Synaptic Pruning in Development: A Computational account

Gal Chechik and Isaac Meilijson
School of Mathematical Sciences
ggal@math.tau.ac.il isaco@math.tau.ac.il
Tel-Aviv University Tel Aviv 69978, Israel
Eytan Ruppin *
Schools of Medicine and Mathematical Sciences
Tel-Aviv University Tel Aviv 69978, Israel
ruppin@math.tau.ac.il

December 2, 1997

Abstract

Research in humans and primates shows that the developmental course of the brain involves synaptic over-growth followed by marked selective pruning. Previous explanations have suggested that this intriguing, seemingly wasteful, phenomenon is utilized to remove 'erroneous' synapses. We prove that this interpretation is wrong if synapses are Hebbian. Under limited metabolic energy resources restricting the amount and strength of synapses, we show that memory performance is maximized if synapses are first overgrown and then pruned following optimal "minimal-value" deletion. This optimal strategy leads to interesting insights concerning childhood amnesia.

*To whom correspondence should be addressed.

1 Introduction

One of the fundamental phenomena in brain development is the reduction in the amount of synapses that occurs between early childhood and puberty. In recent years, many studies have investigated the temporal course of changes in synaptic density in primates, revealing the following picture. Beginning at early stages of the fetal development, synaptic density rises at a constant rate, until a peak level is attained (at 2-3 years of age in humans). Then, after a relatively short period of stable synaptic density (until the age of 5 in humans), an elimination process begins: synapses are being constantly removed, yielding a marked decrease in synaptic density. This process proceeds until puberty, when synaptic density stabilizes at adult levels which are maintained until old age. The peak level of synaptic density in childhood is $50\% - 100\%$ higher than adult levels, depending on the brain region. The phenomenon of synaptic over-growth and pruning was found in humans [Huttenlocher, 1979, Huttenlocher *et al.*, 1982, Huttenlocher and Courten, 1987], as well as in other mammals such as monkeys [Eckenhoff and Rakic, 1991, Bourgeois and Rakic, 1993, Bourgeois, 1993, Rakic *et al.*, 1994], cats [Innocenti, 1995] and rats [J.Takacs and Hamori, 1994]. It was observed throughout widespread brain regions including cortical areas (visual [Bourgeois and Rakic, 1993, Huttenlocher *et al.*, 1982], motor and associative [Huttenlocher, 1979]), the cerebellum [J.Takacs and Hamori, 1994], projection fibers between hemispheres [Innocenti, 1995] and the dentate gyrus [Eckenhoff and Rakic, 1991]. The time scale of synaptic elimination was found to vary between different cortical areas, coarsely following a dorsal to frontal order [Rakic *et al.*, 1994]. The changes in synaptic density are not a result of changes in total brain volume, but reflect true synaptic elimination [Rakic *et al.*, 1994]. In some cases, synaptic elimination was shown to be correlated with experience-dependent activity [Stryker, 1986, Roe *et al.*, 1990].

What advantage could such a seemingly wasteful developmental strategy offer? Some researchers have treated the phenomenon as an inevitable result of synaptic maturation, lacking any computational significance. Others have hypothesized that synapses which are strengthened at an early stage might be later revealed as harmful to overall memory function, when additional memories are stored. Thus, they claimed, synaptic elimination may reduce the interference between memories, and yield better overall performance [Wolff *et al.*, 1995].

This paper shows that in associative memory networks models these previous explanations do not hold, and puts forward a different explanation. Our proposal is based on the assumption that synapses are a costly resource whose efficient utilization is a major optimization goal guiding brain development. This assumption is motivated by the observation that the changes in synaptic density along brain development are highly correlated to the temporal course of changes in energy consumption [Roland, 1993], and by the fact that the brain consumes a large fraction of total energy consumption of the resting adult [Roland, 1993]. By analyzing the network's performance under various synaptic constraints such as limited number of synapses or limited total synaptic strength, we show that if synapses are properly pruned, the performance decrease due to synaptic deletion is small compared to the energy saving. Deriving optimal synaptic pruning strategies, we show that efficient memory storage in the brain requires a specific learning process characterized by initial synaptic over-growth followed by judicious synaptic pruning.

The next section describes the models studied and our analytical results, which are verified numerically in section 3. Section 4 discusses the possible benefits of efficient synaptic elimination, and its implications to the phenomenon of childhood amnesia.

2 Analytical Results

In order to investigate synaptic elimination, we address the more general question of optimal modification of a Hebbian memory matrix. Given previously learned Hebbian synapses we apply a function which changes the synaptic values, and investigate the effect of such a modification function. First, we analyze the way the memory performance depends on a general synaptic modification function. Then, we proceed to derive optimal modification functions under different constraints. Finally, we calculate the dependency of performance on the deletion levels.

2.1 The Models

We investigate synaptic modification in two Hebbian models. The first model is a variant of the canonical Hopfield model. M memories are stored in a N-neuron network forming approximate fixed points of the network dynamics. The initial synaptic efficacy W_{ij} between the jth (pre-synaptic) neuron and the ith (post-synaptic) neuron is

$$W_{ij} = \frac{1}{\sqrt{M}} \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu , \quad 1 \le i \ne j \le N; \quad W_{ii} = 0 \quad , \tag{1}$$

where $\{\xi^\mu\}_{\mu=1}^M$ are ± 1 binary patterns representing the stored memories. The actual synaptic efficacy J_{ij} is

$$J_{ij} = g(W_{ij}) \quad 1 \le i \ne j \le N; \quad J_{ii} = 0 \quad , \tag{2}$$

where g is a general modification function over the Hebbian weights, such that $g(z)$ has finite moment if z is normally distributed. The updating rule for the state X_i^t of the ith neuron at time t is

$$X_i^{t+1} = \theta(f_i), \quad f_i = \sum_{j=1}^{N} J_{ij} X_j^t \quad , \tag{3}$$

where f_i is the neuron's input field, and θ is the function $\theta(f) = sign(f)$. The overlap m^μ (or similarity) between the network's activity pattern X and the memory ξ^μ is $m^\mu = \frac{1}{N}\sum_{j=1}^{N} \xi_j^\mu X_j$.

The second model is a variant of the low activity biologically-motivated model described by [Tsodyks and Feigel'man, 1988], in which synaptic efficacies are described by

$$J_{ij} = g(W_{ij}) = g\left(\frac{1}{p(1-p)\sqrt{M}} \sum_{\mu=1}^{M} (\xi_i^\mu - p)(\xi_j^\mu - p)\right) . \tag{4}$$

where ξ^μ are $\{0,1\}$ memory patterns with coding level p (fraction of firing neurons), and g is a synaptic modification function. The updating rule is similar to Eq.(3),

$$X_i^{t+1} = \theta(f_i), \quad f_i = \sum_j J_{ij} X_j^t - T \, , \tag{5}$$

where θ now denotes the step function $\theta(f) = \frac{1+sign(f)}{2}$, and T is the neuronal threshold, set to its optimal value (see Eq. (26) in the appendix). The overlap m^μ in this model is defined by $m^\mu = \frac{1}{Np(1-p)}\sum_{j=1}^{N}(\xi_j^\mu - p)X_j$.

2.2 Pruning Does Not Improve Performance

To evaluate the impact of synaptic pruning on the network's retrieval performance, we study its effect on the signal to noise ratio (S/N) of the neuron's input field in the modified Hopfield model (Eqs. 2,3). The S/N is known to be the primary determinant of retrieval capacity (ignoring higher order correlations in the neurons input fields) [Meilijson and Ruppin, 1996], and is calculated by analyzing the moments of the neuron's field. The network is started at a state X with overlap m^μ with a specific memory ξ^μ; the overlap with other memories is assumed to be negligible. Therefore $W_{ij} - \frac{\xi_i^\mu \xi_j^\mu}{\sqrt{M}}$ is the sum of $M-1$ independent variables with zero expectation and standard variation $\sqrt{M}$ and is distributed $N(0,1)$. Denoting $\phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$, we use the fact that $\phi'(x) = -x\phi(x)$ and write

$$\begin{aligned} E\left[f_i|\xi_i^\mu\right] &= NE[g(W_{ij})X_j] = Nm^\mu E\left[g(W_{ij})\xi_j^\mu\right] = \\ &= Nm^\mu \frac{1}{2} E\left[g(W_{ij})|\xi_j^\mu = +1\right] - Nm^\mu \frac{1}{2} E\left[g(W_{ij})|\xi_j^\mu = -1\right] = \\ &= Nm^\mu \frac{1}{2}\int_{-\infty}^{\infty} g(W_{ij})\phi(W_{ij} - \frac{\xi_i^\mu}{\sqrt{M}})dW_{ij} - \\ &\quad -Nm^\mu \frac{1}{2}\int_{-\infty}^{\infty} g(W_{ij})\phi(W_{ij} + \frac{\xi_i^\mu}{\sqrt{M}})dW_{ij} \approx \end{aligned} \tag{6}$$

$$
\begin{aligned}
&\approx && N m^{\mu} \frac{1}{2} \int_{-\infty}^{\infty} g(W_{ij}) \left[\phi(W_{ij}) - \frac{\xi_i^{\mu}}{\sqrt{M}} \phi'(W_{ij}) \right] dW_{ij} - \\
& && -N m^{\mu} \frac{1}{2} \int_{-\infty}^{\infty} g(W_{ij}) \left[\phi(W_{ij}) + \frac{\xi_i^{\mu}}{\sqrt{M}} \phi'(W_{ij}) \right] dW_{ij} = \\
&= && N m^{\mu} \int_{-\infty}^{\infty} g(W_{ij}) \frac{\xi_i^{\mu}}{\sqrt{M}} W_{ij} \phi(W_{ij}) dW_{ij} = \\
&= && N m^{\mu} \frac{\xi_i^{\mu}}{\sqrt{M}} E\left[z g(z) \right]
\end{aligned}
$$

where z is a random variable with standard normal distribution. The variance of the field is similarly calculated to be

$$V(f_i|\xi_i^{\mu}) = N E\left[g^2(z) \right] . \tag{7}$$

Hence the signal to noise ratio

$$S/N = \frac{E\left[f_i|\xi = +1\right] - E\left[f_i|\xi = -1\right]}{\sqrt{V\left[f_i|\xi_i\right]}} = \sqrt{\frac{N}{M}} m^{\mu} \frac{E\left[z g(z)\right]}{\sqrt{E\left[g^2(z)\right]}}. \tag{8}$$

As z has standard normal distribution, $E(z^2) = V(z) = 1$. Assuming $g(z)$ is anti-symmetric (or at least has zero expectation) we can use $V\left[g(z)\right] = E\left[g^2(z)\right]$ and write

$$S/N = \frac{1}{\sqrt{\alpha}} m^{\mu} \rho(g(z), z) , \tag{9}$$

where $\alpha = M/N$ is the memory load and ρ denotes the correlation coefficient. The S/N is thus a product of *independent* terms of the load, the initial overlap and a correlation term which depends on the modification function only.

The S/N calculation for the low-activity model is similar but more cumbersome and is described in Appendix A. The resulting S/N can again be separated into a similar product of independent terms

$$S/N = \frac{1}{\sqrt{p\alpha}} m^{\mu} \rho(g(z), z). \tag{10}$$

In both models the only effect of the modification function g on the S/N is through the correlation coefficient, hence, the behavior of the two different models under synaptic modification can be investigated by analyzing $\rho(g(z), z)$ only, regardless of the other parameters [1]. The immediate consequence of Eqs. (9) and (10) is that there is no local synaptic modification function that can improve the performance of the Hebbian network, since ρ has

[1] These results and their following consequences remain valid even when the initial synaptic weights matrix is non zero but has some Gaussian distributed noise, because such a noise can be viewed as additional pre-stored memories.

values in the range $[-1, 1]$, and the identity function $g(z) = z$ already gives the maximal possible value of $\rho = 1$. In particular, no deletion strategy can yield better memory performance than the intact network. A similar result was previously shown by [Sompolinsky, 1988] in the Hopfield model. However, the use here of signal-to-noise analysis enables us to proceed and derive optimal functions under different constraints on modification functions, and evaluate the performance of various modification functions.

When no constraints are involved, pruning has no beneficial effect. However, since synaptic activity is strongly correlated with energy consumption in the brain, its resources may be inherently limited in the adult, and synaptic modification functions should satisfy various synaptic constraints. The following two subsections study deletion under two different synaptic constraints: limited number of synapses, and limited total synaptic efficacy [2].

2.3 Optimal Modification With Limited Number Of Synapses

In this section we find the optimal synaptic modification strategy when the number of synapses is restricted. The analysis consists of the following stages: First we show that under any deletion function, the remaining weights' efficacies should not be changed. Second, we show that the optimal modification function satisfying this rule is minimal-value deletion. Finally, we calculate the S/N and memory capacity of networks deleted with this strategy as a function of the deletion level.

Let g_A be a piece-wise equicontinuous deletion function, which zeroes all weights whose values are not in some set A and possibly modifies the remaining weights. To find the best modification function over the remaining weights we should maximize $\rho(g_A(z), z) = E\left[zg_A(z)\right]/\sqrt{E\left[g_A^2(z)\right]}$, that is invariant to scaling. Therefore, we keep $E\left[g_A^2(z)\right]$ fixed and look for a g_A which maximizes $E\left[zg_A(z)\right] = \int_A zg(z)\phi(z)$. Using the Lagrange method we write (as in [Meilijson and Ruppin, 1996])

$$\int_A zg(z)\phi(z)dz - \gamma(\int_A g^2(z)\phi(z)dz - c_1) \tag{11}$$

for some constant c_1. Denoting $g_i = g(z_i)$ we approximate (11) by

$$\sum_{\{i|z_i \in A\}} z_i g_i \phi(z_i) - \gamma(\sum_{\{i|z_i \in A\}} g_i^2 \phi(z_i) - c_1'). \tag{12}$$

Differentiating with respect to g_i yields that $g_i = \frac{z_i}{2\gamma}, \forall z_i \in A$; hence, g is linear homogeneous in z. We conclude that the optimal function should leave the undeleted weights unchanged (except for arbitrary linear scaling).

[2] It should be noted that we do not derive general optimal synaptic matrices, but optimal modifications of a previously learned Hebbian synapses. A study of the former can be found in [Bouten *et al.*, 1990].

To find the weights that should be deleted, we write the deletion function as $g_A(z) = zR_A(z)$, where

$$R_A(z) = R_A^2(z) = \begin{cases} 1 & when \quad z \in A \\ 0 & otherwise. \end{cases}$$

Since $zg_A(z) = z^2 R_A(z) = g_A^2(z)$, then $E\left[zg_A(z)\right] = E\left[g_A^2(z)\right]$ and $\rho(g_A(z), z) = \sqrt{\int z^2 R_A(z)\phi(z)dz}$. Given a constraint $\int_A \phi(z)dz = const$ which holds the number of synapse fixed, the term $\int_A z^2\phi(z)dz$ is maximized when A supports the larger values of $|z|$. To summarize, if some fraction of the synapses are to be deleted, the optimal (***"minimal value"***) pruning strategy is to delete all synapses whose magnitude is smaller than some threshold, and leave all others intact, as illustrated in figure 1(a).

To calculate $\rho(g(z), z)$ as a function of the deletion level let

$$g_t(z) = zR_t(z), \quad where\, R_t(z) = R_{\{s|t<|s|\}}(z) \tag{13}$$

and t is the threshold below which weights are removed. Using the fact that $\phi'(z) = -z\phi(z)$ and integrating by parts, we obtain

$$E\left[zg_t(z)\right] = E\left[g_t^2(z)\right] = \int_{-\infty}^{\infty} z^2\, R_t(z)\, \phi(z)dz = 2\int_t^{\infty} z\, z\, \phi(z)dz = 2\left[\Phi^*(t) + t\phi(t)\right] \tag{14}$$

and

$$\rho(g_t(z), z) = \sqrt{2t\phi(t) + 2\Phi^*(t)}, \tag{15}$$

where $\Phi^*(t) = P(z > t)$ is the standard normal tail distribution function.

(a) Minimal value deletion (b) Clipping modification (c) Compressed deletion

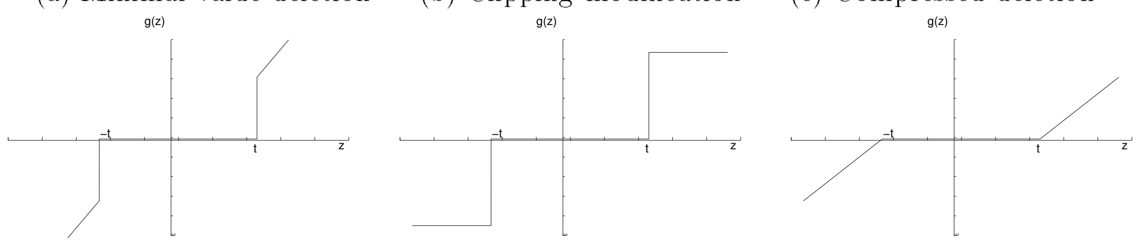

Figure 1: Different synaptic modification strategies. (a) Minimal value deletion: $g(z) = z$ for all $|z| > t$, and zero otherwise (see Eq. 13). (b) Clipping: $g(z) = sign(z)$ for all $|z| > t$ and zero otherwise. (c) Compressed deletion: $g(z) = z - sign(z)t$ for all $|z| > t$ and zero otherwise (see Eq. 16).

2.4 Optimal Modification With Restricted Overall Synaptic Strength

As synapses differ by their strength, a possible different goal may be implied by the energy consumption constraints, minimizing the overall synaptic strength in the network. We thus wish to maximize the S/N while keeping the total synaptic strength $\int |g(z)|$ fixed. Using the Lagrange method (the full derivation is brought in the Appendix A.4) we find that the optimal modification function is

$$g_t(z) = \begin{cases} z - t & when \quad z > t \\ 0 & when \quad |z| < t \\ z + t & when \quad z < -t \end{cases} \tag{16}$$

that is, the absolute value of all synapses with magnitude above some threshold t are reduced by t, and the rest are eliminated. We denote this modification function ***"compressed deletion"*** (figure 1(c)).

The S/N under this strategy is calculated using the function $R_t(z)$ described above (Eq. 13) and then writing $g(z) = (z - t)R_t(z)$ for positive z values. The calculation is done similarly to Eq. (14), yielding

$$\rho(g(z), z) = \frac{2\Phi^*(t)}{\sqrt{2(1+t^2)(\Phi^*(t) - 2t\phi(t))}}. \tag{17}$$

3 Numerical Results

To quantitatively evaluate the performance gain achieved by the strategies described in the previous section, the network's performance is measured by calculating the memory capacity of the network as a function of synaptic deletion levels. The capacity is measured as the maximal number of memories which can be stored in the network and retrieved almost correctly ($m^\mu \geq 0.95$), starting from patterns with an initial overlap of $m_0^\mu = 0.8$. Simulations shown below, are for the more biologically plausible low activity network, with $N = 800$ neurons and coding level $p = 0.1$, and similar results were obtained with Hopfield model simulations. The analytic curves were calculated using the overlap equation (Eqs. 23), which enable to calculate of the overlap after a single step for any memory load, by finding the highest load that yields an overlap ($m^\mu \geq 0.95$). Simulation results are reported both for a single iteration and for ten iterations.

Figure 2 compares the capacity obtained with three modification strategies: minimal value deletion (Eq. 13), random deletion (independent of the connections strength) and a "*clipping*" deletion. In clipping deletion, all weights with magnitude below a threshold value are removed, and the remaining ones are assigned a ± 1 value, according to their sign (see figure 1(b)). Two sets of simulations are presented. The first set was performed with an arbitrary fixed threshold, and the second with a threshold optimally tuned for each deletion level [3]. Minimal-value deletion is indeed significantly better than the other deletion

[3] In one-step simulations, the optimal threshold was determined according to Eq. 26, and in ten-steps simulations, the optimal threshold was found numerically to maximize the network's performance.

strategies, but in high deletion levels, it is almost equaled by the clipping strategy.

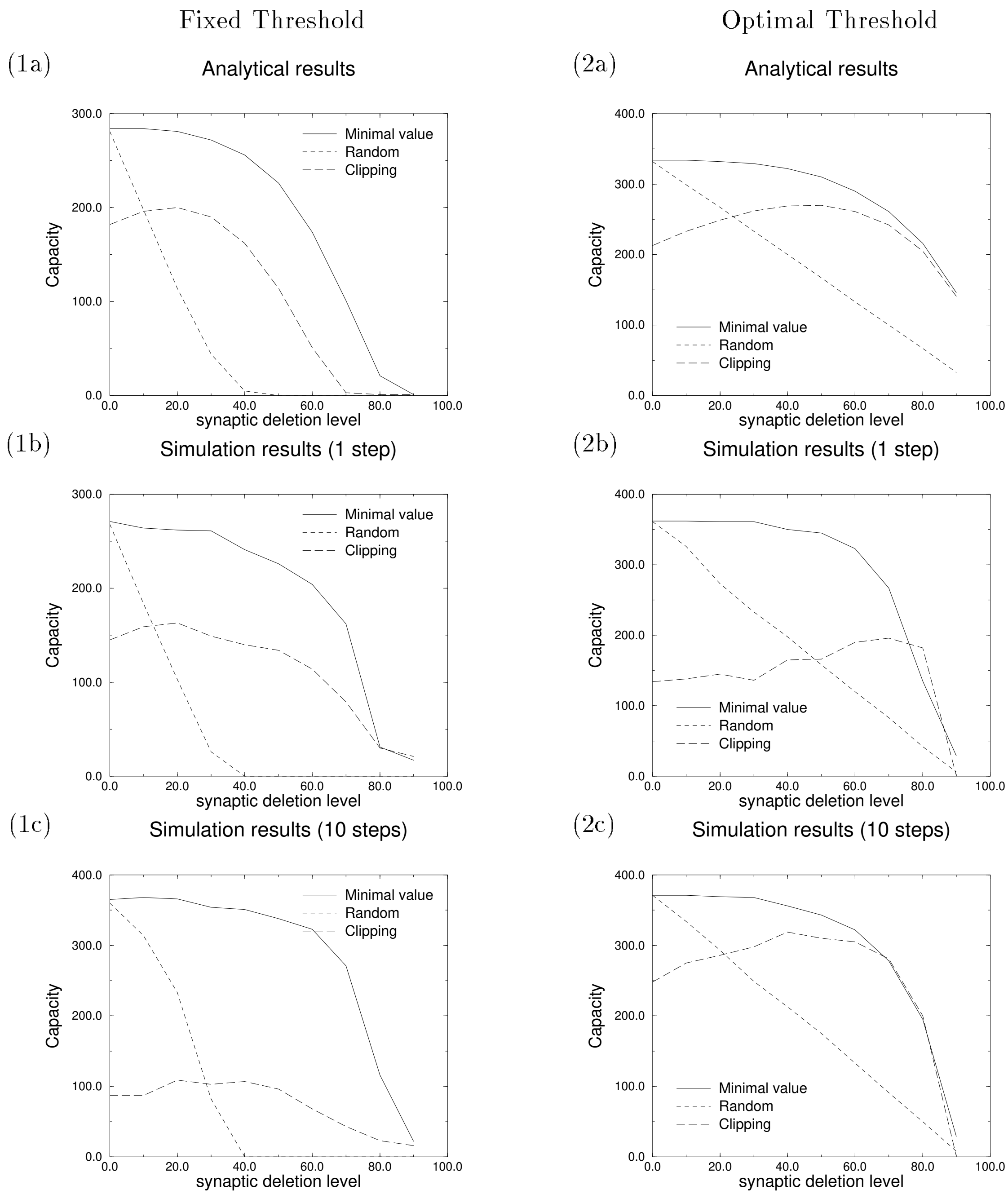

Figure 2: Capacity of a network with different synaptic modification strategies as a function of the synaptic deletion level. The left column shows results of the low activity model with a fixed threshold, while results with optimal neural threshold (i.e. threshold that is varied optimally with the deletion level) are shown in the right column. Both analytical and simulation results of single step and multiple step dynamics are presented, showing a qualitatively similar behavior.

Figure 3 compares the "compressed-deletion" modification strategy (Eq.16) to random deletion, as a function of the fraction of the total synaptic strength that is deleted from the network.

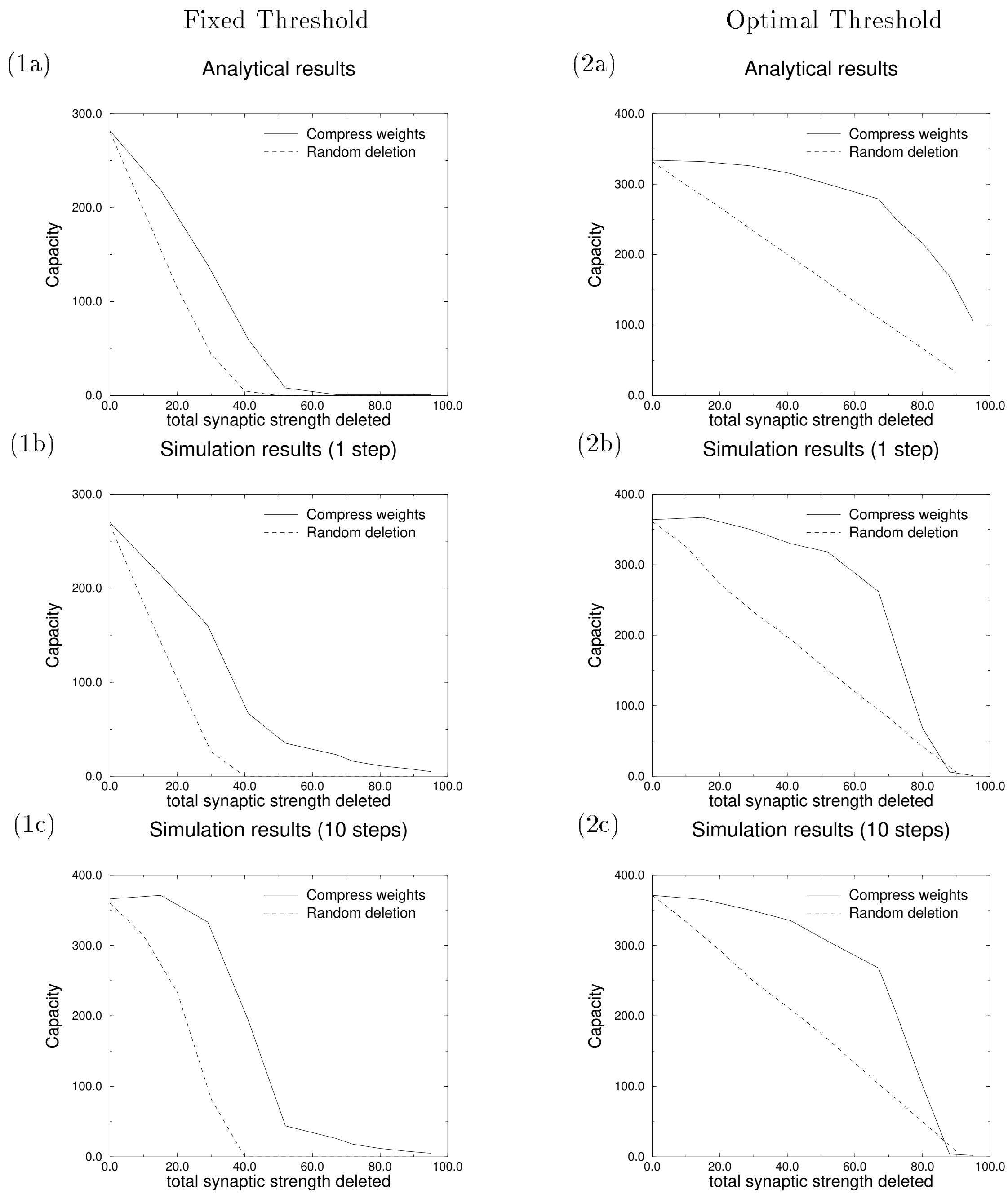

Figure 3: Capacity of a network with different synaptic modification strategies as a function of the fraction of the total synaptic strength removed from the network. The left column shows results of the low activity model with fixed threshold, while results with optimal threshold are shown in the right column.

The above results show that if a network must be subjected to synaptic deletion, minimal value deletion will minimize the damage, yet deletion reduces performance and is hence unfavorable. We now proceed to show that in the case where the amount of synapses is restricted in the adult organism, an initial over-growth of synapses followed by deletion, is beneficial. Figure 4 compares the memory capacity of networks with the same number of synapses, but with a varying number of neurons. The smallest network ($N = 800$) is fully connected while larger networks are pruned by minimal value deletion to end up with the same amount of synapses. The optimal deletion ratio is found around 80% deletion, and improves capacity by 45% . This optimally pruned network, that has more neurons, can store three times more information than the fully connected network with the same number of synapses. When the threshold is sub-optimal or the energy cost for neurons is non-negligible, the optimum drifts to a deletion levels of $50\% - 60\%$.

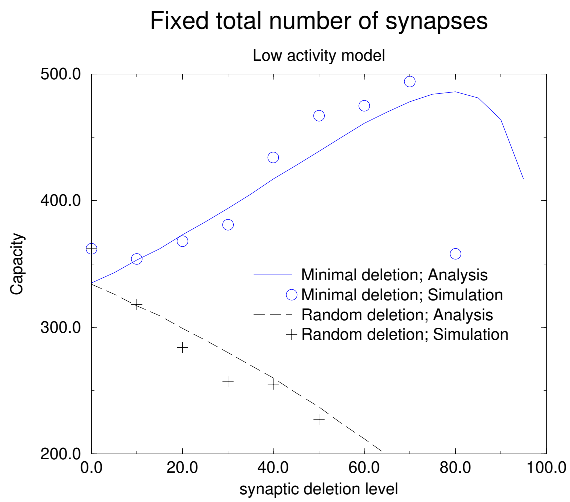

Figure 4: Capacity of networks with different number of neurons but the same total number of synapses as a function of network connectivity. The networks are pruned according to minimal value deletion to keep the total number of synapses (k) constant. Simulation parameters are $k = 800^2$, $p = 0.1$, and T is kept optimal.

The conclusion is that an organism that first over-grows synapses in a large network and then judiciously prunes them, can store much more memories than another adult organism that uses the same synaptic resources, but settles for the adult synaptic density already in infancy.

Until now, we have analyzed synaptic deletion of previously established synaptic matrices (storing a given set of memories). We next turn to simulate the continuous process of learning that is superimposed on the profile of synaptic density changes occurring during human development. These changes naturally define a time step equivalent to one year. Within each time step we store some memories and change the synaptic density following the human data. Synapses are incrementally added, increasing synaptic density until the age of 3 "years". At the age of 5 "years" synaptic pruning begins, lasting until puberty (see the dot-dashed line in figure 5). Addition of new synapses is done at a constant rate, and synaptic efficacies are determined by the new memories stored in the network. The deletion of synapses is done according to the minimal value deletion strategy. The network is tested for recall of the stored memories twice: once, at the age of 3 "years" when synaptic density is at its peak, and again at an age of 15 "years" when synaptic elimination has already removed 40% of the synapses.

Figure 5 traces the networks performance during this experiment, measured by the retrieval acuity (final overlap with the cued memory obtained after one step of the dynamics). It superimposes the synaptic density (dot-dashed line) and memory performance data. Two observations should be noted: First is the inverse temporal gradient in the recall performance of memories stored during the synaptic pruning phase. That is, there is a deterioration in the performance of the "teenage" network as it recalls more recent childhood memories (see the decline in the dashed line). The second is the marked difference between the ability of the "infant" network (the solid line) and the "teenage" network (the dashed line) to recall memories stored at "early childhood"; Older networks totally fail to recall any memory before the age of 3-4 "years", manifesting "childhood amnesia".

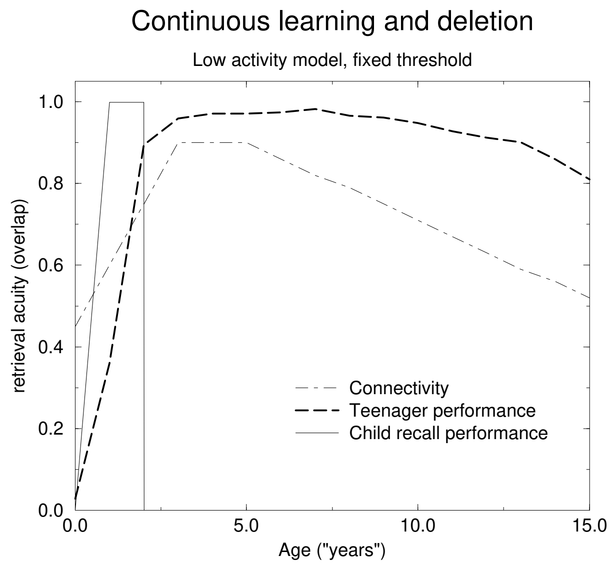

Figure 5: Memory retrieval as a function of storage period. The figure displays both synaptic density and memory performance data. At each time step ("year") m memories are stored in the network, and the network's connectivity is changed following human data (dot-dashed line). The network is tested for retrieval twice: in an early ("infant") stage when network connectivity has reached its peak (solid line), and in a later ("teenage") phase after more memories have been stored in the network (dashed line). In each such test, we look back in time and examine how well does the network currently retrieve memories from its past, as a function of their storage time. Network parameters are $N = 800$, $m = 10$ and $p = 0.1$. The threshold is kept fixed at $T = (1/2 - p)p(1 - p)$.

4 Discussion

We have analyzed the effect of modifying Hebbian synapses in an optimal way that maximizes memory performance, while keeping constant the overall number or total strength of the synapses. The optimal functions found for these criteria use only local information about the synaptic strength, do not depend on the activity level of the network, and are not effected by initial noise in the synaptic matrix. Moreover, they are exactly the same functions in a large family of associative memory networks.

We have shown that under a restricted number of synapses, the optimal local modification function of a given Hebbian matrix is to delete the small weights, and maintain the values of the remaining connections. Under restricted total synaptic strength, the optimal synaptic modification is to delete the small weights and linearly reduce the strength of the remaining synapses. Our results predict that during the elimination phase in the brain synapses undergo weight-dependent pruning in a way that deletes the weak synapses (one should expect that both constraints described above may concomitantly apply in reality, with some relative weighting that would determine the exact desired synaptic modification function).

As we have shown, synaptic deletion cannot improve performance of a given network. What then is its role ? Until now, several computational answers were suggested. Some had hypothesized that synaptic elimination can improve network performance, but this paper proves this argument is incorrect in several associative memory models. Others have claimed that the brain can be viewed as a cascade of filters which can be modeled by feed forward networks models [Sharger and Johnson, 1995]. In these models it is known that a reduction in the amount of free parameters may improve the ability of the network to generalize if the size of the network is too large [Reed, 1993]. This explanation holds when the complexity of the problem is unknown at the time the networks are created (and therefore cannot be pre-programmed genetically), and applies to networks that should generalize well. Another possible argument for justifying synaptic deletion arises if synaptic values are assumed to have ± 1 values only (as in the clipping function described above). Under such an assumption (as can be observed in figure 2), maximal performance is obtained at non-zero deletion levels. However the biological plausibility of uni-valued synapses is in doubt.

Our proposal is that synaptic over-growth and deletion emerge because synaptic resources must be scrupulously utilized, due to metabolic energy consumption constraints. If we have to use a restricted amount of synapses in the adult, better performance is achieved if the synapses are first over-grown, and then cleverly pruned after more memories are stored. The optimally pruned network is not advantageous over the undeleted network (which has many more synapses), but over all other networks, with the same total number of synapses. It should be noted, however, that our results pertain to associative memory networks and that synaptic elimination probably plays different roles in other neural systems. For example, it was shown that in the primary visual cortex, synaptic elimination determines the architecture and the function of the network [Miller *et al.*, 1989]. Interestingly, although rising from different basic principles, recent studies have found that in the neuro-muscular junction synapses are pruned according to their initial synaptic strength, deleting the weaker synapses [Frank, 1997, Colman *et al.*, 1997].

In biological networks, synaptic growth and deletion occur in parallel with memory storage. As shown in Figure 5, the implementation of a minimal value pruning strategy during such process yields two cognitive predictions: one for the rising phase of synaptic density curve and the other for the descending phase. At the descending phase of synaptic density an inverse temporal gradient is observed. That is, as long as synapses are eliminated, remote memories are easier to recall than recently stored memories (dashed curve in figure 5). The reason for this inverse gradient is the continuous change in network connectivity: earlier memories are stored into a highly-connected network, while memories stored later are engraved into a sparser network. The early memories take a prominent role in determining the synapses which are pruned by the minimal value algorithm, and therefore are only slightly damaged by the synaptic deletion. The more recent memories are engraved into an already deleted network, and hence have little influence on determining which synapses are deleted. From the "point of view" of recent memories the network hence undergoes random deletion. However, adding accumulative noise to the network or assuming synaptic decay damages remote memory retrieval more than recent ones. Therefore, the model predicts that the plot of human memory retrieval as a function of storage time within the synaptic elimination period should have a U-shaped form. Interestingly, such a result can be observed in previous studies of long term memory ([Sheingold and Tenney, 1982]) but was unnoticed before.

A comparison of retrieval quality of early memories by the teenager network versus the infant network (see figure 5 at the age of 2 years) shows a sharp change in retrieval quality of early memories. This is reminiscent of the infantile amnesia phenomenon, which is the inability of the adult to recall events from infancy that he could previously recall [Bachevalier *et al.*, 1993, Markievwicz *et al.*, 1986]. In our model, this pattern arises from the fact that earlier memories are stored in sparsely-connected networks (that is, embedded in less synapses) and hence are more sensitive to the noise known to accumulate in the network as additional memories are stored in it. This scenario may provide a network-level perspective to infantile-amnesia, complementing the previous theories suggesting that maturation of memory related structures such as the Hippocampus are responsible for the amnesia [Nadel, 1986].

Synaptic elimination is a broad phenomenon found throughout different brain structures and not restricted to associative memory areas. We believe that our explanation may be generalized to other network models. For example, feed forward Hebbian projections between consecutive networks share similar properties with a single step of synchronous dynamics of associative memory networks analyzed here. There is also some evidence that synaptic growth followed by deletion is not limited to the developmental stage, but may have a more general scope, and can be found in adults [Greenough *et al.*, 1987]. These interesting open issues await further studies in the future.

References

[Bachevalier *et al.*, 1993] J. Bachevalier, M. Brickson, and C. Hagger. Limbic dependent recognition memory in monkeys develops early in infancy. *Neuroreport*, 4(1):77–80, 1993.

[Bourgeois and Rakic, 1993] J.P. Bourgeois and P. Rakic. Changing of synaptic density in the primary visual cortex of the Rhesus monkey from fetal to adult age. *J. Neurosci.*, 13:2801–2820, 1993.

[Bourgeois, 1993] J.P. Bourgeois. Synaptogenesis in the prefrontal cortex of the macaque. In B. do Boysson-Bardies, editor, *Developmental Neurocognition: Speech and Face Processing in the First Year of Life*, pages 31–39. Kluwer Academic Publishers, 1993.

[Bouten *et al.*, 1990] M. Bouten, A. Engel, A. Komoda, and R. Serneel. Quenched versus annealed dilution in neural networks. *J. Phys. A: Math Gen.*, 23:4643–4657, 1990.

[Colman *et al.*, 1997] H. Colman, J. Nabekura, and J.W. Lichtman. Alteration in synaptic strength preceding axon withdrawal. *Science*, 275:356–361, 1997.

[Eckenhoff and Rakic, 1991] M.F. Eckenhoff and P. Rakic. A quantitative analysis of synaptogenesis in the molecular layer of the dentate gyrus in the resus monkey. *Developmental Brain Research*, 64:129–135, 1991.

[Frank, 1997] E. Frank. Synapse elimination: For nerves it's all or nothing. *Science*, 275:324–325, 1997.

[Greenough *et al.*, 1987] W.T. Greenough, J.E. Black, and C.S. Wallace. Experience and brain development. *Child Development*, 58:539–559, 1987.

[Huttenlocher and Courten, 1987] P.R. Huttenlocher and C. De Courten. The development of synapses in striate cortex of man. *J. Neuroscience*, 1987.

[Huttenlocher *et al.*, 1982] P.R. Huttenlocher, C. De Courten, L.J. Garey, and H. Van der Loos. Synaptogenesis in human visual cortex - evidence for synapse elimination during normal development. *Neuroscience letters*, 33:247–252, 1982.

[Huttenlocher, 1979] P.R. Huttenlocher. Synaptic density in human frontal cortex. Development changes and effects of age. *Brain Res.*, 163:195–205, 1979.

[Innocenti, 1995] G.M. Innocenti. Exuberant development of connections and its possible permissive role in cortical evolution. *Trends Neurosci*, 18:397–402, 1995.

[J.Takacs and Hamori, 1994] J.Takacs and J. Hamori. Developmental dynamics of Purkinje cells and dendritic spines in rat cerebellar cortex. *J. of Neuroscience Research*, 38:515–530, 1994.

[Markievwicz *et al.*, 1986] B. Markievwicz, D. Kucharski, and N.E. Spear. Ontogenic comparison of memory for Pavlovian conditioned aversions. *Developmental psychobiology*, 19(2):139–54, 1986.

[Meilijson and Ruppin, 1996] I. Meilijson and E. Ruppin. Optimal firing in sparsely-connected low-activity attractor networks. *Biological cybernetics*, 74:479–485, 1996.

[Miller *et al.*, 1989] K.D. Miller, J.B. Keller, and M.P. Stryker. Ocular dominance column development: analysis and simulation. *Science*, 245:605–615, 1989.

[Nadel, 1986] L. Nadel. Infantile amnesia: A neurobiological perspective. In M. Moscovitch, editor, *Infant Memory; Its Relation To Normal And Pathological Memory In Humans And Other Animals*. Plenum Press, 1986.

[Rakic *et al.*, 1994] P. Rakic, J.P. Bourgeois, and P.S. Goldman-Rakic. Synaptic development of the cerebral cortex: implications for learning, memory and mental illness. *Progress in Brain Research*, 102:227–243, 1994.

[Reed, 1993] R. Reed. Pruning algorithms - a survey. *IEEE transactions on neural networks*, 4(5):740–747, 1993.

[Roe *et al.*, 1990] A.W. Roe, S.L. Pallas, J.O. Hahm, and M. Sur. A map of visual space induced in primary auditory cortex. *Science*, 250:818–820, 1990.

[Roland, 1993] Per E. Roland. *Brain Activation*. Willey-Liss, 1993.

[Sharger and Johnson, 1995] J. Sharger and M.H. Johnson. Modeling development of cortical functions. In B.Julesz I. Kovacs, editor, *Maturational windows and cortical plasticity*. The Santa Fe institute press, 1995.

[Sheingold and Tenney, 1982] K. Sheingold and J. Tenney. Memory for a salient childhood event. In U. Neisser, editor, *Memory observed*. W.H. Freeman and co., 1982.

[Sompolinsky, 1988] H. Sompolinsky. Neural networks with non linear synapses and static noise. *Phys Rev A.*, 34:2571–2574, 1988.

[Stryker, 1986] M.P. Stryker. Binocular impulse blockade prevents the formation of ocular dominance columns in cat visual cortex. *J. of Neuroscience*, 6:2117–2133, 1986.

[Tsodyks and Feigel'man, 1988] M.V. Tsodyks and M. Feigel'man. Enhanced storage capacity in neural networks with low activity level. *Europhys. Lett.*, 6:101–105, 1988.

[Wolff *et al.*, 1995] J.R. Wolff, R. Laskawi, W.B. Spatz, and M. Missler. Structural dynamics of synapses and synaptic components. *Behavioural Brain Research*, 66:13–20, 1995.

A Appendix: Signal To Noise Ratio In A Low Activity Model

A.1 Field Moments

The network is initialized with activity p and overlap m_0^μ with memory μ. Let $\epsilon = P(X_i = 0|\xi_i = 1)$ (which implies an initial overlap of $m_0 = \frac{(1-p-\epsilon)}{(1-p)}$). We write

$$E(f_i|\xi_i) = NP(\xi_j = 1)E\left[\frac{1}{N}g(W_{ij})|\xi_j = 1\right] + NP(\xi_j = 0)E\left[\frac{1}{N}g(W_{ij})|\xi_j = 0\right] \quad (18)$$

The first term is calculated as follows

$$\begin{aligned} NP(\xi_j = 1) \quad & E\left[g(W_{ij})|\xi_j = 1\right] = && (19)\\ = \quad & Np(1-\epsilon)\int g(W_{ij})\phi(W_{ij} - \frac{(1-p)(\xi_i - p)}{p(1-p)\sqrt{M}})dW_{ij} \approx \\ \approx \quad & Np(1-\epsilon)\int g(W_{ij})\left[\phi(W_{ij}) - \frac{(1-p)(\xi_i - p)}{p(1-p)\sqrt{M}}\phi'(W_{ij})\right]dW_{ij} = \\ = \quad & Np(1-\epsilon)E\left[g(z)\right] + Np(1-\epsilon)\frac{(1-p)(\xi_i - p)}{p(1-p)\sqrt{M}}E\left[zg(z)\right] = \\ = \quad & NpE\left[g(z)\right] + Np(1-p-\epsilon)\frac{(\xi_i - p)}{p(1-p)\sqrt{M}}E\left[zg(z)\right] = \end{aligned}$$

A similar calculation for the second term, where g is anti-symmetric yields

$$E(f_i|\xi_i) = Np(1-p-\epsilon)\frac{(\xi_i - p)}{p(1-p)\sqrt{M}}E\left[zg(z)\right] - T. \quad (20)$$

The variance is calculated following

$$V(f_i|\xi_i) = NE(g^2(W_{ij})X_j^2) - NE^2(g(W_{ij})X_j) + N(N-1)Cov(g(W_{ij})X_j, g(W_{ik})X_k), \quad (21)$$

in a similar manner to yield

$$\frac{E(f_i|\xi_i)}{\sqrt{V(f_i|\xi_i)}} = \frac{\frac{N}{\sqrt{M}}m_o(\xi_i - p)E\left[zg(z)\right] - T}{\sqrt{NpE\left[g^2(z)\right]}}. \quad (22)$$

A.2 The Overlap Equation

Given the overlap m_0 between the network's initial state and a pattern ξ^μ we calculate m_1, the overlap in the next step, by

$$\begin{aligned} m_1 \quad = \quad & \frac{1}{Np(1-p)}N(1-p)P(\xi_i = 1)P(X_i = 1|\xi_i = 1) && (23)\\ & -\frac{1}{Np(1-p)}NpP(\xi_i = 0)P(X_i = 1|\xi_i = 0) = \\ = \quad & P(X_i = 1|\xi_i = 1) - P(X_i = 1|\xi_i = 0) = \\ = \quad & \Phi(\frac{E(f_i|\xi_i)}{\sqrt{V(f_i|\xi_i)}}|\xi_i = 1) - \Phi(\frac{E(f_i|\xi_i)}{\sqrt{V(f_i|\xi_i)}}|\xi_i = 0), \end{aligned}$$

where $\Phi = 1 - \Phi^*$ is the standard Gaussian cumulative distribution function.

A.3 Optimal Threshold

In order to find the threshold that maximizes the overlap we differentiate m_1 (Eq. 23) with respect to T,

$$\frac{\partial m_1}{\partial T} = \frac{\partial \left[\Phi(\frac{E(f_i|\xi_i)}{\sqrt{V(f_i|\xi_i)}}|\xi_i = 1) - \Phi(\frac{E(f_i|\xi_i)}{\sqrt{V(f_i|\xi_i)}}|\xi_i = 0)\right]}{\partial T} = 0 \tag{24}$$

which yields

$$\phi(\frac{\frac{N}{\sqrt{M}} m_o (1-p) E\left[zg(z)\right] - T}{\sqrt{NpE\left[g^2(z)\right]}}) = \phi(\frac{\frac{N}{\sqrt{M}} m_o (0-p) E\left[zg(z)\right] - T}{\sqrt{NpE\left[g^2(z)\right]}}) \tag{25}$$

and

$$T = \frac{N}{\sqrt{M}}(\frac{1}{2} - p)\, m_0\, E\left[zg(z)\right]. \tag{26}$$

Using the optimal threshold in Eq. 22 yields

$$S/N = \sqrt{\frac{N}{M}}\, m_0 \frac{1}{\sqrt{p}}\, \rho(g_t(z), z) \tag{27}$$

Similarly to the case of Hopfield model, the S/N of the neuron i can be expressed as a product of independent factors: the load M/N, the deletion strategy g, the activity level p and the activity of the neuron ξ_i.

A.4 Derivation of the compressed deletion

Using the Lagrange method we have

$$\int_{-\infty}^{\infty} zg(z)\phi(z)dz - \gamma_1(\int_{-\infty}^{\infty} g^2(z)\phi(z)dz - c_1) - \gamma_2(\int_{-\infty}^{\infty} |g(z)|\phi(z)dz - c_2) =$$

$$= \int_{-\infty}^{\infty} |z||g(z)|\phi(z)dz - \gamma_1(\int_{-\infty}^{\infty} |g(z)|^2\phi(z)dz - c_1) - \gamma_2(\int_{-\infty}^{\infty} |g(z)|\phi(z)dz - c_2) \tag{28}$$

which is approximated by

$$\sum_i |z_i||g_i|\phi(z_i) - \gamma_1(\sum_i |g_i|^2\phi(z_i) - c_1') - \gamma_2(\sum_i |g_i|\phi(z_i) - c_2') \tag{29}$$

Assuming $g(z)$ to be piece-wise equicontinuous and equating to zero the derivative with respect to $|g_i|$ we obtain

$$|z_i|\phi(z_i) - \gamma_1 2|g_i|\phi(z_i) - \gamma_2\phi(z_i) = 0 \tag{30}$$

or

$$|g(z)| = \frac{1}{2\gamma_1}(|z| - \gamma_2) \; , \tag{31}$$

from where

$$g_t(z) = \begin{cases} z - t & when \quad z > t \\ 0 & when \quad |z| < t \\ z + t & when \quad z < -t \end{cases} \tag{32}$$