Tel Aviv University
The Raymond and Beverly Sackler
Faculty of Exact Sciences
School of Mathematical Sciences

Computational Aspects of Synaptic Elimination

Thesis submitted in partial fulfillment
of graduate requirements for the degree
"Master of Sciences" in Tel Aviv University
Department of Computer Science

by

Gal Chechik

Prepared under the supervision of
Prof. Isaco Meilijson
and
Dr. Eytan Ruppin

June 1997

Acknowledgments

First, I am happy to thank my tutor Eytan Ruppin. His continuous support, friendship and inspiration, have turned the year of work on this thesis to an exciting period. He taught me a lot and was more than any student can hope for.

I would like to express my gratitude to my tutor Isaco Meilijson for his most helpful guidance. His patience and remarks helped me to cross the numerous mathematical obstacles.

My deepest thanks are due to Michal, for the infinite love and support.

Finally I wish to thank Nir Levy, Jonathan Cohen, Eyal Cohen (zoro), and Oran Singer for their wise comments and technical help.

Abstract

Research in humans and primates shows that the developmental course of the brain involves synaptic over-growth followed by marked selective pruning which eliminates about half of the synapses of the child. Previous explanations have suggested that this intriguing, seemingly wasteful, phenomenon is utilized to remove 'erroneous' synapses which were studied at an early stage. This thesis proves that this interpretation is wrong in a large family of associative memory network models. We study modifications of Hebbian synapses under different synaptic constraints emerging from metabolic energy restrictions, and derive optimal modification functions under these constraints. Under restricted number or strength of synapses, we show that memory performance is significantly enhanced if synapses are first overgrown and then pruned following optimal deletion strategies. These results predict that during the elimination phase in the brain synapses undergo weight-dependent pruning in a way that deletes the weak synapses. Implementing the derived optimal pruning strategies to a continuous process of memory storage and synaptic pruning, roughly mimicking the developmental process of humans, leads to interesting insights concerning some long term memory phenomena. This work predicts an inverse temporal gradient of long term memory recall during synaptic elimination phase, and suggests a first network-level explanation for the phenomenon of childhood amnesia.

Contents

Chapter 1

Introduction

One of the fundamental phenomena in normal brain development is the intensive reduction in the amount of synapses which occurs between early childhood and puberty. Although this process is well described, its underlying reasons are still unknown.

This thesis investigates the computational aspects of synaptic pruning, in the computational paradigm of associative memory networks models. We discuss possible motivations for elimination of synapses, and suggest a novel computational explanation for this phenomenon, studying its consequences and predictions. The thesis studies possible cognitive implications of synaptic elimination, such as network-based explanation of childhood amnesia, and the inverse temporal gradient of long term memory during childhood.

The following subsections describe the background of this work from biological, computational and cognitive perspectives. Subsection 1.1 describes the phenomenon of synaptic elimination during normal brain development. Subsection 1.2 reviews previous computational studies of partially connected associative networks, and subsection 1.3 describes the phenomenon of childhood amnesia.

1.1 Synaptic elimination during development

In recent years, many studies have investigated the temporal course of changes in synaptic density in primates, revealing the following picture (see figure 1.1). Beginning at early stages of pregnancy, synaptic density rises at a constant rate, until a peak level is attained (at humans this happens between the ages of 2-3 years). 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 to an adult level which is maintained until old age. The peak level of synaptic density at childhood is $50\% - 100\%$ higher than the adult level, depending on the brain region. These findings were found in humans [Huttenlocher, 1979, Huttenlocher ***et al.***, 1982, Huttenlocher and Courten, 1987], as well as in other vertebrates 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].

This puzzling elimination process was observed throughout different areas of the brain including widespread cortical areas (visual [Bourgeois and Rakic, 1993, Huttenlocher ***et al.***, 1982], motor [Rakic ***et al.***, 1994] and associative [Huttenlocher, 1979, Rakic ***et al.***, 1994]), 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]. Larger differences were found between species: in some species, the peak level of synaptic density is obtained at a very early age after birth (e.g. 2 weeks at the macaque monkeys

[Bourgeois, 1993]).

The experimental method used to investigate synaptic development is electronic microscopy of brain tissues stained with different biochemical markers. This method makes it possible to study synaptic structure and density. It leads to the finding that the changes in synaptic density are not a result of changes in total brain volume which occur during this period, but reflect true synaptic elimination. However, the changes in synaptic strength along development are yet to be characterized and the mechanism underlying these changes is unknown. In some cases, synaptic elimination was shown to be correlated with experience dependent activity [Roe *et al.*, 1990, Stryker, 1986].

The changes in synaptic density with age are correlated with the changes in metabolic energy consumption of the brain along life, as measured by monitoring the concentration of glucose in the cerebral fluid. Figure 1.1 (taken from [Roland, 1993]) shows this correlation: glucose rCMRgl data was taken from [Chugani *et al.*, 1987] and synaptic density data from [Huttenlocher, 1979]. Further more, there is evidence that the majority of brain metabolic energy consumption can be attributed to the synapses. Several studies have shown that the changes in metabolic rate associated with neuronal activity are localized in the regions with dense synaptic contacts, and are related to axonal processes but not to cell bodies. For example, [Kadekaro *et al.*, 1985] measured glucose consumption in the spinal cord after stimulation of nerve afferents. The cell bodies, localized outside the cord, did not show any increase in the metabolic rate while an increase of 150% was detected in areas where the neurons made synapses with local dendrites. Other studies [Roland, 1993] showed that the metabolic energy in the neuron is mainly consumed by the ion pumps. To summarize, there is fair evidence that synapses are the main energy consumers in the brain. As the brain itself consumes about 25% of the energy in the resting adult, synapses are a costly resource that should be efficiently utilized.

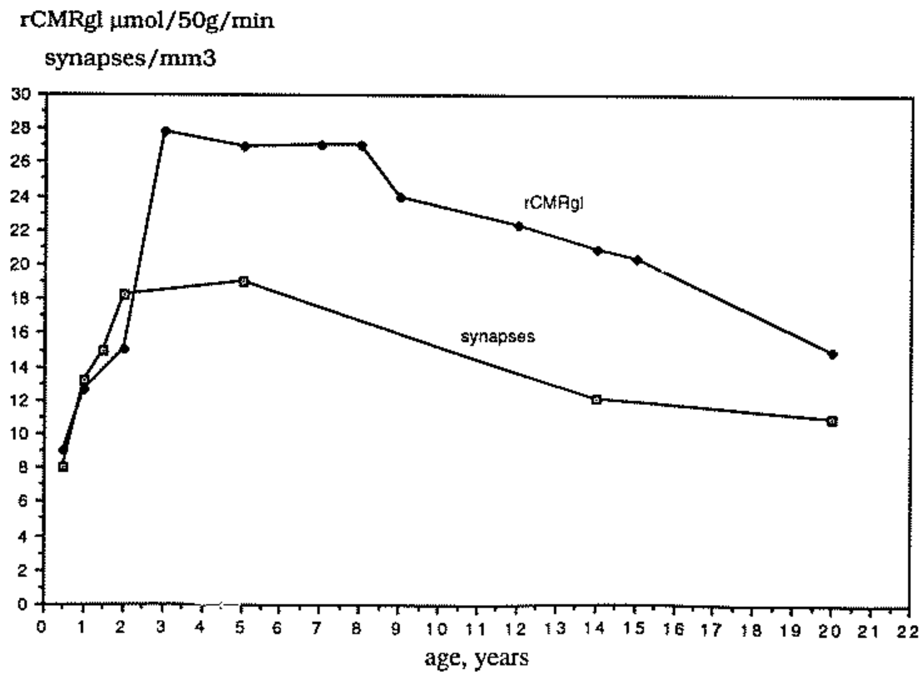

Figure 1.1: Correlation between number of synapses in the human frontal cortex and regional cerebral glucose consumption (rCMRgl) as a function of age in the developmental phase. The scale for the density of synapses must be multiplied with 10^8.

1.2 Computational studies of diluted networks

Partly connected associative memory neural networks were previously investigated in two main routes. Most of the studies dealt with random deletion of connections, i.e. connections removal that is independent of the memories stored in the network, and hence independent of the strength of the connection [Evans, 1989, Bouten *et al.*, 1990, Van-Hemmen, 1987, Tsodyks, 1988]. Other studies investigated pruning of connections in a way that depends on the set of memory patterns stored in the network. These studies may be sub-classified, by distinguishing between the creation of optimal synaptic matrices for storage of given memory patterns in a partially connected network [Bouten *et al.*, 1990], and the modification of pre-existing synaptic storage matrices. The work presented in this thesis can be mapped to this last subcategory.

Previous analysis of modifications of a Hebbian matrices in the Hopfield model was done by [Sompolinsky, 1988]. Using replica techniques he showed that introducing a non-linear function over the weights is equivalent to adding static noise, therefore reducing network's performance. Van Hemmen [Van-Hemmen, 1987] showed that such a non-linear function cannot be beneficial even when it is allowed to use more information than the global weight of a connection. He proved that even when all specific values of the pairs (ξ_i^μ, ξ_j^μ) (which determine the connection's weight by $W_{ij} = \sum \xi_i^\mu \xi_j^\mu$) are known, non-linear functions reduce performance. The work presented here extends these studies: our analysis shows that synaptic deletion reduces performance in a broad family of associative memory network models. The effect on performance of few modification functions such as clipping synaptic weights to a small set of possible values was studied by [Sompolinsky, 1988]. He showed that the capacity of the network under such modification of the synapses is only slightly reduced. Deriving optimal modification strategies under constraints, our analysis shows that such a strategy is near optimal in high deletion levels when the number of synapses in the network is restricted.

The method of Gardner [Gardner, 1987, Gardner, 1988] was used to evaluate the capacity of an optimal synaptic matrix of a previously pruned network and of an optimal synaptic matrix undergoing random deletion [Bouten *et al.*, 1990]. This method enables to find upper bounds on the network's capacity, but does not enable an explicit calculation of the synaptic matrices which achieve these bounds. To reach the upper bound, an optimal matrix must include global information about all memory patterns stored in the network, hence can hardly be realized in a biological system, and is used as a theoretical bound only. The distribution of the synaptic weights in the optimal matrix for partially connected networks was found to have a Gaussian distribution with zero mean from which the middle section has been cut out. Interestingly, though investigating a very different optimization task, our analysis derives a similar distribution of weights (under one of the constraints studied). In the case of networks randomly diluted before the calculation of the synaptic matrix, capacity depends linearly on the average number of connections per neuron.

Many studies have investigated the effect of synaptic and neural pruning on the performance of feed forward networks models (see [Reed, 1993] for a review). One of the main results is that reducing the number of free parameters may enhance the ability of the network to generalize, that is, to extract simple first order properties of the training set. Too large networks generalize poorly: they perform well with the set they are trained with, but their performance with new inputs is poor. Hence, pruning the network may improve overall performance when the size of the network is bigger than the complexity of the problem.

1.3 Childhood amnesia

One of the most striking aspects of human memory is that virtually no person can recall events from the first years of life, though this is the time when experience is at its richest. This curious phenomenon was first discussed by Freud, who called it ***childhood amnesia***. Freud discovered this phenomenon by observing that his patients were generally unable to recall events from their first 3-5 years of age. One of the interesting properties of childhood amnesia, is its dichotomous nature (demonstrated in figure 6.1): One recalls well his 5th year of age, but has almost no recollection of several months earlier. There is evidence that young children can recall these events (for example, 2 years children recall well events which occurred when they were 1 year old [McDonough and Mandler, 1994]). These findings were substantiated in studies with other mammals such as monkeys [Bachevalier ***et al.***, 1993] and rats [Markievwicz ***et al.***, 1986, Schweitzer and Green, 1982]. It is important to note that the strong memory decay in animals' infants was measured through various experimental paradigms all testifying that the phenomenon of infantile amnesia is not restricted to the human species.

What could be the causes of childhood amnesia ? Clearly, this phenomenon cannot be explained by simple forgetting (memory decay over time), as most adults recall well teenage events regardless of the long time passed since they occurred, while teenagers cannot recollect their infant memories after a similar period of time has elapsed. The Explanations suggested are numerous, originating from the different scientific disciplines. Freud thought that childhood amnesia is due to the repression of sexual and aggressive feelings that a child experiences towards his parents. He even argued that the few memories that can be recalled are screen memories made up to hide the emotionally disturbing true memories. Cognitive psychologists suggest childhood amnesia is due to a basic difference between the way that children encode experience and the way adults organize their memories. Some argue that children have limited or no ability to store information in verbal form and they tend to lack a detailed and organized frame of knowledge to which they can relate their experiences. Neurological explanations (mostly based on lesion studies) point to a maturation of specific memory structures in the brain (such as the Hippocampus) as being responsible for the amnesia [Nadel, 1986], but recent studies have questioned these views [Bachevalier ***et al.***, 1993]. This thesis suggests that global developmental processes of the brain effect the quality of memory embedding in the neural networks later yielding a breakdown of early childhood memories.

1.4 Summary of previous work

As described above, a massive pruning of synapses occurs along normal brain development in vertebrates. This work suggests a computational account for this phenomenon which is based on the study of functions over synaptic weights. Motivated by the strong evidence that synapses are a major energy consumer and hence a costly resource, our analysis extends the previous computational studies and considers the case of networks with limited synaptic resources. The following chapters show that the optimal synaptic modification strategies derived, provide a computational reasoning for the over-growth and pruning process. The implementation of these strategies in a continuous process of memory storage and deletion suggests a first explanation for childhood amnesia which operates at the network level.

Chapter 2

Analysis

In order to investigate synaptic elimination, we address the more general question of optimizing the synaptic learning rule. Given previously learned Hebbian synapses we apply a function which changes the synaptic values, and investigate the effect of such a modification function. In this section, we first analyze the way the network's performance depends on a general synaptic modification function in several Hebbian models; 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

In this thesis, synaptic modification is investigated mainly in two Hebbian models described below: Modified Hopfield model and modified low activity model proposed by [Tsodyks and Feigel'man, 1988].

2.1.1 Modified Hopfield model

The first model is a variant of the canonical model suggested by Hopfield. M memories are stored in a N-neuron network forming approximate fixed points of the network dynamics. The synaptic efficacy J_{ij} between the jth (pre-synaptic) neuron and the ith (post-synaptic) neuron is

$$J_{ij} = g(W_{ij}) = g(\frac{1}{\sqrt{M}} \sum_{\mu=1}^{M} \xi_i^\mu \xi_j^\mu)\,, \quad 1 \leq i \neq j \leq N; \quad J_{ii} = 0 \quad , \tag{2.1}$$

where $\{\xi^\mu\}_{\mu=1}^{M}$ are ± 1 binary patterns representing the stored memories, and 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

$$X_i^{t+1} = \theta(f_i), \quad f_i = \frac{1}{N} \sum_{j=1}^{N} J_{ij} X_j^t \quad , \tag{2.2}$$

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

2.1.2 Low activity model

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

$$J_{ij} = g(W_{ij}) = g\left(\frac{1}{p(1-p)\sqrt{M}} \sum_{\mu=1}^{M} (\xi_i^\mu - p)(\xi_j^\mu - p)\right) \quad 1 \leq i \neq j \leq N; \quad J_{ii} = 0\,. \tag{2.3}$$

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 for the state of the network is similar to Eq.(2.2), with $\theta(f) = \frac{1+sign(f)}{2}$ and

$$f_i = \frac{1}{N} \sum_j J_{ij} X_j^t - T \quad , \tag{2.4}$$

where T is the neuronal threshold set to its optimal value (see Eq. (2.29). 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 Signal to noise analysis

To evaluate the impact of synaptic pruning on the network's performance, we study its effect on the signal to noise ratio (S/N) of the neuron's input field. The S/N is known to be the primary determinant of the retrieval capacity (ignoring higher order correlations in the neurons input fields) [Meilijson and Ruppin, 1996]. The network is started at a state X with overlap m_0^μ with memory ξ^μ; the overlap with other memories is assumed to be negligible. We show that in all the models analyzed, the S/N can be separated to similar ***independent*** factors, thus enabling to investigate the effect of synaptic modification independent of the model and the activity level in the network.

2.2.1 Hopfield model

In the modified Hopfield model (Eqs. 2.1,2.2) the weights $W_{ij} - \frac{\xi_i^\mu \xi_j^\mu}{\sqrt{M}}$ are 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

$$\begin{aligned} E\left[f_i|\xi_i^\mu\right] &= NE\left[\frac{1}{N}g(W_{ij})X_j\right] = m_0^\mu E\left[g(W_{ij})\xi_j^\mu\right] = \\ &= m_0^\mu \frac{1}{2} E\left[g(W_{ij})|\xi_j^\mu = +1\right] - m_0^\mu \frac{1}{2} E\left[g(W_{ij})|\xi_j^\mu = -1\right] \end{aligned} \tag{2.5}$$

The first term can be written as

$$\begin{aligned} & m_0^\mu \frac{1}{2} E\left[g(W_{ij})|\xi_j^\mu = +1\right] = \\ = \;& m_0^\mu \frac{1}{2}\int_{-\infty}^{\infty} g(W_{ij})\phi(W_{ij} - \frac{\xi_i^\mu}{\sqrt{M}})d(W_{ij} - \frac{\xi_i^\mu}{\sqrt{M}}) \approx \\ \approx \;& m_0^\mu \frac{1}{2}\int_{-\infty}^{\infty} g(W_{ij})\left[\phi(W_{ij}) - \frac{\xi_i^\mu}{\sqrt{M}}\phi'(W_{ij})\right] d(W_{ij}) = \\ = \;& m_0^\mu \frac{1}{2}\int_{-\infty}^{\infty} g(W_{ij})\left[\phi(W_{ij}) + \frac{\xi_i^\mu}{\sqrt{M}}\phi(W_{ij})W_{ij}\right] d(W_{ij}) = \\ = \;& m_0^\mu \frac{1}{2}\int_{-\infty}^{\infty} g(W_{ij})\phi(W_{ij})d(W_{ij}) + m_0^\mu \frac{1}{2}\frac{\xi_i^\mu}{\sqrt{M}}\int_{-\infty}^{\infty} g(W_{ij})W_{ij}\phi(W_{ij})d(W_{ij}) = \end{aligned} \tag{2.6}$$

$$= \quad m_0^\mu \frac{1}{2} E\left[g(z)\right] + m_0^\mu \frac{1}{2} \frac{\xi_i^\mu}{\sqrt{M}} E\left[z g(z)\right]$$

where z is a random variable with standard normal distribution. Repeating the same calculation for the second term yields

$$m_0^\mu \frac{1}{2} E\left[g(W_{ij}) | \xi_j^\mu = -1\right] \quad = \quad -m_0^\mu \frac{1}{2} E\left[g(z)\right] + m_0^\mu \frac{1}{2} \frac{\xi_i^\mu}{\sqrt{M}} E\left[z g(z)\right] , \qquad (2.7)$$

thus the expectation is

$$E(f_i | \xi_i^\mu) = m_0^\mu \frac{\xi_i^\mu}{\sqrt{M}} E\left[z g(z)\right] . \qquad (2.8)$$

The variance of the field can be written as the sum of three terms :

$$\begin{aligned} V\left[f_i | \xi_i\right] \quad &= \quad N E\left[\frac{1}{N^2} g^2(W_{ij}) X_j\right] - N E^2\left[\frac{1}{N} g(W_{ij}) X_j\right] + \qquad (2.9) \\ &+ \quad N(N-1) COV\left[g(W_{ij}) X_j, g(W_{ik}) X_k\right] . \end{aligned}$$

The first term is calculated similarly to Eq. 2.8:

$$\begin{aligned} & \frac{1}{N} E^2\left[g(W_{ij}) X_j\right] = \qquad (2.10) \\ = \quad & \frac{1}{N} m_0^\mu \int g^2(W_{ij}) \phi(W_{ij}) d(W_{ij}) - \frac{1}{N} m_0^\mu \int g^2(W_{ij}) \frac{\xi_i}{\sqrt{M}} W_{ij} \phi(W_{ij}) d(W_{ij}) = \\ & \frac{1}{N} m_0^\mu E\left[g^2(z)\right] - \frac{1}{N} m_0^\mu \frac{\xi_i}{\sqrt{M}} E\left[z g^2(z)\right] \stackrel{M\to\infty}{\longrightarrow} \frac{1}{N} m_0^\mu E\left[g^2(z)\right] \end{aligned}$$

The second term was calculated in Eq. 2.8, and could be neglected compared to the first term :

$$\frac{1}{N} \frac{m_0^2}{M} E^2\left[z g(z)\right] \stackrel{M\to\infty}{\longrightarrow} 0 \qquad (2.11)$$

The covariance is calculated by

$$COV\left[g(W_{ij}) X_j, g(W_{ik}) X_k\right] = E\left[g(W_{ij}) X_j g(W_{ik}) X_k\right] - E^2\left[g(W_{ij}) X_j\right] \qquad (2.12)$$

Denoting $W_{ij}^* = W_{ij} - \frac{\xi_i^\mu \xi_j^\mu}{\sqrt{M}}$, the first term can be written as

$$E\left[g(W_{ij}) X_j g(W_{ik}) X_k\right] = \qquad (2.13)$$

$$
\begin{aligned}
&= E\left[g(W_{ij}^* + \frac{\xi_i^\mu \xi_j^\mu}{\sqrt{M}})X_j g(W_{ik}^* + \frac{\xi_i^\mu \xi_k^\mu}{\sqrt{M}})X_k\right] \approx \\
&\approx E\left[\left(g(W_{ij}^*) + g'(W_{ij}^*)\frac{\xi_i^\mu \xi_j^\mu}{\sqrt{M}}\right) X_j \left(g(W_{ik}^*) + g'(W_{ik}^*)\frac{\xi_i^\mu \xi_k^\mu}{\sqrt{M}}\right) X_k\right] = \\
&= E\left[g(W_{ij}^*)g(W_{ik}^*)X_j X_k\right] + E\left[g'(W_{ij}^*)\frac{\xi_i^\mu \xi_j^\mu}{\sqrt{M}} g(W_{ik}^*)X_j X_k\right] + \\
&+ E\left[g'(W_{ik}^*)\frac{\xi_i^\mu \xi_k^\mu}{\sqrt{M}} g(W_{ij}^*)X_j X_k\right] + E\left[g'(W_{ij}^*)\frac{\xi_i^\mu \xi_j^\mu}{\sqrt{M}} g'(W_{ik}^*)\frac{\xi_i^\mu \xi_k^\mu}{\sqrt{M}} X_j X_k\right]
\end{aligned}
$$

As X_j, X_k are independent of $g'(W_{ij}^*)$ and $g'(W_{ik}^*)$, we can separate the product of the first three terms. Assuming g has zero expectation these terms vanish and the above equation reduces to

$$
E\left[\frac{\xi_j^\mu \xi_k^\mu}{M} X_j X_k\right] E\left[g'(W_{ij}^*)g'(W_{ik}^*)\right] \frac{1}{M} = \frac{1}{M} E^2\left[\frac{\xi_j^\mu}{\sqrt{M}} X_j\right] E\left[g'(W_{ij}^*)g'(W_{ik}^*)\right] \quad (2.14)
$$

The second term of Eq. 2.12 can be written as

$$
\begin{aligned}
E^2\left[g(W_{ij})X_j\right] &\approx E^2\left[\left(g(W_{ij}^*) + g'(W_{ij}^*)\frac{\xi_i^\mu \xi_j^\mu}{\sqrt{M}}\right) X_j\right] = \\
&= E^2\left[g'(W_{ij}^*)\frac{\xi_i^\mu \xi_j^\mu}{\sqrt{M}} X_j\right] = \\
&= \frac{1}{M^2} E^2\left[g'(W_{ij}^*)\xi_j^\mu X_j\right] = \\
&= \frac{1}{M} E^2\left[\frac{\xi_j^\mu}{\sqrt{M}} X_j\right] E^2\left[g'(W_{ij}^*)\right]
\end{aligned} \quad (2.15)
$$

The above covariance thus equals

$$
COV\left[g(W_{ij})X_j, g(W_{ik})X_k\right] = \frac{1}{M} E^2\left[\frac{\xi_j^\mu}{\sqrt{M}} X_j\right] COV\left[g'(W_{ij}^*), g'(W_{ik}^*)\right]. \quad (2.16)
$$

But the last covariance equals zero, as

$$
\begin{aligned}
E\left[g'(W_{ij}^*), g'(W_{ik}^*)\right] &= \int\int g'(W_{ij}^*)g'(W_{ik}^*)d(W_{ij}^*)d(W_{ik}^*) \\
&= \int g'(W_{ij}^*)d(W_{ij}^*) \int g'(W_{ik}^*)d(W_{ik}^*) \\
&= E\left[g'(W_{ij}^*)\right] E\left[g'(W_{ik}^*)\right]
\end{aligned}
$$

Therefore, the variance of the field is

$$V(f_i|\xi_i^\mu) = NE\left[g^2(z)\right] \quad . \tag{2.17}$$

Hence

$$\frac{S}{N}(\xi_i) = \frac{E(f_i|\xi_i^\mu = 1) - E(f_i|\xi_i^\mu = 0)}{\sqrt{V(f_i|\xi_i^\mu)}} = \sqrt{\frac{N}{M}} m_0^\mu \frac{E\left[zg(z)\right]}{E\left[g^2(z)\right]}. \tag{2.18}$$

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

$$\frac{S}{N}(\xi_i) = \frac{1}{\sqrt{\alpha}} m_0^\mu \, \rho(g(z), z)\,, \tag{2.19}$$

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.

2.2.2 Low activity model

In the low activity model (Eqs. 2.3 and 2.4) the network is initialized with activity p and a noise probability $P(X_i = 0|\xi_i = 1) = \epsilon$. This induces the following probabilities:

$$\begin{cases} P(X_i = 0|\xi_i = 0) & = & 1 - \frac{p\epsilon}{1-p} \\ P(X_i = 0|\xi_i = 1) & = & \epsilon \\ P(X_i = 1|\xi_i = 0) & = & \frac{p\epsilon}{1-p} \\ P(X_i = 1|\xi_i = 1) & = & 1 - \epsilon \end{cases}$$

This implies an initial overlap of

$$\begin{aligned} m_0 &= \frac{1}{Np(1-p)} \sum_{i=1}^{N} (\xi_i - p) X_i = \qquad (2.20) \\ &= \frac{1}{Np(1-p)} \left[\sum_{\xi_i=1}^{N} (\xi_i - p) X_i + \sum_{\xi_i=0}^{N} (\xi_i - p) X_i \right] = \\ &= \frac{1}{Np(1-p)} \left[(1-p) p N P(X_i = 1|\xi_i = 1) - p(1-p) N P(X_i = 1|\xi_i = 0) \right] = \end{aligned}$$

$$
\begin{aligned}
&= P(X_i = 1|\xi_i = 1) - P(X_i = 1|\xi_i = 0) = \\
&= (1-\epsilon) - \frac{p\epsilon}{(1-p)} \\
&= \frac{(1-p-\epsilon)}{(1-p)}
\end{aligned}
$$

between the initial pattern X and the memory pattern ξ^μ. To calculate the fields first moment, we write

$$
E(f_i|\xi_i) = NP(\xi_j = 1)E\left[\frac{1}{N}g(W_{ij})|\xi_j = 1\right] + NP(\xi_j = 0)E\left[\frac{1}{N}g(W_{ij})|\xi_j = 0\right] - T \tag{2.21}
$$

The first term is calculated as follows

$$
\begin{aligned}
& P(\xi_j = 1)E\left[g(W_{ij})|\xi_j = 1\right] = \qquad (2.22)\\
&= p(1-\epsilon)\int g(W_{ij})\phi(W_{ij} - \frac{(1-p)(\xi_i - p)}{p(1-p)\sqrt{M}})d(W_{ij}) = \\
&= p(1-\epsilon)\int g(W_{ij})\left[\phi(W_{ij}) - \frac{(1-p)(\xi_i - p)}{p(1-p)\sqrt{M}}\phi'(W_{ij})\right]d(W_{ij}) = \\
&= p(1-\epsilon)\int g(W_{ij})\left[\phi(W_{ij}) + \frac{(1-p)(\xi_i - p)}{p(1-p)\sqrt{M}}\left[W_{ij}\phi(W_{ij})\right]\right]d(W_{ij}) = \\
&= p(1-\epsilon)E\left[g(z)\right] + p(1-\epsilon)\frac{(1-p)(\xi_i - p)}{p(1-p)\sqrt{M}}E\left[zg(z)\right]
\end{aligned}
$$

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

$$
\begin{aligned}
E(f_i|\xi_i) &= p(1-p-\epsilon)\frac{(\xi_i - p)}{p(1-p)\sqrt{M}}E\left[zg(z)\right] - T \\
&= \frac{(1-p-\epsilon)}{(1-p)}\frac{(\xi_i - p)}{\sqrt{M}}E\left[zg(z)\right] - T \\
&= m_0\frac{(\xi_i - p)}{\sqrt{M}}E\left[zg(z)\right] - T.
\end{aligned} \tag{2.23}
$$

The variance is calculated following

$$
\begin{aligned}
V(f_i|\xi_i) &= NE(\frac{1}{N^2}g^2(W_{ij})X_j) - NE^2(\frac{1}{N}g(W_{ij})X_j) + \\
&+ N(N-1)Cov(\frac{1}{N}g(W_{ij})X_j, \frac{1}{N}g(W_{ik})X_k),
\end{aligned} \tag{2.24}
$$

in a similar manner to yield

$$\frac{E(f_i|\xi_i)}{\sqrt{V(f_i|\xi_i)}} = \frac{\frac{N}{\sqrt{M}} m_o(\xi_i - p) E\left[z g(z)\right] - T}{\sqrt{N p E\left[g^2(z)\right]}}. \tag{2.25}$$

We now proceed to derive the threshold which maximizes the network's performance. We first calculate the overlap after one step as a function of the S/N, and then we find the optimal threshold. Given the overlap m_0 between the network's initial state and a pattern ξ^μ we calculate the overlap in the next step m_1 similarly to Eq. 2.20 by

$$\begin{aligned} m_1 &= P(X_i = 1|\xi_i = 1) - P(X_i = 1|\xi_i = 0) = \\ &= P(f_i > 0|\xi_i = 1) - P(f_i > 0|\xi_i = 0) = \\ &= \Phi(\frac{E[f_i|\xi_i]}{V[f_i|\xi_i]}|\xi_i = 1) - \Phi(\frac{E[f_i|\xi_i]}{V[f_i|\xi_i]}|\xi_i = 0), \end{aligned} \tag{2.26}$$

where Φ is the Gaussian cumulative distribution.

In order to find the threshold that maximizes the overlap, we differentiate m_1 (Eq. 2.26) with respect to T

$$\frac{\partial m_1}{\partial T} = \frac{\partial \left[\Phi(\frac{E[f_i|\xi_i]}{V[f_i|\xi_i]}|\xi_i = 1) - \Phi(\frac{E[f_i|\xi_i]}{V[f_i|\xi_i]}|\xi_i = 0)\right]}{\partial T} = 0 \tag{2.27}$$

Denoting $\frac{E[f_i|\xi_i]}{V[f_i|\xi_i]}|\xi_i = \frac{S_{\xi_i} - T}{R}$, this yields

$$\begin{aligned} \phi(\frac{S_0 - T}{R}) &= \phi(\frac{S_1 - T}{R}) \\ (S_0 - T)^2 &= (S_1 - T)^2 \\ T &= \frac{1}{2}(S_0 + S_1) \end{aligned} \tag{2.28}$$

and

$$T = \frac{N}{\sqrt{M}} m_o(\xi_i - p) E\left[z g(z)\right] - T\,. \tag{2.29}$$

Using the optimal threshold in the signal to noise expression (Eq 2.25) we receive

$$\frac{S}{N} = \frac{E(f_i|\xi_i = 1) - E(f_i|\xi_i = 0)}{\sqrt{V(f_i|\xi_i)}} = \sqrt{\frac{N}{M}}\, m_0 \frac{1}{\sqrt{p}} \rho(g_t(z), z) \quad . \tag{2.30}$$

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.

2.3 Extending the analysis

In this section three further extensions of the above analysis are presented. First, the results are applied to another low activity model. Then, we incorporate noise into the models either by adding a normally distributed noise to the synaptic matrix or by considering a stochastic network's dynamics.

2.3.1 The Tsodyks model

The following low activity model was suggested by [Tsodyks, 1989]. In this model synaptic efficacies are described by:

$$J_{ij} = g(W_{ij}) = g\left(\frac{1}{p\sqrt{M(1-p^2)}} \sum_{\mu=1}^{M} (\xi_i^\mu \xi_j^\mu - p^2)\right) \quad , \tag{2.31}$$

and the updating rule for the network's state remains as in Eq. 2.4. Analysis similar to the above (Eq. 2.21 to 2.30) yields that in this model too, the signal to noise can be separated to independent factors at the following way:

$$\frac{S}{N} = \sqrt{\frac{M}{N}} \frac{1-\epsilon}{\sqrt{p(1-p^2)}} \rho(g(z), z). \tag{2.32}$$

Therefore, the S/N has the same form as in the models described above (except for the difference in the coefficient of ρ), and can be studied in an identical way.

2.3.2 Noisy synaptic matrix

Further extension of the above analysis arises when relaxing the assumption that the synaptic matrix is determined solely by the memory patterns. The results remain valid even when the initial synaptic matrix contains some normally distributed noise. The analysis brought here is for the low activity model with zero mean noise, but it also applies to non-zero mean as the neuronal threshold may be adjusted accordingly. We modify Eq. 2.3 and write

$$J_{ij} = g(W_{ij}) = g\left(\sum_{\mu=1}^{M}(\xi_i^\mu - p)(\xi_j^\mu - p) + \eta_{ij}\right), \tag{2.33}$$

where η_{ij} is a noise factor normally distributed with variance σ^2 and zero expectancy. W_{ij} is now distributed $N(0, p^2(1-p)^2 M + \sigma^2)$, hence we denote

$$\widehat{W_{ij}} = W_{ij}/\sqrt{\sigma^2 + p^2(1-p)^2 M} \tag{2.34}$$

which has standard normal distribution, and repeat Eqs. 2.21-2.23 with $\widehat{W_{ij}}$ instead of W_{ij} as if g operates on the normalized weights, yielding

$$S/N = \sqrt{\frac{N}{M}}\, m_0 \frac{1}{\sqrt{p}} \left[\sqrt{\frac{Mp^2(1-p)^2}{Mp^2(1-p)^2 + \sigma^2}}\right] \rho(g_t(z), z) \tag{2.35}$$

2.3.3 Stochastic dynamics

In this subsection we consider a noisy network's dynamics motivated by the fact that neurons operate in a noisy environment. We discuss the conditions under which the stochastic dynamics can be studied similarly to the previous analysis. We replace the update rule of the network (Eq. 2.2, 2.4) with

$$X_i^{t+1} = \theta(f_i + \eta_i) \tag{2.36}$$

where η_i is a noise term with standard normal distribution. Previous studies ([Amit, 1989]) showed that this dynamics is equivalent to the following dynamics

$$X_i^{t+1} = S(f_i) \tag{2.37}$$

where $S(x)$ is the stochastic sigmoid function:

$$S(x) = \begin{cases} 1 & \text{with probability} \quad \frac{1}{1+e^{-x\beta}} \\ 0 & \text{otherwise} \end{cases}$$

with noise level β. Others have shown that the S/N in this case depends on $\frac{E}{\sqrt{V+c/\beta^2}}$ instead of $\frac{E}{\sqrt{V}}$ [Horn and Ruppin, 1995].

We now assume that increasing the synaptic values increases the related noise. This view is in correlation with the experimental results studying the sources of the noise: synapses were shown to release quantal vesicles with Poisson distribution which depends on the synaptic strength [Katz and Miledi, 1967] and the size of the synaptic vesicles has normal distribution [Amit, 1989]. Under these assumptions we can write

$$\frac{S}{N} = c' \frac{E}{\sqrt{V + Vc/\beta^2}} = c''_\beta \frac{E}{\sqrt{V}}. \tag{2.38}$$

Hence the stochastic case differs from the zero-temperature case by a constant factor which depends on the temperature, and the above analysis can be applied to this case too.

2.4 Summary of the signal to noise analysis

As shown above (Eqs. 2.19, 2.30 and 2.35), in all three models presented the only effect of the modification function g on the S/N is through the correlation coefficient. The analysis applies both to the cases of zero noise and when static and dynamic noise are incorporated into the models. The effect of synaptic deletion is independent of the network's activity level and of the initial overlap between the memory pattern stored, and the pattern presented to the network. Hence, the behavior of the different models under synaptic modification can be investigated by analyzing $\rho(g(z), z)$ only, regardless of the other parameters.

Chapter 3

Optimal modification functions

Following the above analysis, the presentation of the signal to noise as a product of independent terms, allows the investigation of the modification function independently of other parameters. We proceed to consider different possible synaptic constraints and to study modification functions that optimize performance under these constraints.

3.1 Pruning does not improve performance

The immediate consequence of Eqs. (2.19) and (2.30) is that there is no local synaptic modification function that can improve the performance of the Hebbian network, since ρ has 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 performance than the intact network. A similar result was previously shown by [Sompolinsky, 1988] in the Hopfield model. The current use of signal-to-noise analysis enables us to proceed and derive optimal functions under different constraints on modification functions, and evaluate the performance of non-optimal functions.

When no constraints are involved, pruning has no beneficial effect. However, since synaptic activity is a major consumer of energy 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 constraints originating from the assumed energy restriction: limited number of synapses, and limited total synaptic efficacy [1].

3.2 Optimal modification with limited number of synapses

In this section we find the optimal synaptic modification strategy when the amount 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 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])

$$\int_A zg(z)\phi(z)dz - \gamma(\int_A g^2(z)\phi(z)dz - c_1) \tag{3.1}$$

for some constant c_1. Denoting $g_i = g(z_i)$ we approximate (3.1) by

$$\sum_{\{i|z_i\in A\}} z_i g_i \phi(z_i) - \gamma(\sum_{\{i|z_i\in A\}} g_i^2\phi(z_i) - c_1'). \tag{3.2}$$

[1] 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].

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).

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

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

Since $zg_A(z) = z^2 R_A(z) = g_A^2(z)$, $E\left[zg_A(z)\right] = E\left[g_A^2(z)\right]$ 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.

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

$$g_t(z) = zR_t(z) \quad \text{where} R_t(z) = \begin{cases} 1 & \text{when } |z| > t \\ 0 & otherwise \end{cases} \tag{3.3}$$

where t is the threshold beyond which weights are removed. Using the fact that $\phi'(z) = -z\phi(z)$, and integrating by parts, we use the following equations

$$\int_t^\infty z\phi(z)dz = -\int_t^\infty d(\phi(z)) = \phi(t) \tag{3.4}$$

$$\int_t^\infty z^2\phi(z)dz = t\phi(t) + \int_t^\infty \phi(z)dz = t\phi(t) + \Phi^*(t) \tag{3.5}$$

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

$$E\left[zg_t(z)\right] = E\left[g_t^2(z)\right] = \tag{3.6}$$

$$\begin{aligned} &= \int_{-\infty}^\infty z^2\, R_t(z)\, \phi(z)dz = \\ &= 2\int_t^\infty z\, z\, \phi(z)dz = \\ &= 2\left[\Phi^*(t) + t\phi(t)\right] \end{aligned}$$

and

$$\rho(g_t(z), z) = \sqrt{2t\phi(t) + 2\Phi^*(t)}. \tag{3.7}$$

The resulting minimal value deletion strategy $g_t(z)$ is illustrated in figure 3.1(a).

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

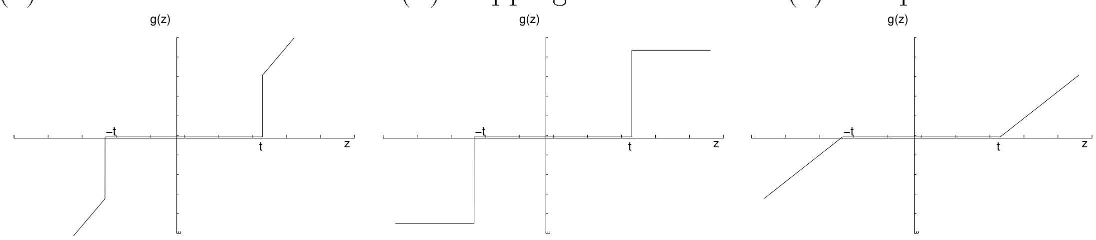

Figure 3.1: Different synaptic modification strategies. (a) Minimal value deletion: $g(z) = z$ for all $|z| > t$ and zero otherwise (see Eq. 3.3). (b) Clipping: $g(z) = sign(z)$ for all $|z| > t$ and zero otherwise (see Eq. 3.16). (c) Compressed synapses: $g(z) = z - sign(z)t$ for all $|z| > t$ and zero otherwise (see Eq. 3.12).

3.3 Optimal modification with limited 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. To derive the optimal modification function for this criterion, we wish to maximize S/N while keeping the total synaptic strength $\int |g(z)|$ fixed. Using the Lagrange method we have

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

$$= \int_{-\infty}^{\infty} |z||g(z)|\phi(z)dz - \gamma_1(\int_{-\infty}^{\infty} |g(z)|^2\phi(z)dz - c_1) - \gamma_2(\int_{-\infty}^{\infty} |g(z)|\phi(z)dz - c_2) \tag{3.8}$$

which is approximated by

$$\sum_i |z_i||g_i|\phi(z_i) - \gamma_1(\sum_i |g_i|^2\phi(z_i) - c_1') - \gamma_2(\sum_i |g_i|\phi(z_i) - c_2') \tag{3.9}$$

Assuming $g(z)$ to be piece-wise equicontinuous and equating to zero the derivative with respect to $|g_i|$ we obtain [2]

$$|z_i|\phi(z_i) - \gamma_1 2|g_i|\phi(z_i) - \gamma_2\phi(z_i) = 0 \tag{3.10}$$

or

$$|g(z)| = \frac{1}{2\gamma_1}(|z| - \gamma_2)\,, \tag{3.11}$$

from where

$$g_t(z) = \begin{cases} z - t & \text{when} \quad z > t \\ 0 & \text{when} \quad |z| < t \\ z + t & \text{when} \quad z < -t \end{cases} \tag{3.12}$$

that is, the absolute value of all synapses with magnitude above some threshold t is reduced by t, and the rest are eliminated. We denote this modification function ***"compressed deletion"***, and it is illustrated in figure 3.1(c).

The S/N under this strategy is calculated using $R_t(z)$ described above by writing

$$\begin{aligned} E\left[zg_t(z)\right] &= \int_{-\infty}^{\infty}(z-t)R_t(z)\phi(z)dz = \\ &= 2\int_t^{\infty}(z^2 - tz)\phi(z)dz = \\ &= 2\left[\Phi^*(t) + t\phi(t)\right] - 2t\phi(t) = \\ &= 2\Phi^*(t) \end{aligned} \tag{3.13}$$

$$\tag{3.14}$$

$$\begin{aligned} E\left[g_t^2(z)\right] &= \int_{-\infty}^{\infty}(z-t)^2 R_t^2(z)\phi(z)dz = \\ &= 2\int_t^{\infty}(z^2 - 2tz + t^2)\phi(z)dz = \\ &= 2\left[\Phi^*(t) + t\phi(t)\right] - 4t\phi(t) + 2t^2\Phi^*(t) = \\ &= (1+t^2)2\Phi^*(t) - 2t\phi(t) \end{aligned}$$

[2] As $|g_i|$ is not derivable at 0, a special treatment is needed at the neighborhood of zero. See [Meilijson and Ruppin, 1996] for details.

yielding

$$\rho(g_t(z), z) = \frac{2\Phi^*(t)}{\sqrt{(1+t^2)2\Phi^*(t) - 2t\phi(t)}}. \tag{3.15}$$

3.4 Clipped synapses

The assumption that synapses can hold values with arbitrary precision is implausible in biological networks. In order to investigate the behavior of a network whose synaptic values are limited to a small set of discrete values, [Sompolinsky, 1988] has investigated modification of Hebbian synapses in the Hopfield model, by clipping them to ± 1 values (see 3.1(b)). In our formalism this modification function is written as

$$g_t(z) = \begin{cases} +1 & \text{when} \quad t < z \\ 0 & \text{when} \quad |z| < t \\ -1 & \text{when} \quad z < -t \end{cases} \tag{3.16}$$

Using Eqs. 3.4 and 3.5 we calculate the signal to noise ratio under the clipping strategy :

$$E\left[zg_t(z)\right] = \int_{-\infty}^{\infty} zg_t(z)\phi(z)dz = 2\int_t^{\infty} z\phi(z)dz = 2\phi(t) \tag{3.17}$$

$$\begin{aligned} E\left[g_t^2(z)\right] &= \int_{-\infty}^{\infty} g_t^2(z)\phi(z)dz = \\ &= \int_{-\infty}^{0} g_t(z)\phi(z)dz + \int_0^{\infty} -g_t(z)\phi(z)dz = \\ &= 2\int_t^{\infty} \phi(z)dz = 2\Phi^*(t), \end{aligned} \tag{3.18}$$

hence

$$\rho(g_t(z), z) = \frac{2\phi(t)}{\sqrt{2\Phi^*(t)}}. \tag{3.19}$$

In the following chapters we use this derivation to compare the clipping strategy with the optimal strategies derived above.

Chapter 4

Networks with limited number of synapses

The above analysis shows that if the number of synapses in the network must be reduced, minimal value deletion will minimize the damage, yet deletion reduces performance and is hence unfavorable. We now proceed to investigate the case where the amount of synapses is restricted in the adult but a process of initial synaptic overgrowth followed by synaptic deletion can be implemented. We study such a process by comparing networks that have the same amount of synaptic resources, yet differ by their size and connectivity. As we have shown, in a network with N neurons, the S/N is proportional to $\sqrt{N}\rho(g_t(z), z)$. Denoting r_t as the percentage of remaining synapses after the deletion function $g_t(z)$ is applied to the fully connected network, the adult network is left with $r_t N^2$ synapses. In order to maximize the performance, we maximize

$$S/N \propto \sqrt{N}\rho(g_t(z), z) \quad \text{with} \quad r_t N^2 = const \tag{4.1}$$

As $N \propto \sqrt{const/r_t}$, we choose the t which maximizes $\frac{\rho(g_t(z),z)}{\sqrt[4]{r_t}}$ for any given modification function. Figure 4.1 shows $\frac{\rho(g_t(z),z)}{\sqrt[4]{r_t}}$ as a function of r_t for the modification functions analyzed above. A high gain of performance in bigger (though sparser) networks can be observed in all three synaptic modification functions.

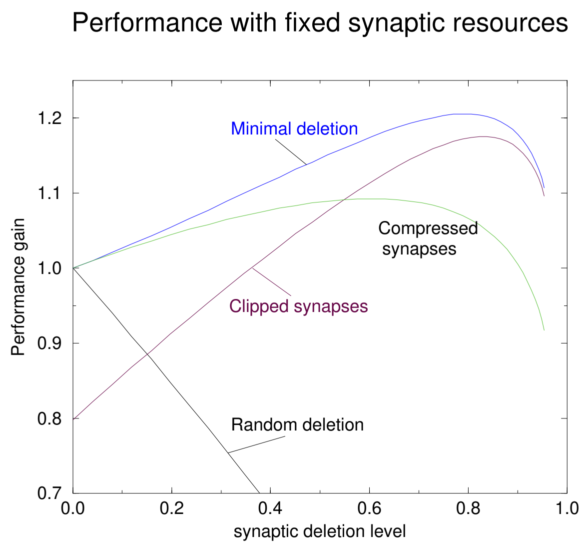

Figure 4.1: The performance in networks with fixed synaptic resources (total number of synapses), but different number of neurons under different synaptic modification functions. The data is derived using Eq. 4.1 which is applied to the modification functions studied in the previous section.

The resulting conclusion is that given a limited number of synapses it may be better to have them sparsely connect neurons in a large network. This is true if indeed a clever pruning takes place, but not if deletion is random. It hence follows that to achieve efficient networks, one must start with densely connected overgrown networks and judiciously prune them.

Chapter 5

Numerical results

5.1 Capacity under deletion

To quantitatively evaluate the way performance is effected by the strategies described in the previous section, we measure the network's performance by calculating the 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 $m_0^\mu = 0.8$, after one or ten iterations. Below are shown simulations performed in the modified Hopfield model with $N = 800$, and simulations performed in the low activity network with $N = 800$ neurons and coding level $p = 0.1$.

Figures 5.1 to 5.3 compare three modification strategies: minimal value deletion (Eq. 3.3), random deletion (independent of the weights strengths) and a clipping deletion strategy. In clipping deletion, all weights with magnitude below some threshold value are removed, and the remaining ones are assigned a ± 1 value, according to their sign (see figure 3.1(b) and Eq. 3.16). In all simulations presented, minimal-value deletion is indeed significantly better than the other deletion strategies. In high deletion levels, it is almost equaled by the clipping strategy.

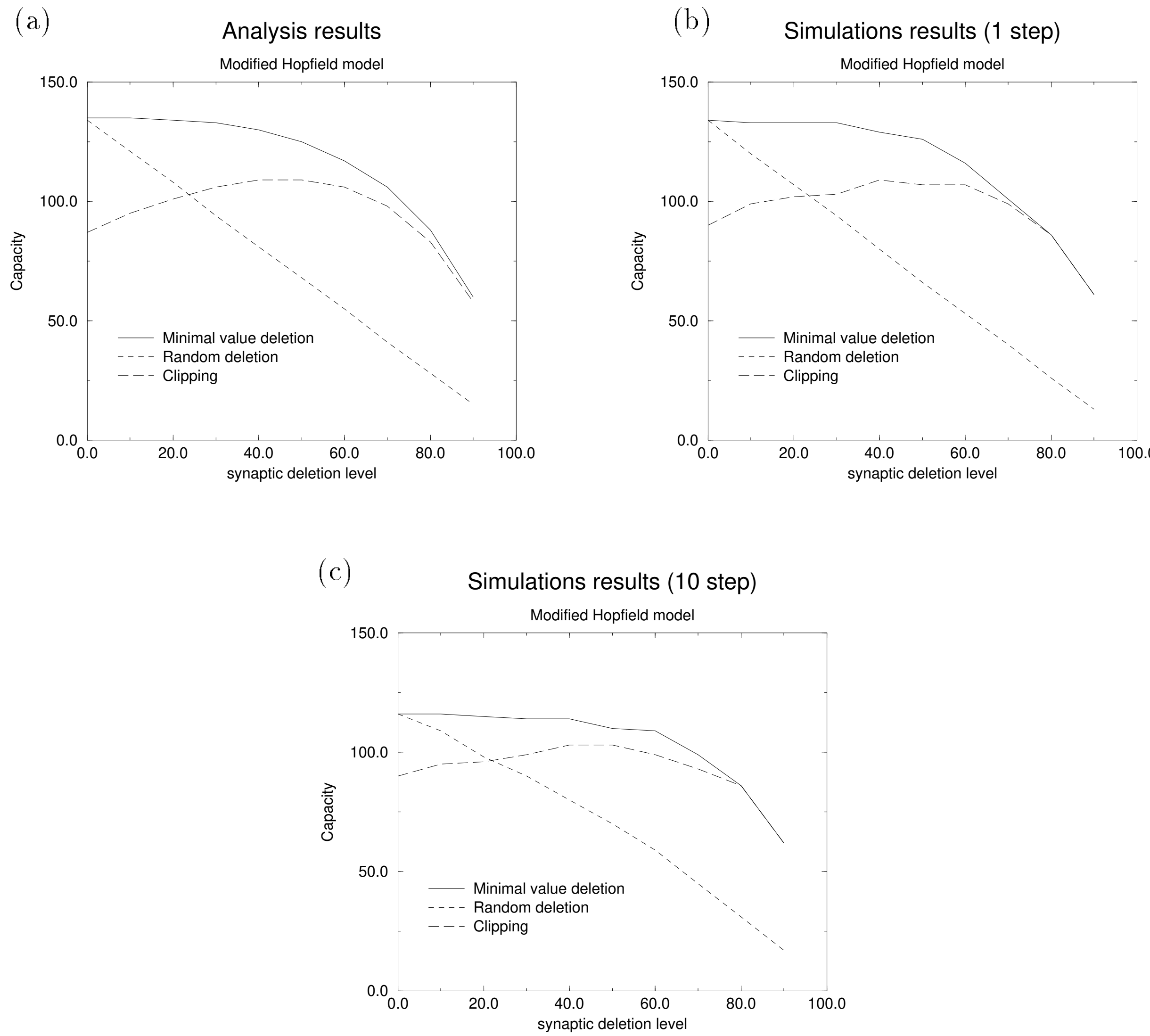

Figure 5.1: Capacity of a modified Hopfield network with different synaptic modification strategies as a function of the synaptic deletion level. Both analytic and simulation results of single step and multiple step dynamics are presented, showing a close correspondence.

In the low activity model, two sets of simulations are presented. Figure 5.2 displays simulations performed in the low activity model with an arbitrary fixed threshold. This threshold was chosen to maximize the performance of the fully connected network in the cases of minimal deletion and random deletion strategies and to maximize the performance in optimal deletion level in the clipping strategy.

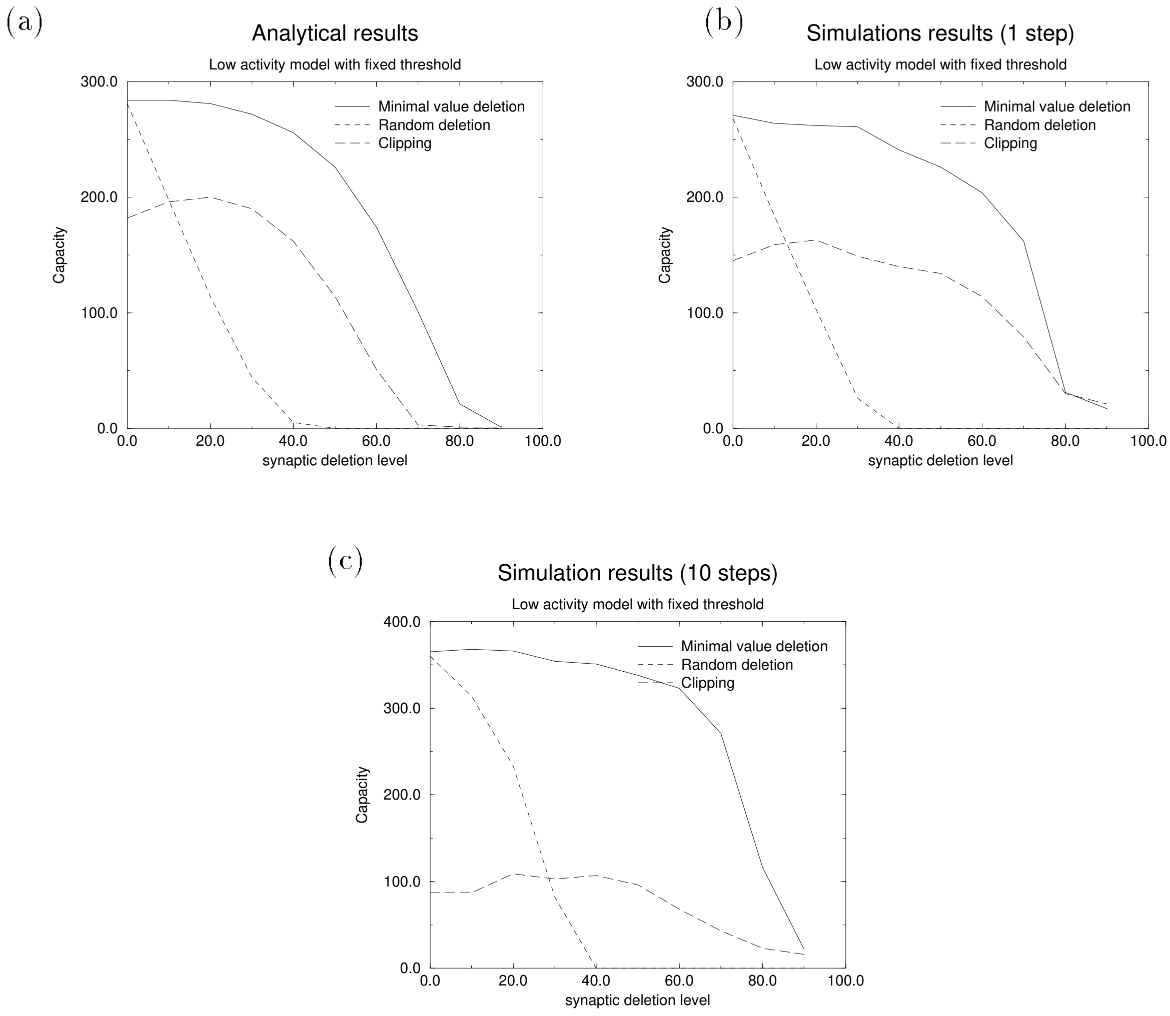

Figure 5.2: Capacity of a network with different synaptic modification strategies as a function of the synaptic deletion level. The figure shows results of the low activity model with fixed threshold. Both analytic and simulation results of single step and multiple step dynamics are presented, showing a fairly close correspondence.

Figure 5.3 displays simulations performed in the low activity model with a threshold optimally tuned for each deletion level. In one-step simulations, the optimal threshold was determined according to Eq. 2.29, and in ten-steps simulations, the optimal threshold was found numerically to maximize the network's performance.

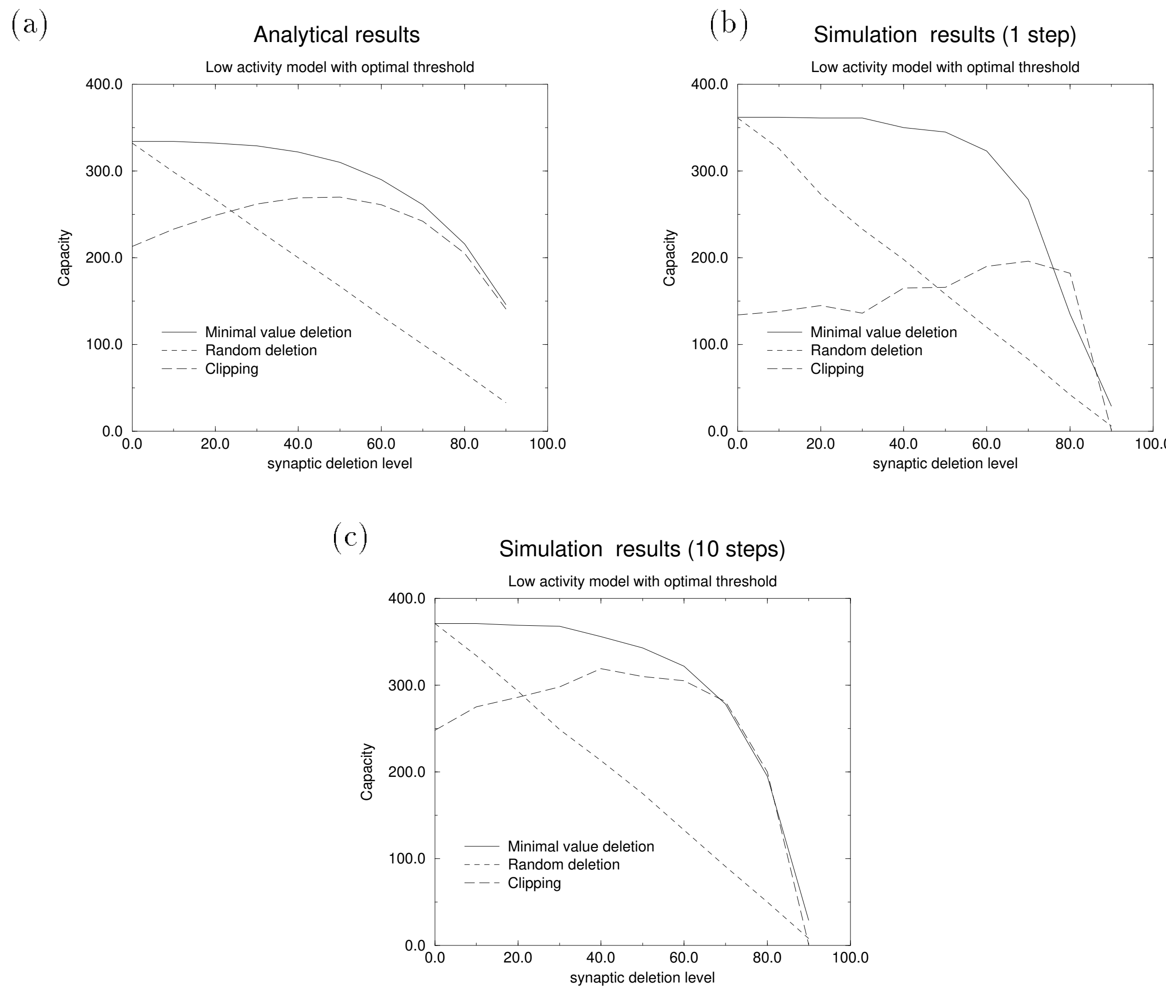

Figure 5.3: Capacity of a network with different synaptic modification strategies as a function of the synaptic deletion level. The figure shows results with optimal neural threshold (i.e. threshold that is varied with the deletion level). Both analytic and simulation results of single step and multiple step dynamics are presented, showing a fairly close correspondence.

Figure 5.4 compares the "compressed-deletion" modification strategy (Eq. 3.12) to random deletion, as a function of the total synaptic strength of the network.

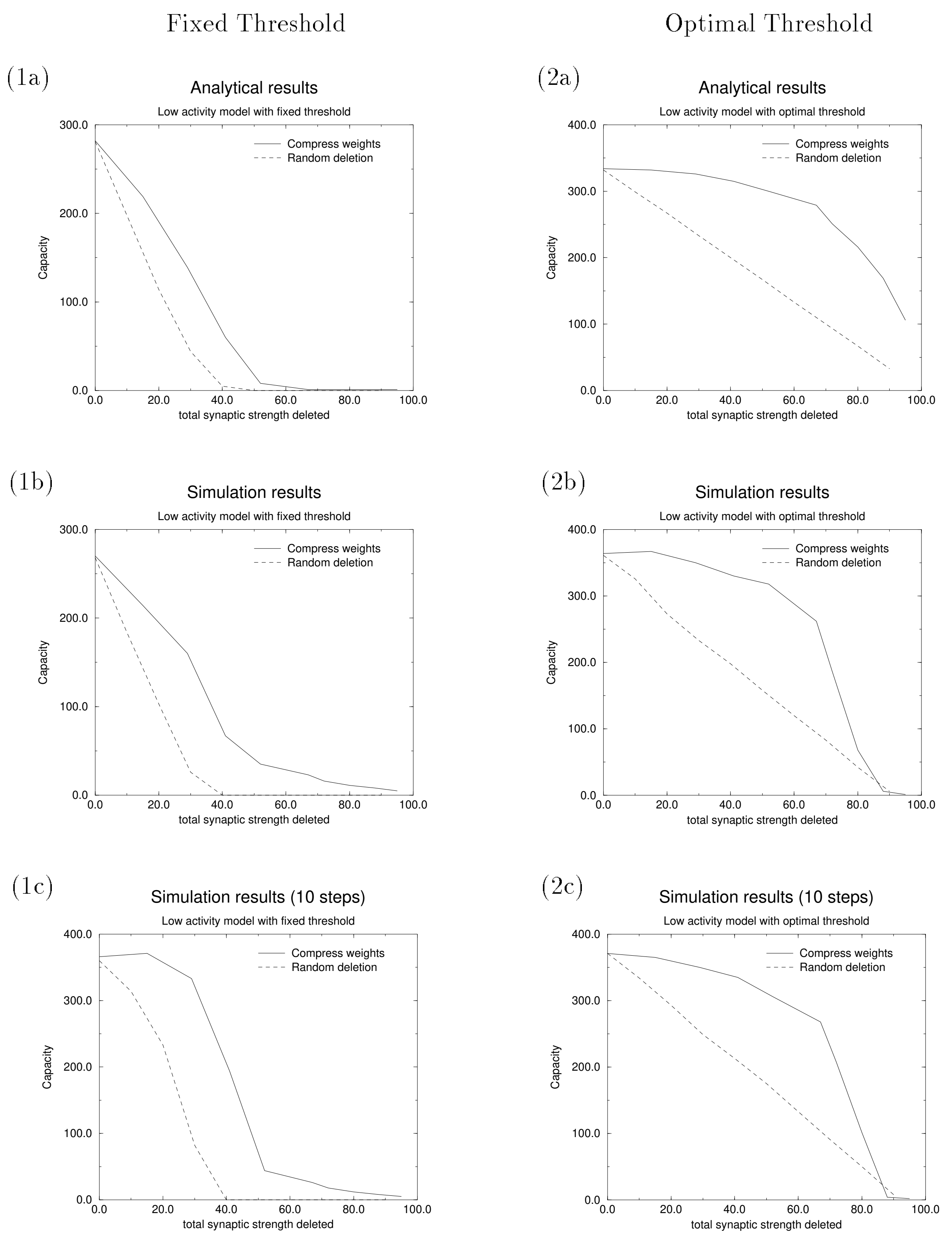

Figure 5.4: Capacity of a network with different synaptic modification strategies as a function of the total synaptic strength in the network. The left column shows results of the low activity model with fixed threshold, while results with optimal threshold are shown in the right column.

5.2 Optimal deletion level

The above results demonstrate 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, an initial over-growth of synapses then followed by deletion, reveals itself as beneficial. Figure 5.5(a) compares the performance of networks with the same synaptic resources, but with varying number of neurons. The smallest network ($N = 800$) is fully connected while larger networks are pruned according to the minimal value deletion strategy to end up with the same amount of synapses. The optimal deletion ratio is found around 80% deletion, and improves performance by 45% . This optimal 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 5.5(b) shows a similar comparison for the limited synaptic strength criterion, comparing networks with same total synaptic strength. Here the optimal capacity is obtained at 70% deletion yielding an improvement of more than 20% in capacity and 120% in information storage.

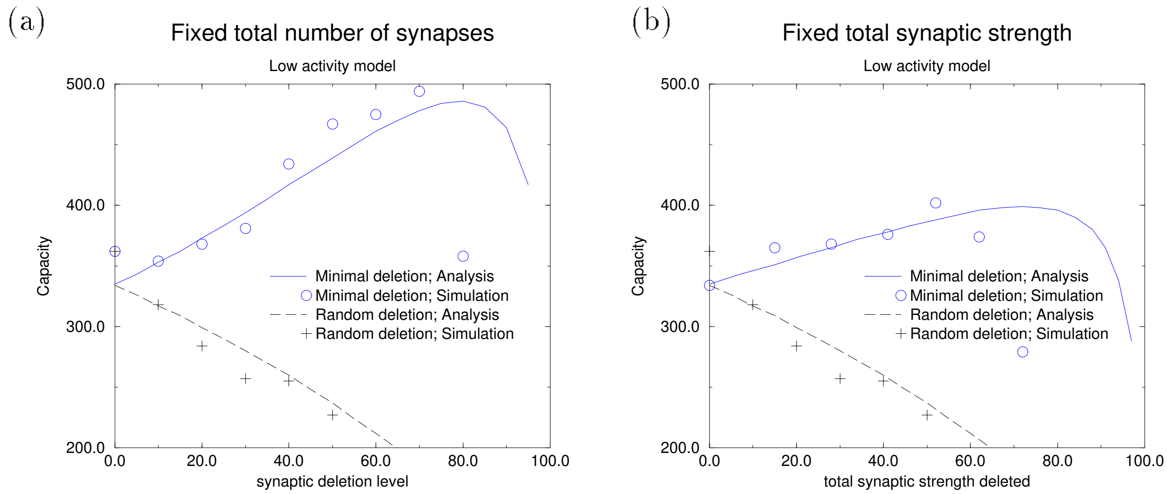

Figure 5.5: (a) Capacity of networks with different number of neurons but the same total number of synapses as a function of network connectivity. The bigger networks (networks with more neurons) are pruned according to minimal value deletion to keep the total number of synapses ($\boldsymbol{k}$) constant. Capacity at the optimal deletion range is 45% higher than in the fully connected network. (b) Capacity of networks with the same total synaptic strength but different sizes. The bigger networks are pruned according to the compressed deletion strategy to remain with the same total synaptic strength as the fully connected network. Simulation parameters are $\boldsymbol{k} = 800^2$, $\boldsymbol{p} = 0.1$, and T is kept optimal.

5.3 Continuous storage and deletion

Until now, we have analyzed synaptic deletion of previously wired synaptic matrices (storing a fixed set of memories). To simulate the continuous process of learning and deletion occurring in the brain, we perform an experiment that is geared at mimicking the profile of synaptic density changes occurring during human development and maturation. These changes naturally define a time step equivalent to one year, such that within each time step we store some memories and changed 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 5.6). Addition of synapses is done randomly (in agreement with experimental results of [Bourgeois ***et al.***, 89] testifying that it occurs in an experience independent manner),while synaptic deletion 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 5.6 traces the networks performance during this experiment. It superimposes the synaptic density (dot-dashed line) and memory performance data. Two observations should be noted: the 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".

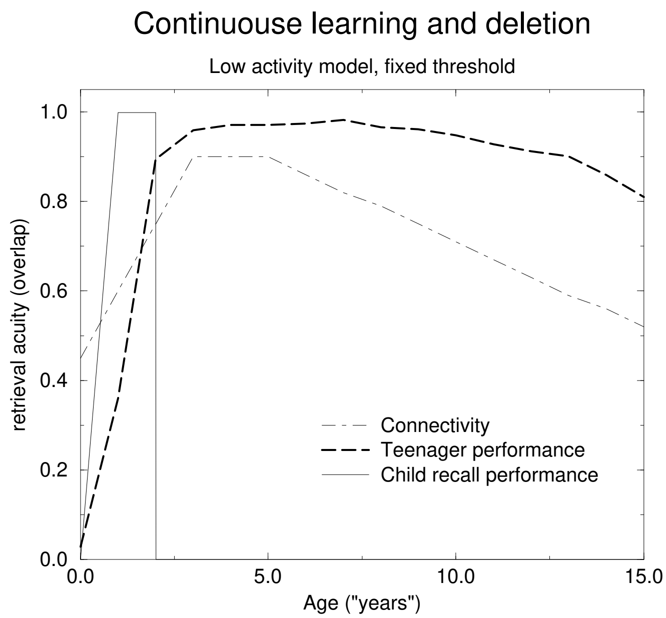

Figure 5.6: Memory retrieval as a function of storage period. The figure displays synaptic density and memory performance data. At each time step ("year") m memories are stored in the network, and network's connectivity is changed following human data (dot-dashed line). Network is tested for retrieval twice: both 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). Network parameters are $N = 800$, $m = 10$ and $p = 0.1$. The threshold is kept fixed at $T = (1/2 - p)p(1 - p)$.

Chapter 6

Discussion

We have analyzed the effect of modifying Hebbian synapses in an optimal way that maximizes memory performance, either when no synaptic constraints are involved or when 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 efficacy, and are not affected by initial noise in the synaptic matrix or in the network dynamics. They do not depend on the activity level of the network, or on the initial overlap between the pattern presented to the network and the memory pattern stored in the network. Moreover, they are exactly the same functions in a large family of associative memory networks.

The above analysis and simulations show, that when synapses are not limited, the optimal modification function of the original Hebbian learning rule is the identity function. Thus elimination of synapses cannot improve the network's performance. Under a restricted number of synapses, the optimal local modification function of a given Hebbian matrix is to delete the small valued weights, and maintain the values of the remaining connections. When the total synaptic strength of the network is limited, the optimal modification function involves deletion of small-magnitude synapses and reduction in the strength of the remaining ones. Clearly, the actual constraints cannot be measured, and must be estimated (e.g. according to energy consumption), yielding a weighted mixture of the constraints described above.

Our results predict that during the elimination phase in the brain synapses undergo weight-dependent pruning in a way that deletes the weak synapses. The experimental data in this regard is yet inconclusive, but interestingly, recent studies have found that in the neuro-muscular junction synapses are indeed pruned according to their initial synaptic strength [Frank, 1997]. The described dynamics of the synaptic changes resembles a strategy which leaves two possible values for synapses : zeroed synapses (which are eliminated), and strong synapses.

6.1 Why are synapses deleted?

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 for this question. Some researchers ([Wolff *et al.*, 1995]), had hypothesized that synaptic elimination can improve network performance, by removing interfering synapses, but this paper proves this argument to be incorrect in a large family of associative memory network 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 if the size of the network is too large, a reduction in the amount of free parameters may improve the ability of the network to generalize [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 figures 5.2 and 5.3), maximal performance is obtained at non-zero deletion levels. However the biological plausibility of uni-valued synapses is in doubt, and the performance gain is mild.

This thesis proposes that synaptic over-growth and deletion emerge from synaptic constraints. As energy consumption in the brain is highly correlated with synaptic density and attributed to axonal processes, (but roughly independent of the number of neurons), synapses may be a restricted resource that must be scrupulously utilized. The deletion strategies described above damage the performance only slightly compared to the energy saved. Therefore, if we have to use a restricted amount of synapses in the adult, better performance is achieved if the synapses are first overgrown, investing more energy for a limited period, and then are 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 other pruned networks, with the same total number of synapses. Figure 5.5 shows that the optimal deletion ratio under the optimal minimal value deletion strategy is around 80 %, increasing memory capacity by 45% and tripling the information storage. However, assuming sub-optimal thresholds, or non-zero energy cost for neurons, the optimum drifts to lower values of $50 - 60\%$ which fits the experimental data.

Synaptic elimination is a broad phenomenon found throughout different brain structures and not restricted to associative memory areas. We believe however 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.

6.2 Cognitive implications of changes in synaptic density

In biological networks, synaptic growth and deletion occur in parallel with memory storage. As shown in Figure 5.6, the implementation of a minimal value pruning strategy during such a 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 5.6). 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 has a U-shaped form. It is interesting to compare these predictions to the experimental cognitive data. Figure 6.1 shows the performance of an adult subject when asked to recall a salient childhood event as a function of the subjects' age at the time of the event ([Sheingold and Tenney, 1982]). Indeed a U-shaped form can be observed at the ages of synaptic elimination but was not noticed before. However, no experimental studies have investigated this phenomenon in details, and it awaits further research.

The rising phase of synaptic density demonstrated in figure 5.6 yields another observation. While the "younger" network is able to recall memories from its remote "childhood", the "older" network fails to recall memories which were stored at an early stage (compare the solid and dashed lines in figure 5.6 at the age of three "years"). The explanation for this effect is as follows: at infancy the network is very sparse, and the stored memories are weakly stored. As the "infant" grows, more synapses are added and the newly stored memories are engraved on a more densely connected synaptic matrix, and hence are stored more strongly. As more and more memories are stored, the synaptic noise level rises, and a collapse of the weaker memories is inevitable, while later memories are retained. Even though synaptic density decreases during later development, a similar collapse will not occur for childhood memories, as synaptic elimination only slightly damages performance.

This scenario provides a new insight into the well studied childhood amnesia phenomenon, that is, the inability to recall events from early childhood described in section 1.3. Current neurological explanations suggest that maturation of memory related structures such as the Hippocampus are responsible for the amnesia [Nadel, 1986], but recent studies have questioned these views [Bachevalier *et al.*, 1993]. Our model is the first to realize an explanation for childhood amnesia which operates on the network level. The model also manifests the sharp transition in retrieval quality observed experimentally, emerging from the catastrophic breakdown phenomenon known to be a property of associative memory networks.

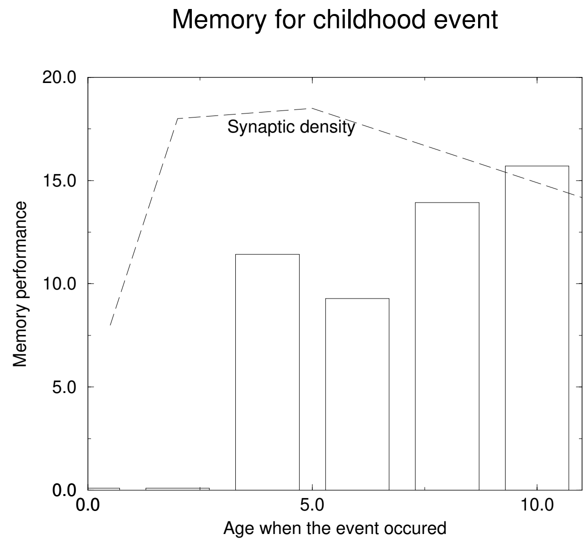

Figure 6.1: Memory recall of a salient childhood event as a function of the age when the event occurred as found by [Sheingold and Tenney, 1982]. 26 subjects with an average age of 20, were asked about a birth of a little brother, and their memories were cross checked against parental memories. Both the sharp transition from no-recall (ages 0-3) to pretty good recall, and the U-shaped curve at the synaptic elimination phase can be observed, matching the changes in synaptic density in human development (the dashed curve)[Huttenlocher, 1979].

Bibliography

[Amit, 1989] D.J. Amit. ***Modeling Brain Function***. Cambridge, 1989.

[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 ***et al.***, 89] J.P. Bourgeois, P.J. Jastreboff, and P. Rakic. Synaptogenesis in visual cortex of normal and preterm monkeys: Evidence for intrinsic regulation of synaptic overproduction. ***Proc. Natl. Acad. Sci. USA***, 86:4297–4301, 89.

[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.

[Chugani *et al.*, 1987] H.T. Chugani, M.E. Phelps, and Mazziotta J.C. Positron emission tomography study of human brain functional development. *Ann. Neurol.*, 22:487–497, 1987.

[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.

[Evans, 1989] M.R. Evans. Random dilutions in a neural network for biased patterns. *J. Phys. A. Math. Gen.*, 22:2103–2118, 1989.

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

[Gardner, 1987] E. Gardner. Maximum storage capacity in neural networks. *Europhys. Letters*, 4:481–485, 1987.

[Gardner, 1988] E. Gardner. The space of interactions in neural networks models. *J. Phys. A. Math. Gen.*, 21:257–270, 1988.

[Horn and Ruppin, 1995] D. Horn and E. Ruppin. Compensatory mechanisms in an attractor neural network model of schizophrenia. *Neural Computation*, 7:1494–1517, 1995.

[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.

[Kadekaro *et al.*, 1985] M. Kadekaro, A.M. Crain, and L. Sokoloff. Differential effects of electrical stimulation of sciatic nerve on metabolic activity in spinal cord and dorsal root ganglion in the rat. *Proc. Natl. Acad. Sci. USA*, 82:6010–6013, 1985.

[Katz and Miledi, 1967] B. Katz and R. Miledi. The study of synaptic transmition in the absence of nerve impulse. *J. Physiol.*, 192:407–436, 1967.

[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.

[McDonough and Mandler, 1994] L. McDonough and J.M. Mandler. Very long term recall in infants: Infantile amnesia. *Memory*, 2(4):339–352, 1994.

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

[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.

[Schweitzer and Green, 1982] L. Schweitzer and L. Green. Extended retention in preweanling rats. *J. Comp. Physiol. Psychol.*, 96(5):791–806, 1982.

[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.

[Tsodyks, 1988] M.V. Tsodyks. Associatice memory in asymetric diluted networks with low level of activity. ***Europhys. Lett.***, 7(3):203–208, 1988.

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

[Van-Hemmen, 1987] J.L Van-Hemmen. Nonlinear neural networks near saturation. ***Physical Review A.***, 36:1959–1969, 1987.

[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.

List of Figures