# The Role of Constraints in Hebbian Learning

Kenneth D. Miller[†]
Division of Biology
Caltech 216-76
Pasadena, CA 91125
ken@phy.ucsf.edu

David J.C. MacKay[*]
Computation and Neural Systems
Caltech 139-74
Pasadena CA 91125
mackay@ras.phy.cam.ac.uk

Published as *Neural Computation* **6**, 100-126 (1994)

**Abstract**

Models of unsupervised correlation-based (Hebbian) synaptic plasticity are typically unstable: either all synapses grow until each reaches the maximum allowed strength, or all synapses decay to zero strength. A common method of avoiding these outcomes is to use a constraint that conserves or limits the total synaptic strength over a cell. We study the dynamical effects of such constraints.

Two methods of enforcing a constraint are distinguished, multiplicative and subtractive. For otherwise linear learning rules, multiplicative enforcement of a constraint results in dynamics that converge to the principal eigenvector of the operator determining unconstrained synaptic development. Subtractive enforcement, in contrast, typically leads to a final state in which almost all synaptic strengths reach either the maximum or minimum allowed value. This final state is often dominated by weight configurations other than the principal eigenvector of the unconstrained operator. Multiplicative enforcement yields a "graded" receptive field in which most mutually correlated inputs are represented, whereas subtractive enforcement yields a receptive field that is "sharpened" to a subset of maximally-correlated inputs. If two equivalent input populations (*e.g.* two eyes) innervate a common target, multiplicative enforcement prevents their segregation (ocular dominance segregation) when the two populations are weakly correlated; whereas subtractive enforcement allows segregation under these circumstances.

These results may be used to understand constraints both over output cells and over input cells. A variety of rules that can implement constrained dynamics are discussed.

[†]Current address: Depts. of Physiology and Otolaryngology, University of California, San Francisco, CA 94143-0444.
[*]Current address: Radio Astronomy, Cavendish Laboratory, Madingley Road, Cambridge, CB3 0HE, United Kingdom.

Development in many neural systems appears to be guided by "Hebbian" or similar activity-dependent, correlation-based rules of synaptic modification (reviewed in Miller, 1990a). Several lines of reasoning suggest that constraints limiting available synaptic resources may play an important role in this development.

Experimentally, such development often appears to be *competitive*. That is, the fate of one set of inputs depends not only on its own patterns of activity, but on the activity patterns of other, competing inputs. A classic example is given by the experiments of Wiesel and Hubel (1965) on the effects of monocular versus binocular visual deprivation in young animals (see also Guillery, 1972). If neural activity is reduced in one eye, inputs responding to that eye lose most of their connections to the visual cortex, while the inputs responding to the normally active, opposite eye gain more than their normal share of connections. If activity is reduced simultaneously in both eyes for a similar period of time, normal development results: each eye's inputs retain their normal cortical innervation. Such competition appears to yield a roughly constant final total strength of innervation regardless of the patterns of input activity, although the distribution of this innervation among the inputs depends upon neural activities. Evidence of competition for a limited number of synaptic sites exists in many biological systems (*e.g.* Bourgeois et al., 1989; Hayes and Meyer, 1988a,b, 1989a,b; Murray et al., 1982; Pallas and Finlay, 1991).

The existence of constraints limiting synaptic resources is also suggested on theoretical grounds. Development under simple correlation-based rules of synaptic modification typically leads to instability. Either all synapses grow to the maximum allowed value, or all synapses decay to zero strength. To achieve the results found biologically, a Hebbian rule must instead lead to the development of selectivity, so that some synaptic patterns grow in strength while others shrink. Von der Malsburg (1973) proposed the use of constraints conserving the total synaptic strength supported by each input or output cell to achieve selectivity; related proposals were also made by others (Perez et al., 1975; Rochester et al., 1956; Rosenblatt, 1961).

A constraint that conserves total synaptic strength over a cell can be enforced through nonspecific decay of all synaptic strengths, provided the rate of this decay is set for the cell as a whole to cancel the total increase due to specific, Hebbian plasticity. Two simple types of decay can be considered. First, each synapse might decay at a rate proportional to its current strength; this is called multiplicative decay. Alternatively, each synapse might decay at a fixed rate, independent of its strength; this is called subtractive decay. The message of this paper is that the dynamical effects of a constraint depend significantly on whether it is enforced via multiplicative or subtractive decay. We have noted this briefly in previous work (MacKay and Miller, 1990a,b; Miller, 1990a; Miller et. al, 1989).

## 1 Simple Examples of the Effects of Constraints

A few simple examples will illustrate that strikingly different outcomes can result from the subtractive or multiplicative enforcement of a constraint. The remainder of the paper will present a systematic analysis of these differences.

Consider synaptic plasticity of a single postsynaptic cell. Let $\mathbf{w}$ be the vector of synaptic weights onto this cell; the $i^{\text{th}}$ component, $w_i$, is the synaptic weight from the $i^{\text{th}}$ input. We assume synaptic weights are initially randomly distributed with mean $w_{\text{init}}$, and are limited to remain between a maximum value $w_{\text{max}}$ and minimum value $w_{\text{min}}$. We consider the effect of a constraint that conserves the total synaptic strength, $\sum_i w_i$, implemented either multiplicatively or subtractively.

Consider first a simple equation for Hebbian synaptic plasticity, $\frac{d}{dt}\mathbf{w} = \mathbf{C}\mathbf{w}$, where $\mathbf{C}$ is a matrix describing correlations among input activities (derived for example in MacKay and Miller, 1990a). Suppose this correlation is a Gaussian function of the separation of two inputs. We assume first that $w_{\text{min}} = 0$. The final outcomes of development under this equation are shown in Fig. 1A. With no constraint (column 1), all synapses saturate at $w_{\text{max}}$, so all selectivity in the cell's response is lost. Under a multiplicative constraint (column 2), synaptic strengths decrease gradually from center to periphery. The final synaptic pattern in this case is proportional to the principal eigenvector of $\mathbf{C}$. Under a subtractive constraint (columns 3 and 4), a central core of synapses saturate at or near strength $w_{\text{max}}$, while the remaining synapses saturate at $w_{\text{min}}$. If $w_{\text{max}}$ is increased, or the total conserved synaptic strength $w_{\text{tot}}$ is decreased by decreasing $w_{\text{init}}$, the receptive field is sharpened (column 4). In contrast, under the multiplicative constraint or without constraints, the shape of the final receptive field is unaltered by such changes in $w_{\text{max}}$ and $w_{\text{tot}}$.

This sharpening of the receptive field under subtractive constraints occurs because all synapses saturate, so the final number of nonzero synapses is approximately $w_{\text{tot}}/w_{\text{max}}$. Such sharpening under subtractive constraints can create a precise match between two spatial maps, for example maps of auditory and of visual space, despite spatially broad correlations between the two maps (Miller and MacKay, 1992).

If $w_{\text{min}}$ is decreased below zero, center-surround receptive fields can result under subtractive constraints (Fig. 1B). In contrast, the results under multiplicative constraints or unconstrained dynamics are unaltered by this change. This mechanism of developing center-surround receptive fields underlies the results of Linsker (1986), as explained in section 2.5. Again, an increase in $w_{\text{max}}$ or decrease in $w_{\text{tot}}$ leads to sharpening of the positive part of the receptive field under subtractive constraints (column 4).

Finally, consider ocular dominance segregation (Miller et al., 1989) (Fig. 1C). We suppose the output cell receives two equivalent sets of inputs: left-eye inputs and right-eye inputs. A Gaussian correlation function $\mathbf{C}$ describes correlations within each eye as before, while between the two eyes there is zero correlation; and $w_{\text{min}} = 0$. Now results are much as in (A), but a new distinction emerges. Under subtractive constraints, ocular dominance segregation occurs: the output cell

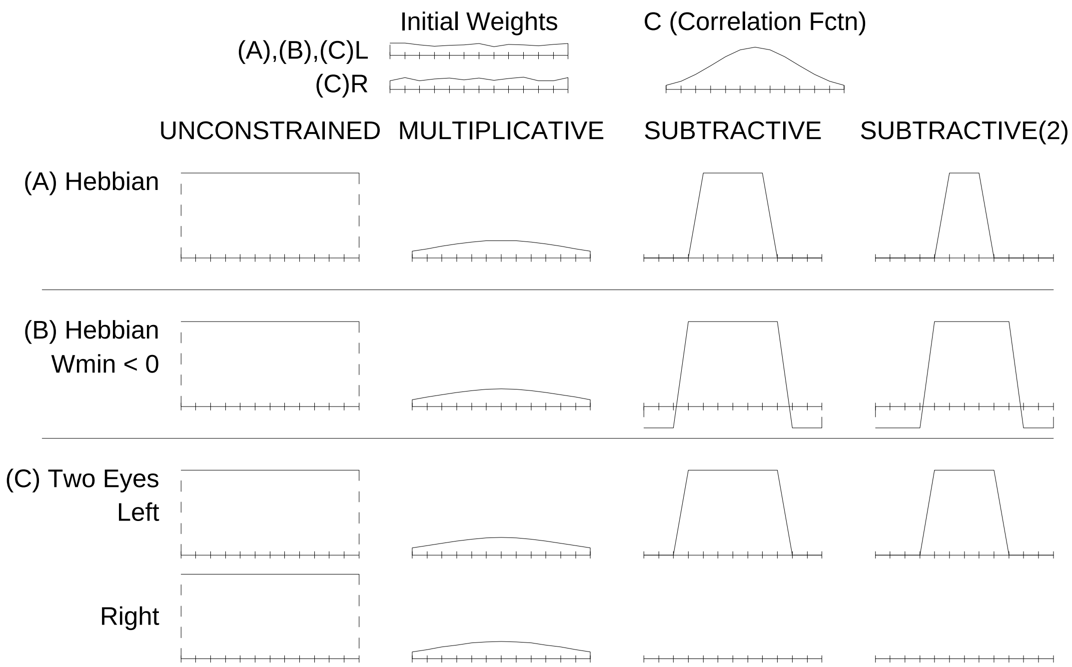


Figure 1: **Outcomes of Development Without Constraints and Under Multiplicative and Subtractive Constraints**
(A), (B): Outcome of a simple Hebbian development equation: unconstrained equation is $\frac{d}{dt}\mathbf{w} = \mathbf{C}\mathbf{w}$. Initial synaptic weights are shown at the top left. The correlation matrix $\mathbf{C}$ is a Gaussian function of the separation between two synapses (shown at top right). (A): $w_{\min} = 0$; (B): $w_{\min} = -2$. (C): Outcome of a similar equation but with two identical sets of inputs, representing left- and right-eye inputs. Within each eye, correlations are the same as in (A); between the eyes there is zero correlation. Unconstrained equations are $\frac{d}{dt}\mathbf{w}^L = \mathbf{C}\mathbf{w}^L$; $\frac{d}{dt}\mathbf{w}^R = \mathbf{C}\mathbf{w}^R$.
All results are from simulations of a two-dimensional receptive field consisting of a diameter-13 circle of inputs drawn from a 13 by 13 square grid. The resulting receptive fields were approximately circularly symmetric; the figures show a slice horizontally through the center of the field. All simulations used $w_{\max} = 8$; all except (B) used $w_{\min} = 0$. The left three columns show results for $w_{\text{init}} = 1$. The right column (subtractive(2)) uses $w_{\text{init}} = 0.5$, which halves the conserved total synaptic strength $w_{\text{tot}}$.

becomes monocular, receiving input from only a single eye. Under multiplicative constraints, there is no ocular dominance segregation: the two eyes develop equal innervations to the output cell. Segregation can only occur under multiplicative constraints if there are anticorrelations between the two eyes, as will be explained in section 2.6.

In summary, unconstrained Hebbian equations often lead all synapses to saturate at the maximal allowed value, destroying selectivity. Multiplicative constraints then lead the inputs to develop graded strengths. Subtractive constraints lead synapses to saturate at either the maximal or minimal allowed value, and can result in a sharpening to a few best-correlated inputs. They also can allow ocular dominance segregation to develop in circumstances where multiplicative constraints do not. These differences between subtractive and multiplicative constraints are easily understood, as we now show.

# 2 Multiplicative and Subtractive Constraints for a Single Output Cell

We begin with a general linear synaptic plasticity equation without decays, $\frac{d}{dt}\mathbf{w}(t) = \mathbf{C}\mathbf{w}(t)$. We assume that the matrix $\mathbf{C}$ is symmetric: in Hebbian learning, $C_{ij}$ represents the correlation in activity between inputs $i$ and $j$, so $C_{ij} = C_{ji}$.[1] Thus, $\mathbf{C}$ has a complete set of orthonormal eigenvectors $\mathbf{e}^a$ with corresponding eigenvalues $\lambda^a$ (that is, $\mathbf{C}\mathbf{e}^a = \lambda^a \mathbf{e}^a$). Typically most or all of the eigenvalues of $\mathbf{C}$ are positive; for example, if $\mathbf{C}$ is the covariance matrix of the input activities then all its eigenvalues are positive.

We use indices $i, j$ to refer to the synaptic basis, and $a, b$ to refer to the eigenvector basis. The strength of the $i^{\text{th}}$ synapse is denoted by $w_i$. The weight vector $\mathbf{w}$ can also be written as a combination of the eigenvectors, $\mathbf{w} = \sum_a w_a \mathbf{e}^a$, where the components of $\mathbf{w}$ in the eigenvector basis are $w_a = \mathbf{w} \cdot \mathbf{e}^a$. We assume as before that the dynamics are linear up to hard limits on the synaptic weights, $w_{\min} \leq w_i(t) \leq w_{\max}$; we will not explicitly note these limits in subsequent equations.

## 2.1 Formulation of multiplicative and subtractive constraints

By a multiplicative or subtractive constraint, respectively, we refer to a time-varying decay term $\gamma(t)\mathbf{w}$ or $\epsilon(t)\mathbf{n}$ that moves $\mathbf{w}$, after application of $\mathbf{C}$, toward a constraint surface. We assume $\gamma$ and $\epsilon$ are determined by the current weight vector $\mathbf{w}(t)$ and do not otherwise depend on $t$, so we

[1] We work in a representation in which each synapse is represented explicitly, and the density of synapses is implicit. Equivalently, one may use a representation in which the synaptic density or "arbor" function is explicit (MacKay and Miller, 1990a, App. A). Then, although the equation governing synaptic development may appear nonsymmetric, it can be symmetrized by a coordinate transformation. Thus, the present analysis also applies in these representations, as further described in (Miller and MacKay, 1992, App. B).

| | |
|---|---|
| **Type 1 constraint** | Conserves the total synaptic strength, $\sum_i w_i = \mathbf{w}\cdot\mathbf{n}$. |
| Zero-sum vector | A vector $\mathbf{w}$ with zero total synaptic strength: $\sum_i w_i = \mathbf{w}\cdot\mathbf{n} = 0$. |
| M1 | Multiplicatively enforced type 1 constraint |
| S1 | Subtractively enforced type 1 constraint, using perpendicular projection onto the constraint surface |

| | |
|---|---|
| **Type 2 constraint** | Conserves the length of the weight vector, $\sum_i w_i^2 = \mathbf{w}\cdot\mathbf{w}$. |
| M2 | Multiplicatively enforced type 2 constraint |

Table 1: **Abbreviations Used.** The typical case $\mathbf{n} = (1,1,\ldots,1)^{\mathrm{T}}$ is used to describe Type 1 constraints.

write them as $\gamma(\mathbf{w})$ or $\epsilon(\mathbf{w})$. Thus, the constrained equations are[2]

$$\frac{d}{dt}\mathbf{w}(t) = \mathbf{C}\mathbf{w}(t) - \gamma(\mathbf{w})\mathbf{w}(t) \qquad \text{(Multiplicative Constraint)} \tag{1}$$

$$\frac{d}{dt}\mathbf{w}(t) = \mathbf{C}\mathbf{w}(t) - \epsilon(\mathbf{w})\mathbf{n} \qquad \text{(Subtractive Constraint)} \tag{2}$$

The vector $\mathbf{n}$ is a constant. Typically, all synapses have equal subtractive decay rate, so $\mathbf{n} = (1,1,\ldots,1)^{\mathrm{T}}$ in the synaptic basis.

Multiplicative or subtractive constraints represent two methods of *enforcing* a constraint, that is, of maintaining the weight vector on some constraint surface. We now consider the type of constraint to be enforced. We will focus on two types. First, a constraint may conserve the total synaptic strength $\sum_i w_i$, as in section 1. We refer to this as a type 1 constraint, and to a multiplicative or subtractive constraint of this type as M1 or S1 respectively. These are frequently used in modeling studies (*e.g.* M1: Grajski and Merzenich, 1990; von der Malsburg, 1973, 1979; von der Malsburg and Willshaw, 1976; Perez et. al, 1975; Rochester et. al, 1956; Whitelaw and Cowan, 1981; Willshaw and von der Malsburg, 1976, 1979; S1: Miller, 1992; Miller et. al, 1989).[3] We define a type 1 constraint more generally as one that conserves the total weighted synaptic strength, $\sum_i w_i n_i = \mathbf{w}\cdot\mathbf{n}$, where $\mathbf{n}$ is a constant vector. Typically, $\mathbf{n} = (1,1,\ldots,1)^{\mathrm{T}}$ and the conserved quantity is the total synaptic strength. A type 1 constraint corresponds to a hyperplane constraint surface.

For an S1 constraint, we choose the subtracted vector $\mathbf{n}$ in Eq. 2 to be the same as this constraint vector $\mathbf{n}$. This means we consider only subtractive constraints that project perpendicularly onto

[2]To understand why the term $-\gamma(\mathbf{w})\mathbf{w}(t)$ represents a multiplicative constraint, consider a multiplicatively constrained equation $\mathbf{w}(t+\Delta t) = \beta(\mathbf{w})\,[\mathbf{w}(t) + \mathbf{C}\mathbf{w}(t)\Delta t]$, where $\beta(\mathbf{w})$ achieves the constraint. This is identical to $\frac{\mathbf{w}(t+\Delta t)-\mathbf{w}(t)}{\Delta t} = \mathbf{C}\mathbf{w}(t) - \gamma(\mathbf{w})\mathbf{w}(t+\Delta t)$ where $\gamma(\mathbf{w}) = \frac{1-\beta(\mathbf{w})}{\beta(\mathbf{w})\Delta t}$. For $\Delta t \to 0$ this becomes Eq. 1.

[3]Many of these models used nonlinearities other than hypercube limits on synaptic weights. Our results nonetheless appear to correctly characterize the outcomes in these models.

the constraint surface. Then type 1 constraints can be achieved by choosing

$$\text{M1:} \qquad \gamma(\mathbf{w}) = \mathbf{n}\cdot\mathbf{Cw}/\mathbf{n}\cdot\mathbf{w} \qquad (\text{with } \mathbf{n}\cdot\mathbf{w}(t{=}0) \neq 0) \tag{3}$$

$$\text{S1:} \qquad \epsilon(\mathbf{w}) = \mathbf{n}\cdot\mathbf{Cw}/\mathbf{n}\cdot\mathbf{n} \tag{4}$$

These choices yield $\mathbf{n}\cdot\frac{d}{dt}\mathbf{w} = \frac{d}{dt}(\mathbf{n}\cdot\mathbf{w}) = 0$ under Eqs. 1 or 2 respectively.

Second, we consider a constraint that conserves the sum–squared synaptic strength, $\sum_i w_i^2 = \mathbf{w}\cdot\mathbf{w}$. This corresponds to a hypersphere constraint surface. We refer to this as a type 2 constraint (the numbers '1' and '2' refer to the exponent $p$ in the constrained quantity $\sum w_i^p$). This constraint, while not biologically motivated, is often used in theoretical studies, *e.g.* (Kohonen, 1989; Oja, 1982). We will consider only multiplicative enforcement of this constraint,[4] called M2. M2 can be achieved by choosing

$$\text{M2:} \qquad \gamma(\mathbf{w}) = \mathbf{w}\cdot\mathbf{Cw}/\mathbf{w}\cdot\mathbf{w} \tag{5}$$

This yields $2\mathbf{w}\cdot\frac{d}{dt}\mathbf{w} = \frac{d}{dt}(\mathbf{w}\cdot\mathbf{w}) = 0$ under Eq. 1.

The abbreviations introduced in this section are summarized in Table 2.1.

## 2.2 Projection operators

Each form of constrained dynamics can be written $\frac{d}{dt}\mathbf{w} = \mathbf{PCw}$, where $\mathbf{P}$ is a projection operator that projects the unconstrained dynamics onto the constraint surface. For S1, the projection operator is $\mathbf{P} = \mathbf{1} - \frac{\mathbf{nn}^{\mathrm{T}}}{\mathbf{n}\cdot\mathbf{n}}$; for M1, it is $\mathbf{P} = \mathbf{1} - \frac{\mathbf{wn}^{\mathrm{T}}}{\mathbf{w}\cdot\mathbf{n}}$; and for M2, it is $\mathbf{P} = \mathbf{1} - \frac{\mathbf{ww}^{\mathrm{T}}}{\mathbf{w}\cdot\mathbf{w}}$. We can write these operators as $\mathbf{P} = \mathbf{1} - \frac{\mathbf{sc}^{\mathrm{T}}}{\mathbf{s}\cdot\mathbf{c}}$, where $\mathbf{s}$ is the *subtracted* vector, $\mathbf{c}$ the *constraint* vector, and $\mathbf{1}$ the identity matrix (Fig. 2). The projection operator removes the $\mathbf{c}$ component of the unconstrained derivative $\mathbf{Cw}$, through subtraction of a multiple of $\mathbf{s}$. Thus, the subtracted vector $\mathbf{s}$ represents the method of constraint enforcement: $\mathbf{s} = \mathbf{w}$ for multiplicative constraints, while $\mathbf{s} = \mathbf{n}$ for subtractive constraints. The constraint vector $\mathbf{c}$ determines the constraint that is enforced: the dynamics remain on the constraint surface $\mathbf{w}\cdot\mathbf{c} = \text{constant}$.

Given a constraint surface, there are two "natural" methods of constraint enforcement: projection perpendicular to the surface ($\mathbf{s} = \mathbf{c}$), or projection towards the origin ($\mathbf{s} = \mathbf{w}$). For a type 1 constraint, these lead to different dynamics: S1 is perpendicular projection, while M1 is projection along $\mathbf{w}$. For a type 2 constraint, these are identical: M2 is both perpendicular projection and projection along $\mathbf{w}$.

[4]Subtractive enforcement, S2, does not work in the typical case in which the fixed points are unstable. The constraint fails where $\mathbf{n}$ is tangent to the constraint hypersphere, *i.e.* at points where $\mathbf{w}\cdot\mathbf{n} = 0$. Such points form a circumference around the hypersphere. The S2 dynamics flow away from the unstable fixed points, at opposite poles of the hypersphere, and flow into this circumference unless prevented by the bounds on synaptic weights.

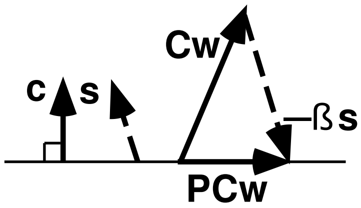


Figure 2: **Projection Onto the Constraint Surface**
The projection operator is $\mathbf{P} = \mathbf{1} - \frac{\mathbf{s}\mathbf{c}^{\mathrm{T}}}{\mathbf{s}\cdot\mathbf{c}}$. This acts on the unconstrained derivative $\mathbf{Cw}$ by removing its $\mathbf{c}$ component, projecting the dynamics onto the constraint surface $\mathbf{c} \cdot \mathbf{PCw} = \mathbf{c} \cdot \frac{d}{dt}\mathbf{w} = 0$. This constraint surface is shown as the line perpendicular to $\mathbf{c}$. The constraint is enforced through subtraction of a multiple of $\mathbf{s}$: $\mathbf{PCw} = \mathbf{Cw} - \beta\mathbf{s}$ where $\beta = \mathbf{c} \cdot \mathbf{Cw}/\mathbf{c} \cdot \mathbf{s}$. For multiplicative constraints, $\mathbf{s} = \mathbf{w}$; for subtractive S1 constraints, $\mathbf{s} = \mathbf{c} = \mathbf{n}$.

### 2.3 Dynamical effects of multiplicative and subtractive constraints

In this section, we characterize the dynamics under M1, S1, and M2 constraints. In section 2.3.1, we demonstrate that under S1, the dynamics typically have no stable fixed point, and flow until all synapses are saturated; while under multiplicative constraints, the principal eigenvector $\mathbf{e}^0$ of $\mathbf{C}$ is a stable fixed point of the dynamics, provided that it satisfies the constraint. In section 2.3.2, we characterize the conditions under which multiplicatively constrained dynamics flow to the principal eigenvector fixed point. In section 2.3.3, we characterize the outcome under S1 constraints in terms of the eigenvectors of $\mathbf{C}$.

We begin by illustrating in Fig. 3 the typical dynamics under M1, S1, and M2 in the plane formed by the principal eigenvector of $\mathbf{C}$, $\mathbf{e}^0$, and one other eigenvector with positive eigenvalue, $\mathbf{e}^1$. Figure 3 illustrates the main conclusions of this section, and may be taken as a visual aid for the remainder:

- In Fig. 3A we illustrate M1 and S1 dynamics in the case in which the principal eigenvector $\mathbf{e}^0$ is close in direction to the constraint vector $\mathbf{n}$ ($\mathbf{n}$ is the vector perpendicular to the constraint surface). This is typical for Hebbian learning when there are only positive correlations and the total synaptic sum is conserved, as in the examples of section 1. Positive correlations lead to a principal eigenvector in which all weights have a single sign; this is close in direction to the usual constraint vector $(1, 1, \ldots, 1)^{\mathrm{T}}$, which conserves total synaptic strength. In this case, growth of $\mathbf{e}^0$ would violate the constraint. Multiplicative and subtractive constraints lead to very different outcomes: multiplicative constraints lead to convergence to $\mathbf{e}^0$, whereas subtractive constraints lead to unstable flow in a direction perpendicular to $\mathbf{n}$. The outcome in this case was illustrated in Figs. 1A,B.

- In Fig. 3B we illustrate M1 and S1 dynamics in the case in which the principal eigenvector

$\mathbf{e}^0$ is parallel to the constraint surface: $\mathbf{e}^0 \cdot \mathbf{n} = 0$. We call such vectors $\mathbf{w}$, for which $\mathbf{w} \cdot \mathbf{n} = 0$, *zero-sum* vectors. Growth of a zero-sum vector does not violate the type 1 constraint. For practical purposes, any vector that is approximately parallel to the constraint surface, so that it intersects the surface far outside the hypercube that limits synaptic weights, may be treated as a zero-sum vector. The principal eigenvector is typically a zero-sum vector in Hebbian learning when correlations among input activities oscillate in sign as a function of input separation (Miller, 1990a). Such oscillations lead to a principal eigenvector in which weights oscillate in sign, and sum approximately to zero; such a vector is approximately perpendicular to the constraint vector $(1, 1, \ldots, 1)^{\mathrm{T}}$.
In this case, growth of $\mathbf{e}^0$ does not violate the constraint. The type of constraint enforcement makes little difference: the weight vector typically flows to a saturated version of $\mathbf{e}^0$.

- Under M2 constraints (Fig. 3C), the principal eigenvector $\mathbf{e}^0$ is always perpendicular to the constraint surface, and its growth would always violate the constraint. The dynamics converge to $\mathbf{e}^0$.

### 2.3.1 General differences between multiplicative and subtractive constraint enforcement: Fixed points and stability

We now establish the essential difference between multiplicative and subtractive constraints. To do so, we examine the locations and stability of the fixed points that are in the interior of the hypercube of allowed synaptic weights ("interior fixed points"). A fixed point $\mathbf{w}^{\mathrm{FP}}$ is a point where the flow $\frac{d}{dt}\mathbf{w} = 0$. For $\mathbf{C}$ symmetric, the constrained dynamics must either flow to a stable interior fixed point or else flow to the hypercube.

We will show that the only stable fixed point of multiplicatively constrained dynamics is the intersection of the principal eigenvector $\mathbf{e}^0$ of $\mathbf{C}$ with the constraint surface; and that if $\mathbf{e}^0$ intersects the constraint surface, then the dynamics typically converge to $\mathbf{e}^0$, as was illustrated in section 1. Under subtractive S1 constraints, there is generally no stable fixed point within the hypercube. S1 dynamics typically are only stabilized once all synapses (or all but one) reach saturation at $w_{\max}$ or $w_{\min}$.

The locations of the interior fixed points follow trivially from Eqs. 1-2 and the fact that the dynamics remain on the constraint surface:

- The fixed points under a multiplicatively enforced constraint are the intersections of the eigenvectors of $\mathbf{C}$ with the constraint surface, that is, the points $\mathbf{w}$ on the constraint surface at which $\mathbf{C}\mathbf{w} \propto \mathbf{w}$.

- The fixed points under a subtractively enforced constraint are the points $\mathbf{w}$ on the constraint surface at which $\mathbf{C}\mathbf{w} \propto \mathbf{n}$.

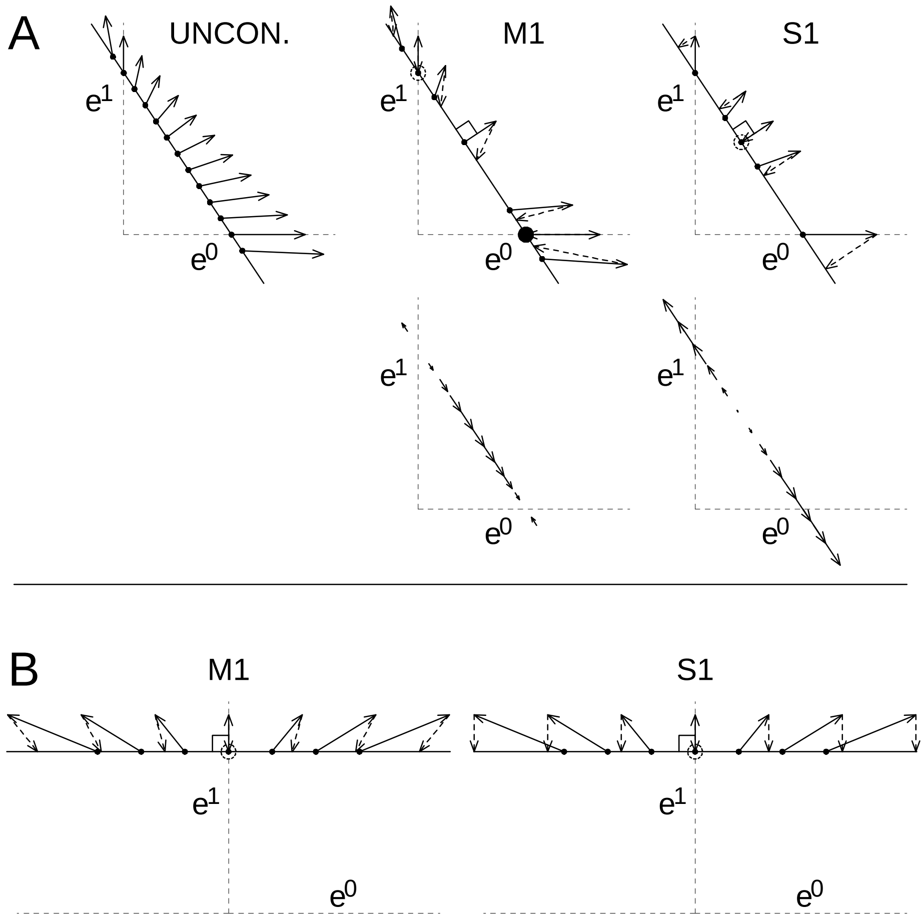


Figure 3: **Dynamics Under Multiplicative and Subtractive Constraints**
Dynamics in the plane formed by the principal eigenvector of $\mathbf{C}$, $\mathbf{e}^0$, and one other eigenvector with positive eigenvalue, $\mathbf{e}^1$.
**A. M1 and S1 constraints when $e^0$ is close in direction to $\mathbf{n}$.** Diagonal lines indicate the constraint surface on which $\mathbf{n} \cdot \mathbf{w}$ is constant. Unconstrained: arrows show the unconstrained derivative $\mathbf{Cw}$ from the point $\mathbf{w}$ at the base of the arrow. M1: Solid arrows show the unconstrained flow; dashed arrows show the return path to the constraint surface (as in Fig. 2). Return path is in the direction $\mathbf{w}$. Open circle indicates unstable fixed point, large filled circle indicates stable fixed point. The fixed points are the eigenvectors, where $\mathbf{Cw} \propto \mathbf{w}$. S1: Return path is in the direction $\mathbf{n}$. The fixed point is the point where $\mathbf{Cw} \propto \mathbf{n}$ (indicated by perpendicular symbol), and is unstable. 2nd Row: The resulting constrained flow along the constraint surface for M1 and S1.
**B. M1 and S1 constraints when $e^0$ is perpendicular to $\mathbf{n}$.** The constraint surface does not intersect $\mathbf{e}^0$. M1 and S1 lead to similar outcomes: unstable growth occurs, predominantly in the $\mathbf{e}^0$ direction, until the hypercube that limits synaptic weights is reached. The outcome is expected to be a saturated version of $\pm\mathbf{e}^0$. Note that the unconstrained dynamics also flow predominantly in the $\pm\mathbf{e}^0$ direction and so should lead to a similar outcome. For convenience, we have chosen the constraint direction $\mathbf{n}=\mathbf{e}^1$.

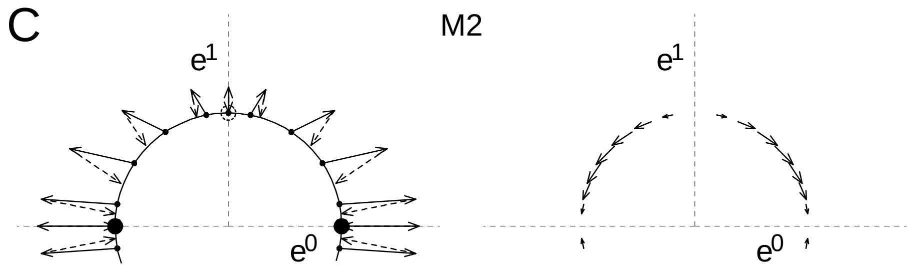


Figure 3: (continued from previous page)
**C. M2 constraints.** The return path is in the direction of $\mathbf{w}$, as for M1. Thus, locally (for example, near the fixed points) the dynamics are like M1. On a large scale, the dynamics differ because of the difference in constraint surface. Left, unconstrained derivative and return path; right, constrained flow. Figures were drawn using eigenvalues $\lambda^0/\lambda^1 = 3$, constraint vector in A $\mathbf{n} \cdot \mathbf{e}^0/\mathbf{n} \cdot \mathbf{e}^1 = 1.5$.

The stability of a fixed point can be determined as shown in Fig. 3A, by determining whether a point perturbed from the fixed point is taken further away by the dynamics. Generalizing the reasoning illustrated there, it is easy to prove (Appendix A):

**Theorem 1** *Under a multiplicatively enforced constraint, if the principal eigenvector of* $\mathbf{C}$ *is an interior fixed point it is stable. Interior fixed points that are non-principal eigenvectors are unstable.*

**Theorem 2** *Under an S1 constraint, if* $\mathbf{C}$ *has at least two eigenvectors with positive eigenvalues, then any interior fixed point is unstable.*

A case of theorem 1 for M2 constraints was proven by Oja (1982).

Theorem 2 shows that S1 dynamics are unstable when no synapse is saturated. If in addition the following condition holds, as occurs when $\mathbf{C}$ is a correlation matrix, then these dynamics remain unstable until all synapses have saturated (Appendix A):

**Theorem 3** *Let $i$ and $j$ be indices in the synaptic basis. Suppose that for all $i$ and $j$ with $i \neq j$, $C_{ii} > |C_{ij}|$. Then under an S1 constraint, either all synapses or all but one are saturated in a stable final condition.*

This condition is satisfied for Hebbian models, because $C_{ij}$ represents the correlation in activities between input $i$ and input $j$. The result is sharply different from that for multiplicative constraints: in that case, the principal eigenvector may be a stable fixed point with no synapse saturated. This theorem generalizes a result proven for a particular model by Linsker (1986).

Theorem 3 explains the sharpening of the receptive field that occurs under an S1 constraint (Fig. 1). A practical implication is that an upper limit on synaptic strengths, $w_{\max}$, is needed for stability under an S1 constraint (whereas no such limit is needed under a multiplicative constraint). If there is no upper synaptic limit, eventually one synapse will acquire all of the allowed synaptic strength, while all other synapses will become saturated at $w_{\min}$.

### 2.3.2 The outcome under multiplicative constraints

From the previous section, we conclude that a multiplicatively enforced constraint results in convergence to the principal eigenvector $\mathbf{e}^0$ of $\mathbf{C}$ provided: (1) The principal eigenvector intersects the constraint surface within the hypercube of allowed weights, forming a stable fixed point; and (2) The initial weight vector is within the basin of attraction of this fixed point. We ignore the possible effect of the hypercube, and assess when these conditions are met.

Under M2 constraints, both of these conditions are always satisfied (Fig. 3C), so M2 constraints always lead to convergence.

Under M1 constraints, condition (1) is satisfied when $\mathbf{e}^0$ is non-zero-sum (Fig. 3A). Then, for $\mathbf{n} = (1, 1, \ldots, 1)^{\mathrm{T}}$, condition (2) is also satisfied in at least two typical cases: if both the initial weight vector and the principal eigenvector have no changes in sign, or if weights are initialized as small fluctuations about a nonzero mean.[5] Thus, M1 constraints typically converge to $\mathbf{e}^0$ when $\mathbf{e}^0$ is non-zero-sum.

### 2.3.3 The outcome under S1 constraints

S1 constraints lead to altered, linear dynamics. To see this, write the subtractively constrained equation as $\frac{d}{dt}\mathbf{w}(t) = \mathbf{PCw}(t)$ with $\mathbf{P} = \mathbf{1} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\mathrm{T}}$; here, $\hat{\mathbf{n}} \equiv \mathbf{n}/|\mathbf{n}|$. Write $\mathbf{w}$ as the sum $\mathbf{w}(t) = \mathbf{Pw}(t) + w_n\hat{\mathbf{n}}$, where $w_n \equiv \mathbf{w} \cdot \hat{\mathbf{n}}$ is conserved. Then the dynamics can be written:

$$\frac{d}{dt}\mathbf{w}(t) = \mathbf{PCPw}(t) + w_n\mathbf{PC}\hat{\mathbf{n}}. \tag{6}$$

$\mathbf{PCP}$ is the operator $\mathbf{C}$, restricted to the subspace of zero-sum vectors. $w_n\mathbf{PC}\hat{\mathbf{n}}$ is a constant vector.

Thus, S1 constraints lead to linear dynamics driven by $\mathbf{PCP}$ rather than by $\mathbf{C}$. These dynamics have been characterized in (MacKay and Miller, 1990a, Appendices B and E). To understand them, consider first the eigenvectors of $\mathbf{PCP}$. These are of two types:

1. Any zero-sum eigenvector of $\mathbf{C}$ is also an eigenvector of $\mathbf{PCP}$ with identical eigenvalue. So zero-sum eigenvectors of $\mathbf{C}$ grow freely, at the same rate as they would in the absence of constraints.

2. Each non-zero-sum eigenvector of $\mathbf{C}$ is replaced by a corresponding zero-sum eigenvector of $\mathbf{PCP}$ with smaller eigenvalue:[6] for example, an all-positive, centrally peaked non-zero-sum

[5]Let $\beta\mathbf{e}^0$ be the stable fixed point, and let $\mathbf{w}_0$ be the initial weight vector on the constraint surface. Condition (2) is satisfied if $\mathbf{w}_0 \cdot (\beta\mathbf{e}^0) > 0$. Suppose the constraint conserves $\mathbf{w} \cdot \hat{\mathbf{n}} = w_n$, so that $\mathbf{w}_0 \cdot \hat{\mathbf{n}} = \beta\mathbf{e}^0 \cdot \hat{\mathbf{n}} = w_n$. Then if $\mathbf{w}_0$ and $\mathbf{e}^0$ are each single-signed, they must have the same sign so $\mathbf{w}_0 \cdot (\beta\mathbf{e}^0) > 0$. If weights are initialized as small fluctuations about a nonzero mean, then $\mathbf{w}_0 \approx w_n\hat{\mathbf{n}}$. so $\mathbf{w}_0 \cdot (\beta\mathbf{e}^0) \approx w_n^2 > 0$.

[6]There is one exception: the non-zero-sum eigenvector of $\mathbf{C}$ with smallest eigenvalue is replaced by $\mathbf{n}$, which is an eigenvector of $\mathbf{PCP}$ with eigenvalue 0.

eigenvector of $\mathbf{C}$ may be replaced by a center-surround (positive center and negative surround) zero-sum eigenvector of $\mathbf{PCP}$. Eigenvalue order of the non-zero-sum eigenvectors is preserved under this correspondence.

Now consider the constant term $w_n\mathbf{PC}\hat{\mathbf{n}}$. This term boosts the growth rate of the eigenvectors of $\mathbf{PCP}$ that compose it. These are the eigenvectors derived from the non-zero-sum eigenvectors of $\mathbf{C}$.

Thus, under S1 constraints, the dynamics may be dominated either by the principal zero-sum eigenvector of $\mathbf{C}$, or by a zero-sum vector that replaces the principal eigenvector. Both vectors may be very different from the principal eigenvector of $\mathbf{C}$.

#### 2.3.4 Summary: The outcome under M1, S1 and M2

Multiplicative and subtractive constraints lead to dramatically different outcomes in many cases. In particular, under a type 1 constraint, multiplicative constraints converge to a dominant non-zero-sum pattern, whereas subtractive constraints suppress such a pattern in favor of a zero-sum pattern. We may summarize as follows:

1. If the principal eigenvector $\mathbf{e}^0$ of $\mathbf{C}$ is a non-zero-sum vector and intersects the constraint surface within the hypercube, as is typical for Hebbian learning when there are only positive correlations, then
   (a) M1 constraints lead to a stabilized version of $\mathbf{e}^0$;
   (b) S1 constraints lead to a zero-sum vector that grows to complete saturation, superimposed on the constrained background $(\mathbf{w} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$. The dominant zero-sum vector may be either
      i. A zero-sum vector derived from $\mathbf{e}^0$, as in Fig. 1A,B; or
      ii. The principal zero-sum eigenvector of $\mathbf{C}$.
2. If the principal eigenvector $\mathbf{e}^0$ of $\mathbf{C}$ is a zero-sum vector, as is typical for Hebbian learning when correlations among input activities oscillate in sign, then a type 1 constraint has little effect; the unconstrained dynamics or M1 or S1 constraints all lead to saturated versions of $\mathbf{e}^0$.
3. M2 constraints always lead to a stabilized version of the principal eigenvector of $\mathbf{C}$, unless the hypercube limiting synaptic weights interferes with the dynamics.

### 2.4 What is maximized under multiplicative and subtractive constraints?

Under multiplicative constraints, the weight vector tends to a multiple of the principal eigenvector $\mathbf{e}^0$ of $\mathbf{C}$. This is the direction in weight space that maximizes $\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{C}\hat{\mathbf{w}}$ over all directions $\hat{\mathbf{w}}$. This maximizes the mutual correlations among the weights; so most mutually correlated inputs are expected to retain representation.

Under S1 constraints, for $\mathbf{C}$ symmetric, the dynamics maximize $E = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{PCPw} + w_n\mathbf{w}^{\mathrm{T}}\mathbf{PCn}$. (section 2.3.3). For $w_n$ sufficiently small, the first term dominates, so the weight vector is dominated by the principal eigenvector $\mathbf{e}^0_{\mathbf{P}}$ of $\mathbf{PCP}$. This is the direction in weight space that maximizes $\hat{\mathbf{w}}^{\mathrm{T}}\mathbf{C}\hat{\mathbf{w}}$ over all *zero-sum* directions $\hat{\mathbf{w}}$. When $\mathbf{n} = (1, 1, \ldots, 1)^{\mathrm{T}}$, $\mathbf{e}^0_{\mathbf{P}}$ is a vector in which some subset of maximally correlated weights are set to positive values, and remaining weights are set to negative values. In the final weight structure, positive weights in $\mathbf{e}^0_{\mathbf{P}}$ tend to $w_{\max}$, while negative weights tend to $w_{\min}$. The receptive field thus becomes sharpened to a subset of maximally correlated inputs.

## 2.5 Application to simple Hebbian learning, including Linsker's simulations

The results just derived explain the outcome for simple Hebbian learning with a positive correlation function (Fig. 1A,B). M1 constraints lead to convergence to the principal eigenvector $\mathbf{e}^0$, which is all-positive. S1 constraints instead lead to growth of a zero-sum vector; in Fig. 1A,B, this vector is a center-surround vector derived from $\mathbf{e}^0$.

The results of section 2.3.3 explain some of the results found by Linsker (1986). He explored Hebbian dynamics under S1 constraints with a Gaussian correlation function, using $w_{\min} = -w_{\max}$ and a spatially Gaussian distribution of inputs. Then, as analyzed in (MacKay and Miller, 1990a,b), the leading eigenvector of $\mathbf{C}$ is an all-positive eigenvector, and the zero-sum vector derived from this is a center-surround vector. The leading zero-sum eigenvectors of $\mathbf{C}$, and leading eigenvectors of the constrained operator $\mathbf{PCP}$, are two vectors that are bilobed, half positive and half negative. The all-positive eigenvector dominates the unconstrained development. For small values of $w_n$, the bilobed vectors dominate the constrained development. For larger values of $w_n$, the contribution of the $w_n\mathbf{PC}\hat{\mathbf{n}}$ term to the growth of the center-surround vector allows the center-surround vector to dominate the constrained development within the hypercube, despite its having a smaller eigenvalue than the bilobed vectors under $\mathbf{PCP}$.

## 2.6 Extension to two input layers

We now consider the case in which two equivalent input layers innervate a common output cell (Fig. 1C). For example, in the visual system, inputs serving the left eye and right eye each project to the visual cortex, in an initially completely overlapping manner (Miller et al., 1989). Similarly, ON-center and OFF-center cells make initially equivalent projections to visual cortex (Miller, 1992).

Let $\mathbf{w}^1$, $\mathbf{w}^2$ respectively be the synaptic weight vector from each input projection. Define the sum, $\mathbf{w}^{\mathrm{S}} \equiv \mathbf{w}^1 + \mathbf{w}^2$, and the difference, $\mathbf{w}^{\mathrm{D}} \equiv \mathbf{w}^1 - \mathbf{w}^2$. Because of the symmetry between the two input layers, the eigenvectors of the unconstrained equation can be divided into sum eigenvectors,

$\mathbf{w}^{\rm S} = \mathbf{e}_a^{\rm S}$, $\mathbf{w}^{\rm D} = \mathbf{0}$, with eigenvalues $\lambda_a^{\rm S}$; and difference eigenvectors, $\mathbf{w}^{\rm D} = \mathbf{e}_a^{\rm D}$, $\mathbf{w}^{\rm S} = \mathbf{0}$, with eigenvalues $\lambda_a^{\rm D}$ (Miller, 1990a).

Now an additional critical distinction emerges between subtractive and multiplicative constraints. A type 1 constraint conserves the total synaptic strength $\sum_i w_i^{\rm S}$. Patterns of $\mathbf{w}^{\rm D}$ have zero total synaptic strength, *i.e.* are zero-sum vectors. Therefore, $\mathbf{w}^{\rm D}$ grows freely under an S1 constraint, whereas under a multiplicative constraint growth of $\mathbf{w}^{\rm D}$ is suppressed unless a difference eigenvector is the principal eigenvector of the unconstrained development.

In models of ocular dominance segregation, the principal eigenvector is typically determined as follows (Miller, 1990a). Let $\lambda_{\max}^{\rm S}$ and $\lambda_{\max}^{\rm D}$ be the largest sum and difference eigenvalues respectively. If there are positive correlations between the activities of the two eyes, then $\lambda_{\max}^{\rm S} > \lambda_{\max}^{\rm D}$. If there are no between-eye correlations, then $\lambda_{\max}^{\rm S} = \lambda_{\max}^{\rm D}$. If these correlations are negative, then $\lambda_{\max}^{\rm D} > \lambda_{\max}^{\rm S}$.

Thus, under a multiplicative constraint, $\mathbf{w}^{\rm D}$ cannot grow, and ocular dominance segregation cannot occur, unless the two eyes are negatively correlated[7] (Miller et al., 1989). Such anticorrelations could be produced by intralaminar inhibition within the LGN. However, it seems unlikely that ocular dominance segregation depends on anticorrelations, since ocular dominance develops in the presence of vision in some animals, and vision should partially correlate the two eyes.

Under an S1 constraint, $\mathbf{w}^{\rm D}$ will grow if $\lambda_{\max}^{\rm D} > 0$. The dynamics under an S1 constraint may be dominated either by $\mathbf{e}_{\max}^{\rm D}$, or by the zero-sum vector that derives from $\mathbf{e}_{\max}^{\rm S}$, depending on which has the faster growth rate (section 2.3.3). In practice, ocular dominance segregation develops under an S1 constraint even if there are positive between-eye correlations of moderate size relative to the within-eye correlations (unpublished observations). Thus, subtractive rather than multiplicative enforcement of constraints appears more appropriate for modeling Hebbian development in visual cortex.

# 3 Constraints Given a Full Layer of Output Cells

When modeling Hebbian learning in a full layer of output cells, two differences arise compared to the case of an isolated cell. First, constraints may be applied to the total innervation onto each output cell (M1: Grajski and Merzenich, 1990; von der Malsburg, 1973; von der Malsburg and Willshaw, 1976; Willshaw and von der Malsburg, 1976; S1: Miller, 1992); or to the total innervation from each input cell (M1: von der Malsburg, 1979; Willshaw and von der Malsburg, 1979); or to both (M1: Whitelaw and Cowan, 1981; S1: Miller et. al, 1989). Second, there is usually coupling between the weight changes on different output cells. For example neighboring cells' activities may interact through intralaminar synaptic connections, causing the evolution of their weights to be coupled;

[7] If between-eye correlations are zero, then under multiplicative constraints the ratio of the principal eigenvector components, $w_{\max}^{\rm D}/w_{\max}^{\rm S}$, does not change under time development, while all other components are suppressed. Typically this ratio is initially small, so ocular dominance segregation does not occur.

or modulatory factors may diffuse, directly affecting neighboring synapses. Both types of coupling may take the mathematical form of an output layer "lateral interaction function" (Miller, 1990b; Miller et al., 1989). Formulation of constraints in the case of a full layer is discussed in (Miller and MacKay, 1992, Appendix C).

We have not studied constraints over a full layer in detail. However, the following heuristics, based on the single cell studies of section 2, appear to be compatible with the studies cited in the previous paragraph and with unpublished observations. We refer to the projection to a single output cell as a receptive field or RF, and the projection from a single input location as a projective field or PF. The eigenvectors are patterns of weights across the entire network, not just across individual RFs or PFs. In simple Hebbian models, the dominant, fastest-growing patterns can often be characterized as follows.

First, in the absence of constraints, the RFs of the dominant patterns are primarily determined by a particular input correlation function, and the PFs of the dominant patterns are similarly determined by the output layer lateral interaction function (Miller, 1990a). If the correlations are all positive, the RFs have a single sign; if correlations oscillate in sign with input separation, the RFs oscillate in sign with a similar wavelength. A single-signed RF can be regarded as one that oscillates with an infinite wavelength, so we may summarize: in the absence of constraints, the RFs of the dominant patterns vary between positive and negative values with a wavelength corresponding to the peak of the Fourier transform of the appropriate input correlation function. Similarly, the PFs of the dominant patterns vary between positive and negative values with a wavelength corresponding to the peak of the Fourier transform of the output layer lateral interaction function.

Second, constraints on output cells appear only to affect the form of the individual RFs, while constraints on input cells only affect the form of the individual PFs. Consider the case of two layers that are topographically connected: each input cell initially makes synapses onto cells over a certain diameter ("arbor diameter") in the output layer, and adjacent input cells project adjacent arbors. Then output cells also receive connections over an arbor diameter from the input layer. Suppose that output or input cell constraints conserve total synaptic strength over the cell. Then an RF or PF that alternates in sign with a wavelength less than or equal to the arbor diameter is approximately zero-sum, that is, it has summed synaptic strength near 0. An RF or PF that alternates with longer wavelength is non-zero-sum. Subtractive constraints selectively suppress the growth of non-zero-sum patterns, whereas multiplicative constraints stabilize the growth of a dominant non-zero-sum pattern.

Thus, we arrive at the following heuristic rules for the wavelength with which RFs or PFs alternate in sign (Fig. 4):

1. If the dominant pattern in the absence of constraints has (RF,PF) wavelength larger than an arbor diameter, then

   (a) Subtractive (output,input) constraints suppress this pattern in favor of a pattern with

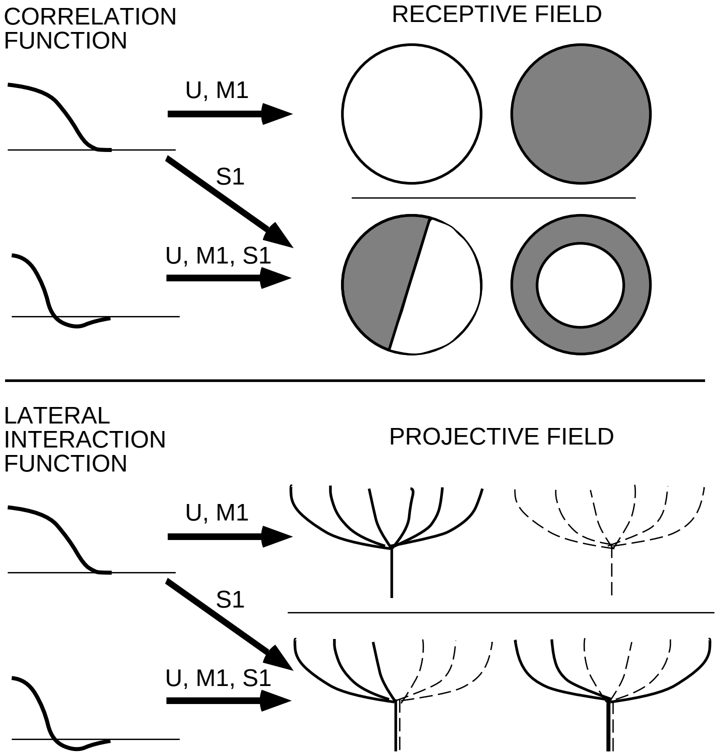


Figure 4: The Role of Constraints on Input and Output Cells: a Heuristic Approach
TOP: Output cell receptive fields (RFs) expected to develop under unconstrained dynamics ('U'), or under M1 or S1 constraints on output cells. White regions in receptive fields indicate positive weights; dark regions indicate zero or negative weights, depending on whether $w_{\mathrm{min}}$ is zero or negative. Correlations between input activities are shown as a function of input separation. Without constraints or under M1, the weights vary in sign to match the oscillation of the correlation function. Under S1, the weights always alternate in sign, with wavelength no larger than an arbor diameter. Note that this approach does not distinguish between different weight structures with similar wavelength of alternation, such as the two lower RFs. BOTTOM: Input cell projective fields (PFs) are determined in the same manner as RFs, except that: (1) the determining function is the output layer lateral interaction function; and (2) the determining constraints are those on input cells. Here, solid lines indicate positive weights, dotted lines indicate zero or negative weights.

(RF,PF) wavelength of an arbor diameter;

(b) Multiplicative (output,input) constraints do not alter the (RF,PF) wavelength of this dominant pattern, but only stabilize its amplitude.

2. If the dominant pattern in the absence of constraints has (RF,PF) wavelength smaller than an arbor diameter, then (output,input) constraints, whether enforced multiplicatively or subtractively, will have little effect.

In all cases, saturation of all synapses is expected without constraints or under subtractive constraints, but not under multiplicative constraints.

Several cautions must be emphasized about this approach. First, it predicts only the characteristic wavelength of weight alternation, and does not distinguish between different weight structures with similar wavelength. Second, the approach is heuristic: its validity must be checked in any particular case. In particular, the final weight pattern is expected to be one in which the dominant PF and RF patterns are "knitted together" into a compatible overall pattern. If such a "knitting" is not possible, the heuristics will fail.

This analysis can be applied to understand the effects of subtractive input constraints on ocular dominance segregation. Consider the development of the difference $\mathbf{w}^{\mathrm{D}}$ between projections from the two eyes (section 2.6). An RF across which $\mathbf{w}^{\mathrm{D}}$ is all-positive or all-negative corresponds to a monocular receptive field. Subtractive *output* constraints have no effect on the development of $\mathbf{w}^{\mathrm{D}}$: such constraints affect only the sum, not the difference, of the two projections. When RFs are monocular, an oscillation across PFs of $\mathbf{w}^{\mathrm{D}}$ corresponds to the oscillation between ocular dominance columns (Fig. 4). Subtractive *input* constraints separately conserve the total strength from the left-eye input and from the right-eye input at each position, and so conserve the total difference $\mathbf{w}^{\mathrm{D}}$ from each input position. Thus, these constraints ensure that there is an oscillation of $\mathbf{w}^{\mathrm{D}}$ across PFs with wavelength no larger than an arbor diameter. Subtractive input constraints thus determine the width of a left-eye plus a right-eye ocular dominance column to be an arbor diameter when the unconstrained dynamics would lead to larger columns, but have little effect on ocular dominance segregation otherwise (Miller et al., 1989).

# 4 How Can Constraints Be Implemented?

## 4.1 Learning rules that converge to constrained dynamics

The formulations in Eqs. 3-5 confine the dynamics to the constraint surface that contains the initial weight vector. Alternatively, constraints may be formulated so that the dynamics converge from an arbitrary initial weight vector to one particular constraint surface, and remain on that constraint surface thereafter. In this case the dynamics are described by Eqs. 3–5 after an initial transient in which the constraint surface is reached.

Such a formulation of S1 constraints is obtained by setting $\epsilon(\mathbf{w}) = |k_2|(\mathbf{n} \cdot \mathbf{w} - k_1)$ for constants $k_1$ and $k_2$ in Eq. 2. When $|k_2|$ is large, this term enforces the constraint $\mathbf{n} \cdot \mathbf{w} = k_1$ (Linsker, 1986) and is equivalent to an S1 constraint (MacKay and Miller, 1990a, Appendix E). Multiplicative constraints can be similarly formulated.

Dynamics that converge to a multiplicative constraint can also be obtained by substituting a constant $k > 0$ for the denominator of $\gamma(\mathbf{w})$ in Eqs. 3 or 5. Let $\mathbf{c}$ be the constraint vector (for M1, $\mathbf{c} = \mathbf{n}$; for M2, $\mathbf{c} = \mathbf{w}$) and set $\gamma(\mathbf{w}) = \mathbf{c} \cdot \mathbf{C}\mathbf{w}/k$. Then, if the initial condition and dynamics maintain $\mathbf{c} \cdot \mathbf{C}\mathbf{w} > 0$, the dynamics will flow to the constraint surface $\mathbf{c} \cdot \mathbf{w} = k$ and remain stable to perturbations off it thereafter (as can be seen by examining $\mathbf{c} \cdot \frac{d}{dt}\mathbf{w}$). Oja (1982) studied such M2 constraints with $k = 1$ and proved convergence to the principal eigenvector.

Finally, if the principal eigenvalue $\lambda^0$ of $\mathbf{C}$ is positive, convergent multiplicative dynamics can also be formulated by using any $\gamma(\mathbf{w})$ in Eq. 1 that grows with $|\mathbf{w}|$ and takes values both smaller and larger than $\lambda^0$ (B. Pearlmutter, unpublished manuscript). This leads to convergence to a multiple of the principal eigenvector, $\beta\mathbf{e}^0$, satisfying the constraint $\gamma(\beta\mathbf{e}^0) = \lambda^0$. An example is $\gamma(\mathbf{w}) = |\mathbf{w}|^2$ (Yuille et al., 1989).

### 4.2 Use of thresholds to achieve constraints

Consider a linear Hebbian rule:

$$\frac{d}{dt}\mathbf{w}(t) = \lambda\,[y(t) - y_\theta]\,[\mathbf{x}(t) - \mathbf{x}_\theta] \tag{7}$$

Here $y$ is the activation of the output cell, $\mathbf{x}$ the vector of input activities, and $\mathbf{x}_\theta$ and $y_\theta$ are threshold activity levels for Hebbian plasticity. Assume a linear activation rule, $y = \mathbf{w} \cdot \mathbf{x}$. We average Eq. 7 over input patterns, assuming that $\mathbf{x}_\theta$ and $y_\theta$ are constant over input patterns. The resulting equation is:

$$\frac{d}{dt}\mathbf{w}(t) = \mathbf{Q}\mathbf{w} + \lambda\Big[\,\langle y\rangle - y_\theta\Big]\Big[\,\langle \mathbf{x}\rangle - \mathbf{x}_\theta\Big] \tag{8}$$

where $\mathbf{Q}$ is the input covariance matrix, $\mathbf{Q} = \lambda\,\langle(\mathbf{x} - \langle\mathbf{x}\rangle)\,(\mathbf{x} - \langle\mathbf{x}\rangle)^{\mathrm{T}}\rangle$, and $\langle y\rangle = \mathbf{w} \cdot \langle\mathbf{x}\rangle$. The second term is a decay term, and can enforce a constraint.

If the elements of $[\langle\mathbf{x}\rangle - \mathbf{x}_\theta]$ are large and negative, the type 1 constraint $\langle y\rangle = y_\theta$, *i.e.* $\mathbf{w} \cdot \langle\mathbf{x}\rangle = y_\theta$, will be enforced. If $\mathbf{x}_\theta$ is independent of $\mathbf{w}$, this constraint is enforced subtractively; if furthermore all inputs have the same mean activity level and threshold, this is an S1 constraint, as discussed in section 4.1 and as used in (Linsker, 1986). The presynaptic threshold $\mathbf{x}_\theta$ can also enforce a constraint if its elements increase with those of $\mathbf{w}$. For example, an M2 constraint that converges to $\mathbf{w} \cdot \mathbf{w} = 1$ (section 4.1), when applied to the unconstrained equation $\frac{d}{dt}\mathbf{w} = y\mathbf{x}$, yields the rule proposed by Oja (1982): $\frac{d}{dt}\mathbf{w} = y\mathbf{x} - \mathbf{w}(\mathbf{w} \cdot y\mathbf{x})$, or $\frac{d}{dt}\mathbf{w} = y(\mathbf{x} - \mathbf{w}y)$. This is $\mathbf{x}_\theta = \mathbf{w}y$.[8]

[8]Note: Eq. 8 is not valid for this case, because $\mathbf{x}_\theta$ varies with input patterns.

Both of these mechanisms require that inputs activated at an average or below-average level lose synaptic strength when the postsynaptic cell is highly activated. This is not the case in at least one biological system, LTP in hippocampus (Gustafsson et al., 1987). This difficulty is avoided by the choice $\mathbf{x}_\theta = \mathbf{w}$, which yields the multiplicatively stabilized rule $\frac{d}{dt}\mathbf{w} = y(\mathbf{x} - \mathbf{w})$. This rule does not ensure a selective outcome: as noted in (Kohonen, 1989, Section 4.3.2), it converges either to the principal eigenvector of $\mathbf{C} = \left\langle \mathbf{x}\mathbf{x}^T \right\rangle$, or else to $\mathbf{w} = \mathbf{0}$. However, with a nonlinear, competitive rule for $y$ that ensures localized activation, this rule is that of the self-organizing feature map of Kohonen (1989) and does achieve selectivity.

The postsynaptic threshold, $y_\theta$, can enforce a constraint if it increases faster than linearly with the average postsynaptic activation, $\langle y \rangle$, and if the elements of $[\langle \mathbf{x} \rangle - \mathbf{x}_\theta]$ are positive. Bienenstock *et al.* (1982) proposed the rule $y_\theta = \langle y \rangle^2 / y_{\text{set}}$, where $y_{\text{set}}$ is a cell's "preset" desired activity level. They combined this with a nonlinear Hebbian rule, *e.g.* $y(y - y_\theta)$ in place of $(y - y_\theta)$ in Eq. 7. With either Hebbian rule, this "sliding threshold" has the effect of adjusting synaptic strengths to achieve $\langle y \rangle \approx y_\theta$, or $\langle y \rangle = \mathbf{w} \cdot \langle \mathbf{x} \rangle \approx y_{\text{set}}$. Thus, it provides another method of achieving a type 1 constraint. Recent results both in the peripheral auditory system (Yang and Faber, 1991) and in hippocampus (Huang et al., 1992) are suggestive that increased neural activity may elevate a threshold for modification, but in a manner specific to those inputs whose activity is increased. This is consistent with an increase in the elements of $\mathbf{x}_\theta$ corresponding to activated inputs, but not with an increase in $y_\theta$, which would elevate thresholds for all inputs.

Covariance models (Sejnowski, 1977a,b) have been proposed to solve the problem that chance coincidences drive synapses to saturate at $w_{\max}$. This problem was known to occur in Hebb models of the form $\frac{d}{dt}\mathbf{w} \propto y\mathbf{x}$. In a covariance model, $\frac{d}{dt}\mathbf{w} \propto (y - \langle y \rangle)(\mathbf{x} - \langle \mathbf{x} \rangle)$. This is a Hebb model with linearly sliding thresholds: $y_\theta = \langle y \rangle$, $\mathbf{x}_\theta = \langle \mathbf{x} \rangle$. In this case, the decay term in Eq. 8 is zero, so synaptic growth is driven by the unconstrained equation $\frac{d}{dt}\mathbf{w} = \mathbf{Q}\mathbf{w}$. Thus, the problem of loss of selectivity under a Hebbian rule is not avoided by a linear covariance rule.

### 4.3 Biological implementation of constraints

Rules that conserve synaptic strength have been criticized as nonlocal (*e.g.* Bienenstock et al., 1982). Thus, it is important to note that multiplicative or subtractive constraints in their general form (Eqs. 1–2) can be implemented locally if each of a cell's synaptic weights undergoes decay, either at a fixed rate (subtractive decay) or at a rate proportional to its strength (multiplicative decay); and if the overall gain of this decay, $\gamma(\mathbf{w})$ or $\epsilon(\mathbf{w})$, is set for the cell as a whole, and increases with the cell's total synaptic strength. Such a cellular increase in decay, implemented locally at each synapse, might be achieved in at least two ways. First, a cell might have a limited capacity to metabolically supply its synapses, so that greater total synaptic strength means less supply and thus faster decay for each synapse. Second, the overall rate of decay might increase with a cell's average degree of activation, which in turn would increase with the total synaptic strength received by a cell. Increased activation could increase release of a molecule that degrades synapses, such as a protease, or decrease release of a molecule that supports synapses, such as a trophic, adhesion, or sprouting factor (evidence for such mechanisms is reviewed in Van Essen et. al, 1990). Increased activation might also increase decay due to thresholds for synaptic modification, as just discussed.

# 5 Discussion

We have demonstrated that multiplicative and subtractive constraints can lead to fundamentally different outcomes in linear learning. Under multiplicative constraints, the weight vector tends to the principal eigenvector of the unconstrained time development operator. This is a "graded" receptive field in which most mutually correlated inputs are represented. Thus, when two equally active eyes compete, both retain equal innervation unless the two eyes are anticorrelated. Under subtractive constraints, the weight vector tends to a receptive field that is "sharpened" to a subset of maximally-correlated inputs: the weights of these inputs reach the maximum allowed strength, while all other weights reach the minimum allowed strength. When two eyes compete, subtractive constraints can lead to domination by one eye (ocular dominance segregation) provided only that correlations within one eye are stronger than those between the eyes.

The instability of subtractive constraints depends on the unconstrained operator having at least two positive eigenvalues, which is typical for Hebbian learning. An interesting alternative is *anti*-Hebbian learning (Mitchison, 1991): in this case, all unconstrained eigenvalues are reversed in sign from the Hebbian case, so typically no eigenvalue is positive. Our analysis applies to this case also: multiplicatively constrained dynamics flow to the principal eigenvector, which is the vector that would have the *smallest* eigenvalue under Hebbian dynamics (Mitchison, 1991); while subtractively constrained dynamics flow *to* the fixed point, which is stable.

Multiplicative and subtractive constraints represent two fundamentally different methods of controlling the size of the weight vector. Multiplication equally rescales all weight patterns, while subtraction directly acts on only a single weight pattern. Because this difference is general, many of the results we have found for the linear case may generalize to cases involving nonlinear rules.

Biologically, there is as yet little evidence as to the mechanisms that lead activity-dependent plasticity to be competitive or to achieve selectivity. Among the two choices of subtractive and multiplicative constraints, subtractive seem to resemble biology more closely in systems where sharp receptive fields are achieved, and in visual cortex where ocular dominance columns are likely to develop without requiring anticorrelation between the eyes; while multiplicative constraints might resemble biology more closely in situations like adult cortical plasticity where continually moving and graded representations may occur (Kaas, 1991). We do not advocate that one or the other of these is the biologically correct choice. Rather, we wish: (1) to point out that different choices of competitive mechanism can yield different outcomes, so it is important for the modeler to know whether and how their results depend on these choices; and (2) to begin to distinguish and characterize different classes of such mechanisms, which might then be compared to biology.

# A Proofs of Mathematical Results

We study dynamics confined to a constraint surface and governed by a general multiplicative constraint (Eq. 1) or by an S1 subtractive constraint (Eq. 4). As in the text, we use indices $a, b, \ldots$ to refer to the eigenvector basis of $\mathbf{C}$. We assume that $\mathbf{C}$ is symmetric and thus has a complete set of orthonormal eigenvectors $\mathbf{e}^a$ with corresponding eigenvalues $\lambda^a$.

Write the constrained equation as $\frac{d}{dt}\mathbf{w} = \mathbf{f}(\mathbf{w})$. To determine the stability of a fixed point $\mathbf{w}^{\text{FP}}$ (where $\mathbf{f}(\mathbf{w}) = \mathbf{0}$), we linearize $\mathbf{f}(\mathbf{w})$ about the fixed point. Call this linearized operator $\mathbf{D}$; in the eigenvector basis of $\mathbf{C}$, it is a matrix with elements $D_{ab} = \frac{\partial f_a(\mathbf{w})}{\partial w_b}\big|_{\mathbf{w}=\mathbf{w}^{\text{FP}}}$. For an S1 constraint, $\mathbf{f}(\mathbf{w}) = \mathbf{PCw}$ is linear (here, $\mathbf{P} = [1 - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\text{T}}]$), so $\mathbf{D} = \mathbf{PC}$. We define the constraint plane to be the hyperplane tangent to the constraint surface at the fixed point, and the constraint vector $\mathbf{c}$ to be the vector normal to the constraint plane. $\mathbf{c}$ is a left eigenvector of $\mathbf{D}$.[9]

The stability of a fixed point is determined by the eigenvalues of $\mathbf{D}$ (Hirsch and Smale, 1974). If one eigenvalue is positive, the fixed point is unstable; if all eigenvalues are negative, the fixed point is stable. In assessing the outcome of the constrained dynamics, we are only concerned with stability of the fixed point to perturbations within the constraint surface. Thus, if all eigenvalues are negative except one zero eigenvalue corresponding to a direction outside the constraint surface, then the fixed point is stable.

**Theorem 1** *Proof.* We consider a multiplicatively constrained equation, $\frac{d}{dt}\mathbf{w} = \mathbf{Cw} - \gamma(\mathbf{w})\mathbf{w}$. We assume that multiplicative confinement of the dynamics to the constraint surface means two things. First, $\mathbf{D}$ has one zero or negative eigenvalue corresponding to the enforcement of the constraint, with associated left eigenvector $\mathbf{c}$. Therefore any right eigenvector of $\mathbf{D}$ with positive eigenvalue is parallel to the constraint plane. Second, the constraint plane is not parallel to the subtracted vector $\mathbf{w}$ that enforces the constraint.

A fixed point is an eigenvector of $\mathbf{C}$: $\mathbf{w}^{\text{FP}} = w_a\mathbf{e}^a$ for some $a$, $\gamma(\mathbf{w}^{\text{FP}}) = \lambda^a$. The linearized operator is $\mathbf{D} = \mathbf{C} - \lambda^a\mathbf{1} - w_a\mathbf{e}^a\left[\nabla\gamma(\mathbf{w}^{\text{FP}})\right]^{\text{T}}$, where $\nabla$ is the gradient operator defined by $\nabla g(\mathbf{x}) \equiv \sum_a \mathbf{e}^a \frac{\partial g(\mathbf{w})}{\partial w_a}\big|_{\mathbf{w}=\mathbf{x}}$. In the eigenvector basis of $\mathbf{C}$, $\mathbf{D}$ is a diagonal matrix with the addition of one row of off-diagonal elements; such a matrix has the same eigenvalues as the diagonal matrix alone (because the characteristic equation, $\det(\mathbf{D} - \lambda\mathbf{1}) = 0$, is unchanged by the additional row). The diagonal part of $\mathbf{D}$ is the matrix $\mathbf{C} - \lambda^a\mathbf{1} - \kappa\mathbf{e}^a[\mathbf{e}^a]^{\text{T}}$, where $\kappa \equiv w_a\frac{\partial\gamma(w)}{\partial w_a}\big|_{\mathbf{w}=w_a\mathbf{e}^a}$. This is the operator $\mathbf{C}$ with eigenvalue $\lambda^a$ reduced to $-\kappa$ and all other eigenvalues reduced by $\lambda^a$. Note that $\mathbf{e}^a$ is the right eigenvector of $\mathbf{D}$ with eigenvalue $-\kappa$; since $\mathbf{e}^a$ is not parallel to the constraint plane, $-\kappa \leq 0$.

Now we determine whether $\mathbf{D}$ has positive eigenvalues. If $\mathbf{e}^a$ is not the principal eigenvector of $\mathbf{C}$, then $\mathbf{D}$ has a positive eigenvalue and the fixed point is unstable. If $\mathbf{e}^a$ is the principal eigenvector of $\mathbf{C}$, and it is nondegenerate (no other eigenvector has the same eigenvalue), then all eigenvalues of $\mathbf{D}$ except perhaps a zero corresponding to $\mathbf{e}^a$ are negative; so the fixed point is stable. If $\mathbf{C}$ has $N$ degenerate principal eigenvectors, and $\mathbf{e}^a$ is one of them, then $\mathbf{D}$ has $N-1$ zeros corresponding to perturbations within the degenerate subspace: the principal eigenvector fixed points are thus marginally stable (eigenvalue 0) to perturbations within this subspace, and stable to other perturbations. •

**Lemma 1** *Under an S1 constraint, $\frac{d}{dt}\mathbf{w} = \mathbf{PCw}$ with $\mathbf{P} = (\mathbf{1} - \hat{\mathbf{n}}\hat{\mathbf{n}}^{\text{T}})$, if there is a vector $\mathbf{v}$ parallel to the constraint plane such that $\mathbf{v}^{\text{T}}\mathbf{PCv} > 0$, then $\mathbf{PC}$ has a positive eigenvalue.*

[9] *Proof*: For any $\Delta\mathbf{w}$ in the constraint plane, $\mathbf{D}\Delta\mathbf{w}$ must remain within the constraint plane; that is, $\mathbf{c}^{\text{T}}\mathbf{D}\Delta\mathbf{w} = 0$ for all $\Delta\mathbf{w}$ satisfying $\mathbf{c}^{\text{T}}\Delta\mathbf{w} = 0$. Therefore, $\mathbf{c}^{\text{T}}\mathbf{D} \propto \mathbf{c}^{\text{T}}$.

*Proof.* $\mathbf{v}$ is parallel to the constraint plane, so $\mathbf{v} \cdot \mathbf{n} = 0$ and $\mathbf{Pv} = \mathbf{v}$. Thus, from $\mathbf{v}^{\mathrm{T}}\mathbf{PCv} > 0$ we conclude that $\mathbf{v}^{\mathrm{T}}\mathbf{PCPv} > 0$. Since $\mathbf{PCP}$ is symmetric, this implies that $\mathbf{PCP}$ has a positive eigenvalue; call this eigenvalue $\lambda^0$, with corresponding eigenvector $\mathbf{e}^0$. $\mathbf{e}^0$ is parallel to the constraint plane, *i.e.* $\mathbf{Pe}^0 = \mathbf{e}^0$ (because $e^0 = \mathbf{PCPe}^0/\lambda^0$, and $\mathbf{P}^2 = \mathbf{P}$). So $\mathbf{PCe}^0 = \mathbf{PCPe}^0 = \lambda^0 e^0$. •

**Theorem 2** *Proof.* We consider the S1 constrained equation $\frac{d}{dt}\mathbf{w} = \mathbf{PCw}$. Let $\mathbf{v}$ be a linear combination of the two eigenvectors of $\mathbf{C}$ with positive eigenvalues, such that $\mathbf{v}$ is parallel to the constraint plane: $\mathbf{v} \cdot \mathbf{n} = 0$. Then $\mathbf{v}^{\mathrm{T}}\mathbf{PCv} = \mathbf{v}^{\mathrm{T}}\mathbf{Cv} > 0$. So, by lemma 1, $\mathbf{PC}$ must have a positive eigenvalue. •

**Theorem 3** *Proof.* This is a generalization of a similar proof in (Linsker, 1986). Suppose $\mathbf{PCw}^{\mathrm{FP}} =$ "$\mathbf{0}$" when synapses $w_i^{\mathrm{FP}}$ and $w_j^{\mathrm{FP}}$ are both not saturated. By "$\mathbf{0}$", we mean that each component of the vector is either 0, or else of a sign that would take an already saturated synapse beyond its limiting value. Let $\hat{\mathbf{u}}^i$ be the unit vector with $i^{\mathrm{th}}$ weight 1 and all other elements 0, and similarly for $\hat{\mathbf{u}}^j$. Consider stability of the fixed point to a perturbation confined to the $\hat{\mathbf{u}}^i/\hat{\mathbf{u}}^j$ plane. The action of $\mathbf{C}$ in this plane is given by the submatrix $\mathbf{C}_{\hat{\mathbf{u}}^i\hat{\mathbf{u}}^j} = \begin{pmatrix} C_{ii} & C_{ij} \\ C_{ji} & C_{jj} \end{pmatrix}$. The eigenvalues of $\mathbf{C}_{\hat{\mathbf{u}}^i\hat{\mathbf{u}}^j}$ are both real and positive when the conditions of the theorem are met. Let $\mathbf{e}_1$ and $\mathbf{e}_2$ be the two orthonormal eigenvectors of $\mathbf{C}_{\hat{\mathbf{u}}^i\hat{\mathbf{u}}^j}$, with all synaptic components other than $i$ and $j$ set to zero. As in the proof of theorem 2, let $\mathbf{v}$ be a linear combination of $\mathbf{e}_1$ and $\mathbf{e}_2$ that is parallel to the constraint plane, $\mathbf{v} \cdot \mathbf{n} = 0$. Then $\mathbf{v}^{\mathrm{T}}\mathbf{PCv} = \mathbf{v}^{\mathrm{T}}\mathbf{Cv} > 0$. So, by lemma 1, the fixed point is unstable. •


## Acknowledgements

K.D.M. thanks C. Koch, Caltech, and M.P. Stryker, UCSF for supporting this work, which was performed in their laboratories. K.D.M. was supported by a Del Webb fellowship and a Markey Foundation internal grant, both from Caltech Division of Biology, and by an N.E.I. Fellowship at UCSF.
D.J.C.M. was supported by a Caltech Fellowship and a Studentship from SERC, UK.
We thank Bartlett Mel and Terry Sejnowski for helpful comments on the manuscript. This collaboration would have been impossible without the internet/NSFnet.


## References


Bienenstock E.L., Cooper L.N., and Munro P.W. (1982). Theory for the development of neuron selectivity: Orientation specificity and binocular interaction in visual cortex. *J. Neurosci.*, 2:32–48.

Bourgeois J.-P., Jastreboff P.J., and Rakic P. (1989). Synaptogenesis in visual cortex of normal and preterm monkeys: Evidence for intrinsic regulation of synaptic overproduction. *Proc. Natl. Acad. Sci. USA*, 86:4297–4301.

Grajski K.A. and Merzenich M.M. (1990). Hebb-type dynamics is sufficient to account for the inverse magnification rule in cortical somatotopy. *Neural Comp.*, 2:71–84.

Guillery R.W. (1972). Binocular competition in the control of geniculate cell growth. *Journal of Comparative Neurology*, 144:117–130.

Gustafsson B., Wigstrom H., Abraham W.C., and Huang Y.-Y. (1987). Long-term potentiation in the hippocampus using depolarizing current pulses as the conditioning stimulus to single volley synaptic potentials. *J. Neurosci.*, 7:774–780.

Hayes W.P. and Meyer R.L. (1988a). Optic synapse number but not density is constrained during regeneration onto surgically halved tectum in goldfish: HRP-EM evidence that optic fibers compete for fixed numbers of postsynaptic sites on the tectum. *J. Comp. Neurol.*, 274:539–559.

Hayes W.P. and Meyer R.L. (1988b). Retinotopically inappropriate synapses of subnormal density formed by misdirected optic fibers in goldfish tectum. *Devel. Brain Res.*, 38:304–312.

Hayes W.P. and Meyer R.L. (1989a). Impulse blockade by intraocular tetrodoxin during optic regeneration in goldfish: HRP-EM evidence that the formation of normal numbers of optic synapses and the elimination of exuberant optic fibers is activity independent. *J. Neurosci.*, 9:1414–1423.

Hayes W.P. and Meyer R.L. (1989b). Normal numbers of retinotectal synapses during the activity-sensitive period of optic regeneration in goldfish: HRP-EM evidence implicating synapse rearrangement and collateral elimination during map refinement. *J. Neurosci.*, 9:1400–1413.

Hirsch M.W. and Smale S. (1974). *Differential equations, dynamical systems and linear algebra.* Academic Press, New York.

Huang Y.Y., Colino A., Selig D.K., and Malenka R.C. (1992). The influence of prior synaptic activity on the induction of long-term potentiation. *Science*, 255:730–733.

Kaas J.H. (1991). Plasticity of sensory and motor maps in adult mammals. *Ann. Rev. Neurosci.*, 14:137–167.

Kohonen T. (1989). *Self-Organization and Associative Memory.* Springer-Verlag, Berlin, 3rd edition.

Linsker R. (1986). From basic network principles to neural architecture: Emergence of spatial-opponent cells. *Proc. Natl. Acad. Sci. USA*, 83:7508–7512.

MacKay D.J.C. and Miller K.D. (1990a). Analysis of Linsker's applications of Hebbian rules to linear networks. *Network*, 1:257–298.

MacKay D.J.C. and Miller K.D. (1990b). Analysis of Linsker's simulation of Hebbian rules. *Neural Comp.*, 2:173–187.

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

Malsburg C. von der (1979). Development of ocularity domains and growth behavior of axon terminals. *Biol. Cyber.*, 32:49–62.

Malsburg C. von der and Willshaw D.J. (1976). A mechanism for producing continuous neural mappings: ocularity dominance stripes and ordered retino-tectal projections. *Exp. Brain Res.*, Supp. 1:463–469.

Miller K.D. (1990a). Correlation-based models of neural development. In Gluck M.A. and Rumelhart D.E., editors, *Neuroscience and Connectionist Theory*, pages 267–353. Lawrence Erlbaum Ass., Hillsdale, NJ.

Miller K.D. (1990b). Derivation of linear Hebbian equations from a nonlinear Hebbian model of synaptic plasticity. *Neural Comp.*, 2:321–333.

Miller K.D. (1992). Development of orientation columns via competition between ON- and OFF-center inputs. *NeuroReport*, 3:73–76.

Miller K.D., Keller J.B., and Stryker M.P. (1989). Ocular dominance column development: Analysis and simulation. *Science*, 245:605–615.

Miller K.D. and MacKay D.J.C. (1992). The role of constraints in Hebbian learning. Technical Report Memo 19, Caltech, Pasadena, CA.

Mitchison G. (1991). Removing time variation with the anti-Hebbian differential synapse. *Neural Comp.*, 3:312–320.

Murray M., Sharma S., and Edwards M.A. (1982). Target regulation of synaptic number in the compressed retinotectal projection of goldfish. *J. Comp. Neurol.*, 209:374–385.

Oja E. (1982). A simplified neuron model as a principal component analyzer. *J. Math. Biol.*, 15:267–273.

Pallas S.L. and Finlay B.L. (1991). Compensation for population-size mismatches in the hamster retinotectal system: Alterations in the organization of retinal projections. *Vis. Neurosci.*, 6:271–281.

Perez R., Glass L., and Shlaer R. (1975). Development of specificity in the cat visual cortex. *J. Math. Biol.*, 1:275–288.

Rochester N., Holland J.H., Haibt L.H., and Duda W.L. (1956). Tests on a cell assembly theory of the action of the brain, using a large digital computer. *IRE Trans. Info. Theory*, IT-2:80–93.

Rosenblatt F. (1961). *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms.* Spartan Books, Washington, D.C.

Sejnowski T.J. (1977a). Statistical constraints on synaptic plasticity. *J. Theor. Biol.*, 69:385–389.

Sejnowski T.J. (1977b). Storing covariance with nonlinearly interacting neurons. *J. Math. Biol.*, 4:303–321.

Van Essen D.C., Gordon H., Soha J.M., and Fraser S.E. (1990). Synaptic dynamics at the neuromuscular junction: Mechanisms and models. *J. Neurobiol.*, 21:223–249.

Whitelaw V.A. and Cowan J.D. (1981). Specificity and plasticity of retinotectal connections: a computational model. *J. Neurosci.*, 1:1369–1387.

Wiesel T.N. and Hubel D.H. (1965). Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *J. Neurophysiol.*, 28:1029–1040.

Willshaw D.J. and Malsburg C. von der (1976). How patterned neural connections can be set up by self-organization. *Proc. R. Soc. Lond. B.*, 194:431–445.

Willshaw D.J. and Malsburg C. von der (1979). A marker induction mechanism for the establishment of ordered neural mappings: its application to the retinotectal problem. *Phil. Trans. R. Soc. Lond. B.*, 287:203–243.

Yang X.D. and Faber D.S. (1991). Initial synaptic efficacy influences induction and expression of long-term changes in transmission. *Proc. Natl. Acad. Sci. USA*, 88:4299–4303.

Yuille A.L., Kammen D.M., and Cohen D.S. (1989). Quadrature and the development of orientation selective cortical cells by Hebb rules. *Biological Cybernetics*, 61:183–194.