Stability analysis of a neural field self-organizing map

Georgios Is Detorakis, Antoine Chaillet, Nicolas P. Rougier

Stability analysis of a neural field self-organizing map

Georgios Is. Detorakis[1, †], Antoine Chaillet[2,3, †], and Nicolas P. Rougier[4,5]

[1]adNomus Inc., San Jose, CA, USA
[2]CentraleSupélec, Univ. Paris Saclay, Laboratoire des Signaux et Systèmes, Gif-sur-Yvette, France
[3]Institut Universitaire de France
[4]Inria Bordeaux Sud-Ouest, Bordeaux, France
[5]Institut des maladies neurodégénératives, CNRS , Université de Bordeaux, France
[†]These authors contributed equally to this work

Abstract This work provides theoretical conditions guaranteeing that a self-organizing map efficiently develops representations of the input space. The study relies on a neural fields model of spatiotemporal activity in area 3b of the primary somatosensory cortex. We rely on Lyapunov's theory for neural fields to derive theoretical conditions for stability. The theoretical conditions are verified by numerical experiments. The analysis highlights the key role played by the balance between excitation and inhibition of lateral synaptic coupling and the strength of synaptic gains in the formation and maintenance of self-organizing maps.

Keywords self-organizing maps, neural fields, Lyapunov function, asymptotic stability, neural networks.

1 Introduction

Self-organizing maps (SOMs) are neural networks mapping a high-dimensional space to a low-dimensional one, through unsupervised learning. They were first introduced by Grossberg (see [14] for a review), and later by Kohonen [19]. SOMs are widely used in computer science and data analysis for quantization and visualization of high dimensional data [37, 24]. They also constitute a suitable tool in computational neuroscience to study the formation and maintenance of topographic maps in primary sensory cortices such as the visual cortex [30, 23] and the somatosensory cortex [13, 33]. Many variations and applications of Kohonen's SOM algorithm can be found in [16] and [26].

A type of self-organizing map based on neural fields theory has been introduced in [8], where neural fields are used to drive the self-organizing process. Neural fields are integrodifferential equations that describe the spatiotemporal dynamics of a cortical sheet [3, 4, 5]. The SOM proposed in [8] describes the topographic organization of area 3b of the primary somatosensory cortex of monkeys [21, 27]. The model relies on an earlier work [28] known as Dynamic SOM (DSOM) algorithm. DSOM provides an online SOM learning algorithm where the Kohonen's SOM time-dependent learning rate and neighborhood function have been replaced by time-invariant ones. DSOM's neighborhood function and learning rate solely depend on the distance of the winner unit (*i.e.*, the most active neuron) from the input. The model proposed in [8, 9] combines the DSOM time-invariant learning rate and neighborhood function with Oja's learning rule [25]. As thoroughly described in [8, 9], the model is compatible with anatomical evidence of how area 3b in monkeys develops, maintains and reorganizes topographic representations of a skin patch of the index finger.

In this work, we provide theoretical insights on the stability and convergence of the neural field SOM algorithm proposed in [8, 9] by studying a more general class of systems than the one proposed originally in [8]. We use Lyapunov's stability theory adapted to neural field dynamics [10]. Since typical activation functions employed in the model (such as absolute values or rectification functions) are not necessarily differentiable, we do not rely on linearization techniques but rather directly assess the stability of the original nonlinear dynamics. Yet, the obtained results are local, meaning they are valid only for initial conditions in the vicinity of the considered equilibrium. Nonetheless, we show that they agree with numerical simulations. The stability conditions derived in this work can be used towards the direction of tuning neural field models such that they achieve the best possible results in developing self-organizing maps and thus more generalized representations. Moreover, the conditions

we propose indicate that the balance between lateral excitation and inhibition keeps the system stable, thus ruling out possible configurations in which learning does not take place properly. These findings are in line with both experimental observations [29, 18] and computational modeling [34, 35, 36].

The paper is organized as follows. In Section 2, we recall the SOM model under concern and its basic mechanisms. In Section 3, we present our main theoretical results, which we confront to numerical simulations in Section 4. A discussion on the obtained results is provided in Section 5. Mathematical proofs are given in Section 6.

2 Self-organizing neural fields

2.1 Neural population dynamics

We consider the following neural fields equation

$$\tau \frac{\partial u}{\partial t}(r,t) = -u(r,t) + \int_\Omega w_l(|r-r'|)\mathrm{rect}(u(r',t))dr' + I, \tag{1}$$

where Ω is a connected and compact subset of $\mathbb{R}^q$ ($q = 1, 2, 3$). For $q = 2$, the integral of a function $g : \Omega = \Omega_1 \times \Omega_2 \to \mathbb{R}$ is to be understood as $\int_\Omega g(r)dr = \int_{\Omega_1}\int_{\Omega_2} g(r_1, r_2)dr_2dr_1$ with $r = (r_1, r_2)$, and similarly for $q = 3$. $u(r,t)$ represents the mean membrane potential at position $r \in \Omega$ and time $t \geq 0$. τ is a positive decay time constant and I denotes an external input. w_l is a function that represents the strength of lateral synaptic coupling. It is given by

$$w_l(x) = w_e(x) - w_i(x), \tag{2}$$

where the excitation and inhibition synaptic weights are typically given by

$$w_e(x) = K_e e^{-x^2/2\sigma_e^2} \tag{3a}$$

$$w_i(x) = K_i e^{-x^2/2\sigma_i^2} \tag{3b}$$

with $K_e, K_i, \sigma_e, \sigma_i > 0$. In [8, 9], the input is provided through a two-dimensional skin model. The skin model is composed of a two-dimensional grid and receptors. The receptors are points distributed on the surface of the grid (uniformly). When a stimulus is applied on the grid, the receptors sample the input signal and convey the information to the cortical model. The skin stimulus is a noisy Gaussian-like function and the input to the neural fields model is provided by the following function I:

$$I(r,p,t) = 1 - \frac{|w_f(r,t) - s(p)|_1}{m}, \tag{4}$$

where $|\cdot|_1$ denotes the 1-norm: $|x|_1 = \sum_{i=1}^m |x_i|$, and $s : \mathbb{R}^2 \to [0,1]^m$ is a function that maps the raw input from the two-dimensional skin space to $[0,1]^m$. For instance, for a tactile stimulus at position $p \in \mathbb{R}^2$ on the skin, $s(p) \in \mathbb{R}^m$ could be defined as the normalized distance from p to each receptor's location, thus potentially of much higher dimension than 2. For a more detailed description of receptors' model please see [9]. $w_f : \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}^m$ represents feed-forward synaptic weights whose value is updated according to

$$\frac{\partial w_f}{\partial t}(r,t) = \gamma\left(s(p) - w_f(r,t)\right)\int_\Omega w_e(|r-r'|)\mathrm{rect}(u(r',t))dr', \tag{5}$$

where γ is a positive constant that represents the learning rate and $\mathrm{rect}(x) = \max\{x, 0\}$. It is worth observing that, since $s(p) \in [0,1]^m$, $w_f(r,t) \in [0,1]^m$ for all $r \in \Omega$ and all $t \geq 0$ given any initial conditions satisfying $w_f(r,0) \in [0,1]^m$ for all $r \in \Omega$ (this can be seen by observing that the entries of $\frac{\partial w_f}{\partial t}(r,t)$ are negative as soon as the corresponding entries of $w_f(r,t)$ become greater than 1; similarly, they are positive when the corresponding entries of $w_f(r,t)$ get below 0: see (5)). Hence,

$\frac{|w_f(r,t)-s(p)|_1}{m} \in [0,1]$ at all times. The expression (4) can thus be interpreted as a high input when the feedforward weights are close to $s(p)$ and a lower input when these are more distant.

The overall model Eq. (1, 4, 5) reflects the dynamics of a cortical neural population in combination with a learning rule of the feed-forward connections w_f, which convey information from receptors to the cortical sheet. As described in [8, 9], this model can express a variety of different behaviors, depending on the lateral connectivity kernels w_e and w_i.

The main advantage of the learning rule given by Eq. (5) is that it is a biologically plausible modification of the DSOM learning rule [28]. In DSOM the learning rate and neighborhood function are time-invariant and can adapt to the input according to one single parameter, called elasticity. This particular modification leads to the following behavior: if the winner neuron (*i.e.*, the neuron that has the shortest distance from the input stimulus to its corresponding codebook–weight) is close to the stimulus, then the neighborhood function shrinks around it. This results in making the weights of neurons within the dynamic neighborhood stronger and the weights of the other units weaker. However, when the winning unit is very far from the current input, the neighborhood function exhibits a broad activity pattern, promoting learning of every unit in the network. Therefore, in [8], the neighborhood function has been replaced by the term $\int_\Omega w_e(|r-r'|)\text{rect}(u(r',t))dr'$ providing a more realistic and biological plausible learning algorithm for self-organizing maps in the context of neuroscience.

2.2 Self-organizing maps

We start by briefly describing how the SOM model introduced in [8] and [9] works. The algorithm starts by initializing the feed-forward weights randomly (usually uniformly) and the neural field activity $u(r,0)$ is set to zero. The second step is to sample the input space by randomly drawn samples of dimension m from an input distribution. At every epoch one sample is given to the neural field Eq. (1) and (5) through Eq. (4). This first step is depicted in Figure 1 **A**, where a two-dimensional point $p = (p_1, p_2)$ is sampled from a uniform distribution $p_1, p_2 \sim \mathcal{U}(0,1)$. The samples are mapped to the neural space through the function s and then are passed to Eq. (4). At this point we should point out that there are two ways of presenting stimuli while training a self-organizing map. The first is to predetermine an amount of input samples and present one at each epoch (on-line learning) and the second is to collect all the input samples into a batch and give all of them at once to the network (batch learning). In this work we use the former (on-line learning) since it is biologically plausible.

Then the algorithm proceeds with computing the numerical solution of Eq. (1) and (5). To that aim, Eq. (1) and Eq. (5) are discretized and solved numerically using Euler's forward method. The numerical solution of Eq. (1) is typically a bell-shaped curve (bump) centered on the neuron which is the closest unit to the input sample and therefore is called winner neuron or best matching unit (BMU). In Figure 1**B** this is depicted as a black disc on a discrete lattice. The lattice represents a discretization of the field where each tile corresponds to a neuron. Neurons that lie within the vicinity (within the black disc in Figure 1 **B**) defined by the solution of Eq. (1) update their weights based on Eq. (5). The rest of the neurons feed-forward weights remain in their previous state. Once the temporal integration of Eq. (1) and Eq. (5) is complete the activity of the field is reset to its baseline activity. Then another input sample is drawn and the whole process repeats itself. Once the number of epochs has been exhausted, the learning stops and the mapping process has been completed.

To make the aforementioned algorithm directly comparable to Kohonen SOM [19] we provide some insights. First, in Kohonen's SOM we compute the distance between the input and the codebooks. Here we do the same using Eq. (4). The neighborhood function that Kohonen's SOM uses to update the feed-forward weights is replaced here by the numerical solution of the neural field (Eq. (1)) and more precisely by the term $\int_\Omega w_e(|r-r'|)\text{rect}(u(r',t))dr'$. Both the learning rate and the width of the neighborhood function are time-independent in our case as opposed to Kohonen's SOM where they are both time-dependent. Our learning rule is different since we use a modified Oja rule [25], which is based on Hebbian learning [15] and it is therefore biologically plausible [1]. The dimensionality reduction in both models, the Kohonen and ours, takes place at the level of the learning rule. This means that Eq. (5) is responsible for learning the representations and mapping the input distribution

(of dimensions m) on a manifold of lower dimension $q \in \{1, 2, 3\}$.

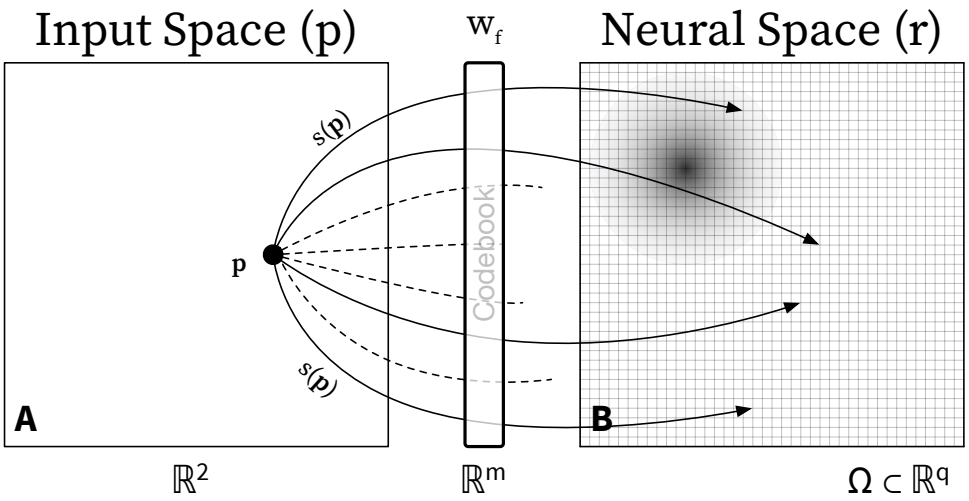

Figure 1: **Neural field self-organizing map.** Graphical representation of the learning algorithm introduced in [8, 9]. **A** Tactile bi-dimensional stimuli are mapped to an m-dimensional space through a function s that involves the skin receptors. This function is used to update the codebooks, which are then mapped to the neural space Ω of lower dimension. The input I receives the mapped input sample and provides input to the neural field and codebooks equations. **B** The numerical steady-state solution of Eq. (1) (*i.e.*, bump) defines the neighborhood (the group of neurons) that will have its neurons updating their codebooks based on Eq. (5).

3 Explicit conditions for stability

The most important question when one trains a self-organizing map is: *Will the learning process converge and properly map the input space to the neural one?* In most of the cases, it is not possible to predict this. However, in the specific case of the self-organizing algorithm provided by [8], we show here that it is possible to obtain an analytical condition that guarantee the stability of the equilibrium point of system (1)-(5). Stability during learning is a pre-requisite to generate a meaningful mapping and thus a proper topographic map. Moreover, a byproduct of deriving such a stability condition is to provide some insights on how to properly tune model parameters.

To this end, we now proceed to the mathematical analysis of the model. For the sake of generality, the adopted mathematical framework is slightly wider than merely Eq. (1, 4, 5) and encompasses more general classes of activation functions and synaptic kernels. We start by introducing the considered class of systems, and then provide sufficient conditions for its stability and convergence.

3.1 Model under study

The self-organizing neural field Eq. (1, 4, 5) is a particular case of the more general dynamics

$$\tau \frac{\partial u}{\partial t}(r,t) = -u(r,t) + \int_\Omega w_l(r,r') f_l(u(r',t))dr' + f_s(w_f(r,t) - s(p)) \tag{6a}$$

$$\frac{\partial w_f}{\partial t}(r,t) = \gamma \left(s(p) - w_f(r,t)\right) \int_\Omega w_e(r,r') f_e(u(r',t))dr', \tag{6b}$$

where $\tau, \gamma > 0$, $w_l, w_e \in L_2(\Omega^2, \mathbb{R})$, the set of all square-integrable functions from Ω^2 to $\mathbb{R}$, and f_e, f_l and f_s are Lipischitz continuous functions.

3.2 Existence of equilibrium patterns

Assuming that $\inf_{r\in\Omega} \int_\Omega w_e(r,r')f_e(u^*(r'))dr' \neq 0$, any equilibrium pattern (u^*, w_f^*) of (6) satisfies the following equations:

$$u^*(r) = f_s(0) + \int_\Omega w_l(r,r')f_l(u^*(r'))dr' \tag{7a}$$

$$w_f^*(r) = s(p). \tag{7b}$$

Since $\omega_l \in L_2(\Omega^2, \mathbb{R})$, [11, Theorem 3.6] ensures the existence of at least one such equilibrium pattern.

3.3 Stability analysis of Eq. (6)

We recall that an equilibrium x^* of a system $\dot{x}(t) = f(x(t))$, where $x(t) : \Omega \to \mathbb{R}^n$ for each fixed $t \geq 0$, is called *globally exponentially stable* if there exist $k, \varepsilon > 0$ such that, for all admissible initial conditions, it holds that

$$\|x(t) - x^*\| \leq k\|x(0) - x^*\|e^{-\varepsilon t}, \quad \forall t \geq 0, \tag{8}$$

where $\|\cdot\|$ denotes the spatial L_2-norm. This property thus ensures that all solutions go to the equilibrium configuration x^* in the L_2 sense (global convergence), and that the transient overshoot is proportional to the L_2-norm of the distance between the initial configuration and the equilibrium (stability). The equilibrium pattern x^* is said to be *locally exponentially stable* if (8) holds only for solutions starting sufficiently near from it (in the L_2 sense). We refer the reader to [10] for a deeper discussion on the stability analysis of neural fields.

Our main result proposes a sufficient condition for the local exponential stability of Eq. (6). Its proof is given in Section 6.1.

Theorem 1. *Let Ω be a compact connected set of $\mathbb{R}^q$, $w_l \in L_2(\Omega^2, \mathbb{R})$ and $w_e : \Omega^2 \to \mathbb{R}$ be a bounded function. Assume further that f_l, f_s and f_e are Lipschitz continuous functions, and let ℓ_l denote the Lipschitz constant of f_l. Let (u^*, w_f^*) denote any equilibrium of Eq. (6), as defined in Eq. (7). Then, under the conditions that*

$$\sqrt{\int_\Omega \int_\Omega w_l(r,r')^2 dr'dr} < \frac{1}{\ell_l} \tag{9}$$

$$\inf_{r\in\Omega} \int_\Omega w_e(r,r')f_e(u^*(r'))dr' > 0, \tag{10}$$

the equilibrium pattern (u^*, w_f^*) is locally exponentially stable for Eq. (6).

Condition (9) imposes that the synaptic weights of the lateral coupling w_l be sufficiently small: stronger lateral synaptic weights can be tolerated if the maximum slope ℓ_l of the activation function f_l is low enough, meaning that the system given by Eq. (6a) is less self-excitable. Recall that, if f_l is a differentiable function, then ℓ_l can be picked as the maximum value of its derivative. Nonetheless, Theorem 1 does not impose such a differentiability requirement, thus allowing to consider non-smooth functions such as absolute values, saturations, or rectification functions. Note that it was shown in [32] that condition (9) ensures that the system owns a single equilibrium pattern. It is also worth stressing that the slopes of the functions f_s and f_e do not intervene in the stability conditions.

Condition (10) requires a sufficient excitation in the vicinity of the equilibrium u^*. Roughly speaking, it imposes that the considered equilibrium pattern u^* does not lie in a region where f_e is zero.

3.4 Stability analysis of the SOM neural fields

Theorem 1 provides a stability condition for the model described by Eq. (6). We next apply it to the model given in [8] in order to derive more explicit and testable stability conditions. More

precisely, the self-organizing neural fields Eq. (1, 4, 5) can be put in the form of Eq. (6) by letting $f_e(x) = f_l(x) = \mathrm{rect}(x)$, $f_s(x) = 1 - \frac{|x|_1}{m}$ and

$$w_e(r, r') = K_e e^{-|r-r'|^2/2\sigma_e^2} \tag{11a}$$

$$w_i(r, r') = K_i e^{-|r-r'|^2/2\sigma_i^2} \tag{11b}$$

$$w_l(r, r') = w_e(r, r') - w_i(r, r'). \tag{11c}$$

In view of (7), the equilibrium patterns of Eq. (1, 4, 5) are given by

$$u^*(r) = 1 + \int_\Omega w_l(r, r')\mathrm{rect}(u^*(r'))dr' \tag{12a}$$

$$w_f^*(r) = s(p). \tag{12b}$$

The Lipschitz constant of f_l is $\ell_l = 1$. Based on this, we can also derive the following corollary, whose proof is provided in Section 6.2.

Corollary 1. *Assume that Ω is a compact and connected set of $\mathbb{R}^q$ and let w_e, w_i and w_l be as in* (11)*. Then, under the condition that*

$$\int_\Omega \int_\Omega \left(K_e e^{-|r-r'|^2/2\sigma_e^2} - K_i e^{-|r-r'|^2/2\sigma_i^2} \right)^2 dr' dr < 1, \tag{13}$$

the equilibrium (u^*, w_f^*), as defined in Eq. (12)*, is locally exponentially stable for Eq. (1)-(5).*

A particular case for which local exponential stability holds is when the excitation and inhibition weight functions are sufficiently balanced. Indeed, it appears clearly that Eq. (13) is fulfilled if $K_e \simeq K_i$ and $\sigma_e \simeq \sigma_i$. See the discussion in Section 5 for further physiological insights on this condition.

The integral involved in (13) can be solved explicitly. For instance, in the two-dimensional case ($q = 2$), the condition boils down to the following.

Corollary 2. *Assume that $\Omega = [a, b] \times [a, b]$ for some $a, b \in \mathbb{R}$ with $b \geq a$ and let w_e, w_i and w_l be as in* (11)*. Define*

$$\xi_{a,b}(\sigma) := \left(2\sigma^2 \left(e^{-\frac{(a-b)^2}{2\sigma^2}} - 1 \right) + \sigma\sqrt{2\pi}(a-b) Erf\left(\frac{a-b}{\sigma\sqrt{2}} \right) \right)^2, \quad \forall \sigma > 0, \tag{14}$$

where $Erf : \mathbb{R} \to (-1, 1)$ denotes the Gauss Error Function. Then, under the condition that

$$K_e^2 \xi_{a,b}(\sigma_e/\sqrt{2}) + K_i^2 \xi_{a,b}(\sigma_i/\sqrt{2}) - 2K_e K_i \xi_{a,b}\left(\frac{\sigma_e \sigma_i}{\sqrt{\sigma_e^2 + \sigma_i^2}} \right) < 1, \tag{15}$$

the equilibrium (u^*, w_f^*), as defined in Eq. (12)*, is locally exponentially stable for Eq. (1)-(5).*

Plenty of approximations are available for the Erf function in the literature. For instance, the following expression approximates it with a 5.10^{-4} error:

$$\mathrm{Erf}(x) \simeq 1 - \frac{1}{(1 + a_1 x + a_2 x^2 + a_3 x^3 + a_4 x^4)^4},$$

with $a_1 = 0.278393$, $a_2 = 0.230389$, $a_3 = 0.000972$, $a_4 = 0.078108$; see for instance [2]. The Erf function is also commonly implemented in mathematical software, thus making Eq. (15) easily testable in practice.

4 Numerical assessment on a two-dimensional map

In order to numerically assess whether the above stability condition correctly predicts the performance of the learning process, we focus on a simple example of a two-dimensional map ($q = 2$) and a two-dimensional input space ($n = 2$). Furthermore, we choose $s(p)$ to be the identity function since we do not consider any receptors: the position of the tactile stimuli is assumed to be directly available. This choice is motivated by the fact that the presence or absence of a receptors grid does not affect the theoretical results of the current work. We refer to [8, 9] for a more complex application of the neural field self-organizing algorithm.

We sample two-dimensional inputs from a uniform distribution. Therefore we have $s_i(p) = (p_1, p_2)$ where i indicates the i-th sample, and $p_1, p_2 \sim \mathcal{U}(0, 1)$. In all our simulations we use 7000 sample points and we train the self-organizing map over each of them (7000 epochs). It is worth stressing the difference between the training time (epochs) and the simulation time. The former refers to the iterations over all the input samples (stimuli): one such input is presented to the model at each epoch. The latter is attributed to the numerical temporal integration of Eq. (1)-(5). Each epoch thus corresponds to a predefined number of simulation steps. At the end of each epoch the activity of the neural field is reset to baseline activity before proceeding to the next epoch.

4.1 Parameters and simulation details

The neural fields equations are discretized using $k = 40 \times 40$ units. Accordingly, the two-dimensional model Eq. (1)-(5) is simulated over a spatial uniform discretization Ω_d of the spatial domain $\Omega = [0, 1] \times [0, 1]$, namely $\Omega_d = \bigcup_{i,j=1}^{40} (\frac{i}{40}, \frac{j}{40})$. The considered input space, over which the stimuli are uniformly distributed, is also $[0, 1] \times [0, 1]$ (two-dimensional input vectors). The temporal integration is performed using the forward Euler method, whereas the spatial convolution in Eq. (1)-(5) is computed via the Fast-Fourier Transform (FFT). The learning process runs for 7000 epochs. The components of the feed-forward weights are initialized from a uniform distribution $\mathcal{U}(0, 0.01)$ and the neural field activity is set to zero. At each epoch, we feed a stimulus to Eq. (1)-(5) and the system evolves according to its dynamics while the feed-forward weights are being updated. Then we reset the neural fields activity to zero. We run each experiment ten times using a different pseudo-random number generator (PRNG) seed each time (the PRNG seeds are given in Appendix 6.3: the same initial conditions and set of PRNG seeds sequence were used in each experimental condition).

The source code is written in Python (Numpy-, Numba-, Sklearn, and Matplotlib-dependent) and are freely distributed under the GPL 3-Clause License (`https://github.com/gdetor/som_stability`). All the parameters used in numerical simulations are summarized in Table 1. All simulations ran on an Intel NUC machine equipped with an Intel i7-10th generation processor and 32 GB of physical memory, running Ubuntu Linux (20.04.1 LTS, Kernel: 5.4.0-47-generic). The simulation of one self-organizing map consumes 493 MB of physical memory and it took 2671 seconds to run the 7000 epochs.

	K_e	σ_e	K_i	σ_i	τ	dt	t	γ	epochs
Figure 2	0.90	0.11	0.86	1.0	1.0	0.015	25.0	0.002	7 000
Figure 3	3.0	0.11	2.80	1.0	1.0	0.015	25.0	0.002	7 000

Table 1: **Simulation parameters.** K_e and K_i are the amplitudes of excitatory and inhibitory lateral connections, respectively. σ_e and σ_i are the variances of excitatory and inhibitory lateral connections, respectively. τ is the decay time constant, dt is the integration time step in ms, t is the simulation time in seconds. γ is the learning rate. In each epoch one stimulus is presented to the model.

4.2 SOM's quality measures

We measure the quality of the self-organizing maps using two performance indicators: the distortion $\mathcal{D}$ [6] and the $\delta x - \delta y$ representation [7]. We recall here that Ω_d is the spatial uniform discretization of $\Omega = [0, 1] \times [0, 1]$ and $k = 40 \times 40$ is the number of nodes (neurons). Furthermore, for each

$j \in \{1, \ldots, k\}$, $w_f^j(t^*)$ denotes the steady-state value of the feed-forward weights at the j-th node of the spatial discretization and t^* corresponds to the time at the end of an epoch.

The distortion assesses the quality of a self-organizing map. It measures the loss of information over the learning process. In other words, it indicates how good a reconstruction of an input will be after the mapping of all inputs to a lower-dimensional neural map. In a sense, distortion measures how well a SOM algorithm "compresses" the input data with respect to the neighborhood structure. Mathematically, the distortion is computed according to its discrete approximation

$$\mathcal{D} = \frac{1}{n} \sum_{i=1}^{n} \min_{j \in \{1,\ldots,k\}} |s_i(p) - w_f^j(t^*)|^2, \tag{16}$$

where n is the number of samples we use during the training of the self-organizing map.

Distortion is essentially an indicator of the map convergence, but it is not a reliable tool for assessing its quality. To gauge the quality of the map, we use the $\delta x - \delta y$ representation [7]. It shows when a map preserves the topology of the input space and hence how well a topographic map is formed. In order to estimate the $\delta x - \delta y$, we compute all the pairwise distances between the feed-forward weights, $\delta x = \delta x(i,j) = |w_f^i(t^*) - w_f^j(t^*)|$, and all the distances between the nodes of the uniform discretization of the input space $[0,1]^2$, $\delta y(i,j) = |y_i - y_j|$, for each $i, j = 1, \ldots, k$, where y_i are the discrete nodes of Ω_d. We plot the $\delta x - \delta y$ (*i.e.*, δx is the ordinate and δy the abscissa) along with a straight line, named $\mathcal{L}_{\delta x - \delta y}$, that crosses the origin and the mean of δx points. If the point cloud representation of $\delta x - \delta y$ closely follows the line $\mathcal{L}_{\delta x - \delta y}$ then the map is considered well-formed and preserves the topology of the input space.

In order to quantify the $\delta x - \delta y$ representation through a scalar performance index, we perform a linear regression on the point cloud of $\delta x - \delta y$ without fitting the intercept (magenta line in figures) and we get a new line named $\mathcal{L}_\Delta$. Then we define the measure $\mathcal{P} = \sqrt{\sum_{i=1}^{k} (a_i - b_i)^2}$, where $a_i \in \mathcal{L}_{\delta_x - \delta y}$, and $b_i \in \mathcal{L}_\Delta$. Naturally, $\mathcal{P}$ should approach zero as the two lines are getting closer, indicating that the self-organizing map respects the topology of the input space and thus it is well-formed.

4.3 Stable case

We start by simulating the model described by Eq. (1)-(5) with the parameters given in first line of Table 1. With these parameters, Condition (15) is fulfilled ($0.47 < 1$) and Corollary 2 predicts that the equilibrium is exponentially stable over each epoch. Accordingly, the model succeeds in building up a self-organizing map as shown in panel **A** of Figure 2. The white discs indicate the feed-forward weights after learning and the black dots indicate the input data points (two-dimensional rectangular uniform distribution).

Panels **B** and **C** show the $\delta x - \delta y$ representation and the distortion, respectively. We observe that the $\delta x - \delta y$ representation indicates a correlation between the feed-forward weights and the rectangular grid points (aligned with the mean of δx–red line). This means that the self-organizing map is well-formed and conserves the topology of the input. Moreover, the distortion declines and converges towards 0.0025 pointing out first that the loss of information during learning is low and that the structure in the self-organizing map is preserved. However, the boundary effects (the density of points is higher at the boundary of the map in panel **A**) affect both the distortion (it does not converge to zero – see panel **C** in Figure 2) and the $\delta x - \delta y$ representation (it is not perfectly aligned with the red line – see panel **B** in Figure 2). In spite of these boundary effects, the obtained $\delta x - \delta y$ performance indicator is good ($\mathcal{P} = 0.01$).

The evolution of the norm-2 of feed-forward weights of three randomly chosen units ($r^* = (0.25, 0.25)$, $(0.1, 0.225)$, $(0.35, 0.075)$) is shown in the panel **D** of Figure 2. This implies that the weights converge to an equilibrium after a transient period of about 2000 epochs. The oscillations around the equilibrium are due to a repeated alteration of the input stimulus which causes a shift to the feed-forward weights values of each winner neuron (see [8] for more details).

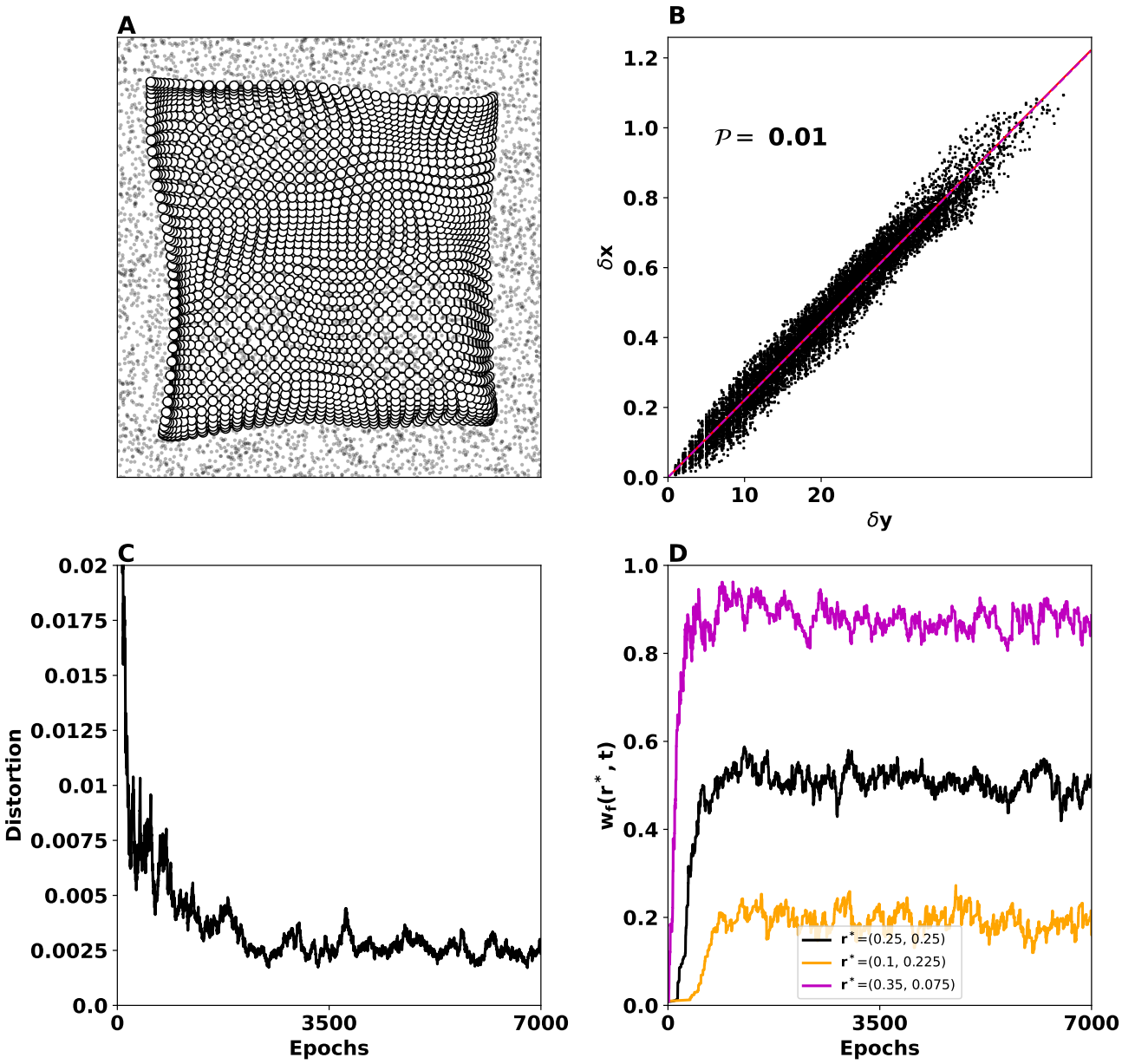

Figure 2: **Two-dimensional SOM performance in the stable case. A** Feed-forward weights (white discs) as they have been organized into a topographic map after 7000 epochs. The input in this case is a two-dimensional rectangular uniform distribution (black dots). **B** $\delta x - \delta y$ representation (black cloud), mean of δx (red line), and the linear regression of the $\delta x - \delta y$ representation (magenta line). The fact that the cloud is aligned around the red line indicates that the topographic map is well-organized, as confirmed by a good index performance $\mathcal{P}$ = 0.01. **C** Distortion indicates that the loss of information during the learning process decreases and the mapping of the input data to a two-dimensional self-organizing map respects the structure of the neighborhoods. **D** Temporal evolution of norm-2 of feed-forward weights of three neurons placed at $r^* = (0.25, 0.25)$, $(0.1, 0.225)$, and $(0.35, 0.075)$). Condition (15) is fulfilled and therefore the weights converge to an equilibrium giving rise to a well-formed topographic map.

4.4 Unstable case

The second line of Table 1 provides parameters for which Condition (15) is violated ($5.25 > 1$). According to our theoretical predictions, the model might not be stable and thus may not be able to develop any self-organizing map at all. In order to make sure that this is the case (and not merely a transient effect), we have let the training take more epochs (20 000). Nevertheless, we present here only the 7 000 first epochs for consistency with the rest of our experiments. This situation is illustrated in Figure 3, where the self-organizing process has failed to generate a well-formed map (panel **A**). In this case it is apparent that self-organization process has failed to generate a topographic map.

The $\delta x - \delta y$ representation in panel **B** of Figure 3 looks like a diffused cloud, indicating that there is no correlation between the grid points and the feed-forward weights meaning that there is no preservation of the topology of the input space. Accordingly the performance index reaches the value $\mathcal{P} = 0.41$, thus higher than the stable case. Moreover, the distortion in panel **C** of Figure 3 oscillates without converging to an equilibrium pointing out that the loss of information remains high and therefore the mapping is not successful. Finally, the norm-2 of feed-forward weights of three units ($r^* = (0.25, 0.25)$, $(0.1, 0.225)$, $(0.35, 0.075)$) are shown in panel **D**: it is apparent that they do not converge to an equilibrium. Instead they oscillate violently and they never stabilize around an equilibrium configuration.

4.5 Numerical Assessment of Corollary 2

Finally, we numerically tested Condition (15) of Corollary 2 for different values of the parameters K_e and K_i (all other parameters remained the same as in Table 1). For each pair (K_e, K_i) we computed the left-hand side of Eq. (15), the distortion $\mathcal{D}$ (averaged over the last 10 epochs), and the $\delta x - \delta y$ performance index $\mathcal{P}$: see Figure 4. We observe that, for high values of K_e and K_i, the stability condition of Corollary 2 is violated (the black solid line overpasses the black dashed line). The distortion (orange curve) closely follows the left-hand side of Condition (15) (up to a scaling factor), suggesting that distortion can serve as a measure of stability of the system (1)-(5). Furthermore, the distortion and the $\delta x - \delta y$ performance index $\mathcal{P}$ indicate that the learning process degrades for high values of (K_e, K_i), in line with the fact that Condition (15) is violated. Figure 5 confirms this good alignment between the theoretical stability condition and the performance of the self-organizing map: for the first five cases it properly maps the input space to the neural one, whereas the topology of the input space is not preserved in the last two cases and a malformed topographic map is obtained.

5 Conclusion

In this work, we have presented theoretical conditions for the stability of a neural fields system coupled with an Oja-like learning rule [25]. Numerical assessments on a two-dimensional self-organizing map indicate that the theoretical condition is closely aligned with the capacity of the network to form a coherent topographic map.

Previous works have shown through simulations that the dynamical system described by Eq. (1)-(5) can develop topographic maps through an unsupervised self-organization process [8, 9]. The model relies on the activity of a neural field to drive a learning process. This type of models are capable of developing topographic maps and reorganize them in face of several kinds of disturbances. Here, we proceed to a rigorous theoretical analysis of such kind of models by employing neural fields Lyapunov theory.

The obtained stability conditions are reminiscent of those obtained for general neural fields dynamics, in which the spatial L_2-norm of synaptic weights plays an essential role [10, 32, 12]. In our setting, these conditions translate in a good balance between excitation and inhibition for the exponential stability of the model's equilibrium, thus allowing the self-organizing process to develop topographic maps. It is worth stressing that the proof techniques employed here do not rely on a linearization of the system around the considered equilibrium; it thus allows to cover activation functions that are not differentiable (such as classical saturation or rectification functions).

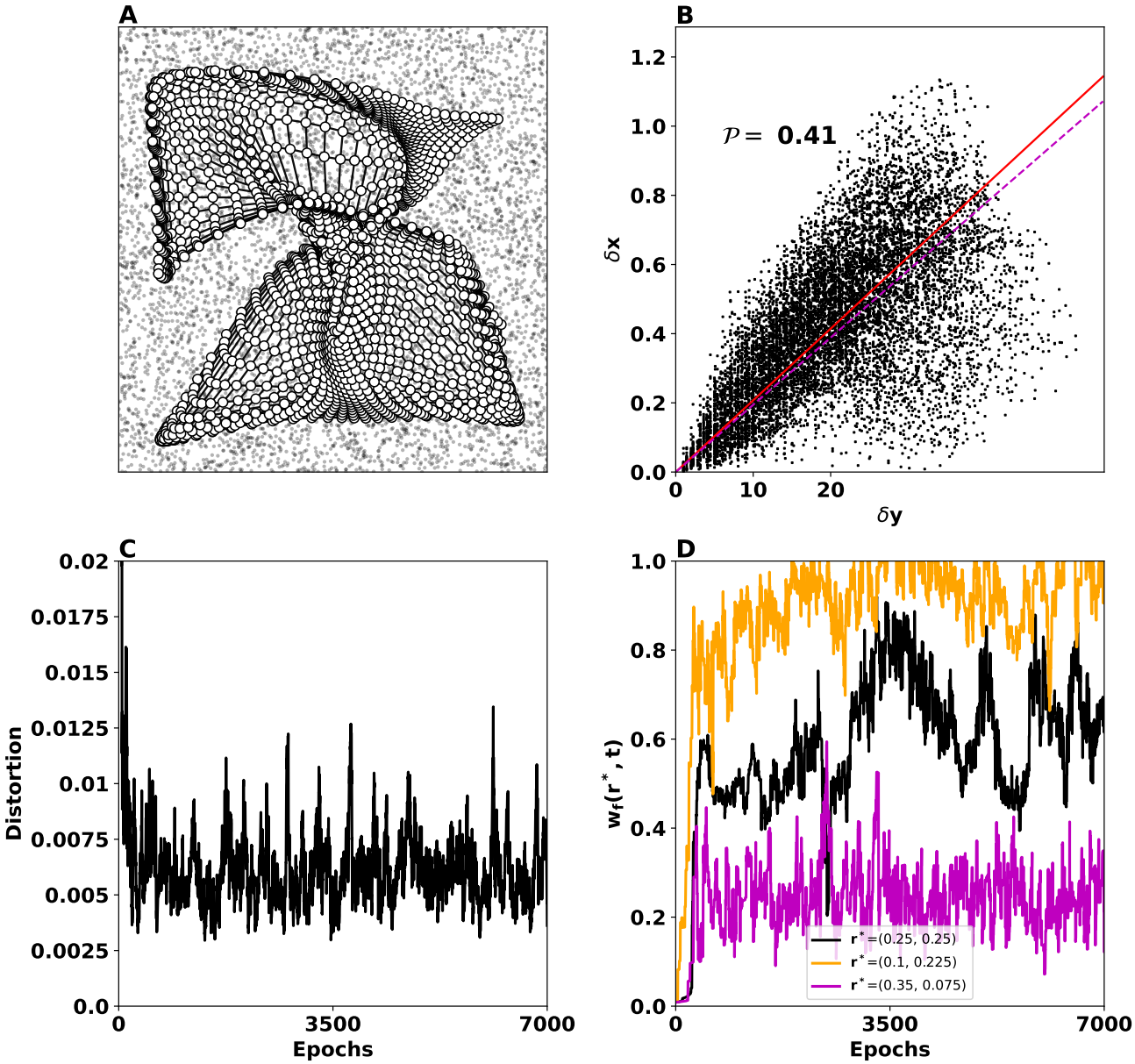

Figure 3: **Two-dimensional SOM performance in the unstable case. A** Feed-forward weights (white discs) as they have failed to organize into a topographic map after 7000 epochs. The input in this case is a two-dimensional rectangular uniform distribution (black dots). **B** $\delta x - \delta y$ representation (black cloud), mean of δx (red line), and the linear regression of the $\delta x - \delta y$ representation (magenta line). The fact that the cloud looks diffused indicates that the topographic map is not well-organized, as confirmed by a high value of $\mathcal{P}$ = 0.41. **C** Distortion indicates that the loss of information during the learning process never drops (converges to an equilibrium) instead it oscillates. This means that the mapping of the input data to a two-dimensional map has failed. **D** Temporal evolution of norm-2 of feed-forward weights of three neurons placed at r^* = $(0.25, 0.25)$, $(0.1, 0.225)$, and $(0.35, 0.075)$). The condition (15) is violated and accordingly the weights do not converge to an equilibrium.

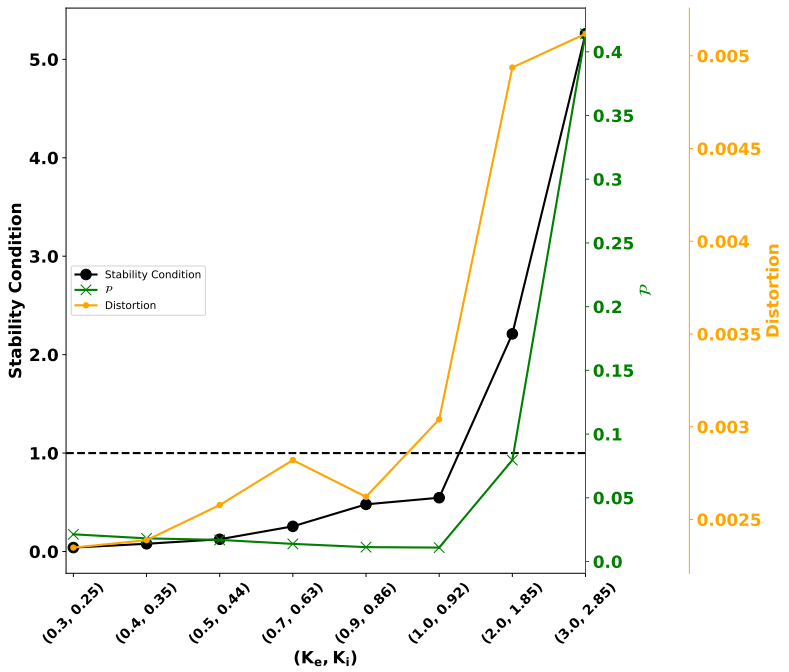

Figure 4: **Numerical investigation of Corollary 2.** Eight different pairs of the parameters (K_e, K_i) were used to investigate the conservativeness of the stability condition given by Corollary 2. We ran eight different simulations for 7 000 epochs keeping always the rest of the parameters same as in Table 1 and the same PRNG seed as before (7659). The black curve indicates the numerical value of the left-hand side of (15): stability is guaranteed if it is below the black dashed line. The green curves indicates the $\delta x - \delta y$ performance index $\mathcal{P}$. The orange curve represents the distortion $\mathcal{D}$ averaged of the 10 last epochs. It is apparent that as the values of (K_e, K_i) increase the Corollary 2 becomes violated and the self-organizing map is fails to map the input space to the neural one (see Figure 5 for more details).

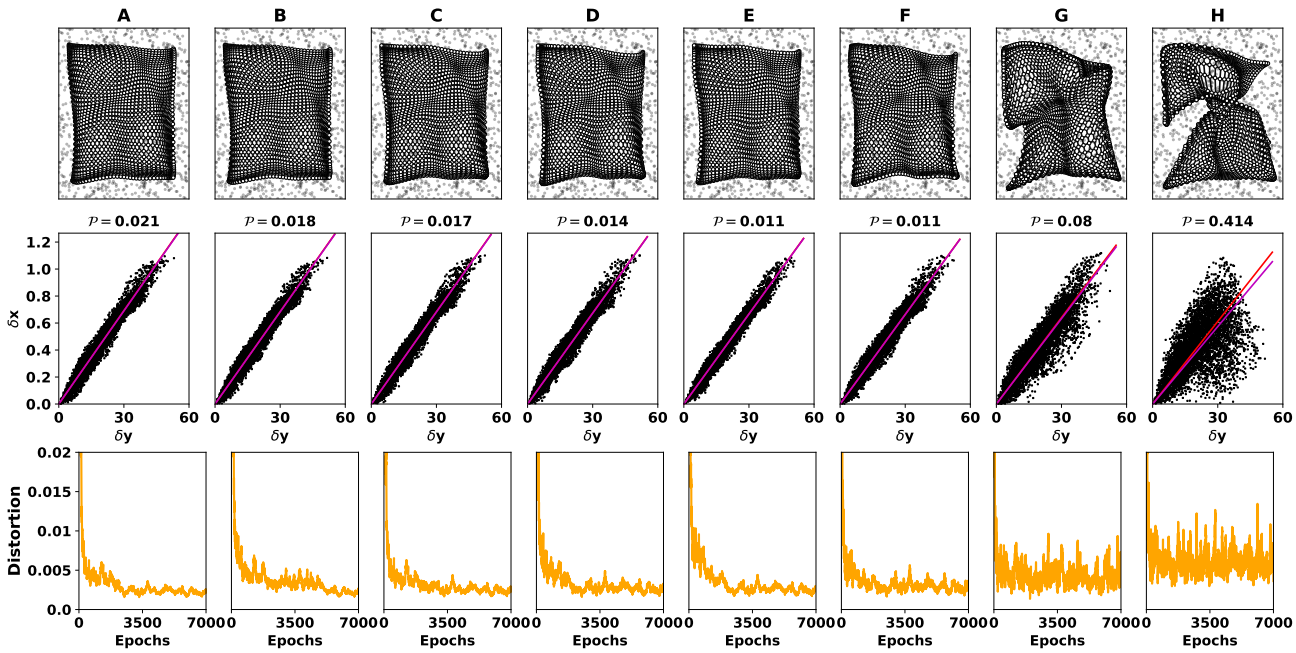

Figure 5: **Numerical Investigation of Corollary 2.** For the same eight experiments as in Figure 4, the obtained self-organizing map is provided (first line), together with its $\delta x - \delta y$ representation (second line) and the evolution of the distortion (third line). The mean δx is represented as a red line, whereas the slope of the linear regression is given as a magenta line. **A**: $(K_e = 0.30, K_i = 0.25)$, **B**: $(K_e = 0.4, K_i = 0.35)$, **C**: $(K_e = 0.5, K_i = 0.45)$, **D**: $(K_e = 0.7, K_i = 0.63)$, **E**: $(K_e = 0.9, K_i = 0.86)$, **F**: $(K_e = 1.0, K_i = 0.92)$. In line with Figure 4, a relevant map is obtained for the first five experiments (for which Condition (15) is fulfilled), whereas for the two last self-organizing maps **G**: $(K_e = 2, K_i = 1.85)$, **H**: $(K_e = 3, K_i = 2.85)$ the stability condition (15) is violated. This violation results in a non-stable neural field equation and thus the self-organizing maps do not learn properly the representations.

These stability conditions provide a means to identify the parameters set within which the unsupervised learning works efficiently, and thus provides an indication on how to tune them in practice. In particular, they can be used to further investigate how the dynamics of an underlying system affects the learning process during an unsupervised training process and what is the effect of the parameters on the final topographic map: as Figure 4 indicates, the parameters of the model directly affect the quality of the topographic map. However, a limitation of the present work is that it does not offer a way to choose the parameters in an optimal way. Furthermore, although the conditions provided by Theorem 1 guarantee stability of the neural field, they do not predict the quality of the obtained map: stability ensures that the learning will converge to an equilibrium, but the quality of the obtained map strongly depends on the structure of this equilibrium, hence on the chosen initial values of the feed-forward weights. This is a well-known problem with self-organizing maps [19] that is generally solved using a decreasing neighborhood, starting from a very wide one. In our case, the neighborhood function is directly correlated with the profile of the field activity and is *fixed* (stereotyped). We thus cannot always ensure the proper unfolding of the map. It is to be noted that when the neighborhood of a Kohonen is kept fixed, it suffers from similar problems. Nevertheless, the numerical assessment of the proposed theoretical stability conditions suggest that the stability condition accurately predicts the emergence of topographic maps through unsupervised learning: see Figures 4 and 5.

Other works have studied stability conditions for Kohonen maps and vector quantization algorithms, using methods from linear systems stability theory [31] or through energy functions [**?**]. However, these works focus on the learning rule for the Kohonen self-organizing maps [20] and the dynamics are not explicitly given by dynamical systems. Our work goes beyond by taking into account not only the learning dynamics, but also the neural dynamics that drives the self-organizing process.

Last but not least, it has been shown that neural adaptation is crucial in the development of the neocortex [22] and neurons tend to adapt their input/output relation according to the statistics of the input stimuli. Our theoretical results provide conditions under which this input/output adaptation successfully takes place at least at a computational level.

6 Proof of the theoretical results

6.1 Proof of Theorem 1

In order to place the equilibrium at the origin, we employ the following change of variables:

$$\begin{aligned}\tilde{u}(r,t) &= u(r,t) - u^*(r)\\ \tilde{w}_f(r,t) &= w_f(r,t) - w_f^*(r),\end{aligned}$$

where u^* and w_f^* denote the equilibrium patters of Eq. (6), as defined in Eq. (7). Then, the system (6) can be written as

$$\tau\frac{\partial\tilde{u}}{\partial t}(r,t) = -\tilde{u}(r,t) + \int_\Omega w_l(r,r')\tilde{f}_l(r',\tilde{u}(r',t))dr' + \tilde{f}_s(\tilde{w}_f(r,t)) \tag{17a}$$

$$\frac{\partial\tilde{w}_f}{\partial t}(r,t) = -\gamma\tilde{w}_f(r,t)\int_\Omega w_e(r,r')\tilde{f}_e(\tilde{u}(r',t))dr'. \tag{17b}$$

where, for all $x \in \mathbb{R}$ and all $r \in \Omega$,

$$\begin{aligned}\tilde{f}_l(r,x) &= f_l(x + u^*(r)) - f_l(u^*(r))\\ \tilde{f}_s(x) &= f_s(x) - f_s(0)\\ \tilde{f}_e(r,x) &= f_e(x + u^*(r)).\end{aligned}$$

With this notation, it holds that $\tilde{f}_l(r,0) = \tilde{f}_s(0) = 0$ for all $r \in \Omega$, meaning that (17) owns an equilibrium at zero. The stability properties of the origin of (17) thus determines those of the equilibria of (6).

First observe that, since w_e is a bounded function and Ω is compact, there exists $\bar{\omega}_e > 0$ such that

$$\int_\Omega w_e(r,r')^2 dr' \leq \bar{w}_e^2, \quad \forall r \in \Omega. \tag{18}$$

In order to assess the stability of (17), one may be tempted to rely on linearization techniques. Nevertheless, the linearized system (17) around the origin would necessarily involve the derivative of f_s at zero, which may be undefined if f_s is not differentiable at zero (which is the case for the system of interest (1)-(5) where f_s involves an absolute value). Consequently, the proof we propose here relies on Lyapunov methods [17] that were extended to neural fields in [10].

Consider the following Lyapunov functional:

$$V(t) := \frac{\tau}{2} \int_\Omega \tilde{u}(r,t)^2 dr + \frac{\rho}{2\gamma} \int_\Omega \tilde{w}_f(r,t)^2 dr, \tag{19}$$

where $\rho > 0$ denotes a parameter whose value will be decided later. First observe that the following bounds hold at all $t \geq 0$:

$$\underline{\alpha}\left(\|\tilde{u}(\cdot,t)\|^2 + \|\tilde{w}_f(\cdot,t)\|^2 \right) \leq V(t) \leq \overline{\alpha}\left(\|\tilde{u}(\cdot,t)\|^2 + \|\tilde{w}_f(\cdot,t)\|^2 \right), \tag{20}$$

where $\underline{\alpha} := \frac{1}{2}\min\{\tau\,;\,\rho/\gamma\} > 0$ and $\overline{\alpha} := \frac{1}{2}\max\{\tau\,;\,\rho/\gamma\} > 0$. The derivative of V along the solutions of (6) reads

$$\begin{aligned}
\dot{V}(t) &= \tau \int_\Omega \tilde{u}(r,t) \frac{\partial \tilde{u}(r,t)}{\partial t} dr + \frac{\rho}{\gamma} \int_\Omega \tilde{w}_f(r,t) \frac{\partial \tilde{w}_f(r,t)}{\partial t} dr \\
&= \int_\Omega \tilde{u}(r,t) \left[-\tilde{u}(r,t) + \int_\Omega w_l(r,r') \tilde{f}_l(r', \tilde{u}(r',t)) dr' + \tilde{f}_s(\tilde{w}_f(r,t)) \right] dr \\
&\quad - \rho \int_\Omega \left[\tilde{w}_f(r,t)^2 \int_\Omega w_e(r,r') \tilde{f}_e(r', \tilde{u}(r',t)) dr' \right] dr.
\end{aligned} \tag{21}$$

Moreover, denoting by ℓ_s, ℓ_e and ℓ_l the Lipschitz constants of f_s, f_e and f_l respectively, it holds that, for all $x \in \mathbb{R}$ and all $r \in \Omega$.

$$|\tilde{f}_l(r,x)| \leq \ell_l |x| \tag{22}$$

$$|\tilde{f}_s(x)| \leq \ell_s |x| \tag{23}$$

$$|\tilde{f}_e(r,x) - \tilde{f}_e(r,0)| \leq \ell_e |x|. \tag{24}$$

Applying the Cauchy-Schwarz inequality and using Eq. (22), it follows that

$$\begin{aligned}
\int_\Omega w_l(r,r') \tilde{f}_l(r', \tilde{u}(r',t)) dr' &\leq \int_\Omega |w_l(r,r')| \, |\tilde{f}_l(r', \tilde{u}(r',t))| dr' \\
&\leq \ell_l \int_\Omega |w_l(r,r')| \, |\tilde{u}(r',t)| dr' \\
&\leq \ell_l \sqrt{\int_\Omega w_l(r,r')^2 dr'} \sqrt{\int_\Omega \tilde{u}(r',t)^2 dr'}.
\end{aligned}$$

Hence, using again Cauchy-Schwarz inequality,

$$\begin{aligned}
&\int_\Omega \tilde{u}(r,t) \left[\int_\Omega w_l(r,r') \tilde{f}_l(r', \tilde{u}(r',t)) dr' \right] dr \\
&\quad \leq \ell_l \int_\Omega |\tilde{u}(r,t)| \left[\sqrt{\int_\Omega w_l(r,r')^2 dr'} \sqrt{\int_\Omega \tilde{u}(r',t)^2 dr'} \right] dr \\
&\quad \leq \ell_l \sqrt{\int_\Omega \tilde{u}(r,t)^2 dr} \sqrt{\int_\Omega \left[\int_\Omega w_l(r,r')^2 dr' \int_\Omega \tilde{u}(r',t)^2 dr' \right] dr}.
\end{aligned}$$

Observing that $\int_\Omega \tilde{u}(r',t)^2 dr'$ is independent of r and defining

$$\bar{w}_l := \sqrt{\int_\Omega \int_\Omega w_l(r,r')^2 dr' dr}, \tag{25}$$

it follows that

$$\int_\Omega \tilde{u}(r,t)\left[\int_\Omega w_l(r,r')\tilde{f}_l(r',\tilde{u}(r',t))dr'\right]dr \le \ell_l \bar{w}_l \int_\Omega \tilde{u}(r,t)^2 dr. \tag{26}$$

Furthermore, using Eq. (23), we have that

$$\begin{aligned}\int_\Omega \tilde{u}(r,t)\tilde{f}_s(w_f(r,t))dr &\le \ell_s \int_\Omega |\tilde{u}(r,t)|\,|w_f(r,t)|dr \\ &\le \ell_s \sqrt{\int_\Omega \tilde{u}(r,t)^2 dr}\sqrt{\int_\Omega w_f(r,t)^2 dr}\end{aligned}$$

Invoking the inequality $2ab \le (a^2/\lambda + \lambda b^2)$ for all $a, b \in \mathbb{R}$ and all $\lambda > 0$, we obtain that

$$\int_\Omega \tilde{u}(r,t)\tilde{f}_s(\tilde{w}_f(r,t))dr \le \frac{\ell_s}{2}\left(\lambda \int_\Omega \tilde{u}(r,t)^2 dr + \frac{1}{\lambda}\int_\Omega \tilde{w}_f(r,t)^2 dr\right), \tag{27}$$

for any $\lambda > 0$.

Now, assumption (10) ensures that $\inf_{r\in\Omega} \int_\Omega w_e(r,r')\tilde{f}_e(r',0)dr' > 0$. It follows that there exists $c > 0$ such that

$$\int_\Omega w_e(r,r')\tilde{f}_e(r',0)dr' \ge 2c, \quad \forall r \in \Omega.$$

Consequently, using (24) and Cauchy-Schwarz inequality, we get that, for any $v \in L_2(\Omega,\mathbb{R})$,

$$\begin{aligned}\int_\Omega w_e(r,r')\tilde{f}_e(r',v(r'))dr' &= \int_\Omega w_e(r,r')\tilde{f}_e(r',0)dr' + \int_\Omega w_e(r,r')\left(\tilde{f}_e(r',v(r')) - \tilde{f}_e(r',0)\right)dr' \\ &\ge 2c - \int_\Omega \left|w_e(r,r')\right|\,\left|\tilde{f}_e(r',v(r')) - \tilde{f}_e(r',0)\right|dr' \\ &\ge 2c - \ell_e \int_\Omega \left|w_e(r,r')\right|\,\left|v(r')\right|dr' \\ &\ge 2c - \ell_e \sqrt{\int_\Omega w_e(r,r')^2 dr'}\sqrt{\int_\Omega v(r')^2 dr'} \\ &\ge 2c - \ell_e \bar{w}_e \|v\|,\end{aligned}$$

where the last bound comes from (18). Let $\mathcal{B}_\varepsilon$ denote the ball (in L_2-norm) of radius $\varepsilon > 0$, that is: $\mathcal{B}_\varepsilon := \{v \in L_2(\Omega,\mathbb{R}) : \|v\| < \varepsilon\}$. Letting $\varepsilon := c/\ell_e\bar{w}_e$, we conclude from the above expression that

$$\int_\Omega w_e(r,r')\tilde{f}_e(r',v(r'))dr' \ge c, \quad \forall r \in \Omega,\ \forall v \in \mathcal{B}_\varepsilon. \tag{28}$$

Consider an initial condition such that $\tilde{u}(\cdot,0) \in \mathcal{B}_\varepsilon$ and let $T \in [0,+\infty]$ denote the time needed for $\tilde{u}(\cdot,t)$ to leave $\mathcal{B}_\varepsilon$. Then it holds by definition that $\tilde{u}(\cdot,t) \in \mathcal{B}_\varepsilon$ for all $t \in [0,T)$ and $\tilde{u}(\cdot,T) \notin \mathcal{B}_\varepsilon$ if T is finite. Note that, by continuity of solutions, $T > 0$. Moreover, in view of (28),

$$\int_\Omega w_e(r,r')\tilde{f}_e(r',\tilde{u}(r',t))dr' \ge c, \quad \forall t \in [0,T), \forall r \in \Omega. \tag{29}$$

Combining Eq.(21), (22), (23), and (29), we obtain that, for all $t \in [0,T)$,

$$\dot{V}(t) \le -\left(1 - \ell_l\bar{w}_l - \frac{\lambda\ell_s}{2}\right)\int_\Omega \tilde{u}(r,t)^2 dr - \left(\rho c - \frac{\ell_s}{2\lambda}\right)\int_\Omega \tilde{w}_f(r,t)^2 dr$$

Pick $\lambda = (1 - \ell_l \bar{w}_l)/\ell_s$. Note that $\lambda > 0$ since $\ell_l \bar{w}_l < 1$ by assumption (see Eq. (9)). Then the choice $\rho = \frac{\ell_s}{c\lambda} = \frac{\ell_s^2}{c(1-\ell_l \bar{w}_l)} > 0$ leads to:

$$\begin{aligned}\dot{V}(t) &\leq -\frac{1}{2}\|\tilde{u}(\cdot,t)\|^2 - \frac{\rho c}{2}\|\tilde{w}_f(\cdot,t)\|^2 \\ &\leq -\frac{1}{2}\min\{1\,;\,\rho c\}\left(\|\tilde{u}(\cdot,t)\|^2 + \|\tilde{w}_f(\cdot,t)\|^2\right).\end{aligned}$$

Using (20) and letting $\alpha := \frac{1}{2\overline{\alpha}}\min\{1\,;\,\rho c\} > 0$, we finally obtain that

$$\dot{V}(t) \leq -\alpha V(t), \quad \forall t \in [0,T).$$

Integrating, this gives $V(t) \leq V(0)e^{-\alpha t}$ for all $t \in [0,T)$, which yields, using (20),

$$\|\tilde{u}(\cdot,t)\|^2 + \|\tilde{w}_f(\cdot,t)\|^2 \leq \frac{\overline{\alpha}}{\underline{\alpha}}\left(\|\tilde{u}(\cdot,0)\|^2 + \|\tilde{w}_f(\cdot,0)\|^2\right)e^{-\alpha t}, \quad \forall t \in [0,T). \tag{30}$$

Thus, if initial conditions are picked within the L_2-ball of radius $\frac{\sqrt{\underline{\alpha}}}{\varepsilon\sqrt{\overline{\alpha}}}$, then $\|\tilde{u}(\cdot,t)\| + \|\tilde{w}_f(\cdot,t)\| < \varepsilon$ at all times $t \geq 0$. This means that, for these initial conditions, solutions never leave the ball $\mathcal{B}_\varepsilon$, hence $T = +\infty$. Eq. (30) thus ensures exponential stability on this set of initial conditions.

6.2 Proof of Corollary 1

Assumption described by Eq. (13) is equivalent to requiring $\bar{w}_l < 1$, with $\bar{w}_l$ defined in Eq. (25). Since the Lipschitz constant of the rectification is $\ell_l = 1$, this makes Eq. (9) fulfilled.

Moreover, we claim that the solution u^* of the implicit Eq. (12) is necessarily positive on some subset of Ω with non-zero measure. To see this, assume on contrary that $u^*(r) \leq 0$ for almost all $r \in \Omega$. Then it holds that $\mathrm{rect}(u^*(r)) = 0$ for almost all $r \in \Omega$, which implies that

$$\int_\Omega w_l(r,r')\mathrm{rect}(u^*(r'))dr' = 0, \quad \forall r \in \Omega.$$

In view of Eq. (12), this implies that $u^*(r) = 1$ for all $r \in \Omega$, thus leading to a contradiction. Consequently, as claimed, u^* is necessarily positive on some subset Ω^+ of Ω with non-zero measure. Recalling that Ω is here assumed to be a compact set, it follows that

$$\inf_{r\in\Omega}\int_\Omega e^{-|r-r'|^2/2\sigma_e^2}\mathrm{rect}(u^*(r))dr \geq \inf_{r\in\Omega}\int_{\Omega^+} e^{-|r-r'|^2/2\sigma_e^2}u^*(r)dr > 0,$$

which makes Eq. (10) satisfied. The conclusion then follows from Theorem 1.

6.3 Proof of Corollary 2

The following one-dimensional relation holds:

$$\int_a^b\int_a^b e^{-\frac{|x-y|^2}{2\sigma^2}}dxdy = \sqrt{\xi_{a,b}(\sigma)}. \tag{31}$$

In order to compute its two-dimensional counterpart, let $r = (r_1, r_2)$ and $r' = (r_1', r_2')$. Then, for $\Omega = [a,b] \times [a,b]$,

$$\int_\Omega\int_\Omega e^{-\frac{|r-r'|^2}{2\sigma^2}}dr'dr = \int_a^b\int_a^b\int_a^b\int_a^b \exp\left(-\frac{(r_1-r_1')^2 + (r_2-r_2')^2}{2\sigma^2}\right)dr_1'dr_2'dr_1dr_2.$$

Using Fubini's theorem, it follows that

$$\int_\Omega\int_\Omega e^{-\frac{|r-r'|^2}{2\sigma^2}}dr'dr = \int_a^b\int_a^b \exp\left(-\frac{(r_1-r_1')^2}{2\sigma^2}\right)\left(\int_a^b\int_a^b \exp\left(-\frac{(r_2-r_2')^2}{2\sigma^2}\right)dr_2dr_2'\right)dr_1dr_1'$$

From Eq. (31), this gives:

$$\int_\Omega \int_\Omega e^{-\frac{|r-r'|^2}{2\sigma^2}} dr'dr = \int_a^b \int_a^b \exp\left(-\frac{(r_1 - r_1')^2}{2\sigma^2}\right) \sqrt{\xi_{a,b}(\sigma)} dr_1 dr_1' \\ = \xi_{a,b}(\sigma). \tag{32}$$

The left-hand term of Eq. (13) then reads:

$$\int_\Omega \int_\Omega \left(K_e e^{-\frac{|r-r'|^2}{2\sigma_e^2}} - K_i e^{-\frac{|r-r'|^2}{2\sigma_i^2}}\right)^2 dr'dr = \int_\Omega \int_\Omega \left(K_e^2 e^{-\frac{|r-r'|^2}{\sigma_e^2}} + K_i^2 e^{-\frac{|r-r'|^2}{\sigma_i^2}} - 2K_e K_i e^{-\frac{|r-r'|^2 \sigma_e \sigma_i}{2\sqrt{\sigma_e^2+\sigma_i^2}}}\right) dr'dr$$

$$= K_e^2 \xi_{a,b}(\sigma_e/\sqrt{2}) + K_i^2 \xi_{a,b}(\sigma_i/\sqrt{2}) - 2K_e K_i \xi_{a,b}\left(\frac{\sigma_e \sigma_i}{\sqrt{\sigma_e^2 + \sigma_i^2}}\right),$$

which concludes the proof.

PRNG Seed

We ran both the stable and non-stable experiments ten times with different PRNG seeds. All the PRNG seeds we used are: $10, 74, 433, 721, 977, 1330, 3433, 5677, 9127, 7659$.

Acknowledgments

Not applicable.

Funding

This work was partially funded by grant ANR-17-CE24-0036.

Abbreviations

SOM Self-organizing Map

DSOM Dynamic Self-organizing Map

FFT Fast Fourier Transform

PRNG Pseudo Random Number Generator

Availability of data and materials

The source code used in this work for running the simulations, analysing the results and plotting the figures, is freely distributed under the GPL-3 License and can be found here: `https://github.com/gdetor/som_stability`.

Ethics approval and consent to participate

Not applicable.

Competing interests

All three authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Consent for publication

Not applicable.

Author's contributions

GID conceived the idea, wrote the code, designed and ran the numerical experiments, AC performed the mathematical derivation and analysis. GID, AC, and NPR contributed to preparing, writing, and revising the manuscript.

References

[1] Larry F Abbott and Sacha B Nelson. Synaptic plasticity: taming the beast. *Nature neuroscience*, 3:1178–1183, 2000.

[2] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions with Formula, Graphs and Mathematical Tables*. Dover Publications, 1983.

[3] S. Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological cybernetics*, 27(2):77–87, 1977.

[4] Paul C Bressloff. Spatiotemporal dynamics of continuum neural fields. *Journal of Physics A: Mathematical and Theoretical*, 45(3):033001, 2012.

[5] S. Coombes. Waves, bumps, and patterns in neural field theories. *Biological Cybernetics*, 93(2):91–108, 2005.

[6] Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[7] Pierre Demartines. Organization measures and representations of kohonen maps. In *First IFIP Working Group*, volume 10. Citeseer, 1992.

[8] Georgios Is Detorakis and Nicolas P Rougier. A neural field model of the somatosensory cortex: formation, maintenance and reorganization of ordered topographic maps. *PloS one*, 7(7):e40257, 2012.

[9] Georgios Is Detorakis and Nicolas P Rougier. Structure of receptive fields in a computational model of area 3b of primary sensory cortex. *Frontiers in Computational Neuroscience*, 8(76), 2014.

[10] Olivier Faugeras, François Grimbert, and Jean-Jacques Slotine. Absolute stability and complete synchronization in a class of neural fields models. *SIAM Journal on Applied Mathematics*, 69(1):205–250, 2008.

[11] Olivier Faugeras, Romain Veltz, and François Grimbert. Persistent neural states: stationary localized activity patterns in nonlinear continuous n-population, q-dimensional neural networks. *Neural computation*, 21(1):147–187, 2009.

[12] Mathieu N Galtier, Olivier D Faugeras, and Paul C Bressloff. Hebbian learning of recurrent connections: a geometrical perspective. *Neural computation*, 24(9):2346–2383, 2012.

[13] Kamil A Grajski and Michael Merzenich. Neural network simulation of somatosensory representational plasticity. In *Advances in neural information processing systems*, pages 52–59, 1990.

[14] Stephen Grossberg. Physiological interpretation of the self-organizing map algorithm. 1994.

[15] Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. Psychology Press, 2002.

[16] Samuel Kaski, Jari Kangas, and Teuvo Kohonen. Bibliography of self-organizing map (som) papers: 1981–1997. *Neural computing surveys*, 1(3&4):1–176, 1998.

[17] H. Khalil. *Nonlinear systems*. Macmillan Publishing Co., 2nd ed., New York, 1996.

[18] Robert T Knight, W Richard Staines, Diane Swick, and Linda L Chao. Prefrontal cortex regulates inhibition and excitation in distributed neural networks. *Acta psychologica*, 101(2):159–178, 1999.

[19] Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological cybernetics*, 43(1):59–69, 1982.

[20] Teuvo Kohonen. *Self-organizing maps*, volume 30. Springer, 2001.

[21] Leah A Krubitzer and Jon H Kaas. The organization and connections of somatosensory cortex in marmosets. *The Journal of Neuroscience*, 10(3):952–974, 1990.

[22] Rebecca A Mease, Michael Famulare, Julijana Gjorgjieva, William J Moody, and Adrienne L Fairhall. Emergence of adaptive computation by single neurons in the developing cortex. *The Journal of Neuroscience*, 33(30):12154–12170, 2013.

[23] Risto Miikkulainen, James A Bednar, Yoonsuck Choe, and Joseph Sirosh. *Computational maps in the visual cortex*. Springer Science & Business Media, 2006.

[24] Nasser M Nasrabadi and Yushu Feng. Vector quantization of images based upon the kohonen self-organizing feature maps. In *Proc. IEEE Int. Conf. Neural Networks*, volume 1, pages 101–105, 1988.

[25] Erkki Oja. Simplified neuron model as a principal component analyzer. *Journal of mathematical biology*, 15(3):267–273, 1982.

[26] Merja Oja, Samuel Kaski, and Teuvo Kohonen. Bibliography of self-organizing map (som) papers: 1998-2001 addendum. *Neural computing surveys*, 3(1):1–156, 2003.

[27] HX Qi, TM Preuss, and JH Kaas. Somatosensory areas of the cerebral cortex: architectonic characteristics and modular organization. *The senses: A comprehensive reference*, 6:143, 2008.

[28] Nicolas Rougier and Yann Boniface. Dynamic self-organising map. *Neurocomputing*, 74(11):1840–1847, 2011.

[29] Michael Schaefer, Hans-Jochen Heinze, and Michael Rotte. Task-relevant modulation of primary somatosensory cortex suggests a prefrontal–cortical sensory gating system. *Neuroimage*, 27(1):130–135, 2005.

[30] Joseph Sirosh and Risto Miikkulainen. Ocular dominance and patterned lateral connections in a self-organizing model of the primary visual cortex. *Advances in Neural Information Processing Systems*, pages 109–116, 1995.

[31] Mauro Tucci and Marco Raugi. Stability analysis of self-organizing maps and vector quantization algorithms. In *The 2010 International Joint Conference on Neural Networks (IJCNN)*, pages 1–5. IEEE, 2010.

[32] Romain Veltz and Olivier Faugeras. Local/global analysis of the stationary solutions of some neural field equations. *SIAM Journal on Applied Dynamical Systems*, 9(3):954–998, 2010.

[33] Stuart P Wilson, Judith S Law, Ben Mitchinson, Tony J Prescott, and James A Bednar. Modeling the emergence of whisker direction maps in rat barrel cortex. *PloS one*, 5(1):e8778, 2010.

[34] Jing Xing and George L Gerstein. Networks with lateral connectivity. i. dynamic properties mediated by the balance of intrinsic excitation and inhibition. *Journal of neurophysiology*, 75(1):184–199, 1996.

[35] Jing Xing and George L Gerstein. Networks with lateral connectivity. ii. development of neuronal grouping and corresponding receptive field changes. *Journal of neurophysiology*, 75(1):200–216, 1996.

[36] Jing Xing and George L Gerstein. Networks with lateral connectivity. iii. plasticity and reorganization of somatosensory cortex. *Journal of neurophysiology*, 75(1):217–232, 1996.

[37] Hujun Yin. Learning nonlinear principal manifolds by self-organising maps. In *Principal manifolds for data visualization and dimension reduction*, pages 68–95. Springer, 2008.