Journal of Theoretical Biology ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Hebbian errors in learning: An analysis using the Oja model

Anca Rădulescu [a,b,*], Kingsley Cox [b,c], Paul Adams [b,c]

[a] University of Colorado, UCB 526 Boulder, CO 80309-0526, USA
[b] Stony Brook University, Stony Brook, NY 11794, USA
[c] Kalypso Institute, Stony Brook, USA

ARTICLE INFO

Article history:
Received 26 March 2008
Received in revised form
27 November 2008
Accepted 23 January 2009

Keywords:
Synaptic plasticity
Information transfer
Neural computation
Learning algorithm
Dynamical system

ABSTRACT

Background: Recent work on long term potentiation in brain slices shows that Hebb's rule is not completely synapse-specific, probably due to intersynapse diffusion of calcium or other factors. We previously suggested that such errors in Hebbian learning might be analogous to mutations in evolution.

Methods and findings: We examine this proposal quantitatively, extending the classical Oja unsupervised model of learning by a single linear neuron to include Hebbian inspecificity. We introduce an error matrix **E**, which expresses possible crosstalk between updating at different connections. When there is no inspecificity, this gives the classical result of convergence to the first principal component of the input distribution (PC1). We show the modified algorithm converges to the leading eigenvector of the matrix **EC**, where **C** is the input covariance matrix. In the most biologically plausible case when there are no intrinsically privileged connections, **E** has diagonal elements Q and off-diagonal elements $(1-Q)/(n-1)$, where Q, the quality, is expected to decrease with the number of inputs n and with a synaptic parameter b that reflects synapse density, calcium diffusion, etc. We study the dependence of the learning accuracy on b, n and the amount of input activity or correlation (analytically and computationally). We find that accuracy increases (learning becomes gradually less useful) with increases in b, particularly for intermediate (i.e., biologically realistic) correlation strength, although some useful learning always occurs up to the trivial limit $Q = 1/n$.

Conclusions and significance: We discuss the relation of our results to Hebbian unsupervised learning in the brain. When the mechanism lacks specificity, the network fails to learn the expected, and typically most useful, result, especially when the input correlation is weak. Hebbian crosstalk would reflect the very high density of synapses along dendrites, and inevitably degrades learning.

1. Introduction

Various brain structures, such as the neocortex, are believed to use unsupervised synaptic learning to form neural representations that capture and exploit statistical regularities of an animal's world. Most neural models of unsupervised learning use some form of Hebb rule to update synaptic connections. Typically, this rule is implemented by updating a connection according to the product of the input and output firing rates. Other forms of the update rule are sometimes used, but they are still typically local and activity dependent, and often Hebbian in the sense that they depend on both input and output activity. Biological networks may also use spike-timing dependent rules, but these are also Hebbian in the sense that they depend on the relative timing of pre- and postsynaptic spiking. The key element in Hebbian learning is that the update should depend on the extent to which the input appears to "take part in" firing the output (Hebb, 1949; we added "appears" to emphasize that a single neural connection knows nothing about actual causation, and merely responds to a statistical coincidence of pre- and postsynaptic spikes).

One of us has previously proposed, in this journal (Adams, 1998), that such Hebbian errors might be analogous to genetic mutations. Specifically, we are interested in the possibility that if the Hebb rule is not completely local (in the sense there might be some, possibly very weak, dependence of the local update on activity at other connections) unsupervised learning might fail catastrophically, not only preventing new learning, but wiping out previous learning. We have proposed that the basic task of the neocortex is to avoid such hypothetical learning catastrophes (Adams and Cox, 2002a, b, 2006; Cox and Adams, 2000). In this paper we modify a classical model of unsupervised learning, the Oja single neuron principal component analyzer (Oja, 1982), to include Hebbian inaccuracy. By "inspecificity", or "inaccuracy", we mean that part of the local update calculated using a Hebb rule (for example, proportional to the product of input and output

* Corresponding author at: University of Colorado, UCB 526 Boulder, CO 80309-0526, USA. Tel.: +1 303 786 1616; fax: +1 303 492 4066.
E-mail address: radulesc@colorado.edu (A. Rădulescu).

0022-5193/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2009.01.036

firing rates) is assigned to connections other than the one at which the product was calculated. We also refer to this postulated nonlocality as "leakage", "crosstalk" or simply "error". Some other papers on this topic have appeared (Adams and Cox, 2002a; Botelho and Jamison, 2002, 2004), and we will try to clarify the relationship between these various studies. We conclude that while the modified Oja model, and perhaps others that are only sensitive to second-order statistics, do not show a true error catastrophe at finite network size, their behaviour gives important clues to understanding the difficulties that brains might encounter in learning higher-order statistics.

Recent experimental work has shown that long term potentiation (LTP), a biological manifestation of the Hebb rule, is indeed not completely synapse specific (Engert and Bonhoeffer, 1997; Schuman and Madison, 1994; Bi, 2002). For example, Engert and Bonhoeffer have shown that LTP induced at a local set of connections on a CA1 pyramidal cell "spills over" to induce LTP at a nearby set of inactive connections on the same cell. In earlier work using less refined methods, it had been concluded that LTP was synapse-specific (Andersen et al., 1977; Levy and Steward, 1979). Even in the Engert–Bonhoeffer experiments (Engert and Bonhoeffer, 1997), it is likely that, because the "pairing" method used to induce LTP was rather crude, the inspecificity was far greater than would ever be actually seen in an awake brain. More recent work has shown that at least one type of Hebbian inspecificity, induced by theta burst stimulation of retinotectal connections, reflects dendritic spread of calcium (Tao et al., 2001). Even more recent LTP experiments at single synapses (Tao et al., 2001) have shown that, while LTP is only expressed at the synapse where it is induced (Matsuzaki et al., 2004), the threshold for LTP induced at neighboring synapses is reduced (Harvey and Svoboda, 2007). Thus, some degree of Hebbian inspecificity is probably inevitable, and its effects on learning need to be evaluated.

2. Overview

Here we briefly review the classical Oja model (Oja, 1982; Oja and Karhunen, 1985; Diamantaras and Kung, 1996; Hertz et al., 1991) and define terms as background to the new analysis. The model network consists of a single output neuron receiving n signals $x_1, x_2, \ldots, x_n$ from a set of n input neurons via connections of corresponding strengths $\omega_1, \ldots, \omega_n$ (see Fig. 1). We assume throughout that the inputs have zero-mean (see Discussion).

The resulting output y is defined as the weighted sum of the inputs:

$$y = \sum_{i=1}^{n} x_i \omega_i \tag{2.1}$$

The input column vector $\mathbf{x} = (x_1 \ldots x_n)^T$ is randomly drawn from a probability distribution $\mathscr{P}(\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^n$ (where T denotes transposition of vectors).

In accordance to Hebb's postulate of learning, a synaptic weight ω_i will strengthen proportionally with the product of x_i and y:

$$\omega_i(t+1) = \omega_i(t) + \gamma y(t) x_i(t) \tag{2.2}$$

Here γ is a time independent learning rate and the argument t represents the dependence on time (or on the input draw). The relation between this formulation and neural processes such as LTP is considered in the Discussion.

Oja (1982) modified this by normalizing the weight vector $\boldsymbol{\omega}$ with respect to the Euclidean metric on $\mathbb{R}^n$:

$$\omega_i(t+1) = \frac{\omega_i(t) + \gamma y(t) x_i(t)}{\|\boldsymbol{\omega}(t) + \gamma y(t)\mathbf{x}(t)\|} \tag{2.3}$$

Expanding in Taylor series with respect to γ, making $\|\boldsymbol{\omega}\| = 1$ and ignoring the $\mathscr{O}(\gamma^2)$ term for γ sufficiently small, the result is

$$\boldsymbol{\omega}(t+1) = \boldsymbol{\omega}(t) + \gamma y(t)[\mathbf{x}(t) - y(t)\boldsymbol{\omega}(t)] \tag{2.4}$$

Henceforth, we omit the variable t whenever there is no ambiguity. The equation can be then rewritten as:

$$\boldsymbol{\omega}(t+1) = \boldsymbol{\omega} + \gamma[\mathbf{x}\mathbf{x}^T\boldsymbol{\omega} - (\boldsymbol{\omega}^T\mathbf{x}\mathbf{x}^T\boldsymbol{\omega})\boldsymbol{\omega}] \tag{2.5}$$

Consider the covariance matrix of the distribution $\mathscr{P}(\mathbf{x})$, defined by $C = \langle \mathbf{x}\mathbf{x}^T \rangle = \langle \mathbf{x}(t)\mathbf{x}^T(t) \rangle$ (where $\langle\,\rangle$ stands for expectation). Clearly $\mathbf{C}$ is symmetric and semipositive definite. With the following additional assumptions:

- the learning process is slow enough for $\boldsymbol{\omega}$ to be treated as stationary,
- $\mathbf{x}(t)$ and $\boldsymbol{\omega}(t)$ are statistically independent.

One can take conditional expectation over $\mathscr{P}(\mathbf{x})$ and rewrite the learning rule as:

$$\langle \boldsymbol{\omega}(t+1) | \boldsymbol{\omega}(t) \rangle = \mathbf{w} + \gamma[\mathbf{C}\mathbf{w} - (\mathbf{w}^T\mathbf{C}\mathbf{w})] \tag{2.6}$$

Oja concluded that, if $\boldsymbol{\omega}(t)$ converges as $t \to \infty$, the limit is expected to be one of the two opposite normalized eigenvectors corresponding to the maximal eigenvalue of $\mathbf{C}$ (i.e., the "principal component" of the matrix $\mathbf{C}$—Oja, 1982; Oja and Karhunen, 1985). If the input elements are Gaussian, the output of the Oja neuron provides the statistically optimal representation of the current input vector.

Three main types of synaptic learning error could occur. First ("Type 0"), one could add uncorrelated noise to the input elements, for example reflecting fluctuations in transmitter release, which would simply add to the variances along the diagonal of $\mathbf{C}$. Second ("Type 1"), the updates themselves may be imprecise (for example because of spine head calcium fluctuations), with these fluctuations occurring independently at different synapses. Third ("Type 2"), the updates could be inaccurate, in the sense of depending on updates occurring at other synapses (e.g., because of intersynapse calcium diffusion). This is the case we study here. To introduce such inspecificity into the learning equation, we assume that, on average, only a fraction Q of the intended update reaches the appropriate connection, the remaining fraction $1 - Q$ being distributed amongst the other connections according to a defined and biologically plausible rule. The actual update at a given connection thus includes contributions from erroneous or inaccurate updates from other connections. The erroneous updating process is formally described by a (possibly time-dependent, but input-independent) error matrix $\mathscr{E} = \mathscr{E}(t)$, whose elements reflect at each time step t the fractional contribution that the activity across weight ω_i makes to the update of ω_j.

If one wanted to ensure that the weight vector retains the same norm as the error-free rule, we would introduce $\mathscr{E}$ into Eq. (2.2), obtaining as a first order approximation for the single weight one pattern rule[1]:

$$\omega_i(t+1) = \omega_i + \gamma y([\mathscr{E}\mathbf{x}]_i - \boldsymbol{\omega}^T\mathscr{E}\mathbf{x}\omega_i) \tag{2.7}$$

However, while the averaged form of this rule has the same fixed points as the rule we actually used (see Eq. (2.8)), we could not prove their stability. Furthermore, the rule would imply biologically that the normalizing component "knew" the pattern-to-pattern form of $\mathscr{E}$, which is highly implausible. More generally we could assume that the Hebbian and normalizing steps would have different error matrices, reflecting their different physical

[1] Here the notation $[\mathscr{E}\mathbf{x}]_i$ stands for the i-th component of the vector $[\mathscr{E}\mathbf{x}]$.

Please cite this article as: Rădulescu, A., et al., Hebbian errors in learning: An analysis using the Oja model. J. Theor. Biol. (2009), doi:10.1016/j.jtbi.2009.01.036

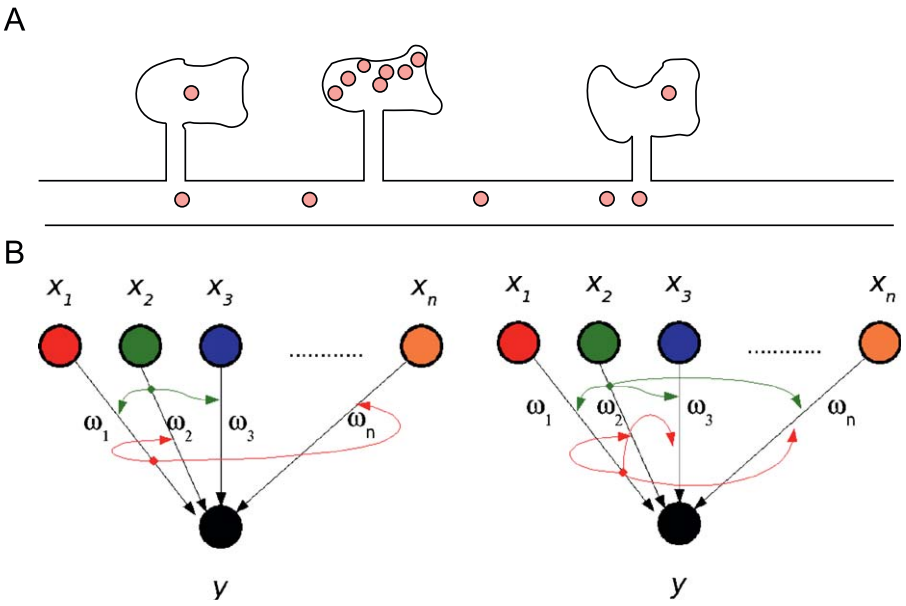

Fig. 1. (A) Three spines on a short dendritic segment are shown. If Hebbian adjustment occurs at the middle synapse, a factor (red dots), such as calcium, diffuses to nearby synapses and affects Hebbian adjustment there. (B) Input neurons (activities x_i) converge on an output neuron (output y) via weights ω_i. Coincident activity at the synapses comprising a weight (e.g. ω_1 or ω_2) leads to modification of that weight and of other weights. The left diagram shows the case where only the immediate neighboring connections (each made up of one synapse) are affected. The right diagram shows the case where all connections are equal neighbors (either because each has many synapses dispersed randomly over the dendrite, or because synapses move around; see Figure S1). The curved red arrows from ω_1 to ω_n shows that periodic boundary conditions are assumed (i.e., ω_1 affects ω_2 and ω_n equally). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

implementation. We assumed for simplicity that the normalizing step is error free, so the rule becomes:

$$\omega_i(t+1) = \omega_i + \gamma y([\mathscr{E}\mathbf{x}]_i - y\omega_i) \tag{2.8}$$

Taking conditional expectation of both sides and rewriting the equation in matrix form leads to (as a "meanfield" approximation, assuming that the errors are independent from the inputs and the weights)

$$\langle \boldsymbol{\omega}(t+1)|\boldsymbol{\omega}(t)\rangle = \mathbf{w} + \gamma[\mathbf{ECw} - (\mathbf{w}^T\mathbf{Cw})\mathbf{w}] \tag{2.9}$$

where we defined $\mathbf{w} = \langle \boldsymbol{\omega} \rangle$ and $\mathbf{E} = \langle \mathscr{E} \rangle$. $\mathbf{E}$ is then a symmetric circulant matrix; in the zero error case of $Q = 1$, $\mathbf{E}$ would become the identity matrix.

3. Methods

As a prelude to analyzing the dynamics of inspecific learning, we revisit the Oja original model with zero error, and the methods used to establish its asymptotic behavior (Botelho and Jamison, 2004; Oja, 1982; Oja and Karhunen, 1985; Hertz et al., 1991). In other words: for a size $n \in \mathbb{N}$, $n \geqslant 2$, we want to know whether or not a vector $\mathbf{w} \in \mathbb{R}^n$ stabilizes under iterations of the function:

$$f\colon \mathbb{R}^n \to \mathbb{R}^n, \quad f(\mathbf{w}) = \mathbf{w} + \gamma[\mathbf{Cw} - (\mathbf{w}^T\mathbf{Cw})\mathbf{w}] \tag{3.1}$$

A nonzero vector $\mathbf{w} \in \mathbb{R}^n$ is a fixed point for f if

$$f(\mathbf{w}) = \mathbf{w} + \gamma[\mathbf{Cw} - (\mathbf{w}^T\mathbf{Cw})\mathbf{w}] = \mathbf{w} \;\Leftrightarrow\; \mathbf{Cw} = (\mathbf{w}^T\mathbf{Cw})\mathbf{w} \tag{3.2}$$

An equivalent set of conditions is:

$$\begin{cases} \mathbf{Cw} = \lambda_{\mathbf{w}}\mathbf{w} \\ \lambda_{\mathbf{w}} = \mathbf{w}^T\mathbf{Cw} \end{cases} \Leftrightarrow \begin{cases} \mathbf{Cw} = \lambda_{\mathbf{w}}\mathbf{w} \\ \lambda_{\mathbf{w}} = \lambda_{\mathbf{w}}\mathbf{w}^T\mathbf{w} \end{cases} \tag{3.3}$$

These conditions translate as: "$\mathbf{w}$ is an eigenvector of $\mathbf{C}$". In case $\mathbf{C}$ is invertible (i.e. all its eigenvalues are nonzero), $\mathbf{w}$ is a unit eigenvector of $\mathbf{C}$ with the Euclidean norm.

Consider an orthonormal basis $\mathscr{B}$ of eigenvectors of $\mathbf{C}$ (with respect to the Euclidean norm on $\mathbb{R}^n$). An eigenvector $\mathbf{w} \in \mathscr{B}$ with eigenvalue $\lambda_{\mathbf{w}}$ is a hyperbolic attractor for f if all eigenvalues of the $n \times n$ Jacobian matrix $(Df_{\mathbf{w}})_{ij} = ((\partial f_i/\partial w_j)(\mathbf{w}))$ are less than one in absolute value.

We calculate the Jacobian matrix $Df_{\mathbf{w}}$, for a fixed vector $\mathbf{w} \in \mathscr{B}$:

Lemma 3.1. $Df_{\mathbf{w}} = \mathbf{I} + \gamma[\mathbf{C} - 2\mathbf{w}(\mathbf{Cw})^T - (\mathbf{w}^T\mathbf{Cw})\mathbf{I}]$

Proof. Call $g(\mathbf{w}) = (\mathbf{w}^T\mathbf{Cw})\mathbf{w}$, so $f(\mathbf{w}) = \mathbf{w} + \gamma(\mathbf{Cw} - g(\mathbf{w}))$

$g_i(\mathbf{w}) = (\mathbf{w}^T\mathbf{Cw})w_i$

If $i \neq j$:

$$\frac{\partial g_i}{\partial w_j}(\mathbf{w}) = \frac{\partial}{\partial w_j}\left(\sum_{k,l} C_{kl}w_k w_l\right)w_i = 2\left(\sum_k C_{kj}w_k\right)w_i = 2[\mathbf{Cw}]_j w_i$$

If $i = j$:

$$\frac{\partial g_i}{\partial w_i}(\mathbf{w}) = \frac{\partial}{\partial w_i}\left(\sum_{k,l} C_{kl}w_k w_l\right)w_i + \sum_{k,l} C_{kl}w_k w_l = 2\left(\sum_k C_{ki}w_k\right)w_i + \mathbf{w}^T\mathbf{Cw} = 2[\mathbf{Cw}]_i w_i + \mathbf{w}^T\mathbf{Cw}$$

So

$$Dg_{\mathbf{w}} = 2\mathbf{w}(\mathbf{Cw})^T + (\mathbf{w}^T\mathbf{Cw})\mathbf{I} \qquad \square \tag{3.4}$$

Pick any $\mathbf{v} \in \mathscr{B}, \mathbf{v} \neq \mathbf{w}$. Call $\lambda_{\mathbf{w}}$ and $\lambda_{\mathbf{v}}$ their corresponding eigenvalues.

$$\begin{aligned} Df_{\mathbf{w}}(\mathbf{v}) &= v + \gamma[\mathbf{Cv} - 2\mathbf{w}(\mathbf{Cw})^T\mathbf{v} - (\mathbf{w}^T\mathbf{Cw})\mathbf{v}] \\ &= \mathbf{v} + \gamma[\mathbf{Cv} - 2\mathbf{w}\mathbf{w}^T\mathbf{Cv} - (\mathbf{w}^T\mathbf{Cw})\mathbf{w}] \\ &= v + \gamma[\lambda_{\mathbf{v}}\mathbf{v} - 2\mathbf{w}\mathbf{w}^T\lambda_{\mathbf{v}}\mathbf{v} - \lambda_{\mathbf{w}}\mathbf{w}] = (1 - \gamma[\lambda_{\mathbf{w}} - \lambda_{\mathbf{v}}])\mathbf{v} \end{aligned} \tag{3.5}$$

$$\begin{aligned} Df_{\mathbf{w}}(\mathbf{w}) &= \mathbf{w} + \gamma[\mathbf{Cw} - 2\mathbf{w}(\mathbf{Cw})^T\mathbf{w} - (\mathbf{w}^T\mathbf{Cw})] \\ &= \mathbf{w} + \gamma[\lambda_{\mathbf{w}}\mathbf{w} - 2\mathbf{w}\mathbf{w}^T\lambda_{\mathbf{w}}\mathbf{w} - \lambda_{\mathbf{w}}\mathbf{w}] \\ &= \mathbf{w} + \gamma[-2\lambda_{\mathbf{w}}\|\mathbf{w}\|\mathbf{w}] = [1 - 2\gamma\lambda_{\mathbf{w}}]\mathbf{w} \end{aligned} \tag{3.6}$$

So $\mathscr{B}$ is also a basis of eigenvectors for $Df_{\mathbf{w}}$. The corresponding eigenvalues are $1 - 2\gamma\lambda_{\mathbf{w}}$ (for the eigenvector $\mathbf{w}$) and $1 - \gamma(\lambda_{\mathbf{w}} - \lambda_{\mathbf{v}})$ (for any other eigenvector $\mathbf{v} \in \mathscr{B}$, $\mathbf{v} \neq \mathbf{w}$). Therefore, a set of

Please cite this article as: Rădulescu, A., et al., Hebbian errors in learning: An analysis using the Oja model. J. Theor. Biol. (2009), doi:10.1016/j.jtbi.2009.01.036

equivalent conditions for $\mathbf{w}$ to be a hyperbolic attractor for f is

$$|1-\gamma(\lambda_{\mathbf{w}}-\lambda_{\mathbf{v}})|<1 \quad \text{for all } \mathbf{v}\in\mathscr{B},\ \mathbf{v}\neq\mathbf{w} \tag{3.7}$$

$$|1-2\gamma\lambda_{\mathbf{w}}|<1 \tag{3.8}$$

In other words, $\mathbf{w}$ is a hyperbolic fixed point of f if and only if:

(i) $\lambda_{\mathbf{w}}>\lambda_{\mathbf{v}}$ for all $\mathbf{v}\neq\mathbf{w}$ (i.e. $\lambda_{\mathbf{w}}$ is the maximal eigenvalue)
(ii) $\gamma<1/\lambda_{\mathbf{w}}$ (in particular $\gamma<2/(\lambda_{\mathbf{w}}-\lambda_{\mathbf{v}})$, for all $\mathbf{v}\neq\mathbf{w}$).

These conditions are always satisfied provided: (i) $\mathbf{C}$ has a maximal eigenvalue of multiplicity one and (ii) γ is small enough ($\gamma<1/\lambda_{\mathbf{w}}$).

In conclusion, under conditions (i) and (ii), the network learns the first principal component (PC1) of the distribution $\mathscr{P}(\mathbf{x})$. The learning of the principal component requires a relationship between the rate of learning γ and the input distribution $\mathscr{P}(\mathbf{x})$: if the maximal eigenvalue of the correlation matrix $\mathbf{C}$ is large (i.e. if the variance of the input patterns' projections on PC1 is high), the network has to learn slowly in order to achieve convergence. Moreover, the convergence time along each eigendirection is given by the inverse of the magnitude of the corresponding eigenvalue of $Df_{\mathbf{w}}$ (see the simulations in Fig. 2; see also Wyatt and Elfadel, 1995).

To formalize learning inspecificity we introduced an error matrix $\mathbf{E}\in\mathscr{M}_n(\mathbb{R})$ that has positive entries, is symmetric and equal to the identity matrix $\mathbf{I}\in\mathscr{M}_n(\mathbb{R})$ when the error is zero. We studied the asymptotic behavior of the new system, using a similar approach. As shown before (Eq. (2.9)), the inspecific learning iteration function becomes:

$$f^{\mathbf{E}}(\mathbf{w})=\mathbf{w}+\gamma[\mathbf{ECw}-(\mathbf{w}^T\mathbf{Cw})\mathbf{w}]$$

Here also, $\mathbf{w}\neq\mathbf{0}$ is a fixed point of $f^{\mathbf{E}}$ if and only if it is an eigenvector of $\mathbf{EC}$ with eigenvalue $\lambda_{\mathbf{w}}=\mathbf{w}^T\mathbf{Cw}$. Furthermore, $\mathbf{w}$ is a hyperbolic attractor of $f^{\mathbf{E}}$ if and only if $\lambda_{\mathbf{w}}$ is the principal eigenvalue of $\mathbf{EC}$ and $\gamma<1/\lambda_{\mathbf{w}}$. (see the Supplementary Material for proofs).

The error-free rule maximizes the variance of the output neuron $\lambda_{\mathbf{w}}$ and therefore, with Gaussian inputs, also maximizes the mutual information (MI) between inputs and outputs (see Discussion). Although the erroneous rule no longer maximizes the output variance, it tolerates a faster learning rate. Conversely, at a fixed γ, learning is slowed by error.

3.1. The error matrix

One way in which an incorrect strengthening of a silent synapse can occur is by diffusion of a messenger such as calcium from one spine head to another, as illustrated in Fig. 1A.

If we assume that the output neuron is connected (at least potentially) to all the input neurons (Stepanyants et al., 2002, 2008) then the amount of error depends on the number of synapses each input neuron makes with the output neuron (relative to the dendritic length L) as well as factors such as the space constant for dendritic calcium diffusion λ_c (Zador and Koch, 1994), the Hill coefficient for calcium action h (DeKoninck and Schulman, 1998; Lisman, 1989), and the amount of head/shaft/head calcium attenuation a. If an intracellular factor other than calcium is responsible for crosstalk (Harvey and Svoboda, 2007), analogous parameters would still apply. We can define a per "synapse error factor" $b\in[0,1]$.

$$b\sim\frac{a^h\lambda_c}{L} \tag{3.9}$$

or equivalently a "synaptic quality" $q\in[0,1]$, $q=1-b$ (see Supplementary Material, Appendix 1 for definitions and details).

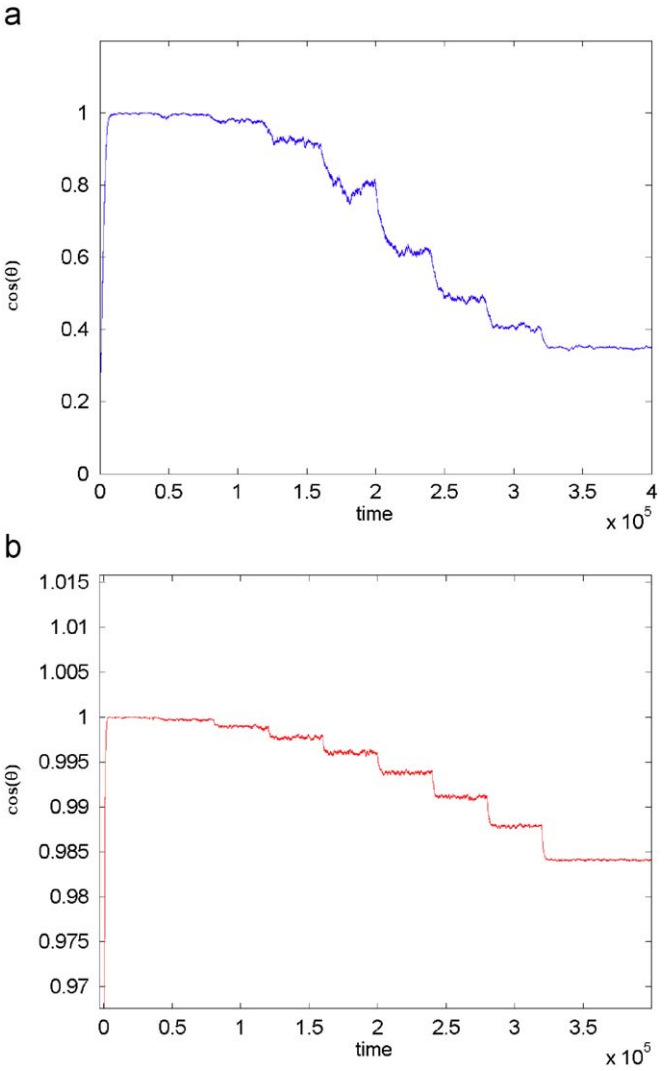

Fig. 2. Effect of errors on performance in the Oja network for $n=10$ neurons. The plots represent the cosine of the angle θ between the weight vector and the principal component, at each time-step of the updating process. Upper plot: The inputs are uncorrelated Gaussian vectors with one of the sources at a variance of 2 compared to the others which are all 1. The learning rate is fixed as $\gamma=5\times10^{-4}$. Lower plot: Correlated inputs were generated by mixing independent Gaussian sources with equal variance with a random mixing matrix, with elements distributed uniformly between 0 and 1. The learning rate for correlated inputs is $\gamma=25\times10^{-6}$. In both cases, the total error $\varepsilon=1-Q=(n-1)\varepsilon$ was initially set to zero and the weight vector converged very quickly to the first PC ($\cos(\theta)=1$) and remained there with minor fluctuations. The error was then increased from zero to 0.8 in steps of 0.1 each 4×10^4 epochs, producing approximately stepwise decreases in performance. The equilibration time increased as error increased; the step heights and the associated fluctuation increased and then decreased. Note that in the correlated case error produces only small decreases in performance, since the principal component already points approximately in the direction $(1,1,\ldots,1)$.

This formula says that the per synapse error b is proportional to two factors: the ratio of the length constant for calcium spread λ_c to the dendritic length L and the effective calcium coupling constant a^h between two adjoining spines. It assumes that as extra inputs are added, the dendritic length remains the same.

The probability Q of the correct synapse being strengthened depends on b and on the network size n. In Appendix 2 in the Supplementary Material we analyze a plausible model and we develop two approximations for $Q=Q(n,b)$:

- the *continuous model*, where weights adjust continuously and $Q=1/(nb+1)$,

- the *discrete model*, where weights adjust discretely and $Q = (1-b)^n$.

3.2. Error spread

We now consider different possibilities for the way that the part of the Hebbian update $x_i y$ could spread to different connections, presumably as a result of intracellular diffusion of messengers such as calcium. In general this will reflect the particular anatomical relationships between synapses, expressed by $\mathscr{E}$, which could change as learning proceeds. We examine two extreme cases. First, each connection is made of a single fixed synapse (e.g., a parallel fiber-Purkinje cell connection—Llinas and Walton, 1998). In the second case, all connections are equivalent ("tabula rasa"—Le Be and Markram, 2006; Elman et al., 1996).

1. The "*nearest neighbor*" model: Each connection consists of a single fixed synapse, and calcium only spreads to two nearest neighbor synapses. **E** then has diagonal elements Q and off diagonal elements $(1-Q)/2$.

$$\mathbf{E} = \begin{pmatrix} Q & \varepsilon & 0 & \cdot & \cdot & \varepsilon \\ \varepsilon & Q & \varepsilon & 0 & \cdot & 0 \\ 0 & \varepsilon & Q & \varepsilon & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & \cdot & \varepsilon & Q & \varepsilon \\ \varepsilon & 0 & \cdot & \cdot & \varepsilon & Q \end{pmatrix} \tag{3.10}$$

The appearance of ε in the top right and bottom right corners reflects periodic boundary conditions. We can define a "trivial" error (see next paragraph) $\varepsilon = \frac{1}{3}$ for which Hebbian adjustments lacks specificity, which is marked in Fig. 3a as a red dot on each curve.

2. The "*error-onto-all*" model: All connections are equally "distant" from each other, so that there are no privileged connections. All offdiagonal elements of **E** are then equal to $(1-Q)/(n-1)$.

$$\mathbf{E} = \begin{pmatrix} Q & \varepsilon & \varepsilon & \cdot & \cdot & \varepsilon \\ \varepsilon & Q & \varepsilon & \varepsilon & \cdot & \varepsilon \\ \varepsilon & \varepsilon & Q & \varepsilon & \cdot & \varepsilon \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \varepsilon & \cdot & \cdot & \varepsilon & Q & \varepsilon \\ \varepsilon & \varepsilon & \cdot & \cdot & \varepsilon & Q \end{pmatrix} \tag{3.11}$$

It is important to notice that the error matrix in this case becomes singular when $Q = \varepsilon$, i.e. when the update leak to each erroneous connection is as large as the update at the right connection. We call this value the "trivial error value" $\varepsilon_0(n)$, which corresponds to $b_0(n) = 1 - q_0(n) = 1 - 1/\sqrt[n]{n}$ in the discrete model, and to $b_0(n) = 1 - q_0(n) = 1/n$ in the continuous model. For all biological purposes, we need only consider errors smaller than the trivial value.

This arrangement could arise in two nonexclusive different ways (see Supplementary Material, Appendix 2 for details):

(a) Each connection is composed of a very large number α of fixed synapses, such that all possible configurations of synapses occur.
(b) Synapses do not have fixed locations, but appear and disappear randomly at all possible locations (i.e. "touch-points"—Le Be and Markram, 2006; Stepanyants et al., 2002) where axons approach the dendrite close enough that a new spine can create a synapse. In this case, assuming the dendrite and axonal geometry are fixed (Holtmaat et al., 2005; Grutzendler et al., 2002; Lendvai et al., 2000), "potential" synapses (Stepanyants et al., 2002) are composed of two shifting subsets: anatomically existing synapses and a reservoir of "incipient" synapses where spines could form (see Supplementary Material, Appendix 2). In order to maintain constant weights (in the absence of learning), each disappearing synapse would have to be replaced by another synapse of equal strength and, possibly, connectivity. This could be done most simply if only zero-strength ("silent") synapses appear and disappear (since then connectivity would not have to be conserved). There is evidence this is the case (Alvarez and Sabatini, 2007).

If synapses appear and disappear, one has the problem that if all synapses are equally plastic (have the same learning rate γ),

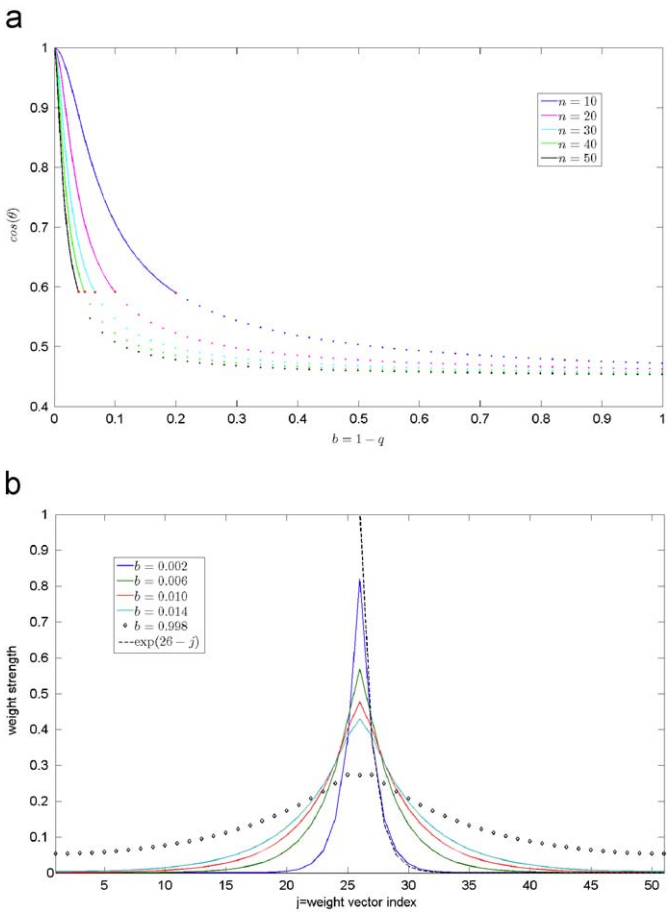

Fig. 3. Dependence of $\cos(\theta)$ on the error factor b in the case of uncorrelated inputs with $\lambda = 2$ for the continuous error, nearest neighbour model. Each curve corresponds to a different n, as shown in the legends. Upper plot: Continuous error, nearest neighbor model. Note that for increasing n values, the value of the total error ε increases at any fixed per synapse error b, reducing performance ($\cos(\theta)$). The curves are shown as solid lines up to the trivial error value where $Q = \varepsilon = \frac{1}{3}$ (i.e., $b = 2/n$) at which learning is inspecific (red dots); beyond this point the curves are unbiological and are shown dotted. Lower plot: The distribution of weights of the asymptotically stable weight vector (the principal eigenvector of **EC**), for fixed network size $n = 51$, and fixed variance $\lambda = 1.1$, but different values of the per synapse error b. The lower the quality, the more similar the weights become. All the b values shown are less than trivial, except for the curve marked with diamonds, where almost all the updates are transferred to neighbors. With the exception of the weight on the high variance neuron labeled #26, the weights decay approximately exponentially as a function of distance from the high variance neuron. This is illustrated by the black dashed curve, which is a shifted exponential with space constant of one unit (neuron). The space constant calculated from Equation 7 in Adams and Cox (2002a) for the corresponding values of b, n and λ is 0.7 neurons. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Please cite this article as: Rădulescu, A., et al., Hebbian errors in learning: An analysis using the Oja model. J. Theor. Biol. (2009), doi:10.1016/j.jtbi.2009.01.036

stochastic changes in the overall number of synapses comprising a connection will change the overall learning rate at a connection. (In the simplest case, if a number of new silent plastic synapses happen to appear at a connection, while the overall weight is unchanged, the learning rate will be increased). One way to prevent this would be to ensure that only one of the synapses comprising a connection is plastic (Adams and Cox, 2002a) but this is a nonlocal rule. Another way would be for the average number of potential synapses comprising a connection to be reasonably high (perhaps ~50) so that fluctuations are relatively small. In several cases the average number of actual synapses at a connection is around five (Markram et al., 1997a, b; Barbour et al., 2007) and since these may only form about 10% of the total (potential) synapses, learning rates at different connections would be fairly similar (and of course identical when time-averaged). Of course for this rough-and ready solution to work, axons and dendrites would have to intersect sufficiently often, implying a high degree of branching. Although there have been some claims that weak synapses are more plastic (Matsuzaki et al., 2004), other evidence suggests that all synapses are equally plastic (Kopec et al., 2006); this could be achieved if strengthening added a new plastic "unit" to each synapse, with previously added units all rendered implastic (Adams and Cox, 2002a; Lisman and Raghavachari, 2007). A related issue is that the synapses comprising a connection will be at different electrotonic distances along the dendrite, and therefore will influence spiking differently, and have different effective learning rates. Rumsey and Abbott (2004) have proposed that a separate antiSTDP can be used to equalize efficacy of synapses.

There is strong evidence for both these ways to achieve complete connectivity (Le Be and Markram, 2006; Alvarez and Sabatini, 2007), and we think this equal error-onto-all is the most biologically plausible assumption, and it will be the principal target of our analysis.

The stabilized weight vector of the modified (inaccurate) Oja model will differ from the principal component of $\mathbf{C}$. If the matrix $\mathbf{EC}$ has a unique maximal eigenvalue, we can talk about its principal eigenvector $\mathbf{w_{EC}}$. (This assumption is not a very strong additional requirement; in the error-to-all case, for example, we show in the Supplementary Material that it is almost always the case.) If $\mathbf{E}$ is invertible, the inspecific learning algorithm will now converge to $\mathbf{w_{EC}}$ rather than to the principal component $\mathbf{w_C}$ of the input distribution. Indeed, for $\mathbf{E}$ symmetric and positive definite, one can define $\mathbf{F}=\sqrt{\mathbf{E}}$. Under the change of variables $\mathbf{x}'=\mathbf{Fx}$, $\mathbf{w}'=\mathbf{F}^{-1}\mathbf{w}$, one recovers the standard Oja rule 2.6, with a new covariance matrix $\mathbf{C}'=\mathbf{FCF}$, so the conditions for convergence are easily derived from those obtained in the noiseless case (see also Botelho and Jamison, 2002, 2004). The analysis in Appendix 1 shows that the convergence follows more generally, i.e. when $\mathbf{E}$ does not necessarily have all nonzero eigenvalues. Since the output of the Oja neuron allows optimal input reconstruction (at least in the least squares sense), Hebbian infidelity would lead to suboptimal performance. We quantified the effect of inaccuracy as the cosine of the angle θ between the principal component $\mathbf{w_C}$ of $\mathbf{C}$ and the principal eigenvector $\mathbf{w_{EC}}$ of $\mathbf{EC}$, which are the stabilized weight vectors in absence and, respectively, presence of error:

$$\cos(\theta)=\frac{\mathbf{w_C^T w_{EC}}}{\|\mathbf{w_C}\|\cdot\|\mathbf{w_{EC}}\|} \tag{3.12}$$

We examined how this measure of error depends on parameters such as the size n and the input error ε for a given $\mathbf{C}$. As the analysis for arbitrary input distributions is rather intractable (because $\mathbf{E}$ and $\mathbf{C}$ do not commute), we detail only a few simple cases of uncorrelated (Section 4.1) and correlated inputs (Section 4.2), illustrating the results with plots and simulations.

4. Results

We start with examples of simulations of the behavior of the erroneous rule in the error-onto-all case, using either uncorrelated (Fig. 2a) or correlated (Fig. 2b) inputs. In all cases the network is initialized with random weights. In the uncorrelated case and in the absence of error, the correct principal component is learned rapidly and accurately. The small fluctuations away from PC1 reflect the nonzero learning rate; they are most obvious at error rates for which the dependence of performance on control parameters is steepest (see Fig. 5a). In all cases, performance (measured by $\cos(\theta)$) gradually deteriorates with progressive increase in error, although the magnitude of the decrease depends on error and on correlation. The remainder of our results explore these effects in more detail, using calculations and analysis. Fig. 2 also shows that learning is somewhat slowed by inaccuracy, as expected; however, we do not analyze learning kinetics further here.

4.1. Uncorrelated inputs

This section shows how network performance depends on the quality factor $q\in[0,q_0(n)]$ (or alternatively on the error factor $b=1-q$) in the case of uncorrelated inputs. We illustrate this dependence by a combination of plots and analytical results.

For the uncorrelated inputs case, we consider a diagonal $\mathbf{C}$ with higher variance on the first component:

$$\mathbf{C}=\begin{pmatrix}\lambda & 0 & 0 & \cdot & 0\\ 0 & 1 & 0 & \cdot & 0\\ 0 & 0 & \cdot & \cdot & \cdot\\ \cdot & \cdot & \cdot & 1 & 0\\ 0 & 0 & \cdot & 0 & 1\end{pmatrix} \tag{4.1}$$

where $\lambda>1$, so that $\mathbf{w_C}=(1,0,\ldots,0)^T$. In this case,

$$\cos(\theta)=\frac{|(\mathbf{w_{EC}})_1|}{\|\mathbf{w_{EC}}\|} \tag{4.2}$$

is our measure of the system's performance.

We studied how $\cos(\theta)$ changes with the error (either ε or b), n and λ. We numerically calculated $\cos(\theta)$ as a function of the error in the two cases where the error is apportioned to two neighbors (nearest neighbor model, Fig. 3a) or to all other connections (error onto all model, Fig. 4a).

The curves in the nearest neighbor model (Fig. 3a) can be understood in the following way. First consider a curve at a given network size. As outlined in the Methods, in the absence of error the high variance connection grows more rapidly than the low variance connections, eventually completely winning, so the final weight vector points in that direction. However, in the presence of error, the immediate neighbors strengthen more than they would have in the absence of error, as a result of leakage from the preferred connection (see Fig. 3b); this means that future patterns will produce extra strengthening of those neighboring connections (because they are stronger and so produce larger outputs); this extra strengthening at the neighbors leads to increased strengthening of the neighbors of the neighbors, and so on down the line. Since the weight vector is normalized, these "wrong" strengthenings combine to reduce the preferred weight, although as long as learning shows some specificity, the preferred final weight is always strongest (see Fig. 3b).

Fig. 3b shows the distribution of equilibrium weights as a function of "distance" from the preferred neuron and of error b. The nearest neighbor case corresponds to the "fitness" model we simulated in previous work, and analyzed in the large n limit (Adams and Cox, 2002a), with "fitnesses" being the input

Please cite this article as: Rădulescu, A., et al., Hebbian errors in learning: An analysis using the Oja model. J. Theor. Biol. (2009), doi:10.1016/j.jtbi.2009.01.036

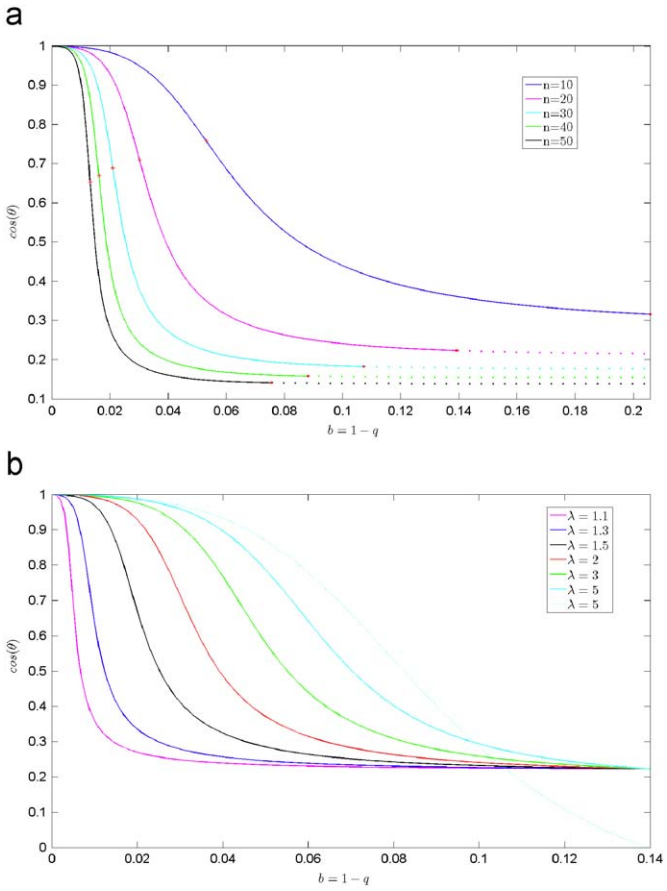

Fig. 4. Error-onto-all, discrete updates model. Upper plot: Dependence of $\cos(\theta)$ on the error factor b in the case of uncorrelated inputs with $\lambda = 2$. The performance is measured as $\cos(\theta)$, where θ is the angle between the principal eigenvector of **EC** and the principal component of **C**. Each color-coded curve corresponds to a different network size n, as shown in the legend. The curves were plotted as solid lines for b between zero and the trivial value $b_0(n) = 1/\sqrt[n]{n}$, and as dotted lines for the error b larger than the trivial value, because this range is not biological. The point of steepest downward slope (inflection point) is marked on each graph by a red asterisk. Consistently with our calculations, the inflection point is always situated between zero and the trivial value $b_0(n)$, getting arbitrarily close to zero for large enough n. Note that the $n = 10$ curve agrees almost exactly with the results in Fig. 2a, if the b values are converted to the corresponding ε values. Lower plot: For a fixed network size $n = 20$ and different values of the input variance λ, we show the dependence of the output performance $\cos(\theta)$ on the synaptic error b, for $b \in [0, b_0(n)]$. Each curve corresponds to a different λ value from $\lambda = 1.1$ to 5. When the variance is very close to $\lambda \to 1$, the stable weight vector approaches $\mathbf{w} = (1, 1 \ldots, 1)$ independently of the error, so $\cos(\theta) \to 1/\sqrt{n} \sim 0.22$ for all values of $b \neq 0$. Also, $\cos(\theta) = 1/\sqrt{n}$ for all λ at $b = b_0(n)$. The performance improves with larger variance, which agrees with our analytical results. The dotted line shows the perturbation approximation for $\lambda = 5$, which works well only at low error. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

variances. In that model, the weight distribution on the nonpreferred connections followed a double exponential function of distance. For sufficiently small error (or large λ), the distribution of weights on the nonpreferred (low variance) connections is close to the single exponential distribution found in this limit in the "fitness" model (see dashed curve in Fig. 3b); the "space constant" for the weight distribution varied in the expected manner (Adams and Cox, 2002a) with the Hebbian error (Fig. 3b) and with the variance λ.

In the remainder of the paper we focus on the error-onto-all model, in which the quality of the network is $Q > 0$ and the error is $\varepsilon = (1 - Q)/(n - 1)$. We will present in detail only the discrete update model, since this seems to be more biologically realistic (Petersen et al., 1998; O'Connor et al., 2005; Bagal et al., 2005); the continuous case is rather similar and treated in Appendix 2. Numerical calculations of the performance at different per synapse error values at various network sizes for the uncorrelated case are shown in Fig. 4a. The curve for $n = 10$ is plotted up to the trivial error value $b = 1/\sqrt[10]{10} \sim 0.205$ where learning is completely inspecific. There is a smoothly increasing degradation of performance with error, which drops to a much lower value for inspecific learning than seen in the previous cases, since error affects all nonpreferred inputs equally (for $b = b_0$, the weight vector is parallel to $(1, 1, \ldots 1)^T$, so the limiting cosine for the case $n = 10$ is $1/\sqrt{10} = 0.316$). In the remaining plots in Fig. 4a, the unbiological points of the curves (i.e. beyond the trivial value) are shown dotted.

Fig. 4b shows plots of performance against b for different values of variance λ, all for the case $n = 20$. For $\lambda = 1$, all eigenvalues are equal and the corresponding eigenvectors are saddles. There is a very large change in performance for small increases of λ above one, especially at low error values. A tiny bit of error stabilizes the behavior, so only the preferred weight is selected, although never perfectly.

We obtained an approximation for $\cos(\theta)$ at small ε values using perturbation theory (Kahn, 1990):

$$\cos(\theta) = \frac{\lambda - 1}{\sqrt{(\lambda - 1)^2 + (n - 1)\lambda^2 \varepsilon^2/(1 - n\varepsilon)^2}} \tag{4.3}$$

Fig. 4b shows that this formula agrees well with the exact results at sufficiently small ε.

We now proceed to an analytic treatment of these numerical results. The characteristic polynomial of the error matrix is:

$$P^{\mathbf{E}}(x) = det(\mathbf{E} - x\mathbf{I}) = (Q - \varepsilon - x)^{(n-1)}(1 - x) \tag{4.4}$$

Note that **E** is invertible, except when $Q = \varepsilon$, and Q and ε themselves depend on biological parameters (see Supplementary Material, Appendix 2).

The maximal eigenvalue μ of **EC** is given in this case by the larger solution of the quadratic equation:

$$\mu^2 - \mu[\lambda + 1 + \varepsilon(\lambda - 1 - n\lambda)] + \lambda - n\lambda\varepsilon = 0 \tag{4.5}$$

The maximal eigenvector will be in the direction $(s, 1, \ldots 1)^T$, where $s = s(\varepsilon, n, \lambda)$ is a "selectivity" value which expresses how strongly one of the weights is favored because one input is more active. This outcome reflects the fact that no weight except that corresponding to the high variance input is preferred (there are no privileged neighbor relations), so the behavior boils down to competition between the preferred weight and the set of equivalent nonpreferred weights, leading to the quadratic equation.

We usually estimate the output performance as $\cos(\theta)$, but here it is simplest to calculate $\tan(\theta)$, which is related to $\cos(\theta)$ by: $\cos(\theta) = 1/\sqrt{\tan^2(\theta) + 1}$, for $\theta \in [0, \pi/2]$.

$$h(q) = \tan(\theta(q)) = -\frac{\lambda - 1}{\sqrt{n - 1}} \frac{\lambda - \mu(q)}{\mu(q) - 1} = \frac{s}{\sqrt{1 - s^2}} \tag{4.6}$$

$$h' = -\frac{\lambda - 1}{\sqrt{n - 1}} \frac{\mu'}{(\mu - 1)^2} \tag{4.7}$$

$$h'' = -\frac{\lambda - 1}{\sqrt{n - 1}} \frac{\mu''(\mu - 1) - 2(\mu')^2}{(\mu - 1)^3} \tag{4.8}$$

where all derivatives are with respect to q. As $\mu' > 0$, we have $h' < 0$ for all q. This is consistent with our simulations: performance decays as the quality factor decreases.

Both the discrete and continuous update models show similar features (Section 3.1 and Supplementary Material, Appendix 3). The angle $\theta = \theta(q)$ (measured by its tangent $h(q) = \tan(\theta(q))$)

Please cite this article as: Rădulescu, A., et al., Hebbian errors in learning: An analysis using the Oja model. J. Theor. Biol. (2009), doi:10.1016/j.jtbi.2009.01.036

decreases as q goes from 0 to 1. In both cases $h(1)=0$, which corresponds to perfect performance for perfect quality. Also, $h(0)\to 0$ as $n\to\infty$, which shows that the output degrades more severely with error for larger values of the network size (because of synaptic "crowding"). Moreover, $h'(1)\to\infty$ as $n\to\infty$, which shows that the rate of the angle decay at $q=1$ gets very steep with large n. However, since the slope is always finite at finite ε, there is no "error catastrophe" (see Discussion).

A less obvious observation concerns the inflection point on each graph, where the decay rate (or "error sensitivity") of the performance is steepest (see red asterisks in Figs. 4a and 5). Although an exact estimate is intractable, we obtained, using the above expressions for the derivatives of $\tan(\theta)$, a lower bound: the inflection point is always situated in the interval $[q_0(n),1]$ (or equivalently in $[0,b_0(n)]$, when refering to synaptic error); see Appendix 3. Fig. 5a further suggests that the inflection point always moves to the left in step with the leftward shift in the trivial error value as n gets larger.

In summary, in the uncorrelated case, high per-synapse quality ensures excellent performance except when inputs are numerous (high n), or almost indistinguishable (low λ). Conversely, since performance only improves very slightly when error is further reduced from initially very low values, it would be difficult for evolution to attain very low error rates. We next asked if these features remain true for correlated inputs.

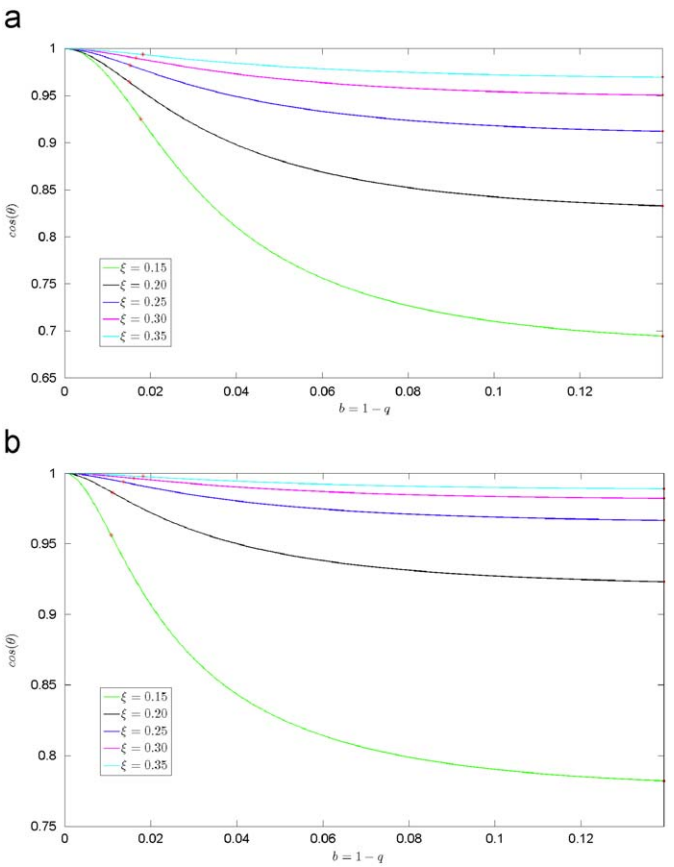

Fig. 5. Dependence of $\cos(\theta)$ on the error factor b, for the error-onto-all, discrete updates model for correlated inputs. The size $n=20$ and $\lambda=4$ have been fixed. Each curve illustrates a different background covariance ξ, as shown in the legend. The inflection point on each curve was marked by a red asterisk. The inflection points are closest to zero for intermediate values of ξ, which agrees with the result in Fig. 7. Upper plot: "high covariance pair" input distribution. Lower plot: "uniform covariance" inputs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.2. Correlated inputs

We now study the equilibrium behavior of the network in response to two simple cases of correlated inputs, in the error-onto-all model, with the following covariance matrices:

$$\mathbf{C}=\begin{pmatrix} 1 & \lambda & \xi & \cdot & \xi \\ \lambda & 1 & \xi & \cdot & \xi \\ \xi & \xi & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \xi \\ \xi & \xi & \cdot & \xi & 1 \end{pmatrix} \tag{4.9}$$

where $1>\lambda>\xi>0$ ($\mathbf{C}$ has higher covariance on one pair) and

$$\mathbf{C}=\begin{pmatrix} \lambda & \xi & \xi & \cdot & \xi \\ \xi & 1 & \xi & \cdot & \xi \\ \xi & \xi & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \xi \\ \xi & \xi & \cdot & \xi & 1 \end{pmatrix} \tag{4.10}$$

where $\lambda>1>\xi>0$ ($\mathbf{C}$ has small uniform background covariance with one high variance input).

Fig. 5 illustrates the dependence of performance on error at various "background correlation" ξ values in a network with 20 inputs, for the two above cases (top plot—higher covariance pair; bottom plot—high variance neuron with uniform background correlation). Since the slopes along the curves corresponding to different ξ values are not simply scalings of each other, it follows that the way the performance degrades with error depends on the background correlation. For intermediate background correlation ξ (e.g., black and blue curves in Fig. 5a), the output shows the highest error-sensitivity at very small error, while for very weak (light blue and pink curves) and for very strong (green curve) background correlation, the maximum error-sensitivity appears at larger values of the error. This results in the inflection point first moving to the left as ξ increases, then moving back to the right (see Figs. 5a and b; the rightward movement is only visible at lower ξ values than those shown in Fig. 5b).

Fig. 6 shows the dependence of performance on error for various network sizes, using fixed λ and ξ values, in both types of background correlation model. In this case the initial effect of error is very strong at large network sizes (because of synaptic crowding), but performance then reaches rather constant levels which are fairly close to the error-free level, because high background correlations tend to equalize all weights even in the absence of error.

We now analyze these numerical results.

4.2.1. Model 1—high covariance on one pair

The principal component of **EC** is the unit vector pointing in the direction of $(s,s,1,\ldots,1)^T$.

Here, the output's "selectivity" $s=s(n,\varepsilon,\lambda,\xi)$ is given by:

$$\frac{1}{s}=1+\frac{(1-n\varepsilon)(\lambda-\xi)}{z_{\mathbf{EC}}^{-}} \tag{4.11}$$

where $z_{EC}^{-}<0$ is the smaller root of a quadratic defined in Appendix 3. Once again, there is competition between the sets of equivalent preferred and nonpreferred weights.

In both models, the selectivity can be used to interpret features of the output performance with various degree of error. As the explicit formula for s is rather complicated, we calculated the upper bound and the lower bound ($r(\varepsilon)$), which are simpler and yet still suggest some of the main features:

$$r(\varepsilon)=1-\frac{(1-n\varepsilon)(\lambda-\xi)}{[1-(n-2)\varepsilon](\lambda-\xi)+n(\varepsilon+\xi-\varepsilon\xi)}\leqslant\frac{1}{s}\leqslant 1 \tag{4.12}$$

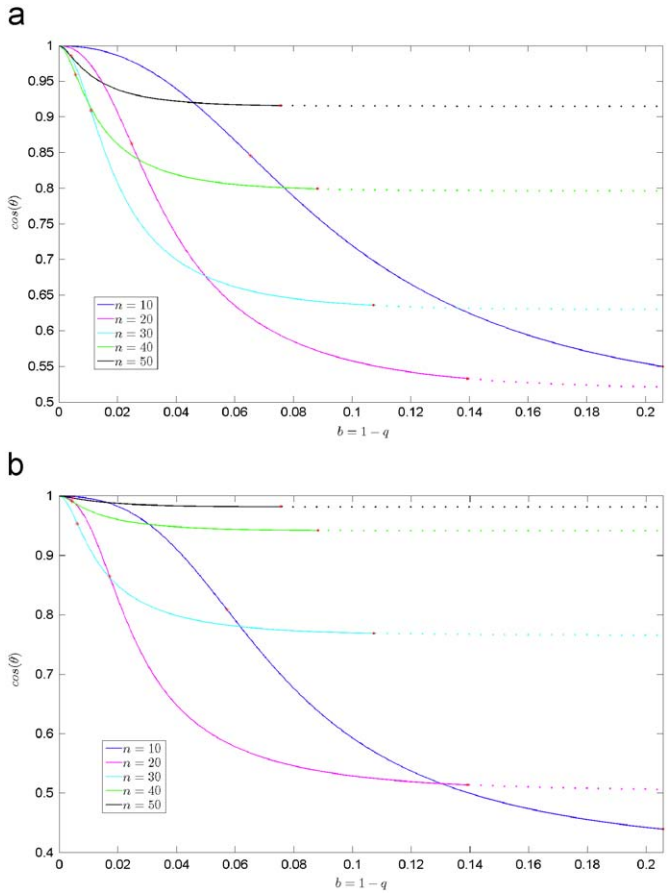

Fig. 6. Dependence of $\cos(\theta)$ on the error factor b, for the error-onto-all, discrete update model. The variances λ and ξ have been fixed to $\lambda = 4$ and $\xi = 0.1$. Each curve corresponds to a different network size n, and the inflection point on each curve is marked by a red asterisk. The infection points approach zero as n gets arbitrarily large. Upper plot: "high variance pair" input distribution. Lower plot: "uniform variance" inputs. The red dots show the trivial error values, and the curves are shown dotted beyond this point because this is a nonbiological range. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

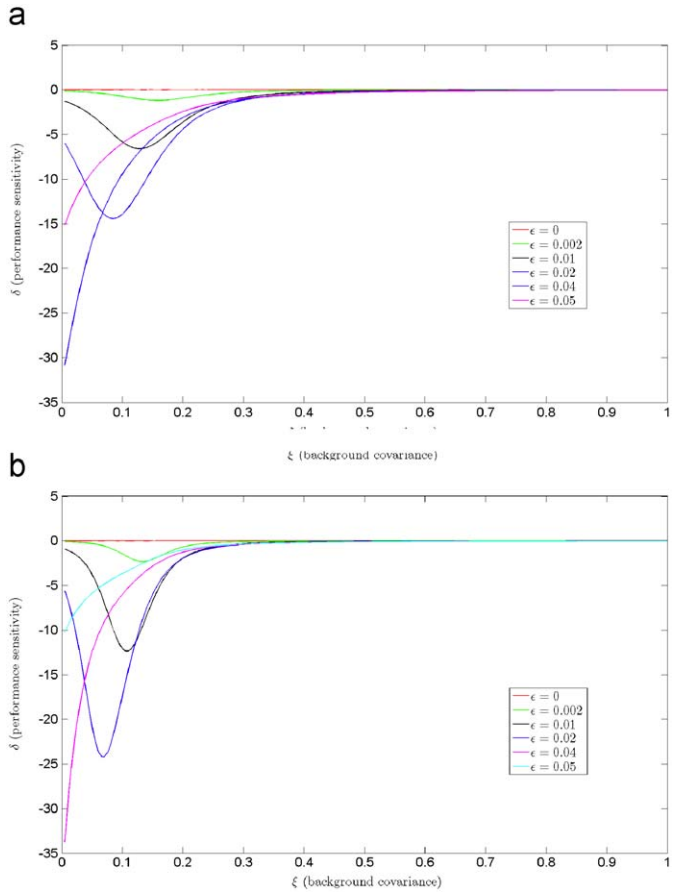

Fig. 7. Dependence of the output sensitivity $\delta = (\partial/\partial\varepsilon)(\cos(\theta))$ on the covariance ξ for $n = 20$, $\lambda = 4$, and five error values $\varepsilon = 0$, 0.002, 0.01, 0.02, 0.04 and 0.05. Each curve corresponds to a different error, as shown in the legend. The output shows the most sensitivity to ε at intermediate error values ($0.02 < \varepsilon < 0.05$) and at low covariance values ($0 < \xi < 0.25$). Upper plot: "high variance pair" input distribution. Lower plot: "uniform variance" inputs.

where $\lim_{\varepsilon\to 1/n} r(\varepsilon) = 1$ and $\lim_{\varepsilon\to 0} r(\varepsilon) = 1 - (\lambda - \xi)/(\lambda + (n-1)\xi)$.

This can be compared with our other measure of output performance: the cosine of the angle $\theta = \theta(n, \varepsilon, \lambda, \xi)$ between the principal eigenvector $(s, s, 1 \ldots 1)$ of **EC** and the principal component of the input $(s_0, s_0, 1 \ldots 1)$ (where $s_0 = s(n, 0, \lambda, \xi)$ is the selectivity in the absence of error).

$$\cos(\theta) = \frac{2ss_0 + n - 2}{\sqrt{2s^2 + n - 2}\sqrt{2s_0^2 + n - 2}} \tag{4.13}$$

4.2.2. Model 2—uniform pairwise covariance

As before, we compute the eigenvector **w** of **EC** corresponding to $\mu_{\mathbf{EC}}$. As expected, we get that **w** is in the direction of $(s, 1, \ldots, 1)^T$.

Here, the output's selectivity s is given by:

$$\frac{1}{s} = 1 + \frac{(1 - n\varepsilon)(\lambda - 1)}{z_{\mathbf{EC}}^-} \tag{4.14}$$

and has upper and lower bounds

$$r(\varepsilon) = 1 - \frac{(1 - n\varepsilon)(\lambda - 1)}{(n-1)[\xi - \varepsilon(\xi - 1)]} \leqslant \frac{1}{s} \leqslant 1 \tag{4.15}$$

Here also $\lim_{\varepsilon\to 1/n} r(\varepsilon) = 1$ and $\lim_{\varepsilon\to 0} r(\varepsilon) = 1 - (\lambda - 1)/(n-1)\xi$.

The relation with $\cos(\theta)$ is given by

$$\cos(\theta) = \frac{ss_0 + n - 1}{\sqrt{s^2 + n - 1}\sqrt{s_0^2 + n - 1}} \tag{4.16}$$

where s_0 is again the selectivity for zero error.

Thus in both models $\cos(\theta)$ has a similar dependence on b and n (see Fig. 6).

4.3. Error sensitivity

We define a quantity δ as the error sensitivity of the performance:

$$\delta = \frac{\partial \cos(\theta)}{\partial \varepsilon} \tag{4.17}$$

Fig. 7 shows plots of the dependence of δ on background covariance, measured at different error rates, for the two correlated cases. δ is always negative (error degrades performance, as in the uncorrelated case), except of course for $\xi = 1$, where the error-free and the erroneous equilibrium eigenvectors already have the same form. Also, δ is very small near zero error, again as in the uncorrelated case. At low error rates, adding background correlation increases the error sensitivity δ. The maximum error sensitivity is greatest at intermediate error rates.

Please cite this article as: Rădulescu, A., et al., Hebbian errors in learning: An analysis using the Oja model. J. Theor. Biol. (2009), doi:10.1016/j.jtbi.2009.01.036

These effects reflect two opposing processes. Background correlation increases the rate of growth of all connections; from a connection's point of view it looks as though the pressure driving selective growth of one (Fig. 6b) or two (Fig. 6a) connections has been reduced (e.g. is equivalent to a reduction in λ in Fig. 6b). But increases in positive background correlation tend to make the weights more equal, synergistic with increases in error. The second effect dominates at high error values.

We also looked at symbolic software computations of the sensitivity of performance to changes in background covariance, $\partial\cos(\theta)/\partial\xi$, at various values of ε and ξ. These can be used to understand the dependence of the sizes of the fluctuations visible in Fig. 2 on parameters. We interpret these fluctuations as small deviations of the input statistics from their average values (i.e. small spontaneous transient perturbations of parameters such as ξ). Their amplitudes should therefore follow $|\partial\cos(\theta)/\partial\xi|$. We found that $|\partial\cos(\theta)/\partial\xi|$ increased as error increased, in agreement with the behavior in Figs. 2b and c.

In Fig. 2b independent and equal variance "sources" were linearly mixed to generate correlated random vectors used as inputs to the erroneous Oja rule. These correlations act as a "background" which tends to equalize the weights even in the absence of error, so adding error has relatively little effect.

4.4. Other models and extensions

Here we consider an input distribution such that the variance is higher, but uneven, on two of the components, while the covariance is uniform (and possibly zero). The correlation matrix will be of the form:

$$\mathbf{C}=\begin{pmatrix}\lambda_1 & \xi & \xi & \cdot & \xi\\ \xi & \lambda_2 & \xi & \cdot & \xi\\ \xi & \xi & 1 & \cdot & \cdot\\ \cdot & \cdot & \cdot & \cdot & \xi\\ \xi & \xi & \cdot & \xi & 1\end{pmatrix} \tag{4.18}$$

with $\lambda_1>\lambda_2>1>\xi$.

The modified correlation matrix **EC** has the eigenvalue $(Q-\varepsilon)(1-\xi)$, with multiplicity $n-3$. The other three eigenvalues μ_1, μ_2 and μ_3 are distinct and lie, respectively, within the intervals:

$$\begin{aligned}&(Q-\varepsilon)(1-\xi)<\mu_1<(Q-\varepsilon)(\lambda_2-\xi)\\ &(Q-\varepsilon)(\lambda_2-\xi)<\mu_2<(Q-\varepsilon)(\lambda_1-\xi)\\ &\max\{(Q-\varepsilon)(\lambda_1-\xi),n\xi+(1-\xi)+\varepsilon(\lambda_1+\lambda_2-2)\}\\ &\quad<\mu_3<n(\xi+\varepsilon-\varepsilon\xi)+\varepsilon(\lambda_1+\lambda_2-2)\end{aligned} \tag{4.19}$$

Clearly, $\mu=\mu_3$ is always the unique maximal eigenvalue of **EC**.

In the case of uncorrelated inputs, for example, $\varepsilon=0$ corresponds to $s_1=1$ and $s_2=0$ (the maximal eigenvalue is $\mu=\lambda_1$, and its corresponding eigenvector is the first element $(1,0,\ldots,0)$ of the standard orthonormal basis in $\mathbb{R}^n$). As the error increases from $\varepsilon=0$ to $\varepsilon=1/n$, the eigenvector $(s_1,s_2,0\cdots.0)^T$ evolves such that the ratio s_1/s_2 decays very dramatically from ∞ (when $\varepsilon=0$) to finite values (see Fig. 8). When $\varepsilon\to 1/n$ (the trivial value) all weights equalize and thus $s_1/s_2\to 1$ (Fig. 8b). Thus in a situation where two highly (but inevitably unequally) active inputs are to be selectively wired by Hebbian learning, the presence of error can promote the desired outcome, at least in the large n case.

When the inputs are correlated, the dependence of μ on parameters is more complicated. The eigenspace of μ is the direction $(s_1,s_2,1,\ldots 1)^T$, where the selectivities s_1 and s_2 themselves depend, via the eigenvalue μ, on all the system parameters:

$$\frac{s_1}{s_2}=1+\frac{(Q-\varepsilon)(\lambda_1-\lambda_2)}{\mu-(Q-\varepsilon)(\lambda_1-\xi)} \tag{4.20}$$

It is easy to observe that, as $\varepsilon\to 1/n$ (the trivial error value), $Q-\varepsilon\to 0$, hence $s_1/s_2\to 1$. As in the uncorrelated case, weights tend to equalize as the error gets close to the trivial value (see Fig. 9). However, the slope of the decay of s_1/s_2 is different from the uncorrelated case, since s_1/s_2 is always finite when the inputs are correlated with $\xi>0$, even for zero ε.

Although these results are not general, they seem to apply to various other situations with increasing degree of background correlation (e.g. Fig. 2). Similar behavior can be observed, for instance, in an Oja network learning from correlated inputs obtained by rotations of n-dimensional normally distributed vectors. Once again one sees that as correlations increase the inflexion points in the performance versus error plots shift to the left and then to the right (Fig. 10; compare with Figs. 5 and 2b), confirming that at low error introducing small correlation increases error sensitivity.

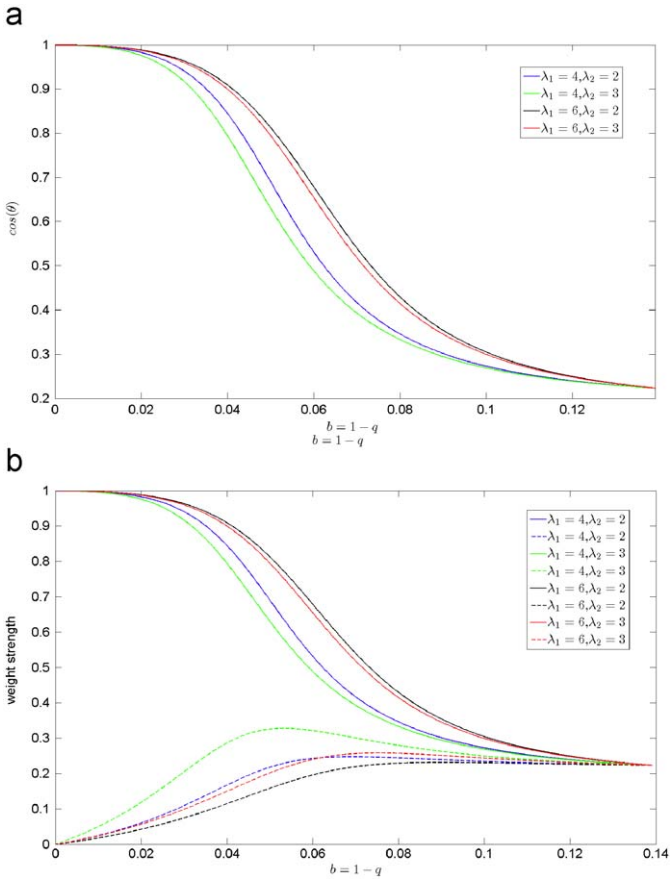

Fig. 8. Error-onto-all, discrete update model for $n=20$ cells receiving uncorrelated inputs with $\lambda_1>\lambda_2>1$. Upper plot: Dependence of $\cos(\theta)$ on the synaptic error b, shown as b increases from zero to the trivial value $b_0(n)=1-1/\sqrt[n]{n}=1-1/\sqrt[20]{20}\sim 0.14$. The trivial error $b_0(n)$ equalizes all weights and makes $\cos(\theta)=1/\sqrt{n}=1/\sqrt{20}\sim 0.22$, independently of the distribution variances λ_1 and λ_2. Lower plot: Evolution of the normalized weights $s_1/\sqrt{s_1^2+s_2^2+(n-2)}$ and $s_2/\sqrt{s_1^2+s_2^2+(n-2)}$ with respect to b. As b increases from zero to $b_0(n)$, the weights equalize and the ratio s_1/s_2 drops from ∞ to 1.

Please cite this article as: Rădulescu, A., et al., Hebbian errors in learning: An analysis using the Oja model. J. Theor. Biol. (2009), doi:10.1016/j.jtbi.2009.01.036

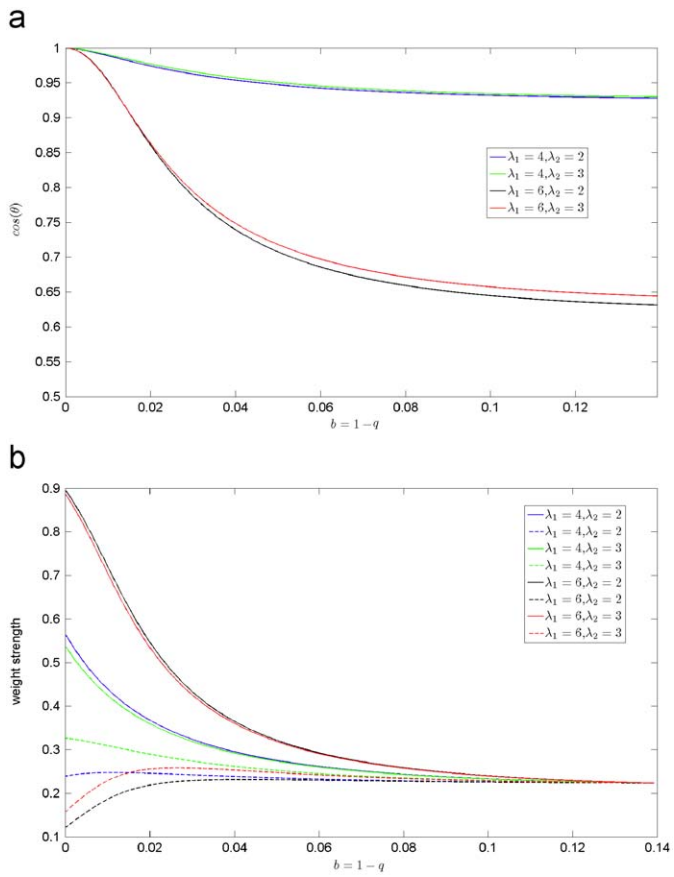

Fig. 9. Error-onto-all, discrete update model for $n = 20$ cells receiving correlated inputs with variances $\lambda_1 > \lambda_2 > 1$ and small uniform covariances $\xi = 0.2$. Upper plot: Dependence of $\cos(\theta)$ on the synaptic error b, shown as b increases from zero to the trivial value $b_0(n) = 1 - 1/\sqrt[n]{n} = 1 - 1/\sqrt[20]{20} \sim 0.14$. The trivial error $b_0(n)$ equalizes all weights, but $\cos(\theta)$ varies at $b_0(n)$. Since the principal component of **C** varies with parameters, so will the angle θ at the trivial error value. Lower plot: Evolution of the normalized weights $s_1/\sqrt{s_1^2 + s_2^2 + (n-2)}$ and $s_2/\sqrt{s_1^2 + s_2^2 + (n-2)}$ with respect to b. As b increases from zero to $b_0(n)$, the weights equalize and the ratio s_1/s_2 drops from an initial finite, parameter-dependent value to 1.

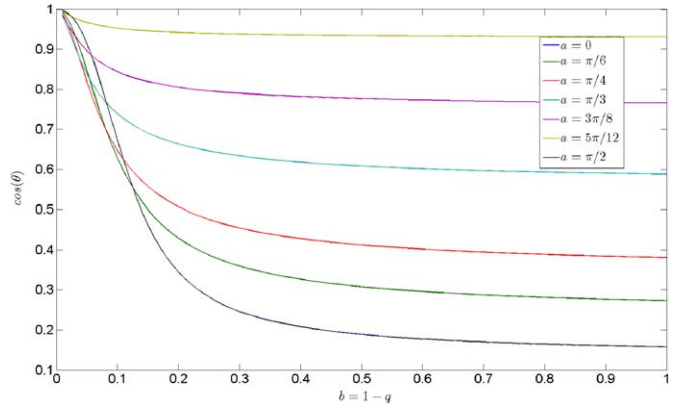

Fig. 10. Oja network learning a distribution of correlated inputs obtained by rotations of n-dimensional normally distributed vectors. Here $\lambda = 10$ and $n = 40$. The amount of rotation alpha was varied as shown in the inset.

5. Discussion

In this paper we analyze and describe the effect of introducing errors into Hebbian learning by a single model neuron, in the form of local or global spread of the updates induced at one connection to other connections. This work was motivated by our earlier suggestion (Adams and Cox, 2002a) that such errors may play roles analogous to mutations in genetic evolution, as well as by recent experimental evidence (Engert and Bonhoeffer, 1997; Harvey and Svoboda, 2007). In previous work we introduced local spread into an even simpler model network, consisting of an input axon making Hebbian connections onto a set of output neurons. Because in that early model each connection operates independently (apart from overall normalization of the weights), it is essentially equivalent to the present model when inputs are uncorrelated. The advantages of the present model are mainly that one can study arbitrary input covariance and error matrices. We obtained some analytic results (for example, we showed that the modified learning rule converges to the duly modified "principal component"), and we further investigated several special, but representative, cases by explicit calculation. Our main finding is that learning remains stable in the presence of crosstalk, under rather general conditions, although it gradually gets worse as crosstalk increases. There appears to be no complete learning failure up to the trivial limit, where synapse spacing becomes so small that the Hebb rule is fully inspecific.

Our goal was to study Hebbian error, or crosstalk, in the simplest possible model of unsupervised learning, since our ultimate aim is to understand circuitry in the neocortex, which seems to be specialized for such learning. We chose the Oja model of a neuron as a principal component analyzer because it is perhaps the most widely known, and simplest, example of a Hebbian model of unsupervised learning. The Oja model is rather unbiological: it uses a rate-coding scheme and a simple multiplicative Hebbian learning rule, together with an elegant, local, but perhaps implausible, normalization procedure. It also supposes that input patterns are zero-mean, with the Hebbian and normalizing parts of the rule both able to trigger ltp or ltd. A more biologistically realistic model would use timed spikes, spike-timing dependent plasticity (STDP) and natural inputs (e.g. movies). However, we wanted to study the effect of one particular aspect of biological realism—plasticity crosstalk—in the context of a model that is otherwise as transparent as possible. Gerstner and Kistler (2002) have developed a model intermediate between the Oja model and a detailed spiking model. It assumes a rate-coding scheme, with Poisson spikes and STDP with ltd and ltp lobes, with postsynaptic spikes triggered by presynaptically generated epsps. This model learns the principal component of the zero-mean inputs, without explicit centering or normalization. It would be interesting to combine this model with ours, by applying an error matrix either to the ltp or ltd lobes. One might anticipate that in the first case the effect would be similar to what we describe, while in the second case normalization might be inaccurate.

Although our numerical results suggest that in a representative range of cases the introduction of crosstalk produces graceful degradation, all we have been able to prove, even in the simple case of uncorrelated inputs, is that learning remains stable. It would be theoretically possible that as error increases the two leading eigenvalues closely approach (while remaining distinct), and that at an "avoided crossing" the nature of the leading eigenvector changes dramatically. This is what happens in the Eigen molecular evolution dynamics (Nowak and Schuster, 1989; Eigen et al., 1989; Swetina and Schuster, 1982; Tarazona, 1992; Volkenstein, 1994), such that at a threshold error value, the leading eigenvector (which represents the relative concentrations of all possible polynucleotide sequences), suddenly switches from a "quasispecies" distribution dominated by the fastest replicating sequence to a binomial distribution (with all sequences equiprobable). This behaviour is only abrupt in the thermodynamic limit

Please cite this article as: Rădulescu, A., et al., Hebbian errors in learning: An analysis using the Oja model. J. Theor. Biol. (2009), doi:10.1016/j.jtbi.2009.01.036

(large dimensionality) and requires a statistical–mechanical analysis. In our model the two leading eigenvalues always seem remain separated (up to the trivial error) and the network performance changes smoothly.

Hebbian learning in the essentially linear Oja model is driven entirely by the input covariances, and, for Gaussian patterns, finds the statistically optimal representation, the first principal component. Sensory input is generated in a very complicated (and essentially unknown) way, and it seems reasonable that it should initially be processed assuming it is approximately Gaussian, i.e. by (pairwise) decorrelation, as in PCA. While such processing cannot capture the underlying real-world generative process, it would provide an efficient way to transmit information to brain structures (such as the neocortex) that could use more sophisticated techniques, involving nonlinear Hebbian learning, sensitive to higher-order correlations. Furthermore, such techniques—e.g. independent component analysis (Bell and Sejnowski, 1995; Hyvarinen et al., 2001)—often work best if inputs are decorrelated. It has been suggested that the center-surround organization of retinal ganglion cells reflects at least partly the fact that the pairwise correlations in retinal images decay radially from a given pixel (Atick and Redlich, 1990, 1992; Bell and Sejnowski, 1997; Srinivasan et al., 1982), so that ganglion cells send optimally compressed information (via thalamus) to cortex. In principle Hebbian mechanisms might underlie this strategy, and our results suggest that crosstalk would compromise efficient decorrelation. While at low error there would be little loss in information (see Fig. 11 for an example of mutual information loss), the slight "dewhitening", combined with the effect of similar crosstalk on subsequent nonlinear learning, might cripple the neocortex. In particular, we have found (paper submitted) that small amounts of crosstalk completely destabilize nonlinear learning in ICA models, especially for nonwhite inputs.

It seems possible that learning inaccuracies of the type we study here could be responsible for some of the apparently "aberrant" wiring seen in the lateral geniculate nucleus, where activity and NMDAR-driven mechanisms lead to the refinement of center-surround visual receptive fields (Shatz, 1996). In particular, as development proceeds, the number of retinal inputs to relay cells decreases dramatically, until only highly correlated inputs (with closely overlapping receptive fields) remain (Chen and Regehr, 2000). However, detailed analysis reveals that occasional weak inputs may remain, even though they seem inappropriate—e.g., *X* cells receiving *Y* input (Wilson et al., 1984). While it is possible that these anomalies may represent some unusual clever strategy (Alonso et al., 2006), our results suggest they could arise from Hebbian crosstalk. Interestingly, the *X* cells, which have the most precise RFs, receive their retinal inputs on spinelike dendritic appendages (Sherman, 2007; Sherman and Guillery, 2001), which may promote calcium isolation and hence minimize crosstalk.

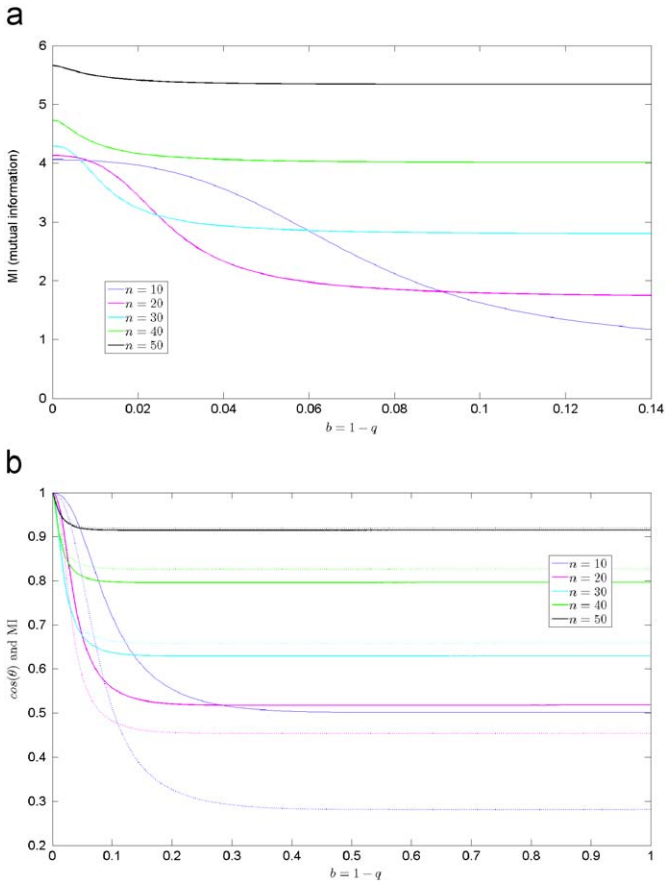

Fig. 11. Mutual information (MI) at different error values and network sizes, for correlated Gaussian inputs. Upper plot: The plots show MI as function of error for different network sizes. The MI depends on output variance; since all the inputs contribute to the output variance, the MI increases with network size. Because error increases, the relative contribution of the poorly correlated inputs to the output variance MI decreases with error. Lower plot: The relative MI (dotted lines) at different error rates is expressed as the fraction of the MI at zero error, and compared with $\cos(\theta)$ at different errors (solid lines) for various network sizes. In all cases error decreases MI, and when $\cos(\theta)$ is small, it closely tracks the MI.

6. Conclusion

Although it is widely appreciated that physics sets ultimate limits to biology (Bialek, 1987), little attention has been paid to the physical limits to the process that is of most interest to humans: learning. The Oja rule is the simplest and best-studied unsupervised learning rule. It captures the key point that linear Hebbian learning is driven by pairwise correlations (in the form of the input covariance matrix). Not surprisingly, when the rule is inaccurate, it fails to accurately learn the expected (and typically most useful) result. Although the failure is graceful, it can be severe when the patterned activity driving growth of particular weights is rather weak. We propose that even though the chemical changes driving Hebbian learning are largely confined to the synapses where learning is induced, the very high density of synapses along dendrites means that significant crosstalk, and therefore somewhat degraded learning, is inevitable. In future work we hope to show that such inevitable crosstalk can completely prevent Hebbian learning of higher-than-pairwise correlations, unless additional interesting machinery, roughly corresponding to the basic neocortical microcircuit, is employed.

While the linear case studied here does not show a learning collapse at a critical error rate, unlike the situation in genetic evolution (Eigen et al., 1989; Swetina and Schuster, 1982; Tarazona, 1992), our recent unpublished results show that a collapse does occur using nonlinear Hebbian rules. This supports the original suggestion (Adams, 1998) that Hebbian errors are analogous to mutations. The advent of the neocortical machinery that we postulate reduces the error rate would thus be analogous to the transition from the RNA world to the DNA/protein world (Orgel, 1994), allowing the emergence of sophisticated learning (and "mind"), a neural analog of Darwinian adaptation ("life").

Appendix A. Supplementary data

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jtbi.2009.01.036.

Please cite this article as: Rădulescu, A., et al., Hebbian errors in learning: An analysis using the Oja model. J. Theor. Biol. (2009), doi:10.1016/j.jtbi.2009.01.036

References

Adams, P., 1998. Hebb and Darwin. J. Theoret. Biol. 195, 419–438.

Adams, P., Cox, K.J.A., 2002a. A new view of thalamocortical function. Plos. Trans. R. Soc. London B 357, 1767–1779.

Adams, P.R., Cox, K.J.A., 2002b. Synaptic Darwinism and neocortical function. Neurocomputing 42, 197–214.

Adams, P., Cox, K.J.A., 2006. A neurobiological perspective on building intelligent devices. Neuromorphic Engineer. 3, 2–10.

Alonso, J.M., Yeh, C.I., Weng, C., Stoelzel, C., 2006. Retinogeniculate connections: a balancing act between connection specificity and receptive field diversity. Progr. Brain Research 154 (Chapter 1).

Alvarez, V.A., Sabatini, B.L., 2007. Anatomical and physiological plasticity of dendritic spines. Ann. Rev. Neurosci. 30, 79–97.

Andersen, P., Sundberg, S.H., Sveen, O., Wigstrom, H., 1977. Specific long-lasting potentiation of synaptic transmission in hippocampal slices. Nature 266, 736–773.

Atick, J., Redlich, N., 1990. Towards a theory of early visual processing. Neural Computation 2, 308–320.

Atick, J., Redlich, A.N., 1992. What does the retina know about natural scenes? Neural Computation 4, 196–210.

Bagal, A.A., Kao, J.P.Y., Tang, C.-M., Thompson, S.M., 2005. Long-term potentiation of exogenous glutamate responses at single dendritic spines. Proc. Natl. Acad. Sci. USA 102, 14434–14439.

Barbour, B., Brunel, N., Hakim, V., Nadal, J.P., 2007. What can we learn from synaptic weight distributions? Trends in Neurosci. 30 (12), 622–629.

Bell, A.J., Sejnowski, T.J., 1995. An information maximization approach to blind separation and blind deconvolution. Neural Computation 7, 1129–1159.

Bell, A.J., Sejnowski, T.J., 1997. The 'independent components' of natural scenes are edge filters. Vision Res. 37 (23), 3327–3338.

Bi, G.-Q., 2002. Spatiotemporal specificity of synaptic plasticity: cellular rules and mechanisms. Biol. Cybern. 87, 319–332.

Bialek, W., 1987. Physical limits on sensation and perception. Ann. Rev. Biophys. Chem. 16, 455–478.

Botelho, F., Jamison, J.E., 2002. A learning rule with generalized Hebbian synapses. J. Math. Anal. Appl. 273, 529–547.

Botelho, F., Jamison, J., 2004. Qualitative behavior of differential equations associated with artificial neural networks. J. Dyn. Differential Equations 16, 179–204.

Chen, C., Regehr, W.G., 2000. Developmental remodeling of the retinogeniculate synapse. Neuron 28, 955–966.

Cox, K.J.A., Adams, P.R., 2000. Implications of synaptic digitisation and error for neocortical function. Neurocomputing 32–33, 673–678.

Diamantaras, K.I., Kung, S.-Y., 1996. Principal Component Neural Networks: Theory and Applications. Wiley, New York.

Eigen, M., McCaskill, J., Schuster, P., 1989. The molecular quasispecies. Adv. Chem. Phys. 75, 149–163.

Elman, J., Bates, E.A., Johnson, M.H., Karmiloff-Smith, A., Parisi, D., Plunkett, K., 1996. Rethinking Innateness: A Connectionist Perspective on Development. MIT Press, Cambridge.

Engert, F., Bonhoeffer, T., 1997. Synapse specificity of long-term potentiation breaks down at short distances. Nature 388, 279–284.

Grutzendler, J., Kasthuri, N., Gan, W.B., 2002. Long-term dendritic spine stability in the adult cortex. Nature 420, 812–816.

Gerstner, W., Kistler, W.M., 2002. Spiking Neuron Models. Cambridge University Press, Cambridge.

Harvey, C.D., Svoboda, K., 2007. Locally dynamic synaptic learning rules in pyramidal neuron dendrites. Nature 450, 1195–1200.

Hebb, D.O., 1949. The Organization of Behavior. Wiley, New York.

Hertz, J., Krogh, A., Palmer, R.G., 1991. Introduction to the Theory of Neural Computation. Lecture Notes Volume in the Santa Fe Institute for Studies in the Sciences of Complexity. Perseus Books Publishing.

Holtmaat, A.J., et al., 2005. Transient and persistent dendritic spines in the neocortex in vivo. Neuron 45, 279–291.

Hyvarinen, A., Karhunen, J., Oja, E., 2001. Independent Component Analysis. Wiley Interscience, New York.

Kahn, P.B., 1990. Mathematical Methods for Scientists and Engineers. Wiley Interscience, New York.

Kopec, C.D., Li, B., Wei, W., Boehm, J., Malinow, R., 2006. Glutamate receptor exocytosis and spine enlargement during chemically induced long-term potentiation. J. Neurosci. 26 (7), 2000–2009.

Le Be, J.V., Markram, H., 2006. Spontaneous and evoked synaptic rewiring in the neonatal neocortex. Proc. Natl. Acad. Sci. USA 103, 13214–13219.

Lendvai, B., Stern, E.A., Chen, B., Svoboda, K., 2000. Experience-dependent plasticity of dendritic spines in the developing rat barrel cortex in vivo. Nature 404, 876–881.

Levy, W.B., Steward, O., 1979. Synapses as associative memory elements in the hippocampal formation. Brain Res. 175, 233–245.

Lisman, J., 1989. A mechanism for the Hebb and the anti-Hebb processes underlying learning and memory. Proc. Natl. Acad. Sci. USA 86, 9574–9578.

Lisman, J., Raghavachari, L., 2007. A unified model of the presynaptic and postsynaptic changes during LTP at CA1 synapses. Science's STKE.

Llinas, R.R., Walton, K.D., 1998. Cerebellum. In: Shepherd, G.M. (Ed.), The Synaptic Organization of the Brain. Oxford University Press, Oxford, pp. 255–288.

Markram, H., Luebke, J., Frotscher, M., Sakmann, B., 1997a. Regulation of synaptic efficacy by coincidence of postsynaptic APs and EPSPs. Science 275, 213–215.

Markram, H., Luebke, J., Frotscher, M., Roth, A., Sakmann, B., 1997b. Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. J. Physiol. 500, 409–440.

Matsuzaki, M., Honkura, N., Ellis-Davies, G.C., Kasai, H., 2004. Structural basis of long-term potentiation in single dendritic spines. Nature 429, 761–766.

Nowak, M.A., Schuster, P., 1989. Error thresholds of replication in finite populations—mutation frequencies and the onset of Muller's ratchet. J. Theoret. Biol. 137, 375–395.

O'Connor, D.H., Wittenberg, G.M., Wang, S.S.-H., 2005. Graded bidirectional synaptic plasticity is composed of switch-like unitary events. Proc. Natl. Acad. Sci. USA 102 (27), 9679–9684.

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

Oja, E., Karhunen, J., 1985. On stochastic approximation of the eigenvectors and eigenvalues of the expectation of a random matrix. J. Math. Anal. Appl. 106, 69–84.

Orgel, L., 1994. The origin of life on earth. Scientific American 271 (4), 81.

Petersen, C.C.H., Malenka, R.C., Nicoll, R.A., Hopfield, J.J., 1998. All-or-none potentiation at CA3–CA1 synapses. Proc. Natl. Acad. Sci. USA 95, 4732–4737.

Rumsey, C.C., Abbott, L.F., 2004. Equalization of synaptic efficacy by activity- and timing-dependent synaptic plasticity. J. Neurophysiol. 91 (5).

Schuman, E.M., Madison, D.V., 1994. Locally distributed synaptic potentiation in the hippocampus. Science 263, 532–536.

Shatz, C.J., 1996. Emergence of order in visual system development. Proc. Natl. Acad. Sci. USA 93, 602–608.

Sherman, S.M., 2007. The thalamus is more than just a relay. Curr. Opin. Neurobiol. 17 (4), 417–422.

Sherman, S.M., Guillery, R.W., 2001. Exploring the Thalamus. Academic Press, San Diego.

Srinivasan, M.V., Laughlin, S.B., Dubs, A., 1982. Predictive coding: a fresh view of inhibition in the retina. Proc. R. Soc. Lond. B Biol. Sci. 216 (1205), 427–459.

Stepanyants, A., Hof, P., Chklovskii, D., 2002. Geometry and structural plasticity of synaptic connectivity. Neuron 34 (2), 275–288.

Stepanyants, A., Hirsch, J.A., Martinez, L.M., Kisvarday, Z.F., Ferecsko, A.S., Chklovskii, D.B., 2008. Local potential connectivity in cat primary visual cortex. Cereb. Cortex 18 (1), 13–28.

Swetina, P., Schuster, P., 1982. Self-replication with error—a model for polynucleotide replication. Biophys. Chem. 16, 329–345.

Tao, H.W., Zhang, L.I., Engert, F., Poo, M., 2001. Neuron 31 (4), 569–580.

Tarazona, P., 1992. Error threshold for molecular quasispecies as phase transitions: from simple landscapes to spin-glass models. Phys. Rev. A 45, 6038–6050.

Volkenstein, M.V., 1994. Physical Approaches to Biological Evolution. Springer, Berlin.

Wilson, J.R., Friedlander, M.J., Sherman, S.M., 1984. Fine structural morphology of identified X- and Y-cells in the cat's lateral geniculate nucleus. Proc. R. Soc. B 221, 411–436.

Wyatt Jr., J.L., Elfadel, I.M., 1995. Time-domain solutions of Oja's equations. Neural Computation 7 (5), 915–922.

Zador, A., Koch, C., 1994. Linearized models of calcium dynamics: formal equivalence to the cable equation. J. Neurosci. 14, 4705–4715.

Please cite this article as: Rădulescu, A., et al., Hebbian errors in learning: An analysis using the Oja model. J. Theor. Biol. (2009), doi:10.1016/j.jtbi.2009.01.036