# Effective Learning Requires Neuronal Remodeling of Hebbian Synapses

**Gal Chechik  Isaac Meilijson  Eytan Ruppin** *
School of Mathematical Sciences
Tel-Aviv University Tel Aviv, Israel
ggal@math.tau.ac.il isaco@math.tau.ac.il ruppin@math.tau.ac.il

**Submitted for oral presentation**

**Category: Neuroscience**

## Abstract

This paper revisits 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 requirement is difficult to achieve in a robust manner by Hebbian synaptic learning, since it depends on network level information. Effective learning can yet be obtained 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.

## 1 Introduction

Synapse-specific changes in synaptic efficacies, carried out by long-term potentiation (LTP) and depression (LTD) 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 showing for the first time that **synaptic learning alone cannot provide effective associative learning in a biologically plausible manner, and must be complemented with neuronally-driven synaptic remodeling**.

The importance of neuronally driven normalization processes has already been

*To whom correspondence should be addressed. (Also affiliated with the Sackler school of medicine)

demonstrated in the context of self-organization of cortical maps [1, 2] and in continuous unsupervised learning as in principal-component-analysis networks [3]. In these scenarios normalization is necessary to prevent the excessive growth of synaptic efficacies that occurs when learning and neuronal activity are strongly coupled. In contradistinction, this paper focuses on associative memory learning where this excessive synaptic runaway growth is mild [4], and shows that even in this simple learning paradigm, normalization processes are essential. Moreover, while numerous normalization procedures can prevent synaptic runaway, our analysis shows that only a specific neuronally-driven correction procedure that preserves the total sum of synaptic efficacies leads to effective associative memory storage.

## 2 Effective Synaptic Learning rules

We study the computational aspects of associative learning in a model of a 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 $i$th neuron updates its 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 $\xi^\eta$

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

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

$$A(\xi_i, \xi_j) \quad = \quad \text{postsynaptic } (\xi_i) \begin{array}{c} \text{presynaptic } (\xi_j) \\ \begin{array}{|c||c|c|} \hline & 1 & 0 \\ \hline 1 & \alpha & \beta \\ \hline 0 & \gamma & \delta \\ \hline \end{array} \end{array} \quad .$$

In conventional biological terms, $\alpha$ denotes an increment following a long-term potentiation (LTP) event, $\beta$ denotes heterosynaptic long-term depression (LTD), and $\gamma$ 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 [5]). These constraints yield the following rule, having two free parameters $(x_D, x_P)$ only

$$A(\xi_i, \xi_j) \quad = \quad \text{postsynaptic } (\xi_i) \begin{array}{c} \text{presynaptic } (\xi_j) \\ \begin{array}{|c||c|c|} \hline & 1 & 0 \\ \hline 1 & x_P & x_D \\ \hline 0 & c & f(x_P, x_D, c) \\ \hline \end{array} \end{array}$$

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

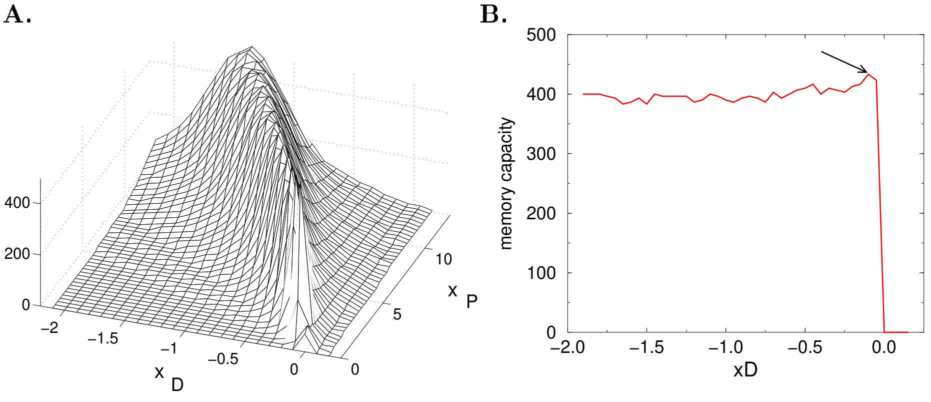


Figure 1: **A.** Memory capacity of a 1000-neurons network with $p = 0.05$ for different values of $x_P$ and $x_D$ as obtained in computer simulations. Capacity is defined as the maximal number of memories that can be retrieved with overlap bigger than 0.95 when presented with a degraded input cue with overlap 0.8. The overlap serves to measure retrieval acuity and is defined as $m^\eta = \frac{1}{p(1-p)N}\sum_{j=1}^{N}(\xi_j^\eta - p)X_j$. **B.** 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 $A(\xi_i, \xi_j) = (\xi_i - p)(\xi_j - p)$ [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

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

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, **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 depicts the memory capacity of the effective synaptic learning rules which 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 [5], which maximizes memory capacity.

The new 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 ($x_D$), and is strongly 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. **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 [2, 1] to occur during self-organization) can maintain a vanishing covariance and turn ineffective Hebbian synapses into effective ones. This enables the brain to utilize inefficient learning rules which use local information only, but still attain effective learning capabilities.

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. We thus propose the following **neuronal weight correction** procedure: During learning, whenever a synapse is modified, its postsynaptic neuron additively modifies all its synapses to maintain the sum of their efficacies at a baseline zero level.

$$W_{ij} \Longrightarrow W_{ij} - \frac{1}{N}\sum_{j=1}^{N} W_{ij} \quad ; \quad \forall j = 1..N \tag{5}$$

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. We prove that this new rule has a zero-covariance learning matrix

$$\begin{array}{|c||c|c|} \hline & 1 & 0 \\ \hline\hline 1 & \alpha & \beta \\ \hline 0 & \gamma & \delta \\ \hline \end{array} \quad \Longrightarrow \quad \begin{array}{|c||c|c|} \hline & 1 & 0 \\ \hline\hline 1 & (\alpha-\beta)(1-p) & (\alpha-\beta)(0-p) \\ \hline 0 & (\gamma-\delta)(1-p) & (\gamma-\delta)(0-p) \\ \hline \end{array} .$$

It should be reemphasized 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.

To demonstrate the beneficiary effects of neuronal weight correction we have first applied it to a common realization of the Hebb rule with inhibition added to obtain a zero-mean input field (otherwise the capacity vanishes) yielding $A(\xi_i, \xi_j) = \xi_i \xi_j - p^2$ [6]. Even though this learning rule has a zero mean synaptic matrix, its postsynaptic covariance is non-zero and is thus still an ineffective rule. Applying neuronal weight correction after learning with the above rule, results in a synaptic matrix which is identical to the one generated by the rule $A(\xi_i, \xi_j) = \xi_i(\xi_j - p)$ without neuronal weight correction, which has both zero mean and zero postsynaptic covariance. Figure 2A 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. After applying neuronal weight correction the originally bounded capacity turns to scale linearly with the network's size.

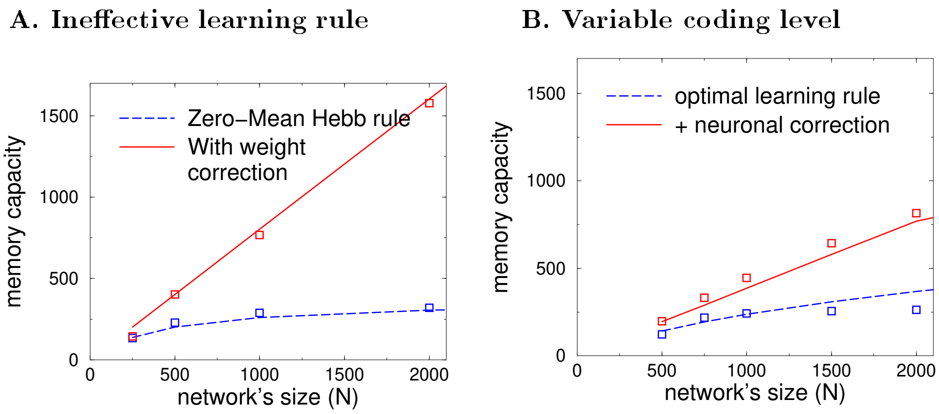


Figure 2: Network memory capacity as a function of network's size. **A.** While the original zero-mean learning rule has bounded memory capacity, the capacity becomes linear in the network's size when the same learning rule is coupled with neuronal-weight-correction. The lines plot analytical results and the squares designate simulation results ($p = 0.05$). **B.** Even the optimal learning rule becomes ineffective when the stored patterns have variable coding levels (coding levels are normally distributed $N(0.1, 0.02^2)$), but neuronal-weight-correction provides successful memory storage of such patterns. Results were obtained in a computer simulations.

As the effectiveness of the learning rule depends on the coding level of the stored patterns, all learning rules turn ineffective when the coding levels of the stored patterns are varying (that is, not all the patterns share exactly the same coding level). Figure 2B compares the memory capacity of a network that uses the optimal learning rule ($A(\xi_i, \xi_j) = (\xi_i - p)(\xi_j - p)$) for a coding level of $p = 0.1$ but actually stores memory patterns with coding levels that are normally distributed around 0.1. Only the application of neuronal weight correction provides effective storage of such patterns while the optimal learning rule does not.

## 4 Neuronal Regulation Implements Weight Correction

Like previous normalization procedures, the proposed neuronal algorithm relies on the availability of explicit information about the total sum of synaptic efficacies at the neuronal level. However, as explicit information on the synaptic sum may not be available, several mechanisms for conservation of the total synaptic strength have been proposed (see [7] for a review). In this paper we focus on one such mechanism, Neuronal Regulation (NR), where the **total synaptic sum is regulated indirectly by estimating the neuron's average postsynaptic potential**. NR is a slow process, continuously modifying synaptic efficacies to maintain the homeostasis of neuronal activity. Such activity-dependent scaling of excitatory synapses, which acts to maintain the homeostasis of neuronal firing, has already been observed in cortical tissues by [8].

We have studied the operation of NR-driven correction compared with additive neuronal weight correction in an excitatory-inhibitory network. Figure 3 plots the memory capacity of networks storing memories according to the Hebb rule, showing how NR approximates the additive neuronal weight correction and succeeds in obtaining a linear growth of memory capacity.

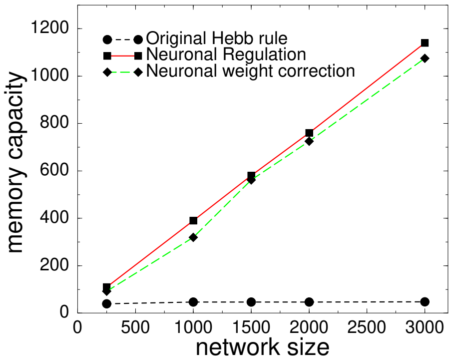


**Figure 3.** Applying NR achieves a linear scaling of memory capacity with a slightly inferior capacity compared with that obtained with neuronal weight correction. Memory patterns were stored according to the Hebb rule $W_{ij} = \sum_{\eta=1}^{M} \xi_i^\eta \xi_j^\eta$.

## 5 Summary

In this paper we have analyzed Hebbian learning rules in associative memory network models, and identified an essential requirement for effective memory storage: a vanishing postsynaptic covariance. We show that this constraint depends on the coding level of the stored memory patterns, thus requiring the use of network level information at the synaptic level. Moreover, when the stored memory patterns do not all share exactly the same coding level, there is no single learning rule that can effectively store all patterns. We further show that applying a neuronally driven mechanism that preserves the total synaptic sum zeroes the catastrophic covariance and provides effective learning even for ineffective synaptic learning rules. **The resulting improvement in memory capacity is drastic: learning rules yielding bounded capacity are transformed into learning rules yielding linear memory capacity as a function of the network's size**. Finally, the normalization mechanism can be carried out by neuronal regulation (NR), a mechanism recently identified in mammalian cortical cultures.

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 [9], and to maintain zero mean synaptic efficacies [10]. Our study shows that effective learning requires proper heterosynaptic depression, but can be obtained regardless of the homosynaptic depression magnitude. The terms potentiation/depression used in the above context should be cautiously interpreted, as 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. These processes may be experimentally segregated as they operate on different time scales ([11, 8]), and their relative weights can be experimentally tested.

While several forms of synaptic constraints were suggested to improve the stability of Hebbian learning [2, 3], our analysis shows that effective memory storage requires that the sum of synaptic strengths which must be preserved, thus predicting that it is this specific form of normalization that occurs in the brain. The utilization of the simple McCullough-Pitts model studied here has enabled us to gain analytical insight to the phenomena in hand. Recent findings of neuronal weight normalization in spiking models [12], lead us to believe that these results will also extent to spiking neurons' networks.

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 is likely to play a fundamental computational role in a variety of brain functions such as perceptual processing and associative learning.

## References

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

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

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

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

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

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

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

[8] 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.

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

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

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

[12] R. Kempter, W. Gergstner, and J.L. van Hemmen. Spike based compared to rate based hebbian learning. *Proceedings of NIPS-98*, 1999.