Abstract

Biological synaptic plasticity exhibits nonlinearities that are not accounted for by classic Hebbian learning rules. Here, we introduce a simple family of generalized nonlinear Hebbian learning rules. We study the computations implemented by their dynamics in the simple setting of a neuron receiving feedforward inputs. These nonlinear Hebbian rules allow a neuron to learn tensor decompositions of its higher-order input correlations. The particular input correlation decomposed and the form of the decomposition depend on the location of nonlinearities in the plasticity rule. For simple, biologically motivated parameters, the neuron learns eigenvectors of higher-order input correlation tensors. We prove that tensor eigenvectors are attractors and determine their basins of attraction. We calculate the volume of those basins, showing that the dominant eigenvector has the largest basin of attraction. We then study arbitrary learning rules and find that any learning rule that admits a finite Taylor expansion into the neural input and output also has stable equilibria at generalized eigenvectors of higher-order input correlation tensors. Nonlinearities in synaptic plasticity thus allow a neuron to encode higher-order input correlations in a simple fashion.

1 Introduction

In Hebbian learning, potentiation of the net synaptic weight between two neurons is driven by the correlation between pre- and postsynaptic activity \( y \). That postulate is a cornerstone of the theory of synaptic plasticity and learning \([2,3]\). In its basic form, the Hebbian model leads to runaway potentiation or depression of synapses, since the pre-post correlation increases with increasing synaptic weight \( y \). That runaway potentiation can be stabilized by supplemental homeostatic plasticity dynamics \( \Delta w \), by weight dependence in the learning rule \([6,7]\), or by synaptic scaling regulating a neuron’s total synaptic weight \( \delta w \). In 1982, Erkki Oja observed that a linear neuron with Hebbian plasticity and synaptic scaling learns the first principal component of its inputs \([10]\). In 1985, Oja and Karhunen proved that this is a global attractor of the Hebbian dynamics \([11]\). This led to a fountain of research on unsupervised feature learning in neural networks \([12,13]\).

Principal component analysis (PCA) describes second-order features of a random variable. Both naturalistic stimuli and neural activity can, however, exhibit higher-order correlations \([14,15]\). Canonical models of retinal and thalamic processing whiten inputs, removing pairwise features \([16,22]\). Beyond-pairwise features, encoded in tensors, can provide a powerful substrate for learning from data \([23,27]\).

The basic Hebbian postulate does not take into account fundamental nonlinear aspects of biological synaptic plasticity in cortical pyramidal neurons. First, synaptic plasticity depends on beyond-pairwise activity correlations \([28,33]\). Second, spatially clustered and temporally coactive synapses exhibit
correlated and cooperative plasticity \cite{34, 31}. There is a rich literature on computationally motivated forms of nonlinear Hebbian learning (section \ref{sec:learning}). Here, we will prove that these biologically motivated nonlinearities allow a neuron to learn higher-order features of its inputs.

We study the dynamics of a simple family of generalized Hebbian learning rules, combined with synaptic scaling (eq. \ref{eq:J}). Equilibria of these learning rules are invariants of higher-order input correlation tensors. The order of input correlation (pair, triplet, etc.) depends on the pre- and postsynaptic nonlinearities of the learning rule. When the only nonlinearity in the plasticity rule is postsynaptic, the steady states are eigenvectors of higher-order input correlation tensors \cite{42, 43}. We prove that these eigenvectors are attractors of the generalized Hebbian plasticity dynamics and characterize their basins of attraction.

Then, we study further generalizations of these learning rules. We show that any plasticity model (with a finite Taylor expansion in the synaptic input, neural output, and synaptic weight) has steady states that generalize those tensor decompositions to multiple input correlations, including generalized tensor eigenvectors. We show that these generalized tensor eigenvectors are stable equilibria of the learning dynamics. Due to the complexity of the arbitrary learning rules, we are unable to fully determine their basins of attraction. We do find that these generalized tensor eigenvectors are in an attracting set for the dynamics, and characterize its basin of attraction. Finally, we conclude by discussing extensions of these results to spiking models and weight-dependent plasticity.

\section{Results}

Take a neuron receiving $K$ inputs $x_i(t)$, $i \in [K]$, each filtered through a connection with synaptic weight $J_i(t)$ to produce activity $n(t)$. We consider synaptic plasticity where the evolution of $J_i$ can depend nonlinearly on the postsynaptic activity $n(t)$, the local input $x_i(t)$, and the current synaptic weight $J_i(t)$. We model these dependencies in a learning rule $f$:

$$J_i(t + dt) = J_i(t) + \frac{dt}{\tau} J_i(t) \left[ f_i(n(t), x_i(t), J_i(t)) \right] \left[ J(t) + \frac{dt}{\tau} f(n(t), x(t), J(t)) \right]$$

where $f_i(t) = n^a_i(t) x^b_i(t) J^c_i(t)$. \label{eq:J}

The parameter $a$ sets the output-dependent nonlinearity of the learning rule, $b$ sets the input-dependent nonlinearity, and $c$ sets its dependence on the current synaptic weight. Eq. \ref{eq:J} assumes a simple form for these nonlinearities; we discuss arbitrary nonlinear learning rules in section \ref{sec:learning}. We assume that $a$ and $b$ are positive integers, as in higher-order voltage or spike timing–dependent plasticity (STDP) models \cite{44, 45}. The scaling by the norm of the synaptic weight vector, $\|J\|$, models homeostatic synaptic scaling \cite{48, 49}. Bold type indicates a vector, matrix, or tensor (depending on the variable) and regular font with lower indices indicates elements thereof. Roman type denotes a random variable ($x$).

We assume that $x(t)$ is drawn from a stationary distribution with finite moments of order $a + b$. Combined with a linear neuron, $n = J^T x$, and a slow learning rate, $\tau \gg dt$, this implies the following dynamics for $J$ (appendix A.1):

$$\tau \dot{J}_i = J_i^c \sum_\alpha \mu_{i,\alpha}(J^@\alpha) - J_i \sum_{j,\alpha} J_j^{c+1} \mu_{j,\alpha}(J^@\alpha). \label{eq:Jdot}$$

In eq. \ref{eq:Jdot}, $\dot{J}_i = dJ_i/dt$, $\alpha = (j_1, \ldots, j_a)$ is a multi-index, and $\otimes$ is the vector outer product; $J^@\alpha$ is the $a$-fold outer product of the synaptic weight vector $J$. $\mu$ is a higher-order moment (correlation) tensor of the inputs:

$$\mu_{i,\alpha} = \langle x_i^\alpha (x^@\alpha) \rangle_x \label{eq:mu}$$

where $\langle \rangle_x$ denotes the expectation with respect to the distribution of the inputs. $\mu$ is an $(a + 1)$-order tensor containing an $(a + b)$-order joint moment of $x$. The order of the tensor refers to its number of indices, so a vector is a first-order tensor and a matrix a second-order tensor. $\mu$ is cubical; each mode of $\mu$ has the same dimension $K$. In the first term of eq. \ref{eq:Jdot} for example, $\sum_\alpha \mu_{i,\alpha}(J^@\alpha)$ takes the dot product of $J$ along modes 2 through $a + 1$ of $\mu$.

\subsection{Steady states of nonlinear Hebbian learning}

If we take $a = b = 1$, then $\mu$ is the second-order correlation of $x$ and $\alpha$ is just the index $j$. With $c = 0$ also, eq. \ref{eq:Jdot} reduces to Oja’s rule and $J$ is guaranteed to converge to the dominant eigenvector
While the top eigenvector of a matrix can be computed efficiently, computing the top eigenvector weight vector and each of the first 10 eigenvectors (components of the Tucker decomposition) of was not symmetric (fig. 1b). The mean squared error of the canonical polyadic (CP) approximation (e.g., fig. 1e, f), around which the weights fluctuated due to the finite learning timescale.

The nonlinear Hebbian learning dynamics converged to an equilibrium from random initial conditions. Our main finding is that with \((b,c) = (1,0)\) in the generalized Hebbian rule, eigenvectors of \(\mu\) are attractors of eq. 2. Contrary to the case when \(a = 1\) (Oja’s rule), the dynamics are thus not guaranteed to converge to the first eigenvector of the input correlation tensor when \(a > 1\). The first eigenvector of \(\mu\) does, however, have the largest basin of attraction.
Theorem 1. In eq. 2 take \((b, c) = (1, 0)\). Let \(\mu\) be a cubical, symmetric tensor of order \(a + 1\) and odeco with \(R\) components:

\[
\mu = \sum_{r=1}^{R} \lambda_r (U_r)^{\otimes(a+1)}
\]  

(6)

where \(U\) is a matrix of unit-norm orthogonal eigenvectors: \(U^T U = I\). Let \(\lambda_i > 0\) for each \(i \in [R]\) and \(\lambda_i \neq \lambda_j\) \((i, j) \in [R] \times [R] \text{ with } i \neq j\). Then for each \(k \in [R]\):

1. With any odd \(a > 1\), \(J = \pm U_k\) are attracting fixed points of eq. 2 and their basin of attraction is \(\bigcap_{i \in [R] \setminus k} \{J : \|U_i^T J / U_k^T J\| < (\lambda_i / \lambda_k)^{(a-1)}/(a-1)\}\). Within that region, the separatrix of \(+U_k\) and \(-U_k\) is the hyperplane orthogonal to \(U_k^T J = 0\).

2. With any even positive \(a\), \(J = U_k\) is an attracting fixed point of eq. 2 and its basin of attraction is \(\{J : U_k^T J > 0\} \cap \bigcap_{i \in [R] \setminus k} \{J : \|U_i^T J / U_k^T J\| < (\lambda_i / \lambda_k)^{(a-1)}/(a-1)\}\).

3. With any even positive \(a\), \(J = 0\) is a neutrally stable fixed point of eq. 2 with basin of attraction \(\{J : \sum_{j=1}^{R} (U_j^T J)^2 < 1 \land U_k^T J < 0 \forall k \in [R]\}\).

Remark. Each component of the orthogonal decomposition of an odeco tensor \(\mu\) (eq. 5) is an eigenvector of \(\mu\). If \(R < (a^{K-1})/(a-1)\), there are additional eigenvectors. The components of the orthogonal decomposition are the robust eigenvectors of \(\mu\): the attractors of its multilinear map \(\mu \circ \circ \cdots \circ \mu\). The non-robust eigenvectors of an odeco tensor are fixed by its robust eigenvectors and their eigenvalues \(\lambda_2, \lambda_3, \ldots, \lambda_R\).
The proof of theorem 1 is given in appendix A.2 To prove theorem 1, we project \( J \) onto the eigenvectors of \( \mu \), and study the dynamics of the loadings \( v = U^T J \). This leads to the discovery of a collection of unstable manifolds: each pair of axes \((i, k)\) has an associated unstable hyperplane \( v_i = v_k (\lambda_k/\lambda_i)^{1/(a-1)} \) (and if \( a \) is odd, also the corresponding hyperplane with negative slope). These partition the phase space into the basins of attraction of the eigenvectors of \( \mu \).

For example, consider a fourth-order input correlation (corresponding to \( a = 3 \) in eq. 1) with two eigenvectors. The phase portrait of the loadings is in fig. 2a, with the nullclines in black and attracting and unstable manifolds in blue. The attracting manifold is the unit sphere. There are two unstable hyperplanes that partition the phase space into the basins of attraction of \((0, \pm 1)\) and \((\pm 1, 0)\), where the synaptic weights \( J \) are an eigenvector of \( \mu \).

For even \( a \), only the unstable hyperplanes with positive slope survive (fig. 2b, blue line). The unit sphere (fig. 2b, blue) is attracting from any region where at least one loading is positive. Its vertices \([1] \times [1]\) are equilibria; \((v_1, v_2) = (1, 0)\) and \((0, 1)\) are attractors and the unstable hyperplane separates their basins of attraction. For the region with all loadings negative that is the basin of attraction of the origin, noise will drive the system away from zero towards one of the eigenvector equilibria.

![Figure 2](image_url)

Figure 2: Example dynamics of the loadings, \( v = U^T J \), for rank-two input correlations. Black curves: nullclines of \( v_1, v_2 \). Blue curves: stable sets and separatrices of the basins of attraction of the sparse attractors for \( v \). a) Odd \( a \) (\( a = 3 \)). The stable set is the unit sphere, and the separatrices are \( v_2 = \pm v_1 (\lambda_1/\lambda_2)^{1/a-1} \). b) Even \( a \) (\( a = 2 \)). The unstable set is \( v_2 = v_1 (\lambda_1/\lambda_2)^{1/a-1} \) (solid blue line). In (a,b), \((\lambda_1, \lambda_2) = (3, 1)\). c) Phase portrait for a two-term learning rule (eq. 2). All parameters of the input correlation tensors \((\lambda_{mri}, A_i)\) are equal to one. Dashed blue curve: \( L = \{ v : \sum_{m,j} \lambda_{mij} v_m v_j^{a-1} = 0 \} \).

By partitioning the phase space of \( J \) into basins of attraction for eigenvectors of \( \mu \), theorem 1 also allows us to determine the volumes of those basins of attraction. The basins of attraction are open sections of \( \mathbb{R}^R \) so we measure their volume relative to that of a large hypercube.

**Corollary 1.1.** Let \( V_k \) be the relative volume of the basin of attraction for \( J = U_k \). For odd \( a > 1 \),

\[
V_k = R^{-1} \prod_{i=1}^{R} \left( \frac{\lambda_k}{\lambda_i} \right)^{1/(a-1)}
\]

(7)

**Corollary 1.2.** Let \( V_k \) be the relative volume of the basin of attraction for \( J = U_k \). For even positive \( a \),

\[
V_k = 2^{1-R} \left( R^{-1} \prod_i \left( \frac{\lambda_k}{\lambda_i} \right)^{1/(a-1)} + (R - 1)^{-1} \sum_{j \neq k} \prod_{l \neq j} \left( \frac{\lambda_k}{\lambda_l} \right)^{1/(a-1)} \right)
\]

\[
+ (R - 2)^{-1} \sum_{j,l \neq k} \prod_{l \neq j} \left( \frac{\lambda_k}{\lambda_l} \right)^{1/(a-1)} + \ldots + 1 \right)
\]

(8)

The calculations for corollaries 1.1, 1.2 are given in appendix A.2 We see that the volumes of the basins of attraction depend on the spectrum of \( \mu \), its rank \( R \), and its order \( a \). The result for odd \( a \) also
We then approximate

We see in theorem 1 that attractors of eq. 2 are points on the unit hypersphere. The volume of the basin of attraction of eigenvector \( k \) is proportional to \( \lambda^k_{\text{max}}^{R/(a-1)} \). An eigenvector with eigenvalue \( \epsilon \) scales the basins of attraction for the other eigenvectors by \( \epsilon^{-1/(a-1)} \). The relative volume of two eigenvectors’ basins of attraction is, however, unaffected by the other eigenvalues whatever their amplitude. With \( a \) odd, the ratio of the volumes of the basins of attraction of eigenvectors \( k \) and \( j \) is \( V_k/V_j = (\lambda_k/\lambda_j)^{R/(a-1)} \).

We see in theorem 1 that attractors of eq. 2 are points on the unit hypersphere, \( S \). For odd \( a \), \( S \) is an attracting set for eq. 2. For even \( a \), the section of \( S \) with at least one positive coordinate is an attracting set (fig. 2a, b, blue circle; see proof of theorem 1 in appendix A.2). We thus next computed the attractors of eq. 2. If \( \mu_k/\mu_j \) does not exhibit the cancellation that \( V_k/V_j \) does for odd \( a \). We saw in simulations with natural image patch inputs and initial conditions for \( J \) chosen uniformly at random on \( S \), the basin of attraction for \( U_j \) was \( \sim 3 \times \) larger than that for \( U_k \) and the higher eigenvectors had negligible basins of attraction (fig. 1).

In Oja’s model, \((a, b, c) = (1, 1, 0)\) in eq. 1 if the largest eigenvalue has multiplicity \( d > 1 \) then the \( d \)-sphere spanned by those codominant eigenvectors is a globally attracting equilibrium manifold for the synaptic weights. The corresponding result for \( a > 1 \) is (for the formal statement and proof, see corollary 1.3 in appendix A.2):

**Corollary 1.3.** (Informal) If any \( d \) robust eigenvalues of \( \mu \) are equal, the \( d \)-sphere spanned by their robust eigenvectors is an attracting equilibrium manifold for the synaptic weights.

### 2.3 Arbitrary learning rules

So far we have studied phenomenological plasticity rules in the particular form of eq. 1. The neural output \( n \), input \( x_i \), and synaptic weight \( J_i \) were each raised to a power and then multiplied together. Changes in the strength of actual synapses are governed by complex biochemical, transcriptional, and regulatory pathways [58]. We view these as specifying some unknown function of the neural output, input, and synaptic weight, \( f(n, x, J) \). That function might not have the form of eq. 1. So, we next investigate the dynamics induced by arbitrary learning rules \( f \). We see here that under a mild condition, any equilibrium of the plasticity dynamics will have a similar form as the steady states of eq. 2. If \( f \) does not depend on \( J \) except through \( n \), equilibria will be generalized eigenvectors of higher-order input correlations.

The Taylor expansion of \( f \) around zero is:

\[
J_i(t) = \sum_{m=1}^{\infty} A_m n^a(t) x_i^b(t) J_i^c(t) = \sum_{m=1}^{\infty} A_m n^a(t) x_i^b(t) J_i^c(t)
\]

(9)

where the coefficients \( A_m \) are partial derivatives of \( f \). We assume that there exists a finite integer \( N \) such that derivatives of order \( N+1 \) and higher are negligible compared to the lower-order derivatives. We then approximate \( f \), truncating its expansion after those \( N \) terms. With a linear neuron, synaptic scaling, and slow learning, this implies the plasticity dynamics

\[
\tau J_i = \sum_{j,m} J_j^m \sum_{\alpha_m} m \mu_{j,x_m} (J^{\otimes \alpha_m})_{\alpha_m} - J_i \sum_{j,m} \sum_{\alpha_m} J_j^{m+1} m \mu_{j,x_m} (J^{\otimes \alpha_m})_{\alpha_m}
\]

(10)

where \( m \mu_{j,x_m} = A_m (x_j^b (x^{\otimes \alpha_m})_{\alpha_m})_{x} \). At steady states where \( J_i \neq 0 \),

\[
\sum_{m} J_j^m \sum_{\alpha_m} m \mu_{j,x_m} (J^{\otimes \alpha_m})_{\alpha_m} = \lambda J_i, \text{ where } \lambda(J, \{m \mu\}) = \sum_{m} \sum_{j,m} \sum_{\alpha_m} J_j^{m+1} m \mu_{j,x_m} (J^{\otimes \alpha_m})_{\alpha_m}
\]

(11)

If each \( c_m = 0 \), this is a kind of generalized tensor eigenvalue:

\[
\sum_{m, \alpha_m} m \mu_{j,x_m} (J^{\otimes \alpha_m})_{\alpha_m} = \lambda J_i
\]

(12)
so that $J$ is invariant under the combined action of the multilinear maps $m\mu$ (which are potentially of different orders). If $a_1 = a_2 = \cdots = a_m$, then this can be simplified to a tensor eigenvector equation by summing the input correlations $m\mu$. If different terms of the expansion of $f$ generate different-order input correlations, however, the steady states are no longer necessarily equivalent to tensor eigenvectors. If there exists a synaptic weight vector $J$ that is an eigenvector of each of those input correlation tensors, $\sum_{\alpha} a_{\alpha} (J^{\otimes \alpha})_{\alpha m} = \lambda J m$ for each $m$, then that configuration $J$ is a steady state of the plasticity dynamics with each $c_m = 0$.

We next investigated whether these steady states were attractors in simulations of a learning rule with a contribution from two-point and three-point correlations ($a = (1, 2)$, $b = 1$, $c = 0$, $A = (1, 1/2)$ in eq. $9$). As before, we used whitened natural image patches for the inputs $x$ (fig. $3a$). The two- and three-point correlations of those image patches had similar first eigenvalues, but the spectrum of the two-point correlation decreased more quickly than the three-point correlation (fig. $3a$). The first three eigenvectors of the different correlations overlapped strongly (fig. $3b$). This was not due to a trivial constant offset since the inputs were whitened. With this parameter set, the synaptic weights usually converged to the (shared) first eigenvector of the input correlations (fig. $3c,d,e$ blue).

![Figure 3](image)

Figure 3: Multi-term Hebbian plasticity rule converges to input correlation tensors’ shared eigenvectors. We study a learning rule with two terms, driven by the second- and third-order input correlations ($a = (1, 2)$, $b = 1$, $c = 0$). a) Tucker spectra of the natural image patches’ correlation tensors. Solid lines: mean over 10 samples of 200 image patches. Shaded regions: standard deviation. b) Overlap of the eigenvectors of the image patches’ two- and three-point correlations (mean over 10 samples of 200 image patches). c, d) Example dynamics for the two-term learning rule. Transparency increases with the eigenvector number in d. e) Distribution of the final alignment to the first eigenvectors over 10 random initial conditions for each of 10 samples of image patches. f) Final alignment to the two- and three-point correlations’ eigenvectors as a function of the weight on the two-point correlation, $A_1$. $A_2 = 1$.

We next asked how the weights of the different input correlations in the learning rule (the parameters $A_1$, $A_2$) affected the plasticity dynamics. When the learning rule weighted the inputs’ three-point correlation more strongly than the two-point correlation ($A \approx (1/2, 1)$), the dynamics converged almost always to the first eigenvector of the three-point correlation (fig. $3a$, blue vs orange). Without loss of generality, we then fixed $A_2 = 1$ and varied the amplitude of $A_1$. As $A_1$ increased, the learning dynamics converged to equilibria increasingly aligned with the top eigenvectors of the input correlations (fig. $3a$). For sufficiently negative $A_1$, the dynamics converged to steady states that were neither eigenvectors of the two-point input correlation nor any of the top 20 eigenvectors of the three-point input correlation (fig. $3a$).

Earlier, we saw that in single-term learning rules, the only attractors were robust eigenvectors of the input correlation (theorem $1$). The dynamics of eq. $1$ usually converged to the first eigenvector...
Then:

While biological synaptic plasticity is certainly more complex than the simple generalized Hebbian rule of eq. 1, neural activity is also more complex than the linear model. We have analyzed biologically motivated plasticity dynamics that generalize the Oja rule. One class of these compute tensor eigenvectors. We proved that without a multiplicative weight-dependence in the plasticity, those eigenvectors are attractors of the dynamics (theorem 1, figs. 1a, b). Contrary to Oja’s rule, the first eigenvector of higher-order input correlations is not a unique attractor. Rather, each eigenvector has a finite basin of attraction, the size of which is proportional to \( \lambda_k^{n/k} \). If there are \( d \) codominant eigenvectors (\( \lambda_1 = \lambda_2 = \ldots = \lambda_d \)), the \( d \)-sphere they span is an attracting equilibrium manifold (corollary 1 in appendix A.2). Furthermore, steady states of any plasticity model with a finite Taylor polynomial in the neural output and inputs are generalized eigenvectors of multiple input correlations. These steady states are stable and attracting (theorem 1, figs. 2, 3).

We next investigated the steady states of multi-term nonlinear Hebbian rules analytically. We focused on the case when the different input correlations generated by the learning rule all have a shared set of eigenvectors. In this case, those shared eigenvectors are all stable equilibria of eq. 10. They are not, however, the only stable equilibria. We see in a simple example that there can be equilibria that are linear combinations of those eigenvectors with all negative weights (in fig. 2, the fixed point in the lower left quadrant on the unit circle). In fact, any stable equilibrium that is not a shared eigenvector must be such a negative combination. Our results are summarized in the following theorem:

**Theorem 2.** In eq. 10 take \( b = 1, c = 0, a \in \mathbb{Z}_N^r \), and consider \( N \) cubical, symmetric tensors, \( m\mu \), each of order \( a_m + 1 \) for \( m \in [N] \), that are mutually odec into \( R \) components:

\[
m\mu = \sum_{r=1}^{R} \lambda_{mr} U_r \otimes U_r \otimes \cdots \otimes U_r
\]

with \( U^T U = I \). Let \( \lambda_{mr} \geq 0 \) and \( \sum_{m} \lambda_{mr} > 0 \) for each \( m, r \in [N] \times [R] \). Let

\[
S(J) = \sum_{i=1}^{R} (U_i^T J)^2 \text{ and } L(J) = \sum_{m=1}^{N} \sum_{i=1}^{R_m} \lambda_{ml}(U_i^T J)^{a_m+1}.
\]

Then:

1. \( S^* = \{ J : S(J) = 1 \land L(J) > 0 \} \) is an attracting set for eq. 10 and its basin of attraction includes \( \{ J : L(J) > 0 \} \).
2. For each \( k \in [R] \), \( J = U_k \) is a stable equilibrium of eq. 10.
3. For each \( k \in [R] \), \( J = -U_k \) is a stable equilibrium of eq. 10 if \( \sum_{m} \lambda_k(-1)^{a_m} < 0 \) (and unstable if \( \sum_{m} \lambda_k(-1)^{a_m} < 0 \)).

4. Any other stable equilibrium must have \( U_k^T J \leq 0 \) for each \( k \in [R] \).

The claims of theorem 2 are proven in appendix A.2. Similar to theorem 1, we see that the robust eigenvectors of each input correlation generated by the learning rule are stable equilibria of the learning dynamics. The simplicity of eq. 10 has kept us from determining their basins of attraction. We can, however, make several guarantees. First, in a large region, the unit sphere is an attracting set for the dynamics of eq. 10. Second, the only stable fixed points are either the eigenvectors of \( \pm \mu \) or combinations of the eigenvectors of \( \mu \) with only nonpositive weights. This is in contrast to the situation where the learning rule has only one term; then theorem 1 guarantees that the only attractors are eigenvectors.

**3 Discussion**

We have analyzed biologically motivated plasticity dynamics that generalize the Oja rule. One class of these compute tensor eigenvectors. We proved that without a multiplicative weight-dependence in the plasticity, those eigenvectors are attractors of the dynamics ( theorem 1, figs. 1a, b). Contrary to Oja’s rule, the first eigenvector of higher-order input correlations is not a unique attractor. Rather, each eigenvector \( k \) has a finite basin of attraction, the size of which is proportional to \( \lambda_k^{n/k} \). If there are \( d \) codominant eigenvectors (\( \lambda_1 = \lambda_2 = \ldots = \lambda_d \)), the \( d \)-sphere they span is an attracting equilibrium manifold (corollary 1 in appendix A.2). Furthermore, steady states of any plasticity model with a finite Taylor polynomial in the neural output and inputs are generalized eigenvectors of multiple input correlations. These steady states are stable and attracting ( theorem 1, figs. 2, 3).

**3.1 Spiking neurons and weight-dependence**

While biological synaptic plasticity is certainly more complex than the simple generalized Hebbian rule of eq. 1, neural activity is also more complex than the linear model \( n = J^T x \). We examined the
simple linear-nonlinear-Poisson spiking model and a generalized spike timing–dependent plasticity (STDP) rule \([45];\) appendix A.3). Similar to eqs. 2 and 10 we can write the dynamical equation for \(J\) as a function of joint cumulant tensors of the input (eq. 56 in appendix A.3). These dynamics have a different structure than eqs. 2 and 10.

We focused here mainly on learning rules with no direct dependence on the synaptic weight \((c = 0\) in eq. \(1\); \(c = 0\) in eq. \(9\)). When \(c \neq 0\), the learning dynamics cannot be simply analyzed in terms of the loading onto the input correlations’ eigenvectors. We studied the learning dynamics with weight-dependence for two simple families of input correlations: diagonal \(\mu\) and piecewise-constant rank one \(\mu\) (appendix A.4). In both cases, we found that with eigenvectors of those simple input correlations were also attractors of the plasticity rule. With diagonal input correlations, sparse steady states with one nonzero synapse are always stable and attracting when \(a + c > 0\), but if \(a + c \leq 0\) synaptic weights converge to solutions where all weights have the same magnitude (fig. A.4.1). With rank one input correlations, multiplicative weight-dependence can interfere with synaptic scaling and lead to an instability for the neurons’ total synaptic amplitude (fig. A.4.2).

3.2 Related work and applications

There is a rich literature on generalized or nonlinear forms of Hebbian learning. We briefly discuss the most closely related results, to our knowledge. The family of Bienenstock, Cooper, and Munro (BCM) learning rules supplement the classic Hebbian model with a stabilizing sliding threshold for potentiation rather than synaptic scaling \([59]\). BCM rules balance terms driven by third- and fourth-order joint moments of the pre- and postsynaptic activity \([60]\). A triplet STDP model with rate-dependent depression and uncorrelated Poisson spiking has BCM dynamics \([45]\) and can develop selective (sparse) connectivity in response to rate- or correlation-based input patterns \([61]\). If the input is drawn from a mixture model then under a BCM rule, the synaptic weights are guaranteed to converge to the class means of the mixture \([62]\).

Learning rules with suitable postsynaptic nonlinearities can allow a neuron to perform independent component analysis (ICA) \([63], [64]\). These learning rules optimize the kurtosis of the neural response. In contrast, we show that a simple nonlinear Hebbian model learns tensor eigenvectors of higher-order input correlations. Those higher-order input correlations can determine which features are learned by gradient-based ICA algorithms \([65]\). Taylor and Coombes showed that a generalization of the Oja rule to higher-order neurons can also learn higher-order correlations \([66]\), which can allow learning independent components \([67]\). In that model, however, the synaptic weights \(J\) are a higher-order tensor.

Computing the robust eigenvectors of an odeco tensor \(\mu\) by power iteration has \(O(K^{a+1})\) space complexity: it requires first computing \(\mu\). The discrete-time dynamics of eq. 1 correspond to streaming power iteration, with \(O(K)\) space complexity \([68], [70]\). Eq. 2 defines limiting continuous-time dynamics for tensor power iteration, exposing the basins of attraction.

Oja’s rule inspired a generation of neural algorithms for PCA and subspace learning \([12], [13]\). Local learning rules for approximating higher-order correlation tensors may also prove useful, for example in neuromorphic devices \([71], [73]\).

Code availability

The code associated with figures one and three is available at https://github.com/gocker/TensorHebb.

Acknowledgments

We thank the Allen Institute founder, Paul G. Allen, for his vision, encouragement and support.
References


