991] of the constraint surface. The limits of the integral required to calculate $\langle\dot{\omega}\rangle$ are thus hard to determine, and the integrand is more complicated. Empirical results have been obtained for this case in the context of a model for ocular dominance segregation [Goodhill 1993]. Figure 4 shows an example for this model (discussed further below), illustrating that qualitatively similar behaviour to the two dimensional case occurs in higher dimensions. See [Miller & MacKay 1992] for analysis of linear learning rules in spaces of general dimension. A more general analysis for the competitive case will be found in [Barrow & Goodhill 1993].

3.3 Representations

How do the representations formed by divisive and subtractive enforcement differ, and for what kinds of problems might each be appropriate? From the two-dimensional results presented here, it appears that divisive enforcement is appropriate if it is desired to spread weight vectors out over the space so as to evenly represent "typical" inputs, or, if the inputs are clustered, to find the cluster centres. Subtractive enforcement on the other hand represents "extreme" inputs: instead of finding the centres of clusters, the weight vectors tend to the axes of the space to which clusters are closest. Subtractive enforcement can be thought of as making harsher decisions about the input distribution.

These properties are illustrated for a high-dimensional case of a similar learning rule in the model of visual map formation and ocular dominance segregation of [Goodhill 1991, Goodhill 1993]. Here, an array of "cortical" units compete in response to inputs from two arrays of "retinal" units. With divisive enforcement of a linear normalization rule, no ocular dominance segregation occurs unless only a small patch of retina in one eye or the other is active at a time [Goodhill 1990]. However with subtractive enforcement segregation

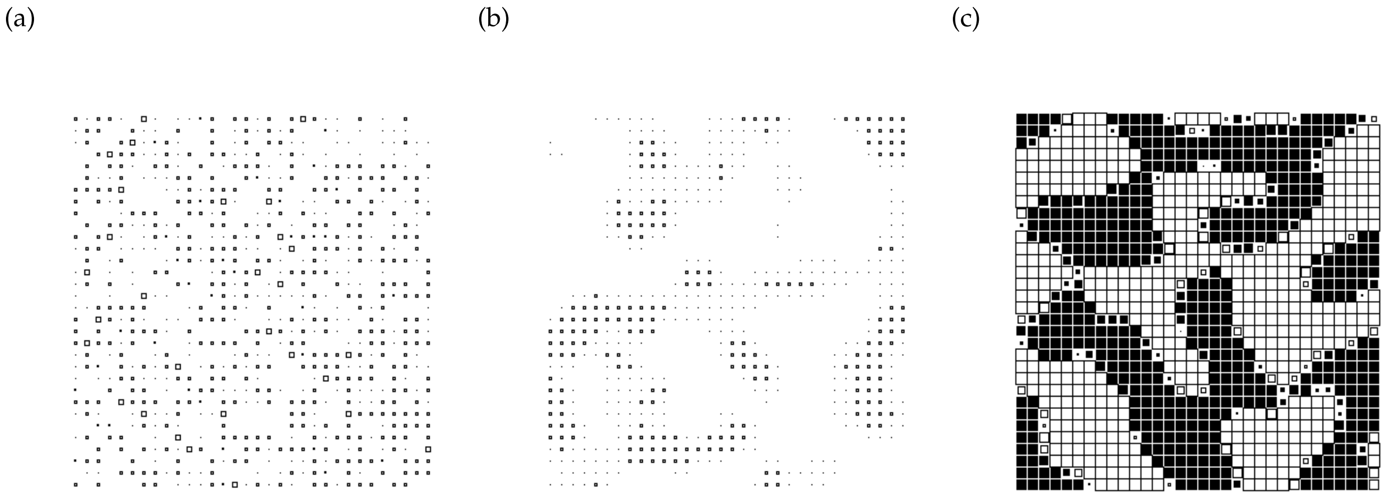

Figure 4: Outcome in a high dimensional case under divisive and subtractive enforcement of a linear normalization constraint ($c_i = 1, \forall i$). The pictures show the ocular dominance of an array of 32 by 32 postsynaptic units, whose weights are updated using a competitive rule in response to distributed, locally correlated patterns of input activity. For each postsynaptic unit the colour of the square indicates for which eye the unit is dominant, and the size of the square represents the degree of dominance. (a) Initial state (random connection strengths): no postsynaptic units are strongly dominated. (b) State after 250,000 iterations for divisive enforcement of a linear constraint: again no postsynaptic units are strongly dominated. (c) State after 250,000 iterations for subtractive enforcement of a linear constraint: segregation has occurred. For further details of this model see [Goodhill 1991, Goodhill 1993].

does occur when all retinal units are simultaneously active, with local correlations of activity within each retina and positive correlations between the two retinae [Goodhill 1991, Goodhill 1993]. This is illustrated in figure 4.

Two other points of note are as follows. (1) Whereas the stable state for divisive enforcement is invariant to affine transformations of the input space, it is not for subtractive enforcement. (2) A natural type of projection onto the constraint surface to consider is an orthogonal projection. For squared constraints orthogonal projection corresponds to divisive enforcement, whereas for linear constraints orthogonal projection corresponds to subtractive enforcement with $c_i = 1 \ \forall i$ in equation 5. Thus applying a rule of orthogonal projection leads to a very different outcome for squared and linear constraints.

4 Conclusions

A simple case of competitive learning has been analysed with respect to whether the normalization constraint is linear or sum-of-squares, and also whether the constraint is enforced divisively or subtractively. It has been shown that the outcome is significantly different depending on the type of enforcement, while being relatively insensitive to the type of constraint. Divisive enforcement causes the weights to represent "typical" inputs, whereas subtractive enforcement causes the weights to represent "extreme" inputs. These results are similar to the linear learning rule case analysed in [Miller & MacKay 1992].

Directions for future work include analysis of normalization in competitive learning systems of higher dimension, and studying the differences in the representations formed by divisive and subtractive enforcement on a variety of problems of both practical and biological interest.

Acknowledgements

We thank Ken Miller both for first introducing us to the difference between divisive and subtractive constraints in the linear case and the two dimensional version of this, and for kindly making available [Miller & MacKay 1992] while still in draft form. GJG thanks David Griffel, Peter Williams, Peter Dayan, Martin Simmen and Steve Finch for helpful comments and suggestions. GJG was supported by an SERC Postgraduate Studentship (at Sussex University), and a Joint Councils Initiative in CogSci/HCI postdoctoral training fellowship.

Bibliography

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

Barrow, H.G. & Goodhill G.J. (1993). In preparation.

Becker, S. (1991). Unsupervised learning: procedures for neural networks. *Int. Jou. Neu. Sys.*, **2**, 17-33.

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

Goodhill, G.J. (1991). *Correlations, Competition and Optimality: Modelling the Development of Topography and Ocular Dominance.* PhD Thesis, University of Sussex.

Goodhill, G.J. (1993). Topography and ocular dominance: a model exploring positive correlations. *Biological Cybernetics*, in press.

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

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

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

Miller, K.D. & MacKay, D.J.C. (1992). The role of constraints in Hebbian Learning. Caltech Computation and Neural Systems Memo 19; Also *Neural Computation*, in press.

Rumelhart, D.E. & Zipser, D. (1986). Feature discovery by competitive learning. In *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, eds. J.L. McClelland & D.E. Rumelhart, 151-193, MIT Press, Cambridge, M.A.