How Lateral Interaction Develops in a Self-Organizing Feature Map

Joseph Sirosh and **Risto Miikkulainen**
Department of Computer Sciences
The University of Texas at Austin, Austin, TX-78712
sirosh,risto@cs.utexas.edu

Abstract— A biologically motivated mechanism for self-organizing a neural network with modifiable lateral connections is presented. The weight modification rules are purely activity-dependent, unsupervised and local. The lateral interaction weights are initially random but develop into a "Mexican hat" shape around each neuron. At the same time, the external input weights self-organize to form a topological map of the input space. The algorithm demonstrates how self-organization can bootstrap itself using input information. Predictions of the algorithm agree very well with experimental observations on the development of lateral connections in cortical feature maps.

I. INTRODUCTION

Two-dimensional topological maps of sensory input are present in various cortices of the brain. They are believed to develop in a self-organizing process based on cooperation and competition between neurons [4, 13, 14]. The Self-Organizing Feature Map (SOFM) algorithm [5, 6] is a computational model of this process. The SOFM algorithm has been applied, for example, into modeling the development of retinotopy, ocular dominance and orientation preference in the visual cortex and somatotopy in the somatosensory cortex [9, 10, 11].

The SOFM algorithm is an abstraction, though biologically inspired. At each step of training, the algorithm finds the neuron whose input synaptic weights are closest to the current input vector and changes the input weights of all neurons in its neighborhood towards the input vector. The size of the neighborhood starts out large, but gradually decreases towards the end of training. The algorithm relies on an external supervisor to find the maximally active unit, and invokes an ad-hoc schedule for decreasing the neighborhood size.

To be biologically realistic, the SOFM algorithm should be reduced to local computations and interactions among neurons of the map. Proposed low-level models of SOFM assume cooperative and competitive lateral interactions through excitatory and inhibitory connections [5, 8]. The lateral connections are non-modifiable and distinct from the external input connections. The connection weight profile in these models is shaped like a "Mexican hat", with short-range excitation and long-range inhibition. Similarly shaped lateral interaction is commonly found in many biological neural networks [5].

How does such lateral interaction arise? Enormous amounts of genetic information would be required to specify each synaptic weight of every neuron in a cortical map. Therefore, it is unrealistic to expect lateral interaction in such networks to be predetermined. All connections of a neuron should be modifiable, and there is no reason why the lateral interaction should have a uniform shape everywhere in the map. It makes sense to assume that the connections initially have random initial weights within a predetermined range. The question is, can these connections self-organize to form global order? Do the random-weight lateral connections develop a biologically realistic profile?

In this paper, we demonstrate through simulations that lateral connections can self-organize simultaneously with external input connections. In the process, the lateral interaction profile becomes a smooth "Mexican hat"-shaped function. The shape varies smoothly from neuron to neuron in the map depending on location. All connections can start out with random weights, and all connections are modified through a version of the Hebb learning rule.

II. THE SELF-ORGANIZING PROCESS

The computations for a self-organizing feature map with lateral connections are described below. The algorithm computes the activity of each neuron in a network as a weighted sum of the external input and refines the activity through lateral interactions between neurons. When the activity stabilizes, all connection weights are modified. The process is repeated for each input. Section A explains and motivates our neuron model. Sections B and C describe the network and the input, and section D delineates

the equations and the computations.

A. The Neuron Model

Each neuron in the network is assumed to have three sets of inputs:

- excitatory input connections that supply external input to the neuron,

- short-range lateral excitatory connections from close neighbors in the map,

- long-range lateral inhibitory connections from within the map.

A connection has a characteristic strength (or weight), which may be any value between zero and a prescribed limit. In a real neuron, these limits would be a property of the synapse.

External inputs to primary cortical areas of the brain synapse differently from intracortical lateral connections [15]. It is possible that the external and lateral connections of the same neuron obey two different rules of weight modification. In our model, the two rules differ only in the normalization. The external connections are normalized to hold the sum of squares of the input weights constant, and the lateral (excitatory/inhibitory) connections are normalized to keep the sum of lateral (excitatory/inhibitory) weights constant.

In the primary cortices, most extrinsic afferents synapse in the dendritic spines of neurons [15]. The dendritic shafts sum the input from the spines approximately linearly. A rule of weight modification proposed by [12] appears realistic when applied to theses synapses. Oja's rule is an approximation of a Hebbian rule in which the synaptic weights are normalized to hold the sum of squares of the weights constant. We use the more general Hebbian rule (5) to modify the external input connections.

In a real neuron, excitatory and inhibitory synaptic transmission are mediated by different neurotransmitters and receptors. The two sets of synapses also have different morphology [15]. A neuron processes each set at different sites with different receptors and secondary messengers, and these resources are limited. It is reasonable to assume that the total synaptic weight for each set is fixed. When lateral connection weights are modified in our model, they are normalized so that the total excitatory weight and the total inhibitory weight are constant (refer to equation 3).

B. The Network

The feature map is a two dimensional $N \times N$ grid of neurons (Fig. 1). Each neuron connects to its neighbors within distance d with excitatory lateral connections and
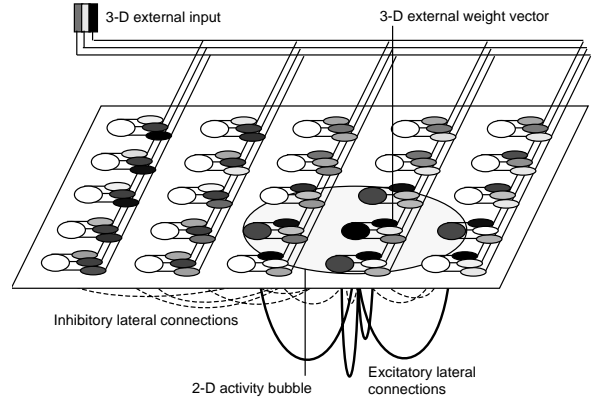

Fig. 1: **The laterally connected self-organizing feature map architecture.** Each neuron receives the same 3-dimensional input vector and computes an initial response based on the similarity with its external weight vector. The response then evolves through propagation along the lateral connections (only a few connections of the most strongly active unit are shown). After a stable activity bubble has formed, weights of the active units are adapted.

to all neighbors within $3d + 1$ with inhibitory connections. Lateral excitation weights are uniformly randomly distributed in the interval $(0, \gamma_e)$ within the excitation radius and are zero outside. Similarly, negative inhibition weights are distributed uniformly in the interval $(\gamma_i, 0)$ within the inhibition radius and are zero outside.

C. The Inputs

The input vectors must be normalized to prevent vectors with large norms from dominating the self-organizing process [8]. For this reason, the 2-D square area used as input in the simulations was laid on the surface of a unit sphere and represented in spherical coordinate system. In effect, such inputs are still 2-dimensional because the radius is constant. Each spherical input vector $(x_1, x_2, 1)$, $(-0.5 \leq x_1, x_2 \leq 0.5)$ was then transformed into a 3-dimensional cartesian vector $x = (\xi_1, \xi_2, \xi_3)$:

$$\begin{cases} \xi_1 = 1 \cdot \cos(x_1) \cos(x_2), \\ \xi_2 = 1 \cdot \sin(x_1) \cos(x_2), \\ \xi_3 = 1 \cdot \sin(x_2). \end{cases} \quad (1)$$

Corresponding to the three input components, each neuron (i, j) has three external input weights $\mu_{ij,h}$, $h = 0, 1, 2$.

D. The Computations

The external and lateral weights are organized through a purely unsupervised learning process. Input items are randomly drawn from the input distribution and presented to the network one at a time. At each training step, the neurons in the network start out with zero activity. The
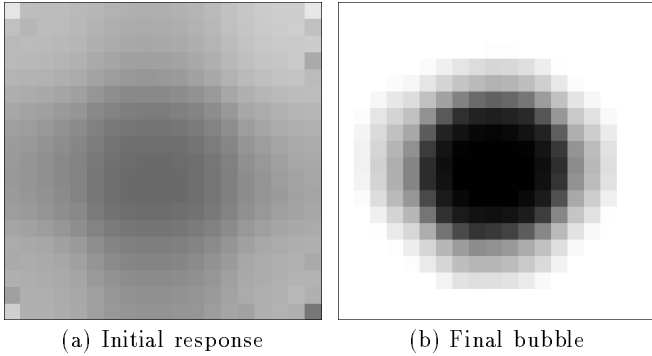
(a) Initial response   (b) Final bubble

Fig. 2: **Focusing the response through lateral inhibition.** The darkness of each square indicates the activity level of the corresponding unit in a 20×20 map.

initial response of each neuron η_{ij} in the map is based on a scalar product of the input and weight vectors:

$$\eta_{ij} = \sigma \left(\sum_h \mu_{ij,h} \xi_h \right), \qquad (2)$$

where the function σ is the familiar sigmoid activation function. The response evolves over time through lateral interaction. At each time step, the neuron combines external activation with lateral excitation and inhibition according to

$$\eta_{ij}(t) = \sigma \, (\; \sum_h \mu_{ij,h} \xi_h + \\ \sum_{k,l} E_{kl,ij} \eta_{kl}(t - \delta t) + \qquad (3) \\ \sum_{k,l} I_{kl,ij} \eta_{kl}(t - \delta t) \;) ,$$

where $E_{kl,ij}$ is the excitatory lateral connection weight on the connection from unit (k, l) to unit (i, j), $I_{kl,ij}$ is the inhibitory connection weight, and $\eta_{kl}(t - \delta t)$ is the activity of unit (k, l) during the previous time step. The activity pattern starts out as diffuse spread over a substantial part of the map and converges iteratively into a stable focused patch of activity, or activity bubble (Fig. 2). After the activity has settled, typically in a few iterations, the connection weights of each neuron are modified.

The lateral weights are modified by a Hebb rule, keeping the sum of weights constant:

$$\gamma_{ij,kl}(t + \delta t) = \frac{\gamma_{ij,kl}(t) + \alpha_L \eta_{ij} \eta_{kl}}{\sum_{kl} [\gamma_{ij,kl}(t) + \alpha_L \eta_{ij} \eta_{kl}]}, \qquad (4)$$

where η_{ij} stands for the activity of the unit (i, j) in the settled activity bubble, the γs are the lateral interaction weights ($E_{ij,kl}$ or $I_{ij,kl}$) and α_L is the learning rate for lateral interaction (α_E for excitatory weights and α_I for inhibitory). The larger the product of pre- and post-synaptic activity $\eta_{ij} \eta_{kl}$, the larger the weight change.

The external input weights are modified according to the normalized Hebbian rule:

$$\mu_{ij,h}(t + \delta t) = \frac{\mu_{ij,h}(t) + \alpha \eta_{ij} \xi_h}{\left\{ \sum_h \left[\mu_{ij,h}(t) + \alpha \eta_{ij} \xi_h \right]^2 \right\}^{1/2}}, \qquad (5)$$

which is otherwise similar to (4), but maintains the sum of squares of external weights constant.

Note that each computation is local to an individual neuron and its connections. The algorithm carries out local computations to achieve global self-organization of the map network. It does not require an external supervisory process.

III. Simulation Results

The algorithm was applied into learning the 2-D structure of a uniform distribution on a square area. The simulations were performed on the Cray Y–MP 8/864 at the University of Texas Center for High-Performance Computing. Fig. 3 illustrates the external input weights of the neurons transformed back to the original spherical coordinates. The weight vectors are initially uniformly distributed on the input square. As the simulation progresses, the network unfolds, and the weight vectors spread out to form a regular topological map and to cover the input space. The course of this process is very similar to the standard SOFM algorithm. After a while, the network reaches a stable equilibrium. Further training causes small fluctuations about the stable equilibrium if the learning rate α is nonzero.

The lateral interconnection weights started random, but evolved into a very smooth "Mexican hat" profile around each neuron. Figures 4 and 5 show the lateral weights before and after training for a neuron at the center of the map. The lateral weights converged faster than the input weights, leading the self-organization. Both the input weights and lateral weights reached a stable dynamic equilibrium after sufficient training.

Interestingly, the lateral interaction profile is not uniform throughout the map, but exhibits boundary effects (Fig. 6). Because the units near the boundary do not have full neighborhoods, normalization of lateral weights results in profiles that are taller and asymmetric. The asymmetry is important because it affects the shape and location of activity bubbles. In simulations with uniform, predetermined interaction [8], activity bubbles formed away from the boundary even for input best stimulating a boundary neuron. In other words, the bubbles were representing the location of the best stimulus on the map incorrectly. In the new model, maximally active areas of the bubble correspond to best input response areas of the map even at the boundary. The shape of the activity bubble resembles the initial input activity, and the profile of
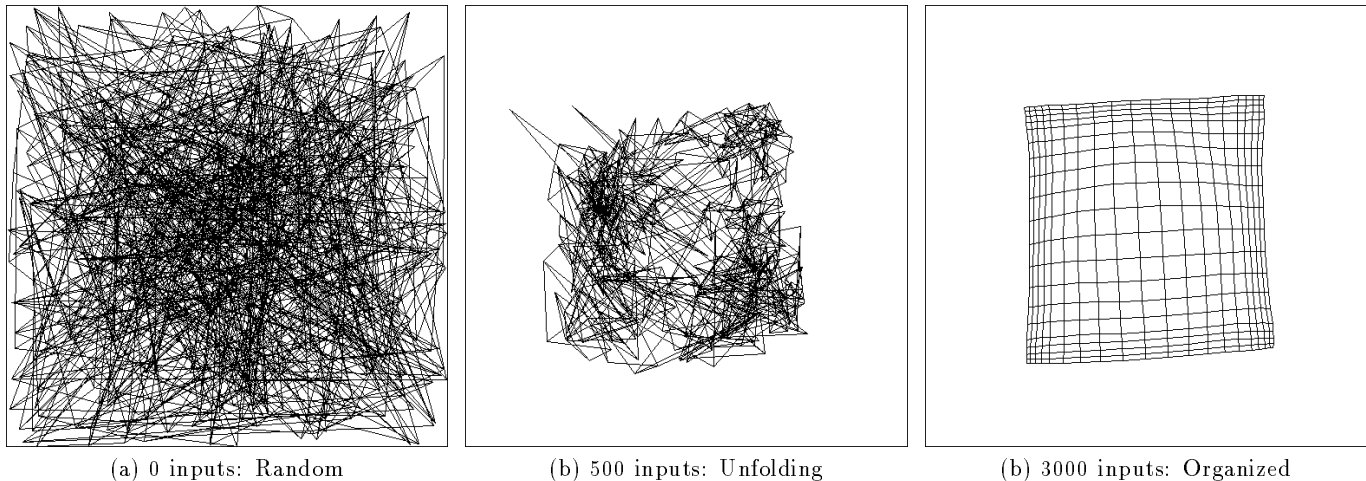
(a) 0 inputs: Random   (b) 500 inputs: Unfolding   (b) 3000 inputs: Organized

Fig. 3: **Self-organization of external input weights of a** 20×20 **map.** The square areas correspond to the input space. Each neuron is represented by its input weight vector plotted in 2D as a point inside the input square. Each neuron is connected to its four immediate neighbors, and the resulting grid depicts the topological organization of the neurons. An equilibrium is reached in about 3000 inputs. Additional inputs cause little change.

activity is very smooth within the bubble. The lateral interaction adapted to capture the smooth distribution of activity patterns on the map.

In the abstract feature map algorithm, metastable states such as twisted maps can form if the neighborhoods are initially too small [1]. The size of the activity bubble corresponds to the neighborhood size. The bubble size is determined by the radius and amount of lateral interaction. If these parameters are not large enough to make initial activity bubble sizes comparable to the size of the map, metastable states may form. The algorithm is robust within a fair range of parameters, and appropriate values can be easily determined.

IV. Discussion

The new self-organization process described above has several important implications. It is a simulation of a realistic physical system that self-organizes based on purely local rules. If we were to actually construct "neurons" with properties described above and form a network as explained, the map would self-organize completely in parallel, without global supervision. This is in contrast to the abstract SOFM process, where the maximally responding unit is chosen through global supervision, and adaptation neighborhoods are reduced according to a preset schedule.

The shape of the lateral interaction is automatically extracted from the statistical properties of the external input. At each input presentation, the input vector is analyzed by the current state of the input weights and represented as an activity bubble on the map. The distribution of bubbles over time shapes the lateral weights, which in turn results in tighter and smoother activity bub-

bles and facilitates the self-organization of the external weights. The self-organization thus "bootstraps" by using external input information to establish the necessary cooperative and competive interactions.

Standard feature maps represent the topology of the input space by the network's grid-like layout. Units close by in the map represent input vectors nearby in the input space, and vice versa. In our model, the neighborhood relations are mediated by the lateral connections, and the model is not limited to strictly 2-dimensional topology. Units that are far apart on the map grid may still belong to the same neighborhood if they are strongly connected laterally. Such topologies are automatically learned as part of the self-organizing process. If several areas of the map are simultaneously active, long range connections between such areas will remain strong. These connections cause the units to behave as if they were neighbors on the map. This property is potentially very significant in representing complex high-dimensional input spaces. While lower-level sensory representation in the brain seems to be organized in 2-dimensional maps (such as retinotopic maps), it is possible that higher representations make use of long-range lateral connections to represent more complex similarity relationships [13]. The laterally connected feature map is a potential computational model for formation of such representations.

Standard SOFM has been used to model the development of input connections to neurons in the primary sensory cortices [9, 10, 11]. With the new algorithm, it should be possible to model the development of both lateral and external input connections in sensory cortices.

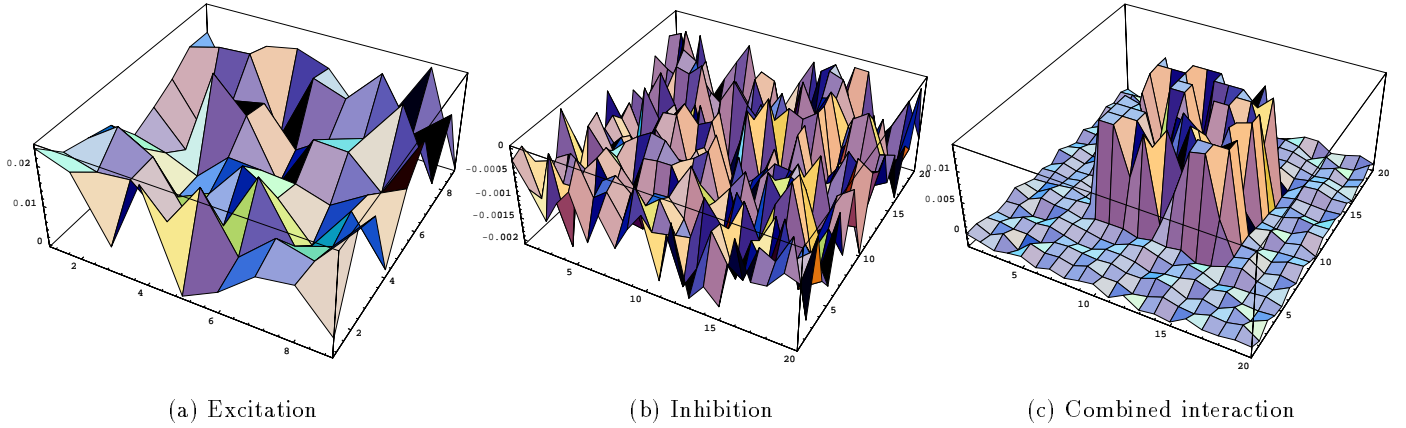(a) Excitation   (b) Inhibition   (c) Combined interaction

Fig. 4: **Initial lateral interaction.** The lateral excitation and inhibition weights and the combined interaction profile are plotted for the neuron at position $(10, 10)$ in the 20×20 map.

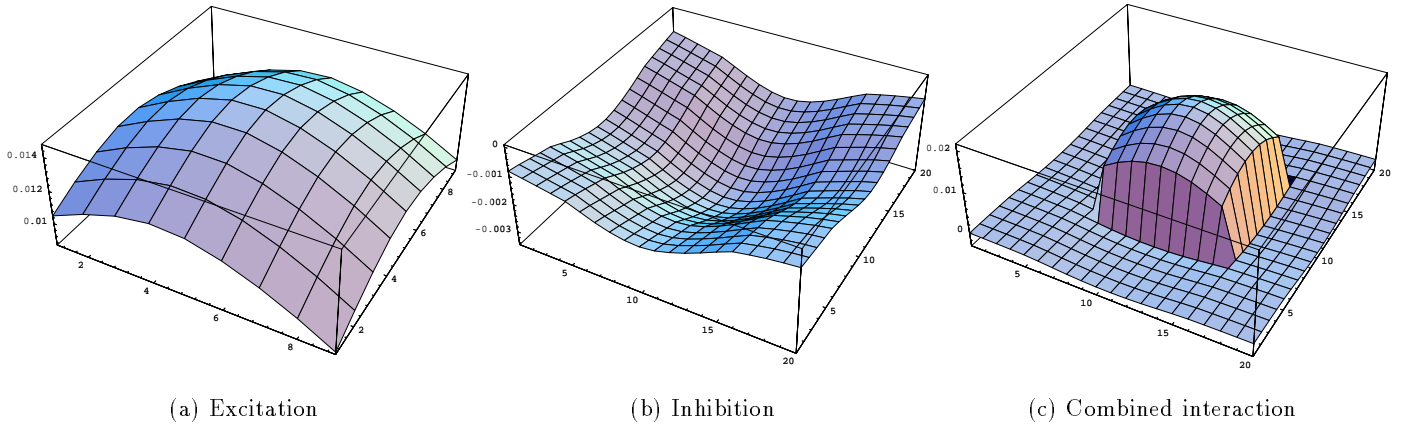(a) Excitation   (b) Inhibition   (c) Combined interaction

Fig. 5: **Final lateral interaction.** A smooth pattern of excitation and inhibition weights has evolved, resulting in a smooth interaction profile.

It has recently been found that horizontal connections in the primary visual cortex mainly connect areas with similar functional properties, such as the same orientation sensitivity or ocular dominance [2, 3]. Assuming that average visual input excites similar feature detectors simultaneously, our model could give a computational account to this phenomenom. Specifically, (1) lateral connections between similar feature detectors strengthen due to correlated activity and (2) connections between dissimilar feature detectors weaken due to limited synaptic resources (through normalization). If weak connections are assumed to die off, the surviving connections would be those that link areas with similar functional properties.

The survival of horizontal connections in the primary visual cortex depend on correlated neuronal activity. If visual input to the cortex from both eyes of the cat is decorrelated by artificially introducing strabismus (squint-

eye) during development, lateral connections preferentially connect ocular dominance columns of the same eye [7]. In normal cats however, the connections prefer orientation columns of the same orientation specificity (as explained above). These results could be explained by the laterally connected model. In strabismic visual input there are more correlations among inputs from the same eye than between eyes. Ocular dominance columns representing the same eye would have highly correlated activity, and connections between them should strengthen. Other lateral connections should weaken due to normalization and eventually die off. Normally, images in the two eyes are very similar on the average, and have significant correlations. Similar orientation detectors in the cortex would have highly correlated activity irrespective of their ocular dominance. The stronger connections should then run between orientation columns of similar specificity.
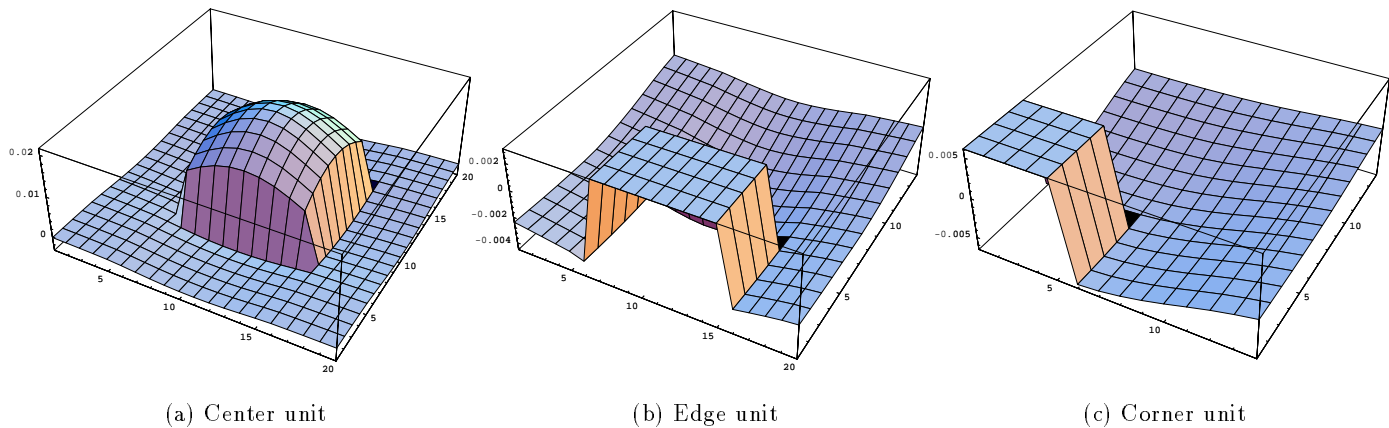
| (a) Center unit | (b) Edge unit | (c) Corner unit |

Fig. 6: **Lateral interaction profiles at various locations of the** 20×20 **map after training:** The profiles of neurons at $(10, 10)$, $(0, 10)$, and $(0, 0)$ are plotted. The profiles are tallest at the corner, shorter at an edge and shortest at the center. This is due to redistribution of synaptic weight by normalization.

V. Conclusion and Future Work

Self-organizing models of cortical map development assume lateral interactions to be predetermined. We are not aware of other published work modelling the development of lateral connections in feature maps, biological or abstract. The algorithm presented here demonstrates simultaneous development of lateral interaction and self-organization of input weights. It is biologically motivated, and its predictions tie in very well with experimental observations in visuo-cortical maps.

Currently, the algorithm does not incorporate a mechanism to reduce the extent of lateral connections. The lateral connections of each neuron cover a substantial part of the map. This keeps the neuronal activity correlated over large distances and only the most significant variances in input (the ones that overcome the threshold of correlation; [9]) become represented. To capture features with lower variance, the extent of lateral connections must decrease gradually. We are currently working on an unsupervised mechanism for pruning lateral connections automatically. Research is also underway into constructing general models of cortical map development based on our algorithm.

References

[1] E. Erwin, K. Obermayer, and K. J. Schulten. "Convergence properties of self-organizing maps," *Proceedings of the International Conference on Artificial Neural Networks* (Espoo, Finland). Amsterdam; New York: North-Holland, 1991, pp. 409–414.

[2] C. D. Gilbert, J. A. Hirsch, and T. N. Wiesel. "Lateral interactions in visual cortex," *Cold Spring Harbor Symposia on Quantitative Biology, Volume LV.* Cold Spring Harbor Laboratory Press, 1990, pp. 663–677.

[3] C. D. Gilbert and T. N. Wiesel. "Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex," *Journal of Neuroscience*, 9:2432–2442, 1989.

[4] D. H. Hubel and T. N. Wiesel. "Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat," *Journal of Neurophysiology*, 28:229–289, 1965.

[5] T. Kohonen. *Self-Organization and Associative Memory*, 3rd ed. New York: Springer, 1989.

[6] T. Kohonen. "The self-organizing map," *Proceedings of the IEEE*, 78:1464–1480, 1990.

[7] S. Lowel and W. Singer. "Selection of intrinsic horizontal connections in the visual cortex by correlated neuronal activity," *Science*, 255:209–212, 1992.

[8] R. Miikkulainen. "Self-organizing process based on lateral inhibition and synaptic resource redistribution," *Proceedings of the International Conference on Artificial Neural Networks* (Espoo, Finland). Amsterdam; New York: North-Holland, 1991, pp. 415–420.

[9] K. Obermayer, G. G. Blasdel, and K. J. Schulten. "Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps," *Physical Review A*, 45:7568–7589, 1992.

[10] K. Obermayer, H. J. Ritter, and K. J. Schulten. Large-scale simulation of a self-organizing neural network. In *Proceedings of the International Conference on Parallel Processing in Neural Systems and Computers (ICNC)*. New York: Elsevier, 1990.

[11] K. Obermayer, H. J. Ritter, and K. J. Schulten. "A principle for the formation of the spatial structure of cortical feature maps," *Proceedings of the National Academy of Sciences, USA*, 87:8345–8349, 1990.

[12] E. Oja. "A simplified neuron model as a principal component analyzer," *Journal of Mathematical Biology*, 15:267–273, 1982.

[13] C. von der Malsburg and W. Singer. "Principles of cortical network organization," P. Rakic and W. Singer, Eds., *Neurobiology of Neocortex*, New York: Wiley, 1988, pp. 69–99.

[14] C. von der Malsburg. "Self-organization of orientation-sensitive cells in the striate cortex," *Kybernetik*, 15:85–100, 1973.

[15] E. L. White. *Cortical Circuits: Synaptic Organization of the Cerebral Cortex—Structure, Function, and Theory*, 1989. Boston: Birkhauser.