Beyond Hebbian plasticity: Effective learning with ineffective Hebbian learning rules

Gal Chechik
The center for neural computation
Hebrew University in Jerusalem, Israel
and the School of Mathematical Sciences
Tel-Aviv University Tel Aviv, Israel
ggal@math.tau.ac.il

Isaac Meilijson
School of Mathematical Sciences
Tel-Aviv University Tel Aviv, Israel
isaco@math.tau.ac.il

Eytan Ruppin *
Schools of Medicine and Mathematical Sciences
Tel-Aviv University Tel Aviv, Israel
ruppin@math.tau.ac.il

May 24, 1999

Abstract

In this paper we revisit the classical neuroscience paradigm of Hebbian learning. We find that a necessary requirement for effective associative memory learning is that the efficacies of the incoming synapses should be uncorrelated. This is difficult to achieve in a robust manner by Hebbian synaptic learning, since it depends on network level information. Effective learning can yet be achieved by a neuronal process that maintains a zero sum of the incoming synaptic efficacies. This normalization drastically improves the memory capacity of associative networks, from an essentially bounded capacity to one that linearly scales with the network's size. Such neuronal normalization can be successfully carried out by activity-dependent homeostasis of the neuron's synaptic efficacies, which was recently observed in cortical tissue. Thus, our findings strongly suggest that effective associative learning with Hebbian synapses alone is biologically implausible and that Hebbian synapses must be continuously remodeled by neuronally-driven regulatory processes in the brain.

*To whom correspondence should be addressed

1 Introduction

Synapse-specific changes in synaptic efficacies, carried out by long-term potentiation (LTP) and depression (LTD) [Bliss and Collingridge, 1993], are thought to underlie cortical self-organization and learning in the brain. In accordance with the Hebbian paradigm, LTP and LTD modify synaptic efficacies as a function of the firing of pre and post synaptic neurons. In this paper we revisit the Hebbian paradigm, studying the role of Hebbian synaptic changes in associative memory storage, and their interplay with neuronally driven processes that modify the synaptic efficacies.

The importance of neuronally driven normalization processes has already been demonstrated in the context of self-organization of cortical maps [Miller and MacKay, 1994, von der Malsburg, 1973] and in continuous unsupervised learning as in principal-component-analysis networks [Oja, 1982]. In these scenarios normalization is necessary to prevent the excessive growth of synaptic efficacies that occurs when learning and neuronal activity are strongly coupled. This paper focuses on associative memory learning where this excessive synaptic runaway growth is mild [Massica and Ruppin, 1998], and shows that even in this more simple learning paradigm, normalization processes are essential. Moreover, while numerous normalization procedures can prevent synaptic runaway, our analysis shows that a specific neuronally-driven correction procedure that preserves the total sum of synaptic efficacies is essential for effective memory storage.

The following section describes the associative memory model and derives constraints on effective synaptic learning rules. Section 3 describes the main result of this paper, a neuronal weight correction mechanism that can modify synaptic efficacies towards maximization of memory capacity. Section 4 presents a biologically plausible realization of the neuronal normalization mechanism in terms of neuronal regulation. Our results are discussed in the last section.

2 Effective Synaptic Learning rules

We study the computational aspects of associative learning in a model of low-activity associative memory network with binary firing $\{0, 1\}$ neurons. M uncorrelated memory patterns

$\{\xi^\mu\}_{\mu=1}^M$ with coding level p (fraction of firing neurons) are stored in an N neurons network. The ith neuron updates its firing state X_i^t at time t by

$$X_i^{t+1} = \theta(f_i^t), \quad f_i^t = \frac{1}{N}\sum_{j=1}^{N} W_{ij} X_j^t - T, \quad \theta(f) = \frac{1 + sign(f)}{2} \quad , \tag{1}$$

where f_i is its input field (postsynaptic potential) and T is its firing threshold. The synaptic weight W_{ij} between the jth (presynaptic) and ith (postsynaptic) neurons is determined by a general additive synaptic learning rule depending on the neurons' activity in each of the M stored memory patterns ξ^η

$$W_{ij} = \sum_{\eta=1}^{M} A(\xi_i^\eta, \xi_j^\eta) \quad , \tag{2}$$

where $A(\xi_i^\eta, \xi_j^\eta)$ is a two-by-two synaptic learning matrix that governs the incremental modifications to a synapse as a function of the firing of the presynaptic (column) and postsynaptic (row) neurons

$A(\xi_i, \xi_j) =$ postsynaptic (ξ_i)

	presynaptic (ξ_j) 1	0
1	α	β
0	γ	δ

.

In conventional biological terms, α denotes an increment following a long-term potentiation (LTP) event, β denotes a heterosynaptic long-term depression (LTD) event, and γ a homosynaptic LTD event.

The parameters $\alpha, \beta, \gamma, \delta$ define a four dimensional space in which all linear additive Hebbian learning rules reside. However, in order to visualize this space, one may represent these Hebbian learning rules in a reduced, two-dimensional space utilizing a scaling invariance constraint and the requirement that the synaptic matrix should have a zero mean (otherwise the synaptic values diverge, the noise overshadows the signal term and no retrieval is possible [Dayan and Willshaw, 1991]). These yield the following rule, having two free parameters (x_D, x_P) only

$A(\xi_i, \xi_j) =$ postsynaptic (ξ_i)

	presynaptic (ξ_j) 1	0
1	x_P	x_D
0	c	$f(x_P, x_D, c)$

where c is a scaling constant and $f(x_P, x_D, c) = \frac{-1}{(1-p)^2}\left[p^2 x_P + p(1-p)(c + x_D)\right]$.

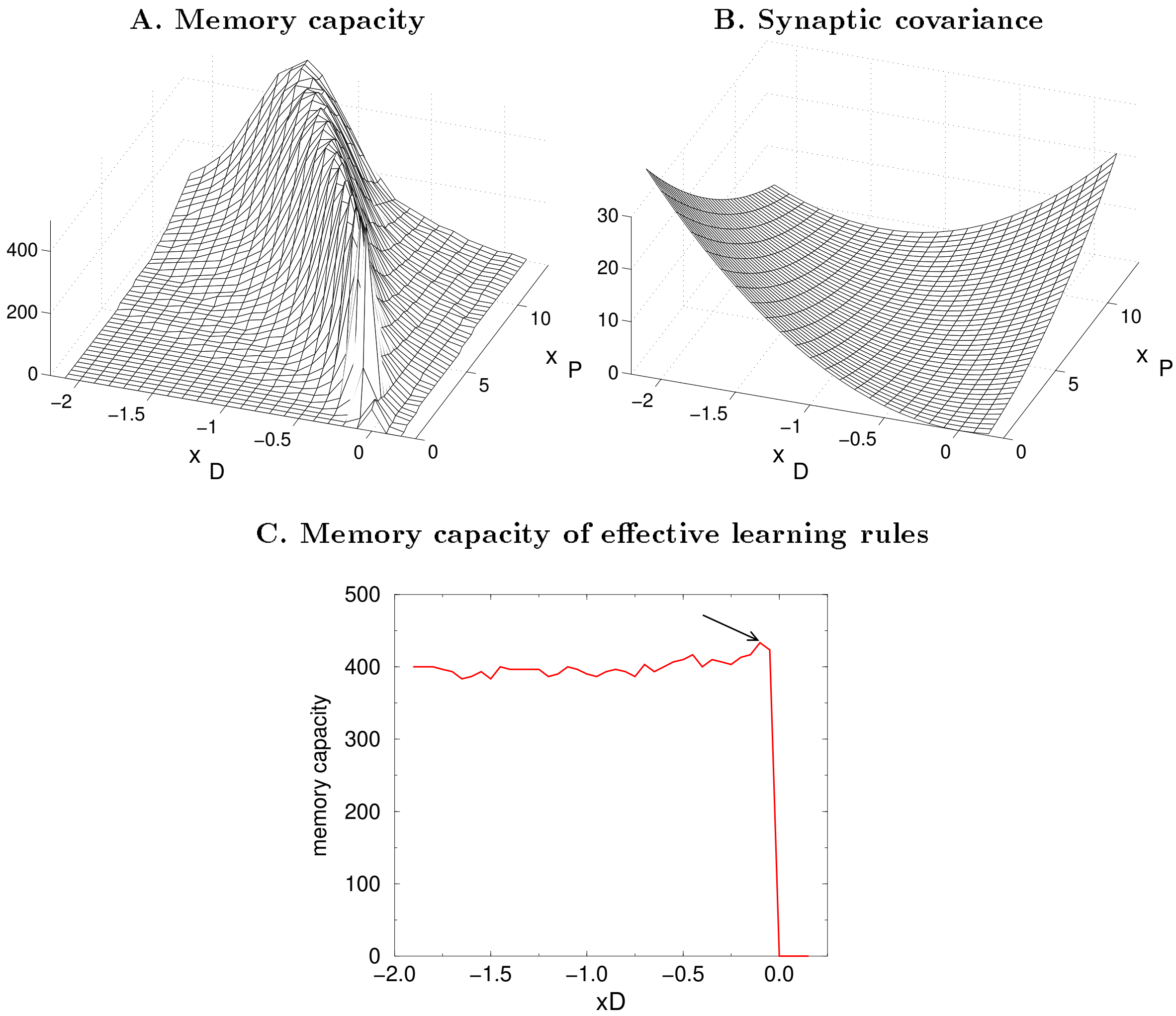

Figure 1: A. Memory capacity of a 1000-neurons network for different values of x_P and x_D as obtained in computer simulations. Capacity is defined as the maximal number of memories that can be retrieved with overlap bigger than $m = 0.95$ when presented with a degraded input cue with overlap $m_0 = 0.8$. The overlap m^η (or similarity) between the current network's activity pattern X and the memory pattern ξ^η serves to measure retrieval acuity and is defined as $m^\eta = \frac{1}{p(1-p)N}\sum_{j=1}^{N}(\xi_j^\eta - p)X_j$. The coding level is $p = 0.05$. B. Covariance between two synapses on the post synaptic neuron, as calculated analytically. Zero covariance is obtained when $x_P = x_D\frac{1-p}{-p}$. C. Memory capacity of the effective learning rules. The peak values on the ridge of Figure A, are displayed by tracing their projection on the x_D coordinate. The optimal learning rule $A(\xi_i, \xi_j) = (\xi_i - p)(\xi_j - p)$ calculated by [5] (marked with an arrow) performs only slightly better than other effective learning rules.

Figure 1A plots the memory capacity of the network as a function of the two free parameters x_P and x_D. It reveals that considerable memory storage may be obtained only along an essentially one dimensional curve, naturally raising the possibility of identifying an additional constraint on the relations between $(\alpha, \beta, \gamma, \delta)$. Such a constraint is revealed by a signal-to-noise analysis of the neuronal input field f_i during retrieval

$$\begin{aligned}\frac{Signal}{Noise} &= \frac{E(f_i|\xi_i = 1) - E(f_i|\xi_i = 0)}{\sqrt{Var(f_i)}} \propto \frac{\sqrt{N}}{\sqrt{Var\left[W_{ij}\right] + NpCOV\left[W_{ij}, W_{ik}\right]}} = \\ &= \frac{\sqrt{N/M}}{\sqrt{Var\left[A(\xi_i, \xi_j)\right] + NpCOV\left[A(\xi_i, \xi_j), A(\xi_i, \xi_k)\right]}}.\end{aligned} \tag{3}$$

As evident from equation (3), when the **postsynaptic covariance** $COV\left[A(\xi_i, \xi_j), A(\xi_i, \xi_k)\right]$ (determining the covariance between the incoming synapses of the postsynaptic neuron) is positive, the network's memory capacity is bounded, i.e., it does not scale with the network size. As the postsynaptic covariance is non negative (see appendix A), **effective learning rules** that obtain linear scaling of memory capacity as a function of the network's size require a vanishing postsynaptic covariance. Intuitively, when the synaptic weights are correlated, adding any new synapse contributes only little new information, thus limiting the number of beneficial synapses that help the neuron estimate whether it should fire or not. Figure 1B plots the postsynaptic covariance for the values of the free parameters (x_P, x_D) enumerated in Figure 1A, showing that it is indeed the covariance that limits the network performance with non-optimal learning rules. Figure 1C depicts the memory capacity of the effective synaptic learning rules that lie on the essentially one-dimensional ridge observed in Figure 1A. It shows that all these effective rules are only slightly inferior to the optimal synaptic learning rule calculated previously by [Dayan and Willshaw, 1991], which maximizes memory capacity.

In addition to the known zero-mean constraint on the neuron's input field, the vanishing covariance constraint on effective learning rules implies a new requirement concerning the balance between synaptic depression and facilitation: $x_D = \frac{-p}{1-p}\, x_P$. Thus, effective memory storage requires a delicate balance between LTP (x_P) and heterosynaptic depression (x_D). These two constraints make effective memory storage explicitly dependent on the coding level p which is a global property of the network. It is thus difficult to see how effective rules can be implemented at the synaptic level. Moreover, as shown in Figure 1A, Hebbian learning rules lack robustness as small perturbations from the effective rules may result in large decrease in memory capacity. Furthermore, it is important to note that these problems cannot be circumvented by introducing a nonlinear Hebbian learning rule of the form $W_{ij} = g\left(\sum_\eta A(\xi_i^\eta, \xi_j^\eta)\right)$ as even for a nonlinear function g the covariance $Cov\left[g(\sum_\eta A(\xi_i^\eta, \xi_j^\eta)), g(\sum_\eta A(\xi_i^\eta, \xi_k^\eta))\right]$ remains positive if $Cov(A(\xi_i, \xi_j), A(\xi_i, \xi_k))$ is positive (see Appendix B). These observations show that effective associative learning with Hebbian rules alone is implausible from a biological standpoint requiring locality of information.

3 Effective Learning via Neuronal Weight Correction

The above results show that in order to obtain effective memory storage, the post-synaptic covariance must be kept negligible. How then may effective storage take place in the brain with Hebbian learning? We now proceed to show that a neuronally-driven procedure (essentially similar to that assumed by [von der Malsburg, 1973, Miller and MacKay, 1994] to take place during self-organization) can maintain a vanishing covariance and enable effective memory storage by acting upon ineffective Hebbian synapses and turning them into effective ones.

3.1 The Neuronal Weight Correction Procedure

The solution emerges when rewriting the signal-to-noise equation (Eq. 3) as

$$\frac{Signal}{Noise} \propto \frac{N}{\sqrt{NVar\left[W_{ij}\right](1-p) + pVar(\sum_{j=1}^{N} W_{ij})}}. \tag{4}$$

showing that the post synaptic covariance can be greatly diminished when the variance of the sum of incoming synapses is vanishing. (see Eq (11) in appendix A) We thus propose that during learning, as a synapse is modified, its postsynaptic neuron additively modifies all its synapses to maintain the sum of their efficacies at a baseline zero level. As this **neuronal weight correction** is additive, it can be performed either after each memory pattern is stored or at a later time after several memories have been stored.

Interestingly, the joint operation of weight correction over a linear Hebbian learning rule is equivalent to the storage of the same set of memory patterns with another Hebbian learning rule. This new rule has a zero-covariance learning matrix, as follows

	1	0
1	α	β
0	γ	δ

$\Longrightarrow$

	1	0
1	$(\alpha-\beta)(1-p)$	$(\alpha-\beta)(0-p)$
0	$(\gamma-\delta)(1-p)$	$(\gamma-\delta)(0-p)$

.

To intuitively see this, focus on a firing neuron in the current memory pattern. When an LTP event occurs, the pertaining synaptic efficacy is strengthened by α, thus all other synaptic efficacies must be reduced by $\frac{\alpha}{N}$ to keep their sum fixed. As there are on average Np LTP events for each memory, all incoming synaptic efficacies will be reduced by αp. This and a similar calculation for quiescent neurons yields the synaptic learning matrix displayed on the right. It should be emphasized that the matrix on the right is not applied at the synaptic level but is the emergent result of the operation of the neuronal mechanism on the matrix on the left, and is used here as a mathematical tool to analyze network's performance. Thus, using a neuronal mechanism that maintains the sum of incoming synapses fixed enables the same level of effective performance as would have been achieved by using a zero-covariance Hebbian learning rule, but without the need to know the memories' coding level. Note also that neuronal weight correction applied to the matrix on the right will result in the same matrix, thus no further changes will occur with its re-application.

3.2 An Example

To demonstrate the beneficiary effects of neuronal weight correction we have applied it to a non-effective rule having non-zero covariance. In particular, we have investigated a common realization of the Hebb rule $A(\xi_i, \xi_j) = \xi_i \xi_j$ with inhibition added to obtain a zero-mean input field (otherwise the capacity vanishes) yielding $A(\xi_i, \xi_j) = \xi_i \xi_j - p^2$ [Tsodyks, 1989], or in matrix form

Zero-mean Hebb rule

	1	0
1	$1-p^2$	$-p^2$
0	$-p^2$	$-p^2$

.

As evident, this learning rule employs both homo-synaptic and hetero-synaptic LTD maintaining a zero mean synaptic matrix, but its postsynaptic covariance is non-zero and is thus still an ineffective rule. Applying neuronal weight correction to the synaptic matrix formed by this rule results in a synaptic matrix which is identical to the one generated without neuronal correction by the following rule

Neuronally corrected Hebb rule

	1	0
1	$1-p$	$-p$
0	0	0

which has both zero mean and zero postsynaptic covariance. Figure 2 plots the memory capacity obtained with the zero mean Hebb rule, before and after neuronal weight correction, as a function of the network's size. The memory capacity of the original zero-mean Hebb rule is essentially bounded, while after applying neuronal weight correction it scales linearly with the network's size.

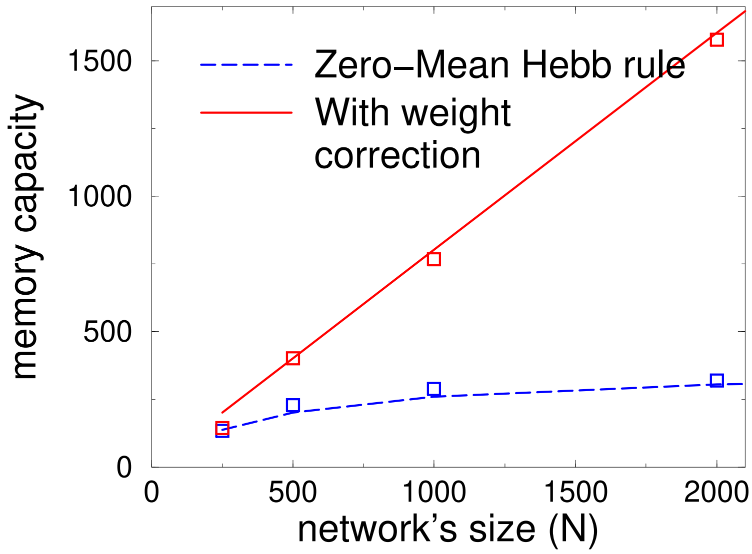

Figure 2: Network capacity as a function of network size. While the original zero-mean learning rule has bounded memory capacity, the capacity becomes linear in the network's size when the same learning rule is coupled with weight correction. The lines plot analytical results and the squares designate simulation results ($p = 0.05$).

Figure 3 shows that the beneficial effect of the neuronal correction remains marked for a wide range of coding level values p.

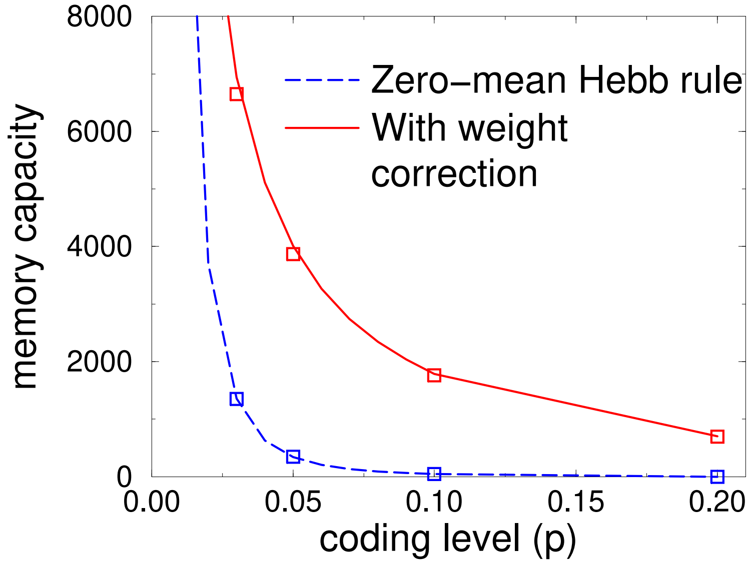

Figure 3: Comparison of network memory capacity for memory patterns with different values of the coding level p. All memories were stored in a network of $N = 5000$ neurons. The effect of neuronal correction is marked for a wide range of the p values, especially in the low coding levels observed in the brain. The lines plot analytical results and the squares designate simulation results.

4 Neuronal Regulation Implements Weight Correction

The proposed neuronal algorithm relies on the availability of explicit information about the total sum of synaptic efficacies at the neuronal level. Several mechanisms for conservation of the total synaptic strength have been proposed [Miller, 1996]. However, as explicit information on the synaptic sum may not be available, we turned to study the possibility that the total synaptic sum is regulated indirectly by estimating the neuronal average postsynaptic potential with a Neuronal Regulation (NR) mechanism [Horn *et al.*, 1998]. NR maintains the homeostasis of neuronal activity by regulating the postsynaptic activity (input field f_i) of the neuron around a fixed baseline. This homeostasis is achieved by multiplying the neuron's incoming synaptic efficacies by a common factor such that changes in the postsynaptic potential are counteracted by inverse changes in the synaptic efficacies. Such activity-dependent scaling of quantal amplitude of excitatory synapses, which acts to maintain the homeostasis of neuronal firing in a multiplicative manner, has already been observed in cortical tissues by [Turrigano *et al.*, 1998, Rutherford *et al.*, 1998]. These studies complement their earlier studies showing that neuronal postsynaptic activity can be kept at fixed levels via activity-dependent regulation of synaptic conductances [LeMasson *et al.*, 1993, Turriango *et al.*, 1994].

We have studied the performance of NR-driven correction in an excitatory-inhibitory memory model where excitatory neurons are segregated from inhibitory ones in the spirit of Dale's law [Horn *et al.*, 1998, Chechik *et al.*,]. This model is similar to our basic model, except that Hebbian learning takes place on the excitatory synapses

$$W_{ij} = \sum_{\eta=1}^{M} A(\xi_i^\eta, \xi_j^\eta) \quad , \tag{5}$$

with a learning matrix A that has a positive mean $E(A) = a$. The input field is now

$$f_i^t = \frac{1}{N} \sum_{j=1}^{N} W_{ij}^{excit} X_j^t - W_i^{inhib} \sum_{j=1}^{N} X_j^t \quad , \tag{6}$$

replacing the original term in Equation (1). When $W_{inhib} = Ma$, this model is mathematically equivalent to the model described above in Eqs. 2 - 1.

NR is performed by repeatedly activating the network with random input patterns, and letting each neuron estimate its input field. During this process, each neuron continuously gauges its average input field f_i^t around a zero mean by slowly modifying its incoming excitatory synaptic efficacies in accordance with

$$\kappa \frac{dW_{ij}^{excit}(t')}{dt'} = -W_{ij}^{excit}(t') f_i^t \ . \tag{7}$$

When all W^{excit} are close to a large mean value, multiplying all weights by a common factor approximates an additive change [1]. Figure 4 plots the memory capacity of networks storing memories according to the Hebb rule $W_{ij}^{excit} = \sum_{\eta=1}^{M} A(\xi_i^\eta, \xi_j^\eta) = \sum_{\eta=1}^{M} \xi_i^\eta \xi_j^\eta$, showing how

[1]The above learning rule results in synapses that are are normally distributed, $N(Mp^2, (\sqrt{M}p(1-p))^2)$, therefor all synapses reside relatively close to their mean when M is large. We may thus substitute $W_{ij}(t') = Mp^2 + \epsilon$ in Eq. (7) yielding $W_{ij}(t'+1) = W_{ij}(t') + \frac{d}{dt'} W_{ij}(t')/\kappa = (Mp^2 + \epsilon)(1 - f_i/\kappa)$. As f_i/κ and ϵ are small, this is well approximated by $W_{ij}(t') - Mp^2 f_i/k$

NR, which approximates the additive neuronal weight correction, succeeds in obtaining a linear growth of memory capacity as long as the inhibitory synaptic weights are close to the mean excitatory synaptic values (i.e., the zero synaptic mean constraint is obeyed).

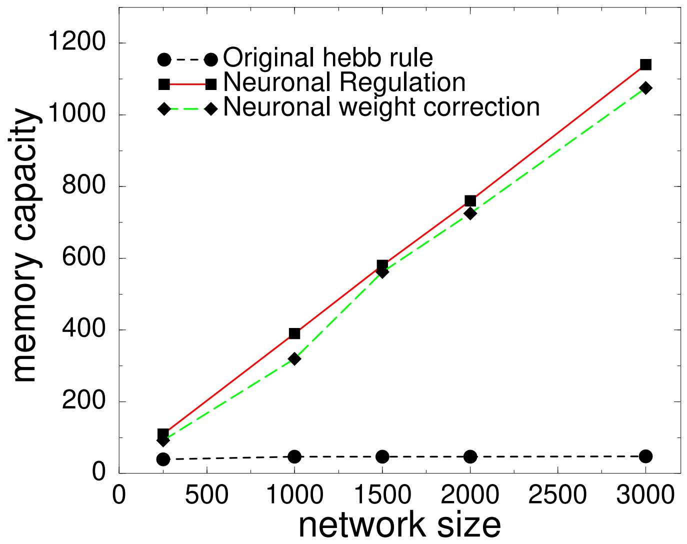

Figure 4: Memory capacity of networks storing patterns via the Hebb rule. Applying NR achieves a linear scaling of memory capacity with a slightly inferior capacity compared with that obtained with neuronal weight correction. Memory capacity is measured as in Figure 2, after the network has reached a stable state. W_i^{inhib} is normally distributed with a mean of $E(W^{excit}) = p^2M$ and a standard deviation of $0.1p^2M^{0.5}$, where $p = 0.1$.

Figure 5 plots the temporal evolution of the retrieval acuity (overlap) and the average postsynaptic covariance, showing that NR slowly removes the interfering covariance, improving memory retrieval.

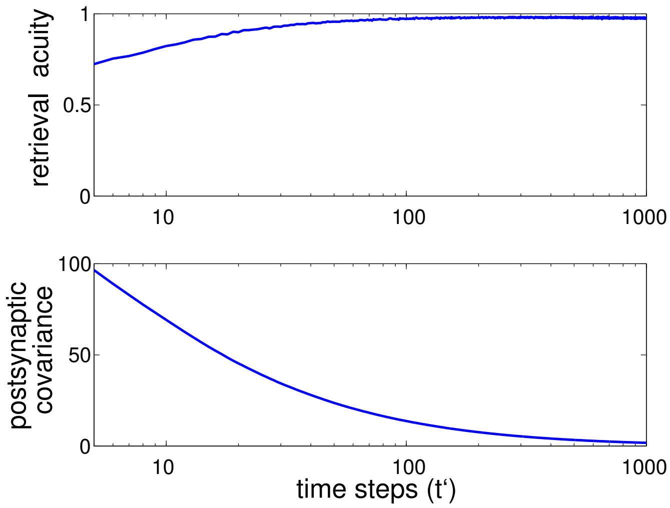

Figure 5: The temporal evolution of retrieval acuity and average postsynaptic covariance in a 1000-neurons network. 250 memories are first stored in the network using the Hebb rule, resulting in a poor retrieval acuity ($m \approx 0.7$ at $t = 0$ in the upper figure). However, as NR is iteratively applied to the network, the retrieval acuity gradually improves as the post-synaptic covariance vanishes. $p = 0.1$, $\kappa = 0.1$, other parameters as in Figure 4.

5 Discussion

The characterization of effective synaptic learning rules reopens the discussion of the computational role of heterosynaptic and homosynaptic depression. Previous studies have shown that long-term synaptic depression is necessary to prevent saturation of synaptic values [Sejnowski, 1977], and to maintain zero mean synaptic efficacies [Willshaw and Dayan, 1990]. Our study shows that proper heterosynaptic depression is needed to enforce zero postsynaptic covariance - an essential prerequisite of effective learning. The zero covariance constraint implies that the magnitude of heterosynaptic depression should be smaller than that of homosynaptic potentiation by a factor of $(1 - p)/p$. However, effective learning can be obtained regardless of the magnitude of the homosynaptic depression changes, as long as the zero mean constraint stated above is satisfied.

The terms potentiation/depression used in the above context should be cautiously interpreted: As neuronal weight normalization may modify synaptic efficacies in the brain, the apparent changes in synaptic efficacies measured in LTD/LTP experiments may involve two kinds of processes: Synaptic-driven processes, changing synapses according to the covariance between pre and post synaptic neurons, and neuronally-driven processes, operating to zero the covariance between incoming synapses of the neuron. Although our analysis pertains to the combined effect of these processes, they may be experimentally segregated as they operate on different time scales and modify different ion channels ([Bear and Abraham, 1996, Turrigano *et al.*, 1998]). Thus, the relative weights of neuronal versus synaptic processes can be experimentally tested by studying the temporal changes in synaptic efficacy following LTP/LTD events, and comparing them with the theoretically predicted potentiation and depression end values.

This paper highlights the role of neuronally-driven synaptic plasticity in remodeling synaptic efficacies during learning. Our findings show that the combined action of synaptic-specific and neuronally-guided synaptic modifications yields a robust learning system. This allows for the usage of biologically feasible but ineffective synaptic learning rules, as long as they are further modified and corrected by neurally driven weight correction. Several forms of synaptic constraints were previously suggested in the literature to improve the stability of Hebbian learning - such as preserving the sum of synaptic strengths or the sum of their squares [von der Malsburg, 1973, Oja, 1982]. Our analysis shows that in order to obtain effective memory storage it is the sum of synaptic strengths which must be preserved, thus predicting that it is this specific form of normalization that occurs in the brain. While the previously suggested normalization techniques can be well approximated by synaptic rules, associative memory learning requires neuronally-driven processes that govern the synaptic changes.

Our results, obtained within the paradigm of autoassociative memory networks, apply also to hetero-associative memory networks. More generally, neuronal weight correction qualitatively improves the ability of a neuron to correctly discriminate between a large number of input patterns. It thus enhances the computational power of the single neuron and may be applied in other learning paradigms. This interplay between cooperative and competitive synaptic changes is likely to plays a fundamental computational role in a variety of brain functions such as visual processing and associative learning.

Appendices

A Signal-To-Noise Calculation Of The Neuronal Input Field For A General Learning matrix

We calculate the signal-to-noise ratio of a network storing memory patterns according to a learning matrix A with zero mean $E(A(\xi_i, \xi_j)) = 0$. The network is initialized in a state with overlap m_0 with memory pattern ξ^1 and with activity p (the overlap with the other patterns is assumed to be negligible). Let $\epsilon = P(X_i = 0|\xi_i = 1)$ implying an initial overlap of $m_0 = \frac{(1-p-\epsilon)}{(1-p)}$. Denoting $W^*_{ij} = W_{ij} - A(\xi^1_i, \xi^1_j)$ the conditional mean of the neuron input field is

$$
\begin{aligned}
E\left[f_i|\xi^1_i\right] &= E\left[\frac{1}{N}\sum_{j=1}^{N} W_{ij}X_j|\xi^1_i\right] = \\
&= E\left[\frac{1}{N}\sum_{j=1}^{N} A(\xi^1_i,\xi^1_j)X_j|\xi^1_i\right] + E\left[\frac{1}{N}\sum_{j=1}^{N} W^*_{ij}X_j|\xi^1_i\right] = \\
&= E\left[\frac{1}{N}\sum_{j=1}^{N} A(\xi^1_i,\xi^1_j)X_j|\xi_i\right] = \\
&= A(\xi^1_i, 1)P(X_j = 1|\xi_j = 1)P(\xi_j = 1) + \\
&\quad + A(\xi^1_i, 0)P(X_j = 1|\xi^1_j = 0)P(\xi^1_j = 0) = \\
&= A(\xi^1_i, 1)(1-\epsilon)p + A(\xi^1_i, 0)\epsilon p \quad .
\end{aligned}
\tag{8}
$$

The variance is

$$
\begin{aligned}
Var\left[f_i|\,\xi^1_i\right] &= Var\left[\frac{1}{N}\sum_{j=1}^{N} W_{ij}X_j\right] \approx Var\left[\frac{1}{N}\sum_{j=1}^{N} W^*_{ij}X_j\right] = \\
&= \frac{1}{N}Var\left[W^*_{ij}X_j\right] + COV\left[W^*_{ij}X_j, W^*_{ik}X_k\right] = \\
&= \frac{M}{N}p\, Var\left[A(\xi_i,\xi_j)\right] + Mp^2\, Cov\left[A(\xi_i,\xi_j), A(\xi_i,\xi_k)\right] \quad .
\end{aligned}
\tag{9}
$$

The neuronal field's noise is thus dominated by the covariance between its incoming synaptic weights. The signal-to-noise ratio of the neurons input field is

$$
\begin{aligned}
\frac{Signal}{Noise} &= \frac{E(f_i|\xi_i = 1) - E(f_i|\xi_i = 0)}{\sqrt{Var(f_i|\xi_i)}} = \\
&= \sqrt{\frac{N}{M}}\frac{[A(1,1) - A(0,1)](1-\epsilon) + [A(1,0) - A(0,0)]\epsilon}{\sqrt{Var\left[A(\xi_i,\xi_j)\right] + NpCov\left[A(\xi_i,\xi_j), A(\xi_i,\xi_k)\right]}}.
\end{aligned}
\tag{10}
$$

When the postsynaptic covariance is zero, the signal-to-noise ratio remains constant as M grows linearly with N, thus implying a linear memory capacity. However, when the covariance is a positive constant, the term on the right is almost independent of N, and the memory capacity is bounded. The variance can also be presented as

$$\begin{aligned}
Var\left[f_i|\xi_i^1\right] &\approx Var\left[\frac{1}{N}\sum_{j=1}^{N} W_{ij}^* X_j\right] = \qquad (11)\\
&= \frac{1}{N^2}\left[\sum_{j=1}^{N}\sum_{k=1,k\neq j}^{N} Cov(W_{ij}^*, W_{ik}^*)\right] E^2(X_j) + \frac{1}{N^2}\left[\sum_{j=1}^{N} Var(W_{ij}^*)\right] E(X_j) = \\
&= \frac{p^2}{N^2} Var(\sum_{j=1}^{N} W_{ij}^*) - \frac{p^2}{N} Var(W_{ij}^*) + \frac{p}{N} Var(W_{ij}^*) p = \\
&= \frac{1}{N} p(1-p) Var\left[W_{ij}^*\right] + \frac{p^2}{N^2} Var(\sum_{j=1}^{N} W_{ij}^*) \quad ,
\end{aligned}$$

thus, keeping the sum of the incoming synapses fixed results in a beneficial effect similar to that of removing postsynaptic covariance, and further improves the signal-to-noise ratio by a factor of $\frac{1}{\sqrt{1-p}}$. The postsynaptic sum ($\sum W_{ij}^*$) remains fixed if each memory pattern has exactly pN firing neurons out of the N neurons of the network.

The covariance of a learning matrix with scaling invariant $c = 1$ and zero mean is

$$Cov\left[A(\xi_i, \xi_j), A(\xi_i, \xi_k)\right] = \frac{p}{(1-p)}\left[p\, x_P + (1-p) x_D\right]^2 \qquad (12)$$

that is always non-negative and equals zero only when

$$x_D = \frac{-p}{1-p}\, x_P \quad . \qquad (13)$$

B The Postsynaptic Covariance of Non-Additive Learning Rules

In this section we show that the postsynaptic covariance cannot be zeroed by introducing a non-additive learning rules of the form

$$W_{ij} = g\left(\sum_{\eta=1}^{M} A(\xi_i^\eta, \xi_j^\eta)\right) \qquad (14)$$

for some nonlinear function g.

To show that, note that when X, Y are positively correlated random variables with marginal standard normal distribution and $E(g(X)) = 0$, we can write (using independent normally distributed random variables U, V, W)

$$
\begin{aligned}
E[g(X)g(Y)] &= E[g(U+V)g(W+V)] = \\
&= E[E(g(U+V)g(W+V)|V)] = \\
&= E\left[E(g(U+V)|V)^2\right] = \\
&= Var[E(g(U+V)|V)] \geq 0.
\end{aligned} \tag{15}
$$

Equality holds only when $\phi(v) = E(g(U+V)|V = v) = E(g(U+v))$ is a constant function, as such it must be zero because $E(g(X)) = 0$. To further show that the equality holds only when g is constant, we look at

$$
\begin{aligned}
0 &= E(g(U+v)) = \\
&= \int \frac{1}{\sqrt{2\pi}\sigma} g(v+u) e^{\frac{-u^2}{2\sigma^2}} du = \\
&= e^{-\frac{v^2}{2\sigma^2}} \frac{1}{\sqrt{2\pi}\sigma} \int e^{\frac{vt}{\sigma^2}} g(t) e^{-\frac{t^2}{2\sigma^2}} dt \quad ,
\end{aligned} \tag{16}
$$

or $\Psi(v) = \int e^{\frac{vt}{\sigma^2}} g(t) e^{-\frac{t^2}{2\sigma^2}} dt \equiv 0$. As $\Psi(v) = 0$ is the Laplace transform of $g(t)e^{-\frac{t^2}{2\sigma^2}}$, g is almost everywhere uniquely determined and the solution $g = 0$ is essentially the only solution.

References

[Bear and Abraham, 1996] M.F. Bear and W.C. Abraham. Long term depression in hippocampus. *Annu. Rev. Neurosci.*, 19:437–462, 1996.

[Bliss and Collingridge, 1993] T.V.P. Bliss and G.L. Collingridge. Syanptic model of memory: long-term potentiation in the hippocampus. *Nature*, 361:31–39, 1993.

[Chechik *et al.*,] G. Chechik, I. Meilijson, and E. Ruppin. Neuronal regulation: A mechanism for synaptic pruning during brain maturation. *Neural Computation*, In press.

[Dayan and Willshaw, 1991] P. Dayan and D.J. Willshaw. Optimizing synaptic learning rules in linear associative memories. *Biol. Cyber.*, 65:253, 1991.

[Horn *et al.*, 1998] D. Horn, N. Levy, and E. Ruppin. Synaptic maintenance via neuronal regulation. *Neural Computation*, 10(1):1–18, 1998.

[LeMasson *et al.*, 1993] G. LeMasson, E. Marder, and L.F. Abbott. Activity-dependent regulation of conductances in model neurons. *Science*, 259:1915–1917, 1993.

[Massica and Ruppin, 1998] A. Grinstein Massica and E. Ruppin. Synaptic runaway in associative networks and the pathogenesis of schizophrenia. *Neural Computation*, 10:451–465, 1998.

[Miller and MacKay, 1994] K.D. Miller and D.J.C MacKay. The role of constraints in Hebbian learning. *Neural Computation*, 6(1):100–126, 1994.

[Miller, 1996] K.D. Miller. Synaptic economics: Competition and cooperation in synaptic plasticity. *Neuron*, 17:371–374, 1996.

[Oja, 1982] Erkki Oja. A simplified neuron model as a principal component analyzer. *Journal of Mathematical Biology*, 15:267–273, 1982.

[Rutherford *et al.*, 1998] L.C. Rutherford, S.B. Nelson, and G.G. Turrigiano. BDNF has opposite effects on the quantal amplitude of pyramidal neuron and interneuron excitatory synapses. *Neuron*, 21:521–530, 1998.

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

[Tsodyks, 1989] M.V. Tsodyks. Associative memory in neural networks with Hebbian learning rule. *Modern Physics letters*, 3(7):555–560, 1989.

[Turriango *et al.*, 1994] G.G. Turriango, L.F. Abbott, and E. Marder. Activity-dependent changes in the intrinsic properties of neurons. *Science*, 264:974–977, 1994.

[Turrigano *et al.*, 1998] G.G. Turrigano, K. Leslie, N. Desai, and S.B. Nelson. Activity dependent scaling of quantal amplitude in neocoritcal pyramidal neurons. *Nature*, 391(6670):892–896, 1998.

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

[Willshaw and Dayan, 1990] D.J. Willshaw and P. Dayan. Optimal plasticity from matrix memories: What goes up must come down. *Neural Computation*, 2(1):85–93, 1990.