Optimal Firing in Sparsely-connected Low-activity Attractor 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.

December 21, 1995

Abstract

We examine the performance of Hebbian-like attractor neural networks, recalling stored memory patterns from their distorted versions. Searching for an activation (firing-rate) function that maximizes the performance in sparsely-connected low-activity networks, we show that the optimal activation function is a *Threshold-Sigmoid* of the neuron's input field. This function is shown to be in close correspondence with the dependence of the firing rate of cortical neurons on their integrated input current, as described by neurophysiological recordings and conduction-based models. It also accounts for the decreasing-density shape of firing rates that has been reported in the literature.

1 Introduction

The reduction of detailed conduction-based models of neuronal firing into simpler, lower-dimensional descriptions has recently received considerable attention (e.g., [Doya and Selverston, 1994, Ermentrout, 1994]). Such reductions lead to frequency-current curves similar to the curve obtained with the more detailed Hodgkin-Huxley model [Doya and Selverston, 1994]. The shape of resulting frequency-current curves, which can be viewed as input/output activation functions governing the neuronal firing rate, can qualitatively be described as *Threshold-Sigmoid (TS)*. That is, a neuron fires only when its integrated input current (denoted also as input field) is above some threshold, and its firing rate is approximately a sigmoidal function of its above-threshold input field [1]. Closely related *Threshold-linear* functions have previously appeared in the literature as a fair approximation of the dependence of the firing rate of a cortical neuron on its integrated input current [Gustafsson and Wigstrom, 1981, Avoli and Olivier, 1989, Mason and Larkman, 1990]. Both Threshold-sigmoid and Threshold-linear activation functions were used by several investigators to construct biologically-motivated neural network models [Treves and Rolls, 1991, Amit and Tsodyks, 1991].

In this work we take a different approach. Instead of deriving the neuronal activation function by a 'bottom-up' dimension reduction of a model describing the dynamics of a single neuron, we examine optimal neuronal signaling at the network level. That is, we search for a neuronal activation function that maximizes the performance of an associative memory network, where the network's performance (defined formally below) is a measure of the accuracy in which it retrieves the correct memorized patterns in response to corresponding input cues. Applying this 'top-down' optimizing approach to an associative memory network with cortical-like characteristics of sparse-connectivity and low activity [Abeles, 1991, Abeles *et al.*, 1990], we show that the activation function yielding maximal performance is the *Threshold-Sigmoid (TS)* $sign(S)max(|S|-d,0)$, where S is the usual logistic sigmoid and the threshold d is dictated by the network's activity level. It is quite surprising and interesting that a qualitatively similar activation function is arrived at using both dimension reduction and performance optimization approaches.

[1] Such a 'sigmoidal-like' function has a first derivative near the origin which is greater than one and its derivative monotonically decreases and goes to zero as its argument increases.

This paper proceeds as follows: First we examine the mathematically convenient simple case of *symmetrically-coded* $\{-1,+1\}$ binary neurons, with no constraint on the activity level permitted in the network. In this case the optimal activation function is the conventional sigmoid (S) function of the neuron's input field. Then, adding the constraint of low activity, we shall show that the Threshold sigmoid is the optimal activation function. Subsequently, S and TS are shown to be optimal also in the more biologically realistic *quiescent-coded* $\{0,1\}$ formulation. Heavily-thresholded TS functions are shown to give rise to an exponential distribution of activation values, which is in accordance with the distribution of relatively low firing rates typically observed in recordings of delayed persistent activity [Miyashita and Chang, 1988, Miyashita, 1988, Abeles *et al.*, 1990].

2 Signaling in Symmetric-coded Networks

2.1 The Model

We examine the dynamics of an analog Hopfield-like attractor neural network (ANN) [Hopfield, 1984] composed of N ± 1 binary neurons. The network stores M memory patterns ξ^ν, $\nu = 1 \ldots M$, each being a randomly generated N-dimensional vector of ± 1's, in the synaptic matrix

$$W_{ij} = \sum_\nu \xi_i^\nu \xi_j^\nu \quad , \quad W_{ii} = 0 \ . \tag{1}$$

A stored memory pattern ξ^c is cued by initializing the firing state of neuron i, $1 \leq i \leq N$, with a value $a_i(0)$ which is close to its correct binary value, i.e., its value in the cued memory pattern. Thereafter the firing states of the neurons evolve in synchronous iterations. Let $a_j(t) \in [-1,+1]$ denote the analog activation (firing) of neuron j at time t. At iteration t neuron i computes its input field $h_i(t)$ as

$$h_i(t) = \sum_j W_{ij} a_j(t-1) \tag{2}$$

where the index j runs over all the neurons which are connected to neuron i, and $a_j(t)$ is some function $a_j(t) = F\left(h_j(s), s \leq t\right)$. The network is likely to converge to a stable steady-state, which is guaranteed under some conditions, such as connections' symmetry ($W_{ij} = W_{ji}$) and asynchronous updatings.

We adopt a Bayesian framework, where each neuron is conceived as a Bayesian decision maker (see [Kononenko, 1989, Lansner and Ekeberg, 1989, Meilijson and Ruppin, 1993,

Meilijson and Ruppin, 1994]). In Bayesian terminology, we assume that on the basis of its previous input fields and independently of everything else, neuron i ($1 \leq i \leq N$) assigns some *prior belief* λ_i to the event $\{\xi^c{}_i = +1\}$, i.e., that it should fire $+1$. Note that this decision does not require the neuron to identify ξ^c, just to dichotomously decide on its sign in the currently cued pattern, regardless of its identity. During the current iteration, the incoming signals each neuron receives sum up to yield its new input field, on the basis of which every neuron formulates a *posterior belief* that it should fire $+1$. This updated posterior belief should determine the neuron's new activation signal to its neighbors in the next iteration.

In this study, we concentrate on analog activation functions. From a a biological point of view, such functions describe the dependence of the neuron's current firing rate on its membrane potential. From a Bayesian perspective, analog functions reflect the notion that the neuron's belief in its correct state is rarely complete. This continuous description must be put on common grounds with the discrete, dichotomous, description of the stored memory patterns; if the activation signal is analog, how should we measure the performance of the network, i.e., the level of accuracy of which the cued patterns are reconstructed during memory retrieval? One approach to this issue has been to use a measure of continuous, analog similarity, as in [Amit and Tsodyks, 1991]. Viewing each neuron as Bayesian decision maker, we take another approach, and let each neuron express its belief in its true state as a dichotomous activation signal. That is, we quantify the current similarity of the network to the cued memory pattern by considering each neuron whose belief in $+1$ is greater (lesser) than 0.5 as if it 'decided' its correct sign is $+1$ (-1) respectively. This technically convenient description does not necessarily imply the existence of some biological mechanism that converts the final value of the local field (or, equivalently, the belief) into a dichotomous state. Alternatively, the belief of the neuron that its true state is $+1$ may be interpreted as the probability of finding it in the $+1$ state at a given 'time slice' during the next iteration.

Within this framework, our goal is to find, over the space of all possible input/output activation functions, the signal function F that maximizes the network's performance. This task, however, is very difficult in the general case since after just a few iterations strong correlations evolve between a neuron's current input field and its previous input field values [Englisch and Xiao, 1991, Meilijson and Ruppin, 1993, Meilijson and Ruppin, 1994]. We

therefore perform our analysis as if the neuron's current input field was independent of its previous fields. This independence assumption is known to be adequate under sparse connectivity [Derrida *et al.*, 1987, Evans, 1989] and low activity [Tsodyks and Feigel'man, 1988]. Due to the assumed independence, each iteration of the network may be viewed on its own (as in [Kinzel, 1985]), and our goal of maximizing the network's performance reduces to that of maximizing the performance increase during a *single* iteration. We henceforth drop the time index t and let F depend exclusively on the current input field. This simplification will be seen to yield explicit optimal activation functions, simple enough to allow for intuitive understanding and biological feasibility judgments.

2.2 The Optimality of the Sigmoid Signal

We now turn to find the optimal signal function F in the symmetrically-coded case. We assume that the network has been cued by an input pattern with some degree of similarity to one of the stored memories, ξ^c, and examine the activation signaling in some arbitrary iteration. It is then convenient to view the current input field of neuron i as a sum of two terms

$$h_i = \sum_j \left[{\xi^c}_i {\xi^c}_j + Z_{ij} \right] a_j = {\xi^c}_i \sum_j {\xi^c}_j a_j + \sum_j Z_{ij} a_j \tag{3}$$

where ${\xi^c}_i {\xi^c}_j$ is the 'signal' term and Z_{ij} is some noise with variance proportional to the number of memories M. The random generation of memories entails that for fixed i the Z_{ij} are independent. This, together with the independence assumption on the neurons' input fields entails that the (a_j)'s are independent of each other and of the Z_{ij}'s. The input field h_i is then a Gaussian variable with mean $s \cdot {\xi^c}_i$ where s is

$$s = \sum_j a_j E({\xi^c}_j) = \sum_j a_j \left[\lambda_j(+1) + (1-\lambda_j)(-1) \right] = \sum_j a_j (2\lambda_j - 1) \tag{4}$$

and standard deviation n (for *noise*) proportional to $\sqrt{\sum_j {a_j}^2}$. In this Gaussian case, [2] it is known that the signal-to-noise ratio s/n determines the performance of the ANN [Kinzel, 1985]. That is, given some fixed initial level of similarity, the similarity to the cued memory pattern after an additional iteration is a monotonically increasing function of the signal-to-noise ratio. (The proof of this well-known observation is analogous to the proof of a

[2] Throughout this paper, a 'Gaussian field' of neuron i means that the field has a conditional Gaussian distribution given any of the two possible memory states ${\xi_i}^c$.

similar claim in the asymmetrically-coded $\{0,1\}$ case, presented in Section 4.1). We shall now prove that

Theorem 1.

The activation function $a_i = 2\lambda_i - 1$ maximizes the signal-to-noise ratio.

Proof:

Since the signal-to-noise ratio remains invariant if the a_i's are multiplied by a positive constant, we fix the standard deviation, and maximize the expected field by differentiating with respect to (a_i) the Lagrangean

$$E(h|\lambda) - \nu Var(h|\lambda) = \sum_j (2\lambda_j - 1)a_j - \nu Var(Z) \sum_j {a_j}^2 \tag{5}$$

to obtain

$$2\lambda_i - 1 = 2\nu Var(Z) a_i \tag{6}$$

from which we see that a_i should be taken as any positive multiple of $2\lambda_i - 1$, such as $2\lambda_i - 1$ itself. □

We see that the optimal activation signal is the difference $\lambda_i - (1 - \lambda_i)$ of the neuron's beliefs in $+1$ and -1 respectively. It is well known that the belief of a Bayesian neuron is a sigmoidal function of its current input field (see e.g., [Hinton and Nowlan, 1990]), and also [Amit, 1989] for an equivalent result in terms of probabilities). It follows then that *the optimal activation function* in the symmetrically-coded case is a *sigmoid function* of the neuron's current input field.

3 Signaling in Low-activity Symmetric-coded Networks

Biological cortical networks have low levels of activity, i.e., at any given moment only a small fraction of the neurons fire at rates significantly higher than their low, spontaneous firing rate (e.g., see [Abeles, 1991]). To study the optimal activation function under this constraint (which may reflect limited metabolic resources), we restrict the sequence (a_j) by compelling it to have low activity; i.e., we put an upper bound on $\sum_j |a_j|$, and search for a sequence (a_j) which maximizes the expected input field under the two constraints of $\sum_j |a_j|$ and $\sum_j {a_j}^2$ bounded from above. We now prove

Theorem 2.

With restricted activity, the activation function $a_i = Sign(2\lambda_i - 1)\ Max(|2\lambda_i - 1| - d, 0)$ maximizes the signal-to-noise ratio.

Proof:

The Lagrangean now becomes

$$\sum_j (2\lambda_j - 1)a_j - \nu_1 Var(Z) \sum_j {a_j}^2 - \nu_2 \sum_j |a_j| \tag{7}$$

Rather than getting into Kuhn-Tucker methodology with non-differentiable objective functions, let us replace the absolute value function $|x|$ by its 'roundoff', illustrated in figure 1,

$$f(x) = \begin{cases} |x| & |x| > \epsilon \\ \frac{x^2+\epsilon^2}{2\epsilon} & |x| \leq \epsilon \end{cases} \tag{8}$$

with

$$f'(x) = \begin{cases} Sign(x) & |x| > \epsilon \\ \frac{x}{\epsilon} & |x| \leq \epsilon \end{cases} \tag{9}$$

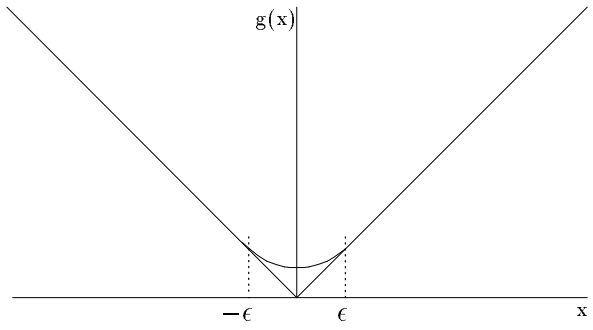

Figure 1: A 'roundoff' of the absolute value function.

and get a modified Lagrangean

$$\sum_j (2\lambda_j - 1)a_j - \nu_1 Var(Z) \sum_j {a_j}^2 - \nu_2 \sum_j f(a_j) \tag{10}$$

from which we see that the solution must satisfy

$$2a_i\nu_1 Var(z) + \nu_2 f'(a_i) = 2\lambda_i - 1\ . \tag{11}$$

A function of the form $\alpha x + \beta f'(x)$ is illustrated in figure 2.

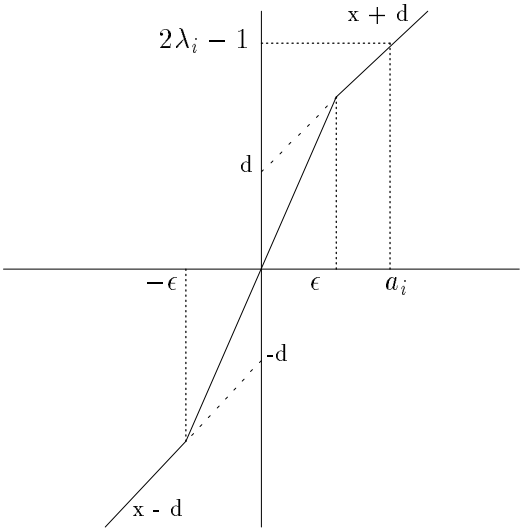

Figure 2: The function $\alpha x + \beta f'(x)$ (solid line).

We see then that for very small ϵ, a_i should be practically of the form

$$a_i = \begin{cases} 0 & -d < 2\lambda_i - 1 < d \\ 2\lambda_i - 1 - d & 2\lambda_i - 1 > d \\ 2\lambda_i - 1 + d & 2\lambda_i - 1 < -d \end{cases} \tag{12}$$

or, $a_i = Sign(2\lambda_i - 1)\ Max(|2\lambda_i - 1| - d, 0)$, completing the proof of Theorem 2. □

Figure 3 displays the optimal activation signal as a function of the neuron's input field h, where λ is sigmoid dependent on h. The parameter d is determined by the required value of $\sum_j |a_j| / \sqrt{\sum_j {a_j}^2}$. As evident, the optimal activation signal a has a TS form.

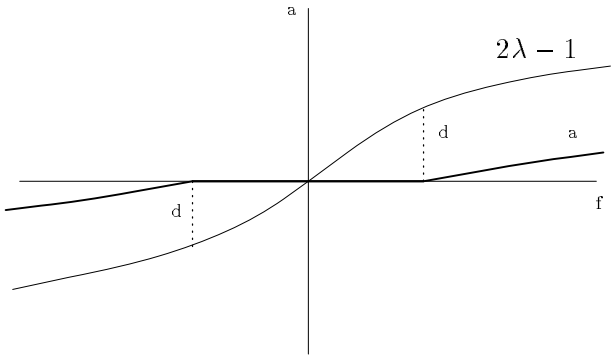

Figure 3: The TS activation function a (thick solid line), and the original S function $2\lambda - 1$ (thin solid line). The non-zero part of the TS function runs parallel (at distance d) to the S function.

4 Signaling in Low-activity Firing/Quiescent Networks

After examining the optimal activation achieved in the mathematically simple symmetrically-coded case, we proceed to the asymmetrically-coded $\{0, 1\}$ formulation, which better models the firing of biological neurons, and assume restricted total activity. The study of optimal activation in the latter formulation is the main goal of this paper. First, we shall present the specific $\{0, 1\}$ model we work in, and prove that in this model, as in the conventional $\{-1, +1\}$ formulation, the signal-to-noise ratio determines the network's performance level. Second, we prove our main result, showing that the Threshold-Sigmoid is indeed the optimal activation function in this case. As we shall see, this derivation requires different analytical methods than those employed in the symmetrically-coded model. Third, we show that the shape of the normalized non-zero part of the TS optimal activation function is practically independent of the total activity level in the network.

4.1 Performance is determined by the s/n ratio

For concreteness, we take the model proposed in [Tsodyks and Feigel'man, 1988] [3]. Each neuron is described by a binary variable $V = \{1, 0\}$ denoting an active (firing) or passive (quiescent) state, respectively. M distributed memory patterns ξ^ν are stored in the network, and each memory pattern contains a proportion p of $1's$. The weights of the synaptic

[3] As discussed in [Tsodyks and Feigel'man, 1988], the transformation from the $\{-1, +1\}$ to the $\{0, 1\}$ formulation in ANNs is not at all trivial as might have been expected.

connections are

$$W_{ij} = \sum_{\nu=1}^{M} (\xi^{\nu}{}_i - p)(\xi^{\nu}{}_j - p) \quad , \quad W_{ii} = 0 \tag{13}$$

and the field of neuron i is

$$h_i = \sum_j W_{ij} a_j = \sum_j \left[(\xi^c{}_i - p)(\xi^c{}_j - p) + Z_{ij} \right] a_j \ . \tag{14}$$

Similarly to the $\{-1,+1\}$ case, we assume that $(\xi^c{}_1, \lambda_1), (\xi^c{}_2, \lambda_2), \cdots, (\xi^c{}_N, \lambda_N)$ are independent. Our first task is to ascertain that the signal-to-noise ratio determines performance in the $\{0,1\}$ case [4]. To this end, it is straightforward to see that, given $\xi^c{}_i$, the field h_i has asymptotically constant variance, and mean of the form

$$E(h_i|\xi^c{}_i) = (\xi^c{}_i - p) E\left(\sum_j (\xi^c{}_j - p) a_j \right) = (\xi^c{}_i - p) E\left(\sum_j (\lambda_j - p) a_j \right) = (\xi^c{}_i - p)\mu \ . \tag{15}$$

Using Bayesian updating of probabilities in the Gaussian case, it can be shown that neuron i, having a prior belief λ_i and receiving a new input field h_i, will update its belief and acquire a new posterior belief $\tilde{\lambda_i}$ in accordance with

$$\tilde{\lambda_i} = P(\xi_i = 1|h_i) = \frac{1}{1 + e^{-\left[\log \frac{\lambda_i}{1-\lambda_i} + \frac{\mu}{\sigma^2}(h_i - \mu(\frac{1}{2} - p)) \right]}} \tag{16}$$

The neuron's current belief is therefore a sigmoid function of a linear combination of a function of its previous belief and the newly observed input field. From (16) one can deduce the probability that neuron i determines its state correctly (i.e., $P\left(\xi_i = 1 \ and \ \tilde{\lambda_i} > \frac{1}{2}\right) + P\left(\xi_i = 0 \ and \ \tilde{\lambda_i} < \frac{1}{2}\right)$), as

$$P(correct|\lambda_i) = \lambda_i \Phi\left(\omega + \frac{\frac{1}{2} \log \frac{\lambda_i}{1-\lambda_i}}{\omega} \right) + (1 - \lambda_i) \Phi\left(\omega - \frac{\frac{1}{2} \log \frac{\lambda_i}{1-\lambda_i}}{\omega} \right) \ , \tag{17}$$

where $2\omega = \frac{\mu}{\sigma}$ is the signal-to-noise ratio, and Φ is the standard normal cumulative distribution function. As can be seen by differentiation, expression (17) is monotonically increasing with ω, which is what we wanted to confirm. Note that if the prior belief λ_i is obtained as the sigmoid function evaluated at a Gaussian-distributed initial field (i.e., at $\log \frac{\lambda_i}{1-\lambda_i}$) then so is the posterior belief $\tilde{\lambda_i}$, whose field $\log \frac{\lambda_i}{1-\lambda_i} + \frac{\mu}{\sigma^2}(h_i - \mu(\frac{1}{2} - p))$ is Gaussian because it is the sum of two independent Gaussian variables. Hence, we are now in a situation similar to the initial one, so we can continue and iterate the process - as long as the independence assumption is adequate.

[4] The following derivation also formalizes the use of of the signal-to-noise ratio in the previous $\{-1,+1\}$ case. For simplicity of exposition, we relied so far on its being well known.

4.2 The Optimality of the TS Function

After establishing that the s/n ratio determines the performance also in the firing/quiescent formulation, we now turn to prove

Theorem 3.

With restricted activity, the activation function $a_i = Max(\lambda_i - b, 0)$ maximizes the signal-to-noise ratio in the $\{0,1\}$ formulation, where b is determined by the activity level allowed.

Proof:

Proceeding in the same manner as in the proof of Theorem 1, we obtain the Lagrangean

$$E(h|\lambda) - \nu Var(h|\lambda) = \sum_j (\lambda_j - p)a_j - \mu_1 \sum_j {a_j}^2 VarZ - \mu_2 \sum_j a_j \qquad (18)$$

Note that a_i should be a monotonically increasing function of λ_i (otherwise exchanging two a_j's would preserve constraints and increase the value of the objective function). Unlike the $\{-1,+1\}$ case where we allowed a_j to have arbitrary real values, we choose to restrict a_j in the $\{0,1\}$ case to have non-negative values only, so as to maintain the analogy between the analog activation signal and the neuron's firing rate. Hence, we cannot proceed as in the $\{-1,+1\}$ case by differentiating the Lagrangean (18). Instead, we prove first

Lemma 3.1.

Rename the λ_j's so that $0 < \lambda_1 \leq \lambda_2 \leq \lambda_3 \ldots \leq \lambda_N$. Let $L > 0$, $S > 0$. The problem

$$Max \sum_{j=1}^{N} (\lambda_j - p)a_j$$

subject to

$$\sum_{j=1}^{N} a_j \leq L$$

$$\sum_{j=1}^{N} {a_j}^2 \leq S$$

$$a_i \geq 0 \ , \ \forall i \in \{1, 2, \ldots, N\}$$

is solved by

$$a_i = cMax(\lambda_i - b, 0) \qquad (19)$$

for some $c > 0$, $b \geq p$.

Proof:

It is clear that for any feasible solution with $a_i > 0$ and $\lambda_i < p$ for some i, if we replace this a_i by zero, the new solution is feasible as well and yields a higher objective value. Hence, $a_i = 0$ whenever $\lambda_i < p$. Since (a_j) is nondecreasing and not all a_j's are zero,

$$k = Min\{j | a_j > 0\}$$

is well defined. Defining the Lagrangean

$$\sum_{j=k}^{N} (\lambda_j - p) a_j - \nu_1 \sum_{j=k}^{N} a_j - \nu_2 \sum_{j=k}^{N} {a_j}^2 \tag{20}$$

and differentiating with respect to each a_j in this range, we obtain

$$2\nu_2 a_j = \lambda_j - p - \nu_1 \tag{21}$$

Let ${\nu_1}^* = \nu_1 + p$. If all λ_j's in this range are equal, so are the a_j's. Otherwise, $\nu_2 \neq 0$ and we see that $a_j = \frac{1}{2\nu_2}(\lambda_j - {\nu_1}^*)$. In both cases (equal or unequal λ_j's), the sequence (a_j) has the claimed pattern (19). □

The proof of Theorem 3 is not yet finished. We must show that if $1 < k < N$ then $\lambda_{k-1} \leq {\nu_1}^*$. Suppose, to the contrary, that $a_{k-1} = 0$ and ${\nu_1}^* < \lambda_{k-1} \leq \lambda_k$, as illustrated in figure 4.

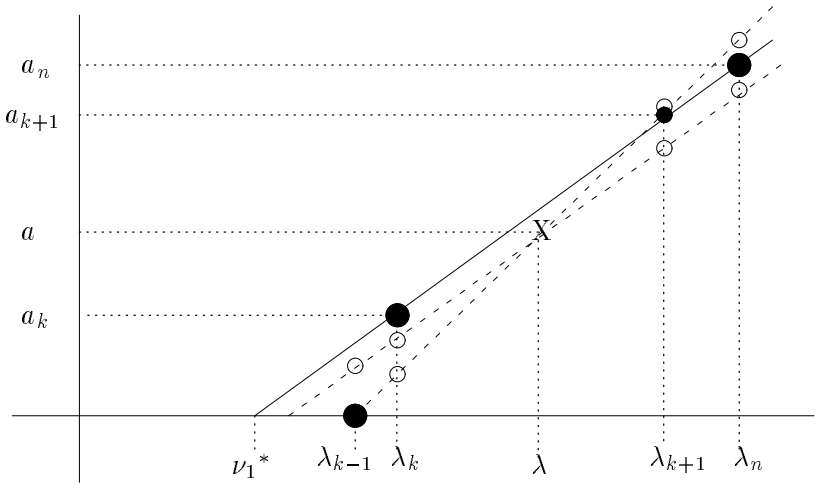

Figure 4: A sketch of the (hypothetical) situation where ${\nu_1}^* < \lambda_{k-1} \leq \lambda_k$.

Let (λ, a), denoted by X in figure 4, be the center of gravity of the set $\{(\lambda_{k-1}, a_{k-1}), (\lambda_k, a_k), \ldots, (\lambda_N, a_N)\}$ and let $L_\gamma(y) = a + \gamma(y - \lambda)$ be any linear function through (λ, a). If we define a new sequence a^γ by

$$a_j{}^\gamma = \begin{cases} 0 & j < k-1 \\ L_\gamma(\lambda_j) & j \geq k-1 \end{cases} \tag{22}$$

and restrict γ so that $L_\gamma(\lambda_{k-1}) \geq 0$ on one hand, while $\gamma \geq c$ on the other, i.e.,

$$c \leq \gamma \leq \frac{a}{\lambda - \lambda_{k-1}} = \delta \tag{23}$$

then

$$\sum_{j=k-1}^{N} a_j{}^\gamma = \sum_{j=k-1}^{N} a_j \tag{24}$$

(clear from linearity) and

$$\sum_{j=k-1}^{N} (a_j{}^c)^2 < \sum_{j=k-1}^{N} a_j{}^2 < \sum_{j=k-1}^{N} (a_j{}^\delta)^2 \tag{25}$$

which we will prove in Lemma 3.2 below. This proves that there is a value of γ for which

$$\sum_{j=k-1}^{N} a_j{}^\gamma = \sum_{j=k-1}^{N} a_j \tag{26}$$

and

$$\sum_{j=k-1}^{N} (a_j{}^\gamma)^2 = \sum_{j=k-1}^{N} a_j{}^2 \tag{27}$$

By Cauchy-Schwartz inequality, for given $\sum_j a_j$ and $\sum_j a_j{}^2$, the covariance $\sum_j a_j \lambda_j$ is maximal if there is a linear relation between the a_j's and λ_j's, such as there exists for $(a_j{}^\gamma)$ but not for (a_j). Hence, such an a_{k-1} could not have existed in the optimal solution. This concludes the proof of theorem 3, and shows that the TS is the optimal activation signal in the biologically-motivated $\{0, 1\}$ model.

To prove the inequalities between the sums of squares (25), we resort to a well known fact in economics of risk, whose proof is omitted.

Lemma 3.2

Let H and G be nondecreasing and let there be an x_0 such that $G(x) \leq H(x)$ for all $x < x_0$ and $G(x) \geq H(x)$ for all $x > x_0$. (H and G 'single cross', and G crosses H from below). Let $x_1 < x_2 < \ldots < x_n$ be such that

$$\sum_{j=1}^{n} H(x_j) = \sum_{j=1}^{n} G(x_j)$$

(but $H(x_i) \neq G(x_i)$ for some i). Then

$$\sum_{j=1}^{n}(H(x_j))^2 < \sum_{j=1}^{n}(G(x_j))^2 \ .$$

The reader can easily ascertain the conditions of Lemma 3.2 for $H = L_c$ and $G =$ (original sequence) as well as for $H =$ (original sequence) and $G = L_\delta$. □

4.3 The Dependence of the Optimal Signal on Activity

We shall now show that the form of the non-zero part of the optimal activation signal is insensitive to the activity level. To see this, normalize and rewrite the optimal activation signal function (for $x > 0$) as

$$q(x) = \frac{Max\left(\frac{1-e^{-2x}}{1+e^{-2x}} - d, 0\right)}{1-d} \tag{28}$$

to achieve $q(\infty) = 1$. Choose $x_0 = \frac{1}{2}\log\frac{1+d}{1-d}$ so that $q(x_0) = 0$, and consider

$$g_\beta(t) = q(x_0 + t) = \frac{1-e^{-2t}}{1+\beta e^{-2t}} \tag{29}$$

where $\beta = e^{-2x_0} = \frac{1-d}{1+d} \in [0,1]$. $g_1(t)$ corresponds to full activity, while $g_0(t)$ corresponds to the limiting case where the activity level approaches zero. All g functions are continuous, increasing and concave, satisfy $g(0) = 0$, $g(\infty) = 1$, and are bounded between $g_1(t)$ and $g_0(t)$ which are close to each other

$$\frac{1}{2}(1-e^{-2t}) \leq \frac{1-e^{-2t}}{1+e^{-2t}} = g_1(t) \leq g_\beta(t) \leq g_0(t) = 1-e^{-2t} \ . \tag{30}$$

An example of g, the non-zero part of the activation function, and its bounds is presented in figure 5. The finding that the non-zero part of the optimal activation function has an almost invariant shape yields the prediction that neuromodulatory changes that alter the excitability of cortical neurons may change the minimal threshold level of input current required to stimulate neural firing, but the dependency of the firing rate on the above-threshold input current should retain its overall form.

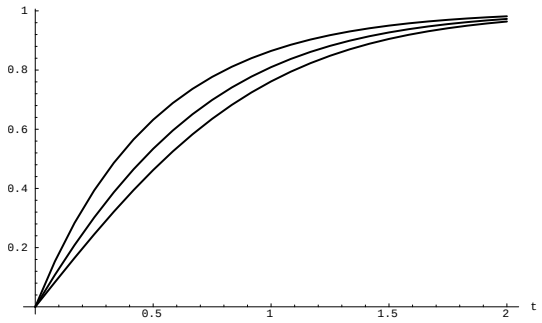

Figure 5: A plot of g (middle line, with $\beta = 0.5$) and its lower ($\beta = 1$) and upper ($\beta = 0$) bounds.

5 The Distribution of Firing Rates

The analog activation signal described for the $\{0,1\}$ formulation should be interpreted as the neuron's firing rate, averaged over some time interval which corresponds paradigmatically to the notion of 'iteration' of the ANN (i.e., a period of time in which, perhaps due to some 'inertial' effects, dynamics do not change by much). We have seen that for Gaussian fields, the neuron's activation is zero below some threshold (determined by the coding level p and by the desired activity level) and is given by the function g_β evaluated at the field value, when this value exceeds the threshold. We shall now study the distribution of field values which exceed a threshold, and the distribution of their corresponding activations.

Suppose that a neuron's field h, properly normalized, is standard normally distributed. We want to assess how are strong, above-threshold field values distributed. Using standard exponential asymptotics of normal distributions we have, for all $d > 0$ (threshold value) and $t > 1$ (incremental field value)

$$P\left(Z > d + \frac{t}{d} \mid Z > d\right) < e^{-t} \ . \tag{31}$$

In fact, the LHS of (31) converges to its RHS as $d \to \infty$. For example, let $t = 3$ and $d = 3$ (or, respectively, $d = 10$). Since $e^{-3} \approx 5\%$, of all field values above 3 (10) standard deviations, at least 95% of the probability mass is concentrated between 3 (10) and 4 (10.3) standard deviations. We see that the excess mass over a given threshold concentrates more and more near the threshold as it increases, and tends to be exponentially distributed. Since a linear homogeneous function of an exponentially distributed random variable is exponentially distributed, the observed (non-zero) firing rates should have an exponential distribution, accounting for the typical decreasing-density shape of cortical firing rates.

6 Summary

Our approach stems from the view that due to the evolutionary pressure of competition, successful species have evolved efficient survival mechanisms. Various criteria may have played part in shaping the structure and dynamics of human cortical memory systems, such as minimization of overall cortical connectivity length [Durbin and Mitchison, 1990, Ruppin *et al.*, 1993], or efficient blood supply requirements. In this paper, we have assumed

that dynamics governing the firing of neurons in cortical associative memory areas have evolved so as to maximize their performance, under the constraint of restricted neural activity levels reflecting limited ongoing metabolic resources. Modeling memory cortical modules as attractor neural networks, we have studied this optimization problem in sparse low-activity ANNs, where the neurons' input fields are uncorrelated. In this simplified modeling framework it is shown that the the form and distribution of cortical neurons' firing can be accounted for.

References

[Abeles *et al.*, 1990] 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.

[Abeles, 1991] M. Abeles. *Corticonics: Neural Circuits of the Cerebral Cortex*. Cambridge University Press, 1991.

[Amit and Tsodyks, 1991] 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.

[Amit, 1989] D.J. Amit. *Modeling brain function: the world of attractor neural networks*. Cambridge University Press, 1989.

[Avoli and Olivier, 1989] M. Avoli and A. Olivier. Electrophysiological properties and synaptic responses in the deep layers of the human epileptogenic neocortex in vitro. *J. Neurophysiol.*, 61:589–606, 1989.

[Derrida *et al.*, 1987] B. Derrida, E. Gardner, and A. Zippelius. An exactly solvable asymmetric neural network model. *Europhys. Lett.*, 4:167–173, 1987.

[Doya and Selverston, 1994] K. Doya and A.I. Selverston. Dimension reduction of biological neuron models by artificial neural networks. *Neural Computation*, 6:696–717, 1994.

[Durbin and Mitchison, 1990] R. Durbin and G. Mitchison. A dimension reduction framework for understanding cortical maps. *Nature*, 343:644–647, 1990.

[Englisch and Xiao, 1991] H. Englisch and Yegao Xiao. Neural networks as perpetual information generators. *Phys. Rev. A.*, 44(2):1382–1385, 1991.

[Ermentrout, 1994] B. Ermentrout. Reduction of conduction-based models with slow synapses to neural nets. *Neural Computation*, 6:679–695, 1994.

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

[Gustafsson and Wigstrom, 1981] B. Gustafsson and H. Wigstrom. Shape of frequency-current curves in ca1 pyramidal cells in the hippocampus. *Brain Res.*, 223:417–421, 1981.

[Hinton and Nowlan, 1990] G.E. Hinton and S.J. Nowlan. The bootstrap widrow-hoff rule as a cluster-formation algorithm. *Neural Computation*, 2:355–362, 1990.

[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.

[Kinzel, 1985] W. Kinzel. Learning and pattern recognition in spin glass models. *Z. Physik*, B60:205–213, 1985.

[Kononenko, 1989] I. Kononenko. Bayesian neural networks. *Biological Cybernetics*, 61:361–370, 1989.

[Lansner and Ekeberg, 1989] A. Lansner and O. Ekeberg. A one-layer feedback artificial neural network with a bayesian learning rule. *International journal of neural systems*, 1 (1):77–87, 1989.

[Mason and Larkman, 1990] A. Mason and A. Larkman. Correlations between morphology and electrophysiology of pyramidal neurons in slices of rat visual cortex: Ii electrophysiology. *J. Neurosci.*, 10:1415–1428, 1990.

[Meilijson and Ruppin, 1993] I. Meilijson and E. Ruppin. History-dependent attractor neural networks. *Network*, 4:195–221, 1993.

[Meilijson and Ruppin, 1994] I. Meilijson and E. Ruppin. Optimal signalling in attractor neural networks. *Network*, 5 (2):277–298, 1994.

[Miyashita and Chang, 1988] Y. Miyashita and H.S. Chang. Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, 331:68–71, 1988.

[Miyashita, 1988] Y. Miyashita. Neuronal correlate of visual associative long-term memory in the primate temporal cortex. *Nature*, 335:817–820, 1988.

[Ruppin *et al.*, 1993] E. Ruppin, E. Schwartz, and Y. Yeshurun. Examining the volume-efficiency of the cortical architecture in a multi-processor network model. *Biological Cybernetics*, 70:89–94, 1993.

[Treves and Rolls, 1991] A. Treves and E. T. Rolls. What determines the capacity of autoassociative memories in the brain? *Network*, 2:371–397, 1991.

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