

Normative Framework for Deriving Neural Networks with Multicompartmental Neurons and Non-Hebbian Plasticity

David Lipshutz^{1,*}, Yanis Bahroun,^{1,2,*} Siavash Golkar^{1,*}, Anirvan M. Sengupta,^{2,3,4} and Dmitri B. Chklovskii^{1,5}

¹Center for Computational Neuroscience, Flatiron Institute, New York, New York 10010, USA

²Center for Computational Mathematics, Flatiron Institute, New York, New York 10010, USA

³Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, USA

⁴Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA

⁵Neuroscience Institute, NYU Langone Medical Center, New York, New York 10017, USA



(Received 22 March 2023; accepted 25 July 2023; published 10 August 2023)

An established normative approach for understanding the algorithmic basis of neural computation is to derive online algorithms from principled computational objectives and evaluate their compatibility with anatomical and physiological observations. Similarity matching objectives have served as successful starting points for deriving online algorithms that map onto neural networks (NNs) with point neurons and Hebbian/anti-Hebbian plasticity. These NN models account for many anatomical and physiological observations; however, the objectives have limited computational power, and the derived NNs do not explain multicompartmental neuronal structures and non-Hebbian forms of plasticity that are prevalent throughout the brain. In this article, we unify and generalize recent extensions of the similarity matching approach to address more complex objectives, including a large class of unsupervised and self-supervised learning tasks that can be formulated as symmetric generalized eigenvalue problems or non-negative matrix factorization problems. Interestingly, the online algorithms derived from these objectives naturally map onto NNs with multicompartmental neurons and local, non-Hebbian learning rules. Therefore, this unified extension of the similarity matching approach provides a normative framework that facilitates understanding multicompartmental neuronal structures and non-Hebbian plasticity found throughout the brain.

DOI: [10.1103/PRXLife.1.013008](https://doi.org/10.1103/PRXLife.1.013008)

I. INTRODUCTION

Advances in theoretical neuroscience are often driven by the development of normative frameworks that explain physiological and anatomical observations from the perspective of computational principles [1–11]. These top-down frameworks start with computational objectives from which physiological and anatomical implications are derived and compared with experimental observations. In the context of understanding the algorithmic basis of neural computation, this approach involves starting with a computational objective, deriving an online algorithm that can be implemented in a neural network (NN), and comparing the NN model with experimental observations.

In a pioneering example of this approach, Oja [4] proposed an online algorithm for principal component analysis (PCA) [12], a popular unsupervised dimensionality reduction method, which can be implemented in a point neuron (i.e., a neuron that only represents its scalar output) with Hebbian plasticity, Fig. 1 (left). Hebbian plasticity, named after

Hebb [13], refers to synaptic updates that are proportional to the product of the pre- and postsynaptic neural outputs. Experimental evidence of Hebbian plasticity came with the discovery of long-term potentiation [14,15], and since then a variety of forms of Hebbian plasticity have been observed [16]. Oja's model of a point neuron thus offers a link between experimentally observed Hebbian plasticity and an unsupervised learning objective. However, in the few decades following Oja's work, efforts to extend Oja's approach to extract multiple principal components resulted in NNs that used nonlocal learning rules [17–19].

Building on Oja's seminal work, Pehlevan and Chklovskii *et al.* recently developed a normative framework to extract multiple principal components using similarity matching objectives [10,20–23], which minimize the difference between the similarity of the NN inputs and that of the NN outputs. Starting from these objectives, they derived online algorithms that map onto multichannel NNs with point neurons and Hebbian plasticity. This normative framework proved useful for linking unsupervised learning objectives to Hebbian plasticity and several anatomical and physiological observations [24–29]. However, the similarity matching objectives have limited computational power, and the derived NNs cannot explain multicompartmental neuronal structures and other forms of synaptic plasticity prevalent throughout the brain [30].

Most neurons in the brain have multicompartmental structures and employ intricate forms of non-Hebbian plasticity. In particular, these neurons represent biophysical quantities

*These authors contributed equally to this work.

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

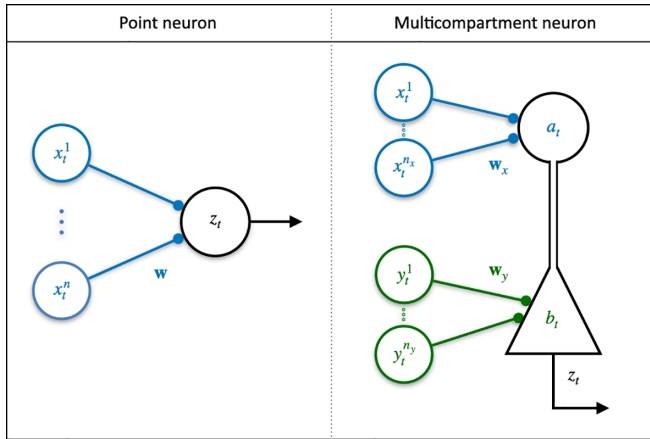


FIG. 1. Left: Oja’s point neuron with inputs \mathbf{x}_t , synaptic weights \mathbf{w} , output $z_t = \mathbf{w}\mathbf{x}_t$, and Hebbian plasticity. Right: Multicompartment model of a pyramidal neuron with a separate distal compartment (circular region) and a proximal compartment (triangular region). Feedback inputs \mathbf{x}_t target the distal portion of the apical tuft to generate the current $a_t = \mathbf{w}_x \mathbf{x}_t$, which drives non-Hebbian plasticity at the proximal synapses \mathbf{w}_y . Feed-forward inputs \mathbf{y}_t target the proximal region to generate the current $b_t = \mathbf{w}_y \mathbf{y}_t$. The output z_t is a function of the currents a_t and b_t , and, depending on the model, it is sometimes represented in a third compartment.

beyond their output, such as local dendritic currents, and these quantities constitute key variables in the neurons’ synaptic learning rules. For example, pyramidal neurons—the primary excitatory neurons of the cortex capable of performing complex computations [31]—receive inputs to their proximal and distal dendrites from distinct neural populations and integrate these inputs in separate compartments [32], Fig. 1 (right). Integrated distal inputs generate calcium plateau potentials that drive non-Hebbian plasticity in the proximal dendrites [33]. What are the computational objectives that lead to these more complex neuronal structures and intricate forms of non-Hebbian plasticity?

In a series of recent works [34–38], we have extended the similarity matching framework to include objectives for more complex learning tasks. Examples include computational objectives for canonical correlation analysis (CCA), slow feature analysis (SFA), independent component analysis (ICA), and contrastive PCA* (cPCA*), which can be interpreted, respectively, as linear instantiations of the following computational principles: associative learning of multimodal inputs, learning temporally invariant features, redundancy reduction, and contrastive learning. Interestingly, the algorithms derived from these objectives naturally map onto NNs with multicompartmental neurons and local, non-Hebbian forms of plasticity. Therefore, these works offer a potential normative account of these anatomical and physiological observations.

In this article, we provide a unified framework that encompasses and generalizes these normative models of NNs with multicompartmental neurons and non-Hebbian plasticity. In particular, we derive an online algorithm for solving a large class of symmetric generalized eigenvalue problems—which includes CCA, SFA, ICA, and cPCA* as special cases—which establishes a precise link between synaptic plasticity rules

and computational objectives. In one direction, this framework can be used to derive NNs for solving other symmetric generalized eigenvalue problems [39,40]. Conversely, given an experimentally observed non-Hebbian synaptic plasticity rule, this framework can potentially be used to predict a guiding computational objective. Therefore, we believe this unified framework will facilitate further development of NNs for solving other relevant learning tasks and advance our understanding of NNs with multicompartmental neurons and non-Hebbian plasticity.

The remainder of this work is organized as follows. We first review prior theoretical results on NNs with point neurons and Hebbian plasticity (Sec. II) and experimental results on multicompartmental neurons and non-Hebbian plasticity (Sec. III). We then present a unified objective for solving a large class of symmetric generalized eigenvalue problems (Sec. IV). Starting from this objective, we derive an online algorithm for solving the objective (Sec. V), and we show that for several examples the algorithm maps onto NNs with multicompartmental linear neurons and non-Hebbian learning rules (Sec. VI). Finally, by modifying the starting objective, we transform the problem from a symmetric generalized eigenvalue problem to a non-negative matrix factorization problem, resulting in NNs with rectified neural outputs (Sec. VII).

II. HEBBIAN NEURAL NETWORKS FOR UNSUPERVISED DIMENSIONALITY REDUCTION

Early sensory processing significantly reduces the dimensionality of the inputs [41,42]. For example, the human retina is a highly convergent pathway with more than 100-fold reduction in dimensionality from photoreceptors to retinal ganglion cells [43]. Therefore, NNs that perform unsupervised dimensionality reduction may be useful models of early sensory processing.

A. Oja’s neuron for principal component analysis

In a seminal work, Oja [4] modeled a single neuron with a PCA algorithm, Fig. 1 (left), which can be derived as a stochastic gradient descent minimizing a reconstruction error objective [44]. At each time point t the neuron receives n inputs, whose activities are encoded in the column vector \mathbf{x}_t . The inputs are multiplied by the corresponding synaptic weights, which are encoded in the row vector \mathbf{w} , and summed to generate the neuron’s scalar output $z_t = \mathbf{w}\mathbf{x}_t$. These synaptic weights are then updated according to the following plasticity rule, referred to as Oja’s rule:

$$\mathbf{w} \leftarrow \mathbf{w} + \eta(z_t \mathbf{x}_t^\top - z_t^2 \mathbf{w}),$$

where $\eta > 0$ denotes the learning rate for the synapses. After multiple iterations, the synaptic weights \mathbf{w} converge to the principal eigenvector of the input covariance matrix $\mathbf{C}_X := \langle \mathbf{x}_t \mathbf{x}_t^\top \rangle$ [4,45,46]. The first term in the synaptic update, $z_t \mathbf{x}_t^\top$, is the product of the pre- and postsynaptic activities, so it is referred to as Hebbian plasticity. The term $-z_t^2 \mathbf{w}$, which is proportional to the synaptic weights, can be viewed as a form of homeostatic plasticity that prevents the synaptic weights from diverging.

Algorithm 1. Online PSP.

```

input  $\{\mathbf{x}_t\}$ ; parameters  $\gamma > 0$  and  $0 < \eta < \tau$ 
initialize  $\mathbf{W} \in \mathbb{R}^{k \times n}$  and  $\mathbf{M} \in \mathbb{S}_{++}^k$ 
for  $t = 1, 2, \dots$  do
  repeat
     $\mathbf{z}_t \leftarrow \mathbf{z}_t + \gamma(\mathbf{W}\mathbf{x}_t - \mathbf{M}\mathbf{z}_t)$ 
  until convergence
   $\mathbf{W} \leftarrow \mathbf{W} + \eta(\mathbf{z}_t\mathbf{x}_t^\top - \mathbf{W})$ 
   $\mathbf{M} \leftarrow \mathbf{M} + \frac{\eta}{\tau}(\mathbf{z}_t\mathbf{z}_t^\top - \mathbf{M})$ 
end for

```

B. Hebbian neural networks derived from similarity matching objectives

Following Oja’s work, several extensions to multichannel NNs were proposed. In one line of work, online algorithms were derived from principal subspace projection (PSP) objectives and mapped onto single-layer neural networks [17–19]; however, the synaptic updates in these NNs are not *local*—they depend on variables that are not represented in the pre- or postsynaptic neurons—so they violate basic biophysical constraints. In another line of work, NNs with local, Hebbian learning rules for synapses were proposed [47–50]; however, the synaptic updates were postulated rather than derived from a principled objective, so the NNs are not normative and lack theoretical understanding.

To enjoy the benefits of both of these two lines of work, a normative approach and local learning rules, Pehlevan *et al.* [10,51] introduced the following similarity matching objective from multidimensional scaling [52]:

$$\min_{\mathbf{z}_1, \dots, \mathbf{z}_T \in \mathbb{R}^k} \frac{1}{T^2} \sum_{t=1}^T \sum_{t'=1}^T (\mathbf{x}_t^\top \mathbf{x}_{t'} - \mathbf{z}_t^\top \mathbf{z}_{t'})^2, \quad (1)$$

where \mathbf{z}_t denotes the output of the NN at time t . The objective minimizes the squared difference between the similarity of the inputs and the similarity of the outputs, where similarity is measured in terms of inner products. The optimal solution $\mathbf{z}_1, \dots, \mathbf{z}_T$ of objective (1) is the projection of the inputs $\mathbf{x}_1, \dots, \mathbf{x}_T$ onto their k -dimensional principal subspace, i.e., the subspace spanned by the top k principal components. Starting from this objective, Pehlevan *et al.* [10,51] derived an online gradient-based algorithm for PSP, Algorithm 1 (see Sec. V for a detailed derivation).

Algorithm 1 can be mapped onto a multichannel single-layer NN with k point neurons and Hebbian plasticity. The neural dynamics are assumed to operate on a fast timescale and equilibrate before the synapses are updated. The synaptic updates to the feed-forward weights \mathbf{W} are naturally viewed as a combination of Hebbian and homeostatic plasticity. Since the first term in the synaptic update for the recurrent connections $-\mathbf{M}$ is *inversely* proportional to the product of the pre- and postsynaptic activities, the update is often referred to as *anti-Hebbian*. Variations of the similarity matching objectives were used as starting points to derive Hebbian NNs for performing a number of unsupervised dimensionality reduction tasks that model aspects of early sensory processing [20]; however, these NNs cannot account for multicompartmental neurons and non-Hebbian forms of plasticity prevalent throughout the brain [30].

III. MULTICOMPARTMENTAL NEURONS AND NON-HEBBIAN PLASTICITY

Most neurons in the brain have multicompartmental structures and learn via non-Hebbian synaptic plasticity rules. For example, pyramidal neurons, which are the main excitatory neurons of the cortex, receive feed-forward excitatory inputs (e.g., lower-level sensory inputs) via their proximal dendrites and feedback excitatory inputs (e.g., from farther up the cortical hierarchy) via dendrites that extend from the distal apical tuft [53–56], Fig. 1 (right). These inputs are integrated in at least two electronically segregated dendritic compartments—a proximal compartment near the soma of the pyramidal neuron, and a distal compartment in the apical tuft [57–60]. Integrated proximal feed-forward inputs are the main drivers of the pyramidal neuron sodium action potential outputs [61,62], while integrated distal feedback inputs generate calcium plateau potentials that are effective drivers of synaptic plasticity in the proximal dendrites [63–67].

There are a number of existing consequential models of both individual pyramidal neurons and other multicompartmental neurons [68–83]. These models provide detailed biophysical descriptions of the neural dynamics and non-Hebbian synaptic plasticity, and, through numerical simulation, they demonstrate computational capabilities of the pyramidal neuron and cortical circuits. However, these models do not provide a *normative* framework for understanding multicompartmental neurons and non-Hebbian forms of plasticity.

IV. CCA AND SYMMETRIC GENERALIZED EIGENVALUE PROBLEMS

To develop a normative framework, we first propose a class of computational objectives. Many linear versions of behaviorally relevant learning tasks can be formulated as symmetric generalized eigenvalue problems. Before stating the general problem, we first present the special case of CCA in the context of a pyramidal neuron.

Consider a pyramidal neuron that receives inputs from two upstream populations of neurons, whose activities at time t are encoded as the components of the column vectors $\mathbf{x}_t \in \mathbb{R}^{n_x}$ and $\mathbf{y}_t \in \mathbb{R}^{n_y}$, Fig. 1 (right). One hypothesis is that the goal of the pyramidal neuron is to learn associations between these high-dimensional data streams. What should the objective be? A relevant associative learning objective is CCA [84], which identifies subspaces of the input data streams such that the corresponding projections of the inputs are maximally correlated. In one dimension, the objective is to find n_x -dimensional and n_y -dimensional row vectors \mathbf{w}_x and \mathbf{w}_y that maximize the covariance $\langle (\mathbf{w}_x \mathbf{x}_t)(\mathbf{w}_y \mathbf{y}_t) \rangle$ subject to the constraint $\langle (\mathbf{w}_x \mathbf{x}_t)^2 \rangle = \langle (\mathbf{w}_y \mathbf{y}_t)^2 \rangle = 1$.

CCA is a special case of the symmetric generalized eigenvalue problem

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{B}\mathbf{v}, \quad \mathbf{A} := \langle \boldsymbol{\xi}_t \boldsymbol{\xi}_t^\top \rangle, \quad \mathbf{B} := \langle \mathbf{B}_t \rangle \quad (2)$$

when $\mathbf{v}^\top = [\mathbf{w}_x, \mathbf{w}_y]$ and

$$\boldsymbol{\xi}_t = \begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix}, \quad \mathbf{B}_t = \begin{bmatrix} \mathbf{x}_t \mathbf{x}_t^\top & \\ & \mathbf{y}_t \mathbf{y}_t^\top \end{bmatrix}. \quad (3)$$

We consider the class of symmetric generalized eigenvalue problems of the form in Eq. (2), where the pair $(\xi_t, \mathbf{B}_t) \in \mathbb{R}^n \times \mathbb{S}_+^n$ is a function of the NN inputs; see Table I in Sec. VI for specific examples. Given such a symmetric generalized eigenvalue problem and $1 \leq k < n$, we refer to the projection of the vector ξ_t onto the k -dimensional subspace spanned by the top k eigenvectors as the generalized principal subspace projection (GPSP) of (ξ_t, \mathbf{B}_t) . Our goal is to derive an online multichannel GPSP algorithm that maps onto a NN with *local* learning rules, i.e., the synaptic updates only depend on variables that are represented in the pre- and postsynaptic neurons as well as globally broadcast variables. There are existing online CCA and GPSP algorithms [85–89]; however, these algorithms cannot be mapped onto NNs with local learning rules and/or they only find the top one-dimensional projection.

V. AN ONLINE GENERALIZED PRINCIPAL SUBSPACE PROJECTION ALGORITHM

Here, we derive an online GPSP algorithm and, in the next section, we show that for many relevant examples the algorithm maps onto a NN with multicompartamental neurons and local non-Hebbian learning rules. The reader who is not interested in the derivation can skip to the end of this section, where we state our algorithm (Algorithm 2).

A. Similarity matching objective

At each time step t , let $\zeta_t \in \mathbb{R}^k$ denote the GPSP of (ξ_t, \mathbf{B}_t) . A useful observation is that ζ_t is equal to the PSP of the normalized data $\sqrt{\mathbf{B}^\dagger} \xi_t$, where \mathbf{B}^\dagger is the Moore-Penrose inverse of \mathbf{B} —to see this, substitute in for \mathbf{A} and \mathbf{v} in Eq. (2) with $\sqrt{\mathbf{B}^\dagger} \mathbf{A} \sqrt{\mathbf{B}^\dagger}$ and $\sqrt{\mathbf{B} \mathbf{v}}$, respectively. Therefore, we can substitute $\sqrt{\mathbf{B}^\dagger} \xi_t$ and ζ_t in for \mathbf{x}_t and \mathbf{z}_t , respectively, in the similarity matching objective (1) for PSP to obtain the GPSP objective

$$\min_{\zeta_1, \dots, \zeta_T \in \mathbb{R}^k} \frac{1}{T^2} \sum_{t=1}^T \sum_{t'=1}^T (\xi_t^\top \mathbf{B}_t^\dagger \xi_{t'} - \zeta_t^\top \zeta_{t'})^2. \quad (4)$$

Every optimal solution ζ_t of the objective (4) is a PSP of the input $\sqrt{\mathbf{B}^\dagger} \xi_t$, which is a GPSP of the input data (ξ_t, \mathbf{B}_t) . When $\xi_t = \mathbf{x}_t$, $\zeta_t = \mathbf{z}_t$, and $\mathbf{B} = \mathbf{I}_n$, we recover the similarity matching objective (1) from [10].

B. Matrix substitutions

The objective (4) does not readily lead to an online GPSP algorithm. For example, direct optimization of the objective via gradient descent with respect to the output ζ_t requires taking gradient steps that depend on the inputs $(\xi_{t'}, \mathbf{B}_{t'})$ from every time point $t' = 1, \dots, T$. Rather, following the approach of Pehlevan *et al.* [51], we substitute in with dynamic matrix variables to obtain a minimax algorithm that can be solved in the online setting. These matrix variables will correspond to feed-forward and lateral synaptic weight matrices in the NN implementations.

For the cross term in Eq. (4), we introduce the synaptic weight matrix \mathbf{W} by substituting in with the Legendre

transform

$$\begin{aligned} & -\frac{1}{T} \sum_{t=1}^T \zeta_t^\top \left[\frac{1}{T} \sum_{t'=1}^T \zeta_{t'} \xi_{t'}^\top \mathbf{B}_t^\dagger \right] \xi_t \\ & = \min_{\mathbf{W} \in \mathbb{R}^{k \times n}} \frac{1}{T} \sum_{t=1}^T [-2\zeta_t^\top \mathbf{W} \xi_t + \text{Tr}(\mathbf{W} \mathbf{B}_t \mathbf{W}^\top)]. \end{aligned}$$

Differentiating the right-hand side of the equality with respect to \mathbf{W} , setting the derivative to zero, and solving for \mathbf{W} , we see that the optimum is achieved at $\frac{1}{T} \sum_{t'=1}^T \zeta_{t'} \xi_{t'}^\top \mathbf{B}_t^\dagger$. To account for the quartic term in (4), we introduce the synaptic weight matrix \mathbf{M} by substituting in with the Legendre transform

$$\begin{aligned} & \frac{1}{T} \sum_{t=1}^T \zeta_t^\top \left[\frac{1}{T} \sum_{t'=1}^T \zeta_{t'} \xi_{t'}^\top \right] \xi_t \\ & = \max_{\mathbf{M} \in \mathbb{S}_{++}^k} \frac{1}{T} \sum_{t=1}^T [2\zeta_t^\top \mathbf{M} \zeta_t - \text{Tr}(\mathbf{M}^2)], \end{aligned}$$

where \mathbb{S}_{++}^k denotes the set of $k \times k$ positive-definite matrices. Differentiating the right-hand side of the equality with respect to \mathbf{M} , setting the derivative to zero, and solving for \mathbf{M} , we see that the optimum is achieved at $\frac{1}{T} \sum_{t'=1}^T \zeta_{t'} \zeta_{t'}^\top$. Substituting the Legendre transformations into the objective (4), interchanging the order of optimization,¹ and dropping terms that do not depend on ζ_t , we arrive at the minimax objective:

$$\min_{\mathbf{W} \in \mathbb{R}^{k \times n}} \max_{\mathbf{M} \in \mathbb{S}_{++}^k} \frac{1}{T} \sum_{t=1}^T \min_{\zeta_t \in \mathbb{R}^k} \ell(\mathbf{W}, \mathbf{M}, \zeta_t, \xi_t, \mathbf{B}_t), \quad (5)$$

where

$$\begin{aligned} & \ell(\mathbf{W}, \mathbf{M}, \zeta_t, \xi_t, \mathbf{B}_t) \\ & := 2\text{Tr}(\mathbf{W} \mathbf{B}_t \mathbf{W}^\top) - \text{Tr}(\mathbf{M}^2) - 4\zeta_t^\top \mathbf{W} \xi_t + 2\zeta_t^\top \mathbf{M} \zeta_t. \end{aligned} \quad (6)$$

As a result of introducing the matrix variables, \mathbf{W} and \mathbf{M} , we have transformed the minimization problem (4) into the minimax objective (5). This objective has the desirable property that for fixed \mathbf{W} and \mathbf{M} , the optimal output ζ_t at time step t only depends on the input ξ_t at time step t .

C. Online algorithm

To derive an online algorithm, we assume there is a separation of timescales between the minimization over the vectors ζ_t , which will correspond to neural activities, and the optimization of the matrices \mathbf{W} and \mathbf{M} , which will correspond to synaptic weights. At each time step t , we minimize $\ell(\mathbf{W}, \mathbf{M}, \zeta_t, \xi_t, \mathbf{B}_t)$ with respect to ζ_t by running gradient descent steps until convergence,

$$\zeta_t \leftarrow \zeta_t + \gamma(\mathbf{W} \xi_t - \mathbf{M} \zeta_t) \Rightarrow \zeta_t = \mathbf{M}^{-1} \mathbf{W} \xi_t. \quad (7)$$

After ζ_t equilibrates, we optimize $\langle \ell(\mathbf{W}, \mathbf{M}, \zeta_t, \xi_t, \mathbf{B}_t) \rangle$ with respect to the matrix variables by taking a stochastic gradient

¹Changing the order of optimization in this problem does not affect the solution due to the saddle point property; see Ref. [90], Sec. 5.4.

Algorithm 2. Online GPSP.

input $\{(\xi_t, \mathbf{B}_t)\}$; parameters $\gamma > 0$ and $0 < \eta < \tau$
initialize $\mathbf{W} \in \mathbb{R}^{k \times n}$ and $\mathbf{M} \in \mathbb{S}_{++}^k$
for $t = 1, 2, \dots$ **do**
 repeat
 $\zeta_t \leftarrow \zeta_t + \gamma(\mathbf{W}\xi_t - \mathbf{M}\zeta_t)$
 until convergence
 $\mathbf{W} \leftarrow \mathbf{W} + 2\eta(\zeta_t \xi_t^\top - \mathbf{W}\mathbf{B}_t)$
 $\mathbf{M} \leftarrow \mathbf{M} + \frac{\eta}{\tau}(\zeta_t \zeta_t^\top - \mathbf{M})$
end for

descent-ascent step in \mathbf{W} and \mathbf{M} :

$$\mathbf{W} \leftarrow \mathbf{W} + 2\eta(\zeta_t \xi_t^\top - \mathbf{W}\mathbf{B}_t), \quad \mathbf{M} \leftarrow \mathbf{M} + \frac{\eta}{\tau}(\zeta_t \zeta_t^\top - \mathbf{M}). \quad (8)$$

Here $\eta > 0$ is the step size for the stochastic gradient descent steps in \mathbf{W} , and $\tau > 0$ denotes the ratio between the learning rate for \mathbf{W} and the learning rate for \mathbf{M} . This yields our online GPSP algorithm, Algorithm 2.

There are a few points worth noting:

(i) Algorithm 2 reduces to Algorithm 1 when $\xi_t = \mathbf{x}_t$, $\zeta_t = \mathbf{z}_t$, and $\mathbf{B}_t = \mathbf{I}_n$ for all t .

(ii) Since ℓ is nonconvex-concave in \mathbf{W} and \mathbf{M} , the minimization over \mathbf{W} cannot be interchanged with the maximization over \mathbf{M} in Eq. (5). Therefore, to ensure convergence of the synaptic weights, the \mathbf{M} updates need to be sufficiently fast relative to the \mathbf{W} updates, i.e., $\tau > 0$ needs to be sufficiently small.

(iii) Since the symmetric generalized eigenvalue problem is defined in terms of the averages $\mathbf{A} := \langle \xi_t \xi_t^\top \rangle$ and $\mathbf{B} := \langle \mathbf{B}_t \rangle$ and the synaptic update rules are in terms of (ξ_t, \mathbf{B}_t) , which are functions on the NN inputs, Algorithm 2 establishes a precise relationship between the symmetric generalized eigenvalue problem and the synaptic learning rules via the variables (ξ_t, \mathbf{B}_t) .

In general, the biological plausibility and biological interpretation of Algorithm 2 depend on the specific form of ξ_t and \mathbf{B}_t .

VI. EXAMPLES OF NEURAL NETWORKS FOR GENERALIZED PRINCIPAL SUBSPACE PROJECTION

We consider several biologically relevant symmetric generalized eigenvalue problems that can be solved using Algorithm 2—for different choices of the vector ξ_t and matrix \mathbf{B}_t , Table I. For each symmetric generalized eigenvalue problem, we map its online algorithm onto a NN with multicompartmental neurons and non-Hebbian learning rules, Fig. 2.

A. Canonical correlation analysis

As discussed in Sec. IV, canonical correlation analysis (CCA) may serve as a useful objective for understanding computation in pyramidal cells and cortical circuits. Using our approach, we derived an online CCA algorithm that maps onto a NN with multicompartmental neurons and non-Hebbian plasticity [34]. Substituting the expressions for (ξ_t, \mathbf{B}_t) from

TABLE I. A list of learning tasks with symmetric generalized eigenvalue problem formulations that can be solved with NNs derived using our framework.

Learning task	ξ_t	\mathbf{B}_t	No. of compartments
PCA	\mathbf{x}_t	\mathbf{I}_n	1
CCA	$\begin{bmatrix} \mathbf{x}_t \\ \mathbf{y}_t \end{bmatrix}$	$\begin{bmatrix} \mathbf{x}_t \mathbf{x}_t^\top & \\ & \mathbf{y}_t \mathbf{y}_t^\top \end{bmatrix}$	3
SFA	$\mathbf{x}_t + \mathbf{x}_{t-1}$	$\mathbf{x}_t \mathbf{x}_t^\top$	2
ICA (FOBI)	\mathbf{x}_t	$\ \mathbf{C}_X^{-1/2} \mathbf{x}_t\ ^2 \mathbf{x}_t \mathbf{x}_t^\top$	2
cPCA*	$\delta_t \mathbf{x}_t$	$(1 - \delta_t) \mathbf{x}_t \mathbf{x}_t^\top$	2

Eq. (3) into Algorithm 2 results in an online algorithm that maps onto a single-layer NN, Fig. 2 (far left).

At each time step t , the NN receives inputs \mathbf{x}_t and \mathbf{y}_t . The inputs are projected onto the feed-forward synaptic weights \mathbf{W}_x and \mathbf{W}_y , which combine to form the feed-forward weight matrix $\mathbf{W} := [\mathbf{W}_x, \mathbf{W}_y]$, to generate dendritic currents $\mathbf{a}_t = \mathbf{W}_x \mathbf{x}_t$ and $\mathbf{b}_t = \mathbf{W}_y \mathbf{y}_t$ that are stored in separate dendritic compartments. The output of the neurons $\mathbf{z}_t = \zeta_t$, which is represented in a third compartment, is computed by running the recurrent neural dynamics:

$$\mathbf{z}_t \leftarrow \mathbf{z}_t + \gamma(\mathbf{a}_t + \mathbf{b}_t - \mathbf{M}\mathbf{z}_t) \Rightarrow \mathbf{z}_t = \mathbf{M}^{-1}(\mathbf{a}_t + \mathbf{b}_t).$$

After the neural dynamic equilibrate, the synaptic weights are updated. The feed-forward synaptic weight updates are given by

$$\mathbf{W}_x \leftarrow \mathbf{W}_x + 2\eta(\mathbf{z}_t - \mathbf{a}_t) \mathbf{x}_t^\top, \quad \mathbf{W}_y \leftarrow \mathbf{W}_y + 2\eta(\mathbf{z}_t - \mathbf{b}_t) \mathbf{y}_t^\top.$$

Since the components of the vectors \mathbf{a}_t , \mathbf{b}_t , and \mathbf{z}_t are represented in the postsynaptic neurons, the synaptic updates are local, but non-Hebbian. The lateral recurrent synaptic weight updates are as in Algorithm 2 with $\zeta_t = \mathbf{z}_t$.

The NN is consistent with certain aspects of experimentally observed physiology and anatomy of pyramidal neurons and cortical circuits. Each neuron includes two dendritic compartments that separately integrate the inputs \mathbf{x}_t and \mathbf{y}_t . Rearranging the formula for the equilibrium neural outputs \mathbf{z}_t of the NN, we see that $\mathbf{b}_t = \mathbf{M}\mathbf{z}_t - \mathbf{a}_t$ and so we can rewrite the proximal synaptic updates as

$$\mathbf{W}_y \leftarrow \mathbf{W}_y + 2\eta(\mathbf{a}_t - [\mathbf{M} - \mathbf{I}_k] \mathbf{z}_t) \mathbf{y}_t^\top.$$

From this formulation of the synaptic update, we can interpret the difference between the distal currents \mathbf{a}_t and the recurrent lateral feedback $-[\mathbf{M} - \mathbf{I}_k] \mathbf{z}_t$ as the calcium plateau potential that drives non-Hebbian plasticity in the proximal synapses, which is consistent with experimental observations that distal currents generate calcium plateau potentials that drive plasticity in the proximal synapses, and these plateaus are mediated by inhibitory inputs [63–67].

The NN for CCA includes direct lateral connections between the pyramidal cells; however, in cortical circuits, lateral communication is typically mediated by local interneurons. By modifying the starting CCA objective to include an output whitening constraint, we can derive an algorithm that faithfully maps onto the wiring diagram of a cortical microcircuit consisting of both pyramidal neurons and interneurons [34].

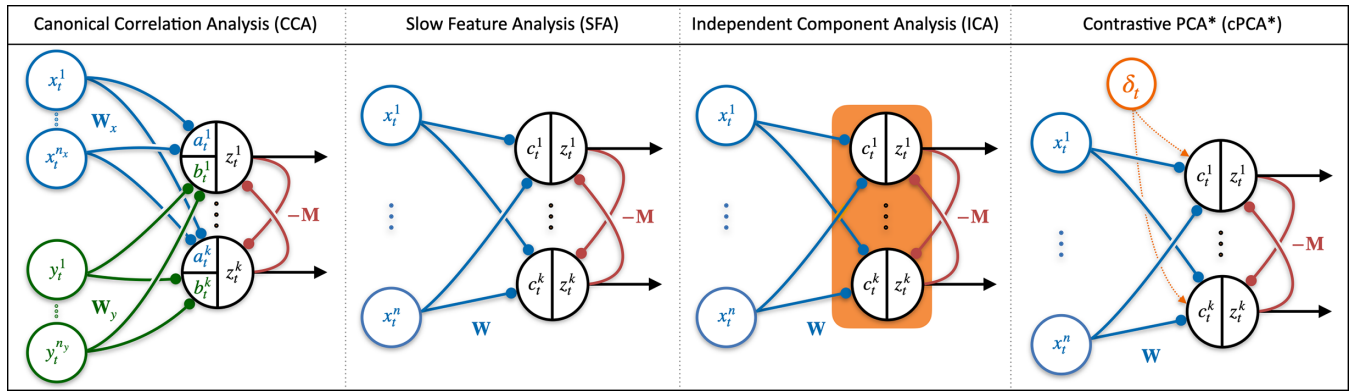


FIG. 2. NNs for GPSP. In each NN, circles denote neurons, and solid lines with circles at the ends are synapses. The synaptic updates depend on the variables inside the circles, which are encoded in different compartments of the neuron, and globally broadcast variables. For ICA, the orange box denotes the presence of a scalar variable (specifically, $\|\mathbf{z}_t\|^2$) that is available to each output neuron. For cPCA*, the scalar δ_t is an indicator variable that gates the output of the NN and is available to each output neuron.

Finally, the output of the neurons in this circuit is symmetric in the integrated currents \mathbf{a}_t and \mathbf{b}_t ; however, experimental evidence suggests that the feed-forward proximal inputs and feedback distal inputs are integrated asymmetrically by the pyramidal neuron [61–67], which is in contrast to our model that treats the inputs symmetrically. In [91,92], we derived an online algorithm for CCA when interpreted as a *supervised* learning task, where the feed-forward inputs are feature vectors and the feedback inputs are supervisory signals. In this case, the output of the circuit is exclusively driven by the feed-forward inputs.

B. Slow feature analysis

Brains are adept at learning meaningful latent representations from noisy, high-dimensional data. Often, the relevant features in the environment (e.g., objects) change slowly compared with noisy sensory data, so temporal slowness has been proposed as a computational principle for identifying latent features [93–95]. A popular approach for extracting slow features, introduced by Wiskott and Sejnowski [95], is called SFA. SFA is an unsupervised learning algorithm that extracts the slowest projection, in terms of discrete time derivative, from a nonlinear expansion of the input signal. When trained on natural images, SFA learns features that resemble properties of complex cells in the primary visual cortex [96]. Further, when trained on a simulated visual stream, a hierarchical version of SFA learns representations of orientation and space similar to those encoded in the hippocampus [97]. Together, these observations suggest that the cortex may use temporal slowness as a computational principle in representation learning.

The projection of the nonlinear expansion can be formulated as a generalized eigenvalue problem of the form (2) with $\xi_t = \mathbf{x}_t + \mathbf{x}_{t-1}$ and $\mathbf{B}_t = \mathbf{x}_t \mathbf{x}_t^\top$. Substituting into Algorithm 2 yields an online algorithm that maps onto a single-layer NN [35], Fig. 2 (middle left). At each time step t , the NN receives inputs \mathbf{x}_t , which are projected onto the weight matrix \mathbf{W} to generate dendritic currents $\mathbf{c}_t = \mathbf{W}\mathbf{x}_t$. The dendritic currents \mathbf{c}_t are stored in a separate compartment from the NN outputs \mathbf{z}_t , so each neuron consists of two compartments. Letting $\zeta_t = \mathbf{z}_t + \mathbf{z}_{t-1}$ and assuming the weights do not significantly

change between time steps $t-1$ and t , we can rewrite the neural dynamics in Algorithm 2 as

$$\mathbf{z}_t \leftarrow \mathbf{z}_t + \gamma(\mathbf{c}_t - \mathbf{M}\mathbf{z}_t) \Rightarrow \mathbf{z}_t = \mathbf{M}^{-1}\mathbf{c}_t.$$

Substituting in with the definitions of ξ_t , \mathbf{B}_t , \mathbf{c}_t , and ζ_t , the feed-forward synaptic weight updates from Algorithm 2 are given by $\Delta\mathbf{W} = \eta((\mathbf{z}_t + \mathbf{z}_{t-1})(\mathbf{x}_t + \mathbf{x}_{t-1})^\top - \mathbf{c}_t \mathbf{x}_t^\top)$. The synaptic update depends on low-frequency signals in both the pre- and postsynaptic neurons; however, dendrites are more likely to store low-frequency signals than axons.

If the input time series is stationary and reversible in time (i.e., $\langle \mathbf{x}_t \mathbf{x}_{t-1}^\top \rangle = \langle \mathbf{x}_{t-1} \mathbf{x}_t^\top \rangle$), we can rewrite the synaptic updates so they only depend on low-frequency signals in the postsynaptic neurons and

$$\mathbf{W} \leftarrow \mathbf{W} + 2\eta(2\mathbf{z}_t + 2\mathbf{z}_{t-1} - \mathbf{c}_t)\mathbf{x}_t^\top.$$

Empirically, this modification extracts slow signals even when the time series is not reversible [35]. Assuming that the postsynaptic neurons represent their low-pass filtered activities $\mathbf{z}_t + \mathbf{z}_{t-1}$ as well as their dendritic currents \mathbf{c}_t , the learning rules only depend on variables that are available in the pre- and postsynaptic neurons and so the synaptic updates are local, but non-Hebbian. Finally, the first term in the synaptic weights update resembles the first term in the following local “trace rule” proposed by Földiák [93] for learning temporally invariant features:

$$\mathbf{W}_{\text{trace}} \leftarrow \mathbf{W}_{\text{trace}} + \eta((\mathbf{z}_t + \mathbf{z}_{t-1})\mathbf{x}_t^\top - \text{diag}(\mathbf{z}_t + \mathbf{z}_{t-1})\mathbf{W}_{\text{trace}}).$$

Therefore, this normative NN model establishes a relationship between a computational objective for SFA and a variant of the proposed trace rule.

C. Independent components analysis

Efficient coding theories of sensory processing posit that early sensory layers transform their inputs to reduce redundancy [98,99]. Independent components analysis (ICA) is a statistical method for reducing redundancy by factorizing the sensory inputs into “independent components,” and it can explain edge detector neurons in area V1 of the visual cortex

[100,101] and receptive fields of cochlear nerve fibers in the auditory system [102].

ICA assumes a generative model $\mathbf{x}_t = \mathbf{A}\mathbf{s}_t$, where \mathbf{s}_t is the n -dimensional source vectors with independent components, \mathbf{A} is an $n \times n$ mixing matrix, and \mathbf{x}_t is the n -dimensional mixture vector. One method for solving ICA, called fourth-order blind identification (FOBI) [103], assumes the components of the sources have distinct kurtosis (i.e., fourth-order moments). FOBI can be solved in three steps: (i) whitening the mixture vector: $\mathbf{h}_t = \mathbf{C}_X^{-1/2}\mathbf{x}_t$; (ii) weighting the whitened mixtures by their norms: $\mathbf{y}_t = \|\mathbf{h}_t\|\mathbf{h}_t$; and (iii) projecting the whitened mixtures \mathbf{h}_t onto the principal components of \mathbf{y}_t . Remarkably, these three steps can be combined and expressed as a single symmetric generalized eigenvalue problem of the form (2) with $\xi_t = \mathbf{x}_t$ and $\mathbf{B}_t = \|\mathbf{h}_t\|^2\mathbf{x}_t\mathbf{x}_t^\top$, so we can apply our framework. Substituting into Algorithm 2 results in an online algorithm that maps onto a single-layer NN with multicompartmental neurons and local non-Hebbian learning rules [36], Fig. 2 (middle right).

At each time step t , the NN receives inputs \mathbf{x}_t , which are projected onto the feed-forward synaptic weights \mathbf{W} to generate dendritic currents $\mathbf{c}_t = \mathbf{W}\mathbf{x}_t$. The dendritic currents \mathbf{c}_t are stored in a separate compartment from the NN outputs $\mathbf{z}_t = \zeta_t$, so each neuron consists of two compartments. The neural dynamics in Algorithm 2 can be written as

$$\mathbf{z}_t \leftarrow \mathbf{z}_t + \gamma(\mathbf{c}_t - \mathbf{M}\mathbf{z}_t) \Rightarrow \mathbf{z}_t = \mathbf{M}^{-1}\mathbf{c}_t.$$

Substituting the expression for \mathbf{B}_t into the feed-forward synaptic weight update in Algorithm 2 results in the update $\Delta\mathbf{W} = \eta(\mathbf{z}_t - \|\mathbf{h}_t\|^2\mathbf{c}_t)\mathbf{x}_t^\top$. As stated in step (iii) of FOBI, the outputs \mathbf{z}_t are equal to the (full-rank) projection of the whitened inputs \mathbf{h}_t onto the principal components of \mathbf{y}_t , which implies that \mathbf{z}_t is an orthogonal transformation of the whitened inputs at each time t . Therefore, $\|\mathbf{h}_t\| = \|\mathbf{z}_t\|$ and we can rewrite the feed-forward synaptic update as

$$\mathbf{W} \leftarrow \mathbf{W} + 2\eta(\mathbf{z}_t - \|\mathbf{z}_t\|^2\mathbf{c}_t)\mathbf{x}_t^\top.$$

Interestingly, the resulting synaptic learning rules are globally modulated by the total activity of the output neurons $\|\mathbf{z}_t\|^2$, which could be accounted for by biophysical quantities such as neuromodulators, extracellular calcium, local field potential, or nitric oxide.

D. Contrastive principal component analysis*

Sensory organs receive an immense amount of information per unit time, but much of it is of little relevance for behavior. A simple approach to process this high-dimensional input is to focus on a lower-dimensional subspace and ignore the directions that are less informative. PCA achieves this by discarding the directions with low variance. Such an approach is, however, inefficient in cases where the irrelevant directions are very noisy, thus having greater variance than the relevant ones. If we have access to representative samples of variability in irrelevant directions (“negative samples”), we can achieve better efficiency by using a contrastive variant of PCA, as in [104]. Contrastive PCA (cPCA) finds the subspace of highest *relevant* variance, associated with “positive samples,” while minimizing the variance associated with irrelevant information, as inferred from negative samples.

In [37], we consider a more robust cPCA method, which we refer to as cPCA*, and we derive an online algorithm with a neural implementation. Assume we have a sequence of centered inputs $(\mathbf{x}_1, \delta_1), \dots, (\mathbf{x}_T, \delta_T) \in \mathbb{R}^n \times \{0, 1\}$. At each time t , the input \mathbf{x}_t is a feature vector that is either a *positive sample* or a *negative sample*, so the positive and negative samples arrive via the same pathway. The scalar variable δ_t is equal to 1 (0) if the \mathbf{x}_t is a positive (negative) sample. The covariance matrices for the positive and negative samples are given, respectively, by $\mathbf{C}_{(+)} := \langle \mathbf{x}_t\mathbf{x}_t^\top | \delta_t = 1 \rangle$ and $\mathbf{C}_{(-)} := \langle \mathbf{x}_t\mathbf{x}_t^\top | \delta_t = 0 \rangle$. The goal of cPCA* is to project the feature vectors \mathbf{x}_t onto vectors \mathbf{v} to maximize the ratio of $\mathbf{v}^\top\mathbf{C}_{(+)}\mathbf{v}$ and $\mathbf{v}^\top\mathbf{C}_{(-)}\mathbf{v}$, which corresponds to the symmetric generalized eigenvalue problem of the form (2) with $\xi_t := \delta_t\mathbf{x}_t$ and $\mathbf{B}_t := (1 - \delta_t)\mathbf{x}_t\mathbf{x}_t^\top$. When the positive samples are measurements of signal + noise and the negative samples are measurements of noise, the problem is closely related to linear discriminant analysis [105] and joint decorrelation methods [106]. Substituting into Algorithm 2 results in an online algorithm that maps onto a single-layer NN with multicompartmental neurons and local non-Hebbian learning rules, Fig. 2 (far right).

At each time step t , the NN receives inputs \mathbf{x}_t and δ_t . The inputs \mathbf{x}_t are projected onto the feed-forward weights \mathbf{W} to generate dendritic currents $\mathbf{c}_t = \mathbf{W}\mathbf{x}_t$. The neural dynamics and synaptic updates for positive and negative samples are given as follows:

δ_t	Fast neural dynamics	Slow synaptic updates
1	$\mathbf{z}_t \leftarrow \mathbf{z}_t + \gamma(\mathbf{c}_t - \mathbf{M}\mathbf{z}_t)$	$\mathbf{W} \leftarrow \mathbf{W} + 2\eta\mathbf{z}_t\mathbf{x}_t^\top$
0	$\mathbf{z}_t = \mathbf{0}$	$\mathbf{W} \leftarrow \mathbf{W} - 2\eta\mathbf{c}_t\mathbf{x}_t^\top$

The scalar δ_t is naturally interpreted as indicated by the presence or absence of a neuromodulator that gates the output \mathbf{z}_t of the NN. Assuming the dendritic currents \mathbf{c}_t and neural outputs \mathbf{z}_t are represented in the postsynaptic neurons, and δ_t is globally available, the synaptic updates are local and non-Hebbian. In both cases, the lateral synapses $-\mathbf{M}$ are updated according to Algorithm 2 with $\zeta_t = \mathbf{z}_t$.

VII. NON-NEGATIVE SIMILARITY MATCHING

Most biological neurons perform nonlinear transformations and have non-negative outputs. In addition, many interesting computations require nonlinear transformations. We can adapt the objective to account for both the nonlinear transformation and the non-negativity of the neural outputs by imposing a non-negativity constraint on the outputs ζ_t in Eq. (4), which results in the non-negative similarity matching objective:

$$\min_{\zeta_1, \dots, \zeta_T \in \mathbb{R}_+^k} \frac{1}{T^2} \sum_{t=1}^T \sum_{t'=1}^T (\xi_t^\top \mathbf{B}^\dagger \xi_{t'} - \zeta_t^\top \zeta_{t'})^2, \quad (9)$$

where \mathbb{R}_+^k denotes the non-negative orthant in \mathbb{R}^k . Imposing the non-negative constraint transforms the problem from a spectral matrix factorization problem to a non-negative matrix factorization problem. For the special case in which $\mathbf{B} = \mathbf{I}_n$, Pehlevan and Chklovskii [107], Bahroun and Soltoggio [108], Sengupta *et al.* [109], and Qin *et al.* [110] explored the

relationship between the objective (9) and clustering, sparse representation learning, manifold tiling, and supervised learning. However, aside from a specific generative model (see Sec. VII B below), it is not clear how to interpret objective (9) when \mathbf{B} is not the identity matrix.

A. An online algorithm

We can derive an online algorithm for solving the objective (9) following the same steps as in Sec. V for deriving an online GPSP algorithm. First, we perform the same matrix substitutions to arrive at the minimax problem

$$\min_{\mathbf{W} \in \mathbb{R}^{k \times n}} \max_{\mathbf{M} \in \mathbb{S}_{++}^k} \min_{\zeta_1, \dots, \zeta_T \in \mathbb{R}_+^k} (\ell(\mathbf{W}, \mathbf{M}, \zeta_t, \xi_t, \mathbf{B}_t)), \quad (10)$$

where ℓ is defined as in Eq. (6). To solve the minimax problem (10) in the online setting, at each time step t , we first minimize ℓ with respect to $\zeta_t \in \mathbb{R}_+^k$ by taking *projected* gradient steps until convergence:

$$\zeta_t \leftarrow [\zeta_t + \gamma(\mathbf{W}\xi_t - \mathbf{M}\zeta_t)]_+,$$

where $[\cdot]_+$ denotes taking the non-negative part elementwise. The minimization is still over a convex set; however, unlike the GPSP setting, we do not have a closed-form expression for the output ζ_t . After ζ_t converges, we update the matrices \mathbf{W} and \mathbf{M} by taking a stochastic gradient descent-ascent step, which results in the exact same updates as in Eq. (8).

B. Non-negative independent component analysis

As discussed in Sec. VIC, ICA is a statistical method for factorizing sensory inputs into independent components. A special case is called non-negative independent component analysis (NICA), which assumes a generative model in which the mixture of stimuli is a linear combination of uncorrelated, non-negative sources; i.e., $\mathbf{x}_t = \mathbf{A}\mathbf{s}_t$, where \mathbf{s}_t denotes the non-negative vector of source intensities, \mathbf{A} is a mixing matrix, and \mathbf{x}_t denotes the vector of mixed stimuli. The goal of NICA is to infer the non-negative source vectors \mathbf{s}_t from the mixture vectors \mathbf{x}_t . Both the linear additivity of stimuli and the non-negativity of the sources are reasonable assumptions in biological applications. For example, in olfaction, concentrations of odorants are both additive and non-negative.

While NICA cannot be expressed as a GPSP problem, it can be solved using the non-negative similarity matching framework. Pehlevan *et al.* [111] solved NICA with an online algorithm that can be implemented in a two-layer network with point neurons and Hebbian/anti-Hebbian learning rules, where each layer is derived from a separate objective function. The two objective functions can be combined into a single non-negative similarity matching objective of the form (9) with $\xi_t = \mathbf{x}_t$ and $\mathbf{B}_t = (\mathbf{x}_t - \langle \mathbf{x}_t \rangle)(\mathbf{x}_t - \langle \mathbf{x}_t \rangle)^\top$. Starting

from the non-negative similarity matching objective, we derived an online algorithm for solving NICA that maps onto a single-layer network with multicompartmental neurons and non-Hebbian plasticity [38].

At each time step t the NN receives inputs \mathbf{x}_t which are projected onto the feed-forward weights to generate the dendritic current \mathbf{c}_t , which is stored in a separate compartment from the NN outputs $\mathbf{z}_t = \zeta_t$. The fast neural dynamics are given by

$$\mathbf{z}_t \leftarrow [\mathbf{z}_t + \gamma(\mathbf{c}_t - \mathbf{M}\mathbf{z}_t)]_+.$$

After the neural dynamics equilibrate, the feed-forward synaptic weights are updated according to the learning rule

$$\mathbf{W} \leftarrow \mathbf{W} + 2\eta(\mathbf{z}_t\mathbf{x}_t^\top - (\mathbf{c}_t - \bar{\mathbf{c}}_t)(\mathbf{x}_t - \bar{\mathbf{x}}_t)^\top),$$

where we have replaced $\mathbf{W}\langle \mathbf{x}_t \rangle$ ($\langle \mathbf{x}_t \rangle$) with the running average $\bar{\mathbf{c}}_t$ ($\bar{\mathbf{x}}_t$) of the dendritic current (inputs), which could be physically represented as local ion concentrations at the synapses.

VIII. DISCUSSION

In this work, we proposed an extension of the similarity matching objective to include a broad class of symmetric generalized eigenvalue problems. Starting from this objective, we derived an online algorithm and showed that for several examples, the algorithm maps onto a NN with multicompartmental neurons and local, non-Hebbian learning rules. Furthermore, we proposed a modification of our framework to solve a broad class of non-negative matrix factorization problems, and we mapped a specific example onto a NN with multicompartmental neurons, local learning rules, and rectified outputs.

Our framework establishes a precise relationship between synaptic learning rules and computational objectives. In particular, the synaptic learning rules in Algorithm 2 are related to the symmetric generalized eigenvalue problem (2) via the variables ξ_t and \mathbf{B}_t . Therefore, given a symmetric generalized eigenvalue problem of the form (2), one can predict the synaptic learning rules for the NN. Conversely, given synaptic learning rules of the form in Algorithm 2, one can predict the computational objective for the NN. We believe this unified framework for relating non-Hebbian synaptic learning rules to computational objectives will be useful for understanding forms of non-Hebbian plasticity found throughout the brain.

Finally, in addition to the framework presented here, there are other normative approaches for deriving online algorithms that map onto NNs with multicompartmental neurons and solve symmetric generalized eigenvalue problems and other related problems [91, 112–115].

ACKNOWLEDGMENT

We thank Pierre-Étienne Fiquet for helpful feedback on an earlier draft of this work.

[1] F. Attneave, Some informational aspects of visual perception, *Psycholog. Rev.* **61**, 183 (1954).
 [2] H. B. Barlow, Possible principles underlying the transformation of sensory messages, in *Sensory Communication*, edited by W. A. Rosenblith (MIT Press, Cambridge, MA, 1961), p. 217.

[3] M. V. Srinivasan, S. B. Laughlin, and A. Dubs, Predictive coding: a fresh view of inhibition in the retina, *Proc. R. Soc. London B* **216**, 427 (1982).
 [4] E. Oja, Simplified neuron model as a principal component analyzer, *J. Math. Biol.* **15**, 267 (1982).

- [5] J. J. Atick and A. N. Redlich, What does the retina know about natural scenes? *Neural Comput.* **4**, 196 (1992).
- [6] J. H. van Hateren, A theory of maximizing sensory information, *Biol. Cybern.* **68**, 23 (1992).
- [7] B. A. Olshausen and D. J. Field, Sparse coding with an overcomplete basis set: A strategy employed by V1? *Vision Res.* **37**, 3311 (1997).
- [8] R. P. N. Rao and D. H. Ballard, Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects, *Nat. Neurosci.* **2**, 79 (1999).
- [9] B. L. Chen, D. H. Hall, and D. B. Chklovskii, Wiring optimization can relate neuronal structure and function, *Proc. Natl. Acad. Sci. USA* **103**, 4723 (2006).
- [10] C. Pehlevan, T. Hu, and D. B. Chklovskii, A Hebbian/anti-Hebbian neural network for linear subspace learning: A derivation from multidimensional scaling of streaming data, *Neural Comput.* **27**, 1461 (2015).
- [11] W. F. Młynarski and A. M. Hermundstad, Efficient and adaptive sensory codes, *Nat. Neurosci.* **24**, 998 (2021).
- [12] K. Pearson, LIII. on lines and planes of closest fit to systems of points in space, *London, Edinburgh Dublin Philos. Mag. J. Sci.* **2**, 559 (1901).
- [13] D. O. Hebb, *The Organisation of Behaviour*, 1st ed. (Wiley & Sons, New York, 1949).
- [14] T. V. P. Bliss and A. R. Gardner-Medwin, Long-lasting potentiation of synaptic transmission in the dentate area of the unanaesthetized rabbit following stimulation of the perforant path, *J. Physiol.* **232**, 357 (1973).
- [15] T. V. P. Bliss and T. Lømo, Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path, *J. Physiol.* **232**, 331 (1973).
- [16] N. Caporale and Y. Dan, Spike timing-dependent plasticity: A Hebbian learning rule, *Annu. Rev. Neurosci.* **31**, 25 (2008).
- [17] E. Oja and J. Karhunen, An analysis of convergence for a learning version of the subspace method, *J. Math. Anal. Appl.* **91**, 102 (1983).
- [18] T. D. Sanger, Optimal unsupervised learning in a single-layer linear feedforward neural network, *Neural Netw.* **2**, 459 (1989).
- [19] T. Leen, M. Rudnick, and D. Hammerstrom, Hebbian feature discovery improves classifier efficiency, in *1990 IJCNN International Joint Conference on Neural Networks* (IEEE, Piscataway, NJ, 1990), pp. 51–56.
- [20] C. Pehlevan and D. B. Chklovskii, Neuroscience-inspired online unsupervised learning algorithms: Artificial neural networks, *IEEE Sign. Proc. Mag.* **36**, 88 (2019).
- [21] D. Obeid, H. Ramambason, and C. Pehlevan, Structure and deep similarity matching via structured and deep Hebbian networks, *Advances in Neural Information Processing Systems*, Vol. 32 (Curran Inc., 2019).
- [22] T. Teşileanu, S. Golkar, S. Nasiri, A. M. Sengupta, and D. B. Chklovskii, Neural circuits for dynamics-based segmentation of time series, *Neural Comput.* **34**, 891 (2022).
- [23] A. Genkin, D. Lipshutz, S. Golkar, T. Teşileanu, and D. B. Chklovskii, Biological learning of irreducible representations of commuting transformations, *Advances in Neural Information Processing Systems*, Vol. 35 (Curran Inc., 2022), pp. 20838–20849.
- [24] C. Pehlevan, A. Genkin, and D. B. Chklovskii, A clustering neural network model of insect olfaction, in *2017 51st Asilomar Conference on Signals, Systems, and Computers* (IEEE, Piscataway, NJ, 2017), pp. 593–600.
- [25] Y. Bahroun, D. Chklovskii, and A. Sengupta, A similarity-preserving network trained on transformed images recapitulates salient features of the fly motion detection circuit, *Advances in Neural Information Processing Systems*, Vol. 32 (Curran Associates, Inc., 2019).
- [26] M. K. Benna and S. Fusi, Place cells may simply be memory cells: Memory compression leads to spatial tuning and history dependence, *Proc. Natl. Acad. Sci. USA* **118**, e2018422118 (2021).
- [27] N. M. Chapochnikov, C. Pehlevan, and D. B. Chklovskii, Normative and mechanistic model of an adaptive circuit for efficient encoding and feature extraction, *Proc. Natl. Acad. Sci. USA* **120**, e2117484120 (2023).
- [28] S. Qin, S. Farashahi, D. Lipshutz, A. M. Sengupta, D. B. Chklovskii, and C. Pehlevan, Coordinated drift of receptive fields in Hebbian/anti-Hebbian network models during noisy representation learning, *Nat. Neurosci.* **26**, 339 (2023).
- [29] D. Lipshutz, C. Pehlevan, and D. B. Chklovskii, Interneurons accelerate learning dynamics in recurrent neural networks for statistical adaptation, *The Eleventh International Conference on Learning Representations* (ICLR, 2023).
- [30] J. C. Magee and C. Grienberger, Synaptic plasticity forms and functions, *Annu. Rev. Neurosci.* **43**, 95 (2020).
- [31] A. Gidon, T. A. Zolnik, P. Fidzinski, F. Bolduan, A. Papoutsi, P. Poirazi, M. Holtkamp, I. Vida, and M. E. Larkum, Dendritic action potentials and computation in human layer 2/3 cortical neurons, *Science* **367**, 83 (2020).
- [32] N. Spruston, Pyramidal neurons: Dendritic structure and synaptic integration, *Nat. Rev. Neurosci.* **9**, 206 (2008).
- [33] H. Takahashi and J. C. Magee, Pathway interactions and synaptic plasticity in the dendritic tuft regions of CA1 pyramidal neurons, *Neuron* **62**, 102 (2009).
- [34] D. Lipshutz, Y. Bahroun, S. Golkar, A. M. Sengupta, and D. B. Chklovskii, A biologically plausible neural network for multichannel canonical correlation analysis, *Neural Comput.* **33**, 2309 (2021).
- [35] D. Lipshutz, C. Windolf, S. Golkar, and D. B. Chklovskii, A biologically plausible neural network for slow feature analysis, *Advances in Neural Information Processing Systems*, Vol. 33 (Curran Associates, Inc., 2020), pp. 14986–14996.
- [36] Y. Bahroun, D. Chklovskii, and A. Sengupta, A normative and biologically plausible algorithm for independent component analysis, *Advances in Neural Information Processing Systems*, Vol. 34 (Curran Associates, Inc., 2021), pp. 7368–7384.
- [37] S. Golkar, D. Lipshutz, T. Tesileanu, and D. B. Chklovskii, An online algorithm for contrastive principal component analysis, *IEEE International Conference on Acoustics, Speech and Signal Processing* (IEEE, Piscataway, NJ, 2023).
- [38] D. Lipshutz, C. Pehlevan, and D. B. Chklovskii, Biologically plausible single-layer networks for nonnegative independent component analysis, *Biol. Cybern.* **116**, 557 (2022).
- [39] J. P. Cunningham and Z. Ghahramani, Linear dimensionality reduction: Survey, insights, and generalizations, *J. Machine Learning Res.* **16**, 2859 (2015).

- [40] B. Ghojogh, F. Karray, and M. Crowley, Eigenvalue and generalized eigenvalue problems: Tutorial, [arXiv:1903.11240](#) (2019).
- [41] J. J. Atick and A. N. Redlich, Towards a theory of early visual processing, *Neural Comput.* **2**, 308 (1990).
- [42] S. Ganguli and H. Sompolinsky, Compressed sensing, sparsity, and dimensionality in neuronal information processing and data analysis, *Annu. Rev. Neurosci.* **35**, 485 (2012).
- [43] D. H. Hubel, *Eye, Brain, and Vision* (Scientific American Library/Scientific American Books, New York, 1995).
- [44] B. Yang, Projection approximation subspace tracking, *IEEE Trans. Signal Proc.* **43**, 95 (1995).
- [45] M. Duflo, *Random Iterative Models* (Springer Science & Business Media, New York, 2013), Vol. 34.
- [46] C.-N. Chou and M. B. Wang, Ode-inspired analysis for the biological version of Oja's rule in solving streaming PCA, in *Proceedings of Machine Learning Research*, Vol. 125 (PMLR, 2020), pp. 1339–1343.
- [47] P. Földiák, Adaptive network for optimal linear feature extraction, in *Proceedings of IEEE/INNS International Joint Conference on Neural Networks* (IEEE, Piscataway, NJ, 1989), Vol. 1, pp. 401–405.
- [48] J. Rubner and P. Tavan, A self-organizing network for principal-component analysis, *Europhys. Lett.* **10**, 693 (1989).
- [49] J. Rubner and K. Schulten, Development of feature detectors by self-organization, *Biol. Cybern.* **62**, 193 (1990).
- [50] T. K. Leen, Learning in linear feature-discovery networks, in *Adaptive Signal Processing* (International Society for Optics and Photonics, 1991), Vol. 1565, pp. 472–481.
- [51] C. Pehlevan, A. M. Sengupta, and D. B. Chklovskii, Why do similarity matching objectives lead to Hebbian/anti-Hebbian networks? *Neural Comput.* **30**, 84 (2018).
- [52] M. A. A. Cox and T. F. Cox, Multidimensional scaling, in *Handbook of Data Visualization* (Springer, New York, 2008), pp. 315–347.
- [53] O. Steward and S. A. Scoville, Cells of origin of entorhinal cortical afferents to the hippocampus and fascia dentata of the rat, *J. Comp. Neurol.* **169**, 347 (1976).
- [54] L. J. Cauller, B. Clancy, and B. W. Connors, Backward cortical projections to primary somatosensory cortex in rats extend long horizontal axons in layer I, *J. Comp. Neurol.* **390**, 297 (1998).
- [55] M. Megias, Z. S. Emri, T. F. Freund, and A. I. Gulyas, Total number and distribution of inhibitory and excitatory synapses on hippocampal CA1 pyramidal cells, *Neuroscience* **102**, 527 (2001).
- [56] L. Petreanu, T. Mao, S. M. Sternson, and K. Svoboda, The subcellular organization of neocortical excitatory connections, *Nature (London)* **457**, 1142 (2009).
- [57] J. C. Magee, Dendritic hyperpolarization-activated currents modify the integrative properties of hippocampal CA1 pyramidal neurons, *J. Neurosci.* **18**, 7613 (1998).
- [58] G. Stuart and N. Spruston, Determinants of voltage attenuation in neocortical pyramidal neuron dendrites, *J. Neurosci.* **18**, 3501 (1998).
- [59] M. T. Harnett, N.-L. Xu, J. C. Magee, and S. R. Williams, Potassium channels control the interaction between active dendritic integration compartments in layer 5 cortical pyramidal neurons, *Neuron* **79**, 516 (2013).
- [60] M. T. Harnett, J. C. Magee, and S. R. Williams, Distribution and function of HCN channels in the apical dendritic tuft of neocortical pyramidal neurons, *J. Neurosci.* **35**, 1024 (2015).
- [61] R. W. Turner, D. E. Meyers, T. L. Richardson, and J. L. Barker, The site for initiation of action potential discharge over the somatodendritic axis of rat hippocampal CA1 pyramidal neurons, *J. Neurosci.* **11**, 2270 (1991).
- [62] G. J. Stuart and B. Sakmann, Active propagation of somatic action potentials into neocortical pyramidal cell dendrites, *Nature (London)* **367**, 69 (1994).
- [63] J. Schiller, Y. Schiller, G. Stuart, and B. Sakmann, Calcium action potentials restricted to distal apical dendrites of rat neocortical pyramidal neurons, *J. Physiol.* **505**, 605 (1997).
- [64] M. E. Larkum, J. J. Zhu, and B. Sakmann, A new cellular mechanism for coupling inputs arriving at different cortical layers, *Nature (London)* **398**, 338 (1999).
- [65] N. L. Golding, N. P. Staff, and N. Spruston, Dendritic spikes as a mechanism for cooperative long-term potentiation, *Nature (London)* **418**, 326 (2002).
- [66] P. J. Sjöström and M. Häusser, A cooperative switch determines the sign of synaptic plasticity in distal dendrites of neocortical pyramidal neurons, *Neuron* **51**, 227 (2006).
- [67] K. C. Bittner, C. Grienberger, S. P. Vaidya, A. D. Milstein, J. J. Macklin, J. Suh, S. Tonegawa, and J. C. Magee, Conjunctive input processing drives feature selectivity in hippocampal CA1 neurons, *Nat. Neurosci.* **18**, 1133 (2015).
- [68] B. W. Mel, Information processing in dendritic trees, *Neural Comput.* **6**, 1031 (1994).
- [69] K. P. Körding and P. König, Learning with two sites of synaptic integration, *Netw., Comput. Neural Syst.* **11**, 25 (2000).
- [70] P. Poirazi, T. Brannon, and B. W. Mel, Pyramidal neuron as two-layer neural network, *Neuron* **37**, 989 (2003).
- [71] P. Poirazi, *Information Processing in Single Cells and Small Networks: Insights from Compartmental Models*, in AIP Conf. Proc. No. 1108 (American Institute of Physics, 2009), pp. 158–167.
- [72] E. B. Hendrickson, J. R. Edgerton, and D. Jaeger, The capabilities and limitations of conductance-based compartmental neuron models with reduced branched or unbranched morphologies and active dendrites, *J. Comput. Neurosci.* **30**, 301 (2011).
- [73] R. Urbanczik and W. Senn, Learning by the dendritic prediction of somatic spiking, *Neuron* **81**, 521 (2014).
- [74] J. Guerguiev, T. P. Lillicrap, and B. A. Richards, Towards deep learning with segregated dendrites, *Elife* **6**, e22901 (2017).
- [75] J. Sacramento, R. Ponte Costa, Y. Bengio, and W. Senn, Dendritic cortical microcircuits approximate the back-propagation algorithm, *Advances in Neural Information Processing Systems*, Vol. 31 (Curran Inc., 2018).
- [76] T. Haga and T. Fukai, Dendritic processing of spontaneous neuronal sequences for single-trial learning, *Sci. Rep.* **8**, 15166 (2018).
- [77] B. A. Richards and T. P. Lillicrap, Dendritic solutions to the credit assignment problem, *Curr. Opin. Neurobiol.* **54**, 28 (2019).

- [78] A. Tzivilaki, G. Kastellakis, and P. Poirazi, Challenging the point neuron dogma: FS basket cells as 2-stage nonlinear integrators, *Nat. Commun.* **10**, 3664 (2019).
- [79] I. S. Jones and K. P. Kording, Might a single neuron solve interesting machine learning problems through successive computations on its dendritic tree? *Neural Comput.* **33**, 1554 (2021).
- [80] A. D. Milstein, Y. Li, K. C. Bittner, C. Grienberger, I. Soltesz, J. C. Magee, and S. Romani, Bidirectional synaptic plasticity rapidly modifies hippocampal representations, *Elife* **10**, e73046 (2021).
- [81] M. E. Larkum, Are dendrites conceptually useful? *Neuroscience* **489**, 4 (2022).
- [82] F. A. Mikulasch, L. Rudelt, M. Wibrals, and V. Priesemann, Where is the error? Hierarchical predictive coding through dendritic error computation, *Trends Neurosci.* **46**, 45 (2023).
- [83] R. Makarov, M. Pagkalos, and P. Poirazi, Dendrites and efficiency: Optimizing performance and resource utilization, [arXiv:2306.07101](https://arxiv.org/abs/2306.07101) (2023).
- [84] H. Hotelling, Relations between two sets of variates, *Biometrika* **28**, 321 (1936).
- [85] R. Arora, T. V. Marinov, P. Mianjy, and N. Srebro, Stochastic approximation for canonical correlation analysis, *Advances in Neural Information Processing Systems*, Vol. 30 (Curran Inc., 2017), p. 4775.
- [86] K. Bhatia, A. Pacchiano, N. Flammarion, P. L. Bartlett, and M. I. Jordan, Gen-Oja: Simple & efficient algorithm for streaming generalized eigenvector computation, *Advances in Neural Information Processing System*, Vol. 31 (Curran Inc., 2018), p. 7016.
- [87] C. Pehlevan, X. Zhao, A. M. Sengupta, and D. Chklovskii, Neurons as canonical correlation analyzers, *Front. Comput. Neurosci.* **14**, 55 (2020).
- [88] Z. Meng, R. Chakraborty, and V. Singh, An online Riemannian PCA for stochastic canonical correlation analysis, *Advances in Neural Information Processing Systems*, Vol. 34 (Curran Inc., 2021), p. 14056.
- [89] I. Gemp, C. Chen, and B. McWilliams, The generalized eigenvalue problem as a Nash equilibrium, *The Eleventh International Conference on Learning Representation (IMLR, 2023)*.
- [90] S. Boyd and L. Vandenberghe, *Convex Optimization* (Cambridge University Press, Cambridge, 2004).
- [91] S. Golkar, D. Lipshutz, Y. Bahroun, A. Sengupta, and D. Chklovskii, A simple normative network approximates local non-Hebbian learning in the cortex, *Advances in Neural Information Processing Systems*, Vol. 33 (Curran Inc., 2020), p. 7283.
- [92] S. Golkar, D. Lipshutz, Y. Bahroun, A. M. Sengupta, and D. B. Chklovskii, A biologically plausible neural network for local supervision in cortical microcircuits, [arXiv:2011.15031](https://arxiv.org/abs/2011.15031) (2020).
- [93] P. Földiák, Learning invariance from transformation sequences, *Neural Comput.* **3**, 194 (1991).
- [94] G. Mitchison, Removing time variation with the anti-Hebbian differential synapse, *Neural Comput.* **3**, 312 (1991).
- [95] L. Wiskott and T. J. Sejnowski, Slow feature analysis: Unsupervised learning of invariances, *Neural Comput.* **14**, 715 (2002).
- [96] P. Berkes and L. Wiskott, Slow feature analysis yields a rich repertoire of complex cell properties, *J. Vision* **5**, 9 (2005).
- [97] M. Franzius, H. Sprekeler, and L. Wiskott, Slowness and sparseness lead to place, head-direction, and spatial-view cells, *PLoS Comput. Biol.* **3**, e166 (2007).
- [98] H. B. Barlow, Unsupervised learning, *Neural Comput.* **1**, 295 (1989).
- [99] J. J. Atick, Could information theory provide an ecological theory of sensory processing? *Netw. Comput. Neural Syst.* **3**, 213 (1992).
- [100] A. Bell and T. J. Sejnowski, Edges are the “independent components” of natural scenes, *Advances in Neural Information Processing Systems*, Vol. 9 (MIT Press, 1996), p. 831.
- [101] A. J. Bell and T. J. Sejnowski, The “independent components” of natural scenes are edge filters, *Vision Res.* **37**, 3327 (1997).
- [102] M. S. Lewicki, Efficient coding of natural sounds, *Nat. Neurosci.* **5**, 356 (2002).
- [103] J.-F. Cardoso, Source separation using higher order moments, in *International Conference on Acoustics, Speech, and Signal Processing* (IEEE, Piscataway, NJ, 1989), pp. 2109–2112.
- [104] A. Abid, M. J. Zhang, V. K. Bagaria, and J. Zou, Exploring patterns enriched in a dataset with contrastive principal component analysis, *Nat. Commun.* **9**, 2134 (2018).
- [105] G. J. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition* (Wiley, Hoboken, NJ, 2005).
- [106] A. de Cheveigné and L. C. Parra, Joint decorrelation, a versatile tool for multichannel data analysis, *NeuroImage* **98**, 487 (2014).
- [107] C. Pehlevan and D. B. Chklovskii, A Hebbian/anti-Hebbian network derived from online non-negative matrix factorization can cluster and discover sparse features, in *2014 48th Asilomar Conference on Signals, Systems and Computers* (IEEE, Piscataway, NJ, 2014), pp. 769–775.
- [108] Y. Bahroun and A. Sottoggio, Online representation learning with single and multi-layer Hebbian networks for image classification, in *Artificial Neural Networks and Machine Learning ICANN 2017*, Lecture Notes in Computer Science, Vol. 10613 (Springer, Cham, 2017), pp. 354–363.
- [109] A. Sengupta, M. Tepper, C. Pehlevan, A. Genkin, and D. Chklovskii, Manifold-tiling localized receptive fields are optimal in similarity-preserving neural networks, *Advances in Neural Information Processing Systems*, Vol. 31 (Curran Inc., 2018).
- [110] S. Qin, N. Mudur, and C. Pehlevan, Contrastive similarity matching for supervised learning, *Neural Comput.* **33**, 1300 (2021).
- [111] C. Pehlevan, S. Mohan, and D. B. Chklovskii, Blind non-negative source separation using biological neural networks, *Neural Comput.* **29**, 2925 (2017).
- [112] S. Golkar, T. Teşileanu, Y. Bahroun, A. Sengupta, and D. B. Chklovskii, Constrained predictive coding as a biologically plausible model of the cortical hierarchy, *Advances in Neural Information Processing Systems*, Vol. 35 (Curran Inc., 2022), pp. 14155–14169.
- [113] J. Friedrich, S. Golkar, S. Farshahi, A. Genkin, A. Sengupta, and D. B. Chklovskii, Neural optimal feedback control with local learning rules, *Advances in Neural*

- Information Processing Systems*, Vol. 34 (Curran Inc., 2021), pp. 16358–16370.
- [114] D. Lipshutz, A. Kashalika, S. Farashahi, and D. B. Chklovskii, A linear discriminant analysis model of imbalanced associative learning in the mushroom body compartment, *PLoS Comput. Biol.* **19**, e1010864 (2023).
- [115] L. R. Duong, D. Lipshutz, D. Heeger, D. B. Chklovskii, and E. P. Simoncelli, Adaptive whitening in neural populations with gain-modulating interneurons, *Proceedings of the 40th International Conference on Machine Learning*, Proceedings of Machine Learning Research (PMLR, 2023), pp. 8902–8921.