

# Self-Organizing Dynamic Neural Fields

Nicolas P. Rougier and Georgios Is. Detorakis  
 INRIA Nancy-Grand Est  
 615 rue du Jardin Botanique, 54600 Villers-lès-Nancy, France  
 e-mail: {Nicolas.Rougier, Georgios.Detorakis}@inria.fr

## Summary

In this paper, we propose a model of cortical self-organization based on the dynamic field theory. Learning is made through the modification of feed-forward connections using a time invariant learning rule that allows for dynamic (or life-long) learning. This preliminary model suggests that cortical plasticity may be conveyed through feed-forward connections only while cortico-cortical connections role would be to ensure dynamic competition among cortical columns.

## Introduction

We introduced in [1] the dynamic self-organized map architecture that is a variation of the self-organizing map algorithm [2] where the original time-dependent (learning rate and neighborhood) learning function has been replaced by a time-invariant learning rule. This modification allows the network to support life-long learning and may explain to some extent cortical plasticity. However, current implementation is not biologically plausible since it requires a central supervisor (to designate the winning unit) and the neighborhood influence is computed using a function.

This article introduces preliminary results concerning a biologically plausible implementation using numerical, distributed and local computations, based on the original dynamic neural field definition [3].

The concept of self-organization using lateral connections is well-known and well-investigated, especially by [4, 5]. In those works authors provide an algorithm for self-organization learning lateral weights using a Hebbian-like learning rule. On the other hand, we put forward a new approach of self-organization using a combination of a dynamic neural field and a Hebbian-like learning rule. Thus, self-organization can be achieved, learning only the feed-forward weights. Because of that property our model is quite simple, straightforward to implementation and it does not require any outlandish handling.

## Model

The neural field theory has been introduced by [6, 7] and latter formalized by [3, 8]. This theory introduces a model at the level of a population of neurons in the form of an integro-differential equation describing the *spatio-temporal evolution of coarse-grained variables such as synaptic or firing rate value* [9]. We will use notations introduced by Amari and consider the membrane potential to be governed by the following equation:

$$\tau \frac{\partial V(\mathbf{x}, t)}{\partial t} = -V(\mathbf{x}, t) + h + I(\mathbf{x}, t) + \int_M W(|\mathbf{x} - \mathbf{y}|) f(V(\mathbf{y}, t)) d\mathbf{y} \quad (1)$$

where  $V(\mathbf{x}, t)$  designates the membrane potential at position  $\mathbf{x}$  and time  $t$ ,  $W(|\mathbf{x} - \mathbf{y}|)$  is the lateral connection weight function between position  $\mathbf{x}$  and  $\mathbf{y}$  (we assume here that the system is spatially homogeneous and isotropic),  $f$  is the mean firing rate function,  $I(\mathbf{x}, t)$  is the input at position  $\mathbf{x}$  and  $h$  is the resting potential.

## Fitting input

Depending on the firing rate function  $f$ , the lateral connectivity function  $W$  and the input  $I$ , such fields are known to exhibit a range of dynamic behaviors going from spatially and/or temporally periodic patterns to localized regions of activity. This latter case has been extensively used in a number of work to model visual attention [10], motor control



[11], working memory [12], etc. In the following, we will consider the lateral connectivity function to be a Difference of Gaussians (DoG) of the form:

$$W(\mathbf{x}) = Ae^{-\mathbf{x}^2/2\sigma_A^2} - Be^{-\mathbf{x}^2/2\sigma_B^2}$$

and the firing rate function to be a simple positive linear function:

$$f(\mathbf{x}) = \begin{cases} \mathbf{x} & \text{if } \mathbf{x} > 0 \\ 0 & \text{if } \mathbf{x} \leq 0 \end{cases}$$

The numerical simulation of such fields requires the spatial discretization of the domain  $M$  into  $n$  spatial elements while the temporal integration can be made using classical integration schemes. Considering  $M = [0..1]$  and  $h = 0$ , we can thus rewrite equation 1 as:

$$\begin{aligned} \tau \frac{\Delta V(\mathbf{x}_i, t)}{\Delta t} = & -V(\mathbf{x}_i, t) + I(\mathbf{x}_i, t) \\ & + \sum_{j=0}^n W(|\mathbf{x}_i - \mathbf{x}_j|) f(V(\mathbf{x}_j, t)) \end{aligned}$$

with  $\mathbf{x}_i = i/(n-1)$ . In the following, we will use the forward Euler integration scheme. Under these assumptions, we have been studying a set of parameters for the lateral weight connection that exhibit the following property: *for any uniform and positive input  $I$ , the neural field converges towards a single localized packet of activity whose maximum is approximately equal to  $I$* . We do not have yet the formal proof of such behavior but we found the property to be very consistent over a wide set of numerical simulations using different parameters ( $n, I, \Delta t$ ). We report on figure (1) such a simulation where the field is able to fit a constant input of level 0.45 after convergence. Instead of a stereotyped packet of activity with a constant maximum, the field activation represents a measure of the input. We will now explain how to use such property to ensure self-organization.

## Self-Organization

Let us now consider the slightly modified equation:

$$\begin{aligned} \tau \frac{\partial V(\mathbf{x}, t)}{\partial t} = & -V(\mathbf{x}, t) + 1 - \int_M |I(t) - W_F(\mathbf{x})| d\mathbf{x} \\ & + \int_M W_L(|\mathbf{x} - \mathbf{y}|) f(V(\mathbf{y}, t)) d\mathbf{y} \quad (2) \end{aligned}$$

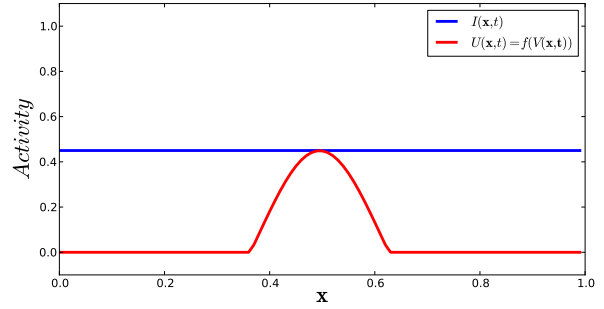


Figure 1: One dimensional neural field using  $n = 100$ ,  $\tau = 10$ ,  $A = 1.5$ ,  $\sigma_A = 0.1$ ,  $B = 0.75$ ,  $\sigma_B = 1.0$ . For any uniform and positive input level  $I$ , the neural field maximum activity is approximately equal to  $I$  after convergence.

where  $I$  is now considered to be uniform over  $M$  while a set of feed-forward weights  $W_F$  has been added such that the actual input for any position  $\mathbf{x}$  is  $1 - |I(t) - W_F(\mathbf{x})|$ .

Considering the set of parameters given in the previous subsection and considering a given input  $I$ , if we have  $W_F(\mathbf{x}) = I$ , the actual input of the field would be 1. In such a case, we explained that the field maximum activity would match this value (1). However, it is not possible to have such equality for any value of  $I$  because it would mean to change all feed-forward weights at once. Nonetheless, we can restrict this equality to the support of the localized packet of activity at the equilibrium point. Said differently, it is sufficient to have  $W_F(\mathbf{x}) = I$  for  $\mathbf{x}$  such that  $V(\mathbf{x}, t)$  is not null at the equilibrium point. The goal of the learning rule is thus to reach such a state.

We considered the learning rule introduced in [1] that reflect two main ideas:

- If a neuron is *close enough* to the data, there is no need for others to learn anything: the winner can represent the data.
- If there is no neuron close enough to the data, any neuron learns the data according to its own distance to the data.

To achieve such behavior, we propose to consider the following learning rule:

$$\frac{\partial W_F(\mathbf{x}, t)}{\partial t} = \eta L_e(\mathbf{x}, t) (I(t) - W_F(\mathbf{x}, t)) \quad (3)$$

where  $L_e(\mathbf{x}, t) = \int_M W_{L_e}(|\mathbf{x} - \mathbf{y}|)f(V(\mathbf{y}, t))d\mathbf{y}$  denotes the excitatory part of the lateral interaction such that  $W_L = W_{L_e} - W_{L_i}$  and  $\eta$  is a constant learning rate. The modification of any feed-forward weights is thus directly correlated with the closeness of weights to the input and this is modulated by the amount of lateral excitation. Since we know the final state of the model is a localized Gaussian-shaped packet of activity, learning occurs maximally in this vicinity.

## Results

### Experimental Setup

A set  $S_k$  of samples is generated by drawing  $k$  evenly spaced values in the interval  $[0..1]$  with special case  $S_\infty$  whose values are drawn uniformly from the interval  $[0..1]$ . A sample is randomly chosen from  $S_k$  and presented to the network which has been previously reset. The network is then simulated and learning occurs until  $\varepsilon$ -convergence has been reached, i.e.  $\forall \mathbf{x}, |V(\mathbf{x}, t + dt) - V(\mathbf{x}, t)| \leq \varepsilon$ . Another sample is then drawn and the procedure is repeated for a given number of epochs.

### Learning discrete values

We trained a network of 100 neurons for 2500 epochs and using as parameters,  $\tau = 10$ ,  $A = 1.5$ ,  $\sigma_A = 0.1$ ,  $B = 0.75$ ,  $\sigma_B = 1.0$ , self-organization was obtained. In figure (2), is illustrated the results of the simulation. The feed-forward weights were randomly initialized (the red line) and after 2500 epochs the feed-forward weights were organized (the blue line), as the network learned the three input values (0, 1/2, 1). Hence, that step-like shape of feed-forward weights were caused because the first 20 neurons learned the value 0, the 20 middle neurons learned the value 1/2, and the last 20 neurons learned the value 1. Moreover, the rest of the neurons learned different values from the input and may be that provide a smooth drift from one batch of neurons to another one.

### Learning continuous values

We used the same network architecture as in introduced in previous paragraph but it has been trained on the  $S_\infty$  set. After 2500 epochs the network has learned the feed-forward weights. In ad-

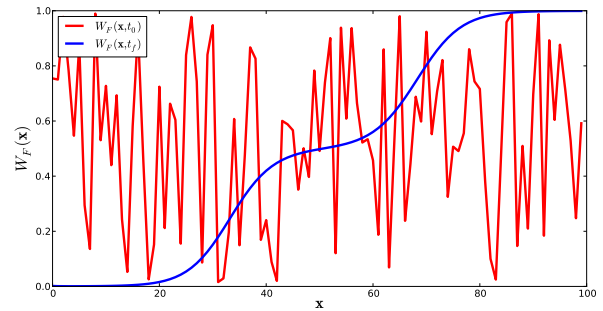


Figure 2: A neural field has been trained for 2500 epochs on  $S_3$  (0, 1/2 and 1). Blue and red curves respectively show initial and final set of weights.

dition, the so-called step-like shape of feed-forward weights was substituted by an almost linear shape. That's because of the continuous nature of the input. It is to be noted that almost all values are represented but the bounds due to the non-toric nature of the network. This side effect is well-known in the case of Kohonen's map. Results are depicted in figure 3 while figure 4 shows the evolution of the receptive fields of unit #50 that slowly drifts from a weak and random response to a sharp localized one.

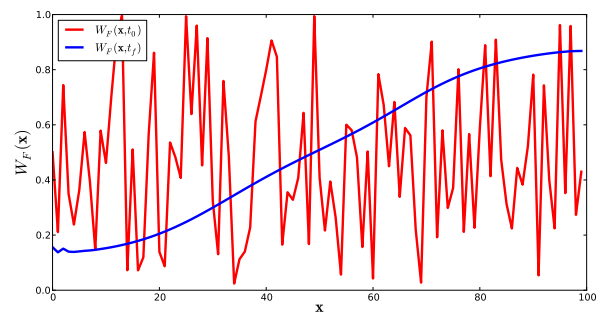


Figure 3: A neural field has been trained for 2500 epochs on  $S_\infty$ . Blue and red curves respectively show initial and final set of weights.

## Discussion

We introduced a one dimensional dynamic neural field that can continuously and dynamically self-organize itself around a set of one-dimensional discrete or uniform values by modifying its feed-forward connections and using the lateral weighted sum of excitation as a modulation signal for learn-

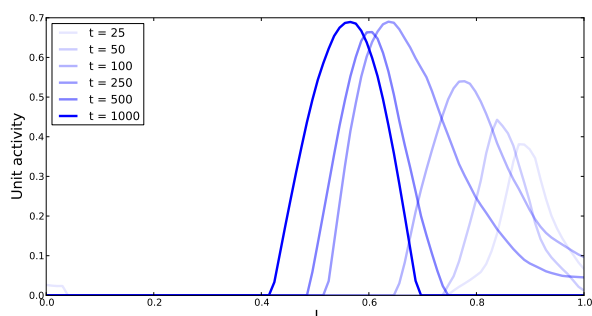


Figure 4: Evolution through time ( $t$ ) of the receptive field of unit #50 from a network learning  $S_\infty$ . At each record time, learning is frozen and the network is presented successively with values ranging from 0 to 1 with step of 0.01. Activity level of unit #50 is recorded (after convergence) for each of these input values and is plotted for each record time.

ing. We are currently investigating the two-dimensional case, due to obtain a more biological plausible model since the cortex can be modeled as a two-dimensional sheet of neurons. In this context, we would like to investigate meta-plasticity and/or homeostatic plasticity as it has been reported in [13].

Finally, we would like to examine the properties of cortical reorganization under the presence of a lesion. Such lesions have been extensively investigated by [14] and others from a neurophysiological point of view and there is consequently a strong and detailed experimental background providing significant data which could feed our computational model.

## References

- [1] N. Rougier and Y. Boniface, “Dynamic self-organising map,” *Neurocomputing*, vol. In Press, Corrected Proof, pp. –, 2011.
- [2] T. Kohonen, “Self-organized formation of topologically correct feature maps,” *Biological Cybernetics*, vol. 43, pp. 59–69, 1982.
- [3] S.-I. Amari, “Dynamic of pattern formation in lateral-inhibition type neural fields,” *Biological Cybernetics*, vol. 27, pp. 77–88, 1977.
- [4] J. Sirosh and R. Miikkulainen, “Self-organizing feature maps with lateral connections: Modeling ocular dominance,” in *Proceedings of the 1993 Connectionist Models Summer School*. Hillsdale, NJ: Erlbaum, Cite-seer, 1994.
- [5] J. Bednar, J. De Paula, and R. Miikkulainen, “Self-organization of color opponent receptive fields and laterally connected orientation maps,” *Neurocomputing*, vol. 65, pp. 69–76, 2005.
- [6] H. Wilson and J. Cowan, “Excitatory and inhibitory interactions in localized populations of model,” *Biophysics*, vol. 12, pp. 1–24, 1972.
- [7] H. Wilson and J. Cowan, “A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue,” *Kybernetik*, vol. 13, pp. 55–80, 1973.
- [8] J. Taylor, “Neural bubble dynamics in two dimensions: foundations,” *Biological Cybernetics*, vol. 80, pp. 5167–5174, 1999.
- [9] S. Coombes, “Neural fields,” *Scholarpedia*, vol. 1, no. 6, p. 1373, 2006.
- [10] N. Rougier and J. Vitay, “Emergence of attention within a neural population,” *Neural Networks*, vol. 19, no. 5, pp. 573–581, 2006.
- [11] W. Erlhagen and E. Bicho, “The dynamic neural field approach to cognitive robotics,” *Journal of Neural Engineering*, 2006.
- [12] J. P. Spencer and G. Schöner, “An embodied approach to cognitive systems: A dynamic neural field theory of spatial working memory,” in *Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci 2006) Society (CogSci 2006)*, (Vancouver, Canada), pp. 2180–2185, 2006.
- [13] D. Buonomano and M. Merzenich, “Cortical plasticity: from synapses to maps,” *Annual review of neuroscience*, vol. 21, no. 1, pp. 149–186, 1998.
- [14] J. Kaas, “Plasticity of sensory and motor maps in adult mammals,” *Annual review of neuroscience*, vol. 14, no. 1, pp. 137–167, 1991.