CALIFORNIA INSTITUTE OF TECHNOLOGY
COMPUTATION AND NEURAL SYSTEMS PROGRAM

OCTOBER 9, 1992 CNS MEMO 19

THE ROLE OF CONSTRAINTS IN HEBBIAN LEARNING

Kenneth D. Miller[†] and David J. C. MacKay[*]

[†] Division of Biology 216-76 and
[*] Computation and Neural Systems Program
California Institute of Technology
Pasadena, CA 91125

The following information is appended on Aug. 30, 1995:
Current address for Kenneth Miller: Department of Physiology, UCSF, SF, CA 94143-0444; ken@phy.ucsf.edu.
Current address for David MacKay: Radio Astronomy, Cavendish Laboratory, Madingley Road, Cambridge, CB3 0HE. United Kingdom; mackay@mrao.cam.ac.uk.
A correction has been appended to the end of Appendix B on this date.
An abridged and rewritten version of this memo appears in *Neural Computation* **6**, 100-126 (1994).

PASADENA, CALIFORNIA 91125

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, 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 few 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. An approach to understanding constraints over input and over output cells is suggested, and some biological implementations are discussed.

Development in many neural systems appears to be guided by "Hebbian" or similar activity-dependent, correlation-based rules of synaptic modification (reviewed in [25]). 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 on the effects of monocular versus binocular visual deprivation in young animals ([42]; see also [6]). 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.

There is no direct biological evidence for the existence of competitive limits on the total synaptic strength supported by a cell. However, some support for the idea that a cell's synaptic resources are limited comes from biological evidence for intrinsic limits to the total *number* of synapses supported by a cell. In the retinotectal system in several species, if retinal cells are forced to innervate only a partial tectum, rather than their usual target of the entire tectum, the number of synapses per tectal cell remains normal [8, 31, 33]. This suggests that there are a limited number of *postsynaptic* sites per target cell, for which incoming neurons must compete. If inputs are given a *larger* than normal target structure, they innervate target cells with less than normal density, indicating intrinsic limits in the total *presynaptic* innervation by an input cell [9] (this also occurs in other systems [3, 4, 37]). In several systems, synapse number appears to be conserved regardless of the patterns of neural activity, even while synaptic rearrangements occur that vary with neural activity patterns. These systems include the monkey visual cortex during ocular dominance column development [2] and the goldfish optic tectum during post-regeneration map refinement [11, 10].

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 correspondingly shrink. One theoretical method of achieving selectivity is to use constraints to enforce a competition. Von der Malsburg [22] first proposed that constraints conserving the total synaptic strength supported by each input or output cell could be used to stabilize a Hebbian rule, a proposal also made by Perez et al. [34]. Much earlier, Rochester et al. [35] used a model that conserved total synaptic strength over all synapses in a Hebbian network. Rosenblatt [36] proposed a synaptic strength conservation rule in the context of perceptron learning. Subsequently, many authors have used one form or another of similar constraints, *e.g.* [5, 17, 18, 28, 32, 41].

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, all synapses might decay at an equal rate; 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 [20, 21, 25, 28, 29].

Intuitively, the difference between these two methods is that multiplicative enforcement of a constraint (a "multiplicative constraint") suppresses the growth of all patterns of synaptic weights, while subtractive enforcement (a "subtractive constraint") only suppresses patterns whose growth would violate the constraint. In each short time Δt, a multiplicative constraint multiplies all weights, or equivalently all weight patterns, by a constant. If the unconstrained dynamics are linear in the synaptic strengths, the dynamical equations are invariant under such a rescaling. Then the dynamics under the multiplicative constraint are equivalent to those of unbounded linear growth. So the synaptic pattern that would grow the fastest in the absence of the constraint (the principal eigenvector of the operator determining unconstrained development) will eventually emerge, while all other components will eventually be suppressed. In contrast, in each short time, a subtractive constraint subtracts a small constant from all weights. This selectively alters the amplitude of weight patterns whose growth would violate the constraint — those that have a nonzero total synaptic strength — while leaving the amplitude of all other weight patterns unaffected.

In this paper, we analyze the effects of subtractive and multiplicative constraints on the dynamics of linear learning rules. We then briefly discuss some theoretical and biological methods of implementing constraints.

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 on a single postsynaptic cell. Let $\mathbf{w}$ be the vector of synaptic weights onto this cell; the i^{th} component, w_i, is the synaptic weight from the i^{th} input. We assume synaptic weights are initially randomly distributed with mean w_{init}, and are limited to remain between a maximum value $w_{\max}$ and minimum value $w_{\min}$. 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 [18, 20, 28, 32]. 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_{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_{max}, while the remaining synapses saturate at w_{min}. If w_{max} is increased, or the total conserved synaptic strength w_{tot} is decreased by decreasing w_{init}, the receptive field is sharpened under a subtractive constraint (column 4). The final number of nonzero synapses in this case (with $w_{\text{min}} = 0$) is approximately $w_{\text{tot}}/w_{\text{max}}$. In contrast, under the multiplicative constraint or without constraints, the shape of the final receptive field is unaltered by such changes in w_{max} and w_{tot}.

If w_{min} is decreased below zero, center-surround receptive fields can result under subtractive constraints (Fig. 1B). Results under multiplicative constraints or without constraints are unaltered by this change. This mechanism of developing center-surround receptive fields underlies the results of Linsker [18], as explained in [20, 21] and in section 2.4. Again, an increase in w_{max} or decrease in w_{tot} leads to sharpening of the positive part of the receptive field under subtractive constraints (column 4).

Next, consider ocular dominance segregation [28] (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 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 2.5.

Finally, consider the problem of map alignment (Fig. 1D). In the barn owl optic tectum, an auditory map comes to be aligned with a visual map under the guidance of visual experience [16]. Consider a toy model of this, as follows. Two input maps innervate the tectum. One map (the auditory) is adaptive, while the other map (the visual) is fixed and serves as a teacher to the auditory map. We consider a single output cell in the tectum, with activity y. Let $\mathbf{w}$ be the auditory map's synaptic weight vector, and $\mathbf{x}_n$ its vector of input activities corresponding to the n^{th} stimulus. Let s_n be the synaptic input from the visual map in response to the same stimulus: the visual map is topographic, so the largest responses s_n will come from stimuli in a particular topographic position

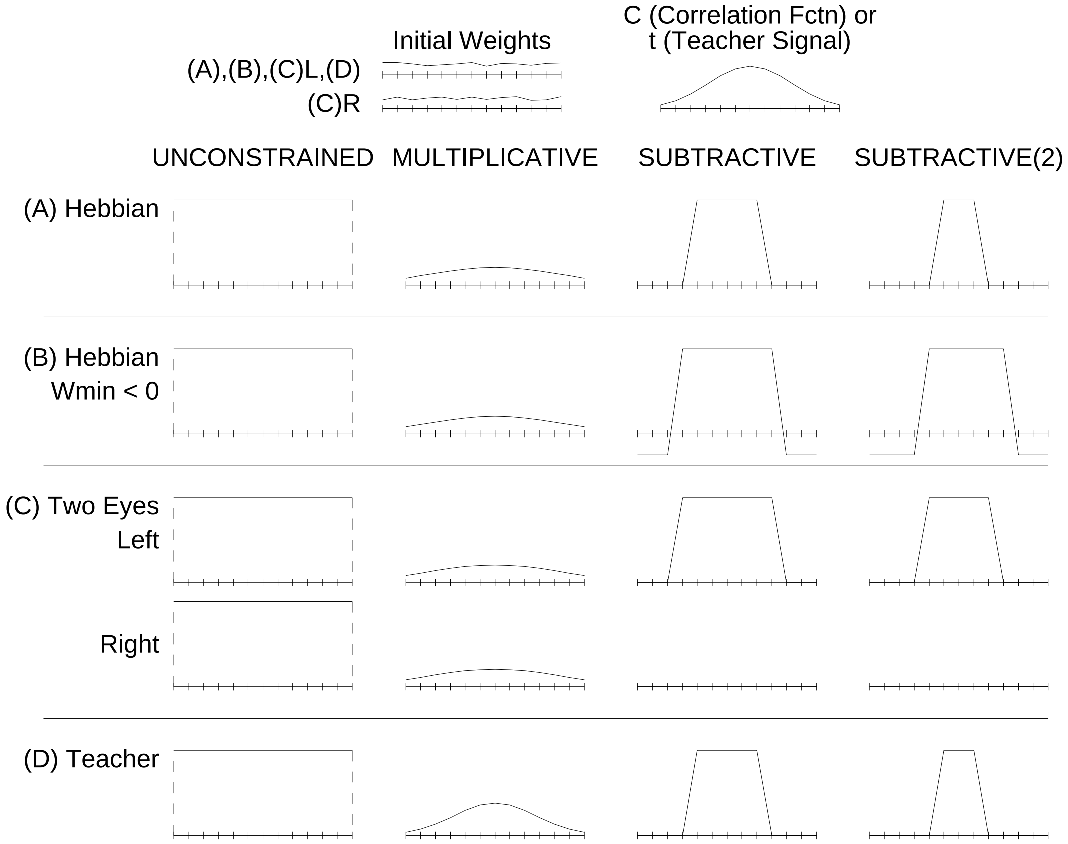

Figure 1: **Outcomes of Development Without Constraints and Under Multiplicative and Subtractive Constraints**

(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 $\frac{d}{dt}\mathbf{w}^L = \mathbf{C}\mathbf{w}^L$; $\frac{d}{dt}\mathbf{w}^R = \mathbf{C}\mathbf{w}^R$. (D). Outcome of a simplified model of the alignment of one map with another: a Hebbian rule in the limit in which the output cell's activity is determined by a "teacher signal". Unconstrained equation is $\frac{d}{dt}\mathbf{w} = \mathbf{t}$ (see text). The constant vector $\mathbf{t}$ is a Gaussian centered at the center of the arbor (shown at top right). Note that the resulting weights under multiplicative constraints are proportional to $\mathbf{t}$.

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_{tot}.

in visual space. Suppose a simple activation rule, $y = \mathbf{w} \cdot \mathbf{x} + s$, is combined with a Hebbian rule, $\frac{d}{dt} w_i = \langle y x_i \rangle$. The brackets indicate an average over input patterns. In the limit in which the visual, teacher signal is dominant, the Hebbian rule becomes $\frac{d}{dt}\mathbf{w} = \langle s\mathbf{x} \rangle$. Here the right-hand side is a constant vector, $\mathbf{t} \equiv \langle s\mathbf{x} \rangle$, which measures the correlation between the auditory inputs and the visual teacher input. We consider the case in which there is initially a rough auditory map, so that the correlation of auditory inputs with the visual teacher decreases as a Gaussian from the center of the arbor. In the absence of constraints, all synapses again reach $w_{\max}$, so no selectivity develops. Multiplicative constraints lead the weights to match the correlations, giving a moderate advantage to the best-correlated inputs. Subtractive constraints lead to a sharpening of the map so that only the best correlated inputs survive.

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, $\frac{d}{dt}\mathbf{w}(t) = \mathbf{C}\mathbf{w}(t)$. $\mathbf{C}$ is a matrix that drives the dynamics. We assume that $\mathbf{C}$ has a complete set of orthonormal eigenvectors $\mathbf{e}^a$ with corresponding eigenvalues λ^a (that is, $\mathbf{C}\mathbf{e}^a = \lambda^a \mathbf{e}^a$), as occurs if $\mathbf{C}$ is symmetric. In Hebbian learning, $\mathbf{C}$ is a correlation matrix: C_{ij} represents the correlation in activity between inputs i and j, so $\mathbf{C}$ is symmetric.[1] 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^{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.

As discussed in the introduction, a constraint can be enforced through time-varying synaptic decay, where the rate of the decay varies to maintain the constraint. Before studying constraints,

[1] We work in a representation in which each synapse is represented explicitly, and the density of synapses is implicit. Equivalently, the dynamics may be written in a representation in which the synaptic density or "arbor" function is explicitly represented. The current analysis also applies in these representations, as described in Appendix B.

it is instructive to review the effects of constant decay terms.

2.1 The effect of constant decay terms

Inclusion of synaptic decay results in the equation

$$\frac{d}{dt}\mathbf{w}(t) = \mathbf{C}\mathbf{w} - \epsilon\mathbf{n} - \gamma\mathbf{w} \tag{1}$$

or, in index notation, $\frac{d}{dt}w_i(t) = \sum_j C_{ij}w_j - \epsilon n_i - \gamma w_i$. The term $-\epsilon\mathbf{n}$ is a subtractive decay term; $\mathbf{n}$ is a constant vector, for example $\mathbf{n} = (1, 1, \ldots, 1)^T$ if all synapses decay at the same rate. The term $-\gamma\mathbf{w}$ is a multiplicative decay term.

To examine the effect of these decay terms, we work in the eigenvector basis. Write the initial condition of Eq. 1 as $\mathbf{w}(t=0) = \sum_a w_a^0 \mathbf{e}^a$. Write the fixed point of the equation, where $\frac{d}{dt}\mathbf{w} = 0$, as $\mathbf{w}^{\mathrm{FP}} = \sum_a w_a^{\mathrm{FP}} \mathbf{e}^a$. Then the solution to Eq. 1 can be written

$$\mathbf{w}(t) - \mathbf{w}^{\mathrm{FP}} = \sum_a \mathbf{e}^a (w_a^0 - w_a^{\mathrm{FP}}) e^{(\lambda^a - \gamma)t} \tag{2}$$

We can compute[2] $\mathbf{w}^{\mathrm{FP}} = \epsilon(\mathbf{C} - \gamma\mathbf{1})^{-1}\mathbf{n} = \sum_a \left(\frac{\epsilon \mathbf{e}^a \cdot \mathbf{n}}{\lambda^a - \gamma}\right) \mathbf{e}^a$, so $w_a^{\mathrm{FP}} = \frac{\epsilon \mathbf{e}^a \cdot \mathbf{n}}{\lambda^a - \gamma}$.

This solution illustrates the basic difference explained in the introduction between multiplicative and subtractive decay terms. The multiplicative decay term lowers the growth rate of all eigenvectors by the same amount γ, thus reducing smaller eigenvalues by a greater proportion. This favors the principal eigenvector. The subtractive term moves the fixed point of the equation, $\mathbf{w}^{\mathrm{FP}}$. This alters the growth rate only of eigenvectors $\mathbf{e}^a$ for which $w_a^{\mathrm{FP}} \neq 0$; these are the eigenvectors with non-zero component in the direction $\mathbf{n}$, that is with $\mathbf{e}^a \cdot \mathbf{n} \neq 0$. That is, constant subtractive decay only alters growth in directions coupled to $\mathbf{n}$, and leaves growth unaltered in all directions perpendicular to $\mathbf{n}$.

Note that neither type of constant decay achieves selectivity: all patterns with $\lambda^a > \gamma$ grow exponentially until the limits on synaptic weights are reached, and all synapses typically reach $w_{\max}$ if the leading eigenvector has no changes of sign. To achieve selectivity, a constraint is needed.

2.2 Formulation of multiplicative and subtractive constraints

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

[2] We have assumed here that $\mathbf{n}$ has no components in the nullspace of $\mathbf{C} - \gamma\mathbf{1}$, so that $\left(\mathbf{C} - \gamma\mathbf{1}\right)^{-1}\mathbf{n}$ is well-defined. This is easily generalized.

them as $\gamma(\mathbf{w})$ or $\epsilon(\mathbf{w})$. Thus, the constrained equations are[3]

$$\frac{d}{dt}\mathbf{w}(t) = \mathbf{C}\mathbf{w}(t) - \gamma(\mathbf{w})\mathbf{w}(t) \qquad \text{(Multiplicative Constraints)} \tag{3}$$

$$\frac{d}{dt}\mathbf{w}(t) = \mathbf{C}\mathbf{w}(t) - \epsilon(\mathbf{w})\mathbf{n} \qquad \text{(Subtractive Constraints)} \tag{4}$$

The vector $\mathbf{n}$ is a constant. For example, if all synapses have equal subtractive decay rate, then $\mathbf{n}=(1,1,\ldots,1)^{\mathrm{T}}$ in the synaptic basis; we refer to this vector, $(1,1,\ldots,1)^{\mathrm{T}}$, as the DC vector.

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 $\sum_i w_i$. 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: [5, 22, 23, 24, 34, 35, 41, 43, 44]; S1: [27, 28]). We define a type 1 constraint more generally as one that conserves the dot product $\mathbf{n}\cdot\mathbf{w}$ for some constant constraint vector $\mathbf{n}$; when $\mathbf{n}$ is the DC vector, the conserved quantity is the total synaptic strength. A type 1 constraint corresponds to a hyperplane constraint surface. We choose the subtracted vector $\mathbf{n}$ in Eq. 4 to be the same as the constraint vector $\mathbf{n}$. Then type 1 constraints can be achieved by choosing

$$\text{M1:} \qquad \gamma(\mathbf{w}) = \mathbf{n}\cdot\mathbf{C}\mathbf{w}/\mathbf{n}\cdot\mathbf{w} \qquad (\text{with } \mathbf{n}\cdot\mathbf{w}(t=0)\neq 0) \tag{5}$$

$$\text{S1:} \qquad \epsilon(\mathbf{w}) = \mathbf{n}\cdot\mathbf{C}\mathbf{w}/\mathbf{n}\cdot\mathbf{n} \tag{6}$$

These choices yield $\mathbf{n}\cdot\frac{d}{dt}\mathbf{w}=\frac{d}{dt}(\mathbf{n}\cdot\mathbf{w})=0$ under Eqs. 3 or 4 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.* [17, 32]. We will consider only multiplicative enforcement of this constraint,[4] called M2. M2 can be achieved by choosing

$$\text{M2:} \qquad \gamma(\mathbf{w}) = \mathbf{w}\cdot\mathbf{C}\mathbf{w}/\mathbf{w}\cdot\mathbf{w} \tag{7}$$

This yields $2\mathbf{w}\cdot\frac{d}{dt}\mathbf{w}=\frac{d}{dt}(\mathbf{w}\cdot\mathbf{w})=0$ under Eq. 3.

[3] To understand why the term $-\gamma(\mathbf{w})\mathbf{w}(t)$ represents multiplicative constraints, 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. 3.

[4] Subtractive enforcement, S2, does not work in the typical case (Theorem 4) 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. S2 constraints work if weights are non-negative ($w_{\min}=0$) and $\mathbf{n}=(1,1,\ldots,1)^{\mathrm{T}}$; but typically type 2 constraints are used in situations in which weights may take either sign.

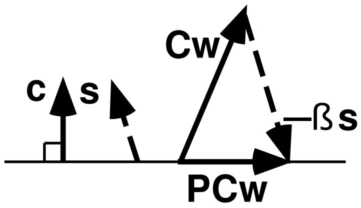

Figure 2: **Projection Onto the Constraint Surface**
The projection operator is $\mathbf{P} = \mathbf{1} - \frac{\mathbf{s}\mathbf{c}^{\mathrm{T}}}{\mathbf{s}\cdot\mathbf{c}}$. This acts on the unconstrained derivative $\mathbf{Cw}$ by removing its $\mathbf{c}$ component, projecting the dynamics onto 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 constraints, $\mathbf{s} = \mathbf{n}$.

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 $\mathbf{P} = \mathbf{1} - \frac{\mathbf{n}\mathbf{n}^{\mathrm{T}}}{\mathbf{n}\cdot\mathbf{n}}$; for M1, it is $\mathbf{P} = \mathbf{1} - \frac{\mathbf{w}\mathbf{n}^{\mathrm{T}}}{\mathbf{w}\cdot\mathbf{n}}$; and for M2, it is $\mathbf{P} = \mathbf{1} - \frac{\mathbf{w}\mathbf{w}^{\mathrm{T}}}{\mathbf{w}\cdot\mathbf{w}}$. We can write these operators as $\mathbf{P} = \mathbf{1} - \frac{\mathbf{s}\mathbf{c}^{\mathrm{T}}}{\mathbf{s}\cdot\mathbf{c}}$, 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 $\mathbf{s}$. 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 $\mathbf{c}\cdot\mathbf{w} = \text{constant}$. When the total synaptic strength is conserved (type 1 constraint), the constraint vector is $\mathbf{c} = \mathbf{n}$; while if the sum–squared strength is conserved (type 2 constraint), the constraint vector is $\mathbf{c} = \mathbf{w}$. Note in Fig. 2 that S1 ($\mathbf{s} = \mathbf{c} = \mathbf{n}$) represents perpendicular projection onto the constraint surface;[5] M1 ($\mathbf{s} = \mathbf{w}$, $\mathbf{c} = \mathbf{n}$) is projection along $\mathbf{w}$; and M2 ($\mathbf{s} = \mathbf{c} = \mathbf{w}$) is both perpendicular projection and projection along $\mathbf{w}$.

2.3 Dynamical effects of multiplicative and subtractive constraints

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, $\mathbf{e}^1$. We have illustrated two typical cases for M1 and S1 constraints:

- In Fig. 3A we have drawn the principal eigenvector $\mathbf{e}^0$ close in direction to the M1 and S1 constraint vector $\mathbf{n}$. This is typical for Hebbian learning when there are only positive correlations

[5]One could also consider a non-perpendicular form of S1 constraints, S1′, in which $\mathbf{s}$ and $\mathbf{c} = \mathbf{n}$ are constants but $\mathbf{s} \neq \mathbf{c}$ (and $\mathbf{s}\cdot\mathbf{c} \neq 0$). The results for S1 constraints can be generalized for S1′ constraints without major changes. S1 constraints may turn into S1′ when viewed in alternative synaptic representations; see Appendix B.

and the total synaptic sum is conserved. Positive correlations lead to a principal eigenvector in which all weights have a single sign; this is close in direction to the DC constraint vector $(1, 1, \ldots, 1)^{\mathrm{T}}$, which conserves total synaptic strength. Multiplicative and subtractive constraints lead to very different outcomes in this case: multiplicative constraints lead to convergence to $\mathbf{e}^0$, whereas subtractive constraints lead to unstable flow. The outcome in this case was illustrated in Figs. 1A,B.

- In Fig. 3B we have drawn the principal eigenvector $\mathbf{e}^0$ perpendicular to the M1 and S1 constraint vector $\mathbf{n}$. This is typical for Hebbian learning when correlations among input activities oscillate in sign as a function of input separation [25]. Such oscillations lead to a principal eigenvector in which weights oscillate in sign and sum approximately to zero. In this case, the type of constraint enforcement makes little difference: the weight vector typically flows to a saturated version of the principal eigenvector.

Under M2 constraints (Fig. 3C), the dynamics converge to $\mathbf{e}^0$.

Figure 3 may be taken as a visual aid for the remainder of this section.

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

The results of this subsection apply to any multiplicatively or subtractively constrained dynamics that are described by Eq. 3 or 4. Our overall conclusions are as follows:

Conclusion 1 *Under a multiplicatively enforced constraint, the weight vector* $\mathbf{w}$ *generally converges to a multiple of the principal eigenvector* $\mathbf{e}^0$ *that satisfies the constraint, provided such a point exists.*

Conclusion 2 *Under a subtractively enforced constraint, the dynamics have no stable outcome within the hypercube of allowed synaptic weights. Generally all weights saturate at* $w_{\max}$ *or* $w_{\min}$.

To establish these points, we will 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 $\frac{d}{dt}\mathbf{w} = 0$. If $\mathbf{C}$ is symmetric the constrained dynamics must either flow to a stable interior fixed point or else flow to the hypercube.

The locations of the fixed points follow trivially from Eqs. 3-4 and the fact that the dynamics remain on the constraint surface:

Theorem 1 *Under a multiplicatively enforced constraint, the interior fixed points 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* $\mathbf{C}\mathbf{w} \propto \mathbf{w}$.

Theorem 2 *Under a subtractively enforced constraint, the interior fixed points are the points* $\mathbf{w}$ *on the constraint surface at which* $\mathbf{C}\mathbf{w} \propto \mathbf{n}$.

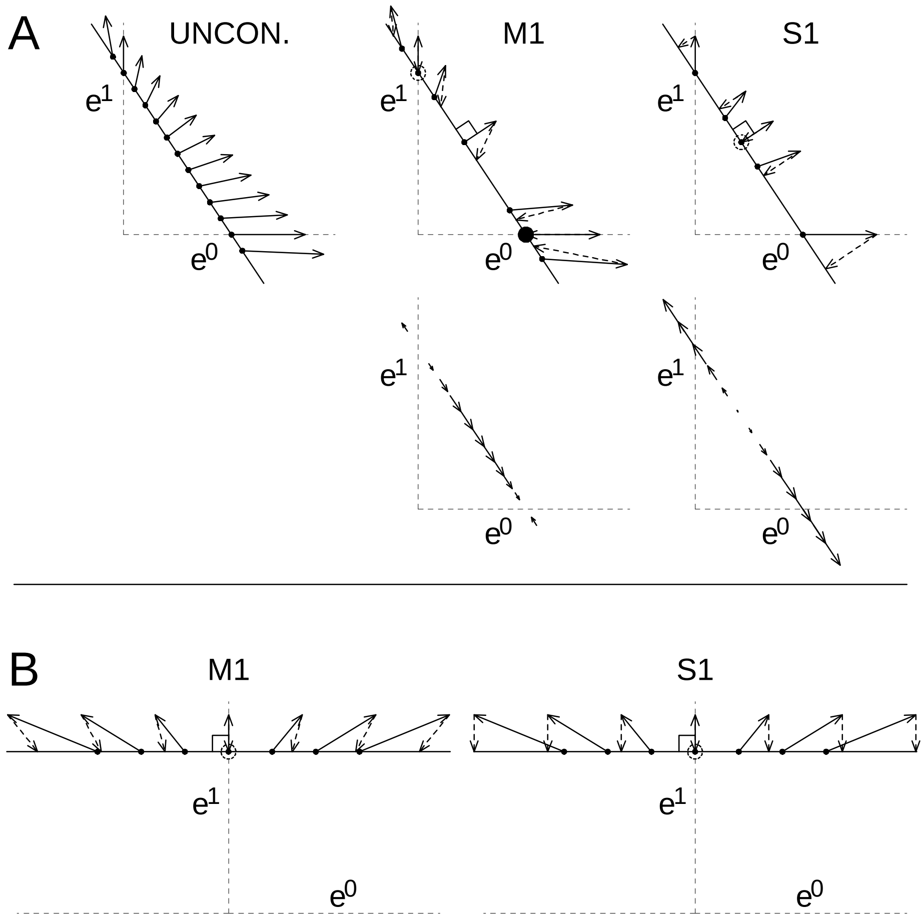

Figure 3: **Dynamics Under Multiplicative and Subtractive Constraints**
Dynamics in the plane formed by the principal eigenvector of $\mathbf{C}$, $\mathbf{e}^0$, and one other eigenvector with positive eigenvalue, $\mathbf{e}^1$.
A. M1 and S1 constraints when e^0 is close in direction to $\mathbf{n}$. 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 in drawing, we have chosen the constraint direction $\mathbf{n} = \mathbf{e}^1$.

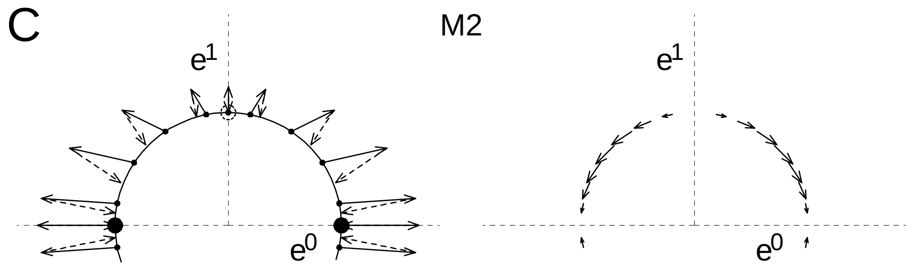

Figure 3: (continued from previous page)
C. M2 constraints. The return path is in the direction of $\mathbf{w}$, as for M1. Thus, locally (for example, near the fixed points) the dynamics are like M1. On a large scale, the dynamics differ because of the difference in constraint surface. Left, unconstrained 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 these fixed points can be determined as shown in Fig. 3. Under multiplicative constraints, a fixed point in the direction of a non-principal eigenvector $\mathbf{e}^1$ is always unstable to a perturbation in the direction of the principal eigenvector $\mathbf{e}^0$, so only $\mathbf{e}^0$ can be a stable fixed point. Subtractive constraints can only stabilize the growth of $\mathbf{w}$ in a single direction, the direction of the constraint vector $\mathbf{n}$. Growth in all directions perpendicular to $\mathbf{n}$, that is, in all directions along the constraint surface, is unaffected by the constraints. Therefore, if there are at least two unstable directions in the absence of constraints, the constrained dynamics will be unstable.

Generalizing this reasoning, it is easy to prove (appendix A):

Theorem 3 *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 4 *Under a subtractively enforced constraint, if* $\mathbf{C}$ *has at least two eigenvectors with positive eigenvalues, then any interior fixed point is unstable.*

A case of theorem 3 for M2 constraints was proven by Oja [32].

Theorem 4 shows that S1 dynamics are unstable when no synapse is saturated. The following condition ensures that subtractively constrained dynamics remain unstable until all synapses have saturated (Appendix A):

Theorem 5 *Let i and j be indices in the synaptic basis. Suppose, for all i and j, $C_{ii} > 0$ and $C_{ii}C_{jj} > C_{ij}C_{ji} > 0$. Then under a subtractively enforced 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; this ensures that (1) $C_{ii} > 0$; (2) $C_{ii} > C_{ij}$ for all i, j; (3) $C_{ij} = C_{ji}$. Theorem 5 generalizes a result proven for a particular model by Linsker [18].

The conditions ensure that whenever at least two synapses are not saturated, the unconstrained dynamics have at least two unstable directions. This 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.

Theorem 5 explains the sharpening of the receptive field that occurs under subtractive constraints when $w_{\max}$ is increased or the total synaptic strength w_{tot} is decreased (Fig. 1: compare column 4 to column 3). Since all synapses reach saturation, the final number of nonzero synapses when $w_{\min} = 0$ is approximately $w_{\text{tot}}/w_{\max}$. 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 $w_{\min}$. Thus, a practical implication of theorem 5 is that an upper limit on synaptic strengths, $w_{\max}$, is needed for stability under subtractive constraints (whereas no such limit is needed under multiplicative constraints).

2.3.2 When constraints do not affect the dynamics: AC vectors under M1 and S1

The principal eigenvector $\mathbf{e}^0$ typically dominates unconstrained development: the outcome is expected to be approximately a saturated version of $\mathbf{e}^0$. Consider the case when the principal eigenvector is parallel to the M1 or S1 constraint surface: $\mathbf{e}^0 \cdot \mathbf{n} = 0$ (Fig. 3B). Then, growth of the principal eigenvector component of the weight vector does not alter $\mathbf{w} \cdot \mathbf{n}$ and so does not violate the type 1 constraint. In this case, addition of M1 or S1 constraints should not substantially alter the unconstrained outcome, although the growth of secondary eigenvectors may be altered.

We define:

Definition 1 *An AC vector* $\mathbf{w}$ *is one that is parallel to the M1 or S1 constraint surface: that is,* $\mathbf{w} \cdot \mathbf{n} = 0$.

When $\mathbf{n}$ is the DC vector $(1, 1, \ldots, 1)^{\mathrm{T}}$, then an AC vector $\mathbf{w}$ is one that varies in sign so that $\sum_i w_i = 0$. Thus, an AC vector can grow without violating the type 1 constraint: it represents the growth of some weights and compensating shrinkage of others, so that the total synaptic strength remains unchanged. 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 regarded as an AC vector.

Conclusion 3 *If the principal eigenvector* $\mathbf{e}^0$ *of* $\mathbf{C}$ *is an AC vector, then under M1 or S1 or without constraints, the dynamics are unstable and the weight vector is expected to flow to a saturated version of* $\mathbf{e}^0$.

When conclusion 3 applies, there may still be important differences between M1 and S1. One difference is that M1 dynamics may be stabilized by a lower weight limit, whereas S1 requires both a lower and an upper limit since all synapses become saturated. Other differences are in the growth rate of $\mathbf{e}^0$ and the fate of non-principal eigenvectors, as discussed in the next two subsections.

2.3.3 The outcome under multiplicative constraints

Conclusion 4 *Under M2 constraints, the weight vector* $\mathbf{w}$ *always converges to the principal eigenvector* $\mathbf{e}^0$ *of* $\mathbf{C}$ *(ignoring possible effects of the hypercube).*

Conclusion 5 *Under M1 constraints, the dynamics typically converge to* $\mathbf{e}^0$ *when* $\mathbf{e}^0$ *is not an AC vector. When* e^0 *is AC, the growth rate of all eigenvectors is suppressed, but not sufficiently to achieve stability.*

A multiplicatively enforced constraint leads to convergence to the principal eigenvector 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. Then for M2 constraints, both of these conditions are always satisfied (Fig. 3C). Under M1 constraints, consider first the case in which condition (1) is satisfied (Fig. 3A). The dynamics then may either flow to infinity or flow to the stable fixed point. Let $\beta\mathbf{e}^0$ be the stable fixed point, and let $\mathbf{w}(t=0)$ be the initial weight vector on the constraint surface. Then the dynamics flow to the stable fixed point if $\mathbf{w}(t=0)\cdot(\beta\mathbf{e}^0) > 0$. This inequality holds in two typical cases: (1) If the initial weight vector and principal eigenvector are both all positive; or (2) If $\mathbf{w}(t=0) \propto \mathbf{n}$ (regardless of the structure of $\mathbf{e}^0$);[6] this is approximately true when $\mathbf{n}$ is the DC vector $(1,1,\ldots,1)^{\mathrm{T}}$ and weights are initialized as small fluctuations about some nonzero mean.

If $\mathbf{e}^0$ is AC (Fig. 3B; conclusion 3), the dynamics under M1 constraints are qualitatively similar to those when $\mathbf{e}^0$ intersects the constraint surface, but growth rates are not suppressed sufficiently to achieve stability. Let $\mathbf{e}^1$ be the principal non-AC eigenvector, with eigenvalue λ^1, and assume that $\mathbf{e}^1$ intersects the constraint surface within the hypercube. Then under M1 constraints, the non-AC part of the weight vector converges to a stabilized version of $\mathbf{e}^1$.[7] Stabilization of $\mathbf{e}^1$ is achieved by lowering the growth rates of all eigenvectors by λ^1. If the principal eigenvector is AC, it will grow but with rate reduced by λ^1.

2.3.4 The outcome under S1 constraints

Conclusion 6 *Under S1 constraints, AC eigenvectors of* $\mathbf{C}$ *grow without constraint; non-AC eigenvectors of* $\mathbf{C}$ *are replaced by AC vectors with decreased growth rate. The dynamics may be dominated either by the principal AC eigenvector of* $\mathbf{C}$*, or by the AC vector derived from the principal eigenvector.*

[6] Let the constraint conserve $\mathbf{w}\cdot\hat{\mathbf{n}} = w_n$. Then $\beta\mathbf{e}^0\cdot\hat{\mathbf{n}} = w_n$. Suppose proportionality holds: $\mathbf{w}(t=0) \approx w_n\hat{\mathbf{n}}$. Then $\mathbf{w}(t=0)\cdot(\beta\mathbf{e}^0) \approx w_n^2 > 0$.

[7] This requires that a generalized basin of attraction condition holds: $\mathbf{w}(t=0)\cdot\beta\mathbf{e}^1 > 0$, where $\beta\mathbf{e}^1$ satisfies the constraint. The proof of the text statement is identical to that of theorem 3, considering the dynamics in the subspace of non-AC vectors.

We 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, $\hat{\mathbf{n}} \equiv \mathbf{n}/|\mathbf{n}|$. Since $w_n \equiv \mathbf{w} \cdot \hat{\mathbf{n}}$ is conserved, we can write $\mathbf{w}(t) = \mathbf{Pw}(t) + w_n\hat{\mathbf{n}}$. So the dynamics can be written:

$$\frac{d}{dt}\mathbf{w}(t) = \mathbf{PCPw}(t) + w_n\mathbf{PC}\hat{\mathbf{n}}. \tag{8}$$

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

Thus, S1 constraints lead to an altered linear dynamics driven by $\mathbf{PCP}$ rather than by $\mathbf{C}$. These dynamics are characterized by three features [20, Appendices B and E]:[8]

1. An AC eigenvector of $\mathbf{C}$ is also an eigenvector of $\mathbf{PCP}$ with identical eigenvalue. Thus, AC eigenvectors grow freely under S1, at the same rate as they would in the absence of constraints;

2. Each non-AC eigenvector of $\mathbf{C}$ is replaced by a corresponding AC eigenvector of $\mathbf{PCP}$ with smaller eigenvalue:[9] for example, an all-positive, centrally peaked non-AC eigenvector of $\mathbf{C}$ may be replaced by a center-surround (positive center and negative surround) AC eigenvector of $\mathbf{PCP}$. Eigenvalue order is preserved under this correspondence.

3. The constant vector $w_n\mathbf{PC}\hat{\mathbf{n}}$ boosts the growth rate of the eigenvectors of $\mathbf{PCP}$ that compose it. These are the eigenvectors derived from the non-AC eigenvectors of $\mathbf{C}$. If w_n is sufficiently large relative to $w_{\max}$, the dynamics will be influenced or dominated by this term; smaller initial weight variance increases this influence [20, section 6.2]. Otherwise, $\mathbf{w}$ grows sufficiently that this term becomes insignificant compared to the $\mathbf{PCPw}$ term.

The leading eigenvector of $\mathbf{PCP}$ may be either the leading AC eigenvector of $\mathbf{C}$, or the AC vector corresponding to $\mathbf{e}^0$. The latter vector's growth rate is increased by the constant term $w_n\mathbf{PC}\hat{\mathbf{n}}$. Both vectors may be very different from $\mathbf{e}^0$. Thus subtractive constraints, unlike multiplicative, may lead to domination by a vector very different from the principal eigenvector of $\mathbf{C}$.

2.3.5 What is maximized under multiplicative and subtractive constraints?

Conclusion 7 *Under multiplicative constraints, the outcome is usually the maximally self-correlated weight vector. Under subtractive constraints, the outcome is often dominated by the maximally self-correlated AC weight vector.*

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 $\hat{\mathbf{w}}$.

[8] The proofs in Appendices B and E of [20] take $\mathbf{n}$ to be the DC vector, but are true for any vector $\mathbf{n}$. They depend only on the facts that (1) the matrix $\mathbf{J} \equiv \mathbf{nn}^T$ has only one eigenvector with non-zero eigenvalue, namely $\hat{\mathbf{n}}$; (2) this nonzero eigenvalue is positive (and equal to $\mathbf{n} \cdot \mathbf{n}$); (3) $\mathbf{PJ} = 0$ and $\mathbf{Pn} = 0$ where $\mathbf{P} = \mathbf{1} - \hat{\mathbf{n}}\hat{\mathbf{n}}^T$.

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

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 $E = \frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{PCPw} + w_n\mathbf{w}^{\mathrm{T}}\mathbf{PCn}$ [20]. 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 AC directions $\hat{\mathbf{w}}$. If $\mathbf{n} = (1, 1, \ldots, 1)^{\mathrm{T}}$, then $\mathbf{e}_0^{\mathbf{P}}$ 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 $\mathbf{e}_0^{\mathbf{P}}$ will tend to $w_{\max}$, while negative weights in $\mathbf{e}_0^{\mathbf{P}}$ will tend to $w_{\min}$.

2.3.6 Summary: The outcome under M1, M2 and S1

Multiplicative and subtractive constraints lead to dramatically different outcomes in many cases. In particular, under a type 1 constraint, multiplicative constraints stabilize a dominant non-AC pattern, whereas subtractive constraints suppress the growth of non-AC patterns. We may summarize as follows:

1. M2 constraints lead to a stabilized version of the principal eigenvector of $\mathbf{C}$;

2. If the principal eigenvector $\mathbf{e}^0$ of $\mathbf{C}$ is a non-AC 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 an AC vector that grows to complete saturation, superimposed on the constrained background $(\mathbf{w} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$. The dominant AC vector may be either

      i. An AC vector derived from $\mathbf{e}^0$, as in Fig. 1A,B; or

      ii. The principal AC eigenvector of $\mathbf{C}$.

3. If the principal eigenvector $\mathbf{e}^0$ of $\mathbf{C}$ is an AC vector, as is typical for Hebbian learning when correlations among input activities oscillate in sign, then the unconstrained dynamics or M1 or S1 constraints all lead to saturated versions of $\mathbf{e}^0$.

2.4 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 an AC vector; in Fig. 1A,B, this vector is a center-surround vector derived from $\mathbf{e}^0$.

The results of section 2.3.4 explain some of the results found by Linsker [18]. 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 [20, 21], the leading eigenvector of $\mathbf{C}$ is an all-positive eigenvector, and the AC vector derived from this is a center-surround vector. The leading AC 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 $w_n\mathbf{PC\hat{n}}$ 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 $\mathbf{PCP}$.

2.5 Extension to two input layers

We now consider the case in which two initially 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 [28]. Similarly, ON-center and OFF-center cells make initially equivalent projections to visual cortex [27].

In this context an additional critical distinction emerges between subtractive and multiplicative constraints. A type 1 constraint conserves the total synaptic strength: this is the sum of the two projections. A pattern of the difference between the two projections has zero total synaptic strength: growth of such a pattern means that one eye's synapses grow wherever the other eye's synapses shrink. Therefore the difference between the two projections is an AC vector. This grows freely under subtractive constraints; whereas under multiplicative constraints, development of this difference is suppressed unless such a pattern is the principal eigenvector of the unconstrained development.

To examine this in more detail, let the synaptic weight vector be $\mathbf{w} \equiv (\mathbf{w}^1, \mathbf{w}^2)^{\mathrm{T}}$ where 1 and 2 label the two input innervations. The equation governing the unconstrained development of $\mathbf{w}$ is $\frac{d}{dt}\mathbf{w} = \mathbf{C}\mathbf{w}$ or

$$\frac{d}{dt}\begin{pmatrix} \mathbf{w}^1 \\ \mathbf{w}^2 \end{pmatrix} = \begin{pmatrix} \mathbf{C}^{11} & \mathbf{C}^{12} \\ \mathbf{C}^{21} & \mathbf{C}^{22} \end{pmatrix} \begin{pmatrix} \mathbf{w}^1 \\ \mathbf{w}^2 \end{pmatrix} \tag{9}$$

Subtractive constraints subtract $\epsilon(\mathbf{w})(\mathbf{n}, \mathbf{n})^{\mathrm{T}}$ from the right side of the equation, while multiplicative constraints subtract $\gamma(\mathbf{w})\mathbf{w} = \gamma(\mathbf{w})(\mathbf{w}^1, \mathbf{w}^2)^{\mathrm{T}}$. By equivalence of the two input projections, $\mathbf{C}^{11} = \mathbf{C}^{22}$, $\mathbf{C}^{12} = \mathbf{C}^{21}$. We can thus diagonalize this equation by converting to sum and difference variables: $\mathbf{w}^{\mathrm{S}} \equiv \mathbf{w}^1 + \mathbf{w}^2$, $\mathbf{w}^{\mathrm{D}} \equiv \mathbf{w}^1 - \mathbf{w}^2$. Defining $\mathbf{C}^{\mathrm{S}} \equiv \mathbf{C}^{11} + \mathbf{C}^{12}$, $\mathbf{C}^{\mathrm{D}} \equiv \mathbf{C}^{11} - \mathbf{C}^{12}$, the unconstrained equations become

$$\frac{d}{dt}\begin{pmatrix} \mathbf{w}^{\mathrm{S}} \\ \mathbf{w}^{\mathrm{D}} \end{pmatrix} = \begin{pmatrix} \mathbf{C}^{S} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^{D} \end{pmatrix} \begin{pmatrix} \mathbf{w}^{\mathrm{S}} \\ \mathbf{w}^{\mathrm{D}} \end{pmatrix} \tag{10}$$

In this representation, subtractive constraints subtract $2\epsilon(\mathbf{w})(\mathbf{n},\mathbf{0})^{\mathrm{T}}$, while multiplicative constraints again subtract $\gamma(\mathbf{w})\mathbf{w}=\gamma(\mathbf{w})(\mathbf{w}^{\mathrm{S}},\mathbf{w}^{\mathrm{D}})$.

If $\mathbf{e}_a^{\mathrm{S}}$, $\mathbf{e}_a^{\mathrm{D}}$ are the eigenvectors of $\mathbf{C}^{\mathrm{S}}$, $\mathbf{C}^{\mathrm{D}}$, respectively, with eigenvalues λ_a^{S}, λ_a^{D}, then the eigenvectors of Eq. 10 are of two types: sum eigenvectors $(\mathbf{w}^{\mathrm{S}},\mathbf{w}^{\mathrm{D}})^{\mathrm{T}}=(\mathbf{e}_a^{\mathrm{S}},\mathbf{0})^{\mathrm{T}}$, with eigenvalue λ_a^{S}; and difference eigenvectors $(\mathbf{w}^{\mathrm{S}},\mathbf{w}^{\mathrm{D}})^{\mathrm{T}}=(\mathbf{0},\mathbf{e}_a^{\mathrm{D}})^{\mathrm{T}}$, with eigenvalue λ_a^{D}. The difference vectors are AC vectors: they are orthogonal to the M1 or S1 constraint vector $(\mathbf{n},\mathbf{0})$.

To describe the effects of constraints on ocular dominance segregation, we assume the following scenario [25]. The principal sum eigenvector $\mathbf{e}_{\mathrm{max}}^{\mathrm{S}}$ and the principal difference eigenvector $\mathbf{e}_{\mathrm{max}}^{\mathrm{D}}$ are both all-positive vectors. If there are positive correlations between the activities of the two eyes, then $\lambda_{\mathrm{max}}^{\mathrm{S}}>\lambda_{\mathrm{max}}^{\mathrm{D}}$. If there are no correlations, then $\lambda_{\mathrm{max}}^{\mathrm{S}}=\lambda_{\mathrm{max}}^{\mathrm{D}}$. If between-eye correlations are negative, then $\lambda_{\mathrm{max}}^{\mathrm{D}}>\lambda_{\mathrm{max}}^{\mathrm{S}}$. For ocular dominance segregation to occur, $\mathbf{e}_{\mathrm{max}}^{\mathrm{D}}$ must dominate the constrained development.

Under multiplicative constraints, $\mathbf{w}^{\mathrm{D}}$ cannot grow unless $\lambda_{\mathrm{max}}^{\mathrm{D}}>\lambda_{\mathrm{max}}^{\mathrm{S}}$.[10] Thus, ocular dominance segregation cannot occur under multiplicative constraints unless the two eyes are negatively correlated, as was noted briefly in [28]. Such anticorrelations could be produced by intralaminar inhibition within the LGN; but it seems unlikely that ocular dominance column development depends on anticorrelations, since ocular dominance develops in the presence of vision in some animals, and vision should partially correlate the two eyes.

Under subtractive constraints, $\mathbf{w}^{\mathrm{D}}$ will grow if $\lambda_{\mathrm{max}}^{\mathrm{D}}>0$. The dynamics under S1 constraints may be dominated either by $\mathbf{e}_{\mathrm{max}}^{\mathrm{D}}$, or by the AC vector that derives from $\mathbf{e}_{\mathrm{max}}^{\mathrm{S}}$, depending on which has the faster growth rate (section 2.3.4). In practice, ocular dominance segregation develops under S1 constraints even if there are positive opposite eye correlations of moderate size relative to the same-eye correlations (unpublished observations). Thus, subtractive rather than multiplicative enforcement of constraints appears more appropriate for modeling Hebbian development in visual cortex.

2.6 Bifurcations under multiplicative constraints

In the ocular dominance example just discussed, consider the effect of a change from negative to positive correlations between the eyes. Suppose within-eye correlations $\mathbf{C}^{11}$ are positive, and let between-eye correlations $\mathbf{C}^{12}=\alpha\mathbf{C}^{11}$ for some parameter α. Then if multiplicative constraints are used, leading to selection of the principal eigenvector, then as α passes through 0 there is a discontinuous, qualitative alteration in behavior from ocular dominance segregation ($\alpha<0$; negative between-eye correlations) to no segregation ($\alpha>0$; positive between-eye correlations).

This illustrates a general property. Multiplicative constraints are nonlinear in $\mathbf{w}$: non-principal

[10] If $\lambda_{\mathrm{max}}^{\mathrm{D}}=\lambda_{\mathrm{max}}^{\mathrm{S}}$, then under multiplicative constraints the ratio of the principal eigenvector components, $w_{\mathrm{max}}^{\mathrm{D}}/w_{\mathrm{max}}^{\mathrm{S}}$, does not change under time development, while all other components are suppressed. Typically this ratio is initially small, so w^{D} cannot develop.

eigenvectors are suppressed in favor of the principal eigenvector. Thus, if there is a change in the set of principal eigenvectors, then there is a sudden, qualitative change in the flow and in the outcome under multiplicative constraints. Such a qualitative change in dynamics as a parameter is varied is known as a bifurcation. In contrast, S1 constraints yield linear equations, so the flow and the outcome change smoothly as parameters are varied (so long as the principal eigenvalues remain positive). For example, under S1 small perturbations of α about $\alpha=0$ would lead to only small changes in the eigenvalues of $\mathbf{PCP}$, so the eigenvectors of $\mathbf{PCP}$ would grow independently with only slightly altered rates.

2.7 The map alignment example

In section 1 we presented the example of an unconstrained equation $\frac{d}{dt}w=\mathbf{t}$ as a simple model for the alignment of one map with another (Fig. 1D). Under M1 constraints this becomes $\frac{d}{dt}w=\mathbf{t}-(t_n/w_n)\mathbf{w}$ where $t_n=\hat{\mathbf{n}}\cdot\mathbf{t}$ and $w_n=\hat{\mathbf{n}}\cdot\mathbf{w}$. The weights flow to $\mathbf{w}=(w_n/t_n)\mathbf{t}$. This is the only fixed point, and it is stable to perturbations on the constraint surface. Under S1 constraints, the equation is $\frac{d}{dt}w=\mathbf{t}-t_n\hat{\mathbf{n}}$. In Fig. 1, $\mathbf{n}$ is the DC vector, so $\mathbf{t}-t_n\hat{\mathbf{n}}$ is the Gaussian $\mathbf{t}$ with a constant subtracted from it such that it sums to zero. Thus, $\mathbf{w}$ flows in the direction of this vector — central synapses growing, peripheral synapses shrinking — until it hits the saturating limits.

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: [5, 22, 24, 43]; S1: [27, 28]); or to the total innervation from each input cell (M1: [23, 44]); or to both (M1: [41]; S1: [28]). 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; 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" [26, 28]. Formulation of constraints in the case of a full layer is discussed in Appendix C.

In the studies of a single output cell (section 2), we have found that multiplication affects all patterns on a cell and hence selects and often stabilizes the dominant unconstrained weight pattern; while subtraction only constrains one pattern per cell, and allows the dominant orthogonal (AC) patterns to grow to complete saturation. When applied to a layer of output cells, both multiplicative and subtractive constraints appear to behave as expected from these single cell studies (based on the studies cited in the previous paragraph, and unpublished observations). We have not studied constraints over a full layer in detail. Such constraints are harder to characterize mathematically because the constraint is imposed separately over each cell, so there is one constraint vector per cell rather than one for the entire network. Theorem 5 continues to hold on a cell-by-cell basis, so subtractive constraints are expected to lead to a fully saturated outcome.

In thinking about constraints on both input and output cells, the following heuristic approach appears to be useful. 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 [25]. 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 on which they act. Then an RF or PF that alternates in sign with a wavelength less than or equal to the arbor diameter is approximately AC, that is, it has summed synaptic strength near 0. An RF or PF that alternates with longer wavelength is not AC. Subtractive constraints selectively suppress the growth of non-AC patterns, whereas multiplicative constraints stabilize the growth of a dominant non-AC 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 (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

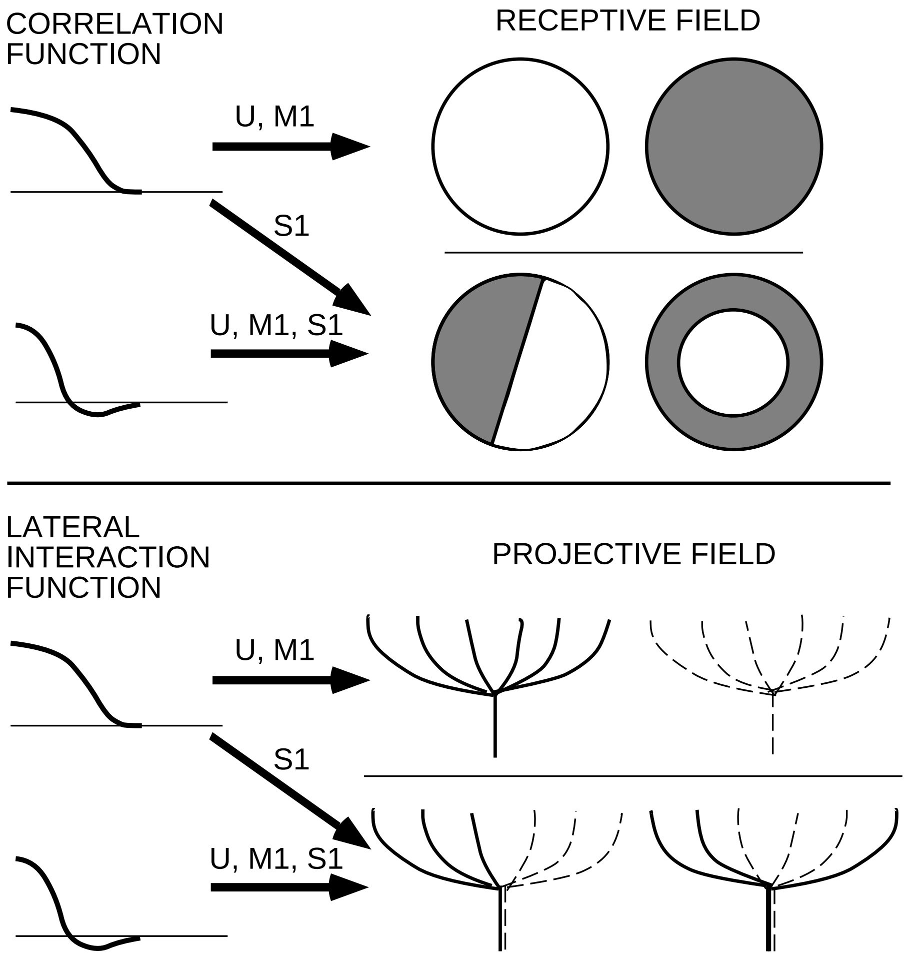

Figure 4: The Role of Constraints on Input and Output Cells: a Heuristic Approach
TOP: Output cell receptive fields (RFs) expected to develop under unconstrained dynamics ('U'), or under M1 or S1 constraints on output cells. White regions in receptive fields indicate positive weights; dark regions indicate zero or negative weights, depending on whether $w_{\rm min}$ 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.

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 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. 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. In some cases such a "knitting" may not be possible: the PF and RF patterns predicted above may be incompatible, in which case the proposed heuristics will fail.

We can apply this analysis 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 2.5). An RF across which $\mathbf{w}^{\mathrm{D}}$ is all-positive or all-negative is a monocular receptive field. When receptive fields are monocular, an oscillation across PFs of $\mathbf{w}^{\mathrm{D}}$ corresponds to the oscillation between ocular dominance columns [25] (Fig. 4). 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 (section 2.5). So, a monocular, DC receptive field pattern of $\mathbf{w}^{\mathrm{D}}$ can develop despite subtractive constraints on output cells, provided that the corresponding correlation function ($C^{\mathrm{D}} \equiv C^{\mathrm{LL}} - C^{\mathrm{LR}}$) has the peak of its Fourier transform at frequency 0. Subtractive *input* constraints do affect $\mathbf{w}^{\mathrm{D}}$: such constraints separately conserve the total strength from the left-eye input and from the right-eye input at each position, and thus conserve the total difference as well as the total sum from each position. So subtractive input constraints ensure that there is an oscillation of $\mathbf{w}^{\mathrm{D}}$ across PFs with wavelength no larger than an arbor diameter. Thus, subtractive input constraints 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 [28].

4 How Can Constraints Be Implemented?

4.1 Learning rules that converge to constrained dynamics

The formulations in Eqs. 5-7 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. 5–7 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. When $|k_2|$ is large, this term enforces the constraint $\mathbf{n} \cdot \mathbf{w} = k_1$. Equivalence of this formulation to that in terms of projection operators, in the limit of large $|k_2|$ and after an initial transient in which the constraint surface is reached, has been shown in [20, Appendix E]. This form of S1 constraint was used by Linsker in [18]. Multiplicative constraints can be similarly formulated: M1 constraints are achieved by setting $\gamma(\mathbf{w}) = k_2(\mathbf{n} \cdot \mathbf{w} - k_1)$ provided $k_2 \mathbf{n} \cdot \mathbf{w} > 0$ always (examine $\mathbf{n} \cdot \frac{d}{dt}\mathbf{w}$). M2 constraints are achieved by setting $\gamma(\mathbf{w}) = |k_2|(\mathbf{w} \cdot \mathbf{w} - k_1)$.

Dynamics that converge to a multiplicative constraint can also be formulated by substituting a constant $k > 0$ for the denominators in Eqs. 5 or 7. Let $\mathbf{c}$ be the constraint vector: for M1, $\mathbf{c} = \mathbf{n}$; for M2, $\mathbf{c} = \mathbf{w}$. Then such convergent dynamics are achieved by setting $\gamma(\mathbf{w}) = \mathbf{c} \cdot \mathbf{C}\mathbf{w}/k$. If the initial condition and dynamics maintain $\mathbf{c} \cdot \mathbf{C}\mathbf{w} > 0$, then 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 [32] used such M2 constraints with $k = 1$ and proved convergence to the principal eigenvector. The self-organizing feature maps of Kohonen [17] also are stabilized in a convergent, multiplicative manner.[11]

4.2 Biological implementations of constraints

Multiplicative or subtractive constraints in their general form (Eqs. 3–4) can be implemented as follows: (1) each of a cell's synaptic weights w_i undergoes decay: the rate of its decay, relative to that of other synapses on the cell, is either proportional to its strength w_i (multiplicative constraints) or given by a constant n_i (subtractive constraints); (2) the overall gain of this decay, $\gamma(\mathbf{w})$ or $\epsilon(\mathbf{w})$, is set for the cell as a whole to achieve the constraint. The separation of the relative decay rate, determined at each synapse, from the overall gain, determined cellularly, allows constraints to be implemented without a cellular assessment of the individual strength of each synapse.

We propose two simple models by which the overall gain of the decay might increase with the total synaptic strength supported by a cell. First, a cell might have a limited capacity to metabolically supply its synapses, so that greater total synaptic strength would mean less supply and thus faster decay for each synapse. This could apply either to the total synaptic strength projected by a cell, or to the total strength received by a cell. 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 lead to greater overall decay

[11] Kohonen's equation for the weights $\mathbf{w}$ onto one output cell is $\frac{d}{dt}\mathbf{w} = y\,[\mathbf{x} - \mathbf{w}]$, where $\mathbf{x}$ is the vector of input activities, y the output cell activity, and y is nonzero only for the most-activated postsynaptic cell and its neighbors. The second, stabilizing term is proportional to $\mathbf{w}$, so the stabilization is multiplicative. The nature of the weight stabilization depends in part on the normalization of input activities. If $\mathbf{n} \cdot \mathbf{x}$=constant, then eventually $\mathbf{n} \cdot \mathbf{w} = \mathbf{n} \cdot \mathbf{x}$ (examine $\mathbf{n} \cdot \frac{d}{dt}\mathbf{w}$). A more common choice is $\mathbf{x} \cdot \mathbf{x}$=constant. Then eventually $\mathbf{w} \cdot \mathbf{w} \leq \mathbf{x} \cdot \mathbf{x}$ (if $\mathbf{w} \cdot \mathbf{w} > \mathbf{x} \cdot \mathbf{x}$, then $\frac{d}{dt}(\mathbf{w} \cdot \mathbf{w}) < 0$); and we expect that eventually $\mathbf{w} \cdot \mathbf{w} \approx \mathbf{x} \cdot \mathbf{x}$ since the winner-take-all activation rule eventually leads only one or a small range of $\mathbf{x}$ to activate the cell, and $\mathbf{w}$ comes to match that $\mathbf{x}$.

by increasing expression or release of a molecule that destabilizes or removes synapses, such as a protease; or by decreasing expression or release of a molecule that stabilizes or adds synapses, such as a trophic, adhesion, or sprouting factor. Many such mechanisms have been proposed (reviewed in [40]). Increased activation might also increase decay by increasing a threshold for synaptic modification; this is discussed in section 4.3.

As simple-minded models of decay, we propose several pictures. First, a synapse's strength might be proportional to its number of postsynaptic receptors. If loss of each of those receptors is an independent event, then decay will occur proportionate to the synapse's strength, yielding multiplicative decay. Replacement would occur at a rate determined by the Hebb rule. Second, a synapse's strength might depend upon the portion of a fixed number of molecules that have been phosphorylated or otherwise activated . Such a model has been proposed in which the molecules are calcium/calmodulin-dependent protein kinase, which is concentrated at the postsynaptic density [19]. Suppose there are m unactivated and m^* activated molecules, $m + m^* = M$. Suppose that Hebbian modification results in promotion of m to m^*,[12] while decay of m^* to m is activity independent and occurs at a rate $-\kappa m^*$. If synaptic strength is linear in m and m^*, $w_i = am_i + bm_i^*$, then this decay is $-\kappa(w_i - aM)/(b-a)$. This is a mixture of subtractive and multiplicative decay. Third, decay might occur when either postsynaptic or presynaptic activity fails to reach a threshold level for positive Hebbian modification. This results in subtractive decay. We examine this in more detail in the next section.

4.3 Use of thresholds to achieve selectivity

Consider a linear activation rule and a linear Hebbian rule for a single output cell:

$$y(t) = \sum_i w_i(t)\,[x_i(t) - x_0] + y_0 \qquad \text{(Activation)} \tag{11}$$

$$\frac{d}{dt}w_i(t) = \lambda[y(t) - y_\theta][x_i(t) - x_\theta] - \gamma w_i - \epsilon \qquad \text{(Hebb)} \tag{12}$$

Here, $y(t)$ is the activation of the output cell, $\mathbf{x}(t)$ the vector of input activations. x_0 and y_0 are baseline levels of input and output activity, respectively. x_θ and y_θ are threshold activity levels for Hebbian plasticity. ϵ and γw_i are synaptic decay terms.

After substituting Eq. 11 into Eq. 12 and averaging over input patterns, and assuming that all inputs have identical mean activity, the thresholds lead to decay terms that are identical for all synapses (subtractive decay). The resulting equation is

$$\frac{d}{dt}w_i(t) = \sum_j Q_{ij} w_j + k_2 \sum_j w_j - n - \gamma w_i \tag{13}$$

where $Q_{ij} = \lambda\,\langle (x_i - \langle x\rangle)\,(x_j - \langle x\rangle)\rangle$, $k_2 = \lambda\,(\langle x\rangle - x_\theta)\,(\langle x\rangle - x_0)$, and $n = \epsilon + \lambda\,(\langle x\rangle - x_\theta)\,(y_\theta - y_0)$. Both the k_2 term and the n term are subtractive decay terms.

[12] Actually this should make Hebbian modification proportional to m, which will complicate the plasticity equation.

Eq. 13 can be rewritten $\frac{d}{dt}w_i(t) = \sum_j Q_{ij}w_j + k_2(\sum_j w_j - k_1) - \gamma w_i$ where $k_1 = n/k_2$. As discussed in section 4.1, for k_2 large and negative this becomes equivalent to S1 constraints conserving $\sum_j w_j = k_1$. Assuming $\langle x \rangle > x_0$, the condition $k_2 < 0$ implies that $\langle x \rangle < x_\theta$. That is, average or weakly activated inputs should lose synaptic strength when the postsynaptic cell is highly activated. This is not the case in at least one system, LTP in hippocampus [7].

Bienenstock *et al.* [1] proposed that the postsynaptic threshold for synaptic plasticity y_θ (Eq. 12) should depend nonlinearly on the average postsynaptic activity $\langle y \rangle$: $y_\theta = \langle y \rangle^2 / y_{\mathrm{set}}$ where y_{set} is a cell's "preset" desired activity level. They refer to this form of postsynaptic threshold as a "sliding threshold". It has the effect of adjusting synaptic strengths to achieve $\langle y \rangle \approx y_{\mathrm{set}}$. They set $y = \mathbf{w} \cdot \mathbf{x}$, so this sliding threshold conserves $\mathbf{w} \cdot \langle \mathbf{x} \rangle \approx y_{\mathrm{set}}$. It thus is an alternative method of achieving a constraint like that of S1 or M1, $\mathbf{w} \cdot \mathbf{n} =$constant, but with constraint vector $\mathbf{n} = \langle \mathbf{x} \rangle$. Bienenstock *et al.* used a nonlinear function $\phi(y - y_\theta)$ in place of $(y - y_\theta)$ in Eq. 12. A more recent formulation [14] sets $y_\theta = \langle y^2 \rangle / y_{\mathrm{set}}$. It will be interesting to determine how each of these rules compares to S1 and M1 on simple problems such as those of Fig. 1. Recent results both in the peripheral auditory system [45] and in hippocampus [13] are suggestive that increased neural activity may elevate a threshold for modification in a manner specific to those inputs whose activity was increased; this is suggestive of an increase in x_θ that is specific to activated inputs, rather than an increase in y_θ, which would elevate thresholds for all inputs.

Covariance models [38, 39] 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 where modification simply depends on the product of input and output activities, $\frac{d}{dt}w_i \propto y x_i$. In a covariance model, one instead takes $\frac{d}{dt}w_i$ proportional to the covariance of y and x_i, $\frac{d}{dt}w_i \propto (y - \langle y \rangle)(x_i - \langle x \rangle)$. This is a Hebb model with linearly sliding thresholds: $y_\theta = \langle y \rangle$, $x_\theta = \langle x \rangle$. With these choices and no intrinsic decay in the Hebb rule, we obtain $k_2 = 0$ and $n = 0$ in Eq. 13, so synaptic growth is driven by the simple Hebbian equation $\frac{d}{dt}\mathbf{w} = \mathbf{Q}\mathbf{w}$ without decays. Thus, the problem of loss of selectivity under a Hebbian rule is not avoided by a linear covariance rule.

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.

Though we have worked in the context of linear Hebbian learning, these results apply to any learning equation for which the unconstrained time development operator has a complete set of eigenvectors and is linear over a significant region. The instability of subtractive constraints also depends on the unconstrained operator having two positive eigenvalues, which is typical for Hebbian learning. An interesting alternative is *anti*-Hebbian learning [30]: 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 [30]; 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 [15]. 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 Eqs. 3 or 4. As in the text, we use indices $a, b, \ldots$ to refer to the eigenvector basis of $\mathbf{C}$. We assume that $\mathbf{C}$ has a complete set of orthonormal eigenvectors $\mathbf{e}^a$ with corresponding eigenvalues λ^a.

Write the constrained equation (Eq. 3 or 4) as $\frac{d}{dt}\mathbf{w} = \mathbf{f}(\mathbf{w})$. To determine the stability of a fixed point $\mathbf{w}^{\mathrm{FP}}$ (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 $D_{ab} = \frac{\partial f_a(\mathbf{w})}{\partial w_b}|_{\mathbf{w}=\mathbf{w}^{\mathrm{FP}}}$. 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 $\mathbf{D}$.[13]

We assume that multiplicative or subtractive 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 that enforces the constraint ($\mathbf{w}$ for multiplicative enforcement, or $\mathbf{n}$ for subtractive).

The stability of a fixed point is determined by the eigenvalues of $\mathbf{D}$ [12]. 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.

We hereafter use eigenvector to mean right eigenvector. We will use the gradient operator, defined by $\nabla g(\mathbf{x}) \equiv \sum_a \mathbf{e}^a \frac{\partial g(\mathbf{w})}{\partial w_a}|_{\mathbf{w}=\mathbf{x}}$.

Theorem 3 *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.*

Proof. We consider a multiplicatively constrained equation (Eq. 3): $\frac{d}{dt}\mathbf{w} = \mathbf{C}\mathbf{w} - \gamma(\mathbf{w})\mathbf{w}$. A fixed point is an eigenvector of $\mathbf{C}$: $\mathbf{w}^{\mathrm{FP}} = w_a \mathbf{e}^a$ for some a, $\gamma(\mathbf{w}^{\mathrm{FP}}) = \lambda^a$. The linearized operator is $\mathbf{D} = \mathbf{C} - \lambda^a \mathbf{1} - w_a \mathbf{e}^a [\nabla\gamma(\mathbf{w}^{\mathrm{FP}})]^{\mathrm{T}}$. 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 $\mathbf{C} - \lambda^a\mathbf{1} - \kappa\mathbf{e}^a[\mathbf{e}^a]^{\mathrm{T}}$, where $\kappa \equiv w_a \frac{\partial\gamma(w)}{\partial w_a}|_{\mathbf{w}=w_a\mathbf{e}^a}$. This is the operator $\mathbf{C}$ with eigenvalue λ^a reduced to $-\kappa$ and all other eigenvalues reduced by λ^a. Note that $\mathbf{e}^a$ is the eigenvector of $\mathbf{D}$ with eigenvalue $-\kappa$; since $\mathbf{e}^a$ is not parallel to the constraint plane, $-\kappa \leq 0$.

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

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. •

Theorem 4 *Under a subtractively enforced constraint, if* $\mathbf{C}$ *has at least two eigenvectors with positive eigenvalues, then any interior fixed point is unstable.*

Proof. We consider a subtractively constrained equation (Eq. 4): $\frac{d}{dt}\mathbf{w} = \mathbf{C}\mathbf{w} - \epsilon(\mathbf{w})\mathbf{n}$. The linearized operator is $\mathbf{D} = \mathbf{C} - \mathbf{n}[\nabla\epsilon(\mathbf{w}^{\text{FP}})]^{\text{T}}$. Let $\mathbf{e}^{\perp}$ be a linear combination of the two eigenvectors of $\mathbf{C}$ with positive eigenvalues, such that $\mathbf{e}^{\perp}\cdot\mathbf{n} = 0$. Then $\mathbf{e}^{\perp}\mathbf{D}\mathbf{e}^{\perp} > 0$; so $\mathbf{D}$ must have a positive eigenvalue. •

Theorem 5 *Let i and j be indices in the synaptic basis. Suppose, for all i and j, $C_{ii} > 0$ and $C_{ii}C_{jj} > C_{ij}C_{ji} > 0$. Then under a subtractively enforced constraint, either all synapses or all but one are saturated in a stable final condition.*

Proof. This is a generalization of a similar proof in [18]. Suppose $\mathbf{C}\mathbf{w}^{\text{FP}} - \epsilon(\mathbf{w})\mathbf{n} =$ "$\mathbf{0}$" when synapses w_i^{FP} and w_j^{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^{th} weight 1 and all other elements 0, and similarly for $\hat{\mathbf{u}}^j$. The action of $\mathbf{C}$ in the $\hat{\mathbf{u}}^i/\hat{\mathbf{u}}^j$ plane is given by the matrix $\begin{pmatrix} C_{ii} & C_{ij} \\ C_{ji} & C_{jj} \end{pmatrix}$. The eigenvalues of this matrix are both real and positive when the conditions of the theorem are met. $\mathbf{D}$ when acting in the $\hat{\mathbf{u}}^i/\hat{\mathbf{u}}^j$ plane is as given in proof of theorem 4 (except that the function $\epsilon(\mathbf{w})$ might be altered when some synapses are saturated). So if $\mathbf{e}^{\perp}$ is a linear combination of $\hat{\mathbf{u}}^i$ and $\hat{\mathbf{u}}^j$ with $\mathbf{e}^{\perp}\cdot\mathbf{n} = 0$, then $\mathbf{e}^{\perp}\mathbf{D}\mathbf{e}^{\perp} > 0$. •

B Constraints in Alternative Synaptic Representations

Suppose that the correlation C_{ij} between two inputs is a function of their spatial positions: $C_{ij} = C(\mathbf{x}_i, \mathbf{x}_j)$ where $\mathbf{x}_i$ is the position of the i^{th} input. The $\mathbf{x}_i$ may not have a regular spatial arrangement: for example, a cell may receive a greater density of inputs from the center of its receptive field than from the periphery. In this case, it is often convenient to use an alternative representation in which the synaptic variables are arrayed in a regular spatial grid, and the density of synapses at each grid position is described by a synaptic density function or 'arbor function'.

Thus, instead of having a label for every individual weight w_i, a set of equally spaced *representative* weights r_n is used, with positions $\mathbf{x}_n$. The value of r_n is the average value of the represented weights. The number of weights represented by r_n is given by the local synaptic density $A_n = A(\mathbf{x}_n)$, where $A(\mathbf{x})$ is the synaptic density function.

Let the dynamics in the $\mathbf{w}$ representation be written $\frac{d}{dt}w_i = \sum_j C^{\mathbf{w}}_{ij} w_j$. Then the dynamics in the $\mathbf{r}$ representation are given by $\frac{d}{dt}r_m = \sum_n C^{\mathbf{r}}_{mn} A_n r_n$. Here, $C^{\mathbf{r}}_{mn} \equiv C(\mathbf{x}_m, \mathbf{x}_n)$, and it is assumed that the function $C(\mathbf{x}_m, \mathbf{x}_n)$ varies slowly on the scale of the spacing between the representative synapses (further discussed in [20, Appendix A]). Define the matrix $\mathbf{A} \equiv \mathrm{diag}\{A_n\}$. Then these dynamics can be written $\frac{d}{dt}\mathbf{r} = \mathbf{L}^{\mathbf{r}}\mathbf{r}$ with $\mathbf{L}^{\mathbf{r}} = \mathbf{C}^{\mathbf{r}}\mathbf{A}$.

One may also consider two alternative bases for the representative weights. First, $\mathbf{v} = \mathbf{A}^{\frac{1}{2}}\mathbf{r}$ is the *symmetrized* basis (the reason for this name follows); here, $\mathbf{A}^{\frac{1}{2}} \equiv \mathrm{diag}\{A_n^{\frac{1}{2}}\}$. Second, $\mathbf{t} = \mathbf{A}\mathbf{r}$ is the *total* represented weight (this representation is used in [25]-[29], where $\mathbf{t}$ is called $\mathbf{S}$). The dynamics in these bases are given by: $\frac{d}{dt}\mathbf{v} = \mathbf{L}^{\mathbf{v}}\mathbf{v}$, with $\mathbf{L}^{\mathbf{v}} = \mathbf{A}^{\frac{1}{2}}\mathbf{C}^{\mathbf{r}}\mathbf{A}^{\frac{1}{2}}$; and $\frac{d}{dt}\mathbf{t} = \mathbf{L}^{\mathbf{t}}\mathbf{t}$, with $\mathbf{L}^{\mathbf{t}} = \mathbf{A}\mathbf{C}^{\mathbf{r}}$. If the operator $\mathbf{C}^{\mathbf{w}}$ is symmetric, the operator $\mathbf{L}^{\mathbf{v}}$ is also symmetric, hence the name "symmetrized" for the $\mathbf{v}$ representation.

The constrained dynamics $\frac{d}{dt}\mathbf{w} = \mathbf{P}^{\mathbf{w}}\mathbf{C}^{\mathbf{w}}\mathbf{w}$ can similarly be translated into constrained dynamics in each representative basis: $\frac{d}{dt}\mathbf{r} = \mathbf{P}^{\mathbf{r}}\mathbf{L}^{\mathbf{r}}\mathbf{r}$, etc. Each projection operator can be written $\mathbf{P} = \mathbf{1} - \frac{\mathbf{s}\mathbf{c}^{\mathrm{T}}}{\mathbf{s}\cdot\mathbf{c}}$ for some subtracted vector $\mathbf{s}$ and constraint vector $\mathbf{c}$ (section 2.2). To describe constraints in the representative bases, we will describe the vectors $\mathbf{s}$ and $\mathbf{c}$ that form the projection operator in each representation.

First, consider a type 1 constraint. Let $\mathbf{n}^{\mathbf{w}}$ by the constraint vector in the $\mathbf{w}$ representation. Define $\mathbf{n}^{\mathbf{r}}$ to be the vector corresponding to $\mathbf{n}^{\mathbf{w}}$ in the $\mathbf{r}$ representation, formed by averaging in the same way that $\mathbf{r}$ is formed from $\mathbf{w}$. Thus, if $\mathbf{n}^{\mathbf{w}}$ is the DC vector $(1,1,\ldots,1)^{\mathrm{T}}$, then $\mathbf{n}^{\mathbf{r}}$ is also $(1,1,\ldots,1)^{\mathrm{T}}$ but with smaller dimension. The constraint vector $\mathbf{c}$ is found by writing the constraint in the appropriate representation: $\mathbf{n}^{\mathbf{w}} \cdot \mathbf{w} = \mathbf{n}^{\mathbf{r}} \cdot \mathbf{A}\mathbf{r} = \mathbf{A}\mathbf{n}^{\mathbf{r}} \cdot \mathbf{r}$ (we have assumed the n_i are approximately constant over a spatial grid position). Thus, in the $\mathbf{r}$ representation, the constraint vector $\mathbf{c}$ is $\mathbf{A}\mathbf{n}^{\mathbf{r}}$. The subtracted vector $\mathbf{s}$ in the $\mathbf{r}$ representation is determined by averaging the vector of the $\mathbf{w}$ representation: thus, the S1 subtracted vector is $\mathbf{n}^{\mathbf{r}}$, while the M1 subtracted vector is $\mathbf{r}$. Equations in the $\mathbf{v}$ and $\mathbf{t}$ bases are found by a change of basis from the $\mathbf{r}$ equations.

M2 constraints cannot be directly translated from the $\mathbf{w}$ representation without knowing the variance in the synaptic weights represented at each spatial grid position. So, we instead consider a similar constraint that conserves $\mathbf{t} \cdot \mathbf{t} = \mathbf{A}^2\mathbf{r} \cdot \mathbf{r}$. Then the $\mathbf{s}$ and $\mathbf{c}$ vectors for each constraint in each representation and basis are:

	S1				M1				M2			
	$\mathbf{w}$	$\mathbf{r}$	$\mathbf{v}$	$\mathbf{t}$	$\mathbf{w}$	$\mathbf{r}$	$\mathbf{v}$	$\mathbf{t}$	$\mathbf{w}$	$\mathbf{r}$	$\mathbf{v}$	$\mathbf{t}$
$\mathbf{s}$	$\mathbf{n}^{\mathbf{w}}$	$\mathbf{n}^{\mathbf{r}}$	$\mathbf{A}^{\frac{1}{2}}\mathbf{n}^{\mathbf{r}}$	$\mathbf{A}\mathbf{n}^{\mathbf{r}}$	$\mathbf{w}$	$\mathbf{r}$	$\mathbf{v}$	$\mathbf{t}$	$\mathbf{w}$	$\mathbf{r}$	$\mathbf{v}$	$\mathbf{t}$
$\mathbf{c}$	$\mathbf{n}^{\mathbf{w}}$	$\mathbf{A}\mathbf{n}^{\mathbf{r}}$	$\mathbf{A}^{\frac{1}{2}}\mathbf{n}^{\mathbf{r}}$	$\mathbf{n}^{\mathbf{r}}$	$\mathbf{n}^{\mathbf{w}}$	$\mathbf{A}\mathbf{n}^{\mathbf{r}}$	$\mathbf{A}^{\frac{1}{2}}\mathbf{n}^{\mathbf{r}}$	$\mathbf{n}^{\mathbf{r}}$	$\mathbf{w}$	$\mathbf{A}^2\mathbf{r}$	$\mathbf{A}\mathbf{v}$	$\mathbf{t}$

The results in the text may be applied to equations in the representative bases in two ways. First, one may consider all analysis to take place in the $\mathbf{w}$ representation, and then translate results to the $\mathbf{r}$, $\mathbf{v}$, or $\mathbf{t}$ representation. Alternatively, one may consider the text equations to be written in a representative basis. In the symmetric basis $\mathbf{v}$, the assumption of a complete orthonormal basis of eigenvectors is satisfied, as is the requirement for S1 constraints that $\mathbf{s} = \mathbf{c}$. Thus, the text equations directly describe S1 or M1 constraints in the $\mathbf{v}$ basis. The formulation of M2 constraints requires that $\mathbf{s} = \mathbf{c}$, which is only satisfied in the $\mathbf{t}$ basis. The $\mathbf{t}$ basis has a complete but not necessarily orthogonal basis of eigenvectors. However, the M2 constraint surface is sufficiently simple that the assumption of orthogonality is not needed for any of our results, so our results directly describe M2 dynamics in the $\mathbf{t}$ basis.

Most of our conclusions are invariant under a change of basis, and hence hold in any of the representative bases since they hold in one. However a few aspects of the results, such as the conditions on $\mathbf{C}$ under which Theorem 5 holds, or the definition of an AC or the DC vector, depend on the choice of basis. Theorem 5 is stated correctly for the $\mathbf{v}$ basis. A vector is AC in the $\mathbf{r}$, $\mathbf{v}$, or $\mathbf{t}$ representations if it is perpendicular to the corresponding type 1 constraint vector $\mathbf{c}$. The DC vector is the type 1 vector $\mathbf{c}$ when $\mathbf{n^r} = (1, 1, \ldots, 1)^{\mathrm{T}}$; if $\mathbf{c}$ is the DC vector, the type 1 constraint conserves total synaptic strength.

Correction added Aug. 30 1995:

The M2 constraint above is incorrectly formulated. The M2 constraint in the $\mathbf{r}$, $\mathbf{v}$, or $\mathbf{t}$ representations should be formulated so as to be identical to the constraint in the $\mathbf{w}$ representation in the case in which there is no variance in the synaptic weights at a given spatial grid position. Let w_n^k, $k = 1, \ldots, A_n$ be the set of weights associated with grid position n. Then the case of no variance is given by $w_n^k \equiv w_n$, independent of k, for all n. Then the M2 constraint conserves $\sum_{n,k}(w_n^k)^2 = \sum_n A_n w_n^2 = \sum_n (A_n^{\frac{1}{2}} w_n)^2 = \mathbf{v} \cdot \mathbf{v}$. So the M2 constraint should have been formulated to conserve $\mathbf{v} \cdot \mathbf{v}$ rather than $\mathbf{t} \cdot \mathbf{t}$. Thus, the constraint vector $\mathbf{c}$ for M2 constraints in the $\mathbf{r}, \mathbf{v}, \mathbf{t}$ representations should have been $\mathbf{A}\mathbf{r}, \mathbf{v}, \mathbf{A}^{-1}\mathbf{t}$ respectively. Since $\mathbf{s} = \mathbf{c}$ in the $\mathbf{v}$ basis also for M2, this means that the text equations directly describe M2, as well as M1 and S1, constraints in the $\mathbf{v}$ representation. [**End of correction.**]

C Constraints on Both Input Cells and Output Cells

In this section we formally describe constraints in the case that a layer of input cells innervates a layer of output cells. We obtain few results from this formalism; we present this section for the use of those who wish to further investigate this problem.

C.1 Formulation of constraints

Let roman letters $x, y, \ldots$ label the output cells, and greek letters α, β label the input cells. The weight vector, $\mathbf{w}$, describes all of the weights in the system: the component $w(x,\alpha)$ corresponds to the synapse from the input at α to the output at x. As in section B, we specify a constraint in terms of a constraint vector $\mathbf{c}$ and a subtracted vector $\mathbf{s}$.

We denote the vector specifying a constraint on the output cell at x_i by $\mathbf{c}_{x_i}$. This vector has nonzero elements only for synapses belonging to the postsynaptic cell at x_i; thus, $c_{x_i}(x,\alpha) = \delta(x, x_i)c_{x_i}(x_i,\alpha)$. Here $\delta(x,y)$ is the Kronecker delta: $\delta(x,y)=1$ if $x=y$; $=0$ otherwise. We adopt the shorthand notation $c_x(\alpha) \equiv c_x(x,\alpha)$. Similarly, $\mathbf{c}_{\alpha_j}$ specifies the constraint on the input cell at α_j, and we write its nonzero elements as $c_{\alpha_j}(x) \equiv c_{\alpha_j}(x,\alpha_j)$. Subtracted vectors $\mathbf{s}$ are treated similarly.

The projection operator that constrains the synaptic inputs onto the cell at x_i is $\mathbf{P}_{x_i} = \mathbf{1} - \frac{\mathbf{s}_{x_i}\mathbf{c}_{x_i}^{\mathrm{T}}}{\mathbf{s}_{x_i}\cdot\mathbf{c}_{x_i}}$. Note that $\mathbf{P}_x\mathbf{P}_y = \mathbf{P}_y\mathbf{P}_x$, so we can unambiguously define the projection operator $\mathcal{P}_{\mathbf{O}}$ that separately constrains each output cell: $\mathcal{P}_{\mathbf{O}} \equiv \Pi_i \mathbf{P}_{x_i} = \mathbf{1} - \sum_i \frac{\mathbf{s}_{x_i}\mathbf{c}_{x_i}^{\mathrm{T}}}{\mathbf{s}_{x_i}\cdot\mathbf{c}_{x_i}}$. Similarly, the operator that constrains the synaptic inputs from the cell at α is $\mathbf{P}_{\alpha_j} = \mathbf{1} - \frac{\mathbf{s}_{\alpha_j}\mathbf{c}_{\alpha_j}^{\mathrm{T}}}{\mathbf{s}_{\alpha_j}\cdot\mathbf{c}_{\alpha_j}}$, and that which separately constrains each input cell is $\mathcal{P}_{\mathbf{I}} \equiv \Pi_j \mathbf{P}_{\alpha_j} = \mathbf{1} - \sum_j \frac{\mathbf{s}_{\alpha_j}\mathbf{c}_{\alpha_j}^{\mathrm{T}}}{\mathbf{s}_{\alpha_j}\cdot\mathbf{c}_{\alpha_j}}$.

Let the unconstrained equation be $\frac{d}{dt}\mathbf{w} = \mathbf{L}\mathbf{w}$; we use $\mathbf{L}$ rather than $\mathbf{C}$ to indicate that this is a linear operator that incorporates both the correlation function $\mathbf{C}$ and the lateral interaction function $\mathbf{I}$ [26, 28]. Then an equation that constrains all output cells is $\frac{d}{dt}\mathbf{w} = \mathcal{P}_{\mathbf{O}}\mathbf{L}\mathbf{w}$, while one that constrains all input cells is $\frac{d}{dt}\mathbf{w} = \mathcal{P}_{\mathbf{I}}\mathbf{L}\mathbf{w}$. The operators $\mathcal{P}_{\mathbf{O}}$ and $\mathcal{P}_{\mathbf{I}}$ do not in general commute (Appendix C.3), so the common practice of sequential application of these two constraints in general only ensures that the last applied is valid: for example, $\frac{d}{dt}\mathbf{w} = \mathcal{P}_{\mathbf{I}}\mathcal{P}_{\mathbf{O}}\mathbf{L}\mathbf{w}$ constrains input cells but not necessarily output cells. A single projection operator can be constructed that enforces both constraints.[14] However in practice, sequential application appears to succeed in giving satisfactory enforcement of both constraints ([25, Fig. 7.16]; [29, Fig. 21]; unpublished observations).

C.2 Fixed points under these constraints

We consider only the case of a constraint over all output cells. Consider first a subtractive constraint, for which $\mathbf{c} = \mathbf{s} = \mathbf{n}$. Typically the conditions of theorem 5 hold on each output cell, because $L_{ij} \propto C_{ij}$ when restricted to a single output cell. Then saturation of all synapses, or all but one, on each output cell is assured, by the same reasoning as for theorem 5. This assures that there are no stable fixed points interior to the hypercube. The interior fixed points (where $\mathcal{P}_{\mathbf{O}}\mathbf{L}\mathbf{w} = 0$) are any points

[14] The operator is created as follows: construct (*e.g.* by Gram-Schmidt orthogonalization) an orthonormal basis that spans the subspace generated by all constraint vectors (both input and output); then use these basis vectors as the constraint vectors in the projection operator.

on the constraint surface where $\mathbf{L}\mathbf{w} = \sum_i \epsilon_i \mathbf{n}_{x_i}$ for some constants ϵ_i. ϵ_i is the amount subtracted by the constraint from synapses on the i^{th} cell.

Next consider multiplicative constraints over all output cells, $\mathbf{c} = \mathbf{w}$. Let $\mathbf{w}_{x_i}$ be the weight vector restricted to the cell at x_i: $w_{x_i}(x, \alpha) = \delta(x, x_i) w(x_i, \alpha)$. Under multiplicative constraints, the interior fixed points are any points on the constraint surface where $\mathbf{L}\mathbf{w} = \sum_i \gamma_i \mathbf{w}_{x_i}$ for some constants γ_i; when all γ_i's are equal, this corresponds to an eigenvector of the system. γ_i is the unconstrained growth rate, and the decay rate due to the constraint, of the i^{th} cell. Note that it may be that no eigenvector is a fixed point: for example if the summed synaptic strength varies between cells in one eigenvector, while the constraint demands that the summed strength be identical over each cell, then the eigenvector cannot satisfy the constraint and cannot intersect the constraint surface.

C.3 Conditions for commutation of $\mathcal{P}_{\mathbf{O}}$ and $\mathcal{P}_{\mathbf{I}}$

The commutator of two operators $\mathbf{X}$ and $\mathbf{Y}$ is defined by $[\mathbf{X}, \mathbf{Y}] \equiv \mathbf{X}\mathbf{Y} - \mathbf{Y}\mathbf{X}$. To determine when $\mathcal{P}_{\mathbf{O}}$ and $\mathcal{P}_{\mathbf{I}}$ commute, we compute the matrix

$$[\mathcal{P}_{\mathbf{O}}, \mathcal{P}_{\mathbf{I}}] = \left[\sum_i \frac{\mathbf{s}_{x_i} \mathbf{c}_{x_i}^{\mathrm{T}}}{\mathbf{s}_{x_i} \cdot \mathbf{c}_{x_i}}, \sum_j \frac{\mathbf{s}_{\alpha_j} \mathbf{c}_{\alpha_j}^{\mathrm{T}}}{\mathbf{s}_{\alpha_j} \cdot \mathbf{c}_{\alpha_j}} \right] = \sum_{i,j} \left\{ \frac{\left(\mathbf{c}_{x_i} \cdot \mathbf{s}_{\alpha_j}\right) \mathbf{s}_{x_i} \mathbf{c}_{\alpha_j}^{\mathrm{T}} - \left(\mathbf{c}_{\alpha_j} \cdot \mathbf{s}_{x_i}\right) \mathbf{s}_{\alpha_j} \mathbf{c}_{x_i}^{\mathrm{T}}}{\left(\mathbf{s}_{x_i} \cdot \mathbf{c}_{x_i}\right) \left(\mathbf{s}_{\alpha_j} \cdot \mathbf{c}_{\alpha_j}\right)} \right\}.$$

Note that $\mathbf{c}_{x_i} \cdot \mathbf{s}_{\alpha_j} = c_{x_i}(\alpha_j) s_{\alpha_j}(x_i)$. For multiplicative constraints, $\mathbf{s}$ is linear in the (variable) weight vector, so this matrix cannot in general be zero. That is, $\mathcal{P}_{\mathbf{O}}$ and $\mathcal{P}_{\mathbf{I}}$ do not in general commute under multiplicative constraints.

To analyze subtractive constraints, we work in the $\mathbf{v}$ representation defined in Appendix B. Then $\mathbf{s} = \mathbf{c} = \mathbf{n}$, so the commutator is $\sum_{i,j} \hat{n}_{x_i}(\alpha_j) \hat{n}_{\alpha_j}(x_i) \left\{ \hat{\mathbf{n}}_{x_i} \hat{\mathbf{n}}_{\alpha_j}^{\mathrm{T}} - \hat{\mathbf{n}}_{\alpha_j} \hat{\mathbf{n}}_{x_i}^{\mathrm{T}} \right\}$, where $\hat{\mathbf{n}} \equiv \mathbf{n}/|\mathbf{n}|$. This is zero iff the matrix $\mathbf{Z} \equiv \sum_{i,j} \hat{n}_{x_i}(\alpha_j) \hat{n}_{\alpha_j}(x_i) \left\{ \hat{\mathbf{n}}_{x_i} \hat{\mathbf{n}}_{\alpha_j}^{\mathrm{T}} \right\}$ is symmetric.

The rows and columns of $\mathbf{Z}$ are each labeled by a particular synapse, so the elements of $\mathbf{Z}$ have indices $(x, \alpha); (y, \beta)$ where (x, α) is the row index and (y, β) is the column index. The only term in the definition of $\mathbf{Z}$ that contributes to the element $Z_{(x,\alpha);(y,\beta)}$ is that with $x_i = x$ and $\alpha_j = \beta$, that is, the single term $\hat{n}_x(\beta) \hat{n}_\beta(x) \left[\hat{\mathbf{n}}_x \hat{\mathbf{n}}_\beta^{\mathrm{T}} \right]$. In particular, $Z_{(x,\alpha);(y,\beta)} = \hat{n}_x(\alpha) \hat{n}_x(\beta) \hat{n}_\beta(x) \hat{n}_\beta(y)$. The transposed element is given by $Z_{(y,\beta);(x,\alpha)} = \hat{n}_y(\alpha) \hat{n}_y(\beta) \hat{n}_\alpha(x) \hat{n}_\alpha(y)$.

A sufficient set of conditions for the symmetry of $\mathbf{Z}$ is that

1. $\hat{n}_\alpha(x) = \hat{n}_\beta(x)$ for all α, β, x;

2. $\hat{n}_x(\alpha) = \hat{n}_y(\alpha)$ for all α, x, y.

The constraint vector for each cell is $\mathbf{A}^{\frac{1}{2}} n^{\mathbf{r}}$ (Appendix B). If $n^{\mathbf{r}}$ is the DC vector $(1, 1, \ldots, 1)$, then $\hat{n}_\alpha(x) = \sqrt{A(x, \alpha)} / \sqrt{\sum_y A(y, \alpha)}$, while $\hat{n}_x(\alpha) = \sqrt{A(x, \alpha)} / \sqrt{\sum_\beta A(x, \beta)}$. Then conditions (1) and (2) are:

1. $\frac{A(x,\alpha)}{\sum_y A(y,\alpha)} = \frac{A(x,\beta)}{\sum_y A(y,\beta)}$ for all α, β, x.

2. $\frac{A(x,\alpha)}{\sum_\gamma A(x,\gamma)} = \frac{A(y,\alpha)}{\sum_\gamma A(y,\gamma)}$ for all α, x, y.

Condition (1) states that an output cell receives the same proportion of every input cell's projection (if x receives 10% of α's projection, it also receives 10% of β's); while condition (2) states that an input cell provides the same proportion of every output cell's synapses (if α provides 10% of x's synapses, it also provides 10% of y's). These conditions for $\mathcal{P}_{\mathbf{O}}$ and $\mathcal{P}_{\mathbf{I}}$ to commute are satisfied if there is uniform connectivity, so that every input cell makes an equal number of synapses on every output cell.

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 for very helpful comments on the manuscript. This collaboration would have been impossible without the internet/NSFnet.

References

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

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

[3] M.C. Brown, J.K.S. Jansen, and D. Van Essen. Polyneuronal innervation of skeletal muscle in new-born rats and its elimination during maturation. *Journal of Physiology*, 261:387–422, 1976.

[4] T. Fladby and J.K.S. Jansen. Postnatal loss of synaptic terminals in the partially denervated mouse soleus muscle. *Acta Physiologica Scandinavia*, 129:239–246, 1987.

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

[6] R.W. Guillery. Binocular competition in the control of geniculate cell growth. *J. Comp. Neurol.*, 144:117–130, 1972.

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

[8] W.P. Hayes and R.L. Meyer. 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, 1988.

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

[10] W.P. Hayes and R.L. Meyer. 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, 1989.

[11] W.P. Hayes and R.L. Meyer. 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, 1989.

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

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

[14] N. Intrator and L.N. Cooper. Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. *Neural Networks*, 5:3–17, 1992.

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

[16] E.I. Knudsen and P.F. Knudsen. Sensitive and critical periods for visual calibration of sound localization by barn owls. *J. Neurosci.*, 10:222–232, 1990.

[17] T. Kohonen. *Self-Organization and Associative Memory.* Springer-Verlag, Berlin, 3rd edition, 1989.

[18] R. Linsker. From basic network principles to neural architecture (series). *Proc. Natl. Acad. Sci. USA*, 83:7508–7512, 8390–8394, 8779–8783, 1986.

[19] J.E. Lisman and M.A. Goldring. Feasibility of long-term storage of graded information by the calcium/calmodulin-dependent protein kinase molecules of the postsynaptic density. *Proc. Nat. Acad. Sci. USA*, 85:5320–5324, 1988.

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

[21] D.J.C. MacKay and K.D. Miller. Analysis of Linsker's simulations of Hebbian rules. *Neural Comput.*, 2:173–187, 1990.

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

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

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

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

[26] K.D. Miller. Derivation of linear Hebbian equations from a nonlinear Hebbian model of synaptic plasticity. *Neural Comput.*, 2:321–333, 1990.

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

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

[29] K.D. Miller and M.P. Stryker. The development of ocular dominance columns: Mechanisms and models. In S.J. Hanson and C.R. Olson, editors, *Connectionist Modeling and Brain Function: The Developing Interface*, pages 255–350. MIT Press/Bradford, Cambridge, MA, 1990.

[30] G. Mitchison. Removing time variation with the anti-Hebbian differential synapse. *Neural Comput.*, 3:312–320, 1991.

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

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

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

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

[35] N. Rochester, J.H. Holland, L.H. Haibt, and W.L. Duda. 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, 1956.

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

[37] G.E. Schneider. Early lesions of superior colliculus: factors affecting the formation of abnormal retinal projections. *Brain, Behav. and Evol.*, 8:73–109, 1973.

[38] T.J. Sejnowski. Statistical constraints on synaptic plasticity. *J. Theor. Biol.*, 69:385–389, 1977.

[39] T.J. Sejnowski. Storing covariance with nonlinearly interacting neurons. *J. Math. Biol.*, 4:303–321, 1977.

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

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

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

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

[44] D.J. Willshaw and C. von der Malsburg. 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, 1979.

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