Tensor decomposition of higher-order correlations by nonlinear Hebbian plasticity

Anonymous Author(s)
Affiliation
Address
email

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. We show that 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 tensor eigenvectors of higher-order input correlations. We prove that each tensor eigenvector is an attractor 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 tensor eigenvectors of its higher-order input correlations. Nonlinearities in synaptic plasticity thus allow a neuron to encode higher-order input correlations in a simple fashion.

1 Introduction

The strength of synaptic connectivity depends on pre- and postsynaptic activity. In Hebbian learning rules, potentiation of the net synaptic weight between two neurons is driven by the correlation between pre- and postsynaptic activity [1]. 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 [4]. That runaway potentiation can be stabilized by supplemental homeostatic plasticity dynamics [5], by weight dependence in the learning rule [6, 7], or by synaptic scaling regulating a neuron's total synaptic weight [8, 9]. In 1982, Erkki Oja observed that a linear neuron with Hebbian plasticity and synaptic scaling learns the first principal component of its inputs [10]. This led to a fountain of research on unsupervised learning in neural networks [11, 12].

Despite its success, the basic Hebbian postulate does not take into account fundamental nonlinear aspects of biological synaptic plasticity. First, synaptic plasticity depends on beyond-pairwise activity correlations [13–18]. Second, spatially clustered and temporally coactive synapses exhibit correlated and cooperative plasticity [19–26]. Here, we study the dynamics of a simple generalized Hebbian learning rule motivated by these observations, combined with synaptic scaling (eq. 1). We show that its equilibrium synaptic weights correspond to low-rank tensor decompositions for higher-order input correlations. The order of input correlation (two-point, 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 tensor eigenvectors [27–29] of higher-order input correlations. We

Submitted to 35th Conference on Neural Information Processing Systems (NeurIPS 2021). Do not distribute.

prove that the higher-order input correlation tensor's 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 eigenvector 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 they are contained 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, and the relation of this work to prior work.

2 Results

We take a neuron receiving K time-varying inputs $\mathrm{x}_i(t)$, $i \in [K]$, 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 $\mathrm{x}_i(t)$, and the current synaptic weight $J_i(t)$. We model these dependencies in a learning rule f:

$$J_i(t+dt) = \frac{J_i(t) + f\left(n(t), \mathrm{x}_i(t), J_i(t)\right)}{\|\boldsymbol{J}(t+dt)\|}, \text{ where } f = n^a(t)\mathrm{x}_i^b(t)J_i^c(t). \tag{1}$$

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. 1 assumes a simple form for these nonlinearities; we discuss arbitrary nonlinear learning rules in section 2.3. We assume that a and b are positive integers, as in higher-order voltage or spike timing-dependent plasticity models [30–32]. The normalization by the norm of the synaptic weight vector, $\|\boldsymbol{J}\|$, models homeostatic synaptic scaling [8–10]. Combined with a linear neuron and a slow learning rate, this implies the following dynamics for $\boldsymbol{J}$ (appendix A.1):

$$\tau \dot{J}_i = J_i^c \sum_{\alpha} \mu_{i,\alpha} (\boldsymbol{J}^{\otimes a})_\alpha - J_i \sum_{j,\alpha} J_j^{c+1} \mu_{j,\alpha} (\boldsymbol{J}^{\otimes a})_\alpha. \tag{2}$$

In eq. 2, $\dot{J}_i = dJ_i/dt$, $\tau \gg 1$ is the learning timescale, $\otimes$ is the vector outer product, and

$$\boldsymbol{\mu}_{i,\alpha} = \langle x_i^b (\mathbf{x}^{\otimes a})_\alpha \rangle_{\mathbf{x}} \tag{3}$$

where $\langle\rangle_{\mathbf{x}}$ denotes the expectation with respect to the distribution of the inputs. $\boldsymbol{\mu}$ is an $(a+1)$-order tensor containing an $(a+b)$-order joint moment of $\mathbf{x}$. 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. Each mode of $\boldsymbol{\mu}$ has the same dimension K. $\alpha = (j_1, \ldots, j_a)$ is a multi-index into $\boldsymbol{\mu}$. $(\boldsymbol{J}^{\otimes a})$ is the a-fold outer product of the synaptic weight vector $\boldsymbol{J}$ with itself. In the first term of eq. 2, for example, $\sum_\alpha \mu_{i,\alpha} (\boldsymbol{J}^{\otimes a})_\alpha$ takes the dot product of $\boldsymbol{J}$ along modes 2 through $a+1$ of $\boldsymbol{\mu}$.

2.1 Steady states of nonlinear Hebbian learning

If we take $a = b = 1$, then $\boldsymbol{\mu}$ is the second-order correlation of $\mathbf{x}$ and α is just the index j. With $c = 0$ also, eq. 2 reduces to Oja's rule and $\boldsymbol{J}$ is guaranteed to converge to the dominant eigenvector of $\boldsymbol{\mu}$ [10, 33]. We next investigate the steady states of eq. 2 for arbitrary $(a, b) \in \mathbb{Z}_+^2$, $c \in \mathbb{R}$. Note that $J_i = 0$ is a trivial steady state. At steady states of eq. 2 where $J_i \neq 0$,

$$\sum_{\alpha} \mu_{i,\alpha} (\boldsymbol{J}^{\otimes a})_\alpha = \lambda J_i^{1-c}, \text{ where } \lambda(\boldsymbol{\mu}, \boldsymbol{J}) = \sum_{j,\alpha} J_j^{c+1} \mu_{j,\alpha} (\boldsymbol{J}^{\otimes a})_\alpha \tag{4}$$

so that $\boldsymbol{J}$ is invariant under the multilinear map of $\boldsymbol{\mu}$ except for a scaling by λ and element-wise exponentiation by $1-c$. For two parameter families (a, b, c), eq. 4 reduces to different types of tensor eigenequation [28, 29]. We next briefly describe these and some of their properties.

First, if $a + c = 1$, we have the tensor eigenvalue/vector equation $\sum_\alpha \mu_{i,\alpha}(\boldsymbol{J}^{\otimes a})_\alpha = \lambda J_i^a$. Qi calls $\lambda, \boldsymbol{J}$ simply the tensor eigenpair [28] and Lim calls them the ℓ^a-norm eigenpair [29]. For any a, there are Ka^{K-1} such eigenvalues [28]. If $\boldsymbol{\mu} \geq 0$ element-wise, then it has a unique largest eigenvalue with a corresponding real, non-negative eigenvector $\boldsymbol{J}$ (analogous to the Perron-Frobenius theorem for matrices) [29, 34]. If $\boldsymbol{\mu}$ is weakly irreducible, that eigenvector is strictly positive [35]. In contrast to matrix eigenvectors, however, for $a > 1$ these tensor eigenvectors are not necessarily invariant under orthogonal transformations [28].

If $c = 0$ we have another variant of tensor eigenvalue/vector equation:

$$\sum_\alpha \mu_{i,\alpha}(\boldsymbol{J}^{\otimes a})_\alpha = \lambda J_i \tag{5}$$

Qi calls these $\lambda, \boldsymbol{J}$ an E-eigenpair [28] and Lim calls them the ℓ^2-eigenpair [29]. In general, a tensor may have infinitely many such eigenpairs. If the spectrum of a K-dimensional tensor of order $a + 1$ is finite, however, there are ${}^{(a^K-1)}/_{(a-1)}$ eigenvalues counted with multiplicity, and the spectrum of a weakly symmetric tensor is guaranteed to be finite [36, 37]. Unlike the steady states when $a + c = 1$, the E-eigenpairs are invariant under orthogonal transformations [28]. For non-negative $\boldsymbol{\mu}$, there exists a positive E-eigenpair [38]. It may not be unique, however, unlike the largest eigenpair for $a + c = 1$ (an anti-Perron-Frobenius result) [37].

The first k E-eigenvectors of $\boldsymbol{\mu}$ define the rank k approximation of $\boldsymbol{\mu}$ with least squared error [27]. The E-eigenvectors thus correspond to the components of the Tucker decomposition of $\boldsymbol{\mu}$ [39]. The E-eigenvalues λ are the norm of sub-tensors of the core tensor in the Tucker decomposition [27]. With no weight-dependence ($c = 0$), components of the best low-rank approximation of the inputs' $(a + b)$-order correlation are thus steady states of the nonlinear Hebbian dynamics (eq. 2). In the remainder of the paper, we will usually focus on parameter sets with $c = 0$, and use "tensor eigenvector" to refer to those of eq. 5.

2.2 Dynamics of nonlinear Hebbian learning

For the linear Hebbian rule, $(a, b, c) = (1, 1, 0)$, Oja and Karhunen proved that the first principal component of the inputs is a global attractor of eq. 2 [33]. We thus asked whether the first tensor eigenvector is a global attractor of eq. 2 when $(a, b) \neq (1, 1)$. We first simulated the nonlinear Hebbian dynamics. For the inputs $\mathbf{x}$, we used images from the Berkeley segmentation dataset [40] and sampled 35×35 pixel patches (fig. 1a). For $b \neq 1$, the correlation of these image patches was not symmetric (fig. 1b). The mean squared error of the orthogonal approximation of these tensors decreased with the approximation rank and was higher for $b = 1$ than $b = 2$ (fig. 1d). It decreased slowly past rank ~ 10, and the rank of the input correlation tensors was at least 30 (fig. 1d).

The nonlinear Hebbian learning dynamics converged to an equilibrium from random initial conditions (e.g., fig. 1e, f), around which the weights fluctuated due to the finite τ. Any equilibrium is guaranteed to be some eigenvector of the input correlation tensor $\boldsymbol{\mu}$ (section 2.1). We next checked which eigenvector the synaptic weights had converged to. For individual realizations of the weight dynamics, we computed the overlap between the final synaptic weight vector (fig. 1f) and each of the first 10 eigenvectors of the corresponding input correlation $\boldsymbol{\mu}$. The dynamics most frequently converged to the first eigenvector. For a non-negligible fraction of initial conditions, however, the nonlinear Hebbian rule converged to sub-dominant eigenvectors (fig. 1g,h). The input correlations $\boldsymbol{\mu}$ did have a unique dominant eigenvector (fig. 3a, blue), but the dynamics of eq. 2 did not always converge to it. This finding stands in contrast to the standard Hebbian rule, which must converge to the first eigenvector if it is unique [33].

To understand this further, we examined the learning dynamics analytically. Our main finding is that with $(b, c) = (1, 0)$ in the generalized Hebbian rule, each eigenvector of $\boldsymbol{\mu}$ is an attractor of eq. 2 with a non-zero basin of attraction. 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 $\boldsymbol{\mu}$ does, however, have the largest basin of attraction.

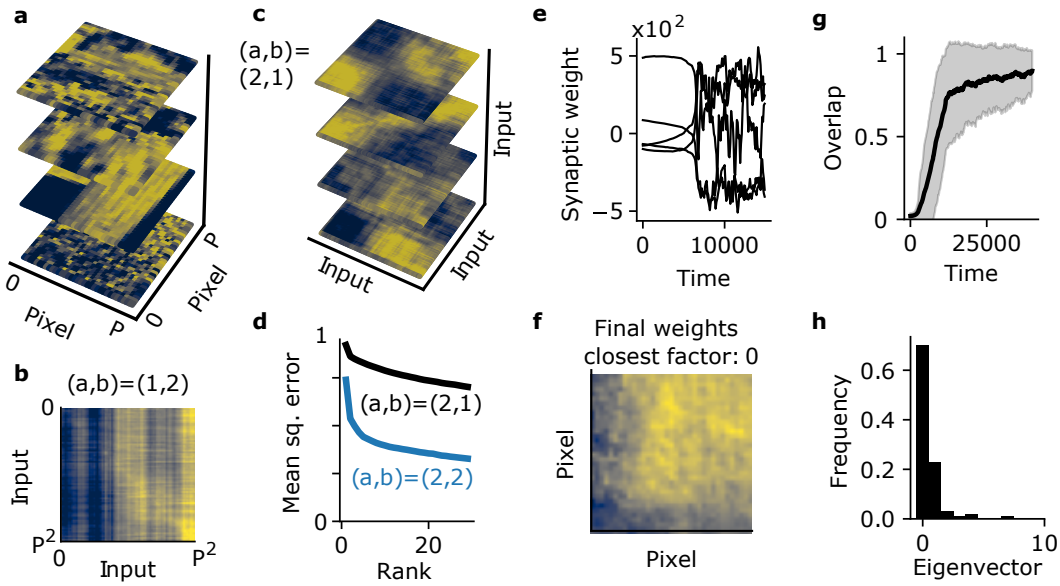

Figure 1: Convergence of higher-order Hebbian plasticity to tensor eigenvectors of natural image correlations. **a)** Example $P \times P$ image patches ($P = 35$ pixels). **b)** Third-order input correlation $\boldsymbol{\mu}$ that drives plasticity for $a = 1, b = 2$. **c)** Third-order input correlation $\boldsymbol{\mu}$ that drives plasticity for $a = 2, b = 1$. **d)** Mean squared error of the canonical polyadic (CP) approximation of the input correlations. (Curves: mean across 4 initializations for the alternating least-squares computation of the CP, with standard deviation within the line thickness.) **e)** Learning dynamics on natural inputs: five randomly selected synaptic weights with $(a, b, c) = (2, 1, 0)$. **f)** Example final synaptic weight configuration with $(a, b, c) = (2, 1, 0)$. **g)** Overlap of the synaptic weights with their final closest singular vector, $\boldsymbol{U}_i^T \boldsymbol{J}$. Solid line: mean over 10 samples of natural image patches and 10 realizations of the weight dynamics for each input patch. Shaded region: standard deviation. Initial conditions for $\boldsymbol{J}$ are chosen uniformly on the unit sphere. **h)** Histogram of the closest singular vector to the final synaptic weights for $(a, b, c) = (2, 1, 0)$.

Theorem 1. *In eq. 2, take $b = 1, c = 0$. Let $\boldsymbol{\mu}$ be a cubical, symmetric tensor of order $a + 1$ with R positive eigenvalues, each unique. That is, the eigen-decomposition of $\boldsymbol{\mu}$ is*

$$\boldsymbol{\mu} = \sum_{r=1}^{R} \lambda_r \left(\boldsymbol{U}_r^T\right)^{\otimes a+1} \tag{6}$$

where $\boldsymbol{U}$ is the matrix of unit-norm orthogonal E-eigenvectors: $\|\boldsymbol{U}_r\|_2 = 1$ for each $r \in [R]$ and $\boldsymbol{U}_i^T \boldsymbol{U}_j = 0$ for $i \neq j$. Let $\lambda_i > 0$ for each $i \in [R]$ and $\lambda_i \neq \lambda_j \; \forall \; (i, j) \in [R] \times [R]$. Then for each $k \in [R]$:

1. *With any odd $a > 1$, $\pm \boldsymbol{U}_k$ are attracting fixed points of eq. 2 and their basin of attraction is $\bigcap_{i \in [R] \setminus k} \left\{ \boldsymbol{J} : \left| \boldsymbol{U}_i^T \boldsymbol{J} / \boldsymbol{U}_k^T \boldsymbol{J} \right| < \left(\lambda_k / \lambda_i \right)^{1/(a-1)} \right\}$. Within that region, the separatrix of $+\boldsymbol{U}_k$ and $-\boldsymbol{U}_k$ is the hyperplane orthogonal to $\boldsymbol{U}_k$: $\{\boldsymbol{J} : \boldsymbol{U}_k^T \boldsymbol{J} = 0\}$.*
2. *With any even positive a, $\boldsymbol{U}_k$ is an attracting fixed point of eq. 2 and its basin of attraction is $\left\{ \boldsymbol{J} : \boldsymbol{U}_k^T \boldsymbol{J} > 0 \right\} \bigcap_{i \in [R] \setminus k} \left\{ \boldsymbol{J} : \boldsymbol{U}_i^T \boldsymbol{J} / \boldsymbol{U}_k^T \boldsymbol{J} < \left(\lambda_k / \lambda_i \right)^{1/(a-1)} \right\}$.*
3. *With any even positive a, $\boldsymbol{J} = \boldsymbol{0}$ is a neutrally stable fixed point of eq. 2 with basin of attraction $\left\{ \boldsymbol{J} : \sum_{j=1}^{R} (\boldsymbol{U}_j^T \boldsymbol{J})^2 < 1 \wedge \boldsymbol{U}_k^T \boldsymbol{J} < 0 \; \forall \; k \in [R] \right\}$.*

The proof is given in appendix A.2. To prove theorem 1, we project $\boldsymbol{J}$ onto the eigenvectors of $\boldsymbol{\mu}$, and study the dynamics of the loadings $\boldsymbol{v} = \boldsymbol{U}^T \boldsymbol{J}$. This leads to the discovery of a collection of unstable manifolds: each pair of axes (i, k) has an associated unstable hyperplane $v_i = v_k \left(\lambda_k / \lambda_i \right)^{1/(a-1)}$ (and

if a is odd, also the corresponding hyperplane with negative slope). These unstable hyperplanes partition the phase space into the basins of attraction of the eigenvectors of $\boldsymbol{\mu}$.

For example, consider a fourth-order input correlation (corresponding to $a = 3$ in eq. 1) of rank two, with two eigenvectors with eigenvalues $(\lambda_1, \lambda_2) = (3, 1)$. 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, and there are two unstable hyperplanes that partition the phase space into the basins of attraction of $(0, \pm 1)$ and $(\pm 1, 0)$.

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 solutions.

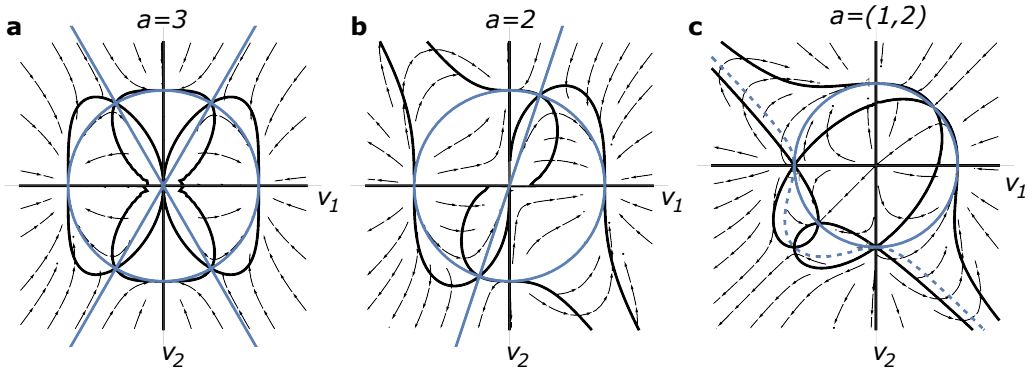

Figure 2: Example dynamics of the loadings, $\boldsymbol{v} = \boldsymbol{U}^T \boldsymbol{J}$, for rank-two input correlation. Black curves: nullclines of $v_1, v2$. Blue curves: stable sets and separatrices of the basins of attraction of the sparse attractors for $\boldsymbol{v}$. **a,b)** $(\lambda_1, \lambda_2) = (3, 1)$. **a)** Odd a ($a = 3$). The stable set is the unit sphere, and the separatrices are $v_2 = \pm v_1 \left(\lambda_1/\lambda_2\right)^{1/a-1}$. **b)** Even a ($a = 2$). The unstable set is $v_2 = v_1 \left(\lambda_1/\lambda_2\right)^{1/a-1}$ (solid blue line). **c)** Phase portrait for a two-term learning rule (eq. 9). All parameters of the input correlation tensors (λ_{mr}, A_i) are equal to one.

By partitioning the phase space of $\boldsymbol{J}$ into basins of attraction of the different eigenvectors of $\boldsymbol{\mu}$, theorem 1 also allows us to determine the volume of their basins of attraction, the topic of the following two corollaries. Since their basins of attraction are open sections of $\mathbb{R}^K$, 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 $\boldsymbol{J}^* = \boldsymbol{U}_k^T$. For odd $a > 1$,*

$$V_k = R^{-1} \prod_{i=1}^{R} \left(\frac{\lambda_k}{\lambda_i}\right)^{1/(a-1)} \tag{7}$$

Corollary 1.2. *Let V_k be the relative volume of the basin of attraction for $\boldsymbol{J}^* = \boldsymbol{U}_k^T$. For even positive a,*

$$\begin{aligned} V_k =& 2^{1-R} \Bigg(R^{-1} \prod_i \left(\frac{\lambda_k}{\lambda_i}\right)^{1/(a-1)} + (R-1)^{-1} \sum_{j \neq k} \prod_{i \neq j} \left(\frac{\lambda_k}{\lambda_i}\right)^{1/(a-1)} \\ &+ (R-2)^{-1} \sum_{j,l \neq k} \prod_{i \neq j,l} \left(\frac{\lambda_k}{\lambda_i}\right)^{1/(a-1)} + \ldots + 1 \Bigg) \end{aligned} \tag{8}$$

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 $\boldsymbol{\mu}$, its rank R, and its order a. The result for odd a also provides a lower bound on the volume for even a. Since the result is considerably simpler for odd a, we focus our discussion here on that case.

While eigenvectors of $\boldsymbol{\mu}$ with small eigenvalues contribute little to values of the input correlation $\boldsymbol{\mu}$, they can have a large impact on the volumes of the basins of attraction. The volume of the basin of attraction of eigenvector k is proportional to $\lambda_k^{R/(a-1)}$. An eigenvector with eigenvalue ϵ 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 $V_k/V_j = \left(\lambda_k/\lambda_j\right)^{R/a-1}$.

We see in theorem 1 that the fixed points of eq. 2 are points on the unit hypersphere, $\mathcal{S}$. With odd a, $\mathcal{S}$ is an attracting set for eq. 2; for even a, the section of $\mathcal{S}$ 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 surface area of the section of $\mathcal{S}$ in the basin of attraction for eigenvector k, A_k (corollary 1.3 in appendix A.2). The result requires knowledge of all non-negligible eigenvalues of $\boldsymbol{\mu}$, and the ratio A_k/A_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 $\boldsymbol{J}$ chosen uniformly at random on $\mathcal{S}$, the basin of attraction for $\boldsymbol{U}_1^T$ was $\sim 3\times$ larger than that for $\boldsymbol{U}_2^T$ and the higher eigenvectors had negligible basins of attraction (Fig. 1h).

In Oja's classic 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 co-dominant 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 1.4 in A.2):

Corollary 1.3. *(Informal) If any d eigenvalues of $\boldsymbol{\mu}$ are equal, the d-sphere spanned by their eigenvectors is an attracting equilibrium manifold and its basin of attraction is defined by each of those d eigenvectors' basins of attraction boundaries with the other $R - d$ eigenvectors.*

2.3 Arbitrary learning rules

So far we have studied phenomenological plasticity rules of a particular form (eq. 1). The neural output n, input x_i, and synaptic weight J_i were each raised to some power and then multiplied together. Changes in the strength of actual synapses are governed by complex biochemical, transcriptional and regulatory pathways [41]. We view these as specifying some unknown function of the neural output, input, and synaptic weight, $f(n, \mathrm{x}_i, J_i)$. That function might not have the form of eq. 1. So, we next investigate the dynamics of arbitrary learning rules f. We see here that under a mild condition, any steady state of the plasticity induced by f will have a similar form as the steady states of eq. 2. If f does not depend on $\boldsymbol{J}$ except through n, steady states will be a kind of generalized eigenvector of higher-order input correlations.

The Taylor expansion of f around zero is:

$$f\left(n(t), \mathrm{x}_i(t), J_i(t)\right) = \sum_{m=1}^{\infty} A_m n^{a_m}(t) \mathrm{x}_i^{b_m}(t) J_i^{c_m}(t) \tag{9}$$

where the coefficients A_m are partial derivates of f with respect to its arguments. We assume that there exists a finite integer N such that f's derivatives of order $N+1, \ldots, \infty$ 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

$$\tau \dot{J}_i = \sum_m J_i^{c_m} \sum_{\alpha_m} {}_m\boldsymbol{\mu}_{i,\alpha_m} (\boldsymbol{J}^{\otimes a_m})_{\alpha_m} - J_i \sum_{j,m} \sum_{\alpha_m} J_j^{c_m+1} {}_m\boldsymbol{\mu}_{j,\alpha_m} (\boldsymbol{J}^{\otimes a_m})_{\alpha_m} \tag{10}$$

where ${}_m\boldsymbol{\mu}_{i,\alpha_m} = A_m \langle \mathrm{x}_i^{b_m} (\mathbf{x}^{\otimes a_m})_{\alpha_m} \rangle_{\mathbf{x}}$. At steady states where $J_i \neq 0$,

$$\sum_m J_i^{c_m} \sum_{\alpha_m} {}_m\mu_{i,\alpha_m} (\boldsymbol{J}^{\otimes a_m})_{\alpha_m} = \lambda J_i, \text{ where } \lambda(\boldsymbol{J}, \{{}_m\boldsymbol{\mu}\}) = \sum_m \sum_{j,\alpha_m} J_j^{c_m+1} {}_m\mu_{j,\alpha} (\boldsymbol{J}^{\otimes a_m})_{\alpha}. \tag{11}$$

If each $c_m = 0$, this is a kind of generalized tensor E-eigenvector equation:

$$\sum_{m,\alpha_m} {}_m\boldsymbol{\mu}_{i,\alpha} (\boldsymbol{J}^{\otimes a_m})_{\alpha} = \lambda J_i \tag{12}$$

so that $\boldsymbol{J}$ is invariant under the *combined* action of the multilinear maps ${}_m\boldsymbol{\mu}$ (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 ${}_m\boldsymbol{\mu}$. 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 $\boldsymbol{J}^*$ that is an eigenvector of each of those input correlation tensors, $\sum_\alpha \mu^{(m)}_{i,\alpha} (\boldsymbol{J}^{*\otimes a_m})_\alpha = \lambda J^*_i$ for each m, then that configuration $\boldsymbol{J}^*$ 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 ($\boldsymbol{a} = (1, 2)$, $\boldsymbol{b} = \boldsymbol{1}$, $\boldsymbol{c} = \boldsymbol{0}$, $\boldsymbol{A} = (1, 1/2)$ in eq. 9). As before, the inputs were whitened natural image patches (fig. 1a). The two-and three-point correlations of those image patches have similar first eigenvalues, but the spectrum of the two-point correlation decreases more quickly than the three-point correlation (fig. 3a). The first three eigenvectors of the different correlations overlap strongly, though not perfectly, with each other (fig. 3b). The inputs were mean-subtracted so this was not due to a trivial constant offset. With this parameter set, the synaptic weights usually converged to the (shared) first eigenvector of the input correlations (fig. 3c,d, e (blue)).

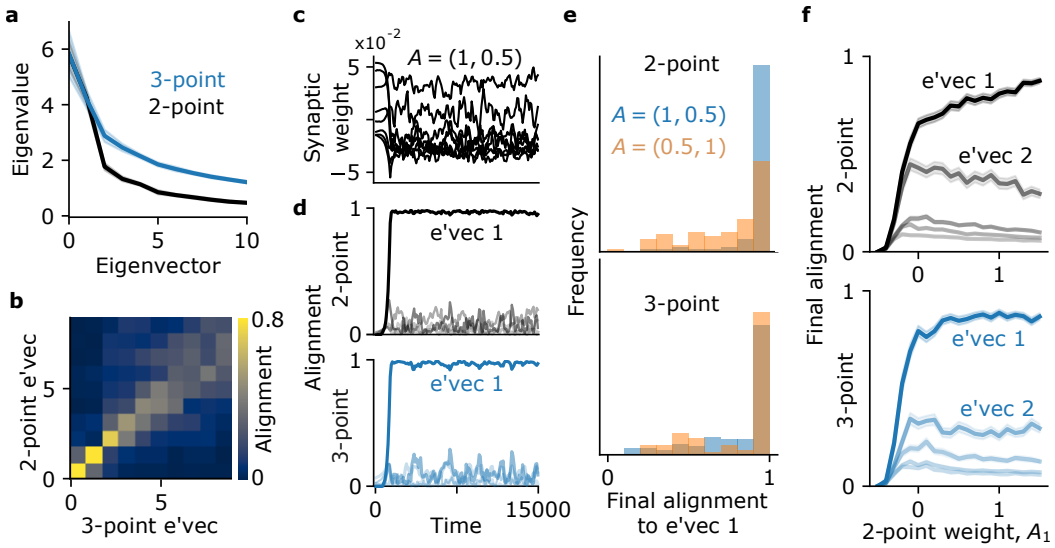

Figure 3: Multi-term Hebbian plasticity rule converges to generalized input correlation tensors' eigenvectors. We study a learning rule with two terms, driven by the second and third-order input correlations ($\boldsymbol{a} = (1, 2)$, $\boldsymbol{b} = \boldsymbol{1}$, $\boldsymbol{c} = \boldsymbol{0}$). **a)** 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 weighed the inputs' three-point correlation more strongly than the two-point correlation ($\boldsymbol{A} = (1/2, 1)$), the dynamics converged almost always to the first eigenvector of the three-point correlation (fig. 3e, 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. 3f). 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. 3f).

Earlier, we saw that in single-term learning rules, the only attractors were the eigenvectors of the input correlation (theorem 1). The dynamics of eq. 2 usually converged to the first eigenvector because it had the largest basin of attraction (corollaries 1.1, 1.2). Here, we saw that at least for some parameter sets, the dynamics of a multi-term generalized Hebbian rule may not converge to an input eigenvector. This suggested the existence of other attractors for the dynamics of eq. 10.

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 (fig. 2c) that there can be equilibria that are linear combinations of those eigenvectors with all negative weights (in fig. 2c, 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 $\boldsymbol{b} = 1, \boldsymbol{c} = 0, \boldsymbol{a} \in \mathbb{Z}_+^N$, and consider N cubical, symmetric tensors, ${}_m\boldsymbol{\mu}$, each of order $a_m + 1$ for $m \in [N]$, that are mutually orthogonally decomposable into R components:*

$$ {}_m\boldsymbol{\mu} = \sum_{r=1}^{R} \lambda_{mr} \boldsymbol{U}_r^T \otimes \boldsymbol{U}_r^T \otimes \cdots \otimes \boldsymbol{U}_r^T \tag{13} $$

with $\|\boldsymbol{U}_r\|_2 = 1$ for each $r \in [R]$ and $\boldsymbol{U}_i^T \boldsymbol{U}_j = 0$ for $i \neq j$. Let $\lambda_{mr} \geq 0$ and $\sum_m \lambda_{mr} > 0$ for each $m, r \in [N] \times [R]$. Let

$$ S(\boldsymbol{J}) = \sum_{i=1}^{R} (\boldsymbol{U}_i^T \boldsymbol{J})^2, \; L(\boldsymbol{J}) = \sum_{m=1}^{N} \sum_{i=1}^{R_m} \lambda_{mi} (\boldsymbol{U}_i^T \boldsymbol{J})^{a_m+1} \tag{14} $$

Then:

1. *$\mathcal{S}^* = \{\boldsymbol{J} : S(\boldsymbol{J}) = 1 \wedge L(\boldsymbol{J}) > 0\}$ is an attracting set for eq. 10 and its basin of attraction includes $\{\boldsymbol{J} : L(\boldsymbol{J}) > 0\}$.*
2. *For each $k \in [R]$, $\boldsymbol{J} = \boldsymbol{U}_k^T$ is a stable equilibrium of eq. 10.*
3. *For each $k \in [R]$, $\boldsymbol{J} = -\boldsymbol{U}_k^T$ is a stable equilibrium of eq. 10 if $\sum_m {}_m\lambda_k (-1)^{a_m} < 0$ (and unstable if $\sum_m {}_m\lambda_k (-1)^{a_m} < 0$)*
4. *Any other stable equilibrium must have $\boldsymbol{U}_k^T \boldsymbol{J} \leq 0$ for each $k \in [R]$.*

The claims of theorem 2 are proven in appendix A.2. Similar to theorem 1, we see that the eigenvectors of each input correlation generated by the learning rule are stable equilibria of the learning dynamics. The complexity 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 $\pm\boldsymbol{\mu}$, or they are combinations of the eigenvectors of $\boldsymbol{\mu}$ with only non-positive 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 the eigenvectors.

3 Discussion

We have analyzed biologically observed plasticity dynamics that generalize the Oja rule. A broad class of these compute the tensor eigenvectors that correspond to multilinear singular vectors [27]. We proved that without a multiplicative weight-dependence in the plasticity, those eigenvectors are attractors of the dynamics (theorem 1, figs. 1, 2a, 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 $\lambda_k^{R/a-1}$. If there are d co-dominant eigenvectors ($\lambda_1 = \lambda_2 = \ldots = \lambda_d$), the d-sphere they span is an attracting equilibrium manifold (corollary 1.4 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, although there can also exist other attractors (theorem 2, figs. 2c, 3). We conclude by discussing extensions of these results to spiking neurons and weight-dependent learning rules ($c \neq 0$), the relation of our findings to prior work, and possible applications.

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 certainly more complex than the linear model $n = \boldsymbol{J}^T \mathbf{x}$ we have used so far. We examined the simple linear-nonlinear-Poisson spiking model and a generalized spike timing-dependent plasticity (STDP) rule (appendix A.3). Similar to eqs. 2 and 10, we can write the dynamical equation for $\boldsymbol{J}$ 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, however, so that the results of this study do not directly apply to that model.

We focused here mainly on learning rules with no direct dependence on the synaptic weight ($c = 0$ in eq. 1, $\boldsymbol{c} = 0$ 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 $\boldsymbol{\mu}$ and piecewise-constant rank one $\boldsymbol{\mu}$ (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 non-zero 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. Learning rules of the form $f(n, x_i, J_i) = x_i g(n)$, with suitable postsynaptic nonlinearities $g(n)$, allow a neuron to learn to estimate independent components of the input [42, 43]. These learning rules typically minimize or maximize the kurtosis of the neural response n. In contrast, we show that a simple nonlinear Hebbian model directly finds tensor eigenvectors of a higher-order correlation of the input $\mathbf{x}$.

The family of Bienenstock, Cooper & Munro (BCM) learning rules supplement the classic Hebbian model ($a = b = 1, c = 0$ in eq. 1) with a stabilizing sliding threshold for potentiation rather than synaptic scaling [44]. BCM rules balance terms driven by third and fourth-order joint moments of the input $\mathbf{x}$ and output n [45]. If $\mathbf{x}$ is drawn from a mixture model, then under a BCM rule, $\boldsymbol{J}$ will converge to the class means of the mixture [46].

From a computational perspective, nonlinear Hebbian learning has attractive properties for tensor decompositions compared to the classic method of tensor power iteration [27], which requires the full tensor $\boldsymbol{\mu}$ to operate on, with $\mathcal{O}(K^{a+1})$ space complexity. Streaming computations that operate on one input sample at a time, such as those of eq. 2, have $\mathcal{O}(K)$ space complexity. The discrete-time dynamics of eq. 1 correspond to streaming tensor power iteration for computing tensor singular vectors [47, 48]. With those parameters, the dynamics of eq. 2 can thus be viewed as the continuous-time limiting dynamics of tensor power iteration. Convergence proofs for the discrete-time dynamics require a learning rate that decreases over time. Biological mechanisms for such a scheme are unclear; this is not necessary in the continuous limiting dynamics. Our study of the limiting differential equations governing the plasticity also expose the basins of attraction of each input correlation's eigenvectors and allows us to study arbitrary learning rules.

Oja's rule inspired a generation of neural algorithms for PCA and subspace learning, with a wealth of applications [11, 12]. Local learning rules for approximating higher-order correlation tensors may also prove useful, for example in neuromorphic devices [49–51]. Incorporating these learning rules into any larger scale effort using larger, curated data sets inherits the promise and peril implicit in such applications.

References

[1] Hebb DO. The organization of behavior: a neuropsychological theory. Mahwah, N.J.: L. Erlbaum Associates; 1949.

[2] Engel A, Van den Broeck C. Statistical Mechanics of Learning. Cambridge: Cambridge University Press; 2001. Available from: https://www.cambridge.org/core/books/statistical-mechanics-of-learning/D10C20B9997048D27EC08348EE851922.

[3] Caporale N, Dan Y. Spike Timing–Dependent Plasticity: A Hebbian Learning Rule. Annual Review of Neuroscience. 2008;31(1):25–46. doi:10.1146/annurev.neuro.31.060407.125639.

[4] Turrigiano Gina G . The dialectic of Hebb and homeostasis. Philosophical Transactions of the Royal Society B: Biological Sciences. 2017;372(1715):20160258. doi:10.1098/rstb.2016.0258.

[5] Zenke F, Gerstner W. Hebbian plasticity requires compensatory processes on multiple timescales. Phil Trans R Soc B. 2017;372(1715):20160259. doi:10.1098/rstb.2016.0259.

[6] Van Rossum MC, Bi GQ, Turrigiano GG. Stable Hebbian learning from spike timing-dependent plasticity. The Journal of Neuroscience. 2000;20(23):8812–8821.

[7] Rubin J, Lee D, Sompolinsky H. Equilibrium Properties of Temporally Asymmetric Hebbian Plasticity. Physical Review Letters. 2001;86(2):364–367. doi:10.1103/PhysRevLett.86.364.

[8] Bourne JN, Harris KM. Coordination of size and number of excitatory and inhibitory synapses results in a balanced structural plasticity along mature hippocampal CA1 dendrites during LTP. Hippocampus. 2011;21(4):354–373. doi:https://doi.org/10.1002/hipo.20768.

[9] Turrigiano G. Homeostatic Synaptic Plasticity: Local and Global Mechanisms for Stabilizing Neuronal Function. Cold Spring Harbor Perspectives in Biology. 2012;4(1):a005736. doi:10.1101/cshperspect.a005736.

[10] Oja E. Simplified neuron model as a principal component analyzer. Journal of Mathematical Biology. 1982;15(3):267–273. doi:10.1007/BF00275687.

[11] Becker S, Plumbley M. Unsupervised neural network learning procedures for feature extraction and classification. Applied Intelligence. 1996;6(3):185–203. doi:10.1007/BF00126625.

[12] Diamantaras KI, Kung SY. Principal component neural networks: theory and applications. USA: John Wiley & Sons, Inc.; 1996.

[13] Sjöström PJ, Turrigiano GG, Nelson SB. Rate, timing, and cooperativity jointly determine cortical synaptic plasticity. Neuron. 2001;32(6):1149–1164.

[14] Bi GQ, Wang HX. Temporal asymmetry in spike timing-dependent synaptic plasticity. Physiology & Behavior. 2002;77(4-5):551–555.

[15] Froemke RC, Dan Y. Spike-timing-dependent synaptic modification induced by natural spike trains. Nature. 2002;416(6879):433–438. doi:10.1038/416433a.

[16] Wang HX, Gerkin RC, Nauen DW, Bi GQ. Coactivation and timing-dependent integration of synaptic potentiation and depression. Nature Neuroscience. 2005;8(2):187–193. doi:10.1038/nn1387.

[17] Froemke RC, Poo Mm, Dan Y. Spike-timing-dependent synaptic plasticity depends on dendritic location. Nature. 2005;434(7030):221–225. doi:10.1038/nature03366.

[18] Froemke RC, Tsay IA, Raad M, Long JD, Dan Y. Contribution of individual spikes in burst-induced long-term synaptic modification. Journal of Neurophysiology. 2006;95(3):1620–1629. doi:10.1152/jn.00910.2005.

[19] Harvey CD, Svoboda K. Locally dynamic synaptic learning rules in pyramidal neuron dendrites. Nature. 2007;450(7173):1195–1200. doi:10.1038/nature06416.

[20] Harvey CD, Yasuda R, Zhong H, Svoboda K. The Spread of Ras Activity Triggered by Activation of a Single Dendritic Spine. Science. 2008;321(5885):136–140. doi:10.1126/science.1159675.

[21] Govindarajan A, Israely I, Huang SY, Tonegawa S. The dendritic branch is the preferred integrative unit for protein synthesis-dependent LTP. Neuron. 2011;69(1):132–146. doi:10.1016/j.neuron.2010.12.008.

[22] Makino H, Malinow R. Compartmentalized versus Global Synaptic Plasticity on Dendrites Controlled by Experience. Neuron. 2011;72(6):1001–1011. doi:10.1016/j.neuron.2011.09.036.

[23] Kleindienst T, Winnubst J, Roth-Alpermann C, Bonhoeffer T, Lohmann C. Activity-Dependent Clustering of Functional Synaptic Inputs on Developing Hippocampal Dendrites. Neuron. 2011;72(6):1012–1024. doi:10.1016/j.neuron.2011.10.015.

[24] Chen J, Villa K, Cha J, So PC, Kubota Y, Nedivi E. Clustered Dynamics of Inhibitory Synapses and Dendritic Spines in the Adult Neocortex. Neuron. 2012;74(2):361–373. doi:10.1016/j.neuron.2012.02.030.

[25] Takahashi N, Kitamura K, Matsuo N, Mayford M, Kano M, Matsuki N, et al. Locally Synchronized Synaptic Inputs. Science. 2012;335(6066):353–356. doi:10.1126/science.1210362.

[26] Lee KFH, Soares C, Thivierge JP, Béïque JC. Correlated Synaptic Inputs Drive Dendritic Calcium Amplification and Cooperative Plasticity during Clustered Synapse Development. Neuron. 2016;89(4):784–799. doi:10.1016/j.neuron.2016.01.012.

[27] De Lathauwer L, De Moor B, Vandewalle J. On the Best Rank-1 and Rank-(R1 ,R2 ,. . .,RN) Approximation of Higher-Order Tensors. SIAM Journal on Matrix Analysis and Applications. 2000;21(4):1324–1342. doi:10.1137/S0895479898346995.

[28] Qi L. Eigenvalues of a real supersymmetric tensor. Journal of Symbolic Computation. 2005;40(6):1302–1324. doi:10.1016/j.jsc.2005.05.007.

[29] Lim LH. Singular values and eigenvalues of tensors: a variational approach. In: 1st IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing, 2005.; 2005. p. 129–132.

[30] Gerstner W, Kistler WM. Mathematical formulations of Hebbian learning. Biological Cybernetics. 2002;87(5-6):404–415. doi:10.1007/s00422-002-0353-y.

[31] Pfister JP, Gerstner W. Triplets of Spikes in a Model of Spike Timing-Dependent Plasticity. The Journal of Neuroscience. 2006;26(38):9673–9682. doi:10.1523/JNEUROSCI.1425-06.2006.

[32] Clopath C, Gerstner W. Voltage and Spike Timing Interact in STDP – A Unified Model. Frontiers in Synaptic Neuroscience. 2010;2. doi:10.3389/fnsyn.2010.00025.

[33] Oja E, Karhunen J. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. Journal of Mathematical Analysis and Applications. 1985;106(1):69–84. doi:10.1016/0022-247X(85)90131-3.

[34] Yang Q, Yang Y. Further Results for Perron–Frobenius Theorem for Nonnegative Tensors II. SIAM Journal on Matrix Analysis and Applications. 2011;32(4):1236–1250. doi:10.1137/100813671.

[35] Friedland S, Gaubert S, Han L. Perron–Frobenius theorem for nonnegative multilinear forms and extensions. Linear Algebra and its Applications. 2013;438(2):738–749. doi:10.1016/j.laa.2011.02.042.

[36] Cartwright D, Sturmfels B. The number of eigenvalues of a tensor. Linear Algebra and its Applications. 2013;438(2):942–952. doi:10.1016/j.laa.2011.05.040.

[37] Chang KC, Pearson KJ, Zhang T. Some variational principles for Z-eigenvalues of nonnegative tensors. Linear Algebra and its Applications. 2013;438(11):4166–4182. doi:10.1016/j.laa.2013.02.013.

[38] Chang KC, Pearson K, Zhang T. Perron-Frobenius theorem for nonnegative tensors. Communications in Mathematical Sciences. 2008;6(2):507–520.

[39] Tucker LR. Some mathematical notes on three-mode factor analysis. Psychometrika. 1966;31(3):279–311. doi:10.1007/BF02289464.

[40] Martin D, Fowlkes C, Tal D, Malik J. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. vol. 2; 2001. p. 416–423 vol.2.

[41] Yap EL, Greenberg ME. Activity-Regulated Transcription: Bridging the Gap between Neural Activity and Behavior. Neuron. 2018;100(2):330–348. doi:10.1016/j.neuron.2018.10.013.

[42] Hyvärinen A, Oja E. Simple neuron models for independent component analysis. International Journal of Neural Systems. 1996;7(6):671–687. doi:10.1142/s0129065796000646.

[43] Bell AJ, Sejnowski TJ. An information-maximization approach to blind separation and blind deconvolution. Neural Computation. 1995;7(6):1129–1159. doi:10.1162/neco.1995.7.6.1129.

[44] Bienenstock EL, Cooper LN, Munro PW. Theory for the development of neuron selectivity: orientation specificity and binocular interaction in visual cortex. The Journal of Neuroscience. 1982;2(1):32–48.

[45] Intrator N, Cooper LN. Objective function formulation of the BCM theory of visual cortical plasticity: Statistical connections, stability conditions. Neural Networks. 1992;5(1):3–17. doi:10.1016/S0893-6080(05)80003-6.

[46] Lawlor M, Zucker SW. Feedforward Learning of Mixture Models. Advances in Neural Information Processing Systems. 2014;27.

[47] Wang Y, Anandkumar A. Online and differentially-private tensor decomposition. In: Proceedings of the 30th International Conference on Neural Information Processing Systems. NIPS'16. Red Hook, NY, USA: Curran Associates Inc.; 2016. p. 3539–3547.

[48] Wang PA, Lu CJ. Tensor Decomposition via Simultaneous Power Iteration. In: International Conference on Machine Learning. PMLR; 2017. p. 3665–3673. Available from: http://proceedings.mlr.press/v70/wang17i.html.

[49] Merolla PA, Arthur JV, Alvarez-Icaza R, Cassidy AS, Sawada J, Akopyan F, et al. A million spiking-neuron integrated circuit with a scalable communication network and interface. Science. 2014;345(6197):668–673. doi:10.1126/science.1254642.

[50] Davies M, Srinivasa N, Lin TH, Chinya G, Cao Y, Choday SH, et al. Loihi: A Neuromorphic Manycore Processor with On-Chip Learning. IEEE Micro. 2018;38(1):82–99. doi:10.1109/MM.2018.112130359.

[51] Zenke F, Neftci EO. Brain-Inspired Learning on Neuromorphic Substrates. Proceedings of the IEEE. 2021;109(5):935–950. doi:10.1109/JPROC.2020.3045625.

[52] Brito CSN, Gerstner W. Nonlinear Hebbian Learning as a Unifying Principle in Receptive Field Formation. PLOS Computational Biology. 2016;12(9):e1005070. doi:10.1371/journal.pcbi.1005070.

[53] Miller KD, Troyer TW. Neural noise can explaine expansive, power-law nonlinearities in neural response functions. J Neurophysiol. 2002;87(2):653–659.

[54] Priebe NJ, Ferster D. Inhibition, Spike Threshold, and Stimulus Selectivity in Primary Visual Cortex. Neuron. 2008;57(4):482–497. doi:10.1016/j.neuron.2008.02.005.

[55] Kempter R, Gerstner W, Van Hemmen JL. Hebbian learning and spiking neurons. Physical Review E. 1999;59(4):4498.

[56] Kolda TG, Bader BW. Tensor Decompositions and Applications. SIAM Review. 2009;51(3):455–500. doi:10.1137/07070111X.

[57] Gerstner W, Kempter R, van Hemmen JL, Wagner H. A neuronal learning rule for sub-millisecond temporal coding. Nature. 1996;383(6595):76–78. doi:10.1038/383076a0.

[58] Markram H. Regulation of Synaptic Efficacy by Coincidence of Postsynaptic APs and EPSPs. Science. 1997;275(5297):213–215. doi:10.1126/science.275.5297.213.

[59] Bi Gq, Poo Mm. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. The Journal of Neuroscience. 1998;18(24):10464–10472.

[60] Hansel D, Vreeswijk Cv. How Noise Contributes to Contrast Invariance of Orientation Tuning in Cat Visual Cortex. Journal of Neuroscience. 2002;22(12):5118–5128. doi:10.1523/JNEUROSCI.22-12-05118.2002.

[61] Priebe NJ, Mechler F, Carandini M, Ferster D. The contribution of spike threshold to the dichotomy of cortical simple and complex cells. Nature Neuroscience. 2004;7(10):1113–1122. doi:10.1038/nn1310.

[62] Priebe NJ, Ferster D. Mechanisms underlying cross-orientation suppression in cat visual cortex. Nature Neuroscience. 2006;9(4):552–561. doi:10.1038/nn1660.

[63] Gütig R, Aharonov R, Rotter S, Sompolinsky H. Learning input correlations through nonlinear temporally asymmetric Hebbian plasticity. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience. 2003;23(9):3697–3714.

Checklist

1. For all authors...
   (a) Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope? **[TODO]**
   (b) Did you describe the limitations of your work? [Yes]
   (c) Did you discuss any potential negative societal impacts of your work? [Yes]
   (d) Have you read the ethics review guidelines and ensured that your paper conforms to them? [Yes]
2. If you are including theoretical results...
   (a) Did you state the full set of assumptions of all theoretical results? [Yes] See appendix A.2.
   (b) Did you include complete proofs of all theoretical results? [Yes] See appendix A.2.
3. If you ran experiments...
   (a) Did you include the code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL)? [Yes] Code will be released on Github following de-anonymization.
   (b) Did you specify all the training details (e.g., data splits, hyperparameters, how they were chosen)? [Yes]
   (c) Did you report error bars (e.g., with respect to the random seed after running experiments multiple times)? [Yes]
   (d) Did you include the total amount of compute and the type of resources used (e.g., type of GPUs, internal cluster, or cloud provider)? [N/A] No HPC resources were used.
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets...
   (a) If your work uses existing assets, did you cite the creators? [Yes] We used images from the Berkeley Segmentation Database.
   (b) Did you mention the license of the assets? [No] We cited the paper describing the dataset.
   (c) Did you include any new assets either in the supplemental material or as a URL? [No]
   (d) Did you discuss whether and how consent was obtained from people whose data you're using/curating? [N/A]
   (e) Did you discuss whether the data you are using/curating contains personally identifiable information or offensive content? [N/A]
5. If you used crowdsourcing or conducted research with human subjects...
   (a) Did you include the full text of instructions given to participants and screenshots, if applicable? [N/A]
   (b) Did you describe any potential participant risks, with links to Institutional Review Board (IRB) approvals, if applicable? [N/A]
   (c) Did you include the estimated hourly wage paid to participants and the total amount spent on participant compensation? [N/A]