A lateral inhibition neural network that emulates a winner-takes-all algorithm

B. Krekelberg & J.N. Kok

Cognitive Artificial Intelligence, Department of Computer Science,
Utrecht University, The Netherlands.
Email: joost@cs.ruu.nl

Abstract. We determine the conditions under which a lateral inhibition type network with binary neurons selects a unit which receives the maximal input. It is shown that the lateral interaction between two neurons needs to be tuned to a critical value. The resulting neural network is a local implementation of the first phase of the Kohonen learning algorithm.

The Kohonen algorithm [6] (which picks the unit with the highest activation and lets a cluster of neurons around it change their weights adaptively) cannot be a neurobiological mechanism for self-organisation. However, it is clear that having a single 'winner' and a surrounding region of neurons change their weights is needed in the context of self-organisation. Usually it is assumed that 'some kind of lateral inhibition phase' can emulate the winner-takes-all algorithm. However a thorough analysis of such networks has not been given for the general case. In [5] a network without inputs and with step-like neighbourhood function is shown to have at most one stable active region. In this paper we extend this work using techniques from [1, 2, 3] to the case with inputs and in which binary neurons are connected through weights that have the typical Mexican hat or difference-of-gaussians (dog) shaped position dependence. This function is formed by adding three functions: a positive gaussian with a maximum m_+ and a half-width σ_+, a gaussian with a negative minimum m_- and half-width σ_- and a negative constant w_∞. We consider one-dimensional networks with a neuron at every position x of the real axis. These networks are called neural *fields*, and are characterised by u (the input to the neuron), τ (the time scale), h (the bias), w (the weights between neurons in the output layer), s (the input a neuron receives from the input layer) and the general differential equation is given by:

$$\tau\frac{\partial u(x,t)}{\partial t} = -u(x,t) + \int_{-\infty}^{\infty} w(x-y)\mathcal{H}[u(y,t)]dy + h + s(x,t),$$

where $\mathcal{H}$ is the Heaviside or step function. Because the weight matrix w is symmetric it is well-known that the network converges to a stable point if it is updated asynchronously [4]. We determine the stationary states in the absence of inputs first. In the absence of inputs the differential equation reduces to:

$$\tau\frac{\partial u(x,t)}{\partial t} = -u(x,t) + \int_{R[u]} w(x-y)dy + h \text{ where } R[u] = \{x|u(x) > 0\}.$$

Furthermore define $W(x) = \int_0^x w(y)dy$, $W_\infty = \lim_{x \to \infty} W(x)$, $W_m = \max W(x)$, let a_m be such that $W(a_m) = W_m$ and let $a_0 > 0$ be such that $W(a_0) = 0$. Stable, stationary solutions of the differential equation have $\frac{\partial u}{\partial t} = 0$ and are therefore functions of position x only. We distinguish three types of stable solutions $u(x)$: ∞-**solutions** if all neurons in the field are active, $\emptyset$-**solutions** if there is no activity in the field, a-**solutions** or **local excitations** if activity in the field extends over a region of length a. Amari has proven the following Theorem [1]:

Theorem 1 *For the differential equation of the lateral interaction neural field in the absence of inputs, an ∞-solution exists if and only if $2W_\infty > -h$, an $\emptyset$-solution exists if and only if $h < 0$, an a-solution exists if and only if $h < 0$ and $W(a) + h = 0$.*

The condition for an a-solution or *local excitation* to exist determines the length a of this solution as follows. For every negative h two solutions with lengths respectively a_1 and a_2 exist. It can be shown, however, that only the larger one $a_2 > a_m$ is stable [1]. Furthermore, because $h < 0$, stable a-solutions only occur for $W(a) > 0$ and therefore $a < a_0$. Concluding, the length a of a stable region satisfies: $a_m \leq a < a_0$.

Now suppose exactly one bubble of length a has formed at the interval $(-a, 0)$ and gives a lateral input to all neurons in the field. This input is given by $\tilde{s}(x) = \int_{-a}^0 w(x-y)dy = W(x+a) - W(x)$. The shape of $\tilde{s}(x)$ can be determined from figure 1. The neural field can be split in two parts: the *far field* for which
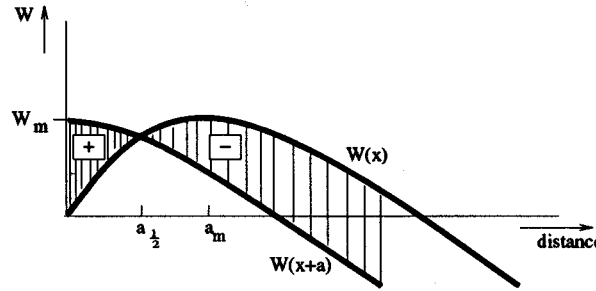

Fig. 1. $W(x)$ and the shifted $W(x + a)$ are sketched to determine $\tilde{s}(x)$.

$\tilde{s}(x) < 0$ (that is $x > a_{\frac{1}{2}}$) and the *near field* for which $\tilde{s}(x) \geq 0$ (the length of the near field we denote by $a_{\frac{1}{2}}$). In the far field the effect of $\tilde{s}$ is to give a negative input to all the neurons. This input can be treated as an addition to the negative bias h. We have a new bias $h'(x) = h + \tilde{s}(x)$. Following Theorem 1 the condition for a local excitation of length a' to occur in the far field now becomes: $W(a') = -h'(x) = -(h + \tilde{s}(x))$. We visualise this condition in figure 2. From this figure it is clear that if h is chosen close enough to W_m *any* negative addition to the bias will prevent the formation of a new excitation in the far field.

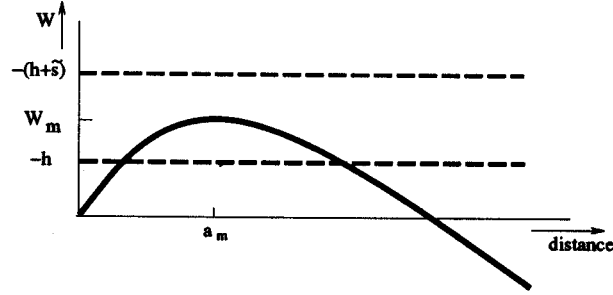**Lemma 2** *If $h = -W_m$ two local excitations in each others far field are not*

10

Fig. 2. Shifting of the bias by the lateral input

stable.

Proof Assume there are two local excitations, one at the origin of length a_m and the other in the far field and of length b: $R[u] = \{x| - a_m < x < 0\} \cup R_f[u]$ where $R_f[u] = \{x|a_{\frac{1}{2}} < (x_f - b) \le x \le x_f\}$. The active neuron at the largest distance from the origin is at $x = x_f$, because of the continuity condition this neuron should receive a total input equal to zero:

$$u(x_f) = \int_{R[u]} w(x_f - y)dy + h = \tilde{s}(x_f) + W(b) - W_m = 0.$$

This can be rewritten to obtain a demand on $W(b)$: $W(b) = -\tilde{s}(x_f) + W_m$. Furthermore, we know that W is bounded by W_m: $W(b) = -\tilde{s}(x_f) + W_m \le W_m$ and therefore $\tilde{s}(x_f) \ge 0$. Contradiction. $\square$

In contrast to the situation in the far field, in the near field the extra input effectively *decreases* $-h$ to $-h'(x)$. In this case another excitation might be possible. However:

Lemma 3 *If $h = -W_m$ a local excitation of length a_m is at $(-a_m, 0)$, the near field is too small to accommodate a stable local excitation of non-zero length.*

Proof Assume there are two local excitations: one at the origin of non-zero length a_m and another of length a' in the near field: $R[u] = \{x| - a_m < x < 0\} \cup R_n[u]$ where $R_n[u] = \{x|0 < x \le a_{\frac{1}{2}} \wedge u(x) \ge 0\}$. We already know that to be stable $a_m \le a' < a_0$. The length of the near field ($a_{\frac{1}{2}}$), can be derived form the shape of the lateral interaction. The function W is a decreasing function for $a \ge a_m$ and increasing for $0 < a < a_m$. Therefore a function $W' = W(x + a_m)$ is decreasing on the positive axis and always smaller than the decreasing part of W. From this it follows that the intersection $a_{\frac{1}{2}}$ of W and W' lies in the region where W is increasing and therefore $a_{\frac{1}{2}} < a_m$. Combining these two facts gives $a_{\frac{1}{2}} < a_m \le a'$. Contradiction. $\square$

Theorem 4 *If the bias h of a lateral interaction type neural field without inputs*

is chosen to be equal to $-W_m$, only one local excitation can exist and if it exists it will have a length a_m.

This theorem is a generalisation of a result in [5] where it is shown that in the absence of inputs and with a sufficiently strong step-like lateral inhibition, a neural network can have only one active region. Here the interaction is no longer step-like but **dog** shaped and therefore closer to neurobiological reality.

Now we generalise further to the case where inputs are present. First we define the sets of inputs we are going to consider. The inputs are bounded by a threshold and the height is coupled to the width of a pattern. The latter condition is imposed to circumvent a problem caused by the use of the Heaviside function as an activation function. The step-function sets the activation of a neuron to one if the stimulus is above some threshold and thereby part of the information in the input is lost. By restricting ourselves to scale-invariant patterns we attain a situation where the width of the excited region carries information about the height of the stimulus.

Definition 5 *A set of continuous inputs s_i is called scale-invariant with respect to a threshold h if each input s_i has only one maximum at x_i, $\forall i,j : s_i(x_i) \leq s_j(x_j) \Rightarrow \forall x : s_i(x_i + x) \leq s_j(x_j + x)$, and $\forall x : \sum_i s_i(x) \leq |h|$.*

Furthermore we will use the following lemma from [2]:

Lemma 6 *The active region of a lateral interaction output layer contains the neuron that receives maximal input from the input layer.*

Proof First, consider a neural field with one local excitation extending from x_1 to x_2. Stable equilibrium ($\frac{\partial u}{\partial t} = 0$) requires from the general differential equation (with inputs) that $u(x)$ satisfies: $u(x) = \int_{x_1}^{x_2} w(x - y)dy + s(x) + h$ and because of continuity of u we have: $u(x_1) = u(x_2) = 0$. This gives: $W(x_1 - x_2) + s(x_1) = -W(x_2 - x_1) + s(x_2)$, which reduces to $s(x_1) = s(x_2)$. This shows that an excited region surrounds an extremum of $s(x)$. Secondly, we give an informal argument that a region surrounding a minimum of $s(x)$ is not stable. Consider neurons which are close to neurons which receive a large input s. These neurons have a tendency to become active because they are excited by nearby neurons. Neurons that are further away will have a lower and even negative lateral input and can be de-activated. This explains the lability of a minimum: a region of active neurons which contains a minimum of the input. is perturbed: some neurons on one side of the region are activated and some on the other side are de-activated. As a consequence more neurons close to high inputvalues become active and through lateral inhibition they de-activate the neurons with lower inputvalues. This looks as if the region *moves* towards a maximum of the input.

12

□

We can now formulate the main theorem:

Theorem 7 *If the parameters of the difference-of-gaussians lateral interaction strength of a neural field with binary neurons are chosen such that*

$$W(2a_m + a_{\frac{1}{2}}) - W(a_{\frac{1}{2}} + a_0) + W(a_m) < 0 \text{ and } h = -W(a_m)$$

the neural field emulates a winner-takes-all algorithm for scale-invariant inputs.

Proof Assume there is one local excitation covering the interval $[-l_0, 0]$ due to an input s_0. We investigate the condition for the formation of a stable local excitation of length l_1 in the far field and due to an input s_1 : $R[u] = \{x| -l_0 < x < 0\} \cup R_f[u]$ where $R_f[u] = \{x|x_f - l_1 \le x \le x_f\}$. The active neuron at the largest distance from the origin is at $x = x_f$ it receives a total input equal to:

$$
\begin{aligned}
u(x_f) &= \int_{R[u]} w(x_f - y)dy + h + s_0(x_f) + s_1(x_f) \\
&= W(x_f + l_0) - W(x_f) + \int_{R_f[u]} w(x_f - y)dy - W(a_m) + s_0(x_f) + s_1(x_f) \\
&= W(x_f + l_0) - W(x_f) + W(l_1) - W(a_m) + s_0(x_f) + s_1(x_f).
\end{aligned}
$$

We determine an upper bound to this expression by taking $R_f[u]$ to be the region of neurons that gets the highest lateral input. This is the region closest to the origin but still in the far field. In this case the critical neuron is a distance l_1 away from the near field $x_f = a_{\frac{1}{2}} + l_1$. All neurons at the far-end of an active region which lies further into the far field will receive less lateral input. Furthermore we use one of the conditions for scale invariant inputs: $\forall x : \sum_i s_i(x) \le |h| = W(a_m)$. The bound is:

$$u(x_f) \le W(l_0 + a_{\frac{1}{2}} + l_1) - W(a_{\frac{1}{2}} + l_1) + W(l_1).$$

Further simplification is obtained by using the fact that stable solutions always have $a_m \le l \le a_0$:

$$u(x_f) \le W(2a_m + a_{\frac{1}{2}}) - W(a_{\frac{1}{2}} + a_0) + W(a_m),$$

and therefore $u(x_f) < 0$ which shows that no neuron in the far field can be active. Just as is the case without inputs, the near field is too small to accommodate a second local excitation of length $a > a_m$. This shows that there can never be more than one region active in the neural field. Together with Lemma 6 this shows that there is a single active region and it surrounds a neuron receiving maximal input. To show that this maximum is a global one, we note that the highest scale-invariant input is also the broadest. This means that in the course of an asynchronous, random and fair updating process the region receiving the

globally maximal input will be the first to become active and thereby prevent others from becoming active. □

The demand on $W(x)$ of Theorem 7 states a relation between the excitatory and inhibitory part of the interaction function w. This demand can always be met by a dog-shaped function by tuning the five parameters. Concluding one can say that we have shown that it is possible, but by no means trivial, to build a neural net with binary neurons that can select a 'winner'. To achieve this result with a Heaviside activation function we had to resort to scale-invariant input patterns to obtain an implicit relation between excitatory and inhibitory lateral weights. In further work we would like to extend the above analysis to a neural field with continuous activation values to make the demands on the interaction function more explicit and to be able to use more general input patterns. Secondly, we want to study the effect of the lateral inhibition phase on the ordering of the self-organising network. It is hoped that some kind of ordering proof along the lines of [7, 8] can be given.

References

[1] S-I. Amari. *Biological Cybernetics*, 27:77–87, 1977.

[2] S-I. Amari. *Bulletin of Mathematical Biology*, 42:339–364, 1980.

[3] S-I. Amari. *Research notes in neural computing*, vol. 1, pages 15–34.

[4] J.J. Hopfield. *Proc. Nat.Ac.Sc.*, 81:3088–3092, 1984.

[5] H. Ritter, T. Martinetz, and K. Schulten. *Neural Computation and Self-Organizing Maps*. Addison-Wesley, 1991.

[6] T. Kohonen *Biological Cybernetics* 43:59-69, 1982

[7] M. Cottrell and J.-C. Fort *Annales de l'institute Henri Poincaré*, 23:1-20, 1987

[8] E.Erwin and K. Obermayer and K. Schulten,*Biological Cybernetics* 67:47-55, 1992.