

# Optimal signalling in Attractor Neural Networks

Isaac Meilijson \* and Eytan Ruppin  
School of Mathematical Sciences  
Raymond and Beverly Sackler Faculty of Exact Sciences  
Tel-Aviv University, 69978 Tel-Aviv, Israel.

February 3, 1994

## Abstract

In [Meilijson and Ruppin, 1993a] we presented a methodological framework describing the two-iteration performance of Hopfield-like attractor neural networks with history-dependent, Bayesian dynamics. We now extend this analysis in a number of directions: input patterns applied to small subsets of neurons, general connectivity architectures and more efficient use of history. We show that the optimal signal (activation) function has a *slanted sigmoidal* shape, and provide an intuitive account of activation functions with a non-monotone shape. This function endows the analytical model with some properties characteristic of cortical neurons' firing.

---

\*Royal Society Scholar, Summer 1992. The generosity of the Royal Society and hospitality of the Department of Statistics at Oxford University is warmly acknowledged.

# 1 Introduction

It is well known that a given cortical neuron can respond with a different firing pattern for the same synaptic input, depending on its firing history and on the effects of modulator transmitters (see [Connors and Gutnick, 1990, Schwindt, 1992] for a review). The time span of different channel conductances is very broad, and the influence of some ionic currents varies with the history of the membrane potential [Lytton, 1991].

Working within the convenient framework of Hopfield-like attractor neural networks (ANN) [Hopfield, 1982, Hopfield, 1984], but motivated by the history-dependent nature of neuronal firing, we continue our previous investigation [Meilijson and Ruppin, 1993a] (henceforth, M & R) describing the short-term, two-iteration performance of feedback neural networks with history-dependent dynamics. Building upon the findings presented in M & R, we now study a more general framework:

- We differentiate between ‘input’ neurons receiving the initial input signal with high fidelity and ‘background’ neurons that receive it with low fidelity.
- Dynamics now depend on the neuron’s history of firing, in addition to its history of input fields.
- The dependence of ANN performance on the network architecture can be explicitly expressed. In particular, this enables the investigation of cortical-like architectures, where neurons are randomly connected to other neurons, with higher probability of connections formed between spatially proximal neurons [Braitenberg and Schuz, 1991, Abeles, 1991].
- While in M & R we have concentrated on discrete, dichotomous signal functions, in this work we extend our analysis to investigate analog, continuous signal functions.

Concentrating on the study of continuous input/output signal functions which govern the firing rate of the neuron (such as the conventional sigmoidal function [Wilson and Cowan, 1972, Pearson *et al.*, 1987]), the notion of a synchronous instantaneous ‘iteration’ is now viewed as an abstraction of the overall dynamics for some short length of time, during which the firing rate does not change significantly. We analyze the performance of the network after two such iterations, or intermediate times spans, a period sufficiently long for some significant neural information to be fed back within the network, but shorter than

those the network may require for falling into an attractor. However, as demonstrated in section 6, the performance of history-dependent ANNs after two iterations is sufficiently high compared with that of memoryless (history-independent) models, that the analysis of two iterations becomes a viable end in its own right (See also Appendix B for a discussion of attractors and the relation between interim and attractor performance). It should also be noted that even memoryless ANNs are known to converge in practically two or three iterations, when a memory pattern is successfully retrieved [Komlós and Paturi, 1988].

Examining this general family of signal functions, we now search for the computationally most efficient history-dependent neuronal signal (firing) function, and study its performance. We derive the optimal analog signal function, having the *slanted sigmoidal* form illustrated in figure 1a, and show that it significantly improves performance, both in relation to memoryless dynamics and versus the performance obtained with the previous dichotomous signalling. The optimal signal function is obtained by subtracting from the conventional sigmoid signal function some multiple of the current input field. As shown in figure 1a (or in figure 1b, plotting the discretized version of the optimal signal function) the neuron’s signal may have a sign opposite to the one it believes in. [Yoshizawa *et al.*, 1993, Morita, 1993] and [Felice *et al.*, 1993] have also observed that the capacity of ANNs is significantly improved by using nonmonotone analog signal functions. They studied the limit (after infinitely many iterations) under dynamics using a nonmonotone function of the current input field, similar in form to the slanted sigmoid. The Bayesian framework we work in provides, for the first time, a clear intuitive account of the non-monotone form and the seemingly bizarre sign reversal behavior. As we shall see, the slanted sigmoidal form of the optimal signal function is mainly a result of collective cooperation between neurons, whose ‘common goal’ is to maximize the network’s performance. It is rather striking that the resulting slanted sigmoid endows the analytical model with some properties characteristic of cortical neurons’ firing; this ‘collectively optimal’ function may be hard-wired into the cellular biophysical mechanisms determining each neuron’s firing function.

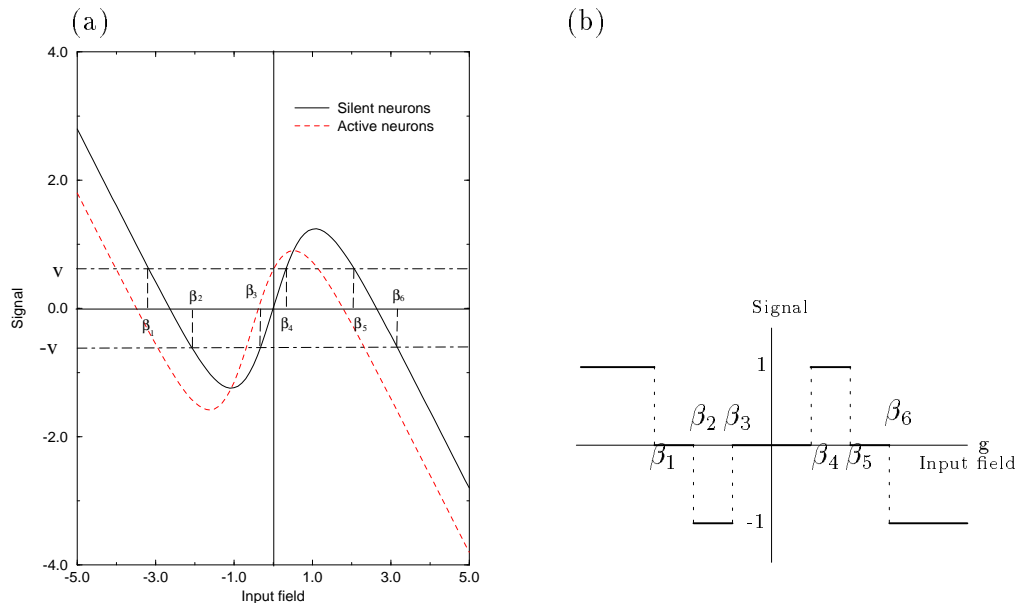


Figure 1: (a) A typical plot of the slanted sigmoid, Network parameters are  $N = 5000$ ,  $K = 3000$ ,  $n_1 = 200$  and  $m = 50$ . (b) A sketch of its discretized version.

This paper is organized as follows. The next section presents the model we work in. Section 3 gives analytical expressions for the performance of the network, and section 4 presents the optimal analog signal function and its discretized version. The derivation of the optimal signal function is described in Appendix A. The performance achieved with optimal signalling in different networks is illustrated in section 5. The biological relevance of our results is discussed in section 6. Finally, Appendix B uses Martingale arguments to infer that the network always converges to a well defined ‘final state’ and that the similarity of the network increases with every Bayesian iteration. This makes our two-iteration results a lower bound on the performance of history-dependent ANNs.

## 2 The model

Our framework is an ANN storing  $m + 1$  memory patterns  $\xi^1, \xi^2, \dots, \xi^{m+1}$ , each an  $N$ -dimensional vector. The network is composed of  $N$  neurons, each of which is randomly connected to  $K$  other neurons. The  $(m + 1)N$  memory entries are independent with equally likely  $\pm 1$  values. The initial pattern  $X$ , synchronously signalled by  $L (\leq N)$  *initially active* neurons, is a vector of  $\pm 1$ 's, randomly generated from one of the memory patterns (say  $\xi = \xi^{m+1}$ ) such that  $P(X_i = \xi_i) = \frac{1+\epsilon}{2}$  for each of the  $L$  initially active neurons and

$P(X_i = \xi_i) = \frac{1+\delta}{2}$  for each *initially quiescent* (non-active) neuron. Although  $\epsilon, \delta \in [0, 1)$  are arbitrary, it is useful to think of  $\epsilon$  as being 0.5 (corresponding to an initial similarity of 75%) and of  $\delta$  as being zero - a quiescent neuron has no prior preference for any given sign. Let  $\alpha_1 = m/n_1$  denote the initial memory load, where  $n_1 = LK/N$  is the average number of signals received by each neuron.

We follow a Bayesian approach under which the neuron's signalling and activation decisions are based on the a-posteriori probabilities assigned to its two possible true memory states,  $\pm 1$ . We distinguish between input fields that model incoming spikes, and generalized fields that model history-dependent, adaptive post-synaptic potentials. Clearly, the *prior probability* that neuron  $i$  has memory state  $+1$  is

$$\lambda_i^{(0)} = P(\xi_i = 1 | X_i, I_i) = \begin{cases} \frac{1+\epsilon}{2} & \text{if } X_i = 1, I_i = 1 \\ \frac{1-\epsilon}{2} & \text{if } X_i = -1, I_i = 1 \\ \frac{1+\delta}{2} & \text{if } X_i = 1, I_i = 0 \\ \frac{1-\delta}{2} & \text{if } X_i = -1, I_i = 0 \end{cases} \quad (1)$$

$$= \frac{1 + (\epsilon I_i + \delta(1 - I_i))X_i}{2} = \frac{1}{1 + e^{-2g_i^{(0)}}}$$

where  $I_i = 0, 1$  indicates whether neuron  $i$  has been active (i.e., transmitted a signal) in the first iteration, and the *generalized field*  $g_i^{(0)}$  is given by

$$g_i^{(0)} = \begin{cases} g(\epsilon)X_i & \text{if } i \text{ is active} \\ g(\delta)X_i & \text{if } i \text{ is quiescent} . \end{cases} \quad (2)$$

where

$$g(t) = \operatorname{arctanh}(t) = \frac{1}{2} \log \frac{1+t}{1-t} \quad ; \quad 0 \leq t < 1 . \quad (3)$$

We also define the *prior belief* that neuron  $i$  has memory state  $+1$

$$O_i^{(0)} = \lambda_i^{(0)} - (1 - \lambda_i^{(0)}) = 2\lambda_i^{(0)} - 1 = \operatorname{tanh}(g_i^{(0)}) \quad (4)$$

whose possible values are  $\pm\epsilon$  and  $\pm\delta$  (The belief is simply a rescaling of the probability from the  $[0, 1]$  interval to  $[-1, +1]$ ).

The input field observed by neuron  $i$  as a result of the initial activity is

$$f_i^{(1)} = \frac{1}{n_1} \sum_{j=1}^N W_{ij} I_{ij} I_j X_j \quad (5)$$

where  $I_{ij} = 0, 1$  indicates whether a connection exists from neuron  $j$  to neuron  $i$  and  $W_{ij}$  denotes its magnitude, given by the Hopfield prescription

$$W_{ij} = \sum_{\mu=1}^{m+1} \xi_i^\mu \xi_j^\mu \quad , \quad W_{ii} = 0. \quad (6)$$

As a result of observing the input field  $f_i^{(1)}$ , which is approximately normally distributed (given  $\xi_i, X_i$  and  $I_i$ ) with mean and variance

$$E(f_i^{(1)}|\xi_i, X_i, I_i) = \epsilon\xi_i \quad (7)$$

$$Var(f_i^{(1)}|\xi_i, X_i, I_i) = \alpha_1 , \quad (8)$$

neuron  $i$  changes its opinion about  $\{\xi_i = 1\}$  from  $\lambda_i^{(0)}$  to the *posterior probability*

$$\lambda_i^{(1)} = P(\xi_i = 1|X_i, I_i, f_i^{(1)}) = \frac{1}{1 + e^{-2g_i^{(1)}}} , \quad (9)$$

with a corresponding *posterior belief*  $O_i^{(1)} = \tanh(g_i^{(1)})$ , where  $g_i^{(1)}$  is conveniently expressed as an additive generalized field (see Lemma 1(II) in M & R)

$$g_i^{(1)} = g_i^{(0)} + \frac{\epsilon}{\alpha_1} f_i^{(1)} . \quad (10)$$

We now get to the second iteration, in which, as in the first iteration, some of the neurons become active and signal to the network. Unlike the first iteration, in which initially active neurons had independent beliefs of equal strength and simply signalled their states in the initial pattern, the preamble to the second iteration finds neuron  $i$  in possession of a personal history  $(X_i, I_i, f_i^{(1)})$ , as a function of which the neuron has to determine the signal to transmit to the network. While the history-independent Hopfield dynamics choose  $\text{sign}(f_i^{(1)})$  as this signal, we model the signal function as  $h(g_i^{(1)}, X_i, I_i)$ . This seems like four different functions of  $g_i^{(1)}$ . However, by symmetry,  $h(g_i^{(1)}, +1, I_i)$  should be equal to  $-h(-g_i^{(1)}, -1, I_i)$ . Hence, we only have two functions of  $g_i^{(1)}$  to define,  $h_1(\cdot)$  for the signals of the initially active neurons and  $h_0(\cdot)$  for the quiescent ones. For mathematical convenience we would like to insert into these functions random variables with unit variance. By (8) and (10), the conditional variance  $Var(g_i^{(1)}|\xi_i, X_i, I_i)$  is  $(\epsilon/\alpha_1)^2\alpha_1 = (\epsilon/\sqrt{\alpha_1})^2$ . We thus define  $\omega = \epsilon/\sqrt{\alpha_1}$  and let

$$h(g_i^{(1)}, X_i, I_i) = X_i h_{I_i}(X_i g_i^{(1)}/\omega) . \quad (11)$$

The field observed by neuron  $i$  following the second iteration (with  $n_2$  updating neurons per neuron) is

$$f_i^{(2)} = \frac{1}{n_2} \sum_{j=1}^N W_{ij} I_{ij} h(g_j^{(1)}, X_j, I_j) , \quad (12)$$

on the basis of which neuron  $i$  computes its posterior probability

$$\lambda_i^{(2)} = P(\xi_i = 1|X_i, I_i, f_i^{(1)}, f_i^{(2)}) \quad (13)$$

and corresponding posterior belief  $O_i^{(2)} = 2\lambda_i^{(2)} - 1$ , which will be expressed in section 4.3 as  $\tanh(g_i^{(2)})$ .

In this paper we stop at the above two information-exchange iterations and let each neuron express its final choice of sign as

$$X_i^{(2)} = \text{sign}(O_i^{(2)}) . \quad (14)$$

The performance of the network is measured by the final similarity

$$S_f = P(X_i^{(2)}) = \frac{1 + \frac{1}{N} \sum_{j=1}^N X_j^{(2)} \xi_j}{2} \quad (15)$$

(where the last equality holds asymptotically).

Our first task is to present (as simple as possible) an expression for the performance under arbitrary architecture and activity parameters, for general signal functions  $h_0$  and  $h_1$ . Then, using this expression, our main goal is to find the best choice of signal functions which maximize the performance attained. We find these functions when there are either no restrictions on their range set or they are restricted to the values  $\{-1, 0, 1\}$ , and calculate the performance achieved in Gaussian, random and multi-layer patterns of connectivity. The optimal choice will be shown to be the *slanted sigmoid*

$$h(g_i^{(1)}, X_i, I_i) = O_i^{(1)} - c f_i^{(1)} \quad (16)$$

for some  $c$  in  $(0, 1)$ . We present explicitly all formulas, providing their derivation in Appendix A.

### 3 The rationale for nonmonotone signalling

The common Hopfield convention is to have neuron  $i$  signal  $\text{sign}(f_i^{(1)})$ . Another possibility, studied in M & R, is to signal the preferred sign only if this preference is strong enough, otherwise remain silent. However, an even better performance was seen to be achieved by counterintuitive signals which are not monotone in  $g_i^{(1)}$  [Yoshizawa *et al.*, 1993, Felice *et al.*, 1993, Meilijson and Ruppin, 1993a]. In fact, precisely those neurons that are *most* convinced of their signs should signal the sign *opposite* to the one they so strongly believe in! We would like to offer now an intuitive explanation for this seeming pathology, and proceed later to the mathematics leading to it.

In the initial pattern, the different entries  $X_i$  and  $X_j$  are conditionally independent given  $\xi_i$  and  $\xi_j$ . This is not the case for the input fields  $f_i^{(1)}$  and  $f_j^{(1)}$ , whose correlation

is proportional to the synaptic weight  $W_{ij}$  (M & R). For concreteness, let  $\epsilon = 0.5$  and  $\alpha_1 = 0.25$  and suppose that neuron  $i$  has observed an input field  $f_i^{(1)} = 3$ . Neuron  $i$  now knows that either its true memory state is  $\xi_i = +1$  in which case the ‘noise’ in the input field is  $3 - \epsilon = 2.5$  (i.e., five standard deviations above the mean) or its true memory state is  $\xi_i = -1$  and the noise is  $3 + \epsilon = 3.5$  (or seven standard deviations above the mean). In a Gaussian distribution, deviations of five or seven standard deviations are very unusual, but seven is so much more unusual than five, that neuron  $i$  is practically convinced that its true state is  $+1$ . However, neuron  $i$  knows that its input field  $f_i^{(1)}$  is grossly inflicted with noise and since the input field  $f_j^{(1)}$  of neuron  $j$  is correlated with its own, neuron  $i$  would want to warn neuron  $j$  that its input field has unusual noise too and should not be believed on face value. Neuron  $i$ , a good student of Regression Analysis, wants to tell neuron  $j$ , without knowing the weight  $W_{ij}$ , to subtract from its field a multiple of  $W_{ij}f_i^{(1)}$ . This is accomplished, to the simultaneous benefit of all neurons  $j$ , by signalling a multiple of  $-f_i^{(1)}$ . We see that neuron  $i$ , out of ‘purely altruistic traits’, has a conflict between the positive act of signalling its assessed true sign and the negative act of signalling the opposite as a means of correcting the fields of its peers. It is not surprising that this inhibitory behavior is the dominant one only when field values are strong enough.

In M & R, we considered signals of the form  $h(g_i^{(1)})$  and obtained that the best such signals, restricted to values  $-1, 0$  or  $1$ , are of the form displayed in figure 1b. This form is consistent with the intuitive reasoning above, coupled with the wise social recommendation that unless a neuron has a strong reason for becoming active, it should remain silent in order not to transmit mostly noise. This orthodox Bayesian paradigm of transmitting a signal determined solely by the posterior belief (codified by the generalized field  $g_i^{(1)}$ ) should now be questioned, in the light of the intuitive explanation above. Although  $g_i^{(1)}$  is the only object relevant for determining how strong does neuron  $i$  feel about itself, this neuron’s earlier firing history  $(X_i, I_i)$  may remain relevant for determining the extent to which the neuron should warn other neurons about the presence of noise. Furthermore, history may help initially active neurons to correct initial signals now assessed as erroneous, as we shall indeed show.

## 4 Performance

### 4.1 Architecture parameters

This subsection introduces and illustrates certain parameters whose relevance will become apparent in section 4.3. There are  $N$  neurons in the network and  $K$  incoming synapses projecting on every neuron. If there is a synapse from neuron  $i$  to neuron  $j$ , the probability is  $r_2$  that there is a synapse from neuron  $j$  to neuron  $i$ . If there are synapses from  $i$  to  $j$  and from  $j$  to  $k$ , the probability is  $r_3$  that there is a synapse from  $i$  to  $k$ . If there are synapses from  $i$  to each of  $j$  and  $k$ , and from  $j$  to  $l$ , the probability is  $r_4$  that there is a synapse from  $k$  to  $l$ .

We saw in M & R that Bayesian neurons are adaptive enough to make  $r_2$  irrelevant for performance, but that  $r_3$  and  $r_4$ , which we took simply to be  $K/N$  assuming *fully random connectivity*, are of relevance. It is clear that if each neuron is connected to its  $K$  closest neighbors, then  $r_2$  is 1 and  $r_3$  and  $r_4$  are large. For *fully connected* networks all three are equal to 1.

For *Gaussian connectivity*, if neurons  $i$  and  $j$  are at a distance  $x$  from each other, then the probability that there is a synapse from  $j$  to  $i$  is

$$P(\text{synapse}) = pe^{-\frac{x^2}{2s^2}} \quad (17)$$

where  $p \in (0, 1]$  and  $s^2 > 0$  are parameters. Since the sum of  $n$  independent and identically distributed Gaussian random vectors is Gaussian with variance  $n$  times as large as that of the summands, we get that in  $d$ -dimensional space

$$\begin{aligned} r_k &= \int \left( pe^{-\frac{1}{2s^2} \sum_{i=1}^d x_i^2} \right) \frac{e^{-\frac{1}{2s^2(k-1)} \sum_{i=1}^d x_i^2}}{(2\pi(k-1)s^2)^{d/2}} dx_1 dx_2 \dots dx_d \\ &= \frac{p}{k^{d/2}} \int \frac{e^{-\frac{1}{2s^2((k-1)/k)} \sum_{i=1}^d x_i^2}}{(2\pi s^2((k-1)/k))^{d/2}} dx_1 dx_2 \dots dx_d = \frac{p}{k^{d/2}}. \end{aligned} \quad (18)$$

Thus, in 3-dimensional space,  $r_2 = p/(2\sqrt{2})$ ,  $r_3 = p/(3\sqrt{3})$ ,  $r_4 = p/8$ , depending on the parameter  $p$  but not on  $s$ .

For *multilayered networks* in which there is full connectivity between consecutive layers but no other connections,  $r_2$  and  $r_4$  are equal to 1 and  $r_3$  is 0 (unless there are three layers cyclically connected, in which case  $r_3 = 1$  as well).

## 4.2 One-iteration performance

Clearly, if neuron  $i$  had to choose for itself a sign on the basis of one iteration, this sign would have been

$$X_i^{(1)} = \text{sign}(O_i^{(1)}) . \quad (19)$$

Hence, letting  $\omega = \epsilon/\sqrt{\alpha_1}$ , if  $P(X_i = \xi_i) = (1+t)/2$  (where  $t$  is either  $\epsilon$  or  $\delta$ ), then after one iteration (similar to [Engelisch *et al.*, 1990]),

$$\begin{aligned} P(X_i^{(1)} = \xi_i) &= P(\lambda_i^{(1)} > 0.5 | \xi_i = 1) = P\left(g(t)X_i + \frac{\epsilon}{\alpha_1}f_i^{(1)} > 0 | \xi_i = 1\right) \quad (20) \\ &= \frac{1+t}{2}P\left(g(t) + \frac{\epsilon}{\alpha_1}(\epsilon + \sqrt{\alpha_1}Z) > 0\right) + \frac{1-t}{2}P\left(-g(t) + \frac{\epsilon}{\alpha_1}(\epsilon + \sqrt{\alpha_1}Z) > 0\right) \\ &= \frac{1+t}{2}\Phi\left(\omega + \frac{g(t)}{\omega}\right) + \frac{1-t}{2}\Phi\left(\omega - \frac{g(t)}{\omega}\right) \end{aligned}$$

where  $Z$  is a standard normal random variable and  $\Phi$  is its distribution function. Letting

$$Q^*(x, t) = \frac{1+t}{2}\Phi\left(x + \frac{g(t)}{x}\right) + \frac{1-t}{2}\Phi\left(x - \frac{g(t)}{x}\right) ; 0 \leq t < 1, x > 0 , \quad (21)$$

we see that (20) is expressible as  $Q^*(\omega, t)$ . Since the proportion of initially active neurons is  $n_1/K$ , the similarity after one iteration is

$$S_1 = \frac{n_1}{K}Q^*(\omega, \epsilon) + \left(1 - \frac{n_1}{K}\right)Q^*(\omega, \delta) . \quad (22)$$

As for the relation between the current similarity  $S_1$  and the initial similarity, observe that  $Q^*(x, t)$  is strictly increasing in  $x$  and converges to  $\frac{1+t}{2}$  as  $x \downarrow 0$ . Hence,  $S_1$  strictly exceeds the initial similarity  $\frac{n_1}{K}\frac{1+\epsilon}{2} + (1 - \frac{n_1}{K})\frac{1+\delta}{2}$ . Furthermore,  $S_1$  is a strictly increasing function of  $n_1$  ( $= m/\alpha_1$ ).

## 4.3 The second iteration

In order to analyze the effect of a second iteration, it is necessary to identify the (asymptotic) conditional distribution of the new input field  $f_i^{(2)}$ , defined by (12), given  $(\xi_i, X_i, I_i, f_i^{(1)})$ . Under a working paradigm that, given  $\xi_i, X_i$  and  $I_i$ , the input fields  $(f_i^{(1)}, f_i^{(2)})$  are jointly normally distributed, the conditional distribution of  $f_i^{(2)}$  given  $(\xi_i, X_i, I_i, f_i^{(1)})$  should be normal with mean depending linearly on  $f_i^{(1)}$  and variance independent of  $f_i^{(1)}$ . More explicitly, if  $(U, V)$  are jointly normally distributed with correlation coefficient  $\rho = \text{Cov}(U, V)/(\sigma_U\sigma_V)$ , then

$$E(V|U) = E(V) + \rho(\sigma_V/\sigma_U)(U - E(U)) \quad (23)$$

and

$$\text{Var}(V|U) = \text{Var}(V)(1 - \rho^2) . \quad (24)$$

Thus, the only parameters needed to define dynamics and evaluate performance are  $E(f_i^{(2)}|\xi_i, X_i, I_i)$ ,  $\text{Cov}(f_i^{(1)}, f_i^{(2)}|\xi_i, X_i, I_i)$  and  $\text{Var}(f_i^{(2)}|\xi_i, X_i, I_i)$ . In terms of these, the conditional distribution of  $f_i^{(2)}$  given  $(\xi_i, X_i, I_i, f_i^{(1)})$  is normal with

$$\begin{aligned} E(f_i^{(2)}|\xi_i, X_i, I_i, f_i^{(1)}) &= \\ &= E(f_i^{(2)}|\xi_i, X_i, I_i) + \frac{\text{Cov}(f_i^{(1)}, f_i^{(2)}|\xi_i, X_i, I_i)}{\text{Var}(f_i^{(1)}|\xi_i, X_i, I_i)} \left( f_i^{(1)} - E(f_i^{(1)}|\xi_i, X_i, I_i) \right) \end{aligned} \quad (25)$$

and

$$\text{Var}(f_i^{(2)}|\xi_i, X_i, I_i, f_i^{(1)}) = \text{Var}(f_i^{(2)}|\xi_i, X_i, I_i) - \frac{\text{Cov}^2(f_i^{(1)}, f_i^{(2)}|\xi_i, X_i, I_i)}{\text{Var}(f_i^{(1)}|\xi_i, X_i, I_i)} . \quad (26)$$

Assuming a model of joint normality, as in M & R, we rigorously identify limiting expressions for the three parameters of the model. Although we do not have as yet sufficient formal evidence pointing to the correctness of the joint normality assumption, the simulation results presented in section 6 fully support the adequacy of this common model.

In M & R we proved that  $E(f_i^{(2)}|\xi_i, X_i, I_i)$  is a linear combination of  $\xi_i$  and  $X_i I_i$ , which we denote by

$$E(f_i^{(2)}|\xi_i, X_i, I_i) = \epsilon^* \xi_i + b X_i I_i . \quad (27)$$

We also proved that  $\text{Cov}(f_i^{(1)}, f_i^{(2)}|\xi_i, X_i, I_i)$  and  $\text{Var}(f_i^{(2)}|\xi_i, X_i, I_i)$  are independent of  $(\xi_i, X_i, I_i)$ . These parameters determine the *regression coefficient*

$$a = \frac{\text{Cov}(f_i^{(1)}, f_i^{(2)}|\xi_i, X_i, I_i)}{\text{Var}(f_i^{(1)}|\xi_i, X_i, I_i)} \quad (28)$$

and the *residual variance*

$$\tau^2 = \text{Var}(f_i^{(2)}|\xi_i, X_i, I_i, f_i^{(1)}) . \quad (29)$$

These facts remain true in the current more general framework. We present in Appendix A formulas for  $a, b, \epsilon^*$  and  $\tau^2$ , whose derivation is cumbersome. The posterior probability that neuron  $i$  has memory state +1 is (see (9) and Lemma 1(II) in M & R)

$$\begin{aligned} \lambda_i^{(2)} &= P(\xi_i = 1 | X_i, I_i, f_i^{(1)}, f_i^{(2)}) = \\ &= \frac{1}{1 + \exp\{-2 [g_i^{(1)} + \frac{\epsilon^* - a\epsilon}{\tau^2} (f_i^{(2)} - a f_i^{(1)} - b X_i I_i)]\}} \end{aligned} \quad (30)$$

from which we obtain the final belief  $O_i^{(2)} = 2\lambda_i^{(2)} - 1 = \tanh(g_i^{(2)})$ , where  $g_i^{(2)}$  should be defined as

$$g_i^{(2)} = \left( \frac{\epsilon}{\alpha_1} - \frac{(\epsilon^* - a\epsilon)a}{\tau^2} \right) f_i^{(1)} + \left( \frac{\epsilon^* - a\epsilon}{\tau^2} \right) f_i^{(2)} + \begin{cases} g(\delta)X_i & \text{if } I_i = 0 \\ \left( g(\epsilon) - \frac{b(\epsilon^* - a\epsilon)}{\tau^2} \right) X_i & \text{otherwise} \end{cases} \quad (31)$$

to yield the final decision  $X_i^{(2)} = \text{sign}(g_i^{(2)})$ . Since  $(f_i^{(1)}, f_i^{(2)})$  are jointly normally distributed given  $(\xi_i, X_i, I_i)$ , any linear combination of the two, such as the one in expression (31), is normally distributed. After identifying its mean and variance, a standard computation reveals that the final similarity  $S_2 = P(X_i^{(2)} = \xi_i)$  - our global measure of performance - is given by a formula similar to expression (22) for  $S_1$ , with heavier activity  $n^*$  than  $n_1$ :

$$S_2 = \frac{n_1}{K} Q^* \left( \frac{\epsilon}{\sqrt{\alpha^*}}, \epsilon \right) + \left( 1 - \frac{n_1}{K} \right) Q^* \left( \frac{\epsilon}{\sqrt{\alpha^*}}, \delta \right) \quad (32)$$

where

$$\alpha^* = \frac{m}{n^*} = \frac{m}{n_1 + m \left( \frac{\epsilon^* / \epsilon - a}{\tau} \right)^2} . \quad (33)$$

In agreement with the ever-improving nature of Bayesian updatings, discussed in Appendix B,  $S_2$  exceeds  $S_1$  just as  $S_1$  exceeds the initial similarity. Furthermore,  $S_2$  is an increasing function of  $|\frac{\epsilon^* / \epsilon - a}{\tau}|$ .

## 5 Optimal signalling and performance

By optimizing over the factor  $|\frac{\epsilon^* / \epsilon - a}{\tau}|$  determining performance, we show in Appendix A that the optimal signal functions are

$$h_1(y) = R^*(y, \epsilon) - 1 , \quad h_0(y) = R^*(y, \delta) \quad (34)$$

where  $R^*$  is

$$R^*(y, t) = \frac{1}{\epsilon} (1 + r_3 \omega^2) [\tanh(\omega y) - c(\omega y - g(t))] \quad (35)$$

and  $c$  is a constant in  $(0, 1)$ .

The nonmonotone form of these functions, illustrated in figure 1, is clear. Neurons that have already signalled +1 in the first iteration have a lesser tendency to send positive signals than quiescent neurons. The signalling of quiescent neurons which receive no prior information ( $\delta = 0$ ) has a symmetric form.

The signal function of the initially active neurons may be shifted without affecting performance: if instead of taking  $h_1(y)$  to be  $R^*(y, \epsilon) - 1$  we take it to be  $R^*(y, \epsilon) - 1 + \Delta$

for some arbitrary  $\Delta$ , we will get the same performance because the effect of such  $\Delta$  on the second iteration input field  $f_i^{(2)}$  would be (see (12)) the addition of

$$\frac{1}{n_2} \sum_{j=1}^N W_{ij} I_{ij} \Delta X_j I_j = \Delta \frac{n_1}{n_2} f_i^{(1)} \quad (36)$$

which history-based Bayesian updating rules can fully adapt to. As shown in Appendix A,  $\Delta$  appears nowhere in  $(\epsilon^*/\epsilon - a)$  nor in  $\tau$  but it affects  $a$ . Hence,  $\Delta$  may be given several roles:

- Setting the ratio of the coefficients of  $f_i^{(1)}$  and  $f_i^{(2)}$  in (31) to a desired value, mimicking the passive decay of the membrane potential.
- Making the final decision  $X_i^{(2)}$  (see (31)) free of  $f_i^{(1)}$ , by letting the coefficient of the latter vanish. A judicious choice of the value of the reflexivity parameter  $r_2$  (which, just as  $\Delta$ , doesn't affect performance) can make the final decision  $X_i^{(2)}$  free of whether the neuron was initially quiescent or active. For the natural choice  $\delta = 0$  this will make the final decision free of the initial state as well and become simply the usual history-independent Hopfield rule  $X_i^{(2)} = \text{sign}(f_i^{(2)})$ , except that  $f_i^{(2)}$  is the result of carefully tuned slanted sigmoidal signalling.
- We may take  $\Delta = 1$  in which case both functions  $h_0$  and  $h_1$  are given simply by  $R^*(y, t)$ , where  $t = \epsilon$  or  $\delta$  depending on whether the neuron is initially active or quiescent. Let us express this signal explicitly in terms of history. By Table 1 and expression (11), the signal emitted by neuron  $i$  (whether it is active or quiescent) is

$$\begin{aligned} h(g_i^{(1)}, X_i, I_i) &= X_i h_{I_i} (X_i g_i^{(1)} / \omega) = \\ &= \frac{1 + r_3 \omega^2}{\epsilon} X_i \left[ \tanh(X_i g_i^{(1)}) - c(X_i g_i^{(1)} - g(t)) \right] = \\ &= \frac{1 + r_3 \omega^2}{\epsilon} \left[ \tanh(g_i^{(1)}) - c(g_i^{(1)} - X_i g(t)) \right] = \frac{1 + r_3 \omega^2}{\epsilon} \left[ \tanh(g_i^{(1)}) - c f_i^{(1)} \right]. \end{aligned} \quad (37)$$

We see that the signal is essentially equal to the sigmoid (see expression (10))  $\tanh(g_i^{(1)}) = 2\lambda_i^{(1)} - 1$ , modified by a correction term depending only on the current input field, in full agreement with the intuitive explanations of section 2. This correction is never too strong; note that  $c$  (see expression (75)) is always less than 1. In a fully-connected network  $c$  is simply

$$c = \frac{1}{1 + \omega^2},$$

i.e., in the limit of low memory load ( $\omega \rightarrow \infty$ ), the best signal is simply a sigmoidal function of the generalized input field.

To obtain a discretized version of the slanted sigmoid, we let the signal be  $sign(h(y))$  as long as  $|h(y)|$  is big enough - where  $h$  is the slanted sigmoid. The resulting signal, as a function of the generalized field, is (see figure 1a and 1b)

$$h_j(y) = \begin{cases} +1 & y < \beta_1^{(j)} \text{ or } \beta_4^{(j)} < y < \beta_5^{(j)} \\ -1 & y > \beta_6^{(j)} \text{ or } \beta_2^{(j)} < y < \beta_3^{(j)} \\ 0 & \text{otherwise} \end{cases} \quad (38)$$

where  $-\infty < \beta_1^{(0)} < \beta_2^{(0)} \leq \beta_3^{(0)} < \beta_4^{(0)} \leq \beta_5^{(0)} < \beta_6^{(0)} < \infty$  and  $-\infty < \beta_1^{(1)} < \beta_2^{(1)} \leq \beta_3^{(1)} < \beta_4^{(1)} \leq \beta_5^{(1)} < \beta_6^{(1)} < \infty$  define, respectively, the firing pattern of the neurons that were silent or active in the first iteration. To find the best such discretized version of the optimal signal, we search numerically for the activity level  $v$  which maximizes performance. Every activity level  $v$ , used as a threshold on  $|h(y)|$ , defines the (at most) twelve parameters  $\beta_i^{(j)}$  (which are identified numerically via the Newton-Raphson method) as illustrated in figure 1b.

## 6 Results

Using the formulation presented in the previous section, we investigate numerically the two-iteration performance achieved in several network architectures with optimal analog and discretized signalling. First we repeat the example of a cortical-like network investigated in M & R, but now with optimal analog and discretized signalling. The nearly identical marked superiority of optimal analog and discretized dynamics over the previous, posterior-probability-based signalling is evident, as shown in figure 2. While low activity is enforced in the first iteration, the number of neurons allowed to become active in the second iteration is not restricted, and best performance is typically achieved when about 70% of the neurons in the network are active (both with optimal signalling and with the previous, heuristic signalling). An initial similarity of 75% leads practically to 100% similarity after two such iterations.

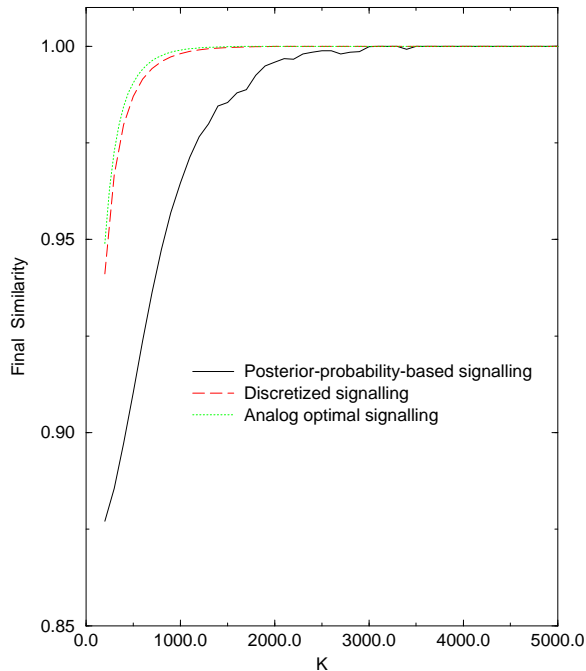


Figure 2: Two-iteration performance as a function of connectivity  $K$ . Network parameters are  $N = 5000$ ,  $n_1 = 200$ , and  $m = 50$ . All neurons receive their input state with similar initial overlap  $\epsilon = \delta = 0.5$ .

Figure 3 displays the performance achieved in the same network, when the input signal is applied only to the small fraction (4%) of neurons which are active in the first iteration (expressing possible limited resources of input information). As in figure 2, the performance

loss due to discretization is not considerable. We see that (for  $K > 1000$ ) near perfect final similarity is achieved even when the 96% initially quiescent neurons get no initial clue as to their true memory state, if no restrictions are placed on the second iteration activity level.

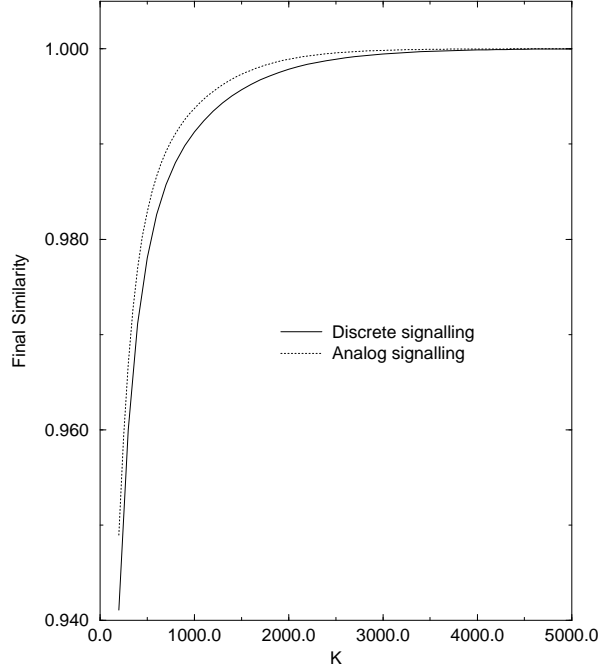


Figure 3: Two-iteration performance in a low-activity network as a function of connectivity  $K$ . Network parameters are  $N = 5000$ ,  $m = 50$ ,  $n_1 = 200$ ,  $\epsilon = 0.5$  and  $\delta = 0$ .

Figure 4 fixes the value of  $\omega = \frac{\epsilon}{\sqrt{\alpha_1}} = 1$ , and contrasts the case ( $n_1 = 200, \epsilon = 0.5$ ) of figure 3 with ( $n_1 = 50, \epsilon = 1$ ). The latter corresponds to applying perfect input patterns to one-fourth of the neurons receiving 75%-similarity input patterns in the former, and no initial information to the remaining three fourths. Since initial similarity is given by  $\frac{n_1 \epsilon}{K} = \frac{m \omega^2}{K \epsilon}$ , which is inversely proportional to  $\epsilon$ , the overall initial similarity under ( $n_1 = 50, \epsilon = 1$ ) is only half its value under ( $n_1 = 200, \epsilon = 0.5$ ). In spite of this, it achieves a slightly higher final similarity. This supports the idea that the input pattern should not be applied as the conventional uniformly distorted version of the correct memory, but rather as a less distorted pattern applied only to a small subset of the neurons.

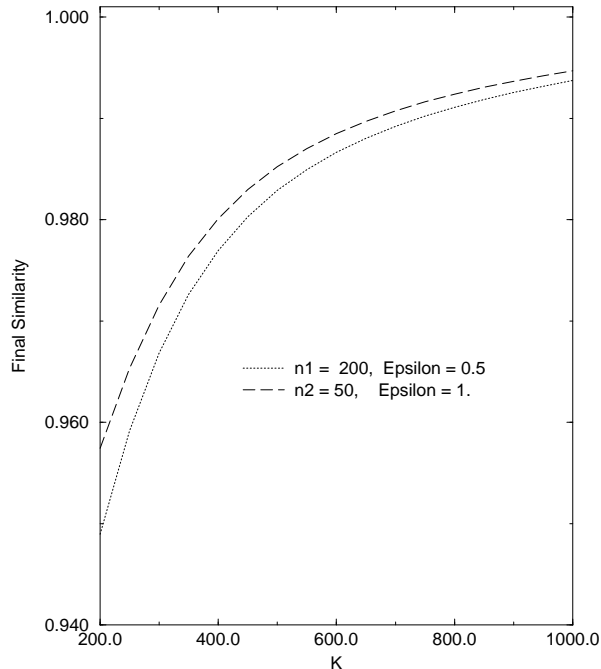


Figure 4: Two-iteration performance of analog signalling in a low-activity network as a function of connectivity  $K$ . Network parameters are  $N = 5000$ ,  $m = 50$ ,  $\omega = 1$  and  $\delta = 0$ .

Figure 5 illustrates the performance when connectivity and the number of signals received by each neuron are held fixed, but the network size is increased. A region of decreased performance is evident at mid-connectivity ( $K \approx N/2$ ) values, due to the increased residual variance. Hence, for neurons capable of forming  $K$  connections on the average, the network should either be fully connected or have a size  $N$  much larger than  $K$ . Since (unavoidable eventually) synaptic deletion would sharply worsen the performance of fully connected networks, cortical ANNs should indeed be sparsely connected. As evident, performance approaches an upper limit (the performance achieved with  $r_3 = 0$  and  $r_4 = 0$ ) as the network size is increased, and any further increase in the network size is unrewarding. The final similarity achieved in the fully connected network (with  $N = K = 200$ ) should be noted. In this case, the memory load (0.2) is significantly above the critical capacity of the Hopfield network [Amit *et al.*, 1985], but optimal history-dependent dynamics still manage to achieve a rather high two-iterations similarity (0.975) from initial similarity 0.75. This is in agreement with the findings of [Morita, 1993, Yoshizawa *et al.*, 1993], who show that nonmonotone dynamics increase capacity.

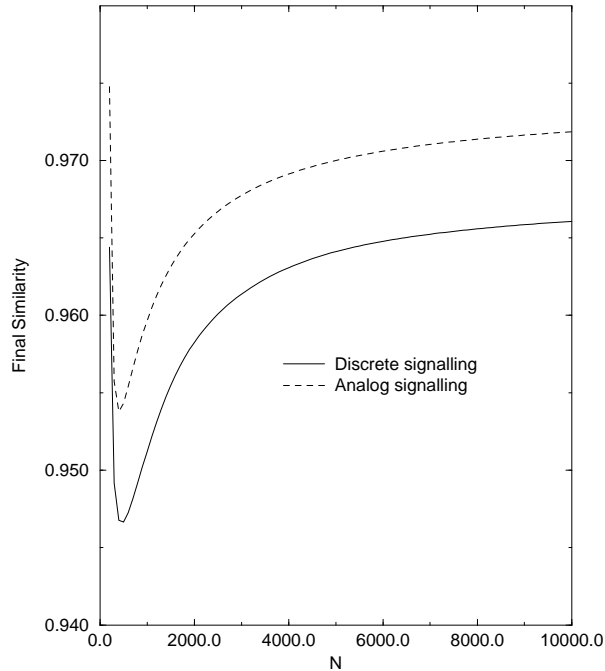


Figure 5: Two-iteration performance in a full-activity network as a function of network size  $N$ . Network parameters are  $n_1 = K = 200$ ,  $m = 40$  and  $\epsilon = 0.5$ .

Our theoretical predictions have been extensively examined by network simulations, and already in relatively small-scale networks close correspondence is achieved. For example, simulating a fully-connected network storing 100 memories with 500 neurons, the performance achieved with discretized dynamics under initial full activity (averaged over 100 trials, with  $\epsilon = 0.5$  and  $\delta = 0$ ) was 0.969 versus the 0.964 predicted theoretically. When  $m$ ,  $n_1$  and  $K$  were reduced by half (i.e.,  $N = 500$ ,  $K = 250$ ,  $m = 50$  and  $n_1 = 250$ ) the predicted performance was 0.947 and that achieved in simulation was 0.946. When  $m$ ,  $n_1$  and  $K$  were further reduced by half (into  $K = 125$ ,  $m = 25$  and  $n_1 = 125$ ) the predicted performance was 0.949 and that actually achieved was 0.953. In a larger network, with  $N = 1500$ ,  $K = 500$ ,  $m = 50$ ,  $n_1 = 250$ ,  $\epsilon = 0.5$  and  $\delta = 0$ , the predicted performance is 0.977 and that obtained numerically was 0.973.

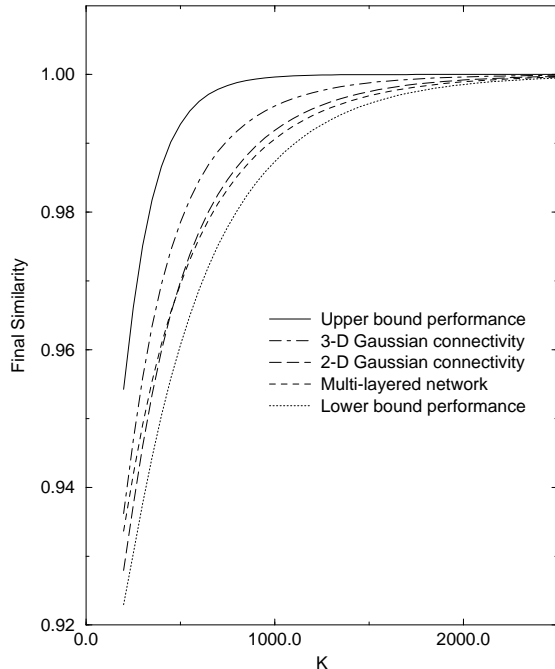


Figure 6: Two-iteration performance achieved with various network architectures, as a function of the network connectivity  $K$ . Network parameters are  $N = 5000$ ,  $n_1 = 200$ ,  $m = 50$ ,  $\epsilon = 0.5$  and  $\delta = 0$ .

Figure 6 illustrates the performance achieved with various network architectures, all sharing the same network parameters  $N, K, m$  and input similarity parameters  $n_1, \epsilon, \delta$ , but differing in the spatial organization of the neurons' synapses. Five different configurations are examined, characterized by different values of the architecture parameters  $r_3$  and  $r_4$ , as described in section 4.1. The upper bound on the final similarity that can be achieved in ANNs in two iterations is demonstrated by letting  $r_3 = 0$  and  $r_4 = 0$ . A lower bound (i.e., the worst possible architecture) on the performance gained with optimal signalling has been calculated by letting  $r_4 = 1$  and searching for  $r_3$  values that yielded the worst performance (such values began around 0.6 and increased to  $\approx 0.8$  as  $K$  was increased). The performance of the Multi-layered architecture was calculated by letting  $r_4 = 1$  and  $r_3 = 0$ . Finally, the worst performance achievable with 2-D and 3-D Gaussian connectivity (corresponding to  $p = 1$  in (17)) has been demonstrated by letting  $r_3 = 1/3, r_4 = 1/4$  and  $r_3 = 1/(3\sqrt{3}), r_4 = 1/8$  respectively. As evident, even in low-activity sparse-connectivity conditions, the decrease in performance with Gaussian connectivity (in relation, say, to the upper bound) does not seem considerable. Hence, history-dependent ANNs can work well

in a cortical-like architecture. It is interesting but not surprising to see that 3-D Gaussian-connectivity architecture is superior to the 2-D one along the whole connectivity range. Random connectivity, with  $r_3 = r_4 = K/N$ , is not displayed but is slightly above the performance achieved with 3-D Gaussian connectivity.

We have found Gaussian connectivity architectures under which the optimal final decision is simply the (history-free) sign of the last input field, but failed to achieve this under random connectivity. For example, this was achieved in a 3-D Gaussian connectivity network of  $N = 5000$ ,  $m = 50$ ,  $K = 1000$ ,  $n_1 = 163$ ,  $\epsilon = 0.55$  and  $\delta = 0$  with the architecture parameters  $p = 1, r_2 = 0.35, r_3 = 0.19, r_4 = 0.125$  and  $\Delta = -0.5$ . The final similarity achieved is 0.995.

## 7 Discussion

We have shown that Bayesian history-dependent dynamics make performance increase with every iteration, and that two iterations already achieve high similarity. The Bayesian framework gives rise to the slanted-sigmoid as the optimal signal function, displaying the nonmonotone shape proposed by [Morita, 1993]. The two-iteration performance has been analyzed in terms of general connectivity architectures, initial similarity and activity level.

The optimal signal function has some interesting biological perspectives. The possibly asymmetric form of the function, where neurons that have been silent in the previous iteration have an increased tendency to fire in the next iteration versus previously active neurons, is reminiscent of the bi-threshold phenomenon observed in biological neurons (see [Tam, 1992] for a review), where the threshold of neurons held at a hyperpolarized potential for a prolonged period of time is significantly lowered. As we have shown in section 5, the precise value of the parameter  $\Delta$  leads to different biological interpretations of the slanted sigmoid signal function. The most obvious one is letting  $\Delta$  set the ratio of the coefficients of  $f_i^{(1)}$  and  $f_i^{(2)}$  so as to mimic the decay of the membrane voltage. Perhaps more important, the finding that history-dependent neurons can maintain optimal performance in face of a broad range of  $\Delta$  values points out that neuromodulators may change the form of the signal function without changing the performance of the network. Obviously, the history-free variant of the optimal final decision is not resilient to such modulatory changes.

In this paper we have concentrated on analog signal functions, which, from a biological point of view, describe the dependence of the neuron's current firing rate on its (past

and present) firing and membrane potential. This continuous description must be put on common grounds with the discrete, dichotomous, description of the stored memory patterns the network converges to. To this end, we have let each neuron express its belief in its true state after the second iteration as a dichotomous signal, which has enabled us to define the final similarity measure upon which we optimize performance. However, this technically convenient description was used for simplicity, and does not necessarily imply the existence of some biological mechanism that converts the final value of the local field (or, equivalently, the posterior belief) into a dichotomous state. Alternatively, the posterior belief of the neuron that its true state is  $+1$ , determining its subsequent firing rate, may be interpreted as the probability of finding it in the  $+1$  state at a given ‘time slice’ during the next iteration.

It is reasonable to assume that no current formal neural model, with memoryless or history-dependent dynamics, with Hebbian or a more complex learning mechanism, is adequate enough to allow for a reliable quantitative description of the dynamics of the nervous system. In light of that, we do not consider the quantitative performance results reported here as being intrinsically important. What we do stress is the fact that the performance of ANN models can be heavily affected by dynamics, as exhibited by the sharp improvements obtained by fine tuning the neuron’s signal function. When there is a sizable evolutionary advantage to fine tuning, theoretical optimization becomes an important research tool: the solutions it provides and the qualitative features it deems critical may have their parallels in reality. In addition to the computational efficiency of nonmonotone signalling, the numerical investigations presented in the previous section point out to a few more features with possible biological relevance:

- In an efficient associative network, input patterns should be applied with high fidelity on a small subset of neurons, rather than spreading a given level of initial similarity as a low fidelity stimulus applied to a large subset of neurons.
- If neurons have some restriction on the number of connections they may form, such that each neuron forms some  $K$  connections on the average, then efficient ANNs, converging to high final similarity within few iterations, should be sparsely connected.
- With a properly tuned signal function, cortical-like Gaussian-connectivity ANNs perform nearly as well as randomly-connected ones.

We have gained some additional insight to the computational merits of non-monotone signal functions, as reported earlier by [Yoshizawa *et al.*, 1993, Morita, 1993]. However, inverse signalling may exist only in symmetric ANNs using the  $\pm 1$  notation. To study the realization of non-monotone signalling in the asymmetric  $0, 1$  formulation, we are currently undertaking the technically more complicated investigation of optimal history-dependent signalling in the latter framework [Meilijson and Ruppin, 1993b]. The non-monotone shape of the optimal signal function remains qualitatively similar. As may be deduced from a simple ‘mapping’ of the  $\pm 1$  sign reversal dynamics to the  $0, 1$  case, only neurons whose field values are greater than some low threshold and smaller than some high threshold should fire. This seemingly bizarre behavior may correspond well to the behavior of biological neurons; neurons with very high field values have most probably fired constantly in the previous ‘iteration’, and due to the effect of neural adaptation are now silenced.

## References

- [Abeles, 1991] M. Abeles. *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge University Press, 1991.
- [Amari and Maginu, 1988] S.I. Amari and K. Maginu. Statistical neurodynamics of associative memory. *Neural Networks*, 1:67–73, 1988.
- [Amit *et al.*, 1985] D. J. Amit, H. Gutfreund, and H. Sompolinsky. Storing infinite numbers of patterns in a spin-glass model of neural networks. *Phys. Rev. Lett.*, 55:1530–1533, 1985.
- [Braitenberg and Schuz, 1991] V. Braitenberg and A. Schuz. *Anatomy of the Cortex: Statistics and Geometry*. Springer-Verlag, 1991.
- [Connors and Gutnick, 1990] B.W. Connors and M.J. Gutnick. Intrinsic firing patterns of diverse neocortical neurons. *TINS*, 13(3):99–104, 1990.
- [Englich *et al.*, 1990] H. Englich, A. Engel, A. Schutte, and M. Stcherbina. Improved retrieval in nets of formal neurons with thresholds and non-linear synapses. *Studia Biophysica*, 137:37–54, 1990.
- [Felice *et al.*, 1993] P. De Felice, C. Marangi, G. Nardulli, G. Pasquariello, and L. Tedesco. Dynamics of neural networks with non-monotone activation function. *Network*, 4:1–9, 1993.
- [Haley, 1952] David C. Haley. Estimation of the dosage mortality relationship when the dose is subject to error. Technical Report TR-15, August 29, Stanford University, 1952.
- [Hopfield, 1982] J.J. Hopfield. Neural networks and physical systems with emergent collective abilities. *Proc. Nat. Acad. Sci. USA*, 79:2554, 1982.
- [Hopfield, 1984] J.J. Hopfield. Neurons with graded response have collective computational properties like those of two-state neurons. *Proc. Nat. Acad. Sci. USA*, 81:3088, 1984.
- [Komlós and Paturi, 1988] J. Komlós and R. Paturi. Convergence results in an associative memory model. *Neural Networks*, 1:239, 1988.
- [Lytton, 1991] W. Lytton. Simulations of cortical pyramidal neurons synchronized by inhibitory interneurons. *J. Neurophysiol.*, 66(3):1059–1079, 1991.

- [Meilijson and Ruppin, 1993a] I. Meilijson and E. Ruppin. History-dependent attractor neural networks. *Network*, 4:195–221, 1993.
- [Meilijson and Ruppin, 1993b] I. Meilijson and E. Ruppin. Optimal signalling in quiescent/firing attractor neural networks. *Preprint*, 1993.
- [Morita, 1993] M. Morita. Associative memory with nonmonotone dynamics. *Neural Networks*, 6:115–126, 1993.
- [Pearson *et al.*, 1987] J.C. Pearson, L.H. Finkel, and G.M. Edelman. Plasticity in the organization of adult cerebral cortical maps: A computer simulation based on neuronal group selection. *Journal of Neuroscience*, 7(12):4209–4223, 1987.
- [Schwindt, 1992] Peter C. Schwindt. Ionic currents governing input-output relations of betz cells. In T. McKenna, J. Davis, and S.F. Zornetzer, editors, *Single neuron computation*, pages 235–258. Academic Press, 1992.
- [Tam, 1992] David C. Tam. Signal processing in multi-threshold neurons. In T. McKenna, J. Davis, and S.F. Zornetzer, editors, *Single neuron computation*, pages 481–501. Academic Press, 1992.
- [Wilson and Cowan, 1972] H.R. Wilson and J.D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12:1–24, 1972.
- [Yoshizawa *et al.*, 1993] S. Yoshizawa, M. Morita, and S.-I. Amari. Capacity of associative memory using a nonmonotonic neuron model. *Neural Networks*, 6:167–176, 1993.

## Appendix A: Optimizing the signal function

We now present formulas for  $a, b, \epsilon^*$  and  $\tau^2$ , omitting the proofs which follow in a similar way to that presented in M & R. Let  $\phi(x) = \exp\{-x^2/2\}/\sqrt{2\pi}$  denote the standard normal density function. For given signal functions  $h_j$  ( $j = 0$  or  $1$ ), let

$$\Psi_j(x) = \int h_j(x+z)\phi(z)dz = \int h_j(y)\phi(y-x)dy \quad (39)$$

be the *integrated signal function*, let

$$\Psi'_j(x) = \frac{d}{dx}\Psi_j(x) = \int h_j(x+z)z\phi(z)dz = \int h_j(y)(y-x)\phi(y-x)dy \quad (40)$$

be its derivative, and

$$\Psi_j^{(a)}(x) = \int |h_j(x+z)|\phi(z)dz = \int |h_j(y)|\phi(y-x)dy \quad (41)$$

$$\Psi_j^{(s)}(x) = \int h_j^2(x+z)\phi(z)dz = \int h_j^2(y)\phi(y-x)dy \quad (42)$$

be, respectively, the integrated *absolute* and *square* signal functions. For any function  $H$ , define the operators

$$(Q^*H)(x, t) = \frac{1+t}{2}H\left(x + \frac{g(t)}{x}\right) + \frac{1-t}{2}H\left(x - \frac{g(t)}{x}\right) ; 0 \leq t < 1, x > 0 \quad (43)$$

$$(Q^+H)(x, t) = \frac{1+t}{2}H\left(x + \frac{g(t)}{x}\right) + \frac{1-t}{2}H\left(-x + \frac{g(t)}{x}\right) ; 0 \leq t < 1, x > 0 \quad (44)$$

$$(Q^-H)(x, t) = \frac{1+t}{2}H\left(x + \frac{g(t)}{x}\right) - \frac{1-t}{2}H\left(-x + \frac{g(t)}{x}\right) ; 0 \leq t < 1, x > 0 . \quad (45)$$

and let

$$\bar{\Psi}^{(a)} = \frac{n_1}{K}(Q^+\Psi_1^{(a)})(\omega, \epsilon) + \left(1 - \frac{n_1}{K}\right)(Q^+\Psi_0^{(a)})(\omega, \delta) \quad (46)$$

be the *level of activity*. If the signal functions  $h_j$  are supported by  $\{-1, 0, 1\}$ , then  $\bar{\Psi}^{(a)} = \frac{n_2}{K}$  is the proportion of neurons active in the second iteration, i.e., those for which  $|h| = 1$ , and if  $h$  has value in more general sets, we let  $\bar{\Psi}^{(a)} \cdot K = n_2$  be the definition of  $n_2$ , which only appears in expression (12) for the input field  $f_i^{(2)}$ . Using further the level of activity  $\bar{\Psi}^{(a)}$  as a natural normalizing factor, we now define

$$\epsilon^* = \left[ \frac{n_1}{K}(Q^-\Psi_1)(\omega, \epsilon) + \left(1 - \frac{n_1}{K}\right)(Q^-\Psi_0)(\omega, \delta) \right] / \bar{\Psi}^{(a)} \quad (47)$$

$$\bar{\Psi}' = \left[ \frac{n_1}{K}(Q^+\Psi_1')(\omega, \epsilon) + \left(1 - \frac{n_1}{K}\right)(Q^+\Psi_0')(\omega, \delta) \right] / \bar{\Psi}^{(a)} \quad (48)$$

$$\bar{\Psi}^{(s)} = \left[ \frac{n_1}{K}(Q^+\Psi_1^{(s)})(\omega, \epsilon) + \left(1 - \frac{n_1}{K}\right)(Q^+\Psi_0^{(s)})(\omega, \delta) \right]^{1/2} / \bar{\Psi}^{(a)} \quad (49)$$

$$\bar{\Psi} = (Q^+ \Psi_1)(\omega, \epsilon) / \bar{\Psi}^{(a)} . \quad (50)$$

In terms of these expressions and the architecture parameters defined in section 4.1,

$$a = \frac{m}{K\alpha_1} \bar{\Psi} + \frac{r_3}{\sqrt{\alpha_1}} \bar{\Psi}' \quad (51)$$

$$b = \sqrt{\alpha_1} r_2 \bar{\Psi}' \quad (52)$$

and

$$\tau^2 = \frac{m}{K} (\bar{\Psi}^{(s)})^2 - \frac{1}{\alpha_1} \left( \frac{m}{K} \right)^2 (\bar{\Psi})^2 + (r_4 - r_3^2) (\bar{\Psi}')^2 . \quad (53)$$

We now express  $\bar{\Psi}^{(a)}$ ,  $\epsilon^*/\epsilon - a$ ,  $\bar{\Psi}^{(s)}$ ,  $\bar{\Psi}$  and  $\bar{\Psi}'$  in a form amenable to analysis. Expressions (28),(47), (48) and (50) show that  $\epsilon^*$  and  $a$  are linear combinations of various integrated signal functions, each divided by the activity level  $\bar{\Psi}^{(a)}$ . We will first find a convenient expression for this activity level. From (44) and (46) we get

$$\begin{aligned} \bar{\Psi}^{(a)} &= \frac{n_1}{K} \int |h_1(y)| \left\{ \frac{1+\epsilon}{2} \phi \left( y - \omega - \frac{g(\epsilon)}{\omega} \right) + \frac{1-\epsilon}{2} \phi \left( y + \omega - \frac{g(\epsilon)}{\omega} \right) \right\} dy \\ &+ \left( 1 - \frac{n_1}{K} \right) \int |h_0(y)| \left\{ \frac{1+\delta}{2} \phi \left( y - \omega - \frac{g(\delta)}{\omega} \right) + \frac{1-\delta}{2} \phi \left( y + \omega - \frac{g(\delta)}{\omega} \right) \right\} dy \\ &= \frac{n_1}{K} \int |h_1(y)| \varphi_1(y) dy + \left( 1 - \frac{n_1}{K} \right) \int |h_0(y)| \varphi_0(y) dy \end{aligned} \quad (54)$$

where  $\varphi_1(y) = \varphi(y, \epsilon)$  and  $\varphi_0(y) = \varphi(y, \delta)$  are given by

$$\begin{aligned} \varphi(y, t) &= \frac{1+t}{2} \phi \left( y - \omega - \frac{g(t)}{\omega} \right) + \frac{1-t}{2} \phi \left( y + \omega - \frac{g(t)}{\omega} \right) \\ &= \sqrt{1-t^2} \exp \left\{ -\frac{1}{2} \left[ \omega^2 + (y - g(t)/\omega)^2 \right] \right\} \cosh(\omega y) / \sqrt{2\pi} . \end{aligned} \quad (55)$$

In the same manner, from (49) we have

$$\bar{\Psi}^{(s)} \bar{\Psi}^{(a)} = \left[ \frac{n_1}{K} \int h_1^2(y) \varphi_1(y) dy + \left( 1 - \frac{n_1}{K} \right) \int h_0^2(y) \varphi_0(y) dy \right]^{1/2} \quad (56)$$

and

$$(\epsilon^*/\epsilon - a) \bar{\Psi}^{(a)} = \frac{n_1}{K} \int h_1(y) R_1(y) \varphi_1(y) + \left( 1 - \frac{n_1}{K} \right) \int h_0(y) R_0(y) \varphi_0(y) \quad (57)$$

where  $R_1(y) = R(y, \epsilon) - 1$ ,  $R_0(y) = R(y, \delta)$  and  $R$  is given by the *slanted sigmoid* (see figure 1a)

$$R(y, t) = \frac{1}{\epsilon} [(1 + r_3 \omega^2) \tanh(\omega y) - r_3(\omega y - g(t))] \quad (58)$$

If (see (53))  $\tau^2$  was simply  $(m/K)(\bar{\Psi}^{(s)})^2$  then performance, measured by  $(\epsilon^*/\epsilon - a)/\tau$ , would have been expressible as

$$\frac{\int h R \varphi}{\sqrt{\frac{m}{K} \int h^2 \varphi}} = \sqrt{\frac{K}{m} \int R^2 \varphi} \frac{\int h R \varphi}{\sqrt{\int h^2 \varphi} \sqrt{\int R^2 \varphi}} . \quad (59)$$

The second factor, being a correlation, is bounded in absolute value by 1 (Cauchy-Schwartz inequality) and can be made equal to 1 (thus, automatically maximal) by letting the signal function be  $h \equiv R$ . If  $h$  is only allowed to have  $\{-1, 0, 1\}$  values, then  $\int h^2 \varphi = \int |h| \varphi = \bar{\Psi}^{(a)}$  and the problem may be seen as a two-stage procedure of maximizing  $\int h R \varphi$  for  $\int |h| \varphi$  constant and then maximizing their ratio as a function of this constant. The first stage is reminiscent of the Most Powerful tests in the theory of Statistics. Here the role of ‘power’ is played by  $\int h R \varphi$  and that of ‘level of significance’ by  $\int |h| \varphi$ . A slight generalization of the Neyman-Pearson lemma teaches us that  $h(y)$  should be nonzero ( $\pm 1$ ) only if  $|R(y)|$  exceeds some threshold, and should be given the sign of  $R$  (see figure 1b).

However, other terms appear in  $\tau^2$  and should be taken into account. Introducing the proper Lagrangian functions for the minimization of  $\tau^2$ , keeping  $\epsilon^*/\epsilon - a$  constant, we came in a lengthy but straightforward way to the realization that it is enough to consider signal functions of the type

$$h_1(y) = R^*(y, \epsilon) - 1, \quad h_0(y) = R^*(y, \delta) \quad (60)$$

where  $R^*$  is obtained from  $R$  by modifying the coefficient of the linear part

$$\begin{aligned} R^*(y, t) &= \frac{1}{\epsilon} [(1 + r_3 \omega^2) \tanh(\omega y) - (r_3 + \gamma)(\omega y - g(t))] \\ &= \frac{1}{\epsilon} (1 + r_3 \omega^2) [\tanh(\omega y) - c(\omega y - g(t))] \end{aligned} \quad (61)$$

We will now present closed form expressions for performance as well as for the best choice of  $c$ . Readers wishing to reproduce our computations should first ascertain the following formulas ((62) through (67)):

$$\int \varphi(y, t) dy = 1 \quad (62)$$

$$\int y \varphi(y, t) dy = \omega t + g(t)/\omega \quad (63)$$

$$\int y^2 \varphi(y, t) dy = 1 + \omega^2 + g^2(t)/\omega^2 + 2g(t)t \quad (64)$$

$$\int \tanh(\omega y) \varphi(y, t) dy = t \quad (65)$$

$$\int y \tanh(\omega y) \varphi(y, t) dy = \omega + g(t)t/\omega \quad (66)$$

$$\hat{\epsilon}(t) = \int (\tanh(\omega y))^2 \varphi(y, t) dy = \int (\tanh(\omega y))^2 [ \cdot ] \cosh(\omega y) dy \quad (67)$$

$$= \int \tanh(\omega y) [ \cdot ] \sinh(\omega y) dy$$

$$\approx \int \left[ 2\Phi\left(\frac{2}{1.702}\omega y\right) - 1 \right] [ \cdot ] \sinh(\omega y) dy$$

$$= 2 \left[ \frac{1+t}{2} \Phi \left( \frac{\omega^2 + g(t)}{\sqrt{.7242 + \omega^2}} \right) + \frac{1-t}{2} \Phi \left( \frac{\omega^2 - g(t)}{\sqrt{.7242 + \omega^2}} \right) \right] - 1$$

where we have used a sigmoidal approximation to the Gaussian cumulative distribution function (see [Haley, 1952])

$$\text{Sup}_x \left| \Phi(x) - \frac{1}{1 + \exp(-1.702x)} \right| < 0.01 . \quad (68)$$

We now present explicit formulas for the best choice of  $c$  (expression (75)) as well as for the ingredients needed for performance evaluation, i.e.,  $\epsilon^*$  (expression (73)),  $a$  (expression (74)) and the components  $\bar{\Psi}^{(s)}$ ,  $\bar{\Psi}$  and  $\bar{\Psi}'$  of  $\tau^2$  in expression (53). Let (see expression (67))

$$\hat{\epsilon} = \frac{n_1}{K} \hat{\epsilon}(\epsilon) + \left(1 - \frac{n_1}{K}\right) \hat{\epsilon}(\delta) \quad (69)$$

Then

$$\left( \bar{\Psi}^{(s)} \bar{\Psi}^{(a)} \right)^2 = \frac{1}{\epsilon^2} \left[ (1 + r_3 \omega^2)^2 \hat{\epsilon} + (r_3 + \gamma)^2 \omega^2 (1 + \omega^2) - 2(r_3 + \gamma) \omega^2 (1 + (r_3 - m/K) \omega^2) \right] \quad (70)$$

$$-(n_1/K)(1 - \Delta)(1 + \Delta + 2r_3 \omega^2)$$

$$\bar{\Psi} \bar{\Psi}^{(a)} = \Delta - \gamma \omega^2 \quad (71)$$

$$\bar{\Psi}' \bar{\Psi}^{(a)} = (\omega/\epsilon) \left[ (1 + r_3 \omega^2)(1 - \hat{\epsilon}) - r_3 - \gamma \right] \quad (72)$$

$$\epsilon^* \bar{\Psi}^{(a)} = \frac{1}{\epsilon} \left[ (1 + r_3 \omega^2) \hat{\epsilon} - (r_3 + \gamma) \omega^2 \right] - \epsilon (n_1/K)(1 - \Delta) \quad (73)$$

$$a \bar{\Psi}^{(a)} = \frac{\omega^2}{\epsilon^2} \left[ r_3 (1 + r_3 \omega^2) (1 - \hat{\epsilon}) - r_3 (r_3 + \gamma) - \epsilon^2 (n_1/K) \gamma \right] - (n_1/K) \Delta . \quad (74)$$

The best value of  $c$ , the multiplier of the input field in the linear correction term of the signal, is

$$c = \frac{(r_4 - r_3^2)(1 - \hat{\epsilon}) + \frac{m}{K} r_3}{r_4 - r_3^2 + \frac{m}{K} (1 + r_3 \omega^2)} . \quad (75)$$

Performance with optimal analog signalling may now be evaluated, since  $(\epsilon^*/\epsilon - a)^2/\tau^2 = (\epsilon^* \bar{\Psi}^{(a)}/\epsilon - a \bar{\Psi}^{(a)})^2/(\tau \bar{\Psi}^{(a)})^2$  does not require knowledge of the level of activity  $\bar{\Psi}^{(a)}$ , but in order to identify  $a$  and  $b$  and to implement the final decision  $X_i^{(2)}$ , we must identify  $\bar{\Psi}^{(a)}$ . For purposes of simulation, it is possible to evaluate numerically the zero-crossings of the signal functions  $h_0$  and  $h_1$  and present  $\bar{\Psi}^{(a)}$  in closed form in terms of these, or to evaluate  $\bar{\Psi}^{(a)}$  (as we have chosen to do) by numerical integration. With regard to the performance achieved with discretized signalling, the functions  $\Psi_j(x)$ ,  $\Psi_j'(x)$  and  $\Psi_j^{(a)}(x)$  (see expressions (39), (40) and (41)) are easily expressed as linear combinations of normal cumulative

distribution and density functions evaluated at the  $\beta$ 's, and performance evaluation follows, to be maximized with respect to the activity threshold  $v$ .

### Appendix B: Neurons as Bayesian decisionmakers

Neuron  $i$  starts with a prior probability  $\lambda_i^{(0)} = P(\xi_i = +1)$  and after observing input fields  $f_i^{(1)}, f_i^{(2)}, \dots, f_i^{(t)}$  computes the posterior probability

$$\lambda_i^{(t)} = P\left(\xi_i = +1 | f_i^{(1)}, f_i^{(2)}, \dots, f_i^{(t)}\right) \quad (76)$$

It now signals

$$h_i^{(t)} = h^{(t)}\left(\lambda_i^{(0)}, f_i^{(1)}, f_i^{(2)}, \dots, f_i^{(t)}\right) \quad (77)$$

and computes the new input field

$$f_i^{(t+1)} = \sum_j W_{ij} I_{ij} h_j^{(t)}. \quad (78)$$

This description proceeds inductively.

The stochastic process  $\lambda_i^{(0)}, \lambda_i^{(1)}, \lambda_i^{(2)}, \dots$  is of the form

$$X_t = E(Z | Y_1, Y_2, \dots, Y_t)$$

where  $Z = I_{\{\xi_i = +1\}}$  is a (bounded) random variable and the Y-process adds in every stage some more information to the data available earlier. Such a process is termed a *Martingale* in Probability theory. The following facts are well known, the first being actually the usual definition

1. For all  $t$ ,

$$E(X_{t+1} | Y_1, Y_2, \dots, Y_t) = X_t \quad a.s.$$

(where a.s. means ‘almost surely’ or ‘except for an event with probability zero’.)

2. In particular,  $E(X_t)$  is the same for all  $t$ .
3. If the finite interval  $[a, b]$  is such that  $P(a \leq X_t \leq b) = 1$  for all  $t$  and  $\Psi$  is a convex function on  $[a, b]$ , then for all  $t$ ,

$$E(\Psi(X_{t+1}) | Y_1, Y_2, \dots, Y_t) \geq \Psi(X_t) \quad a.s.$$

4. In particular, for all  $t$ ,

$$E(\Psi(X_t)) \leq E(\Psi(X_{t+1}))$$

5. (A special case of Doob’s Martingale Convergence Theorem)

For every bounded Martingale  $(X_t)$  there is a random variable  $X$  such that

$$X_t \rightarrow X \text{ as } t \rightarrow \infty, \text{ a.s.}$$

and in fact the Martingale is the sequence of ‘opinions’ about  $X$ : For all  $t$ ,

$$X_t = E(X|Y_1, Y_2, \dots, Y_t) \text{ a.s.}$$

6. In particular,  $E(X) = E(X_t)$  and  $E(\Psi(X)) \geq E(\Psi(X_t))$  for all  $t$ , for any convex function  $\Psi$  defined on  $[a, b]$ .

A neuron with posterior probability  $\lambda_i^{(t)}$  as in (76) decides momentarily that its true state is  $+1$  if  $\lambda_i^{(t)} > 1/2$  and  $-1$  if  $\lambda_i^{(t)} < 1/2$ . The strength of belief, or confidence in the preferred state, is given by the *convex* function  $\Psi(x) = \text{Max}(x, 1 - x)$  applied to the  $[0, 1]$ -bounded Martingale  $(\lambda_i^{(t)})$ . For large  $N$ , the current *similarity* of the network, or proportion of neurons whose preferred state is the correct one, is mathematically characterized as  $E(\Psi(\lambda_i^{(t)}))$ . By the above, Bayesian updatings are always such that every neuron has a well defined final decision about its state (we may call this a ‘fixed point’) and the network’s similarity increases with every iteration, being at the ‘fixed point’ even higher. This holds true for arbitrary signal functions  $h$ , and not only for those that are in some sense optimal.

In the current paper we developed signalling mechanisms and neuron’s decisions, and evaluated global similarity after two iterations. By the above, whatever similarity we achieve is a lower bound for what can be achieved by more iterations, unlike memoryless Hopfield dynamics which are known to do reasonably well at the beginning even below capacity, in which case they converge eventually to random fixed points [Amari and Maginu, 1988].

As for the similarity that could be achieved with memory-dependent iterations (78) and Bayesian updatings (76), it seems to be *in theory*, leaving biological plausibility aside, as high as could be achieved with a computer that is supplied with the entire initial state vector as well as with all weights  $W_{ij}$  and all connectivity data  $I_{ij}$  that define the global architecture of the network. Informally speaking, consider the following scenario: Have the neurons transmit some random noise that will give each neuron an identifying field-value signature. By letting the signal function  $h$  be non-zero in very small intervals that contain the signature of at most one neuron, it is possible to implement asynchronous dynamics, where at any given moment only a single neuron fires. Each neuron can now let all neurons into which

it has a synapse know the weight between the two by simply signalling  $h = 1$ . It can then transmit any information acquired from other neurons (such as a string composed of two neurons' signatures and the weight between them) by signalling any one-to-one function of this information, and have the receiving neuron compute the value of this function as the ratio between the input field and the (already known) synaptic weight between them, and then invert the one-to-one function. It is straightforward to see that via this process each neuron acquires the entire global information of the network's synaptic matrix and the initial state, as long as connectivity is such that for any two neurons  $i$  and  $j$  there is a directed path leading from  $i$  to  $j$ .