Dynamical properties of a randomly diluted neural network with variable activity

S Großkinsky‡ §
Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, NY 11794, USA

Abstract. The subject of study is a neural network with binary neurons, randomly diluted synapses and variable pattern activity. We look at the system with parallel updating using a probabilistic approach to solve the one step dynamics with one condensed pattern. We derive restrictions on the storage capacity and the mutual information content occurring during the retrieval process. Special focus is on the constraints on the threshold for optimal performance. We also look at the effect of noisy updating, giving a dynamical version of the critical temperature, the corresponding threshold and an approximation for the time evolution for small temperatures. The description is applicable to the whole retrieval process in the limit of strong dilution. The analysis is carried out as exactly as possible and over the whole parameter ranges, generalizing some former results.

PACS numbers: 64.60Cn,87.10+e

Submitted to: J. Phys. A: Math. Gen.

‡ Present address: Institut für Angewandte Mathematik, Julius-Maximilians-Universität Würzburg, Am Hubland, D-97074 Würzburg, Germany
§ E-mail: sngrossk@cip.physik.uni-wuerzburg.de
1. Introduction

Simple cognitive functions of the brain, like associative recall of memories, have been studied for 15 years by using models of attractor networks. One of these is the Hopfield-Little model [3], which was analyzed by means of statistical mechanics [4] because of the similarity to spin models. To get a more realistical description of properties in data processing, it was extended to models with variable pattern activity [5]. In this context a network with neurons $S_i \in \{0, 1\}$, instead of $\{-1, 1\}$ was studied in [3, 6] and showed enhanced storage capacity, resembling the upper bound obtained by Gardner [8]. To account for the low connectivity in the brain, this model was extended to random dilution of synapses and analyzed with a dynamical approach [9]. The one step parallel dynamics of a network can be solved exactly, using a probabilistic description [10] with a restriction to the first time step or very high dilution [11], due to feedback loops. More recently this method was used to characterize the influence of the threshold on the retrieval properties [12, 13].

Following this work, we characterize the state of the network by two overlaps with one condensed pattern. We use the probabilistic description of the time evolution of these overlaps to derive conditions for their improvement on the storage capacity. We also consider the mutual information content [12] and the restriction on the threshold in the $(Q, \alpha)$-plane for optimal performance. The effect of positive temperature is studied, leading to an exact expression for the critical value with the corresponding threshold, and an approximation for the time evolution for small temperatures. The analysis is done as exactly as possible, covering the whole range of the pattern activity and the overlaps from a dynamical point of view. In this way we can rederive some former results by looking at special limits, like small pattern activity and the network near retrieval. The main purpose is to demonstrate the possibilities of this relatively simple approach and therefore we study a simple network with realistic features. The most important considerations are confirmed by simulations.

In the next two subsections we will introduce the model and the theoretical description. In chapter 2 we will study restrictions on the pattern loading and the threshold for good retrieval properties at zero temperature. The effect of noisy updating is analyzed in chapter 3 and in section 4 we summarize our results.

1.1. The model

The network consists of $N$ neurons $S_i \in \{0, 1\}$, $i = 1, \ldots, N$ and the state is described by $S = (S_1, \ldots, S_N)$. There are $p$ patterns $\xi^\nu$, $\nu = 1, \ldots, p$, with elements $\xi^\nu_i \in \{0, 1\}$, $i = 1, \ldots, N$. They are stored by using a modified Hebb rule (see [1] and [3, 4]):

$$J_{ij} = \frac{c_{ij}(1-\delta_{ij})}{Nca(1-a)} \sum_{\nu=1}^{p} (\xi^\nu_i - a)(\xi^\nu_j - a)$$

The $\xi^\nu_i$ are independent, identically distributed random variables (IIDRV) with the distribution function $P(\xi^\nu_i) = a\delta(\xi^\nu_i - 1) + (1-a)\delta(\xi^\nu_i)$, $a \in [0, 1]$. Therefore the activity
Dynamical properties of a neural network

of every pattern

\[ a^\nu := \langle \xi^\nu_i \rangle_i = \frac{1}{N} \sum_{i=1}^{N} \xi^\nu_i \in [0, 1] \]

is a random variable with mean \( a \) and variance \( \frac{1}{N} a (1 - a) \) for all patterns \( \nu \in \{1, \ldots, p\} \).

The \( c_{ij} \in \{0, 1\} \) are also IIDRV with the distribution \( P(c_{ij}) = c \delta(c_{ij} - 1) + (1 - c) \delta(c_{ij}) \), \( c \in [0, 1] \). Hence the number of connections per neuron \( i \) over the number of neurons

\[ C^i/N := \langle c_{ij} \rangle_j = \frac{1}{N} \sum_{j=1}^{N} c_{ij} \in [0, 1] \]

is again a random variable with mean \( c \) and variance \( \frac{1}{N} c (1 - c) \) for all neurons \( i \in \{1, \ldots, N\} \).

The normalization factor \( (ca(1-a))^{-1} \) in the definition of the synapses turns out to be useful to keep the local field in the same order of magnitude over the whole range of pattern activity and network connectivity. As usual we have \( J_{ii} = 0 \) for all \( i \in \{1, \ldots, N\} \), enforced by the factor \( (1 - \delta_{ij}) \).

The neurons are updated parallel in discrete time steps \( t \in \mathbb{N} \) and a uniform threshold \( Q \), which may be time dependent, is subtracted from the local field

\[ h_i(t) := \sum_{j=1}^{N} J_{ij} S_j(t). \]

We have the usual activation function \( g(x, \beta) = (1 + e^{-2\beta x})^{-1} \), with the noise parameter \( \beta = T^{-1} \) and temperature \( T \), giving the update rule:

\[ P(S_i(t+1) = 1) = g(h_i(t) - Q, \beta) \]
\[ P(S_i(t+1) = 0) = 1 - g(h_i(t) - Q, \beta) \]

(2)

For the noiseless case the activation function reduces to the Heavyside step function, \( g(x, \beta) \rightarrow \Theta(x) \) for \( \beta \rightarrow \infty \), and the update rule is:

\[ S_i(t+1) = \begin{cases} 
1 & \text{if } h_i(t) - Q > 0 \\
0 & \text{if } h_i(t) - Q < 0 
\end{cases} \]

(3)

1.2. Theoretical description

We describe the network theoretically in the thermodynamic limit \( N, p \rightarrow \infty, \alpha = p/(cN) = \text{const.} \), denoted with \( \text{Lim} \), yielding the following simplifications:

\[ \text{Lim } a^\nu = a, \quad \nu = 1, \ldots, p \]
\[ \text{Lim } C^i/N = c, \quad i = 1, \ldots, N \]

We consider only the case of one condensed pattern \( \mu \in \{1, \ldots, p\} \) and characterize the state of the network by two corresponding overlaps (omitting the index \( \mu \)):

\[ m^\uparrow(t) := \frac{1}{a} \text{Lim } \langle \xi^\mu_i S_i(t) \rangle_i \in [0, 1] \]
\[ m^\downarrow(t) := \frac{1}{1 - a} \text{Lim } \langle (1 - \xi^\mu_i)(1 - S_i(t)) \rangle_i \in [0, 1] \]
With the two observables one can easily get the current network activity \( A(t) \):
\[
A(t) := \lim \langle S_i(t) \rangle_i = a \, m_1(t) + (1 - a)(1 - m_1(t)) \quad \in \quad [0, 1]
\]
The average overlaps with a noncondensed pattern \( \nu \neq \mu \) are \( m_1^\nu(t) = A(t) \) and \( m_1^\nu(t) = 1 - A(t) \). As pattern \( \mu \) is condensed, \( m_1 \) resp. \( m_1 \) have to be substantially bigger than these values and the case \( m_1 = A, m_1 = 1 - A \), which is equivalent with \( m_1 + m_1 = 1 \), corresponds to the failure of retrieval. If the system is in a state with \( m_1(t) + m_1(t) = 1 \) we have \( m_1(t + 1) + m_1(t + 1) = 1 \), which can easily be seen from the evolution equation \( \mathfrak{B} \) derived later. Hence an uncorrelated state will stay uncorrelated.

We will often consider the network in a state where \( A(t) = a \). Given \( m_1 \) this is achieved by setting:
\[
m_1(t) = m_{1A}(t) := 1 - \frac{a}{(1 - a)}(1 - m_1(t))
\]
To model the time evolution of the network we calculate the mean and the variance of the local field \( h_i(t) \) for \( \xi_i^\mu = 0 \) and \( \xi_i^\mu = 1 \), by splitting it in signal part \( h_i^S \) and noise part \( h_i^N \) as usual:
\[
h_i = \lim \langle \frac{c_{ij}(1 - \delta_{ij})}{N \rho(a)}(\xi_i^\mu - a)(\xi_j^\mu - a)S_j \rangle_j + \quad \rightarrow h_i^S
\]
\[
+ \lim \langle \langle \frac{c_{ij}(1 - \delta_{ij})}{N \rho(a)}(\xi_i^\nu - a)(\xi_j^\nu - a)S_j \rangle_j \rangle_{\nu \neq \mu} \quad \rightarrow h_i^N
\]
The indication of the time dependence is omitted. Now we average over the random distributions of the patterns \( \xi_i^\nu \) and the state of the network \( S \) in the thermodynamic limit, fixing the value of \( \xi_i^\mu \). Due to self-averaging the signal part has a concrete value, \( \mu_1 \) resp. \( \mu_1 \), but the noise part is a gaussian random variable with mean 0 and variance \( \sigma^2 \):
\[
\mu_1 = \lim \langle \langle h_i | \xi_i^\mu = 1 \rangle \rangle_{\xi_i^\nu \neq \mu, S} = (1 - a)(m_1 + m_1 - 1)
\]
\[
\mu_1 = \lim \langle \langle h_i | \xi_i^\mu = 0 \rangle \rangle_{\xi_i^\nu \neq \mu, S} = -a(m_1 + m_1 - 1)
\]
\[
\sigma^2 = \lim \langle \langle h_i^N \rangle \rangle_{\xi_i^\nu \neq \mu, S} = \alpha A
\]
The two random variables \( h_i |_{\xi_i^\nu = 0,1} \sim Q \) have mean \( \mu_1 - Q \) resp. \( \mu_1 - Q \), standard deviation \( \sigma \) and the distribution functions
\[
\rho_1(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left( \frac{(x - \mu_1 + Q)^2}{2\sigma^2} \right) \quad \rho_1(x) \quad \text{resp.}
\]
With the dynamical equations \( \mathfrak{B}, \mathfrak{B} \) we can compute \( m_1^\mu \) and \( m_1^\mu \) for the next time step by averaging the activation function over the distribution of the \( h_i - Q \).
\[
m_1(t + 1) = \langle g(h_i(t) - Q, \beta) \rangle_{\rho_1} = \int_{-\infty}^{\infty} g(x, \beta) \rho_1(x) \, dx
\]
\[
m_1(t + 1) = \langle 1 - g(h_i(t) - Q, \beta) \rangle_{\rho_1} = \int_{-\infty}^{\infty} (1 - g(x, \beta)) \rho_1(x) \, dx
\]
For the noiseless case this simplifies to:
\[
m_1(t + 1) = \int_0^\infty \rho_1(x) \, dx = \frac{1}{2} \left( 1 + \text{erf} \left( \frac{\mu_1 - Q}{\sqrt{2\sigma}} \right) \right)
\]
Dynamical properties of a neural network

\[ m_i(t+1) = \frac{1}{2} \left( 1 + \text{erf} \left( \frac{Q - \mu_i}{\sqrt{2\sigma}} \right) \right) \]  

(7)

In the calculations above we considered the neurons to be IID RVs with activity \( A \). This can be realized in the first time step but after one iteration the neurons may be correlated due to feedback loops. In the limit of strong dilution, so that the neurons do not have common ancestors in former time steps, the description is valid for the whole retrieval process. To achieve this the number of connections at each neuron \( C_i \) has to be of order \( \ln N \), which includes \( c \sim N^{-1} \ln N \to 0 \) for \( N \to \infty \). For further details see [10, 11] and references within there. There are \( 1/(aN) \) corrections to \( \sigma \) and \( \mu \) which become relevant in finite systems with small \( a \). Therefore we will only present simulations for relatively high \( a \), which is not a limitation of the theory.

2. Dynamical properties at zero temperature

In the following we look at the one step dynamics and get conditions on the pattern loading and the threshold for improvement of \( m_\uparrow \) and \( m_\downarrow \) at \( T = 0 \). We set \( \Delta m_\uparrow = m_\uparrow(t+1) - m_\uparrow(t) > 0 \) and \( \Delta m_\downarrow > 0 \) in equation (7) and get (inverf \( x \)) is the inverse errorfunction \( \text{erf}^{-1}(x) \):

\[ \frac{\mu_\uparrow - Q}{\sigma} > c_\uparrow := \sqrt{2} \text{inverf}(2m_\uparrow - 1) \]
\[ \frac{Q - \mu_\downarrow}{\sigma} > c_\downarrow := \sqrt{2} \text{inverf}(2m_\downarrow - 1) \]

(8)

2.1. Critical storage capacity and mutual information

By adding the two equations (8) we get the condition \( \mu_\uparrow - \mu_\downarrow > \sigma(c_\uparrow + c_\downarrow) \) for improvement, which has to be satisfied independent of the choice of \( Q \), assuming that it is chosen optimally. With the expressions for \( \mu_\uparrow, \mu_\downarrow \) and \( \sigma \) from equation (4) we can solve for the critical value \( \alpha_c \) where both improvements are zero:

\[ \alpha_c = \frac{(m_\uparrow + m_\downarrow - 1)^2}{(c_\uparrow + c_\downarrow)^2 A} \quad \text{for} \quad m_\uparrow + m_\downarrow \neq 1 \]
\[ = \frac{1}{2\pi m_\uparrow} e^{-c_\uparrow^2/2} \quad \text{for} \quad m_\uparrow + m_\downarrow = 1 \]

(9)

From the inequality we get the following conditions for improvement of \( m_\uparrow \) and \( m_\downarrow \):

\[ \alpha < \alpha_c \quad \text{if} \quad m_\uparrow + m_\downarrow > 1 \quad \alpha > \alpha_c \quad \text{if} \quad m_\uparrow + m_\downarrow < 1 \]

(10)

This is the case because we have \( c_\uparrow + c_\downarrow < 0 \) for \( m_\uparrow + m_\downarrow < 1 \) and in this region we can have improvement for arbitrary high pattern loadings. So if \( \alpha \) is too high the network always develops towards a state uncorrelated with the retrieval pattern, where we have \( m_\uparrow + m_\downarrow = 1 \).

In contrast to the usual storage capacity in equilibrium, \( \alpha_c \) is a dynamical variable,
Dynamical properties of a neural network

Depending on \( m_\uparrow \) and \( m_\downarrow \), as well as on \( a \) and \( c \).

- **Dependence on network connectivity**
  As we defined \( \alpha = p/(cN) \) we see that the maximal number of storable patterns decreases proportional to the number of connections per neuron, because \( \alpha_c \) is independent on \( c \). This is in accordance with former studies of random dilution in equilibrium (e.g. \([3]\)).

- **Dependence on pattern activity**
  For decreasing \( a \) the capacity increases monotonically if \( m_\uparrow + m_\downarrow > 1 \) and decreases if \( m_\uparrow + m_\downarrow < 1 \). If \( m_\uparrow \) and \( m_\downarrow \) are fixed we get the finite value \( \alpha|_{a=0} = \frac{(m_\uparrow + m_\downarrow - 1)^2}{(c_l + c_r)^2(1 - m_\downarrow)} \).
  When the activity of the network is equal to the one of the stored patterns we have \( A = a \), \( m_\downarrow = m_\downarrow A \) and therefore \( c_l|m_\downarrow A = \sqrt{2}inverf[1 - \frac{2m}{1-a}(1 - m_\downarrow)] \). By using the approximation \( inverf(x) \approx (-\ln(1 - x))^{1/2} \) for small \( x \) (see \([14]\)), we get the leading behaviour of \( \alpha_c \) for the limit \( a \to 0 \):
  \[
  \alpha_c \approx \frac{(m_\uparrow + m_\downarrow - 1)^2}{(c_l|m_\downarrow A)^2a} \approx \frac{m_\uparrow^2c}{-2a \ln a} \quad \text{with} \quad m_\uparrow \ll 1 - 2a \quad (11)
  \]
  This is known from former studies \([3, 6, 7, 9]\) as an approximation to the equilibrium storage capacity for small \( a \) and resembles the upper bound by Gardner \([8]\).

- **Dependence on state of the network**
  If one of the two parameters, \( m_\uparrow \) and \( m_\downarrow \), is equal to 1 or 0, \( \alpha_c \) is zero, because for \( m_\uparrow, m_\downarrow \to 1 \) we have \( c_\uparrow, c_\downarrow \to \infty \). This means we can have perfect retrieval only for a finite number of patterns.

For very small pattern activities the storage capacity may increase but the number of active neurons and the information represented by a single pattern decreases. Therefore the maximal amount of information storable in the network gives a more sensible characterization of performance. To account for the loss of information due to retrieval errors, i.e. \( m_\uparrow, m_\downarrow \neq 1 \), one uses the mutual information \([12]\). It can be defined as the negative logarithm of the conditional probability of choosing a network state with \( m_\uparrow \) and \( m_\downarrow \), given the activity \( A \):

\[
\text{Lim} \frac{I_m}{N} = -\text{Lim} \frac{1}{N} \ln \left[ \left( \frac{aN}{m_\uparrow aN} \right) \left( \frac{(1 - a)N}{(1 - m_\downarrow)(1 - a)N} \right) \right] = a(m_\uparrow \ln m_\uparrow + (1 - m_\uparrow) \ln(1 - m_\uparrow)) - (1 - A) \ln(1 - A) - A \ln A + (1 - a)(m_\downarrow \ln m_\downarrow + (1 - m_\downarrow) \ln(1 - m_\downdownarrow))
\]

This is the average amount of information per neuron of the retrieval pattern, that can be obtained from the network in a given state \( m_\uparrow, m_\downarrow \). For the maximal mutual information content \( i_m \) per synapse in units of bits we get:

\[
i_m = \text{Lim} I_m \frac{p_{\max}}{cN^2 \ln 2} = \frac{\alpha_c}{c \ln 2} \left\{ a \left[ m_\uparrow \ln \frac{m_\uparrow}{A} + (1 - m_\uparrow) \ln \frac{1 - m_\uparrow}{1 - A} \right] + (1 - a) \left[ m_\downarrow \ln \frac{m_\downarrow}{1 - A} + (1 - m_\downarrow) \ln \frac{1 - m_\downdownarrow}{A} \right] \right\} \quad (12)
\]
Dynamical properties of a neural network

This gives the maximal amount of information per synapse storable in the network, subject to the constraint of improvement of \( m_\uparrow \) and \( m_\downarrow \). It has more reasonable properties for low pattern activity than \( \alpha_c \):

- If the network is in a state uncorrelated with the retrieval pattern, i.e. \( m_\uparrow + m_\downarrow = 1 \), it is \( i_{m|m_\uparrow+m_\downarrow=1} = 0 \).
- We have \( i_{m|a=0,1} = 0 \) for any fixed \( m_\uparrow \) and \( m_\downarrow \) with \( A \neq a \), as there is certainly no information stored when all pattern elements have the same value.
- For \( m_\downarrow = m_\downarrow A \) or \( A = a \), using the approximation of \( \alpha_c \) for small \( a \), we get:

  \[
  i_m \approx \frac{m_\uparrow^2}{-2a \ln a \ln 2} (-am_\uparrow \ln a) \quad \rightarrow \quad \frac{m_\uparrow^3}{2 \ln 2} \quad \text{for} \quad a \rightarrow 0, \quad m_\uparrow \ll 1 - 2a \quad (13)
  \]

  That means if we keep the network activity fixed to \( a \) during the retrieval process we can obtain a nonzero amount of information from the network for \( a \rightarrow 0 \), although the information content of a single pattern vanishes. This result is in accordance with former studies in equilibrium (see e.g. [7, 8]).

The rederivation of \( \alpha_c \) and \( i_m \) in the limit \( a \rightarrow 0 \) shows that the enhanced properties for \( a \rightarrow 0 \) are only given during the retrieval process if we have \( A = a \). Further studies showed that this fixed relation between \( m_\uparrow \) and \( m_\downarrow \) is not optimal for maximizing the storage capacity. The optimal relation can be found, but the resulting capacity is of the same order, as the maximum according to Gardner is already reached.

2.2. Involving the threshold

Now we look at the two conditions (8) separately. After inserting \( \sigma = \sqrt{\alpha A} \) from (4) we can solve for \( \alpha_\uparrow \) resp. \( \alpha_\downarrow \) where \( \Delta m_\uparrow = 0 \) resp. \( \Delta m_\downarrow = 0 \):

\[
\alpha_\uparrow = \frac{(\mu_\uparrow - Q)^2}{c_\uparrow^2 A} \quad \alpha_\downarrow = \frac{(Q - \mu_\downarrow)^2}{c_\downarrow^2 A} \quad (14)
\]

Because the conditions were squared only one branch of the parabolas \( \alpha_\uparrow(Q) \) and \( \alpha_\downarrow(Q) \) imposes a condition on \( \alpha \), depending on the values of \( m_\uparrow \) and \( m_\downarrow \). For \( m_\uparrow > 0.5 \) we have \( c_\uparrow > 0 \) and the condition for improvement of \( m_\uparrow \) (8) gives an upper bound on \( \alpha \). For \( m_\uparrow < 0.5 \) the sign of \( c_\uparrow \) is changed and therefore we have a lower bound on \( \alpha \) for improvement of \( m_\uparrow \):

\[
m_\uparrow > 0.5 \quad \Rightarrow \quad \alpha < \alpha_\uparrow \quad \text{for} \quad Q < \mu_\uparrow \quad \alpha = 0 \quad \text{for} \quad Q \geq \mu_\uparrow \quad m_\downarrow < 0.5 \quad \Rightarrow \quad \alpha > \alpha_\downarrow \quad \text{for} \quad Q > \mu_\downarrow \quad \alpha \geq 0 \quad \text{for} \quad Q \leq \mu_\downarrow \quad (15)
\]

For \( m_\downarrow = 0.5 \) the parabola reduces to a vertical line at \( Q = \mu_\uparrow \) and we have improvement if \( Q < \mu_\uparrow \) for unlimited pattern loading and no improvement for \( Q > \mu_\uparrow \). The conditions for improvement of \( m_\downarrow \) can be obtain in the same way.

For good retrieval qualities we would like to have improvement of both overlaps, which is ensured in a region of the \((Q, \alpha)\)-plane limited by the valid branches of \( \alpha_\uparrow \) and \( \alpha_\downarrow \) (see figure [1]). Hence we have a lower and upper bound for the region of the threshold,
Dynamical properties of a neural network

$Q \in [\mu_c + c_\uparrow \sigma, \mu_\downarrow - c_\downarrow \sigma]$, that will lead to improvements on $m_\uparrow$ and $m_\downarrow$ depending on $\alpha$. Unless $m_\uparrow = m_\downarrow = 0.5$ there is always an intersection of the valid branches of $\alpha_\uparrow$ and $\alpha_\downarrow$, where both improvements are equal to zero. Here the interval for $Q$ reduces to a single point and we have $\alpha_\uparrow = \alpha_\downarrow = \alpha_c$. The value of $Q$ at this point is:

$$Q_c = \left( \frac{c_\downarrow}{c_\downarrow + c_\uparrow} - a \right) (m_\uparrow + m_\downarrow - 1) \quad \text{for} \quad m_\uparrow + m_\downarrow \neq 1$$

$$= \frac{c_\downarrow}{\sqrt{2\pi}} e^{-c^2/2} \quad \text{for} \quad m_\uparrow + m_\downarrow = 1 \quad (16)$$

The dependence on $a$ is linear and for $m_\uparrow = m_\downarrow = 1$ it reduces to $Q_c = 0.5 - a$ which is known to be the optimal threshold near retrieval [1, 7]. For $m_\uparrow + m_\downarrow > 1$ the region of simultaneous improvement is bounded below and above and the maximal possible storage is $\alpha_c$. For $m_\uparrow + m_\downarrow < 1$ the region is only bounded below and $\alpha_c$ is the minimal possible pattern loading for improvement, in accordance with (10). In figure 2 our considerations are confirmed by simulations for $a = 0.3$, $m_\uparrow = m_\downarrow = 0.9$ and $m_\downarrow = 0.3$, $m_\updownarrow = 0.9$.

To illustrate the use of the derived restrictions in the $(Q, \alpha)$-plane we discuss some choices of thresholds as functions of the state of the network which ensure $Q|_{\alpha = \alpha_c} = Q_c$, i.e. they allow for the critical storage capacity $\alpha_c$ (see figure 3).

- **Critical threshold**
  The easiest thing is of course to choose $Q = Q_c$ which certainly meets the requirement. But if $m_\uparrow$ resp. $m_\downarrow < 0.5$ it does not improve if $\alpha < \alpha_c$ (see $m_\uparrow$ in figure 2). This choice of $Q$ maximizes the term $\Delta c_\uparrow + \Delta c_\downarrow$. 

![Figure 1](image-url)
Dynamical properties of a neural network

Figure 2. Simulations for $\alpha_1(Q)$ and $\alpha_1(Q)$ with $a = 0.3, T = 0$ and $m_\uparrow = m_\downarrow = 0.9$ —— (Data filled), $m_\uparrow = 0.3, m_\downarrow = 0.9$ ---- (Data unfilled). Data obtained with $N = 600$ and average over 10 networks each with at least 150 stimuli. The average improvement of $m_\uparrow$ and $m_\downarrow$ was recorded as a function of $\alpha$ for different values of $Q$. At $\alpha = \alpha_1$ resp. $\alpha_1$, $\Delta m_\uparrow$ resp. $\Delta m_\downarrow$ vanishes.

- **Maximizing $\Delta m_\uparrow + \Delta m_\downarrow$**
  Just for comparison we will also look at $Q = Q_m := \frac{1}{2}(\mu_\uparrow + \mu_\downarrow)$ which maximizes $\Delta m_\uparrow + \Delta m_\downarrow$. This seems to be a somehow ‘natural’ choice, but it only fulfills $Q_m|_{\alpha=\alpha_c} = Q_c$ when $m_\uparrow = m_\downarrow$, where we have $Q_m = Q_c$.

- **Ensuring $A(t + 1) = a$**
  Another possibility is to choose $Q = Q_a$ in order to get the network activity $A(t + 1) = a$, ensuring an enhanced storage capacity for small $a$. In the retrieval state we also have $A = a$, so we can reach it with this threshold dependence. However the condition $Q_a|_{\alpha=\alpha_c} = Q_c$ is only obeyed for $A(t) = a$.

- **Preserving $m_\uparrow/m_\downarrow$**
  We can also choose $Q = Q_r$ to preserve the ratio $m_\uparrow(t + 1)/m_\downarrow(t + 1) = m_\uparrow(t)/m_\downarrow(t) = r$. We only look at this choice because it ensures $Q_r|_{\alpha=\alpha_c} = Q_c$, but if we start out with a ratio considerably different from 1 the attractor reached with this threshold won’t be very close to the retrieval state. So the only case where this is really interesting is for $r = 1$ where we have $Q_r = Q_c = Q_m$.

In figure 3 we illustrate the four choices in the $(Q, \alpha)$-plane with $a = 0.1$, $m_\uparrow = 0.6$ and $m_\downarrow = m_{\downarrow A} = 0.956$. Therefore we have $Q_a|_{\alpha=\alpha_c} = Q_c$ and the curve of $Q_m$ does not cross the critical point $(Q_c, \alpha_c)$. One can compare the different choices of thresholds by the improvements $\Delta m_\uparrow$ and $\Delta m_\downarrow$, which can easily be calculated. As $m_\uparrow = 0.6$ the upper limit for $\Delta m_\uparrow$ is 0.4 and for $\Delta m_\downarrow$ it is $1 - m_{\downarrow A} = 0.044$. For small $\alpha$ these limits are reached by $Q_c$ and $Q_a$. As expected $Q_a$ gives the highest improvements because it
has the best position between the two parabolas in figure 3. The improvement of \( m_\uparrow \) resp. \( m_\downarrow \) is a monotonically increasing function of the distance between \( Q \) and \( Q|_{\alpha=\alpha_1} \) resp. \( Q|_{\alpha=\alpha_1} \). \( Q_c \) and \( Q_r \) give the highest possible storage capacity \( \alpha_c \) in any case but the improvements \( \Delta m_\uparrow \) look rather poor, especially of \( Q_r \).

If we consider a case with \( m_\uparrow = m_\downarrow \) the picture is symmetric with respect to the axis \( Q = Q_c \) and therefore \( Q_m = Q_r = Q_c \) give the best possible improvements \( \Delta m_\uparrow = \Delta m_\downarrow \). Thus the optimal choice of the threshold during the retrieval process depends on the situation. If one has \( m_\uparrow(0) \approx m_\downarrow(0) \) and \( a \approx 0.5 \) the best choice for fast retrieval is \( Q_c \). But in the case of small pattern activity this limits the storage capacity and it is better to choose \( Q = Q_a \), independent of the initial condition.

3. Dynamics for positive temperature

With increasing temperature the critical storage capacity, as defined in 2.1, decreases and we take \( T_c \) as the value of \( T \) where \( \alpha_c = 0 \). The allowed region for the threshold reduces to a single point \( Q_c|_{T=T_c} \), which we will also calculate. After that we make an approximation for \( \alpha_\uparrow \) and \( \alpha_\downarrow \) for small temperatures.

3.1. Critical temperature

We look at the dynamical equation (3) in the limit \( \alpha \to 0 \), where the distribution functions \( \rho_\uparrow \) resp. \( \rho_\downarrow \) reduce to delta functions, because \( \sigma^2 \sim \alpha \):

\[
\rho_\uparrow(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu_\uparrow + Q)^2}{2\sigma^2}\right) \xrightarrow{\alpha \to 0} \delta(x - (\mu_\uparrow - Q))
\]

\( \rho_\downarrow(x) \) resp.
Dynamical properties of a neural network

Figure 4. Critical temperature \( T_c(m_\uparrow) \) for \( m_\uparrow = 0.5 \) - - - , \( m_\uparrow = 0.9 \) - - - , \( m_\uparrow = 0.999 \) — — — . Data with \( N = 2000 \), \( \alpha = 0.0005 \) and \( m_\uparrow = 0.5 \) (⋆, \( a = 0.3 \) and \( c = 0.5 \)), \( m_\uparrow = 0.9 \) (♦, \( a = 0.3 \) and \( c = 1 \)) and \( m_\uparrow = 0.999 \) (⊓ □, \( a = 0.5 \) and \( c = 1 \)).

In this limit we can evaluate the integrals in the evolution equation (6) and after solving for \( \beta \) and \( Q \) we get the critical temperature \( T_c = \beta_c^{-1} \) and the corresponding threshold \( Q_c|_{T=T_c} \).

For \( m_\uparrow + m_\downarrow \neq 1 \):

\[
T_c = \frac{-2(m_\uparrow + m_\downarrow - 1)}{\ln(\frac{1}{m_\downarrow} - 1) + \ln(\frac{1}{m_\uparrow} - 1)}
\]

\[
Q_c|_{T=T_c} = \left( \frac{\ln(\frac{1}{m_\downarrow} - 1)}{\ln(\frac{1}{m_\downarrow} - 1) + \ln(\frac{1}{m_\uparrow} - 1)} - a \right) (m_\uparrow + m_\downarrow - 1)
\]

For \( m_\uparrow + m_\downarrow = 1 \):

\[
T_c = 2m_\uparrow(1 - m_\uparrow) \quad Q_c|_{T=T_c} = m_\uparrow(1 - m_\uparrow) \ln\left( \frac{1}{m_\uparrow} - 1 \right)
\]

The critical temperature is independent of the pattern activity and the network connectivity. It is easy to see that the maximum value is \( T_c = 0.5 \) at \( m_\uparrow = m_\downarrow = 0.5 \), and if at least one of the two observables is equal to zero or one we have \( T_c = 0 \). We see \( T_c(m_\downarrow) \) in figure 4 for several values of \( m_\uparrow \), confirmed by simulations which were done at different values of \( a \) and \( c \) to demonstrate that it is independent on these parameters.

The critical threshold at \( T = T_c \) has the same linear dependence on \( a \) as for zero temperature, except for an additive constant depending on \( m_\uparrow \) and \( m_\downarrow \). Only for \( m_\uparrow, m_\downarrow \approx 1 \) with \( m_\uparrow \neq m_\downarrow \) there is a considerable difference between the two, in the case \( m_\uparrow = m_\downarrow \) both are equal for every \( a \). This indicates that for increasing temperature the critical point in the \((Q, \alpha)\)-plane is mainly shifted to lower \( \alpha \) with fixed \( Q \), which is true to high accuracy as we will see next.
3.2. Expansion for small temperatures

For $T > 0$ we solve the conditions $m_\dagger(t + 1) = m_\dagger(t)$ and $m_\dagger(t + 1) = m_\dagger(t)$ for $\alpha_\dagger$ and $\alpha_\dagger$ by using an approximation for low temperatures, as $T_c \in [0, 0.5]$. In the following we look at the first condition in order to get $\alpha_\dagger$, the same can be done for $\alpha_\dagger$. After some algebra given in the appendix we have the following approximation which is accurate for $T \ll 2\sigma^2/(\mu_\dagger - Q)$:

$$m_\dagger(t + 1) \approx m_\dagger(t + 1)|_{T=0} - T^2 \frac{\mu_\dagger - Q}{2\sqrt{2\pi}\sigma^3} e^{-\left(\frac{\mu_\dagger - Q}{\sqrt{2}\sigma}\right)^2} \frac{\pi^2}{12}$$

(19)

By looking at the numerical solutions of $\alpha_\dagger(Q)$ we see that the shape of the parabola is more or less unaltered, it is just shifted down. Therefore we make the ansatz $\alpha_\dagger = \alpha_\dagger|_{T=0} - f(T)$ where the function $f(T)$ accounts for the downshift of $\alpha_\dagger$ as an additional source of noise, like in [1]. Now we replace $\alpha_\dagger|_{T=0}$ by $\alpha_\dagger + f(T)$ in the dynamical equation for zero temperature (10) and expand at $f(T) = 0$ to first order:

$$m_\dagger(t + 1) \approx m_\dagger(t + 1)|_{T=0} - \frac{\mu_\dagger - Q}{2\sqrt{2\pi}\sigma^3} e^{-\left(\frac{\mu_\dagger - Q}{\sqrt{2}\sigma}\right)^2} Af(T)$$

This term looks very similar to the expansion of the dynamical equation and by comparing the two we get the following approximation:

$$\alpha_{\dagger 1} \approx \alpha_\dagger|_{T=0} - \gamma_1 T^2$$

(20)

$$\alpha_{\dagger 1} \approx \alpha_\dagger|_{T=0} - \gamma_1 T^2$$

where $\gamma_1 = \frac{\pi^2}{12A}$

The calculation for $\alpha_{\dagger 1}$ gives the same correction to this order. We labeled the approximation with 1 because we can also think of another one. By assuming a quadratic $T$-dependence of $f$ we can make directly the ansatz $\alpha_\dagger = \alpha_\dagger|_{T=0} - \gamma_2 T^2$. Knowing the critical temperature we get the condition $\alpha_c|_{T=T_c} = \alpha_c|_{T=0} - \gamma_2 T_c^2 = 0$, yielding another approximation:

$$\alpha_{\dagger 2} \approx \alpha_\dagger|_{T=0} - \gamma_2 T^2$$

$$\alpha_{\dagger 2} \approx \alpha_\dagger|_{T=0} - \gamma_2 T^2$$

where $\gamma_2 = \frac{1}{4A} \left[ \ln \left( \frac{1}{m_\dagger} - 1 \right) + \ln \left( \frac{1}{m_\dagger} - 1 \right) \right]^2 \left( c_\dagger + c_\dagger \right)^2$.

(21)

In figure 1 both approximations are compared with the numerical solution for $T = 0.2$, 0.3 and the case $T = 0$ for $m_\dagger = 0.6$ and $a = A = 0.3$. Therefore we have $\gamma_1 = 0.822 \frac{1}{A} > \gamma_2 = 0.679 \frac{1}{A}$ and the first approximation (20) gives the lower curves, the second (21) the upper ones. As expected (21) is better for small $\alpha$, as $T_c$ was evaluated at $\alpha = 0$. For $\alpha_{\dagger 11}$ and $\alpha_{\dagger 11}$ to be accurate we have the condition $\alpha \gg \max(\mu_\dagger - Q, Q - \mu_\dagger) \frac{1}{2A} T \approx 0.1$ from above, making it better for higher storage levels.

We also see that with increasing temperature the ansatz of a simple downshift of the parabolas $\alpha_\dagger(Q)$ and $\alpha_{\dagger}(Q)$ becomes more and more inaccurate. But even for $T = 0.3$, which is relatively high, the approximations are pretty good. Therefore $\alpha_c$ decreases proportional to $T^2$ for small temperatures, which is in accordance with former studies in equilibrium (see e.g. [1]). From the $1/A$-dependence of $\gamma_1$ and $\gamma_2$ we know that the effect of positive temperature is much more drastic for small network activities than for large ones.
Dynamical properties of a neural network

Figure 5. \( \alpha_\uparrow(Q) \) and \( \alpha_\downarrow(Q) \) for \( m_\uparrow = 0.6, A = a = 0.3 \) and \( T = 0 \), \( T = 0.2 \) —— (Data filled), \( T = 0.3 \) — — — (Data unfilled). Approximations —— (grey lines): \( \alpha_{1\uparrow} \) and \( \alpha_{1\downarrow} \) lower curves, \( \alpha_{1\uparrow} \) and \( \alpha_{1\downarrow} \) upper curves.

4. Conclusion

In this work we described a randomly diluted neural network model with variable pattern activity using a probabilistic approach to solve the one step dynamics with one condensed pattern. By carrying out the analysis as exactly as possible we were able to confirm with this relatively simple approach many previous results on this model, derived with different techniques and often restricted to special cases. Most of our results are valid for the whole range of the different parameters, except for the dilution level, generalizing some former studies.

We got new insight in the dynamical properties of the network, studying the critical storage capacity, information content and critical temperature for arbitrary network states. Special focus was on the resulting constraints on the threshold to realize the critical values, a feature that was often overlooked in former studies. We used this to analyze the effects of choices of threshold functions during retrieval. We also showed that we have to impose conditions during the retrieval process to get the enhanced properties derived in former studies.

To demonstrate the possibilities of the probabilistic approach we chose a neural network which is relatively simple, but has many realistic features. The analysis is not restricted to this model, it can be used for all networks with parallel updating. Further important is the site-independent description of the local field, in our case represented by \( \mu_\uparrow, \mu_\downarrow \) and \( \sigma \), excluding for example site-dependent thresholds. But the method is easily extendable to other models with graded response neurons (see \[\square\]), groups of patterns with different activities, a finite number of condensed patterns or sequential patterns
Dynamical properties of a neural network

It may be interesting to extend the presented analysis to these cases, making it possible to describe various network properties exactly with relatively easy computations. Especially the restrictions we derived on the parameters during retrieval may be very helpful for simulations.

Acknowledgments

I am thankful to professor Robert Shorck for advising my work on this subject and to Miguel Maravall for very useful discussions. This work was supported in part by the german academic exchange program (DAAD) and the Swartz Foundation at Stony Brook.

Appendix. Calculation for $T > 0$

In section 3.2 we want to calculate $\alpha_\uparrow$ and $\alpha_\downarrow$ from the dynamical equations for $T > 0$. In the following we perform the calculation for $\alpha_\uparrow$, where $\alpha_\downarrow$ can be evaluated in exactly the same way. First we split the integral in the dynamical equation (6):

$$m_\uparrow(t + 1) = \int_{-\infty}^{0} \frac{\exp(2\beta x)}{1 + \exp(2\beta x)} \rho_\uparrow(x) \, dx + \int_{0}^{\infty} \frac{1}{1 + \exp(-2\beta x)} \rho_\uparrow(x) \, dx$$

By using the geometric series $1/(1 + y) = \sum_{n=0}^{\infty} (-1)^n y^n$ for $y < 1$ and replacing $x$ by $-x$ in the first integral we get:

$$m_\uparrow(t + 1) = \sum_{n=0}^{\infty} (-1)^n \int_{0}^{\infty} \left[ e^{-2\beta x n} \rho_\uparrow(x) - e^{-2\beta x (n+1)} \rho_\uparrow(-x) \right] \, dx$$

Inserting $\rho_\uparrow(x)$ from equation (5), we have to calculate two gaussian integrals of the form

$$\int_{0}^{\infty} e^{-q_1 x - q_2 x^2} \, dx = \frac{\sqrt{\pi}}{2\sqrt{q_2}} \exp \left( \frac{q_1^2}{4q_2} \right) \left[ 1 - \text{erf} \left( \frac{q_1}{2\sqrt{q_2}} \right) \right].$$

After this we see that $m_\uparrow(t + 1)|_{T=0}$ is the term in the sum for $n = 0$ and we get the expression

$$m_\uparrow(t + 1) = \frac{1}{2} \left( 1 + \text{erf} \left( \frac{\mu_\uparrow - Q}{\sqrt{2}\sigma} \right) \right) - \Delta = m_\uparrow(t + 1)|_{T=0} - \Delta$$

with the correction term

$$\Delta = \frac{1}{2} \sum_{n=1}^{\infty} (-1)^n e^{2\beta^2 \sigma^2 n^2} \left[ e^{2\beta(\mu_\uparrow - Q)n} \left( 1 - \text{erf} \left( \sqrt{2}\sigma \beta n + \frac{\mu_\uparrow - Q}{\sqrt{2}\sigma} \right) \right) - e^{-2\beta(\mu_\uparrow - Q)n} \left( 1 - \text{erf} \left( \sqrt{2}\sigma \beta n - \frac{\mu_\uparrow - Q}{\sqrt{2}\sigma} \right) \right) \right].$$
This expression is still exact but in order to do the sum we use the approximation 
\[ 1 - \text{erf}(x) = \exp(-x^2)/(\sqrt{\pi}x) \] [14]. This is accurate for \( x \gg 1 \) which means in our case 
\( \beta \gg (\mu^\uparrow - Q)/(2\sigma^2) \). The expression for \( \Delta \) simplifies to

\[
\Delta = \frac{\mu^\uparrow - Q}{2\sqrt{2\pi}\beta^2\sigma^3} e^{-\left(\frac{\mu^\uparrow - Q}{\sqrt{2\sigma^2}}\right)^2} \sum_{n=1}^{\infty} (-1)^n \left( n^2 - \left( \frac{\mu^\uparrow - Q}{2\beta\sigma^2} \right)^2 \right)^{-1}.
\]

The sum can be done exactly and we end up with the following, writing \( T = \beta^{-1} \):

\[
m^\uparrow(t + 1) = m^\uparrow(t + 1)|_{T=0} - T^2 \frac{\mu^\uparrow - Q}{2\sqrt{2\pi}\sigma^3} e^{-\left(\frac{\mu^\uparrow - Q}{\sqrt{2\sigma^2}}\right)^2} \Gamma(T\pi \frac{\mu^\uparrow - Q}{2\sigma^2}),
\]

\[
\Gamma(x) = \frac{x - \sin x}{2x^2 \sin x} = \frac{\pi^2}{12} + \frac{7\pi^4}{720} x^2 + \ldots
\]

For \( T \ll 2\sigma^2/(\mu^\uparrow - Q) \) the argument of \( \Gamma \) is small and we use the zero order approximation \( \Gamma \approx \frac{\pi^2}{12} \).

References

[1] Amit D J 1989 Modelling Brain Function (New York: Cambridge University Press)
[2] Hertz J, Krogh A and Palmer R G 1991 Introduction to the theory of neural computation (Reedwood City, CA: Adison Wesley)
[3] Hopfield J J 1982 Proc. Natl. Acad. Sci. USA 79 2554 - 2558
[4] Little W A 1974 Math. Biosci. 19 101 - 120
[5] Amit D J, Gutfreund H and Sompolinsky H 1985 Phys. Rev. A 32(2) 1007 - 1018
[6] Tsodyks M V, Feigel’man M V 1988 Europhys. Lett. 6(2) 101 - 105
[7] Buhmann J, Divko R and Schulten K 1989 Phys. Rev. A 39(5) 2689 - 2692
[8] Gardner E 1988 J. Phys. A: Math. Gen. 21 257 - 270
[9] Tsodyks M V 1988 Europhys. Lett. 7(3) 203 - 208
[10] Bollé D, Vinck B and Zagrebnov V A 1993 J. Stat. Phys. 70 1099 - 1119
[11] Derrida B, Gardner E and Zippelius A 1987 Europhys. Lett. 4(2) 167 - 173
[12] Dominguez D R C and Bollé D 1998 Phys. Rev. Lett. 80 2961
[13] Kitano K and Aoyagi T 1998 J. Phys. A: Math. Gen. 31 613
[14] Philip J R 1960 Australian Journal of Physics 13 13 - 20