

History-dependent 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.

March 31, 1997

Abstract

We present a methodological framework enabling a detailed description of the performance of Hopfield-like attractor neural networks (ANN) in the first two iterations. Using the Bayesian approach, we find that performance is improved when a history-based term is included in the neuron's dynamics. A further enhancement of the network's performance is achieved by judiciously choosing the *censored* neurons (those which become active in a given iteration) on the basis of the magnitude of their post-synaptic potentials. The contribution of biologically plausible, censored, history-dependent dynamics is especially marked in conditions of low firing activity and sparse connectivity, two important characteristics of the mammalian cortex. In such networks, the performance attained is higher than the performance of two 'independent' iterations, which represents an upper bound on the performance of history-independent networks.

1 Introduction

Associative Attractor Neural Network (ANN) models provide a theoretical background for the understanding of human memory processes. Considerable effort has been devoted recently to narrow the gap between the original ANN Hopfield model [1, 2] and the realm of the structure and dynamics of the brain. This had led to the development of ANN models with low firing rates [3, 4, 5], to the investigation of the influence of various canonical connectivity architectures and various coding regimes on the network’s memory capacity [6], and to a model of ‘spiking’ neurons [7]. In this paper, we contribute to the examination of the performance of ANNs under cortical-like architectures, where neurons are typically connected to only a fraction of their neighboring neurons [8], and have a low firing activity [9]. We develop a general framework for examining various *signalling* mechanisms (firing functions) and *activation* rules (the mechanism for deciding which neurons are active in some interval of time). In cortical-like sparsely-connected low-activity networks we find that high performance is achieved when the neurons’ dynamics are governed by a threshold-censored function of their post-synaptic potential, considered by some authors to actually represent the firing of cortical neurons [5].

This brings us to the central theme of this paper, neurons’ dynamics governed by post-synaptic potentials. The Hopfield model is based on *memoryless dynamics*, which identify the notion of ‘post-synaptic potential’ with the input field received by a neuron from the neurons active in the current iteration. We follow a Bayesian approach under which the neuron’s signalling and activation decisions are based on the current a-posteriori probabilities assigned to its two possible true memory states, ± 1 . As we shall see, the a-posteriori belief in $+1$ is the sigmoidal function evaluated at a neuron’s *generalized field*, a linear combination of present and past input fields. As will be shown, these *history-dependent dynamics* can considerably outperform the memoryless ones, even to the point of making the two-iterations performance strictly better than the performance of two *independent* iterations, known to be an upper bound on the performance of memoryless dynamics [10]. As mentioned earlier, performance can be significantly enhanced if active neurons are chosen by the magnitude of the post-synaptic potential rather than randomly. However, performance under this more biologically appealing activation regime is sensitive to connectivity.

The neural network model presented is characterized as follows. There are m ‘random memories’ ξ^μ , $1 \leq \mu \leq m$, and one ‘true’ memory $\xi^{m+1} = \xi$. The $(m+1)N$ entries of these

memories are independent and identically distributed, with equally likely values of $+1$ or -1 . The initial state X has *similarity* $P(X_i = \xi_i) = (1 + \epsilon)/2$, $P(X_i = -\xi_i) = (1 - \epsilon)/2$, independently of everything else. The weight of the synaptic connection between neurons i and j ($i \neq j$) is given by the simple Hebbian law

$$W_{ij}^* = \sum_{\mu=1}^{m+1} \xi_i^\mu \xi_j^\mu . \quad (1)$$

Each neuron receives incoming synaptic connections from a random choice of K of the N neurons in the network in such a way that if a synapse exists, the synapse in the opposite direction exists with probability r , the *reflexivity* parameter. In the first iteration, a random sample of L_1 neurons become *active* (i.e., ‘fire’), thus on the average $n_1 = L_1 K/N$ neurons update the state of each neuron. The field $f_i^{(1)}$ of neuron i in the first iteration is

$$f_i^{(1)} = \frac{1}{n_1} \sum_{j=1}^N W_{ij}^* I_{ij} I_j^{(1)} X_j , \quad (2)$$

where I_{ij} denotes the indicator function of the event ‘neuron i receives a synaptic connection from neuron j ’, and $I_j^{(t)}$ denotes the indicator function of the event ‘neuron j is active in the t ’th iteration’. Under the Bayesian approach we adopt, neuron i assigns an a-priori probability $\lambda_i^{(0)} = P(\xi_i = +1 | X_i) = (1 + \epsilon X_i)/2$ to having $+1$ as the correct memory state and evaluates the corresponding a-posteriori probability $\lambda_i^{(1)} = P(\xi_i = +1 | X_i, f_i^{(1)})$, which turns out to be expressible as the sigmoidal function $1/(1 + \exp(-2x))$ evaluated at some linear combination of X_i and $f_i^{(1)}$.

In the second iteration the belief $\lambda_i^{(1)}$ of a neuron determines the probability that the neuron is active. We illustrate two extreme modes for determining the active updating neurons, or *activation*: the *random* case where L_2 active neurons are randomly chosen, independently of the strength of their fields, and the *censored* case, which consists of selecting the L_2 neurons whose belief belongs to some set. The most appealing censoring rule from the biological point of view is *tail-censoring*, where the active neurons are those with the strongest beliefs. Performance, however, is improved under *interval-censoring*, where the active neurons are those with mid-range beliefs, and even further by combining tail and interval censoring into a *hybrid* rule.

Let $n_2 = L_2 K/N$. The activation rule is given by a function $C : [\frac{1}{2}, 1] \rightarrow [0, 1]$. Neuron j , with belief $\lambda_j^{(1)}$ in $+1$, becomes active with probability $C(\max(\lambda_j^{(1)}, 1 - \lambda_j^{(1)}))$, independently of everything else. For example, the random case corresponds to $C \equiv \frac{L_2}{N}$ and

the tail-censored case corresponds to $C(\lambda) = 1$ or 0 depending on whether $\max(\lambda, 1 - \lambda)$ exceeds some threshold. The output of an active neuron j is a *signal function* $S(\lambda_j^{(1)})$ of its current belief. The field $f_i^{(2)}$ of neuron i in the second iteration is

$$f_i^{(2)} = \frac{1}{n_2} \sum_{j=1}^N W_{ij} * I_{ij} I_j^{(2)} S(\lambda_j^{(1)}) . \quad (3)$$

Neuron i now evaluates its a-posteriori belief $\lambda_i^{(2)} = P(\xi_i = +1 | X_i, I_i^{(1)}, f_i^{(1)}, f_i^{(2)})$. As we shall see, $\lambda_i^{(2)}$ is, again, the sigmoidal function evaluated at some linear combination of the neuron's history $X_i, X_i I_i^{(1)}, f_i^{(1)}$ and $f_i^{(2)}$. In contrast to the common history-independent Hopfield dynamics where the signal emitted by neuron j in the t' th iteration is a function of $f_j^{(t-1)}$ only, Bayesian history-dependent dynamics involve signals and activation rules which depend on the neuron's generalized field, obtained by adaptively incorporating $f_j^{(t-1)}$ to its previous generalized field. The final state $X_i^{(2)}$ of neuron i is taken as -1 or $+1$, depending on which of $1 - \lambda_i^{(2)}$ and $\lambda_i^{(2)}$ exceeds $1/2$.

For $n_1/N, n_2/N, m/N, K/N$ constant, and N large, we develop explicit expressions for the performance of the network, for any signal function (e.g., $S_1(\lambda) = \text{Sgn}(\lambda - 1/2)$ or $S_2(\lambda) = 2\lambda - 1$) and activation rule. Performance is measured by the final *overlap* $\epsilon'' = \frac{1}{N} \sum \xi_i X_i^{(2)}$ (or equivalently by the final similarity $(1 + \epsilon'')/2$). Various possible combinations of activation modes and signal functions described above are then examined under varying degrees of connectivity and neuronal activity.

Previous research investigating the dynamic evolution of ANNs has distinguished between two general classes of sparse networks. In strongly diluted networks, where the number of connections per neuron is smaller than the logarithm of the number of neurons, the dynamics become exactly soluble due to the elimination of correlations between the inputs to different neurons [11]. In the second class, which we address, sparseness is not necessarily so extensive, correlations exist and should be accounted for. We follow the approach originated by [12] and developed further in [13], who derived explicit formulas for the evolution of the first two or three iterations. This approach was extended by [14] to multiple iterations, postulating an approximate multi-step transition equation based on the assumption that the noise term is correlated with history only through its variance and its macroscopic overlap. Further developments saw a self-stimulus hysteretic term included in the dynamics [15, 16], and introduced arbitrary connectivity in the non-hysteretic case [17]. Extending the idea of hysteresis to the broader notion of history-dependent dynamics, we find that the performance of history-dependent ANNs 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.

From a biological perspective, the history-dependent approach is strongly motivated by the observation that the time span of the different channel conductances in a given neuron is very broad (see, e.g., [18] for a review). While some channels are active for only microseconds, some slow-acting channels may remain open for seconds. Hence, a synaptic input currently impeding on the neuron may influence both its current post-synaptic membrane potential, and its post-synaptic potential at some future time. We distinguish accordingly between input fields $f_i^{(t)}$ that model incoming spikes, and generalized fields that model the history-dependent, adaptive post-synaptic potentials. As a neuron fires in spikes, we emphasize dichotomous signal functions. However, we perform the mathematical derivations under arbitrary signal functions, for their possible benefits in artificial neural networks. This approach may be biologically justified as well, due to the broad time span upon which a given synaptic input may influence the membrane potential. The notion of ‘iteration’ may be then viewed as an abstraction for the overall dynamics for some length of time, during which some continuous non-threshold input/output function, such as the conventional sigmoidal function [19, 2, 20], governs the firing rate of each neuron. History-dependent terms may also express hysteretic properties of the neuron’s dynamics [15], as manifested by its refractory period and the prevalence of bursting.

This paper is organized as follows. In section 2 we show how single-iteration optimization results in a straightforward manner from the Bayesian approach. Section 3 presents the rationale behind history-dependent dynamics. Section 4 describes the optimization of two iterations, while detailed relevant calculations are relegated to the Appendix. Section 5 capitalizes on the derivations of section 4 and uses Statistical Hypotheses Testing analogies to present a heuristic rationale for the various activation rules studied in this paper. Formulas pertaining to these activation rules are shown in section 6 and illustrated numerically in section 7.

2 Single-iteration optimization: the Bayesian approach

Let us first summarize for ease of reference the following well known basic fact in Bayesian Hypothesis Testing.

Lemma 1

(i) If the prior distribution of ξ is $P(\xi = 1) = P(\xi = -1) = \frac{1}{2}$ and the conditional distribution of the observable X is

$$P(X = \xi|\xi) = \frac{1 + \epsilon}{2}, \quad P(X = -\xi|\xi) = \frac{1 - \epsilon}{2} \quad (4)$$

then the posterior probability that $\xi = 1$ is

$$P(\xi = 1|X) = \frac{1 + \epsilon X}{2} = \frac{1}{1 + e^{-2\epsilon\gamma(\epsilon)X}} \quad (5)$$

where

$$\gamma(\epsilon) = \frac{1}{2\epsilon} \log \frac{1 + \epsilon}{1 - \epsilon}. \quad (6)$$

(ii) Express the prior probability as

$$P(\xi = 1) = \frac{1}{1 + e^{-2x}} \quad (7)$$

and assume an observable Y which, given ξ , is distributed according to

$$Y|\xi \sim N(\mu\xi, \sigma^2) \quad (8)$$

for some constants $\mu \in (-\infty, \infty)$ and $\sigma^2 \in (0, \infty)$. Then the posterior probability is

$$P(\xi = 1|Y = y) = \frac{1}{1 + e^{-2(x + (\mu/\sigma^2)y)}}. \quad (9)$$

This formula is the basis for the definition of a generalized field x which is updated by the addition of a multiple of the current observation.

To analyze the field $f_i^{(1)}$ of neuron i in the first iteration, we express $W_{ij}^* = \xi_i \xi_j + W_{ij}$, where W_{ij} is the contribution of the random memories. Then

$$f_i^{(1)} = \frac{1}{n_1} \sum_{j=1}^N \xi_i \xi_j I_{ij} I_j^{(1)} X_j + \frac{1}{n_1} \sum_{j=1}^N W_{ij} I_{ij} I_j^{(1)} X_j = \epsilon \xi_i + \frac{1}{n_1} \sum_{j=1}^N W_{ij} I_{ij} I_j^{(1)} X_j \quad (10)$$

where $I_j^{(t)} = 1$ or 0 depending on whether or not neuron j is active in the t 'th iteration. In the sequel, Z or Z_j will always denote standard normal variables. $f_i^{(1)}$ is asymptotically normally distributed given X_i and ξ_i , with

$$E(f_i^{(1)}|\xi_i, X_i) = \epsilon \xi_i \quad (11)$$

and

$$\text{Var}(f_i^{(1)}|\xi_i, X_i) = \frac{m}{n_1} \equiv \alpha_1 . \quad (12)$$

If we apply Lemma 1(ii) to $Y = f_i^{(1)}$, with $\mu = \epsilon$ and $\sigma^2 = \alpha_1$, we see that

$$\lambda_i^{(1)} = P(\xi_i = 1|X_i, f_i^{(1)}) = \frac{1}{1 + e^{-2\epsilon(\gamma(\epsilon)X_i + f_i^{(1)}/\alpha_1)}} . \quad (13)$$

Hence, $P(\xi = 1|X_i, f_i^{(1)}) > 1/2$ if and only if $f_i^{(1)} + \alpha_1\gamma(\epsilon)X_i > 0$.

The single-iteration performance is then given by the similarity

$$\begin{aligned} \frac{1 + \epsilon'}{2} &= P\left((f_i^{(1)} + \alpha_1\gamma(\epsilon)X_i)\xi_i > 0|\xi_i\right) = \\ &= \frac{1 + \epsilon}{2}P(f_i^{(1)} + \alpha_1\gamma(\epsilon)X_i > 0|\xi_i = 1, X_i = 1) + \frac{1 - \epsilon}{2}P(f_i^{(1)} + \alpha_1\gamma(\epsilon)X_i > 0|\xi_i = 1, X_i = -1) = \\ &= \frac{1 + \epsilon}{2}P(\epsilon + \sqrt{\alpha_1}Z + \alpha_1\gamma(\epsilon) > 0) + \frac{1 - \epsilon}{2}P(\epsilon + \sqrt{\alpha_1}Z - \alpha_1\gamma(\epsilon) > 0) = \\ &= \frac{1 + \epsilon}{2}\Phi\left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon)\sqrt{\alpha_1}\right) + \frac{1 - \epsilon}{2}\Phi\left(\frac{\epsilon}{\sqrt{\alpha_1}} - \gamma(\epsilon)\sqrt{\alpha_1}\right) \\ &\equiv Q(\epsilon, \alpha_1) \end{aligned} \quad (14)$$

where Φ is the standard normal distribution function. As we see, (14) agrees with Englisch et. al.'s result [21]. Single-iteration performance (reaching overlap ϵ') is seen to be maximized by adding the contribution $m\gamma(\epsilon)X$ to the neuron's unnormalized input field, which has been interpreted by Englisch et al. as the addition of an appropriate self-stimulus diagonal term to the neuron's connectivity. The Hopfield dynamics, modified by redefining W_{ii} as $m\gamma(\epsilon)$ (in the Neural Network terminology) is equivalent (in the Bayesian jargon) to the obvious optimal policy, under which a neuron sets for itself the sign with posterior probability above 1/2 of being correct. This gives the best single-iteration performance, satisfying $\epsilon' > \epsilon$ for all α_1 , (because $Q(\epsilon, \alpha) \rightarrow (1 + \epsilon)/2$ as $\alpha \uparrow \infty$ and $\frac{\partial}{\partial \sqrt{\alpha}}Q(\epsilon, \alpha) = -(1/\alpha)((1 - \epsilon^2)/(2\pi))^{-1/2} \exp\{-(\epsilon^2/\alpha + \alpha(\gamma(\epsilon))^2)/2\}$ is negative) unlike the usual $\gamma = 0$ case, where $\epsilon' < \epsilon$ if α_1 is too large. For $\gamma = 0$ we see that the similarity after one iteration is $(1 + \epsilon')/2 = \Phi(\frac{\epsilon}{\sqrt{\alpha_1}})$, as shown by Kinzel [22]. This corresponds to the classical (non-Bayesian) approach in Statistics that values unbiasedness over the use of prior beliefs. Under both policies the neuron decides on +1 if $f_i^{(1)}$ exceeds some *decision threshold* $T_i^{(1)}$, which is 0 or $\alpha_1\gamma(\epsilon)X_i$.

3 The rationale behind history-dependent dynamics

We continue to follow the paradigm that the neuron is a Bayesian decision maker, and that the input field arising from the currently active neurons is an ‘observation’ in the statistical sense, allowing the neuron to update its belief concerning its correct memory. Since, as we have seen in the previous section (equation (13)), the neuron’s belief in its correct state after a single iteration is a strictly increasing function of its generalized field, this belief typically varies from neuron to neuron. Performing a second independent iteration (viewing the second iteration as a first iteration with the corresponding new initial similarity [10]) corresponds to splitting the neurons into two sets by the preferred true memory state, and then replacing the belief of each neuron by the average belief of its class. We shall show now that heterogeneity of prior beliefs induces higher performance.

Consider a neuron with prior probability λ of having state $+1$, and let h and g be the densities of the observable random variable under states $+1$ and -1 respectively. Then the posterior probability of state $+1$, as given by Bayes formula, is $\lambda^{(1)} = \lambda h(x) / (\lambda h(x) + (1 - \lambda)g(x))$. Recall that the similarity of a neuron is obtained from the belief λ in $+1$ by evaluating the convex function $\psi(x) = \max(x, 1 - x)$ at λ . We now take any twice continuously differentiable convex function ψ on $[0, 1]$ and evaluate the expectation of $\psi(\lambda^{(1)})$:

$$E\psi(\lambda^{(1)}) = \int \psi \left(\frac{\lambda f(x)}{\lambda h(x) + (1 - \lambda)g(x)} \right) (\lambda h(x) + (1 - \lambda)g(x)) d\mu(x) \quad (15)$$

$$\begin{aligned} \frac{d}{d\lambda} E\psi(\lambda^{(1)}) &= \int \psi' \left(\frac{\lambda h(x)}{\lambda h(x) + (1 - \lambda)g(x)} \right) \frac{h(x)g(x)}{\lambda h(x) + (1 - \lambda)g(x)} d\mu(x) + \\ &\int \psi \left(\frac{\lambda h(x)}{\lambda h(x) + (1 - \lambda)g(x)} \right) (h(x) - g(x)) d\mu(x) \end{aligned} \quad (16)$$

$$\frac{d^2}{d\lambda^2} E\psi(\lambda^{(1)}) = \int \psi'' \left(\frac{\lambda h(x)}{\lambda h(x) + (1 - \lambda)g(x)} \right) \frac{h^2(x)g^2(x)}{(\lambda h(x) + (1 - \lambda)g(x))^3} d\mu(x) \quad (17)$$

Since expression (17) is positive, we see that the expectation of a well behaved convex function of the posterior probability is a convex function of the prior probability. Since twice continuously differentiable convex functions on a compact interval are dense in the class of all continuous convex functions on that interval, this fact extends to nondifferentiable convex functions as well, such as $\psi(x) = \max(x, 1 - x)$.

We have seen that similarity is the expectation of some convex function of the belief. We conclude by formalizing the effect of diversity on the expectation of convex functions. The common notion of comparative dispersion of distributions in Probability [23] and Economics

[24] is that G is more dispersed than F if for some random variables X and Y with marginal distributions F and G respectively, $E(Y|X) = X$ almost surely, i.e., if Y is obtained from X by the addition of noise. Jensen’s inequality yields $E(\psi(Y)) = E(E(\psi(Y)|X)) \geq E(\psi(X))$ for any convex function ψ . In particular, for any partitioning of the set of neurons, the network’s final similarity decreases if each neuron’s belief is replaced by the average belief of its partition set.

We began this section with the remark that with history-independent dynamics, after every iteration the neurons’ beliefs are collapsed into two values; the belief of each neuron with prior preference for $+1$ is replaced by the average belief of its class, and similarly regarding those neurons with prior preference for -1 . As we have shown, this implies that for sparse networks (i.e., those for which the sign of different neurons continues to be nearly independent beyond the first iteration) a history-dependent second iteration performs better than a memoryless one.

If beliefs are dependent or if - as we assume throughout - only a fixed fraction of the neurons are allowed to be active in a given iteration, the firing mechanisms should be modified accordingly, by letting activation depend not only on current beliefs but also on whether the particular neuron has been silent or active in the previous iterations. In this paper we allow this activity history to affect the assessment of the final belief $\lambda_i^{(2)}$ but not the (otherwise general) interim signal $S(\lambda_i^{(1)})$. In the same vein, we adopt the common premise that the initial state is a *homogeneously* distorted version of the true memory, in spite of the fact that heterogeneity of prior beliefs induces higher performance. Other patterns of initial states should be studied, such as the opposite extreme where $\epsilon = 0$ for most neurons and $\epsilon = 1$ for the remaining few.

In the next section we develop a framework for quantifying the performance of a history-dependent second iteration in ANNs.

4 Two-iterations optimization

The field observed by neuron i in the second iteration is

$$f_i^{(2)} = \frac{1}{n_2} \sum_j I_{ij} W^*_{ij} I_j^{(2)} S(\lambda_j^{(1)}) . \quad (18)$$

For mathematical convenience, we will relate signals and activation rules to normalized generalized fields rather than to beliefs. We let

$$h(x) = S\left(\frac{1}{1 + e^{-2cx}}\right) , \quad p(x) = C\left(\max\left(\frac{1}{1 + e^{-2cx}}, 1 - \frac{1}{1 + e^{-2cx}}\right)\right) \quad (19)$$

for $c = \epsilon/\sqrt{\alpha_1}$. The signal function h is assumed to be odd, and the activation function p , even. (18) may be then rewritten as

$$f_i^{(2)} = \frac{1}{n_2} \sum_j I_{ij} W^*_{ij} I_j^{(2)} h\left(\frac{f_j^{(1)}}{\sqrt{\alpha_1}} + \gamma(\epsilon)\sqrt{\alpha_1} X_j\right) . \quad (20)$$

where $I_j^{(2)} = 1$ with probability $p\left(\frac{f_j^{(1)}}{\sqrt{\alpha_1}} + \gamma(\epsilon)\sqrt{\alpha_1} X_j\right)$, independently of everything else.

In order to evaluate the belief $\lambda_i^{(2)}$, we need the conditional distribution of $f_i^{(2)}$ given $X_i, I_i^{(1)}$ and $f_i^{(1)}$, for $\xi_i = -1$ or $\xi_i = +1$. We adopt the working assumption that the pair of random variables $(f_i^{(1)}, f_i^{(2)})$ has a bivariate normal distribution given $\xi_i, I_i^{(1)}$ and X_i , with $\xi_i, I_i^{(1)}$ and X_i affecting means but not variances or correlations. Under this working assumption, $f_i^{(2)}$ is conditionally normal given $(\xi_i, I_i^{(1)}, X_i, f_i^{(1)})$, with constant variance and a mean which we will identify. This working assumption allows us to model performance via the following well known regression model.

Lemma 2

If two random variables U and V with finite variances are such that $E(V|U)$ is a linear function of U and $Var(V|U)$ is constant, then

$$E(V|U) = E(V) + \frac{Cov(U, V)}{Var(U)}(U - E(U)) \quad (21)$$

and

$$Var(V|U) = Var(V) - \frac{(Cov(U, V))^2}{Var(U)} . \quad (22)$$

Jointly normally distributed random variables U and V satisfy the assumptions of Lemma 2. Furthermore, the conditional distribution of V given U is itself normal. Letting

$U = f_i^{(1)}$ and $V = f_i^{(2)}$, this lemma will enable us to model posterior probabilities following a second iteration. As $E(U) = E(f_i^{(1)}|\xi_i, X_i, I_i^{(1)})$ and $Var(U) = Var(f_i^{(1)}|\xi_i, X_i, I_i^{(1)})$ have already been calculated (characterizing the first iteration), it remains to identify the mean $E(V) = E(f_i^{(2)}|\xi_i, X_i, I_i^{(1)})$, the variance $Var(V) = Var(f_i^{(2)}|\xi_i, X_i, I_i^{(1)})$ and the covariance term $Cov(U, V) = Cov(f_i^{(1)}, f_i^{(2)}|\xi_i, X_i, I_i^{(1)})$. Lemma 2 permits us to express

$$E(f_i^{(2)}|\xi_i, X_i, I_i^{(1)}, f_i^{(1)}) = E(f_i^{(2)}|\xi_i, X_i, I_i^{(1)}) + a(f_i^{(1)} - \epsilon\xi_i) \quad (23)$$

$$Var(f_i^{(2)}|\xi_i, X_i, I_i^{(1)}, f_i^{(1)}) = \tau^2 \quad (24)$$

for some constants a and τ^2 . We will prove in the sequel that for some constants ϵ^* and b to be identified together with a and τ^2 ,

$$E(f_i^{(2)}|\xi_i, X_i, I_i^{(1)}) = \epsilon^*\xi_i + bX_iI_i^{(1)} \quad (25)$$

We are now in a position to apply Lemma 1(ii) with (see (7) and (13))

$$x = \epsilon(f_i^{(1)}/\alpha_1 + \gamma(\epsilon)X_i)$$

and (see (8), (23) and (24))

$$Y = f_i^{(2)} - bX_iI_i^{(1)} - af_i^{(1)}$$

$$\mu = \epsilon^* - a\epsilon$$

$$\sigma^2 = \tau^2$$

to obtain from (9) the posterior probability

$$\begin{aligned} \lambda_i^{(2)} &= P(\xi_i = 1|X_i, I_i^{(1)}, f_i^{(1)}, f_i^{(2)}) = \\ &= \frac{1}{1 + \exp\{-2 \left[\epsilon \left(f_i^{(1)}/\alpha_1 + \gamma(\epsilon)X_i \right) + \frac{\epsilon^* - a\epsilon}{\tau^2} \left(f_i^{(2)} - bX_iI_i^{(1)} - af_i^{(1)} \right) \right] \}} \\ &= \frac{1}{1 + \exp\{-2 \left[\left(\epsilon\gamma(\epsilon) - \frac{b(\epsilon^* - a\epsilon)}{\tau^2} I_i^{(1)} \right) X_i + \left(\frac{\epsilon}{\alpha_1} - \frac{a(\epsilon^* - a\epsilon)}{\tau^2} \right) f_i^{(1)} + \frac{\epsilon^* - a\epsilon}{\tau^2} f_i^{(2)} \right] \}} \end{aligned} \quad (26)$$

which is, as announced, the sigmoidal function evaluated at some generalized field.

Expression (26) shows that the correct definition of a final state $X_i^{(2)}$, as the most likely value among $+1$ or -1 , is

$$X_i^{(2)} = Sgn \left[\left(\epsilon\gamma(\epsilon) - \frac{b(\epsilon^* - a\epsilon)}{\tau^2} I_i^{(1)} \right) X_i + \left(\frac{\epsilon}{\alpha_1} - \frac{a(\epsilon^* - a\epsilon)}{\tau^2} \right) f_i^{(1)} + \frac{\epsilon^* - a\epsilon}{\tau^2} f_i^{(2)} \right] \quad (27)$$

By our working normality assumption, given $X_i, I_i^{(1)}$ and ξ_i , the random vector $(f_i^{(1)}, f_i^{(2)})$ has a bivariate normal distribution. Hence, any linear combination of the two coordinates, such as $W = \left(\frac{\epsilon}{\alpha_1} - \frac{a(\epsilon^* - a\epsilon)}{\tau^2}\right) f_i^{(1)} + \frac{\epsilon^* - a\epsilon}{\tau^2} f_i^{(2)}$, is normally distributed. As can be seen from (27), the two-iterations performance is measured by the final similarity

$$P(X_i^{(2)} = \xi_i) = P\left(W > \left(\frac{b(\epsilon^* - a\epsilon)}{\tau^2} I_i^{(1)} - \epsilon\gamma(\epsilon)\right) X_i \mid \xi_i = +1\right) \quad (28)$$

To get an explicit expression for the two-iterations performance, it remains to identify the conditional mean and variance of W given $\xi_i, I_i^{(1)}$ and X_i . Given $\xi_i = 1, I_i^{(1)}$ and X_i , W is normally distributed with

$$E(W \mid X_i, I_i^{(1)}, \xi_i = 1) = \left(\frac{\epsilon}{\alpha_1} - a\frac{\epsilon^* - a\epsilon}{\tau^2}\right) \epsilon + \frac{\epsilon^* - a\epsilon}{\tau^2} (\epsilon^* + bX_i I_i^{(1)}) \quad (29)$$

and variance

$$\begin{aligned} \text{Var}(W \mid X_i, I_i^{(1)}, \xi_i = 1) &= \left(\frac{\epsilon}{\alpha_1} - a\frac{\epsilon^* - a\epsilon}{\tau^2}\right)^2 \text{Var}(f_i^{(1)}) + \left(\frac{\epsilon^* - a\epsilon}{\tau^2}\right)^2 \text{Var}(f_i^{(2)}) + \\ &2 \left(\frac{\epsilon}{\alpha_1} - a\frac{\epsilon^* - a\epsilon}{\tau^2}\right) \frac{\epsilon^* - a\epsilon}{\tau^2} \text{Cov}(f_i^{(1)}, f_i^{(2)}) \end{aligned} \quad (30)$$

where, for clarity of exposition, we have omitted the conditioning symbols. By (21) and (23), we can identify $\text{Cov}(f_i^{(1)}, f_i^{(2)})/\text{Var}(f_i^{(1)})$ with a , and by (22) we can identify $\text{Var}(f_i^{(2)})$ with $\tau^2 + (\text{Cov}(f_i^{(1)}, f_i^{(2)}))^2/\text{Var}(f_i^{(1)}) = \tau^2 + \alpha_1 a^2$. We thus get

$$\text{Var}(W \mid X_i, I_i^{(1)}, \xi_i) = \epsilon^2 \left[\frac{1}{\alpha_1} + \left(\frac{\epsilon^*/\epsilon - a}{\tau}\right)^2 \right] = \epsilon^2/\alpha^* \quad (31)$$

where

$$\alpha^* = \frac{m}{n^*} = \frac{m}{n_1 + m \left(\frac{\epsilon^*/\epsilon - a}{\tau}\right)^2}. \quad (32)$$

From (29) and (31),

$$P(X_i^{(2)} = \xi_i \mid X_i, I_i^{(1)}, \xi_i) = P\left(W > \left(\frac{b(\epsilon^* - a\epsilon)}{\tau^2} I_i^{(1)} - \epsilon\gamma(\epsilon)\right) X_i \mid \xi_i = +1\right) = \quad (33)$$

$$P\left(\frac{W - E(W)}{\sigma(W)} > -\left[\frac{\epsilon}{\sqrt{\alpha^*}} + \gamma(\epsilon) X_i \sqrt{\alpha^*}\right]\right) = \Phi\left(\frac{\epsilon}{\sqrt{\alpha^*}} + \gamma(\epsilon) X_i \sqrt{\alpha^*}\right)$$

from which

$$P(X_i^{(2)} = \xi_i \mid \xi_i) = \frac{1 + \epsilon}{2} \Phi\left(\frac{\epsilon}{\sqrt{\alpha^*}} + \gamma(\epsilon) \sqrt{\alpha^*}\right) + \frac{1 - \epsilon}{2} \Phi\left(\frac{\epsilon}{\sqrt{\alpha^*}} - \gamma(\epsilon) \sqrt{\alpha^*}\right) = \quad (34)$$

$$Q(\epsilon, \alpha^*)$$

where the one-iteration performance function Q is defined by (14).

We see that the performance is conveniently expressed as the single-iteration optimal performance, had this iteration involved n^* rather than n_1 sampled neurons. This formula yields a numerical and analytical tool to assess the network's performance with different signal functions, activation proportions and architectures. We see that $\alpha^* < \alpha_1$ (unless $\epsilon^* = a\epsilon$, in which case, by (26), the second iteration is uninformative). By the monotonicity of Q in α proved earlier, two iterations strictly outperform one iteration, just as the latter strictly improves over the initial similarity.

It is worth noticing that the final decision (27) may be expressed as a comparison of the second iteration input field $f_i^{(2)}$ with the decision threshold

$$T_i^{(2)} = \left(a - \frac{\tau^2}{\alpha_1(\epsilon^*/\epsilon - a)} \right) f_i^{(1)} + \left(bI_i^{(1)} - \frac{\tau^2\gamma(\epsilon)}{(\epsilon^*/\epsilon - a)} \right) X_i \quad (35)$$

defined by previous history. If $\epsilon^* > a\epsilon$ then the final state is $X_i^{(2)} = \text{sgn}(f_i^{(2)} - T_i^{(2)})$ while if $\epsilon^* < a\epsilon$, each neuron *reverses* this decision. We shall see that $\epsilon^* - a\epsilon$ is positive under biologically relevant connectivity, but negative values will be illustrated as well, under heavy connectivity.

The identification of the parameters a, b, ϵ^* and τ is technically difficult. We will present here their formulas for the biologically motivated two-valued signal function $h(x) = \text{Sgn}(x)$, and illustrate these formulas for random and censored activation rules. Proofs and generalizations to arbitrary signal and activation functions are given in the Appendix.

Let the *integrated signal function* and the *integrated absolute signal function* be, respectively,

$$\Psi(x) = \int p(x+z)h(x+z)\phi(z)dz \quad (36)$$

and

$$\Psi_a(x) = \int p(x+z)|h(x+z)|\phi(z)dz \quad (37)$$

where $\phi(x) = \exp\{-x^2/2\}/\sqrt{2\pi}$ is the standard normal density function. From now on, let $h(x) = \text{Sgn}(x)$. In analogy to the computation in expression (14), expression (37) provides the number n_2 of updating neurons per neuron, or equivalently the number L_2 of active neurons, by

$$\bar{\Psi}_a \equiv \frac{n_2}{K} = \frac{L_2}{N} = \frac{1+\epsilon}{2}\Psi_a\left(\frac{\epsilon}{\sqrt{\alpha_1}} + \sqrt{\alpha_1}\gamma(\epsilon)\right) + \frac{1-\epsilon}{2}\Psi_a\left(\frac{\epsilon}{\sqrt{\alpha_1}} - \sqrt{\alpha_1}\gamma(\epsilon)\right) \quad (38)$$

and expression (36) provides the *drift* parameter

$$\epsilon^* = \frac{K}{m} \alpha_2 \bar{\Psi}_+ = \frac{\bar{\Psi}_+}{\bar{\Psi}_a} \quad (39)$$

where

$$\bar{\Psi}_+ = \frac{1+\epsilon}{2} \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \sqrt{\alpha_1} \gamma(\epsilon) \right) + \frac{1-\epsilon}{2} \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} - \sqrt{\alpha_1} \gamma(\epsilon) \right). \quad (40)$$

The three parameters a, b and τ^2 are conveniently expressed in terms of

$$\bar{\Psi}' = \frac{1+\epsilon}{2} \Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \sqrt{\alpha_1} \gamma(\epsilon) \right) + \frac{1-\epsilon}{2} \Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}} - \sqrt{\alpha_1} \gamma(\epsilon) \right) \quad (41)$$

$$\bar{\Psi}_- = \frac{1+\epsilon}{2} \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \sqrt{\alpha_1} \gamma(\epsilon) \right) - \frac{1-\epsilon}{2} \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} - \sqrt{\alpha_1} \gamma(\epsilon) \right) \quad (42)$$

and

$$r = P(I_{ij} = 1 | I_{ji} = 1) \quad (43)$$

as

$$a = \frac{m}{K} \frac{1}{\alpha_1} \frac{\bar{\Psi}_-}{\bar{\Psi}_a} + \frac{K}{N} \frac{1}{\sqrt{\alpha_1}} \frac{\bar{\Psi}'}{\bar{\Psi}_a} \quad (44)$$

$$b = \sqrt{\alpha_1} r \frac{\bar{\Psi}'}{\bar{\Psi}_a} \quad (45)$$

$$\tau^2 = \alpha_2 - \frac{1}{\alpha_1} \left(\frac{m}{K} \right)^2 \left(\frac{\bar{\Psi}_-}{\bar{\Psi}_a} \right)^2 + \frac{K}{N} \left(1 - \frac{K}{N} \right) \left(\frac{\bar{\Psi}'}{\bar{\Psi}_a} \right)^2 \quad (46)$$

It is interesting to realize that performance is independent of the *reflexivity* parameter r . This parameter only affects dynamics (see expression 27), by calibrating the field of an initially active neuron - filtering out the information it gave the other neurons from the information received from them.

We easily see that in the sparse limit arrived at by fixing α_1 and α_2 and letting both K/m and N/K go to infinity, we obtain that $a = 0$, $\tau^2 = \alpha_2$, and (see (32)) $n^* = n_1 + n_2(\epsilon^*/\epsilon)^2$. As will be shown, under random activation, $\epsilon^* = \epsilon'$. This implies that n^* strictly exceeds $n_1 + n_2$ because, as we saw at the end of section 2, ϵ' strictly exceeds ϵ . Since the two-iteration performance is equal to the one-iteration performance had this single iteration involved n^* updating neurons per neuron, this means that in the sparse limit it is always better to replace an iteration by two smaller ones. This suggests that Bayesian updating dynamics should be essentially asynchronous.

In reference to the above mentioned limiting case, we saw in the last section that the two-iterations performance $Q \left(\epsilon, \frac{1}{\frac{1}{\alpha_1} + \frac{1}{\alpha_2} \left(\frac{Q(\epsilon, \alpha_1)}{\epsilon} \right)^2} \right)$ is superior to the performance $Q(2Q(\epsilon, \alpha_1) - 1, \alpha_2)$

of two *independent* optimal single iterations. Section 7, presenting some numerical examples of our findings, illustrates this superiority. In section 6, we present the formulas for the parameters $\bar{\Psi}_a$, $\bar{\Psi}_+$, $\bar{\Psi}'$, $\bar{\Psi}_-$, ϵ^* , a and τ^2 , as given by expressions (38) through (46), for each of the activation functions we examine. But first, let us present a heuristic approach for obtaining high performance by efficient activation and signalling.

5 Heuristics on activation and signalling

By (32) and (34), performance is mostly determined by the magnitude of $(\epsilon^* - a\epsilon)^2$. By (38), (39) and (44),

$$(\epsilon^* - a\epsilon)\bar{\Psi}_a = \bar{\Psi}_+ - \frac{m\epsilon}{K\alpha_1}\bar{\Psi}_- - \frac{\epsilon K}{N\sqrt{\alpha_1}}\bar{\Psi}' . \quad (47)$$

Using the integral representations (36) and (37) for Ψ and Ψ_a (and assuming as we did that $|h(x)| \equiv 1$), it is easy to see that

$$(\epsilon^* - a\epsilon)\bar{\Psi}_a = \int_0^\infty p(x)h(x)\phi_1(x)dx \quad (48)$$

and

$$\bar{\Psi}_a = \int_0^\infty p(x)\phi_0(x)dx \quad (49)$$

where ϕ_1 and ϕ_0 are some specific linear combinations of Gaussian densities and their derivatives.

For a given activity level $\bar{\Psi}_a = n_2/K$, high performance is achieved by maximizing over p and possibly over h the absolute value of expression (48) keeping (49) fixed. In complete analogy to Hypothesis Testing in Statistics, where $\bar{\Psi}_a$ takes the role of *level of significance* and $(\epsilon^* - a\epsilon)\bar{\Psi}_a$ the role of *power*, $p(x)$ should be 1 or 0 (activate the neuron or don't) depending on whether the field value x is such that the *likelihood ratio* $h(x)\phi_1(x)/\phi_0(x)$ is above or below a given threshold, determined by (49). Omitting details, the ratio $R(x) = \phi_1(x)/\phi_0(x)$ looks as in figure 1, and converges to $-\infty$ as $x \rightarrow \infty$.

FIGURE 1

We see that there are three reasonable ways to make the ratio $h(x)\phi_1(x)/\phi_0(x)$ large: if we assume that $h(x) = Sgn(x)$, i.e., $h(x) \equiv 1$ since we work on the interval $(0, \infty)$, we can take a negative threshold such as t_1 in figure 1 and activate all neurons with generalized field exceeding β_3 (tail-censoring), or a positive threshold such as t_2 and activate all neurons with field value between β_1 and β_2 (interval-censoring). Figure 1 suggests that tail-censoring

should involve a reversal of decision, as suggested previously following expression (35). Equivalently, active neurons could signal the sign *opposite* the one they believe in, and neuron i would then maintain the decision as $+1$ or -1 depending on whether $f_i^{(2)}$ is *above* or *below* $T_i^{(2)}$.

This inverted signalling rule paves the way to a third reasonable way to make $h(x)R(x)$ large, the *hybrid signalling-censoring rule*: Activate all neurons with absolute field value between β_1 and β_2 , or beyond β_3 . The first group should signal their preferred sign, while those in the second group should signal the sign opposite to the one they so strongly believe in !

In the next section we will develop explicit formulas for $\bar{\Psi}_+$, $\bar{\Psi}_-$, $\bar{\Psi}'$ and $\bar{\Psi}_a$ under random activation, tail-censoring, interval-censoring and the hybrid signalling-activation rule. The following section illustrates numerically their performance.

6 Some specific activation rules

For each rule, we will present explicit formulas for the parameters $\bar{\Psi}_+$, $\bar{\Psi}_-$, $\bar{\Psi}'$ and $\bar{\Psi}_a$, which are the ingredients of the evaluation of the parameters a , b , ϵ^* and τ^2 determining dynamics and performance.

6.1 Random activation

In the case of random activation, where $p(x) \equiv \frac{n_2}{K}$, we get

$$\Psi(x) = \frac{n_2}{K} \left[\int_{-x}^{\infty} \phi(z) dz - \int_{-\infty}^{-x} \phi(z) dz \right] = \frac{n_2}{K} [2\Phi(x) - 1] \quad (50)$$

$$\Psi_a(x) = \frac{n_2}{K} \int \phi(z) dz = \frac{n_2}{K} \quad (51)$$

$$\Psi'(x) = 2 \frac{n_2}{K} \phi(x) \quad (52)$$

$$\frac{\bar{\Psi}'}{\bar{\Psi}_a} = (1 + \epsilon) \phi \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \sqrt{\alpha_1} \gamma(\epsilon) \right) + (1 - \epsilon) \phi \left(\frac{\epsilon}{\sqrt{\alpha_1}} - \sqrt{\alpha_1} \gamma(\epsilon) \right) \quad (53)$$

$$\frac{\bar{\Psi}_+}{\bar{\Psi}_a} = (1 + \epsilon) \Phi \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \sqrt{\alpha_1} \gamma(\epsilon) \right) + (1 - \epsilon) \Phi \left(\frac{\epsilon}{\sqrt{\alpha_1}} - \sqrt{\alpha_1} \gamma(\epsilon) \right) - 1 \quad (54)$$

$$\frac{\bar{\Psi}_-}{\bar{\Psi}_a} = (1 + \epsilon) \Phi \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \sqrt{\alpha_1} \gamma(\epsilon) \right) - (1 - \epsilon) \Phi \left(\frac{\epsilon}{\sqrt{\alpha_1}} - \sqrt{\alpha_1} \gamma(\epsilon) \right) - \epsilon. \quad (55)$$

We see (by (14)) that in this case $\epsilon^* = \epsilon' = 2Q(\epsilon, \alpha_1) - 1$.

6.2 Tail-censoring

In the case of tail-censoring, ϵ^* is even bigger, yielding (at least in the sparse limit) better performance. In this case, the threshold β for which a proportion $\frac{n_2}{K} = \frac{L_2}{N}$ of the neurons are active, i.e., have $|f_i^{(1)} + \alpha_1 \gamma(\epsilon) X_i| > \beta$, should be identified. The integrated signal function is given by

$$\begin{aligned} \Psi(x) &= \int_{-x+\beta/\sqrt{\alpha_1}}^{\infty} \phi(z) dz - \int_{-\infty}^{-x-\beta/\sqrt{\alpha_1}} \phi(z) dz = \\ &\Phi(x - \beta/\sqrt{\alpha_1}) + \Phi(x + \beta/\sqrt{\alpha_1}) - 1 . \end{aligned} \quad (56)$$

Its derivative is

$$\Psi'(x) = \phi(x - \beta/\sqrt{\alpha_1}) + \phi(x + \beta/\sqrt{\alpha_1}) \quad (57)$$

and the integrated absolute signal function is

$$\Psi_a(x) = \Phi(x - \beta/\sqrt{\alpha_1}) - \Phi(x + \beta/\sqrt{\alpha_1}) + 1 . \quad (58)$$

Let

$$y_{ij} = \frac{\epsilon}{\sqrt{\alpha_1}} + \left((-1)^i \beta/\sqrt{\alpha_1} + (-1)^j \gamma(\epsilon) \sqrt{\alpha_1} \right) , \quad i, j \in \{0, 1\} \quad (59)$$

Then

$$\bar{\Psi}' = \frac{1+\epsilon}{2} [\phi(y_{00}) + \phi(y_{10})] + \frac{1-\epsilon}{2} [\phi(y_{01}) + \phi(y_{11})] \quad (60)$$

$$\bar{\Psi}_+ = \frac{1+\epsilon}{2} [\Phi(y_{00}) + \Phi(y_{10}) - 1] + \frac{1-\epsilon}{2} [\Phi(y_{01}) + \Phi(y_{11}) - 1] \quad (61)$$

$$\bar{\Psi}_- = \frac{1+\epsilon}{2} [\Phi(y_{00}) + \Phi(y_{10}) - 1] - \frac{1-\epsilon}{2} [\Phi(y_{01}) + \Phi(y_{11}) - 1] \quad (62)$$

and

$$\bar{\Psi}_a = \frac{1+\epsilon}{2} [\Phi(y_{10}) - \Phi(y_{00}) + 1] + \frac{1-\epsilon}{2} [\Phi(y_{11}) - \Phi(y_{01}) + 1] . \quad (63)$$

β should be chosen as the solution to the equation $\bar{\Psi}_a = \frac{n_2}{K}$.

6.3 Interval censoring

The analysis of tail-censoring permits us to evaluate rather easily the performance of other censoring rules, such as *interval censoring*, which activates all neurons with absolute post-synaptic potential *between* two threshold values β_1 and β_2 with $0 \leq \beta_1 < \beta_2$. Stressing dependence on the general threshold β in (60), (61), (62) and (63) by letting these parameters be denoted by $(\bar{\Psi}')_{\beta}$, $(\bar{\Psi}_+)_{\beta}$, $(\bar{\Psi}_-)_{\beta}$ and $(\bar{\Psi}_a)_{\beta}$ respectively, the corresponding parameters for the interval-censoring rule are given simply by

$$\bar{\Psi}' = (\bar{\Psi}')_{\beta_1} - (\bar{\Psi}')_{\beta_2} \quad (64)$$

$$\bar{\Psi}_+ = (\bar{\Psi}_+)_{\beta_1} - (\bar{\Psi}_+)_{\beta_2} \quad (65)$$

$$\bar{\Psi}_- = (\bar{\Psi}_-)_{\beta_1} - (\bar{\Psi}_-)_{\beta_2} \quad (66)$$

and, respectively,

$$\bar{\Psi}_a = (\bar{\Psi}_a)_{\beta_1} - (\bar{\Psi}_a)_{\beta_2} \quad (67)$$

Thus, n_2 assumes the value $K\bar{\Psi}_a$ and performance is given by (34) in the usual manner.

It should be clear that for random activation and tail-censoring the total number L_2 of active neurons determines the activation rule completely, but there may be many intervals (β_1, β_2) , such as (β, ∞) itself, for which the expected number of active neurons is L_2 . We will only illustrate in the next section the interval achieving the highest performance.

6.4 Hybrid signalling-censoring

The analysis of tail-censoring permits us to evaluate the performance of this rule with the same ease as it facilitated the analysis of interval-censoring. The corresponding parameters in the hybrid case are given by

$$\bar{\Psi}' = (\bar{\Psi}')_{\beta_1} - (\bar{\Psi}')_{\beta_2} - (\bar{\Psi}')_{\beta_3} \quad (68)$$

$$\bar{\Psi}_+ = (\bar{\Psi}_+)_{\beta_1} - (\bar{\Psi}_+)_{\beta_2} - (\bar{\Psi}_+)_{\beta_3} \quad (69)$$

$$\bar{\Psi}_- = (\bar{\Psi}_-)_{\beta_1} - (\bar{\Psi}_-)_{\beta_2} - (\bar{\Psi}_-)_{\beta_3} \quad (70)$$

and, respectively,

$$\bar{\Psi}_a = (\bar{\Psi}_a)_{\beta_1} - (\bar{\Psi}_a)_{\beta_2} + (\bar{\Psi}_a)_{\beta_3} . \quad (71)$$

The heuristic analysis of section 5 ignored the parameter τ^2 and was meant only as a qualitative justification for studying tail, interval and hybrid rules. The parameters β_1 and β_2 for interval censoring and β_1, β_2 and β_3 for the hybrid rule to be illustrated in the next section are obtained by numerically maximizing the two-iteration similarity function (34) over these β 's.

7 Numerical results

As is evident from tables 1,2,3,4, 5,6 and 7, our theoretical performance predictions show good correspondence with simulation results, already at fairly small-scale networks. The superiority of history-dependent dynamics is apparent.

All tables report the similarity after two iterations, starting from initial similarity 0.75, except for the second entry of the experimental Hopfield-zero diagonal dynamics which reports on the final similarity achieved at convergence. All experimental results are averages over 100 trials. The results at different initial similarities are qualitatively similar. When the best interval chosen by interval-censoring coalesces with the interval of tail-censoring, their corresponding performance is given in the same entry, and similarly, for the case when interval-censoring and hybrid signalling-censoring coalesce.

Due to the correlation that evolves in the second iteration, the two-iterations performance of the zero-diagonal Hopfield network is worse than the theoretical two ‘independent iterations’ prediction, and approaches the latter in the sparse limit. As for history-dependent dynamics, table 3 shows that for $m/n_t = 0.25$, the similarity after two independent optimal iterations is 0.96, which is outperformed by history-dependent dynamics (see also table 7).

TABLES 1 THROUGH 7

Figures 2 and 3 illustrate the theoretical two-iterations performance of large, low-activity ‘cortical-like’ networks, as a function of connectivity. We see that interval-censoring can maintain superior performance throughout the connectivity range. The performance of tail-censoring is very sensitive to connectivity, almost achieving the performance of interval censoring at a narrow low-connectivity range, and becoming optimal only at very high connectivity. There is a mid-range value of K for which tail-censoring does not improve on the one-iteration performance.

FIGURES 2 AND 3

Performance is measured (see (32)) by $((\epsilon^*/\epsilon) - a)^2/\tau^2$. Disregarding the dependence of $\frac{\bar{\Psi}_-}{\bar{\Psi}_a}$ and $\frac{\bar{\Psi}'}{\bar{\Psi}'_a}$ on K and assuming that $\bar{\Psi}' > 0$, a is minimal (see (44)) when

$$K = \sqrt{Nm} \sqrt{\frac{\bar{\Psi}_-}{\sqrt{\alpha_1} \bar{\Psi}'}} . \quad (72)$$

Thus, if τ^2 is not too sensitive to K in this range, performance should be high if K is of the order of magnitude of the geometric mean of the network size N and the memory capacity m . This is indeed the range where random activation and tail-censoring perform well. The reason for the high performance of tail-censoring in this range is that it is essentially the same as interval-censoring, in the sense that under the optimal pair (β_1, β_2) , the probability that the field will exceed β_2 is so small, that it is immaterial whether we activate neurons in the *interval* (β_1, β_2) or in the *tail* (β_1, ∞) . The fact that tail-censoring is in this case ‘false’ interval-censoring implies that the neuron’s final decision should indeed be aligned with the second iteration input field, rather than its inverse. These two modes of decision were fully corroborated by our numerical simulations.

Figure 4 illustrates the performance of a columnar-like ANN at geometric-mean connectivity. The architecture is sparse enough so that all three history-dependent dynamics outperform the hypothetical optimal two-independent-iterations network. As is evident, as few as two low-firing iterations may already lead to the retrieval of the correct memory, showing that such ANNs might perform adequately within the short reaction times reported in the literature. As is evident, the superiority of history-dependent dynamics is indeed maintained along a wide range of initial similarities.

FIGURE 4

8 Discussion

In this paper we have presented a computational framework enabling a detailed description of the performance of ANNs in the first two iterations. We have stressed the development of a general Bayesian framework that lays the ground for the analysis of history-dependent dynamics, and have used this analysis to identify efficient activation and signalling strategies.

The neocortex is traditionally described as being composed of modular columns of locally interconnected circuitry [25, 26], whose size is of the order of $N = 10^5$ neurons [9]. As a cortical neuron should receive the concomitant firing of about 200 – 300 neurons in order to be activated [6], we set $n = 200$ as the average number of neurons which update the state of each neuron. The optimal connectivity we have found for biologically plausible activation, a few thousand synapses per neuron, is of the same order of magnitude as actual cortical connectivity [8]. It is interesting to note that similar network (N) and connectivity (K)

sizes have been found also in the hippocampal CA3 network of the rat, its main memory storage network [6]. The actual number nN/K of neurons firing in every iteration is about 5000, which is in close correspondence with the evidence suggesting that about 4% of the neurons in a module fire at any given moment [9].

In cortical-like conditions, censoring is considerably superior to random updating. This naturally raises the possibility that censoring should take place already in the first iteration. This means essentially that the input pattern should not be applied as the conventional uniformly distorted version of the correct memory, but rather as the almost absolutely correct pattern on some small subset of the neurons, and random on the others. We plan to present in forthcoming publications the analysis under this type of initial input pattern, as well as the transition from symmetric ± 1 to the more biological, asymmetric 0 – 1 formulation, where activation and signalling are one entity.

Appendix A: Evaluating a, b and τ^2

Let us first summarize some expectations of products of weights W_{ij} , needed in the sequel.

Lemma 3

(i)

$$E(W_{12}^{odd}W_{13}^{even}W_{23}^{even}) = 0 \quad (73)$$

$$E(W_{12}^2) = m, \quad E(W_{12}^4) = 3m^2 - 2m \approx 3m^2 \quad (74)$$

$$E(W_{12}W_{13}|W_{23}) = W_{23} \text{ a.s.} \quad (75)$$

$$E(W_{12}^2W_{13}^2) = m^2 \quad (76)$$

$$E(W_{12}^2W_{13}^2W_{23}^2) = m^3 + 4m^2 - 4m \approx m^3 \quad (77)$$

$$E(W_{12}W_{13}^2W_{23}^3) = 0 \quad (78)$$

(ii) For large N , the conditional correlation coefficient between $\sum_{j=4}^N W_{2j}I_{2j}I_j^{(1)}X_j$ and $\sum_{j=4}^N W_{3j}I_{3j}I_j^{(1)}X_j$ given all memory entries in neurons 1, 2 and 3 and given that there are synapses from neurons 2 and 3 to neuron 1, becomes

$$P(I_{34} = 1 | I_{12}I_{13}I_{24} = 1) \frac{W_{23}}{m} = \frac{K}{N} \frac{W_{23}}{m} \quad (79)$$

Proof of (i)

(73) follows from symmetry considerations. Since $W_{12} = \sum_{\mu} \xi_1^{\mu} \xi_2^{\mu}$ is the sum of m independent variables J_{μ} with equally likely values of ± 1 , then

$$E(W_{12}^2) = Var \sum J_{\nu} = m Var(J_{\nu}) = m$$

$$E(W_{12}^4) = mE(J_{\nu}^4) + \binom{m}{2} \binom{4}{2} E(J_{\nu}^2)E(J_{\nu}^2) = 3m^2 - 2m$$

$$E(W_{12}W_{13}^2W_{23}^3) = E \left(\sum_{\mu} \xi_1^{\mu} \xi_2^{\mu} \left(\sum_{\mu'} \xi_1^{\mu'} \xi_3^{\mu'} \right)^2 \left(\sum_{\mu''} \xi_2^{\mu''} \xi_3^{\mu''} \right)^3 \right)$$

In every term, some ξ_1^{μ} appears with odd power, so the expectation is zero.

$$E(W_{12}^2W_{13}^2W_{23}^2) = E \left(\left(\sum_{\mu} \xi_1^{\mu} \xi_2^{\mu} \right)^2 \left(\sum_{\mu'} \xi_1^{\mu'} \xi_3^{\mu'} \right)^2 \left(\sum_{\mu''} \xi_2^{\mu''} \xi_3^{\mu''} \right)^2 \right)$$

Consider any μ , any μ' and any μ'' , and take each twice in every sum. These terms $(\xi_1^{\mu} \xi_2^{\mu})^2 (\xi_1^{\mu'} \xi_3^{\mu'})^2 (\xi_2^{\mu''} \xi_3^{\mu''})^2$ contribute m^3 to the sum. Let us study other terms. If

we pick two different values of μ in the first sum, we must repeat them in the second and third sums to get a non-zero contribution. We obtain in this way $\binom{m}{2}2^3 = 4m^2 - 4m$ terms, giving a total expected value $m^3 + 4m^2 - 4m$. We have shown (73), (74) and (77). (76) and (78) are verified similarly. As for (75),

$$\begin{aligned} E(W_{12}W_{13} | \text{all memory entries of neurons 2 and 3}) &= \\ E\left(\sum_{\mu} \xi_1^{\mu} \xi_2^{\mu} \sum_{\mu'} \xi_1^{\mu'} \xi_3^{\mu'} | \text{data}\right) &= E\left(\sum_{\mu} \xi_1^{\mu} \xi_2^{\mu} \xi_1^{\mu} \xi_3^{\mu} | \text{data}\right) = \\ E\left(\sum_{\mu} \xi_2^{\mu} \xi_3^{\mu} | \text{data}\right) &= E(W_{23} | \text{data}) = W_{23}. \end{aligned}$$

Proof of (ii)

$$\begin{aligned} \text{Cov}\left(\sum_{j=4}^N W_{2j} I_{2j} I_j^{(1)} X_j, \sum_{j=4}^N W_{3j} I_{3j} I_j^{(1)} X_j | \text{data}\right) &= \\ (N-3) \text{Cov}(W_{24} I_{24} I_4^{(1)} X_4, W_{34} I_{34} I_4^{(1)} X_4 | \text{data}) &= \\ (N-3) E(I_4^{(1)}) E(I_{24} I_{34} | I_{12} I_{23} = 1) \text{Cov}(W_{24}, W_{34} | \text{data}) &= \\ \frac{N n_1}{K} E(I_{24} I_{34} | I_{12} I_{13} = 1) W_{23} & \end{aligned}$$

by (75). It is clear that $E(I_{24} I_{34} | I_{12} I_{13} = 1) = \frac{K}{N} P(I_{34} = 1 | I_{12} I_{13} I_{24} = 1)$. Although this probability $r_4 = P(I_{34} = 1 | I_{12} I_{13} I_{24} = 1)$ is architecture-dependent, we assume random choice, i.e., replace r_4 by $\frac{K}{N}$ and let the covariance be $\frac{n_1 K}{N} W_{23}$. Now divide by the product $\sqrt{m n_1} \sqrt{m n_1} = m n_1$ of the standard deviations to obtain the correlation coefficient. \square

The following auxiliary lemma will be needed as well.

Lemma 4

Let c, d be constants and (Z_1, Z_2) standard normal with correlation ρ . For a 'well behaved' function f (measurability and boundedness will do), let

$$\Psi(x) = \int f(x+z) \phi(z) dz \tag{80}$$

Then, as $\rho \rightarrow 0$,

$$E f(Z_1 + c) f(Z_2 + d) = \Psi(c) \Psi(d) + \rho \Psi'(c) \Psi'(d) + o(\rho). \tag{81}$$

Proof:

$$E f(Z_1 + c) f(Z_2 + d) = \int \int \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right\} f(x+c) f(y+d) dx dy$$

whose value at $\rho = 0$ is

$$\int \int \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(x^2 + y^2)\right\} f(x+c) f(y+d) dx dy = \Psi(c)\Psi(d)$$

and whose derivative at $\rho = 0$ is

$$\begin{aligned} & \int \int \frac{xy}{2\pi} \exp\left\{-\frac{1}{2}(x^2 + y^2)\right\} f(x+c) f(y+d) dx dy = \\ & \left(-\int f(x+c) d\phi(x)\right) \left(-\int f(y+d) d\phi(y)\right) = \Psi'(c)\Psi'(d) \square. \end{aligned}$$

We now identify $E(f_i^{(2)}|\xi_i, X_i, I_i^{(1)})$ and prove it to be as represented in (25).

Lemma 5

Let (compare with (36))

$$\Psi(x) = \int p(x+z)h(x+z)\phi(z)dz \quad (82)$$

Then

$$E\left(f_i^{(2)}|\xi_i, X_i, I_i^{(1)}\right) = \frac{K}{m}\alpha_2 \left[\bar{\Psi}_+\xi_i + r\sqrt{\alpha_1}\bar{\Psi}'X_iI_i^{(1)}\right] = \epsilon^*\xi_i + bX_iI_i^{(1)} \quad (83)$$

where

$$\bar{\Psi}_+ = \frac{1+\epsilon}{2}\Psi\left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon)\sqrt{\alpha_1}\right) + \frac{1-\epsilon}{2}\Psi\left(\frac{\epsilon}{\sqrt{\alpha_1}} - \gamma(\epsilon)\sqrt{\alpha_1}\right) \quad (84)$$

and

$$\bar{\Psi}' = \frac{1+\epsilon}{2}\Psi'\left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon)\sqrt{\alpha_1}\right) + \frac{1-\epsilon}{2}\Psi'\left(\frac{\epsilon}{\sqrt{\alpha_1}} - \gamma(\epsilon)\sqrt{\alpha_1}\right) \quad (85)$$

Proof:

The field $f_i^{(1)}$ has mean $\epsilon\xi_i$ and variance α_1 . We let $f_i^{(1)} = \epsilon\xi_i + \sqrt{\alpha_1}Z$. The contribution of neuron j to $f_i^{(1)}$ is the infinitesimal term (see (10)) $\Delta_{ij} = W_{ij}I_{ij}I_j^{(1)}X_j/n_1$. Whenever it is important to single out these infinitesimal contributions, we will denote the field by $\epsilon\xi_i + \sqrt{\alpha_1}Z + \Delta_{ij}$.

Without loss of generality, let $i = 1$. We split $E(f_1^{(2)}|\xi_1, X_1, I_1^{(1)})$ into the contribution of the true and random memories. Omitting conditioning symbols, the contribution of the true memory is

$$\begin{aligned}
& E \sum_j \xi_1 \xi_j I_{1j} I_j^{(2)} h \left(\frac{f_j^{(1)}}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} X_j \right) = \\
& K \xi_1 E \xi_2 I_2^{(2)} h \left(\frac{\epsilon}{\sqrt{\alpha_1}} \xi_2 + \gamma(\epsilon) \sqrt{\alpha_1} X_2 + Z \right) = \\
& K \xi_1 E \left[I_2^{(2)} h \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} X_2 \xi_2 + Z \right) \right] = \\
& K \xi_1 \left[\frac{1+\epsilon}{2} \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} \right) + \frac{1-\epsilon}{2} \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} - \gamma(\epsilon) \sqrt{\alpha_1} \right) \right] = K \xi_1 \bar{\Psi}_+ .
\end{aligned} \tag{86}$$

The contribution of the random memories is calculated as follows, where we single out of the field $f_2^{(1)}$ the contribution Δ_{21} of neuron 1.

$$\begin{aligned}
& E W_{12} I_{12} I_2^{(2)} h \left(\frac{f_2^{(1)}}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} X_2 \right) = \\
& E \left[W_{12} I_{12} I_2^{(2)} h \left(\frac{\epsilon}{\sqrt{\alpha_1}} \xi_2 + \gamma(\epsilon) \sqrt{\alpha_1} X_2 + \frac{\sqrt{\alpha_1}}{m} W_{12} I_{21} I_1^{(1)} X_1 + Z \right) \right] = \\
& E \left[W_{12} I_{12} \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} \xi_2 + \gamma(\epsilon) \sqrt{\alpha_1} X_2 + \frac{\sqrt{\alpha_1}}{m} W_{12} I_{21} I_1^{(1)} X_1 \right) \right] = \\
& E \left[W_{12} I_{12} \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} \xi_2 + \gamma(\epsilon) \sqrt{\alpha_1} X_2 \right) \right] + \\
& E \left[W_{12} I_{12} \frac{\sqrt{\alpha_1}}{m} W_{12} I_{21} I_1^{(1)} X_1 \Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}} \xi_2 + \gamma(\epsilon) \sqrt{\alpha_1} X_2 \right) \right] = \\
& 0 + \frac{E(W_{12}^2)}{m} \sqrt{\alpha_1} E(I_{12} I_{21}) X_1 I_1^{(1)} E \left[\Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} X_2 \xi_2 \right) \right] = \\
& \frac{K}{N} r \sqrt{\alpha_1} X_1 I_1^{(1)} \left[\frac{1+\epsilon}{2} \Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} \right) + \frac{1-\epsilon}{2} \Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}} - \gamma(\epsilon) \sqrt{\alpha_1} \right) \right] = \\
& \frac{K}{N} r \sqrt{\alpha_1} X_1 I_1^{(1)} \bar{\Psi}'
\end{aligned} \tag{87}$$

Hence, dividing by n_2 and adding up we get,

$$E \left(f_i^{(2)} | \xi_i, X_i, I_i^{(1)} \right) = \frac{K}{m} \alpha_2 \bar{\Psi}_+ \xi_i + \frac{K}{m} r \alpha_2 \sqrt{\alpha_1} \bar{\Psi}' X_i I_i^{(1)} = \epsilon^* \xi_i + b X_i I_i^{(1)} \square$$

As explained after (30), the parameter a is $Cov \left(f_i^{(1)}, f_i^{(2)} | \xi_i, X_i, I_i^{(1)} \right) / \alpha_1$. Expression (44) for a follows directly from

Lemma 6

$$\text{Cov}(f_i^{(1)}, f_i^{(2)} | \xi_i, X_i, I_i^{(1)}) = \alpha_2 \left[\bar{\Psi}_- + \sqrt{\alpha_1} \frac{K^2}{mN} \bar{\Psi}' \right] \quad (88)$$

where

$$\bar{\Psi}_- = \frac{1+\epsilon}{2} \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} \right) - \frac{1-\epsilon}{2} \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} - \gamma(\epsilon) \sqrt{\alpha_1} \right) \quad (89)$$

Proof:

$$\begin{aligned} & \text{Cov}(f_i^{(1)}, f_i^{(2)} | \xi_i, X_i, I_i^{(1)}) = \\ & \frac{N}{n_1 n_2} E \left[I_{12} I_2^{(1)} I_2^{(2)} W_{12}^2 X_2 h \left(\frac{\epsilon}{\sqrt{\alpha_1}} \xi_2 + \gamma(\epsilon) \sqrt{\alpha_1} X_2 + \frac{\sqrt{\alpha_1}}{m} W_{12} I_{21} I_1^{(1)} X_1 + Z \right) \right] + \\ & \quad \frac{N^2}{n_1 n_2} E I_{12} I_{13} I_3^{(1)} I_2^{(2)} W_{12} W_{13} X_3. \\ & h \left(\frac{\epsilon}{\sqrt{\alpha_1}} \xi_2 + \gamma(\epsilon) \sqrt{\alpha_1} X_2 + \frac{\sqrt{\alpha_1}}{m} W_{12} I_{21} I_1^{(1)} X_1 + \frac{\sqrt{\alpha_1}}{m} W_{23} I_{23} I_3^{(1)} X_3 + Z \right) = \\ & \frac{N}{n_1 n_2} E \left[I_{12} I_2^{(1)} W_{12}^2 X_2 \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} \xi_2 + \gamma(\epsilon) \sqrt{\alpha_1} X_2 + \frac{\sqrt{\alpha_1}}{m} W_{12} I_{21} I_1^{(1)} X_1 \right) \right] + \\ & \quad \frac{N^2}{n_1 n_2} E I_{12} I_{13} I_3^{(1)} W_{12} W_{13} X_3. \\ & \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} \xi_2 + \gamma(\epsilon) \sqrt{\alpha_1} X_2 + \frac{\sqrt{\alpha_1}}{m} W_{12} I_{21} I_1^{(1)} X_1 + \frac{\sqrt{\alpha_1}}{m} W_{23} I_{23} I_3^{(1)} X_3 \right) = \\ & \frac{mN}{n_1 n_2} E(I_{12}) E(I_2^{(1)}) E \left[X_2 \xi_2 \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} X_2 \xi_2 \right) \right] + 0 + 0 + 0 + \\ & \quad \frac{mN^2}{n_1 n_2} E(I_{12} I_{13} I_{23}) E(I_3^{(1)}) \frac{\sqrt{\alpha_1}}{m} E \left[\Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} X_2 \xi_2 \right) \right] = \\ & \quad \alpha_2 \left[\frac{1+\epsilon}{2} \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} \right) - \frac{1-\epsilon}{2} \Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}} - \gamma(\epsilon) \sqrt{\alpha_1} \right) \right] + \\ & \frac{mN^2}{n_1 n_2} \frac{K}{N} \frac{K}{N} P(I_{23} = 1 | I_{12} I_{13} = 1) \frac{n_1 \sqrt{\alpha_1}}{K} \frac{1}{m} \left[\frac{1+\epsilon}{2} \Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} \right) + \frac{1-\epsilon}{2} \Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}} - \gamma(\epsilon) \sqrt{\alpha_1} \right) \right] \equiv \\ & \quad \alpha_2 \bar{\Psi}_- + \frac{K}{m} \alpha_2 P(I_{23} = 1 | I_{12} I_{13} = 1) \sqrt{\alpha_1} \bar{\Psi}' . \end{aligned}$$

Just as in the proof of lemma 3, the parameter $r_3 = P(I_{23} = 1 | I_{12} I_{13} = 1)$ is architecture-

dependent. We model it simply as $\frac{K}{N}$. We thus get, $\text{Cov}(f_i^{(1)}, f_i^{(2)} | \xi_i, X_i, I_i^{(1)}) = \alpha_2 \left[\bar{\Psi}_- + \sqrt{\alpha_1} \frac{K^2}{mN} \bar{\Psi}' \right]$

and $a = (\alpha_2/\alpha_1) \left[\bar{\Psi}_- + \sqrt{\alpha_1} \frac{K^2}{mN} \bar{\Psi}' \right]$. \square

By (22), in order to evaluate $\tau^2 = Var(V|U)$ (see (24)) we need to compute $Var(V) = Var(f_i^{(2)}|\xi_i, X_i, I_i^{(1)})$, as the remaining terms have already been identified. Since $f_i^{(2)}$ is expressed as a sum (see (20)), its variance is expressible as the sum of the individual variances and the individual covariances. Lemmas 7 and 8 are dedicated to these two chores, respectively.

Lemma 7

$$E \left[\left(W_{12} I_{12} I_2^{(2)} h \left(\frac{\epsilon}{\sqrt{\alpha_1}} \xi_2 + \gamma(\epsilon) \sqrt{\alpha_1} X_2 + \frac{\sqrt{\alpha_1}}{m} W_{12} I_{21} I_1^{(1)} X_1 + Z \right) \right)^2 \right] = m \frac{K}{N} \bar{\Psi}_s \quad (90)$$

where

$$\Psi_s = \int p(x+z) h^2(x+z) \phi(z) dz \quad (91)$$

and

$$\bar{\Psi}_s = \frac{1+\epsilon}{2} \Psi_s \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} \right) + \frac{1-\epsilon}{2} \Psi_s \left(\frac{\epsilon}{\sqrt{\alpha_1}} - \gamma(\epsilon) \sqrt{\alpha_1} \right) \quad (92)$$

Proof:

$$\begin{aligned} E \left[\left(W_{12} I_{12} I_2^{(2)} h \left(\frac{\epsilon}{\sqrt{\alpha_1}} \xi_2 + \gamma(\epsilon) \sqrt{\alpha_1} X_2 + \frac{\sqrt{\alpha_1}}{m} W_{12} I_{21} I_1^{(1)} X_1 + Z \right) \right)^2 \right] &= \\ m \frac{K}{N} E \left[I_2^{(2)} h^2 \left(\frac{\epsilon}{\sqrt{\alpha_1}} \xi_2 + \gamma(\epsilon) \sqrt{\alpha_1} X_2 + Z \right) \right] &= \\ m \frac{K}{N} \frac{1+\epsilon}{2} \Psi_s \left(\frac{\epsilon}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} \right) + \frac{1-\epsilon}{2} \Psi_s \left(\frac{\epsilon}{\sqrt{\alpha_1}} - \gamma(\epsilon) \sqrt{\alpha_1} \right) &= m \frac{K}{N} \bar{\Psi}_s \square \end{aligned}$$

We remark that if $h(x) = Sgn(x)$, then Ψ_a (for ‘absolute’) and Ψ_s (for ‘square’) are identical, and in this case expression (90) is $\frac{mn_2}{N}$.

Lemma 8

$$E \left[W_{12} W_{13} I_{12} I_2^{(2)} I_{13} I_3^{(2)} h \left(\frac{f_2^{(1)}}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} X_2 \right) h \left(\frac{f_3^{(1)}}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} X_3 \right) \middle| \xi_1, X_1, I_1^{(1)} \right] = \quad (93)$$

$$\frac{2}{\sqrt{\alpha_1}} \frac{mK^2}{N^3} \bar{\Psi}_- \bar{\Psi}' + (\alpha_1 r^2 I_1^{(1)} + \frac{K}{N}) \left(\frac{K}{N} \right)^2 (\bar{\Psi}')^2$$

Proof:

$$E \left[W_{12}W_{13}I_{12}I_2^{(2)}I_{13}I_3^{(2)}h \left(\frac{f_2^{(1)}}{\sqrt{\alpha_1}} + \gamma(\epsilon)\sqrt{\alpha_1}X_2 \right) h \left(\frac{f_3^{(1)}}{\sqrt{\alpha_1}} + \gamma(\epsilon)\sqrt{\alpha_1}X_3 \right) \middle| \xi_1, X_1, I_1^{(1)} \right] =$$

$$EW_{12}W_{13}I_{12}I_2^{(2)}I_{13}I_3^{(2)}.$$

$$h \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_2 + \gamma(\epsilon)\sqrt{\alpha_1}X_2 + \frac{\sqrt{\alpha_1}}{m}W_{12}I_{21}I_1^{(1)}X_1 + \frac{\sqrt{\alpha_1}}{m}W_{23}I_{23}I_3^{(1)}X_3 + Z_2 \right).$$

$$h \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_3 + \gamma(\epsilon)\sqrt{\alpha_1}X_3 + \frac{\sqrt{\alpha_1}}{m}W_{13}I_{31}I_1^{(1)}X_1 + \frac{\sqrt{\alpha_1}}{m}W_{23}I_{32}I_2^{(1)}X_2 + Z_3 \right).$$

By Lemma 4, Lemma 3(ii) and the remark following its proof, this expression is

$$EW_{12}W_{13}I_{12}I_{13}\Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_2 + \gamma(\epsilon)\sqrt{\alpha_1}X_2 + \frac{\sqrt{\alpha_1}}{m}W_{12}I_{21}I_1^{(1)}X_1 + \frac{\sqrt{\alpha_1}}{m}W_{23}I_{23}I_3^{(1)}X_3 \right).$$

$$\Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_3 + \gamma(\epsilon)\sqrt{\alpha_1}X_3 + \frac{\sqrt{\alpha_1}}{m}W_{13}I_{31}I_1^{(1)}X_1 + \frac{\sqrt{\alpha_1}}{m}W_{23}I_{32}I_2^{(1)}X_2 \right) +$$

$$EW_{12}W_{13}I_{12}I_{13}\frac{K}{Nm}W_{23}$$

$$\Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_2 + \gamma(\epsilon)\sqrt{\alpha_1}X_2 + \frac{\sqrt{\alpha_1}}{m}W_{12}I_{21}I_1^{(1)}X_1 + \frac{\sqrt{\alpha_1}}{m}W_{23}I_{23}I_3^{(1)}X_3 \right).$$

$$\Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_3 + \gamma(\epsilon)\sqrt{\alpha_1}X_3 + \frac{\sqrt{\alpha_1}}{m}W_{13}I_{31}I_1^{(1)}X_1 + \frac{\sqrt{\alpha_1}}{m}W_{23}I_{32}I_2^{(1)}X_2 \right).$$

A careful study of the Taylor expansion of Ψ shows that the only nonnegligible contributions come from the terms involving $\Psi(2)\Psi'(3)\Delta_{32}$, $\Psi(3)\Psi'(2)\Delta_{23}$ and $\Psi'(2)\Psi'(3)\Delta_{21}\Delta_{31}$ in the first summand, and the term involving $\Psi'(2)\Psi'(3)$ in the second summand. These contributions add up to

$$\frac{\sqrt{\alpha_1}}{m}EW_{12}W_{13}W_{23}I_2^{(1)}I_{12}I_{13}I_{32}X_2\Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_2 + \gamma(\epsilon)\sqrt{\alpha_1}X_2 \right) \Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_3 + \gamma(\epsilon)\sqrt{\alpha_1}X_3 \right) +$$

$$\frac{\sqrt{\alpha_1}}{m}EW_{12}W_{13}W_{23}I_3^{(1)}I_{12}I_{13}I_{23}X_3\Psi \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_3 + \gamma(\epsilon)\sqrt{\alpha_1}X_3 \right) \Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_2 + \gamma(\epsilon)\sqrt{\alpha_1}X_2 \right) +$$

$$\frac{\alpha_1}{m^2}W_{12}^2W_{13}^2I_{12}I_{13}I_{21}I_{31}I_1^{(1)}\Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_3 + \gamma(\epsilon)\sqrt{\alpha_1}X_3 \right) \Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_2 + \gamma(\epsilon)\sqrt{\alpha_1}X_2 \right)$$

$$\frac{K}{Nm}EW_{12}W_{13}W_{23}I_{12}I_{13}\Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_3 + \gamma(\epsilon)\sqrt{\alpha_1}X_3 \right) \Psi' \left(\frac{\epsilon}{\sqrt{\alpha_1}}\xi_2 + \gamma(\epsilon)\sqrt{\alpha_1}X_2 \right) =$$

$$2\sqrt{\alpha_1}\frac{n_1}{K}\left(\frac{K}{N}\right)^3\bar{\Psi}_-\bar{\Psi}' + \alpha_1\left(\frac{K}{N}\right)^2r^2I_1^{(1)}(\bar{\Psi}')^2 + \left(\frac{K}{N}\right)^3(\bar{\Psi}')^2 =$$

$$\frac{2}{\sqrt{\alpha_1}}\frac{mK^2}{N^3}\bar{\Psi}_-\bar{\Psi}' + (\alpha_1r^2I_1^{(1)} + \frac{K}{N})\left(\frac{K}{N}\right)^2(\bar{\Psi}')^2$$

□.

We now combine lemmas 7 and 8 and expression (87) to get

$$Var(f_i^{(2)}|\xi_i, X_i, I_i^{(1)}) = \frac{1}{n_2^2} \left[NV ar \left(W_{12} I_{12} I_2^{(2)} h \left(\frac{f_2^{(1)}}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} X_2 \right) \right) \right] + \quad (94)$$

$$\begin{aligned} & \frac{1}{n_2^2} \left[N^2 Cov \left(W_{12} I_{12} I_2^{(2)} h \left(\frac{f_2^{(1)}}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} X_2 \right) , W_{13} I_{13} I_3^{(2)} h \left(\frac{f_3^{(1)}}{\sqrt{\alpha_1}} + \gamma(\epsilon) \sqrt{\alpha_1} X_3 \right) \right) \right] = \\ & \frac{1}{n_2^2} \left\{ N \left[E(\cdot)^2 - (E(\cdot))^2 \right] + N^2 \left[E(\cdot) E(\cdot) - (E(\cdot))^2 \right] \right\} = \\ & \frac{1}{n_2^2} N \left[\frac{mK}{N} \bar{\Psi}_s - \left(\frac{K}{N} \right)^2 r^2 \alpha_1 I_1^{(1)} (\bar{\Psi}')^2 \right] + \\ & \frac{1}{n_2^2} N^2 \left[\frac{2}{\sqrt{\alpha_1}} \frac{mK^2}{N^3} \bar{\Psi}_- \bar{\Psi}' + (\alpha_1 r^2 I_1^{(1)} + \frac{K}{N}) \left(\frac{K}{N} \right)^2 (\bar{\Psi}')^2 - \left(\frac{K}{N} \right)^2 r^2 \alpha_1 I_1^{(1)} (\bar{\Psi}')^2 \right] = \\ & \alpha_2^2 \left\{ \frac{K}{m} \bar{\Psi}_s + \frac{2}{\sqrt{\alpha_1}} \frac{K^2}{Nm} \bar{\Psi}_- \bar{\Psi}' + \frac{K^3}{Nm^2} (\bar{\Psi}')^2 \right\} \end{aligned}$$

from which, via (22) and (24),

$$\begin{aligned} \tau^2 &= \alpha_2^2 \left\{ \frac{K}{m} \bar{\Psi}_s + \frac{2}{\sqrt{\alpha_1}} \frac{K^2}{Nm} \bar{\Psi}_- \bar{\Psi}' + \frac{K^3}{Nm^2} (\bar{\Psi}')^2 \right\} - \left(\alpha_2 \left[\bar{\Psi}_- + \sqrt{\alpha_1} \frac{K^2}{mN} \bar{\Psi}' \right] \right)^2 / \alpha_1 = \\ & \alpha_2^2 \frac{K}{m} \bar{\Psi}_s - \frac{\alpha_2^2}{\alpha_1} (\bar{\Psi}_-)^2 + \alpha_2^2 \frac{K^3}{Nm^2} \left(1 - \frac{K}{N} \right) (\bar{\Psi}')^2 = \\ & \alpha_2 \frac{\bar{\Psi}_s}{\bar{\Psi}_a} - \frac{m^2}{\alpha_1 K^2} \left(\frac{\bar{\Psi}_-}{\bar{\Psi}_a} \right)^2 + \frac{K}{N} \left(1 - \frac{K}{N} \right) \left(\frac{\bar{\Psi}'}{\bar{\Psi}_a} \right)^2 . \end{aligned} \quad (95)$$

Since $h(x) = Sgn(x)$ implies that $\Psi_s = \Psi_a$, we have proved and generalized expression (46).

References

- [1] J.J. Hopfield. Neural networks and physical systems with emergent collective abilities. *Proc. Nat. Acad. Sci. USA*, 79:2554, 1982.
- [2] 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.
- [3] D.J. Amit and A. Treves. Associative memory neural networks with low temporal spiking rates. *Proc. Natl. Acad. Sci.*, 86:7671, 1989.
- [4] J. Buhmann. Oscillations and low firing rates in associative memory neural networks. *Phys. Rev. A.*, 40:4145, 1989.
- [5] D. J. Amit and M. V. Tsodyks. Quantitative study of attractor neural network retrieving at low spike rates: I. substrate–spikes, rates and neuronal gain. *Network*, 2:259–273, 1991.
- [6] A. Treves and E. T. Rolls. What determines the capacity of autoassociative memories in the brain? *Network*, 2:371–397, 1991.
- [7] W. Gerstner and J. Leo van Hemmen. Universality in neural networks: the importance of the ‘mean firing rate’. *Biol. Cybern.*, 67:195–205, 1992.
- [8] C. F. Stevens. How cortical interconnectedness varies with network size? *Neural Computation*, 1:473–479, 1989.
- [9] M. Abeles, E. Vaadia, and H. Bergman. Firing patterns of single units in the prefrontal cortex and neural network models. *Network*, 1:13–25, 1990.
- [10] H. Englisch and Yegao Xiao. Neural networks as perpetual information generators. *Phys. Rev. A.*, 44(2):1382–1385, 1991.
- [11] B. Derrida, E. Gardner, and A. Zippelius. An exactly solvable asymmetric neural network model. *Europhys. Lett.*, 4:167–173, 1987.
- [12] E. Gardner, B. Derrida, and P. Mottishaw. Zero temperature parallel dynamics for infinite range spin glasses and neural networks. *J. Physique*, 48:741–755, 1987.

- [13] A. E. Patrick and V.A. Zagrebnov. On the parallel dynamics for the little-hopfield model. Technical Report CPT-89/P.2311, Centre de Physique Theoretique Marseille, 1989.
- [14] S.I. Amari and K. Maginu. Statistical neurodynamics of associative memory. *Neural Networks*, 1:67–73, 1988.
- [15] H.F. Yanai and Y. Sawada. Associative memory network composed of neurons with hysteretic property. *Neural Networks*, 3:223–228, 1990.
- [16] J.H. Wang, T.F. Krile, and T.L. Jong. On the characteristics of the autoassociative memory with nonzero diagonal terms in the memory matrix. *Neural Computation*, 3:428–439, 1991.
- [17] H.F. Yanai, Y. Sawada, and S. Yoshizawa. Dynamics of an auto-associative neural network model with arbitrary connectivity and noise in the threshold. *Network*, 2:295–314, 1991.
- [18] W. Lytton. Simulations of cortical pyramidal neurons synchronized by inhibitory interneurons. *J. Neurophysiol.*, 66(3):1059–1079, 1991.
- [19] H.R. Wilson and J.D. Cowan. Excitatory and inhibitory interactions in localized populations of model neurons. *Biophysical Journal*, 12:1–24, 1972.
- [20] D. Servan-Schrieber, H. Printz, and J.D. Cohen. A network model of catecholamine effects: gain, signal-to-noise ratio, and behavior. *Science*, 249:892–895, 1990.
- [21] H. Englisch, 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.
- [22] W. Kinzel. Learning and pattern recognition in spin glass models. *Z. Physik*, B60:205–213, 1985.
- [23] G. H. Hardy, J. E. Littlewood, and G. Polya. *Inequalities*. Cambridge University Press, 1934.
- [24] M. Rothschild and J. Stiglitz. Increasing risk i: a definition. *Journal of Economic Theory*, 2:225–243, 1970.

- [25] A. Chernjavsky and J. Moody. Spontaneous development of modularity in simple cortical models. *Neural Computation*, 12:334–354, 1990.
- [26] M. Nelson and J.M. Bower. Brain maps and parallel computers. *Trends Neurosci.*, 13:403–408, 1990.

Performance	predicted	experimental
Single $\gamma(\epsilon)$ iteration	0.893	0.895
History-dependent dynamics	0.903	0.907
Hopfield - zero diagonal	–	0.878,0.867
Independent $\gamma(\epsilon)$ diagonal	0.978	–
Independent zero diagonal	0.95	–

Table 1: Fully connected, full activity network: $N = 500, K = 500, n_1 = n_2 = 500, m = 100$.

Performance	predicted	experimental
Random activation	0.902	0.904
Tail censoring	0.90	0.897
Interval censoring	0.945	0.937
Hybrid censoring/signalling	0.963	0.964
Hopfield - zero diagonal	–	0.865,0.869

Table 2: Fully connected, variable activity network: $N = 500, K = 500, n_1 = 500, n_2 = 300, m = 100$.

Performance	predicted	experimental
Single $\gamma(\epsilon)$ iteration	0.872	0.869
Random activation	0.896	0.903
Tail censoring	0.878	0.873
Interval censoring/activation	0.958	0.951
Hybrid censoring/signalling	0.967	0.968
Hopfield - zero diagonal	–	0.871,0.875
Independent $\gamma(\epsilon)$ diagonal	0.96	–
Independent zero diagonal	0.913	–

Table 3: Fully connected, mid-activity network: $N = 500, K = 500, n_1 = n_2 = 200, m = 50$.

Performance	predicted	experimental
Random activation	0.911	0.907
Tail/Interval censoring	0.976	0.979
Hybrid censoring/signalling	0.986	0.983
Hopfield - zero diagonal	–	0.87,0.871

Table 4: Fully connected, low activity network: $N = 500, K = 500, n_1 = n_2 = 40, m = 10$.

Performance	predicted	experimental
Random activation	0.926	0.928
Tail censoring	0.913	0.91
Interval/Hybrid censoring	0.957	0.952
Hopfield - zero diagonal	–	0.868,0.879

Table 5: Mid-size connectivity, mid-activity network: $N = 500, K = 200, n_1 = n_2 = 100, m = 25$.

Performance	predicted	experimental
Random activation	0.937	0.932
Tail censoring	0.89	0.897
Interval/Hybrid censoring	0.97	0.972
Hopfield - zero diagonal	–	0.854, 0.872

Table 6: Mid-size connectivity, low-activity network: $N = 500, K = 200, n_1 = n_2 = 40, m = 10$.

Performance	predicted	experimental
Random activation	0.955	0.951
Tail censoring	0.972	0.973
Interval/Hybrid censoring	0.975	0.972
Hopfield - zero diagonal	–	0.902,0.973

Table 7: Sparsely connected, low activity network: $N = 1500, K = 50, n_1 = n_2 = 20, m = 5$.

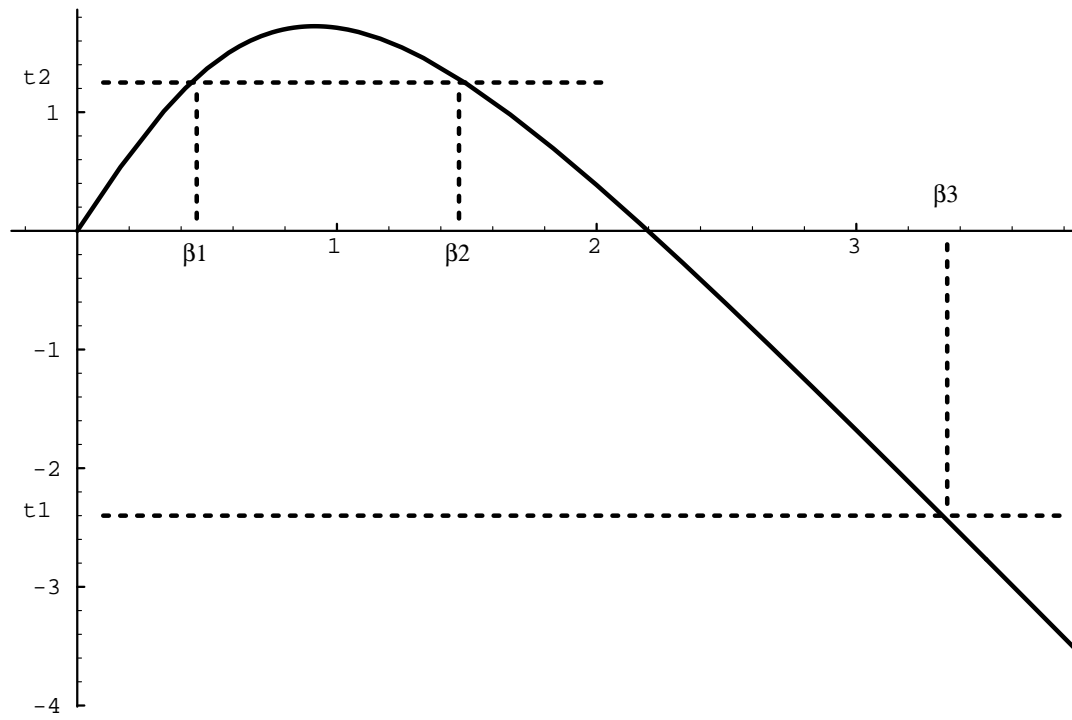


Figure 1: A typical plot of $R(x) = \phi_1(x)/\phi_0(x)$. Network parameters are $N = 500$, $K = 500$, $n_1 = 50$ and $m = 10$.

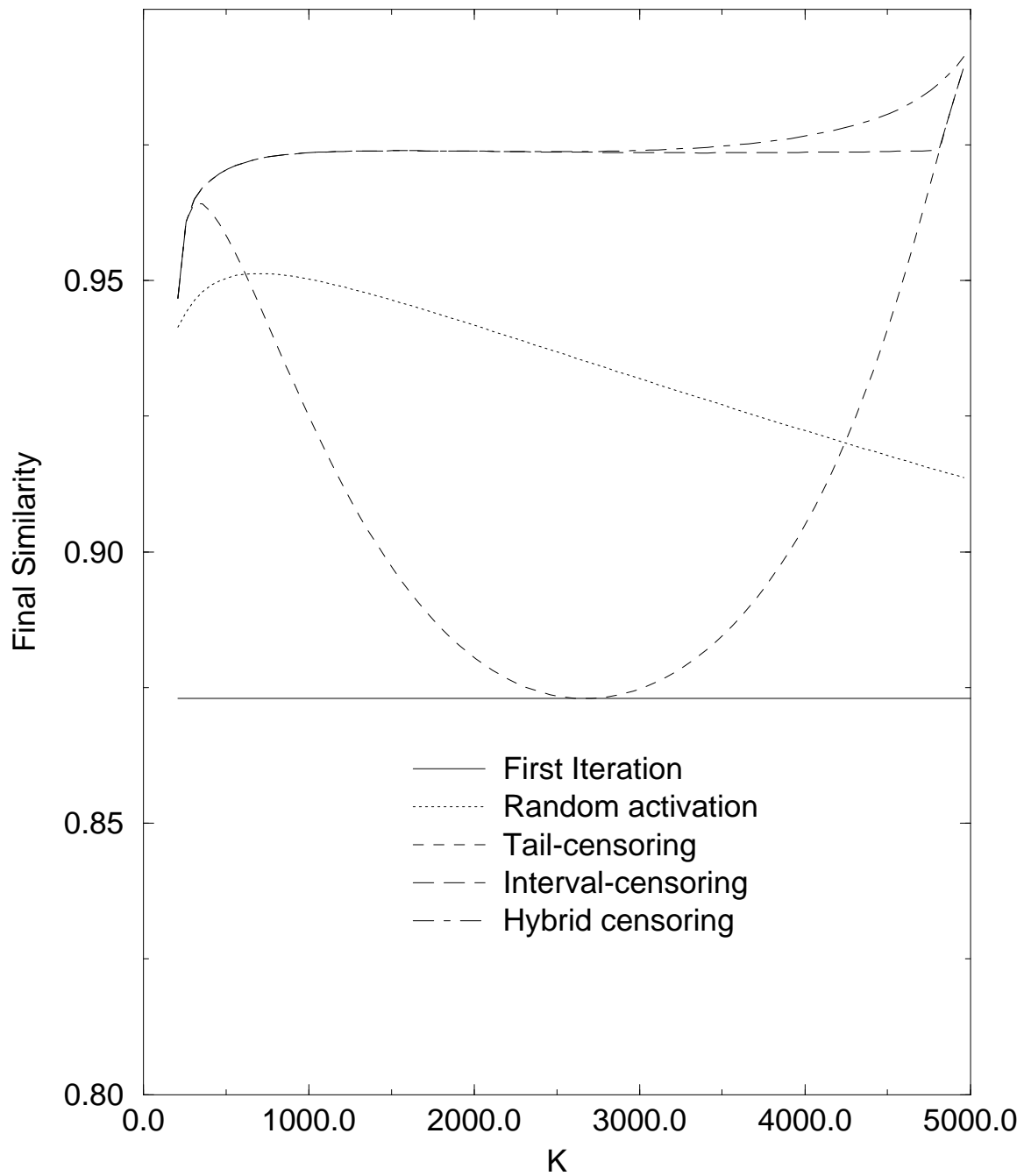


Figure 2: Performance of a small-scale cortical-like ‘columnar’ ANN, at different values of connectivity K , for initial similarity 0.75. $N = 5000$, $n_1 = n_2 = 200$, $m = 50$.

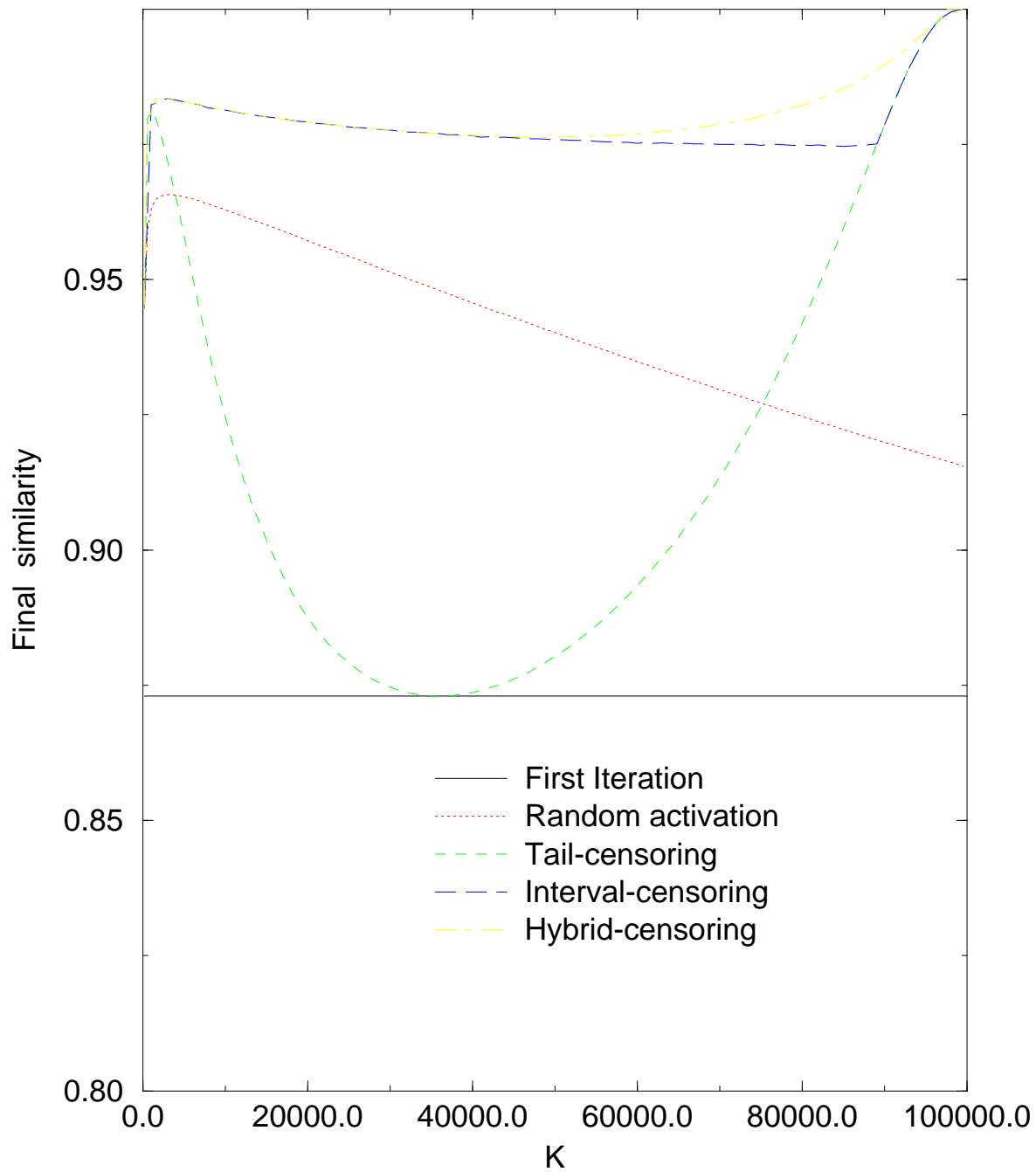


Figure 3: Performance of a large-scale cortical-like ‘columnar’ ANN, at different values of connectivity K , for initial similarity 0.75. $N = 10^5$, $n_1 = n_2 = 200$, $m = 50$.

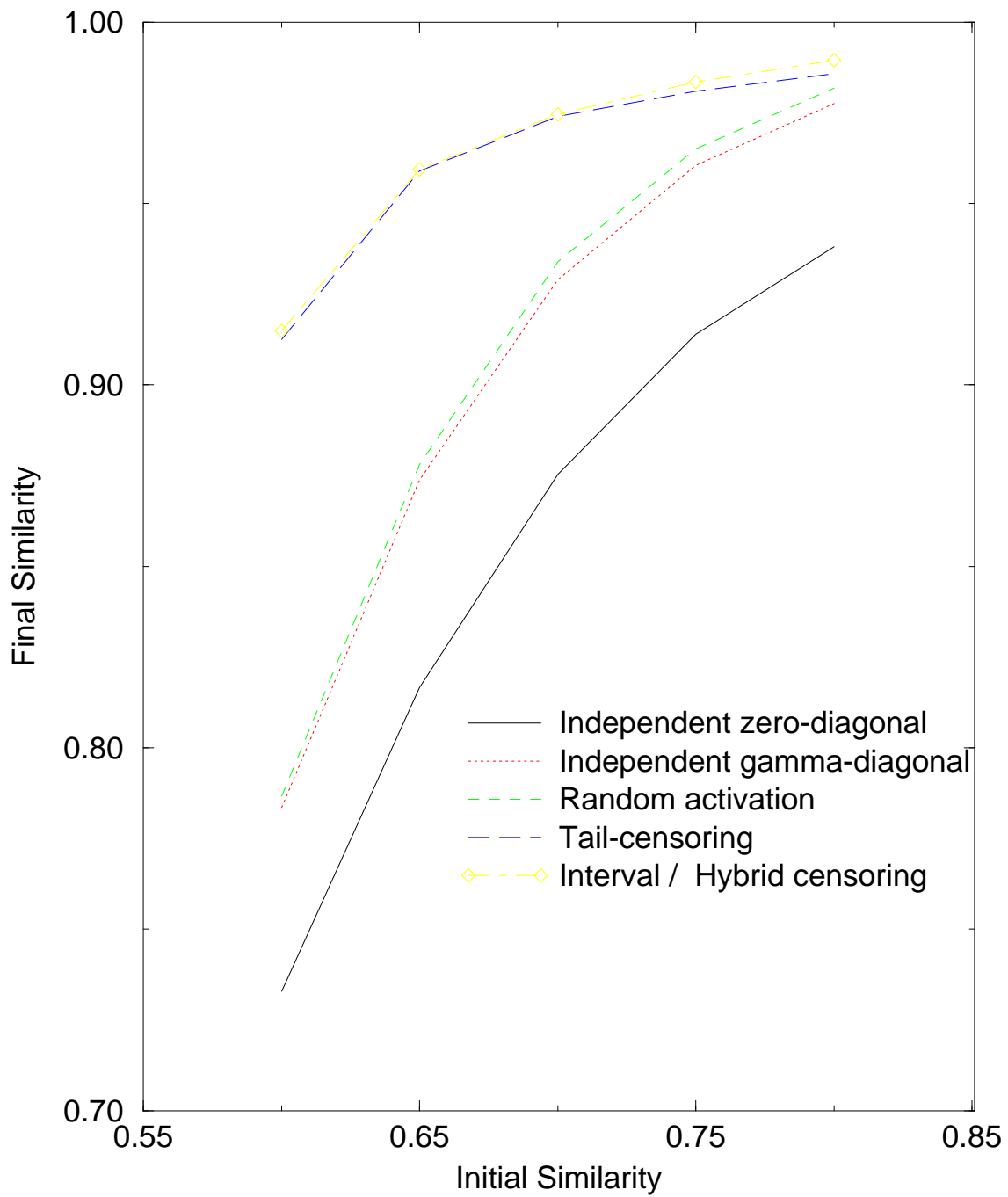


Figure 4: Performance of a large-scale cortical-like ‘columnar’ ANN with geometric mean connectivity K , at various initial similarity values. $N = 10^5$, $K = 2000$, $n_1 = n_2 = 200$, $m = 50$.