The Annals of Applied Probability
1995, Vol. 5, No. 4, 1177–1216

ON THE A.S. CONVERGENCE OF THE KOHONEN ALGORITHM WITH A GENERAL NEIGHBORHOOD FUNCTION

BY JEAN-CLAUDE FORT AND GILLES PAGÈS

Université Nancy I, Lab Samos, and Université Pierre et Marie Curie

Some existence and stability results for the equilibrium points of the one-dimensional Kohonen self-organizing neural network with two neighbors are extended to most nonincreasing neighborhood functions. All the functions mentioned in the neural literature are included. The assumption on the stimuli distribution is weakened, too. In the multidimensional setting, we derive from a general formula various stability and instability results.

Introduction. The Kohonen algorithm was originally devised and studied by Kohonen in 1982 (see Kohonen [9, 11]) as a model of self-organization of the neurotopic maps that lie in various areas of the brain. One may think, for example, of some retinotopic "projections" from the retina to the optic tectum. These maps are made up of nervous connections from the ganglion cells of the retina to the tectum. If we imagine the retina as a square grid, one important feature of these maps is that the "topology" of the retina grid is preserved by this set of connections in the following sense: neighbor cells in the retina are connected to neighbor cells in the tectum. It seems realistic to assume that the most important contribution to the formation of such maps comes from a self-organizing process governed by various selection rules of the spontaneous neural activity of the cells. This led Kohonen to devise a simple self-organizing algorithm, roughly "mimicking" the behavior of neural cells.

Although the algorithm turned out not to be very realistic as a biological model, its striking efficiency as an automatic classifier and quantifier drew the attention of some statisticians. Indeed, it provides a technique that both projects a high-dimensional set of points into a lower-dimensional space (say one, two or three dimensions) and preserves in some way the initial topology of the set. Then a visualization and a classification can be carried out in this lower-dimensional space. Various practical applications in data analysis have been or are being developed (see Varfis and Versino [17], Kohonen [10] and Cottrell, Letrémy and Roy [5]). From a computational point of view, the algorithm relies on the local updating of some unit-to-unit connections using a cooperation/competition rule when stimulated by an input. So, the imple-

Received May 1994; revised July 1995.

AMS 1991 *subject classifications*. Primary 60J05; secondary 60K30.

Key words and phrases. Self-organization, quantization, stochastic algorithms, neural networks.

mentation on a computer is straightforward and numerically robust. That partially explains its popularity.

THE INVESTIGATED MODEL. One considers a set of *units* (or *neurons*) represented by a finite subset I of $\mathbf{Z}^d$ equipped with a neighborhood function $\sigma\colon I - I := \{i - j,\ i, j \in I\} \to [0,1]$ satisfying $\sigma(k) = \sigma(-k)$, $\sigma(0) = 1$. The neighborhood function σ is the key of the self-organization property: it measures the strength of the connection between two units in I; σ is the cooperation *medium* between units. It is a priori defined and describes the topology of the set I. For instance, if $\sigma(i - j) = 1$, then i and j are fully connected, so they strongly influence each other; on the other hand, if $\sigma(i - j) = 0$, i and j are totally disconnected and have no interaction. Some anisotropic neighborhood function $\sigma(i, j)$ could be considered, too.

Next one considers a sequence $(\omega^t)_{t \ge 1}$ of i.i.d. $[0,1]^d$-valued stimuli with distribution μ. The vector $\omega \in [0,1]^d$ represents the emitted activity of a cell lying in a remote area Ω of the nervous system when excited. So the cells in Ω are assumed to get excited at random, independently, with respect to μ. The described phenomenon is that the unit set I intends to "project" onto Ω. The (technical) term *project* roughly means that the neurons in I "get connected to some cells in" Ω.

At this stage of the modeling, the idea is to identify the state space of the connections starting from I to the state space of the stimuli emitted by cells in Ω, that is, $[0,1]^d$ (see below): the closer a connection is to a stimulus ω, the stronger is the response of the corresponding unit $i \in I$. Then each unit i is mapped to a $[0,1]^d$-valued *weight* vector x_i which is simply the connection mentioned above. The weight vector x_i may be understood as the center of gravity of the actual connections of cell $i \in I$ into the area Ω identified with $[0,1]^d$.

The Kohonen algorithm is an adaptive, unsupervised learning process. It builds up from the stimulus sequence $(\omega^t)_{t \ge 1}$ a family of weight vectors $(x_i^*)_{i \in I}$ that both *quantize* the stimuli distribution μ and preserve (in some sense) the neighborhood structure provided by σ on I; that is:

1. the number of units i whose connections (or weight vectors) x_i^* lie in an area $\mathscr{A}$ of $[0,1]^d$ is approximately proportional to $\mu(\mathscr{A})$ (*quantization*);
2. two close units i and j in I—in the sense that $\sigma(i - j) \approx 1$—have close weight vectors, that is, $x_i \approx x_j$ (*organization*).

When $d = 1$, $I := \{1, \dots, n\}$ and $\sigma(k) = \mathbf{1}_{\{|k| \le p\}}$, $p \ge 1$, the algorithm is known as the $2p$-neighbor algorithm. Theoretically speaking, the most investigated settings undoubtedly are the zero- and two-neighbor ones (see Cottrell and Fort [4] and Bouton and Pagès [2, 3]), while long-range functions such as $\sigma(k) := e^{-k/T}$ are often encountered in simulations. When $d = 2$, in most implemented cases, the unit set I is a rectangle. The basic neighborhood functions σ are $\{0,1\}$-valued, for example, the four-neighbor one which makes up a vertical cross parallel to the axis or the eight-neighbor one which surrounds each unit by a square. In higher-dimensional settings, one may consider a (suitable) subset I of $\mathbf{Z}^d$ and define σ as $\sigma(k) := \mathbf{1}_{\{N(k) \le p\}}$ or $\sigma(k) := e^{-N(k)/T}$, where $N(\cdot)$ denotes a norm on $\mathbf{Z}^d$.

THE ALGORITHM. From now on μ will be a *strongly* diffuse Borel probability measure (i.e., μ assigns no mass to hyperplanes) with a $[0,1]^d$ support.

Let $X^0 = (x_i^0)_{i \in I} \in ([0,1]^d)^I$ be the initial weight vector state and X^t be the state of the weight vectors at time $t \in \mathbb{N}$. At time $t+1$, X^t is updated as follows:

(i) *Competition phase.* Computation of the winning unit $i^{t+1} = i(\omega^{t+1}, X^t) := \arg\min_{k \in I} \|\omega^{t+1} - X_k^t\|$. In case of conflict, one takes the lexicographic minimum where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$.

(ii) *Cooperation phase.*

$$\forall j \in I, \qquad X_j^{t+1} = X_j^t - \varepsilon_{t+1}\sigma(i^{t+1} - j)\big(X_j^t - \omega^{t+1}\big), \tag{1}$$

where $(\varepsilon_t)_{t \ge 1}$ is a sequence of $(0,1)$-valued real numbers. The selected or winning unit is the one that gives the maximum response to the current stimulus ω^{t+1}: it is the most connected to the cell in Ω that emitted ω^{t+1}. The effect of (1) is to make X_j^t closer to ω^{t+1} proportionally to $\sigma(i^{t+1} - j)$, so that if $\sigma(i^{t+1} - j) \approx 1$, X_j^t both reinforces its response to ω^{t+1} and gets closer to the weight vector of unit i^{t+1}.

AN APPLICATION. The automatic coding of a continuous set of grey levels with light sensors provides a simple example of application. Each sensor $i \in I$ is sensitive to a small range of light signals (grey levels). We approximate this range by its mean value x_i^0. This value can be modified as time passes. The sensors are connected along a straight line: except for the edges, each sensor is connected to two neighbor sensors (see Figure 1). At time $t+1$ a grey level signal ω^{t+1} is sent and the most sensitive sensor is selected. Then its sensitivity is slightly tuned toward ω^{t+1} and the same is done for its neighbors. At the end of the process the sensitivities $(X_i^\infty)_{i \in I}$ are ordered in an increasing or decreasing way and the number of sensors sensitive to a given range of grey levels is proportional to its frequency in the sequence $(\omega^t)_{t \ge 1}$. Both self-organization and quantization occurred.

A two-dimensional example is displayed in Figure 2: $7 \times 7 = 49$ units are arranged on a square grid, the neighborhood function is set at 1 for the eight nearest neighbors (when in I) and 0 elsewhere. The stimuli distribution μ is $U([0,1]^2)$ and the step ε_t slowly goes from 0.1 down to 0.01. One starts from a random value $(x_i^0)_{i \in I}$.

SOME MATHEMATICAL BACKGROUND AND A FEW RESULTS. Putting the recursive updating of the algorithm (1) in a shorter form yields

$$\forall\, t \in \mathbb{N}, \qquad X^{t+1} = X^t - \varepsilon_{t+1} H^\sigma(X^t, \omega^{t+1}); \tag{2}$$

$(X^t)_{t \ge 0}$ is a Markov chain. The chain is homogeneous if $\varepsilon_t = \varepsilon > 0$ ($\mathbb{P}_x$ will then denote its distribution starting from x). Let $D_I := \{x \in ([0,1]^d)^I | x_i \ne x_j$

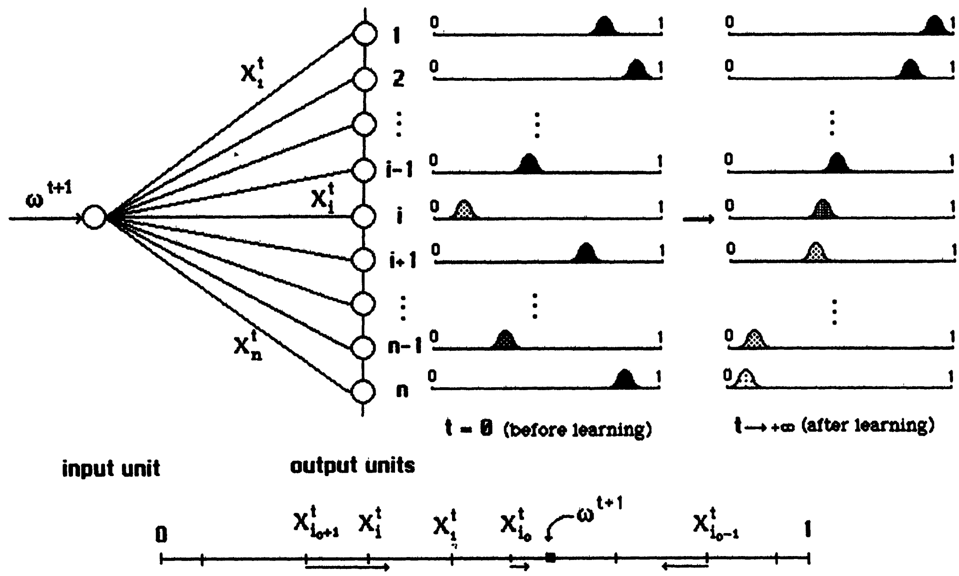


FIG. 1. *Representation of the weights in the stimuli space* (d = 1).

if $i \neq j\}$ be the set of $([0,1]^d)^I$-valued vectors with pairwise distinct components. If $x \in D_I$, then, $\mathbb{P}_x$-a.s., $X^t \in D_I$ for every $t \in \mathbb{N}$.

Referring to the Kushner and Clark approach (see Kushner and Clark [13]), the "conditional" a.s.-convergence of X^t is related to the stability of its so-called ordinary differential equation (ODE) which describes the mean behavior of the stochastic process. To write it down, we need to define the Voronoï tessellation.

DEFINITION 1. The Voronoï tessellation $\{C_i(x)\}_{i \in I}$ of $x \in D_I$ is defined by

$$\forall\, i \in I, \qquad C_i(x) := \left\{\omega \in [0,1]^d \,|\, \|x_i - \omega\| < \|x_k - \omega\| \text{ if } k \neq i\right\}.$$

The average function h^σ of the algorithm is defined by $h^\sigma(x) := \mathbb{E}(H^\sigma(x, \omega^1))$; h^σ is obviously continuous on D_I since μ is strongly diffuse. Then the ODE can be written as

$$\begin{aligned} &\dot{x} = -h^\sigma(x), \qquad x_0 \in D_I, \\ (3) \qquad &h_i^\sigma(x) := \sum_{k \in I} \sigma(k-i) \int_{C_k(x)} (x_i - \omega)\mu(d\omega), \qquad i \in I. \end{aligned}$$

When $\sigma(k) = \mathbf{1}_{\{k=0\}}$, $(X^t)_{t \in \mathbb{N}}$ is the so-called quantization algorithm (or Kohonen algorithm with zero neighbor). It provides a "skeleton" for μ (see [2], [11] and [14]). The self-organization problem is then irrelevant. Moreover, h^σ

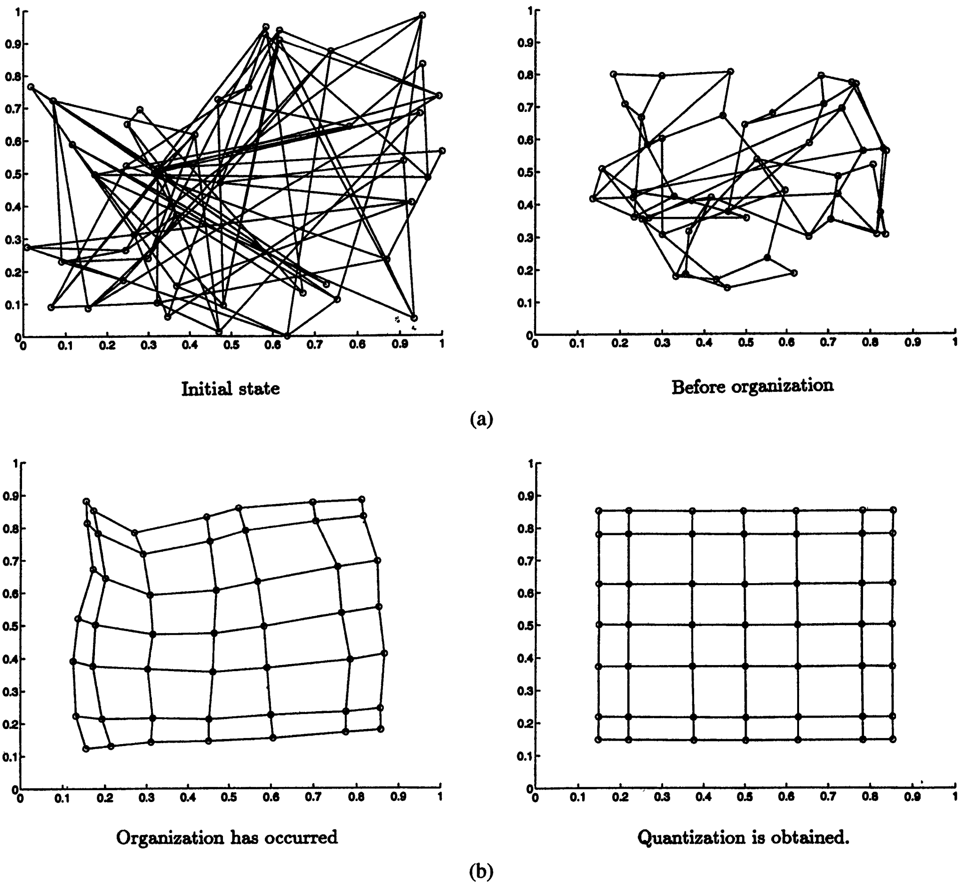


FIG. 2. ∘ *denotes the weight vectors* X_i^t; —*shows the connections* (*i.e., when* 1) *between units.*

is the gradient of

$$E_0(x) := \tfrac{1}{2}\int_{[0,1]^d} \min_{k\in I}\|x_k - \omega\|^2 \mu(d\omega) \quad \text{at every point } x \in D_I,$$

and $(X^t)_{t\in\mathbb{N}}$ is then a D_I-valued stochastic gradient descent whose attractors are the local minima (assumed to be isolated) of E_0 (see [14]). Actually it is easy to prove that h^σ is a gradient iff there is some $a \in [0,1]$ such that $\sigma(0) = 1$ and $\sigma(k) = a$, $k \neq 0$. Its potential E_a is

$$E_a(x) := \frac{1-a}{2}\int_{[0,1]^d} \min_{k\in I}\|x_k - \omega\|^2 \mu(d\omega) + \frac{a}{2}\sum_{i=1}^{n}\int_{[0,1]^d}\|x_i - \omega\|^2 \mu(d\omega).$$

The related algorithm is purely quantifying too and has no self-organizing property.

Let us turn now to some background on the convergence result for stochastic algorithms. Subsequently, the superscript σ will be temporarily dropped, although $(X^t)_{t\in\mathbb{N}}$ is still supposed to be formally defined by an equation similar to (2). We now recall the well-known result of Kushner and Clark about conditional convergence to an (asymptotically) stable equilibrium x^* of the ODE; x^* is said to be a stable equilibrium if it owns a *stable attracting* area Γ_{x^*}. Let $x(x^0,\cdot)$ denote any solution of the ODE starting from x^0 (uniqueness is not required a priori); Γ_{x^*} is defined as a *neighborhood* of x^* satisfying the following:

(i) $\forall\, x^0\in\Gamma_{x^*},\ \forall\, u\in\mathbb{R}_+,\ x(x^0,u)\in\Gamma_{x^*}$;
(ii) $\forall\, x^0\in\Gamma_{x^*},\ \lim_{u\to+\infty} x(x^0,u)=x^*$;
(iii) $\forall\, K\subset\Gamma_{x^*}$, K compact set, $\forall\,\varepsilon>0$, $\exists\,\eta_{\varepsilon,K}>0$ such that

$$\forall\, x^0\in K,\ \|x^0-x^*\|\le\eta_{\varepsilon,K}\Rightarrow \sup_{u\in\mathbb{R}_+}\left\|x(x^0,u)-x^*\right\|\le\varepsilon.$$

A sufficient condition for the existence of such a stable attracting area Γ_{x^*} is that the mean function h is differentiable at x^* with a gradient $\nabla h(x^*)$ satisfying:

all the eigenvalues of $\nabla h(x^*)$ have positive real part.

This comes from the existence of a local strict Liapounov function V for the ODE given by

$$V(x):=\int_0^{+\infty}\left\|e^{-s\nabla h(x^*)}(x-x^*)\right\|^2\,ds.$$

The celebrated Kushner–Clark theorem (see [13]) reads as follows.

THEOREM 1. *Assume that* $\sum_{t\ge0}\varepsilon_t=+\infty$ *and* $\sum_{t\ge0}\varepsilon_t^2<+\infty$. *Let* x^* *be a zero of* h *and let* K *be a compact subset of its stable attracting area* Γ_{x^*}. *Then the sequence* $(X^t)_{t\ge0}$ *"conditionally" (a.s.) converges to* x^*; *that is,*

$$X^t\to x^*\qquad \text{as } t\to+\infty$$

on the event $A_K^{x^*}:=\{(X^t)_{t\ge0}$ *is bounded and* $X^t\in K$ *infinitely often*$\}$.

This leads to the following definition.

DEFINITION 2. If $(X^t)_{t\in\mathbb{N}}$ satisfies Theorem 1 and x^* is a stable equilibrium, $(X^t)_{t\in\mathbb{N}}$ "conditionally" a.s. converges to x^*.

A "regular" a.s. convergence general result is available as well. It relies on some global assumptions on h (see, e.g., Duflo [6]). Some background is provided in Section 2.

However, in actual simulations, the Kohonen algorithm, as well as many stochastic algorithms, is implemented with a (small) nonvanishing step parameter. Thus it is also interesting to pay attention to the fluctuations of the constant step algorithm ($\varepsilon_t=\varepsilon$) around the trajectories of the ODE as $\varepsilon\to0$.

Assume that for every starting value x^0 in a domain D, the ODE admits a unique maximal solution $x(x^0, \cdot)$ living in D (or its closure) up to $+\infty$.

Let $(X^{\varepsilon, t})_{t \geq 0}$ be the algorithm with constant step $\varepsilon > 0$ starting from x^0. One defines the stepwise functions $X^{(\varepsilon)}$ by

$$X_s^{(\varepsilon)} := X^{\varepsilon, t} \quad \text{if } s \in [t\varepsilon, (t+1)\varepsilon[.$$

Then the functions $X^{(\varepsilon)}$ approximate $x(x^0, \cdot)$ in the following sense (see Jacod and Shiryaev [8] and Kushner [12]):

$$(4) \quad X_{\cdot}^{(\varepsilon)} - x(x^0, \cdot) \to_{\mathbb{P}(U_K)} 0 \text{ and } \frac{X_{\cdot}^{(\varepsilon)} - x(x^0, \cdot)}{\sqrt{\varepsilon}} \to_{\mathscr{D}(U_K)} Z_{\cdot} \text{ as } \varepsilon \to 0,$$

where U_K is for the compact convergence topology and Z denotes a (possibly) multidimensional Brownian diffusion process defined by

$$dZ_s = -\nabla h(x(x^0, s))\,ds + \left(\int_0^1 H^t H(x(x^0, s), \omega)\mu(d\omega) - h^t h(x(x^0, s)) \right)^{1/2} dW_s.$$

One verifies once again that the asymptotic behavior of the algorithm is driven by ∇h along the path $x(x^0, \cdot)$ of the ODE.

Let us come back to the original Kohonen algorithm. To situate this work precisely, we finally recall the main previous mathematical results. All of them are related to the one-dimensional case. Let $I = \{1, 2, \dots, n\}$ and $\sigma(k) = \mathbf{1}_{\{|k| \leq 1\}}$ (algorithm with two neighbors). Let $F_n^+ := \{x \in \mathbb{R}^n,\ 0 < x_1 < x_2 < \cdots < x_n < 1\}$ and $F_n^- := \{x \in \mathbb{R}^n,\ 0 < x_n < x_{n-1} < \cdots < x_1 < 1\}$. In [4], [2] and [3] the following results are established.

Self-organization (with constant step $\varepsilon_t = \varepsilon$). The sets F_n^+ and F_n^- are absorbing sets, and the hitting time of $F_n := F_n^+ \cup F_n^-$ is $\mathbb{P}_x$-a.s. finite and admits an exponential moment, uniformly with respect to $x \in [0,1]^n$ (the assumption supp$(\mu) = [0,1]$ can be relaxed; see [2]).

Convergence in distribution (with a constant step $\varepsilon_t = \varepsilon$). There exists a unique probability measure $\nu_\varepsilon^{\pm}$ on $F_n^{\pm}$ s.t., for every $x \in F_n^{\pm}$, $\mathbb{P}_x$-a.s. $X^t \to_{\mathscr{D}} \nu_\varepsilon^{\pm}$ (actually the chain is Doeblin recurrent).

Convergence (with a "decreasing" step). If $\sum_t \varepsilon_t = +\infty$ and $\sum_t \varepsilon_t^2 < +\infty$:

(a) the average function h^σ has (at least) one equilibrium point x^* in F_n^+;

(b) if μ has a strictly log-concave density $f > 0$ on $(0,1)$, then any equilibrium x^* is stable. Hence, $X^t \to x^*$ conditionally a.s.;

(c) if $\mu = U([0,1])$, h^σ has a *unique* equilibrium x^* in F_n^+ and $X^t \to x^*$ a.s.

Therefore, as far as mathematical treatment is concerned, rigorous results are not numerous. The one-dimensional results rely on nontrivial Markov material and are rather technical. No completely rigorous higher-dimensional results were proven so far (see Ritter and Schulten [16] for a physical approach in a two-dimensional setting with uniform stimuli).

This paper is devoted to the study of the Kohonen algorithm in the quantization phase for general functions σ in both one-dimensional and multidimensional settings. The quantization phase corresponds to the converging properties of the algorithm, so we assume that the usual decreasing step assumption holds:

$$\sum_{t\geq 0} \varepsilon_t = +\infty, \qquad \sum_{t\geq 0} \varepsilon_t^2 < +\infty.$$

Our aim is to provide some simple conditions on σ and μ that imply some (a.s.) converging properties for the algorithm. All the one-dimensional results formerly established for some particular functions σ are included in ours and we now reach all the σ-functions mentioned in the neural literature (see Ritter, Martinetz and Schulten [15] and [11]). Furthermore, we relax the assumption on the stimuli distribution μ (compare the existing results above and those of Section 1.2).

The most general result is Theorem 2, which states that (under suitable assumptions) all the equilibrium points of the ODE are stable, so that, once organization occurs, the algorithm conditionally a.s. converges to one of them. Regular a.s. convergence is established for uniformly distributed stimuli (Theorem 3).

In the multidimensional case ($d \geq 2$) several stability and instability results are investigated. First, a general formula for ∇h^σ is established (Theorem 4) which can be used numerically. Then Theorem 5 provides various stability and instability results when μ is a product of independent marginals. We check that if a product-grid made up of marginal equilibrium points is an equilibrium point for the whole algorithm, it is usually not stable [Theorem 5(a)]. The particular case of the zero-neighbor setting and $U([0,1]^2)$ i.i.d. stimuli is completely solved [Theorem 5(b)]. We finally investigate the case of a d_1-dimensional unit set I stimulated by some i.i.d. $[0,1]^{d_1+d_2}$-valued r.v. with a small variance along $[0,1]^{d_2}$ [Theorem 5(c) for the Kohonen string and Theorem 6]. As a conclusion, some simulations illustrate most of the theoretical or numerical results.

The paper is basically divided into two parts. The first part, made up of the first two sections, is devoted to the one-dimensional setting: Section 1 deals with conditional a.s. convergence and Section 2 shows the a.s. convergence in the uniformly distributed case. The second part (Sections 3 and 4) deals with the higher-dimensional results. Section 4 presents the simulations and some provisional remarks.

The canonical inner product in any $\mathbb{R}^p$ will be denoted $(\cdot\,|\,\cdot)$.

1. General results in the one-dimensional setting. Throughout this section, we study the one-dimensional Kohonen algorithm. As $d = 1$ the neighborhood function is simply defined by a function $\sigma: \mathbb{N} \to [0,1]$ with $\sigma(0) = 1$. We set for notational convenience $\sigma(k) = 0$, $k \geq n$. The following proposition, whose proof is obvious, describes the (two) absorbing classes of the algorithm.

PROPOSITION 1. *If σ is nonincreasing, the two convex subsets F_n^+ and F_n^- are left stable by the algorithm.*

From now on, as far as convergence is concerned, we will always assume that the initial value x of the algorithm lies in F_n^+.

Let $x \in F_n^+$. We define the $(n+1)$-tuples $\tilde{x}$ by $\tilde{x}_1 = 0$, $\tilde{x}_k = (x_k + x_{k-1})/2$, $2 \leq k \leq n$, $\tilde{x}_{n+1} = 1$. As $x \in F_n^+$, then the functions H^σ and h^σ are given by

$$
\begin{aligned}
&\forall\, i \in \{1,\dots,n\}, \qquad H_i^\sigma(x,\omega) = \sum_{k=1}^{n} \sigma(|k-i|)(x_i - \omega)\mathbf{1}_{]\tilde{x}_k, \tilde{x}_{k+1}]}(\omega), \\
(5)\qquad &\forall\, i \in \{1,\dots,n\}, \qquad h_i^\sigma(x) \quad = \sum_{k=1}^{n} \sigma(|k-i|) \int_{]\tilde{x}_k, \tilde{x}_{k+1}]} (x_i - \omega)\mu(d\omega).
\end{aligned}
$$

Notice that if μ is diffuse, then h^σ *has a continuous extension on the closure* $\overline{F_n^+}$ of F_n^+ given by (5). From now on, when $d = 1$, h^σ will always denote this extension. After these preliminary results, we are in a position to discuss the existence and the location of the equilibrium points.

1.1. *Existence of an equilibrium point.* The existence of an equilibrium point x^* is provided by the following proposition.

PROPOSITION 2. *If the probability measure μ is diffuse, then there exists at least one equilibrium point x^* in the closure $\overline{F_n^+}$ of F_n^+, that is, satisfying $h^\sigma(x^*) = 0$.*

PROOF. If $\omega \in [0,1]$, F_n^+ is stable under $\mathrm{Id}_{F_n^+} - \frac{1}{2}H^\sigma(\cdot,\omega)$. As F_n^+ is a convex set, $(\mathrm{Id}_{F_n^+} - \frac{1}{2}h^\sigma)(F_n^+) \subset \overline{F_n^+}$. Hence $\overline{F_n^+}$ is left stable by the continuous extension of $\mathrm{Id}_{F_n^+} - \frac{1}{2}h^\sigma$ on $\overline{F_n^+}$. The Brouwer fixed point theorem completes the proof. □

We will now prove that, under suitable assumptions, any equilibrium point x^* actually lies in F_n^+.

PROPOSITION 3. *Assume that μ is diffuse and σ is nonincreasing.*

(a) *If* $\{0,1\} \subset \operatorname{supp}(\mu)$, *then* $0 < x_1^* \leq x_n^* < 1$.

(b) *If* $\operatorname{supp}(\mu) = [0,1]$ *and if σ satisfies assumption* $(\mathscr{S})$,

$$
\begin{aligned}
(\mathscr{S}) \equiv (n \geq 2 \text{ and } \sigma(1) < 1) \text{ or } (n \geq 3 \text{ and } \sigma(2) < 1) \\
\text{or } (n \geq 5 \text{ and } \sigma(3) < 1),
\end{aligned}
$$

then x^* lies in F_n^+.

PROOF. (a) If $x_n^* = 0$, then $h_n^\sigma(x^*) = 0$ amounts to $\sigma(|n-n|)\int_0^1(0-\omega)\mu(d\omega) = 0$, which cannot hold since $\operatorname{supp}(\mu) \neq \{0\}$. So, $x_n^* > 0$; if $x_{n-1}^* = 0$, then $\tilde{x}_n^* = x_n^*/2 > 0$ and $h_{n-1}^\sigma(x^*) = 0$ yields

$$\sigma(0)\int_0^{x_n^*/2}(0-\omega)\mu(d\omega) + \sigma(1)\int_{x_n^*/2}^1(0-\omega)\mu(d\omega) = 0.$$

Both terms are nonnegative; hence $\int_0^{x_n^*/2}\omega\mu(d\omega) = 0$, which is impossible as $\mu([0, x_n^*/2]) > 0$. A descending induction finally yields $x_i^* > 0$, $1 \le i \le n$. An analogous proof yields $x_i^* < 1$, $1 \le i \le n$.

(b) We will investigate all the possible values for the length $p \ge 2$ of the *largest* cluster of packed components. Assume that, for some i, $x_i^* = x_{i+1}^* = \cdots = x_{i+p-1}^*$. Then $1 \le i \le n-p+1$ and $\tilde{x}_{i+1}^* = \cdots = \tilde{x}_{i+p-1}^* = x_i^*$ as well. Equations $h_i^\sigma(x^*) = 0$ and $h_{i+p-1}^\sigma(x^*) = 0$ now give

$$\sum_{k=1}^{i}\int_{\tilde{x}_k^*}^{\tilde{x}_{k+1}^*}(x_i^*-\omega)\mu(d\omega)\sigma(|k-i|)$$
$$= \sum_{k=i+p-1}^{n}\int_{\tilde{x}_k^*}^{\tilde{x}_{k+1}^*}(\omega-x_i^*)\mu(d\omega)\sigma(|k-i|),$$

$$\sum_{k=1}^{i}\int_{\tilde{x}_k^*}^{\tilde{x}_{k+1}^*}(x_i^*-\omega)\mu(d\omega)\sigma(|k-(i+p-1)|)$$
$$= \sum_{k=i+p-1}^{n}\int_{\tilde{x}_k^*}^{\tilde{x}_{k+1}^*}(\omega-x_i^*)\mu(d\omega)\sigma(|k-(i+p-1)|).$$

Subtracting these two equalities yields

$$\sum_{k=1}^{i}\int_{\tilde{x}_k^*}^{\tilde{x}_{k+1}^*}\underbrace{(x_i^*-\omega)}_{\ge 0}\mu(d\omega)\underbrace{[\sigma(i-k)-\sigma(i+p-1-k)]}_{\ge 0}$$
$$= -\sum_{k=i+p-1}^{n}\int_{\tilde{x}_k^*}^{\tilde{x}_{k+1}^*}\underbrace{(\omega-x_i^*)}_{\ge 0}\mu(d\omega)\underbrace{[\sigma(k-i-p+1)-\sigma(k-i)]}_{\ge 0}.$$

The terms on the two sides of the equality are of opposite sign, so they are 0:

$$[\sigma(|k-i|)-\sigma(|k-i-p+1|)]\int_{\tilde{x}_k^*}^{\tilde{x}_{k+1}^*}|x_i^*-\omega|\mu(d\omega) = 0 \tag{6}$$
$$\forall\, k \in \{1,\dots,i\} \cup \{i+p-1,\dots,n\}.$$

(i) If $p = n$, then $x_k^* = x$, $1 \le k \le n$. Then $i = 1$ and (6) reads

$$\int_0^x(x-\omega)\mu(d\omega)(\sigma(0)-\sigma(n-1)) = \int_x^1(\omega-x)\mu(d\omega)(\sigma(0)-\sigma(n-1))$$
$$= 0,$$

which is impossible for $\sigma(0) - \sigma(n-1) > 0$ by assumption ($\mathscr{S}$) and $\mathrm{supp}(\mu) \neq \{x\}$.

(ii) If $p = n-1$, then $n \geq 3$ and $i = 1, 2$. If $i = 1$, then $x_{n-1}^* < x_n^*$ and (6) yields, when $k = n-1, n$,

$$(\sigma(0) - \sigma(n-2)) \int_{\tilde{x}_{n-1}^*}^{\tilde{x}_n^*} |x_1^* - \omega| \, \mu(d\omega) = 0$$

or

$$(\sigma(1) - \sigma(n-1)) \int_{\tilde{x}_n^*}^{\tilde{x}_{n+1}^*} |x_1^* - \omega| \, \mu(d\omega) = 0.$$

Due to assumption ($\mathscr{S}$), one of these two integrals is 0. Since $\mathrm{supp}(\mu) = [0,1]$, either $x_{n-1}^* = x_n^*$ or $x_{n-1}^* = x_n^* = 1$, so $x_1^* = \cdots = x_{n-1}^* = x_n^*$, which is impossible by the induction assumption. If $i = 2$, the proof works the same way round.

(iii) If $p = n-2$, then $n \geq 4$. Still applying (6) shows that there exist $k_1 \in \{i-1, i+p\}$ and $k_2 \in \{i, i+p+1\}$ (the one that belongs to $\{1, \dots, n\}$) such that

$$(\sigma(1) - \sigma(n-2)) \int_{\tilde{x}_{k_1}^*}^{\tilde{x}_{k_1+1}^*} |x_i^* - \omega| \, \mu(d\omega) = 0$$

and

$$(\sigma(0) - \sigma(n-3)) \int_{\tilde{x}_{k_2}^*}^{\tilde{x}_{k_2+1}^*} |x_i^* - \omega| \, \mu(d\omega) = 0.$$

If $n = 4$, then assumption ($\mathscr{S}$) implies $\sigma(1) \neq \sigma(n-2)$ or $\sigma(0) \neq \sigma(n-3)$; the same argument as above implies that at least one more component (e.g., x_{i-1}^* or x_{i+p}^*) belongs to the cluster.

If $n \geq 5$, then $\sigma(0) = \sigma(n-3)$ and $\sigma(1) = \sigma(n-2)$ is impossible unless $\sigma(0) = \sigma(3)$, so we can conclude as with $n = 4$.

(iv) Assume now that $p \leq n-3$ (which implies $n \geq 5$). Then there exist $k_1 \in \{i-2, i+p+2\}$, $k_2 \in \{i-1, i+p\}$ and $k_3 \in \{i, i+p+1\}$ such that

$$|\sigma(2) - \sigma(p+1)| \int_{\tilde{x}_{k_1}^*}^{\tilde{x}_{k_1+1}^*} |x_i^* - \omega| \, \mu(d\omega) = 0,$$

$$|\sigma(1) - \sigma(p)| \int_{\tilde{x}_{k_2}^*}^{\tilde{x}_{k_2+1}^*} |x_i^* - \omega| \, \mu(d\omega) = 0,$$

$$|\sigma(0) - \sigma(p-1)| \int_{\tilde{x}_{k_3}^*}^{\tilde{x}_{k_3+1}^*} |x_i^* - \omega| \, \mu(d\omega) = 0.$$

The three integrals cannot be nonzero simultaneously: if so, σ satisfies $\sigma(0) = \sigma(p-1) \leq \sigma(1) = \sigma(p) \leq \sigma(2) = \sigma(p+1) \leq \sigma(3)$, which would contradict assumption ($\mathscr{S}$) since $n \geq 5$. If the second or the third integral is 0, one concludes as above. If *only* the first integral is 0, then $\sigma(0) =$

$\sigma(p-1) \le \sigma(1) = \sigma(p) \le \sigma(p-1) \le \sigma(0)$, that is, $\sigma(0) = \sigma(p-1) = \sigma(p)$. Assumption $(\mathscr{S})$ implies $p = 2$. Two cases are to be inspected:

Case 1 ($k_1 \ne 1, \eta$). Then $x^*_{k_1-1} = x^*_{k_1} = x^*_{k_1+1}$. It makes up a three-element cluster of packed components, which contradicts $p = 2$.

Case 2 ($k_1 = 1, n$). Assume, for example, that $k_1 = 1$; then $i = 3$ and $\int_0^{\tilde{x}^*_2} |x^*_3 - \omega| \mu(d\omega) = 0$, which implies that $0 = x^*_1 = x^*_2$. This is impossible since $x^* \in (0,1)^n$. □

1.2. *Stability of an equilibrium point.* From now on, we suppose that the probability distribution μ has a density f, *continuous* on $(0,1)$. Then (5) becomes

$$\forall\, i \in \{1, \dots, n\}, \qquad h_i^\sigma(x) = \sum_{k=1}^{n} \sigma(|k-i|) \int_{\tilde{x}_k}^{\tilde{x}_{k+1}} (x_i - \omega) f(\omega)\, d\omega.$$

Without loss of generality on f, we will adopt throughout the text the following convention:

$$f(\tilde{x}_1) = f(0^-) = 0 \quad \text{and} \quad f(\tilde{x}_{n+1}) = f(1^+) = 0.$$

PROPOSITION 4. *If the density f is continuous on* $(0,1)$, *then h is continuously differentiable on F_n^+ and* $\nabla h^\sigma(x) := \mathrm{Diag}[\zeta_1, \dots, \zeta_n] + [\alpha_{ij}]_{1 \le i,j \le n}$, *where*

$$\zeta_i := \sum_{k=1}^{n} \sigma(|k-i|) \int_{\tilde{x}_k}^{\tilde{x}_{k+1}} f(\omega)\, d\omega, \qquad 1 \le i \le n,$$

$$\alpha_{ij} := \frac{\sigma(|j-(i+1)|) - \sigma(|j-i|)}{2} (x_i - \tilde{x}_j) f(\tilde{x}_j) + \frac{\sigma(|j-i|) - \sigma(|j-i+1|)}{2} (x_i - \tilde{x}_{j+1}) f(\tilde{x}_{j+1}). \tag{7}$$

APPLICATION (a first result for the algorithm with constant step). It is straightforward from the above formula that if the density f is *bounded*, ∇h^σ is also bounded on F_n^+; so h^σ is Lipschitz on $\overline{F_n^+}$ which, in turn, implies that the ODE $\dot{x} = -h^\sigma(x)$ admits a unique maximal solution $x(x^0, \cdot)$ starting from $x^0 \in F_n^+$ and living in $\overline{F_n^+}$. On the other hand, all the applications $\mathrm{Id}_{\overline{F_n^+}} - \varepsilon h^\sigma$ leave $\overline{F_n^+}$ stable (cf. Proposition 3), so the stepwise Euler approximations of $x(x^0, \cdot)$ defined for every $\varepsilon > 0$ by

$$x_0^{(\varepsilon)} := x^0 \quad \text{and} \quad x_s^{(\varepsilon)} := x_{\varepsilon t}^{(\varepsilon)} - \varepsilon h\big(x_{\varepsilon t}^{(\varepsilon)}\big), \qquad s \in [\varepsilon t, \varepsilon(t+1)[,$$

live in $\overline{F_n^+}$ up to $+\infty$. As $x^{(\varepsilon)} \to_{U_K} x(x^0, \cdot)$, $x(x^0, \cdot)$ live in $\overline{F_n^+}$ up to $+\infty$. Then the results of functional weak convergence (4) hold for the algorithm with constant step (see the Introduction).

The following theorem—the main result of this section—deals with the algorithm with *decreasing* step.

THEOREM 2. *Assume that the density f is continuous on* $[0,1]$ *and* $f>0$ *on* $(0,1)$. *Let* $x^* \in \overline{F_n^+}$ *be an equilibrium point. If* σ *is nonincreasing and satisfies assumption* $(\mathscr{S})$, *if* $\sum_t \varepsilon_t = +\infty$ *and* $\sum_t \varepsilon_t^2 < +\infty$ *and if one of the following assumptions holds*:

(i) $\log f$ *is concave on* $[0,1]$ *and* $f(0)+f(1)>0$;
(ii) $\log f$ *is strictly concave on* $[0,1]$;

then $x^* \in F_n^+$ *and* X^t *a.s. conditionally converges to* x^* (*see Definition* 2).

The rest of the section is devoted to the proof of this theorem, but before going into technicalities, let us say that the approach basically consists of showing that the real parts of the eigenvalues of the $\nabla h^{\sigma}(x^*)$ are positive. To this end, we will establish a variant of the celebrated Gershgorin lemma on matrices with dominating diagonal (see Gantmacher [7] for the original one).

LEMMA 5. *Let* $A := [a_{ij}]_{1\le i,j\le n}$ *be a real-valued matrix and* $p \in \{1,\dots, n-1\}$ *satisfying*:

(i) $\forall\, i \ne j,\ a_{ij} \le 0$ *and* $\forall\, i,\ \sum_j a_{ij} \ge 0$;
(ii) $a_{i,i\pm p} < 0$ *provided that they exist*;
(iii) $\exists\, i_1,\dots,i_p \in \{1,\dots,n\}$ *s.t.* $i_k \equiv k \bmod p$ *and* $\sum_j a_{i_k j} > 0$.

Then all the eigenvalues of A have a positive real part.

PROOF. Let $x \ne 0$ be an eigenvector with $Ax = \lambda x$ and $k_0 \in \arg\max|x_k|$. Without loss of generality, we may assume that $x_{k_0} = 1$. Assume that $\mathrm{Re}(\lambda) \le 0$. Then, using (i),

$$\sum_{j\ne k_0} \underbrace{a_{k_0 j}(\mathrm{Re}(x_j)-1)}_{\ge 0} = \sum_j a_{k_0 j}\,\mathrm{Re}(x_j) - \sum_j a_{k_0 j} \le \sum_j a_{k_0 j}\,\mathrm{Re}(x_j) = \mathrm{Re}(\lambda) \le 0.$$

In turn, (ii) implies that $\mathrm{Re}(x_j) = 1$ for $j = k_0 \pm p$ (if in $\{1,\dots,n\}$). Iterating the process yields that $\mathrm{Re}(x_j) = 1$ for every $j \equiv k_0 \bmod p$. Now let $i_{k_0} \equiv k_0 \bmod p$ as in (iii). One has

$$\mathrm{Re}(\lambda) = \sum_j a_{i_{k_0} j}\,\mathrm{Re}(x_j) \ge \sum_{j\equiv k_0 \bmod p} a_{i_{k_0} j} \ge \sum_j a_{i_{k_0} j} > 0.$$

Hence $\mathrm{Re}(\lambda) > 0$. □

REMARK. The classical Gershgorin lemma yields the same conclusion whenever

$$\forall\, i \ne j, \qquad a_{ij} \le 0 \quad \text{and} \quad \forall\, i, \qquad \sum_j a_{ij} > 0.$$

In fact, Lemma 5 will also be the key of the next section devoted to the uniformly distributed case where regular a.s. convergence holds (see Theorem 3). The proof relies on the positivity of the symmetrical matrix $(\nabla h^{\sigma} + {}^t\nabla h^{\sigma})(x)$ on F_n^+.

1.2.1. *Proof of Theorem* 1 (*log-concave setting*). Following the hypothesis of Lemma 5, we mainly need to study the sign of both the components and the (sum of the) lines

$$L_i(x) := \sum_{j=1}^{n} \frac{\partial h_i^\sigma}{\partial x_j} \quad \text{of } \nabla h^\sigma(x), \qquad x \in F_n^+.$$

LEMMA 6. (a) *For all* $x \in F_n^+$, $\forall\ i \neq j$, $\partial h_i^\sigma/\partial x_j(x) \leq 0$.
(b) *If* $f > 0$ *on* $(0,1)$ *then, for every* $x \in F_n^+$:

(i)

$$n \geq 2 \text{ and } \sigma(1) < \sigma(0) \quad \Rightarrow \quad \frac{\partial h_i^\sigma}{\partial x_{i-1}}(x) \text{ and } \frac{\partial h_i^\sigma}{\partial x_{i+1}}(x) < 0;$$

(ii)

$$n \geq 3 \text{ and } \sigma(2) < \sigma(1) \quad \Rightarrow \quad \frac{\partial h_i^\sigma}{\partial x_{i-2}}(x) \text{ and } \frac{\partial h_i^\sigma}{\partial x_{i+2}}(x) < 0;$$

(iii) $$n \geq 5 \text{ and } \sigma(3) < \sigma(2) \quad \Rightarrow \quad \frac{\partial h_i^\sigma}{\partial x_{i-3}}(x) \text{ and } \frac{\partial h_i^\sigma}{\partial x_{i+3}}(x) < 0.$$

PROOF. (a) Assume, for example, that $j \leq i-1$. Then the result follows from (7) and from the obvious inequalities $\sigma(|j-(i+1)|) - \sigma(|j-i|) \leq 0$, $\tilde{x}_j < x_j < x_i$, $\sigma(|j-i|) - \sigma(|j+1-i|) \geq 0$ and $\tilde{x}_{j+1} < x_{j+1} \leq x_i$. Hence $\partial h_i^\sigma/\partial x_j(x) \leq 0$.

(b)(i) One just checks, still using (7), that

$$\frac{\partial h_i^\sigma}{\partial x_{i-1}}(x) \leq \frac{\sigma(1) - \sigma(0)}{2}(x_i - \tilde{x}_i) f(\tilde{x}_i) < 0$$

and

$$\frac{\partial h_i^\sigma}{\partial x_{i+1}}(x) \leq \frac{\sigma(0) - \sigma(1)}{2}(x_i - \tilde{x}_{i+1}) f(\tilde{x}_{i+1}) < 0,$$

since $x \in F_n^+$.

Items (ii) and (iii) follow the same way. □

As we are concerned here only with the gradient matrix at x^*, we may use the equilibrium system $h^\sigma(x^*) = 0$:

$$x_i^* = \frac{\sum_{j=1}^n \sigma(|j-i|) \int_{\tilde{x}_j^*}^{\tilde{x}_{j+1}^*} \omega f(\omega)\, d\omega}{\sum_{j=1}^n \sigma(|j-i|) \int_{\tilde{x}_j^*}^{\tilde{x}_{j+1}^*} f(\omega)\, d\omega}, \qquad 1 \leq i \leq n. \tag{8}$$

Thus, plugging (8) into expression (7) for $\nabla h^\sigma(x^*)$ finally yields

$$\forall\, 1 \leq i \leq n, \qquad L_i(x^*) := \frac{D_i(x^*)}{\sum_{j=1}^n \sigma(|j-i|) \int_{\tilde{x}_j^*}^{\tilde{x}_{j+1}^*} f(\omega)\, d\omega}, \tag{9}$$

where $D_i(x^*)$ is given (keeping in mind the conventions $\tilde{x}_1 = 0^-$, $\tilde{x}_{n+1} = 1^+$) by

$$
\begin{aligned}
D_i(x^*) &:= \left(\sum_{j=1}^{n} \sigma(|j-i|) \int_{\tilde{x}_j^*}^{\tilde{x}_{j+1}^*} f(\omega)\, d\omega \right)^2 \\
&\quad - \sum_{j=1}^{n-1} \big(\sigma(|j+1-i|) - \sigma(|j-i|) \big) \\
&\quad \times f(\tilde{x}_{j+1}^*) \sum_{k=1}^{n} \sigma(|k-i|) \int_{\tilde{x}_k^*}^{\tilde{x}_{k+1}^*} (\omega - \tilde{x}_{j+1}^*) f(\omega)\, d\omega.
\end{aligned} \tag{10}
$$

Now the problem clearly amounts to determining the sign of $D_i(x^*)$. To this end, let us introduce on $\overline{F_{n+1}^+}$ some auxiliary functions φ_i^n, defined for every i by

$$
\varphi_i^n(u) := \psi_i^n(u) - \sum_{k=1}^{n+1} \tau(k,i) f(u_k) \sum_{j=1}^{n} \sigma(|j-i|) \int_{u_j}^{u_{j+1}} (\omega - u_k) f(\omega)\, d\omega,
$$

where

$$
\tau(k,i) := \sigma(|k-i|) \mathbf{1}_{\{k \le n\}} - \sigma(|k-1-i|) \mathbf{1}_{\{k \ge 2\}},
$$

$$
\psi_i^n(u) := \left(\sum_{k=1}^{n} \sigma(|k-i|) \int_{u_k}^{u_{k+1}} f(\omega)\, d\omega \right)^2.
$$

Being log-concave, f has right derivatives so φ_i^n has right partial derivatives as well and a little algebra gives

$$
\begin{aligned}
\frac{\partial_+ \varphi_i^n}{\partial u_l}(u) &= - \psi_i^n(u) f(u_l) \tau(l,i) + \left(\sum_{k=1}^{n+1} \tau(k,i) f(u_k)(u_l - u_k) \right) f(u_l) \tau(l,i) \\
&\quad - \tau(l,i) f'_+(u_l) \left(\sum_{j=1}^{n} \sigma(|j-i|) \int_{u_j}^{u_{j+1}} (\omega - u_l) f(\omega)\, d\omega \right),
\end{aligned}
$$

where the subscript + denotes *right* derivative. A second (right) differentiation, this time with respect to a variable u_m, $m \neq l$, finally yields

$$
\begin{aligned}
\forall\, l \neq m, \quad \frac{\partial_+^2 \varphi_i^n}{\partial u_l\, \partial u_m}(u) &= \tau(l,i) \tau(m,i)(u_l - u_m) \\
&\quad \times \big(f'_+(u_m) f(u_l) - f'_+(u_l) f(u_m) \big).
\end{aligned} \tag{11}
$$

The following lemmas show that the derivatives calculated in (11) are sufficient to specify the sign of φ_i^n.

LEMMA 7. (a) *For all* $\alpha \in [0,1]$, $\varphi_i^n(\alpha, \dots, \alpha) = 0$,

$$
\forall\, \alpha \in (0,1), \forall\, l \in \{1, \dots, n+1\}, \qquad \frac{\partial_+ \varphi_i^n}{\partial u_l}(\alpha, \dots, \alpha) = 0.
$$

(b) *Assume that* $f > 0$ *on* $(0,1)$. *If* $\log f$ *is concave and* $u \in \overline{F_{n+1}^{+}} \cap (0,1)^{n+1}$, *then the symmetric matrix* $[\partial_+^2 \varphi_i^n / \partial u_l \, \partial u_m(u)]_{l, m \in \{1, \dots, n+1\}}$ *has a sign structure given by*

$$\begin{array}{cc} & \begin{array}{ccccccc} 1 & & & i+1 & & & n+1 \end{array} \\ \begin{array}{c} \\ \\ \\ i+1 \\ \\ \\ \\ \end{array} & \left[\begin{array}{ccccccc} \times & & \geq 0 & \vdots & & & \\ & \times & & \vdots & & \leq 0 & \\ \geq 0 & & \times & \vdots & & & \\ \cdots & \cdots & \cdots & \times & \cdots & \cdots & \cdots \\ & & & \vdots & \times & & \geq 0 \\ & \leq 0 & & \vdots & & \times & \\ & & & \vdots & \geq 0 & & \times \end{array}\right], \end{array}$$

the sign of the diagonal terms being unknown a priori.

Proof. Part (a) is obvious.

(b) Because $\log f$ is concave, f'_+/f is decreasing and

$$\forall\, u \in \overline{F_{n+1}^{+}} \cap (0,1)^{n+1}, \qquad (u_m - u_l)[f'_+(u_m) f(u_l) - f'_+(u_l) f(u_m)] \leq 0.$$

Simple considerations on the sign of $\tau(l, i)$ and (11) complete the proof. □

Lemma 8. *Assume that* $f > 0$ *on* $(0,1)$ *and* $\log f$ *is concave. Then*:

(a) $\forall\, u \in \overline{F_{n+1}^{+}}$, $\varphi_i^n(u) \geq 0$;

(b) *if* $\log f$ *is strictly concave and* $\sigma(1) < \sigma(0)$, *then*

$$\forall\, u \in \overline{F_{n+1}^{+}}, \qquad u_i < u_{i+1} \Rightarrow \varphi_i^n(u) > 0.$$

Proof. (a) Assume first that $1 \leq m \leq i$ and $u_m \in (0,1)$;

$$\frac{\partial_+ \varphi_i^n}{\partial u_m}(u_m, \dots, u_m) = 0$$

and, following Lemma 7, for every $1 \leq l \leq m-1$,

$$\frac{\partial_+^2 \varphi_i^n}{\partial u_l \, \partial u_m} \geq 0 \quad \text{on } \overline{F_{n+1}^{+}} \cap (0,1)^{n+1}.$$

Applying this, starting from $l = 1$ up to $l = m-1$, we derive by a simple induction that

$$\begin{aligned} & 0 < u_1 \leq u_2 \leq \cdots \leq u_{m-1} \leq u_m < 1 \\ & \Rightarrow \quad \frac{\partial_+ \varphi_i^n}{\partial u_m}(u_1, \dots, u_{m-1}, u_m, \dots, u_m) \leq 0. \end{aligned} \tag{12}$$

Assume now that $m \geq i + 1$. Then, using the inequalities

$$\frac{\partial_+^2 \varphi_i^n}{\partial u_l \, \partial u_m}(u) \geq 0 \quad \text{for } l \geq m + 1$$

and

$$\frac{\partial_+^2 \varphi_i^n}{\partial u_l \, \partial u_m}(u) \leq 0 \quad \text{for } l \leq i,$$

one derives in the same way that, for every u satisfying $0 < u_1 \leq u_2 \leq \cdots \leq u_i \leq u_m \leq u_{m+1} \leq \cdots \leq u_{n+1} < 1$,

$$\frac{\partial_+ \varphi_i^n}{\partial u_m}(u_1, \dots, u_i, u_m, \dots, u_m, u_{m+1}, \dots, u_{n+1}) \geq 0. \tag{13}$$

Combining now inequality (12) and $\varphi_i^n(u_i, \dots, u_i) = 0$ yields, by an induction on $m = 1$ up to $m = i$,

$$0 \leq u_1 \leq u_2 \leq \cdots \leq u_i \leq 1 \Rightarrow \varphi_i^n(u_1, \dots, u_{i-1}, u_i, \dots, u_i) \geq 0.$$

The next step relies on inequality (13) at some special u's satisfying $u_{m-1} = \cdots = u_{i+1} := u_i$, $m \geq i + 1$. Thus, carrying on a descending induction on m, from $m = n + 1$ down to $m = i + 1$ yields this time

$$\forall\, u \in F_{n+1}^+, \qquad \varphi_i^n(u) \geq 0.$$

(b) First notice that if $\sigma(1) > \sigma(0)$ and $0 < u_i < u_{i+1} < 1$, then, for every $i \in \{2, \dots, n-1\}$, using the *strict* concavity of $\log f$:

$$\begin{aligned}
&\frac{\partial_+^2 \varphi_i^n}{\partial u_i \, \partial u_{i+1}}(u) \\
&\quad = \underbrace{(\sigma(1) - \sigma(0))^2 (u_{i+1} - u_i)}_{>0} \underbrace{(f(u_i) f'_+(u_{i+1}) - f'_+(u_i) f(u_{i+1}))}_{<0} < 0.
\end{aligned}$$

One can easily see that the above inequality still holds if $i = 1$ or n. Finally,

$$0 < u_i < u_{i+1} < 1 \Rightarrow \frac{\partial_+^2 \varphi_i^n}{\partial u_{i+1}}(u_1, \dots, u_i, u_{i+1}, \dots, u_{n+1}) > 0$$

and, subsequently, $\varphi_i^n(u) > 0$. $\square$

PROOF OF THEOREM 2. (i) The fact that x^* belongs to F_n^+ follows from Proposition 3(b). The rest of the proof amounts to the study of three different cases.

Case 1 [$n \geq 2$, $\sigma(1) < \sigma(0)$]. Using (10) and the conventions $\tilde{x}_1 = 0^-$ and $\tilde{x}_{n+1} = 1^+$, one easily checks that

$$\begin{aligned}
D_i(x^*) = \varphi_i^n(\tilde{x}^*) &+ f(0)\sigma(i-1) \sum_{k=1}^{n} \sigma(|k - i|) \int_{\tilde{x}_k^*}^{\tilde{x}_{k+1}^*} \omega f(\omega)\, d\omega \\
&+ f(1)\sigma(n-i) \sum_{k=1}^{n} \sigma(|k - i|) \int_{\tilde{x}_k^*}^{\tilde{x}_{k+1}^*} (1 - \omega) f(\omega)\, d\omega.
\end{aligned}$$

Then Lemma 8(a) implies that $D_i(x^*) \geq 0$, $1 \leq i \leq n$, and that either $D_1(x^*) > 0$ [if $f(0) > 0$] or $D_n(x^*) > 0$ [if $f(1) > 0$]. The same conclusion holds for the $L_i(x^*)$'s due to (9). As $\partial h_i^\sigma / \partial x_{i\pm 1}$ are negative, one derives from Lemma 5 (with $p = 1$) that all the eigenvalues of $\nabla h^\sigma(x^*)$ have positive real parts.

Case 2 [$n \geq 3$, $\sigma(2) < \sigma(1) = \sigma(0)$]. The $L_i(x^*)$'s are nonnegative and, this time, both $D_1(x^*)$, $D_2(x^*)$ [resp. $D_{n-1}(x^*)$, $D_n(x^*)$] are positive if $f(0) \neq 0$ [resp. $f(1) \neq 0$]. Furthermore, $\partial h_i^\sigma / \partial x_{i\pm 2}$ are negative. So Lemma 5 works with $p = 2$.

Case 3 [$n \geq 5$ and $\sigma(3) < \sigma(2) = \sigma(1) = \sigma(0)$]. Then $D_1(x^*)$, $D_2(x^*)$, $D_3(x^*)$ [resp. $D_{n-2}(x^*)$, $D_{n-1}(x^*)$, $D_n(x^*)$] > 0 if $f(0)$ [resp. $f(1)$] > 0 and $\partial h_i^\sigma / \partial x_{i\pm 3} < 0$. Lemma 6 works with $p = 3$.

(ii) Again, three subcases are to be discussed in order to fulfill the classical Gershgorin lemma:

Case 1 [$n \geq 2$ and $\sigma(1) < \sigma(0)$]. Since $\tilde{x}_i^* < \tilde{x}_{i+1}^*$ Lemma 8(b) implies $L_i(x^*) > 0$ for every i.

Case 2 [$n \geq 3$, $\sigma(2) < \sigma(1) = \sigma(0)$]. Setting $\bar{\sigma}(i) := \sigma(i+1)$, we have $\bar{\sigma}(1) < \bar{\sigma}(0)$. Let $\bar{\varphi}_i^n$ be the φ_i^n function related to $\bar{\sigma}$. It follows from equality $\sigma(1) - \sigma(0) = 0$,

$$\varphi_1^n(u) = \bar{\varphi}_1^{n-1}(u_1, u_3, u_4, \dots, u_{n+1}),$$

$$\varphi_i^n(u) = \bar{\varphi}_{i-1}^{-2}(u_1, \dots, u_{i-1}, u_{i+2}, \dots, u_{n+1}), \qquad 2 \leq i \leq n-1,$$

$$\varphi_n^n(u) = \bar{\varphi}_{n-1}^{n-1}(u_1, \dots, u_{n-2}, u_n).$$

The inequalities $\tilde{x}_{i-1}^* < \tilde{x}_{i+2}^*$, $2 \leq i \leq n-1$, $\tilde{x}_1^* < \tilde{x}_3^*$ and $\tilde{x}_{n-2}^* < \tilde{x}_n^*$ and Lemma 8(b) imply the positivity of all $L_i(x^*)$'s.

Case 3 [$n \geq 5$, $\sigma(3) < \sigma(2) = \sigma(1) = \sigma(0)$]. Now we set $\bar{\sigma}(i) = \sigma(i+2)$ and $\bar{\varphi}_i^n$ as above. Evaluating φ_i^n, $3 \leq i \leq n-2$, φ_1^n, φ_2^n, φ_{n-1}^n, φ_{n-2}^n with the appropriate $\bar{\varphi}_j^m$ finally gives the expected result: $L_i(x^*) > 0$, $1 \leq i \leq n$.

The classical Gershgorin lemma and Lemma 6(a) complete the proof. The eigenvalues of $\nabla h^\sigma(x^*)$ have positive real part, so x^* is stable. □

2. The uniformly distributed case.

2.1. *Almost sure convergence.* Assume that assumption ($\mathcal{S}$) holds. As $f = 1_{[0,1]}$ has a concave logarithm and $f(0) = f(1) = 1$, we know from Theorem 2 that every equilibrium point x^* is conditionally stable. Actually this result can be substantially improved in this particular setting.

THEOREM 3. *If σ satisfies assumption ($\mathcal{S}$), $\mu := U([0,1])$ and $\sum_t \varepsilon_t = +\infty$, $\sum_t \varepsilon_t^2 < +\infty$, then the equilibrium point x^* is unique and unconditionally stable. That is,*

$$\forall\, x \in F_n^+,\ \mathbb{P}_x\text{-}a.s., \qquad X^t \to x^* \quad as\ t \to +\infty.$$

This result relies on the classical Robbins–Monro theorem (in the compact valued setting; see [6]): if X^t lives in a compact set K and h admits a unique zero x^* in K, a.s. convergence to x^* holds provided that $(h(x)|x - x^*) > 0$ on $K\setminus\{x^*\}$. The following lemma, whose proof is omitted, yields a simple criterion to fulfill this assumption.

LEMMA 9. *Let K be a convex compact set with a nonempty interior $\mathring{K}$, and let h: $K \to \mathbb{R}$ be a continuous function, differentiable on $\mathring{K}$ such that $\nabla h(x) +{}^t \nabla h(x)$ is everywhere positive on $\mathring{K}$. Then*:

(a) $\forall\ x \in K, \forall\ y \in \mathring{K},\ x \neq y \Rightarrow (h(x) - h(y)|x - y) > 0$;

(b) *if* $\{h = 0\} \neq \varnothing$ *and* $\{h = 0\} \subset \mathring{K}$, *then* $\{h = 0\}$ *is reduced to a singleton* $\{x^*\}$.

PROOF OF THEOREM 3. It follows from Propositions 2 and 3 that $\{h^\sigma = 0\} \subset F_n^+$ (keeping in mind that h^σ is continuously extended on $\overline{F_n^+}$). So, according to Lemma 9, it remains to prove that $\nabla h^\sigma +{}^t\nabla h^\sigma$ is positive on F_n^+. To this end, we will use again Lemma 5. Prior to any further calculation, we know (see Lemma 6) that

$$\left(\frac{\partial h_i^\sigma}{\partial x_j} + \frac{\partial h_j^\sigma}{\partial x_i}\right)(x) \leq 0 \quad \text{whenever } i \neq j.$$

Then

$$L_i^{\mathrm{sym}}(x) := \sum_{j=1}^{n} \left(\frac{\partial h_i^\sigma}{\partial x_j} + \frac{\partial h_j^\sigma}{\partial x_i}\right)(x), \qquad 1 \leq i \leq n,$$

has to be calculated in order to determine its sign. If $i \in \{2, \dots, n-1\}$, we first calculate

$$\begin{aligned}
\sum_{j\neq i}\left[\frac{\partial h_i^\sigma}{\partial x_j} + \frac{\partial h_j^\sigma}{\partial x_i}(x)\right] = {}& \sum_{j=1}^{n} \frac{\sigma(|j-1-i|) - \sigma(|j-i|)}{2}\left(x_i - \tilde{x}_j - x_j + \tilde{x}_{i+1}\right) \\
& - \frac{\sigma(i) - \sigma(i-1)}{2}(x_i - \tilde{x}_1) \\
& + \sum_{j=1}^{n} \frac{\sigma(|j-i|) - \sigma(|j+1-i|)}{2}\left(x_i - \tilde{x}_{j+1} - x_j + \tilde{x}_i\right) \\
& - \frac{\sigma(n-i) - \sigma(n+1-i)}{2}(x_i - \tilde{x}_{n+1}) \\
& - \frac{\sigma(1) - \sigma(0)}{2}(\tilde{x}_{i+1} - \tilde{x}_i) - \frac{\sigma(0) - \sigma(1)}{2}(\tilde{x}_i - \tilde{x}_{i+1}),
\end{aligned}$$

so

$$
\begin{aligned}
L_i^{\mathrm{sym}}(x) = {} & \sum_{j=2}^{n} \left(\sigma(|j-1-i|) - \sigma(|j-i|)\right)\left(x_i - 2\tilde{x}_j + \frac{\tilde{x}_i + \tilde{x}_{i+1}}{2}\right) \\
& + \left(\sigma(0) - \sigma(1)\right)\left(\tilde{x}_i - \tilde{x}_{i+1}\right) \\
& + \frac{\sigma(i) - \sigma(|i-1|)}{2}\left(\tilde{x}_{i+1} - x_1\right) \\
& + \frac{\sigma(|n-i|) - \sigma(|n+1-i|)}{2}\left(\tilde{x}_i - x_n\right) + 2\frac{\partial h_i}{\partial x_i}.
\end{aligned}
$$

Using an Abel transform along with expression (7) for $\partial h_i^\sigma / \partial x_i$, we get

$$
\begin{aligned}
L_i^{\mathrm{sym}}(x) = {} & \sigma(i-1) \underbrace{\left(x_i + \frac{\tilde{x}_i + x_1}{2}\right)}_{>0} \\
& + \sigma(n-i) \underbrace{\left(2 - x_i - \frac{\tilde{x}_{i+1} + x_n}{2}\right)}_{>0} \\
& + \frac{\sigma(i)}{2} \underbrace{\left(\tilde{x}_{i+1} - x_1\right)}_{>0} + \frac{\sigma(n+1-i)}{2} \underbrace{\left(x_n - \tilde{x}_i\right)}_{>0} \geq 0 \\
& \forall\, i \in \{2, \dots, n-1\}.
\end{aligned}
\tag{14}
$$

If $i = 1, n$, one checks that some nonpositive terms in the regular formula (7) for $\partial h_i / \partial x_i$ are canceled due to "edge" effect (i.e., $\tilde{x}_1 = 0^-$ and $\tilde{x}_{n+1} = 1^+$). So (14) still holds but as an inequality $\cdots \geq \cdots$ provided that one sets $\sigma(n) = 0$.

Hence $L_1^{\mathrm{sym}}(x) > 0$. When $[n \geq 2$ and $\sigma(1) < \sigma(0)]$ or $[n \geq 3$ and $\sigma(2) < \sigma(1)]$, one checks that

$$
\frac{\partial h_i^\sigma}{\partial x_{i-1}} + \frac{\partial h_{i-1}^\sigma}{\partial x_i} < 0 \quad \text{and} \quad \frac{\partial h_i^\sigma}{\partial x_{i+1}} + \frac{\partial h_{i+1}^\sigma}{\partial x_i} < 0.
$$

So the assumptions of Lemma 5 are fulfilled with $p = 1$, which completes the proof in that case.

If $n \geq 5$ and $\sigma(0) = \sigma(1) = \sigma(2) > \sigma(3)$, (14) implies that $L_1^{\mathrm{sym}}(x)$ and $L_2^{\mathrm{sym}}(x) > 0$. The assumptions of Lemma 5 are fulfilled with $p = 2$ since

$$
\frac{\partial h_i^\sigma}{\partial x_{i-2}} + \frac{\partial h_{i-2}^\sigma}{\partial x_i} < 0 \quad \text{and} \quad \frac{\partial h_i^\sigma}{\partial x_{i+2}} + \frac{\partial h_{i+2}^\sigma}{\partial x_i} < 0. \qquad \square
$$

2.2. *Rate of convergence.* Let $z^t := (X^t - x^*)/\sqrt{\varepsilon_t}$, $t \in \mathbb{N}$, be the normalized error of an algorithm defined by (2). We will inspect some various rates for the step ε_t, following the general results contained in Bouton [1] (or [12]).

1. If $\varepsilon_t := a/(b + t^\alpha)$, $\alpha \in (\frac{1}{2}, 1)$, $0 < a < b$, then z^t is asymptotically Gaussian. More precisely, if one defines the stepwise process

$$Z_s := z^t \quad \text{if } s \in [T_t, T_{t+1}[\text{ with } T_t := \sum_{1 \le i \le t} \varepsilon_i,$$

then the sequence of processes $(u \mapsto Z_{T_t+u},\ t \in \mathbb{N})$ converges in distribution under the U_K-topology to a *stationary* n-dimensional Ornstein–Uhlenbeck process

$$\tilde{Z}_s := \exp(-s\nabla h^\sigma(x^*))\left(\tilde{Z}_0 + \int_0^s \exp(u\nabla h^\sigma(x^*))\sqrt{A}\, dW_u\right),$$

where $A := \int_0^1 (H^{\sigma\,t}H^\sigma)(x^*, \omega)\, d\omega$. Hence,

$$z^t = Z_{T_t}(0) \to_{\mathscr{D}} \tilde{Z}_0, \qquad \tilde{Z}_0 \sim_{\mathscr{D}} \mathscr{N}(0; \Sigma^2)$$

$$\text{with } \Sigma^2 := \int_0^{+\infty} \exp(-\nabla h^\sigma(x^*)s)\, A \exp\left(-{}^t\nabla h^\sigma(x^*)s\right) ds.$$

APPLICATION TO THE KOHONEN ALGORITHM. A straightforward computation yields

$$\begin{aligned} A &:= \left[\sum_{k=1}^{n} (\tilde{x}_{k+1}^* - \tilde{x}_k^*)\sigma(|k-i|)\sigma(|k-j|)\right]_{1 \le i, j \le n} \\ &= S\, \mathrm{Diag}\left[\tilde{x}_{k+1}^* - \tilde{x}_k^*, 1 \le k \le n\right] S, \end{aligned}$$

where $S := [\sigma(|i-j|)]_{1 \le i, j \le n}$ is the "neighborhood matrix." Note that A, S and Σ^2 have the same rank since $x^* \in F_n^+$ and that A is diagonal iff σ is the 0-neighbor function. This result stresses the effect of the neighborhood structure on the asymptotics of the algorithm.

2. If $\varepsilon_t = a/(b + t)$, $0 < a < b$, let $\lambda_{\min}$ denote the lowest eigenvalue of $\mathrm{Sym}(\nabla h(x^*))$. Then, still following [1], a central limit theorem holds whenever $a > 1/2\lambda_{\min}$. That is,

$$\begin{aligned} z^t \to_{\mathscr{D}} \mathscr{N}(0; \Sigma_a^2) \quad \text{with } \Sigma_a^2 := \int_0^{+\infty} &\exp\left[\left(\frac{\mathrm{Id}_n}{2a} - \nabla h^\sigma(x^*)\right)s\right] A \\ &\times \exp\left[{}^t\left(\frac{\mathrm{Id}_n}{2a} - \nabla h^\sigma(x^*)\right)s\right] ds, \end{aligned}$$

where Id_n denotes the identity matrix.

APPLICATION. Unfortunately, only a lower bound is easily available for $\lambda_{\min}$:

$$\begin{aligned} n\lambda_{\min} &\le \mathrm{Tr}(\nabla h^\sigma(x^*)) \\ &= \sum_{1 \le i, k \le n} \sigma(|k-i|)(\tilde{x}_{k+1}^* - \tilde{x}_k^*) - \frac{\sigma(0) - \sigma(1)}{2}, \end{aligned} \tag{15}$$

which yields

$$\lambda_{\min} < \frac{\max_k \Sigma_i \sigma(|k-i|)}{n}.$$

So, any reasonable hope to observe an asymptotically Gaussian behavior requires that

$$b > a \text{``} \gg \text{''} \frac{n}{2\max_k \Sigma_{1 \le i \le n} \sigma(|k-i|)}.$$

In the $2p$-neighbor setting ($p = 0, 1, 2$) that leads to $b > a \gg n/(2(2p+1))$. When $p = 0$ (0-neighbor setting), the true value of $\lambda_{\min}$ is available (see the proof of Theorem 5, Section 3.3):

$$\lambda_{\min} := \frac{1}{n}\sin^2\left(\frac{\pi}{2n}\right) \sim \frac{\pi^2}{4n^3},$$

which shows how loose the above general bound (15) is.

3. Some multidimensional results. We are interested in the Kohonen algorithm with $[0,1]^d$-valued stimuli space, a d-dimensional *unit* set I with $|I| = n$ and a d-dimensional neighborhood structure σ defined on I. For the sake of clarity, we will denote from now on a unit $\mathbf{i} \in I$ with bold letters. We will denote by $K_d(I, \mu, \sigma)$ the d-dimensional Kohonen algorithm.

3.1. *Grid equilibrium points of a product Kohonen algorithm.* In this section we introduce a more specific d-dimensional model: the *product* Kohonen algorithm.

DEFINITION 3. Let $n := n_1 \cdots n_d$ be a decomposition of n. For every $l \in \{1, \dots, d\}$, one sets $I_l := \{1, 2, \dots, n_l\}$, σ_l a neighborhood function on I_l and μ_l a probability measure on $[0,1]$. Then $I := I_1 \times \cdots \times I_d$, $\mu = \mu_1 \otimes \cdots \otimes \mu_d$ and the neighborhood function $\sigma := \sigma_1 \otimes \cdots \otimes \sigma_d$ is written

$$\forall\,(\mathbf{i},\mathbf{j}) \in I^2, \mathbf{i} = (i_1, \dots, i_d), \mathbf{j} = (j_1, \dots, j_d), \quad \sigma(\mathbf{i}-\mathbf{j}) := \prod_{1 \le l \le d} \sigma_l(|i_l - j_l|).$$

The related "product" Kohonen algorithm will be denoted $\otimes_{l=1}^d K_1(I_l, \mu_l, \sigma_l)$.

Among all the points in D_I, some play a special role in that setting: the grids.

DEFINITION 4. We say that $x \in D_I$ is an $(n_1, \dots, n_d)$-grid if there exist some [0, 1]-valued n_l-tuples x^l, $1 \le l \le d$, such that $\forall\ \mathbf{i} \in I$, $x_{\mathbf{i}} := (x^1_{i_1}, \dots, x^d_{i_d})$. The symbol x denotes $x := x^1 \otimes \cdots \otimes x^d$.

PROPOSITION 10. *Let x^{l*} be an equilibrium point of the $K_1(I_l, \sigma_l, \mu_l)$-Kohonen algorithm, $1 \le l \le d$. Then the grid $x^* := x^{1*} \otimes \cdots \otimes x^{d*}$ is an equilibrium point of the $\otimes_{l=1}^d K_1(I_l, \sigma_l, \mu_l)$-Kohonen algorithm. These are the only grid equilibrium points provided that* $\operatorname{supp}(\mu_l) = [0,1]$, $1 \le l \le d$.

PROOF. Let $\mathbf{i} := (i_1, \ldots, i_d) \in I$. Once it is noticed that $C_{\mathbf{i}}(x^*) = \prod_{l=1}^d C_{i_l}(x^{l*})$, it is straightforward that the component $h_{\mathbf{i}}^{\sigma, l}$, $1 \le l \le d$, $\mathbf{i} \in I$, of the average function h^σ reads

$$\prod_{l' \neq l} \left(\sum_{j_{l'}=1}^{n_{l'}} \sigma_{l'}(|j_{l'} - i_{l'}|) \mu_{l'}\big(C_{j_{l'}}(x^{l'*})\big) \right) \tag{16}$$
$$\times \underbrace{\sum_{j_l=1}^{n_l} \sigma_l(|j_l - i_l|) \int_{C_{j_l}(x^{l*})} (x_{il}^{l*} - \omega^l) \mu_l(d\omega^l)}_{= h_{i_l}^{\sigma_l, \mu_l}(x^{l*}) = 0} = 0. \qquad \square$$

Some stability results for such equilibrium grids are proved in Section 3.3.

3.2. *General computation of* $\nabla h^\sigma(x)$, $x \in D_I$. All the multidimensional stability results rely on the calculation of the gradient of h^σ at equilibrium points. This calculation follows from the following slightly technical lemma. Some additional notation must be introduced. For every $\mathbf{i}, \mathbf{j} \in I$, we define

$$\vec{n}_x^{\mathbf{ij}} := \frac{x_{\mathbf{j}} - x_{\mathbf{i}}}{\|x_{\mathbf{j}} - x_{\mathbf{i}}\|}, \qquad M_x^{\mathbf{ij}} := \left\{ u \in \mathbb{R}^d / \left(u - \frac{x_{\mathbf{i}} + x_{\mathbf{j}}}{2} | x_{\mathbf{i}} - x_{\mathbf{j}} \right) = 0 \right\},$$

the median hyperplane, and let $\lambda_x^{\mathbf{ij}}(d\omega)$ be the Lebesgue measure on $M_x^{\mathbf{ij}}$.

LEMMA 11. *Let φ be a continuous $\mathbb{R}$-valued function defined on $[0,1]^d$. For every $x \in D_I$, let $\Phi_{\mathbf{i}}(x) := \int_{C_{\mathbf{i}}(x)} \varphi(\omega)\, d\omega$. Then $\Phi_{\mathbf{i}}$ is continuously differentiable on D_I and*

$$\forall\, \mathbf{i} \neq \mathbf{j}, \qquad \frac{\partial \Phi_{\mathbf{i}}}{\partial x_{\mathbf{j}}}(x) = \int_{\overline{C}_{\mathbf{i}}(x) \cap \overline{C}_{\mathbf{j}}(x)} \varphi(\omega) \left\{ \frac{1}{2} \vec{n}_x^{\mathbf{ij}} + \frac{1}{\|x_{\mathbf{j}} - x_{\mathbf{i}}\|} \times \left(\frac{x_{\mathbf{i}} + x_{\mathbf{j}}}{2} - \omega \right) \right\} \lambda_x^{\mathbf{ij}}(d\omega) \tag{17}$$

and

$$\frac{\partial \Phi_{\mathbf{i}}}{\partial x_{\mathbf{i}}}(x) = - \sum_{\mathbf{j} \in I, \mathbf{j} \neq \mathbf{i}} \frac{\partial \Phi_{\mathbf{i}}}{\partial x_{\mathbf{j}}}(x).$$

PROOF. For the sake of understanding, we set $d = 2$ throughout the proof. Only the first statement is to be proved since $\sum_{\mathbf{i} \in I} \Phi_{\mathbf{i}} = \int_{[0,1]^2} \varphi(\omega)\, d\omega$. Let us calculate $\partial \Phi_{\mathbf{i}} / \partial x_{\mathbf{j}}(x)$. First, we prove that if $\overline{C}_{\mathbf{i}}(x) \cap \overline{C}_{\mathbf{j}}(x) = \varnothing$ [i.e., $C_{\mathbf{i}}(x)$ and $C_{\mathbf{j}}(x)$ have no vertex in common], then $\partial \Phi_{\mathbf{i}} / \partial x_{\mathbf{j}}(x) = 0$. Indeed, if $x'_{\mathbf{k}} := x_{\mathbf{k}} + \delta_{\mathbf{kj}} \delta$, $\delta > 0$, then $C_{\mathbf{i}}(x) = C_{\mathbf{i}}(x')$ for small enough δ.

We assume now that $\overline{C}_{\mathbf{i}}(x) \cap \overline{C}_{\mathbf{j}}(x) \neq \varnothing$ (i.e., they have a common vertex). Let $\vec{\tau}_x^{\mathbf{ij}} := \wedge\, \vec{n}_x^{\mathbf{ij}}$. As a first step, we will calculate the partial derivatives in the directions $\vec{n}_x^{\mathbf{ij}}$ and $\vec{\tau}_x^{\mathbf{ij}}$ and show that they are continuous. Then, as $x \mapsto (\vec{n}_x^{\mathbf{ij}}, \vec{\tau}_x^{\mathbf{ij}})$ is continuously differentiable, $\Phi_{\mathbf{i}}$ will be, in turn, continuously differentiable in the direction $x_{\mathbf{j}}$. Finally, $\Phi_{\mathbf{i}}$ will be $\mathscr{C}^1$. Figures 3a and 3b display the geometrical situation.

Let $\vec{\delta} := (\delta_{\mathbf{k}})_{\mathbf{k}\in I}$ with $\delta_{\mathbf{j}} := \delta\vec{n}_x^{\mathbf{ij}}$ and $\delta_{\mathbf{k}} := 0$ for $\mathbf{k} \neq \mathbf{j}$. Then we have (see Figure 1a)

$$\int_{C_{\mathbf{i}}(x+\vec{\delta})}\varphi(\omega)\,d\omega - \int_{C_{\mathbf{i}}(x)}\varphi(\omega)\,d\omega = \frac{\delta}{2}\int_{\overline{C}_{\mathbf{i}}(x)\cap\overline{C}_{\mathbf{j}}(x)}\varphi(\omega)\,d\omega + o(\delta).$$

Thus $\Phi_{\mathbf{i}}$ has a continuous derivative in the direction $\vec{n}_x^{\mathbf{ij}}$ given by

$$\frac{\partial\Phi_{\mathbf{i}}}{\partial\vec{n}_x^{\mathbf{ij}}}(x) = \frac{1}{2}\int_{\overline{C}_{\mathbf{i}}(x)\cap\overline{C}_{\mathbf{j}}(x)}\varphi(\omega)\,d\omega.$$

The calculation of the derivative in the direction $\vec{\tau}_x^{\mathbf{ij}}$ is slightly different (see Figure 1b). Let $\vec{\delta} := (\delta_{\mathbf{k}})_{\mathbf{k}\in I}$ with $\delta_{\mathbf{j}} := \delta\vec{\tau}_x^{\mathbf{ij}}$ and $\delta_{\mathbf{k}} := 0$ if $\mathbf{k} \neq \mathbf{j}$. Then we have

$$\begin{aligned}&\Phi_{\mathbf{i}}(x+\vec{\delta}) - \Phi_{\mathbf{i}}(x)\\ &\quad= \int_{\overline{C}_{\mathbf{i}}(x)\cap\overline{C}_{\mathbf{j}}(x)}\varphi(\omega)\big(m(\vec{\delta}) - \omega|\vec{\tau}_x^{\mathbf{ij}}\big)\tan\big(\alpha(\vec{\delta})\big)\lambda_x^{\mathbf{ij}}(d\omega) + o(\delta),\end{aligned}$$

where $m(\vec{\delta})$ is the middle of $x_{\mathbf{i}}$ and $x_{\mathbf{j}} + \delta\vec{\tau}_x^{\mathbf{ij}}$ and $\alpha(\vec{\delta})$ denotes the angle between $M_x^{\mathbf{ij}}$ and $M_{x+\vec{\delta}}^{\mathbf{ij}}$. The continuity of $m(\vec{\delta})$ and $\alpha(\vec{\delta})$ as functions of δ yields

$$\begin{aligned}&\Phi_{\mathbf{i}}(x+\vec{\delta}) - \Phi_{\mathbf{i}}(x)\\ &\quad= \delta\int_{\overline{C}_{\mathbf{i}}(x)\cap\overline{C}_{\mathbf{j}}(x)}\varphi(\omega)\frac{1}{\|x_{\mathbf{i}} - x_{\mathbf{j}}\|}\left(\frac{x_{\mathbf{i}} + x_{\mathbf{j}}}{2} - \omega|\vec{\tau}_x^{\mathbf{ij}}\right)\lambda_x^{\mathbf{ij}}(d\omega) + o(\delta).\end{aligned}$$

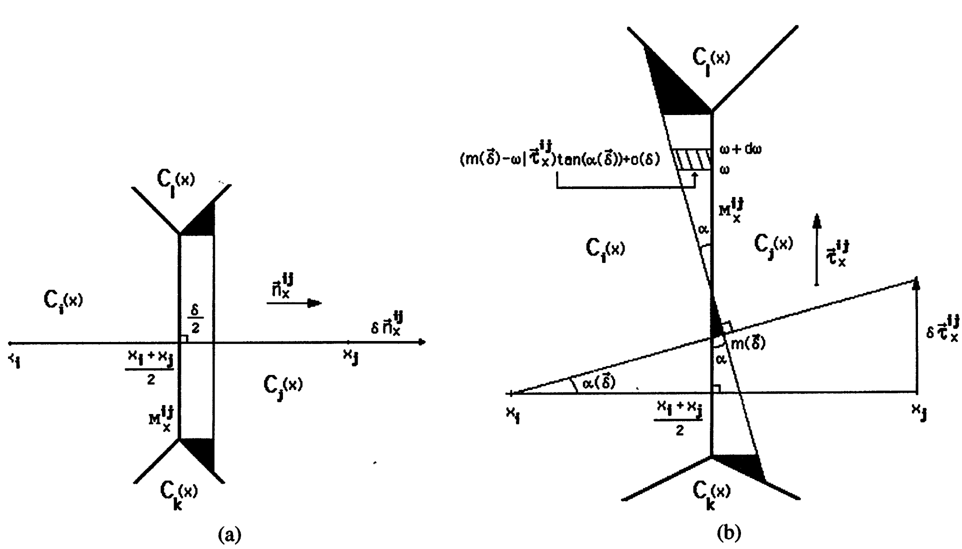


FIG. 3. *Calculation of the gradient of* Φ_i *at* x: (a) *in the direction* $\overrightarrow{x_i x_j}$; (b) *in the direction of the median hyperplane* M_{ij}.

It follows that $\Phi_{\mathbf{i}}$ has a continuous derivative in the direction $\vec{\tau}_x^{\mathbf{ij}}$, which completes the proof. □

Now, using (17) in Lemma 11 makes straightforward the calculation of ∇h^{σ} at any point $x \in D_I$.

THEOREM 4. *If $\mu(d\omega) = f(\omega)\,d\omega$, f continuous, then $h^{\sigma} \in C^1(D_I)$ and we have*

$$
\begin{aligned}
\frac{\partial h_{\mathbf{i}}^{\sigma,l}}{\partial x_{\mathbf{j}}}(x) = {} & \sum_{\mathbf{k}\in I} \sigma(\mathbf{i}-\mathbf{k})\mu(C_{\mathbf{k}}(x))\delta_{\mathbf{ij}}\vec{e}^{\,l} \\
& + \sum_{\mathbf{k}\neq\mathbf{j}} (\sigma(\mathbf{i}-\mathbf{k}) - \sigma(\mathbf{i}-\mathbf{j})) \\
& \times \int_{\overline{C}_{\mathbf{k}}(x)\cap\overline{C}_{\mathbf{j}}(x)} \left(x_{\mathbf{i}}^{l} - \omega^{l}\right)\left(\frac{1}{2}\vec{n}_x^{\mathbf{kj}} + \frac{1}{\|x_{\mathbf{k}} - x_{\mathbf{j}}\|}\left(\frac{x_{\mathbf{k}} + x_{\mathbf{j}}}{2} - \omega\right)\right) \\
& \times f(\omega)\lambda_x^{\mathbf{kj}}(d\omega), \qquad \forall\, \mathbf{i},\mathbf{j}\in I,
\end{aligned}
\tag{18}
$$

where $h_{\mathbf{i}}^{\sigma,l}$, $l = 1,\dots,d$, denotes the lth component of $h_{\mathbf{i}}^{\sigma}$, $(\vec{e}^{\,1},\dots,\vec{e}^{\,d})$ is the canonical basis of $\mathbb{R}^d$ and $\delta_{\mathbf{ij}}$ is the Kronecker symbol.

REMARK. Note that (18) is valid for any Kohonen algorithm regardless of the structure of μ or σ (product or not).

APPLICATION TO THE GRIDS OF A PRODUCT KOHONEN ALGORITHM.

PROPOSITION 12. *Let $K_1(I_1,\mu_1,\sigma_1) \otimes K_1(I_2,\mu_2,\sigma_2)$ be a two-dimensional (n_1,n_2)-product Kohonen algorithm. Assume that μ_1 and μ_2 have continuous densities f_1 and f_2. Then at any (n_1,n_2)-grid $x := x^1 \otimes x^2$, one has*

$$
\nabla h^{\sigma}(x) = \left[A_{(i_2-1)n_2+i_1,(j_2-1)n_2+j_1}\right]_{(i_1,i_2),(j_1,j_2)\in I},
$$

with

$$
A_{(i_2-1)n_1+i_1,(j_2-1)n_2+j_1}
= \begin{bmatrix}
\rho_{i_2j_2}^{(2)}\alpha_{i_1j_1}^{(1)} + \nu_{i_1j_1}^{(1)}\gamma_{i_2j_2}^{(2)} + \zeta_{i_1j_1}^{(1)}\zeta_{i_2j_2}^{(2)}\delta_{i_1j_1}\delta_{i_2j_2} & \frac{1}{2}\left(\theta_{i_1j_1}^{(1)}\beta_{i_2j_2}^{(2)} + \chi_{i_2j_2}^{(2)}\eta_{i_1j_1}^{(1)}\right) \\
\frac{1}{2}\left(\theta_{i_2j_2}^{(2)}\beta_{i_1j_1}^{(1)} + \chi_{i_1j_1}^{(1)}\eta_{i_2j_2}^{(2)}\right) & \rho_{i_1j_1}^{(1)}\alpha_{i_2j_2}^{(2)} + \nu_{i_2j_2}^{(2)}\gamma_{i_1j_1}^{(1)} + \zeta_{i_1}^{(1)}\zeta_{i_2}^{(2)}\delta_{i_1j_1}\delta_{i_2j_2}
\end{bmatrix},
$$

where $[\zeta_i^{(l)}\delta_{ij} + \alpha_{ij}^{(l)}] = \nabla h^{\mu_l,\sigma_l}$ [*see* (7)] *is the gradient of h^{μ_l,σ_l} related to $K_1(I_l,\mu_l,\sigma_l)$,*

$$
\begin{aligned}
\beta_{ij}^{(l)} := {} & (\sigma_l(|i-j+1|) - \sigma_l(|i-j|))f_l\left(\tilde{x}_j^l\right) \\
& + (\sigma_l(|i-j|) - \sigma_l(|i-j-1|))f_l\left(\tilde{x}_{j+1}^l\right),
\end{aligned}
$$

$$\gamma_{ij}^{(l)} := \frac{1}{2}\Bigg[(\sigma_l(|i-j+1|) - \sigma_l(|i-j|))\frac{f_l(\tilde{x}_j^l)}{x_j^l - \tilde{x}_j^l}$$

$$-(\sigma_l(|i-j|) - \sigma_l(|i-j-1|))\frac{f_l(\tilde{x}_{j+1}^l)}{x_{j+1}^l - \tilde{x}_{j+1}^l}\Bigg],$$

$$\eta_{ij}^{(l)} := (\sigma_l(|i-j+1|) - \sigma_l(|i-j|))f_l(\tilde{x}_j^l)\frac{x_i^l - \tilde{x}_j^l}{x_j^l - \tilde{x}_j^l}$$

$$-(\sigma_l(|i-j|) - \sigma_l(|i-j-1|))f_l(\tilde{x}_{j+1}^l)\frac{x_i^l - \tilde{x}_{j+1}^l}{x_{j+1}^l - \tilde{x}_{j+1}^l},$$

$$\rho_{ij}^{(l)} := \sigma_l(|i-j|)\mu_l(C_j(x^l)),$$

$$\theta_{ij}^{(l)} := \sigma_l(|i-j|)\int_{C_j(x^l)}(x_i^l - \omega^l)\mu_l(d\omega_l),$$

$$\nu_{ij}^{(l)} = \sigma_l(|i-j|)\int_{C_j(x^l)}(x_i^l - \omega^l)(x_j^l - \omega^l)\mu_l(d\omega^l),$$

$$\chi_{ij}^{(l)} := \sigma_l(|i-j|)\int_{C_j(x^l)}(x_j^l - \omega^l)\mu_l(d\omega^l).$$

PROOF. The proof consists of translating Lemma 11 when x is a grid $x = x^1 \otimes x^2$, $\sigma(\mathbf{i}-\mathbf{j}) = \sigma_1(|i_1 - j_1|)\sigma_2(|i_2 - j_2|)$ and $f(\omega) = f_1(\omega^1)f_2(\omega^2)$. The details are relegated to the Appendix. □

3.3. *About the (in)stability of equilibrium grids.* Due to these rather lengthy and somewhat tedious calculations, we are now in a position to derive some stability and instability properties for the grid equilibrium points of the Kohonen algorithm with multidimensional stimuli.

THEOREM 5. *Let $K_2(I_1 \times I_2, \mu_1 \otimes \mu_2, \sigma_1 \otimes \sigma_2)$ be a two-dimensional product Kohonen algorithm, μ_1 and μ_2 having continuous densities f_1 and f_2 on* $[0,1]$.

(a) *A general case of instability. Assume that $f_l > 0$ and set $a_l := \min_{[0,1]} f_l > 0$. If there is some nondecreasing function $\psi\colon \mathbb{N} \to \mathbb{R}_+^*$, $\psi(k)\uparrow +\infty$ satisfying*

$$\sum_{j=0}^{k} \sigma_l(j) = o(\psi(k)), \qquad l = 1,2, \tag{19}$$

and

(i) $\sigma_1(1) < 1$ *and* $n_2 := O\left(\dfrac{n_1}{\psi(n_1)}\right)$ *as* $n_1 \to +\infty$

or

$$\text{(ii)} \qquad \sigma_2(1) < 1 \quad \textit{and} \quad n_1 := O\left(\frac{n_2}{\psi(n_2)}\right) \quad \textit{as } n_2 \to +\infty,$$

then for large enough n_1 [*if* (i)] *or* n_2 [*if* (ii)], *every* (n_1, n_2)*-grid equilibrium is unstable.*

(b) *If* $\sigma(k) = \mathbf{1}_{\{k=0\}}$ (*0-neighbor algorithm*) *and* $\mu_1 = \mu_2 = U([0,1])$, *then*

$$x^* = \left(\left(\frac{2_{i_1} - 1}{2n_1}, \frac{2_{i_2} - 1}{2n_2}\right)\right)_{(i_1, i_2) \in I}$$

is the unique (n_1, n_2)*-grid equilibrium and*

$$(20) \qquad x^* \textit{ is stable iff } (n_1, n_2) \in \{(1,1), (1,2), (2,1), (2,2)\}.$$

Furthermore, the $(3,3)$*-grid has nonnegative eigenvalues, one being* 0.

(c) *A stability case. If* $n_2 = 1$, $K_2(I, \mu, \sigma)$ *is a Kohonen string. Assume that* x^{1*} *is a stable equilibrium for* $K_1(I_1, \mu_1, \sigma_1)$. *Let* $x^{2*} = \int_0^1 \omega \mu_2(d\omega) \in [0,1]$. *Then there is some* $\eta > 0$ *such that* $\operatorname{var}(\mu_2) < \eta$ *implies* $x^* = x^{1*} \otimes x^{2*}$ *is a stable equilibrium of* $K_2(I, \mu, \sigma)$.

APPLICATIONS AND REMARKS. The assumption [see (19)] in statement (a) is fulfilled by the usual "long-range" neighborhood functions proposed in the neural literature ([15]) as $\sigma(k) := \exp(-k/T)$ or $\sigma(k) := \exp(-k^2/2T)$, $T > 0$.

A more accurate bound than (19), for example, for numerical purposes, is available if considering directly condition (21) in the proof below.

If $\sigma_1(1) = \sigma_2(1) = 1$, the whole proof of case (a) falls apart. So no result can be derived on the stability of grids for the standard algorithm with eight, or a fortiori $3^p - 1$, $p \geq 1$, neighbors (i.e., p square "layers" of neighbors) by this method.

PROOF OF THEOREM 5. (a) To prove this instability property, we will first calculate $\operatorname{Tr}(\nabla h^\sigma(x^*))$ (Tr for trace). When it is negative, the gradient has an eigenvalue with a negative real part. Set $l' := 2$ if $l = 1$, $l' = 1$ if $l = 2$:

$$\begin{aligned}
\operatorname{Tr}(\nabla h^\sigma(x^*)) &= \sum_{l=1}^{2} \sum_{i_l, i_{l'}} \left(\rho_{i_{l'} i_{l'}}^{(l')} \alpha_{i_l i_l}^{(l)} + \nu_{i_l i_l}^{(l)} \gamma_{i_{l'} i_{l'}}^{(l')} + \zeta_{i_l i_l}^{(l)} \zeta_{i_{l'} i_{l'}}^{(l')} \right) \\
&= \sum_{l=1}^{2} \left(\operatorname{Tr}(\rho^{(l')}) \operatorname{Tr}(\alpha^{(l)}) + \operatorname{Tr}(\nu^{(l)}) \operatorname{Tr}(\gamma^{(l')}) + \operatorname{Tr}(\zeta^{(l)}) \operatorname{Tr}(\zeta^{(l')}) \right).
\end{aligned}$$

The expressions for these traces are

$$\operatorname{Tr}(\rho^{(l)}) = \sum_{i=1}^{n_l} \sigma_l(0) \mu_l(C_i(x^{l*})) = 1,$$

$$\operatorname{Tr}(\nu^{(l)}) = \sum_{i=1}^{n_l} \sigma_l(0) \int_{C_i(x^{l*})} (x_i^{l*} - \omega^l)^2 \mu_l(d\omega^l),$$

$$
\begin{aligned}
\operatorname{Tr}(\alpha^{(l)}) &= \frac{1}{2}\sum_{i=1}^{n_l} (\sigma_l(1) - \sigma_l(0))(x_i^{l*} - \tilde{x}_i^{l*}) f_l(x_i^{l*}) \\
&\quad + (\sigma_l(0) - \sigma_l(1))(x_i^{l*} - \tilde{x}_{i+1}^{l*}) f_l(\tilde{x}_{i+1}^{l*}) \\
&= -(1 - \sigma_l(1)) \sum_{i=1}^{n_l} (x_i^{l*} - \tilde{x}_i^{l*}) f_l(\tilde{x}_i^{l*})
\end{aligned}
$$

$$
\left[\text{recall that } f_l(\tilde{x}_1) = f_l(\tilde{x}_{n+1}) = 0\right],
$$

$$
\operatorname{Tr}(\zeta^{(l)}) = \sum_{k=1}^{n_l} \mu_l(C_k(x^{l*})) \sum_{i=1}^{n_l} \sigma_l(|i-k|),
$$

$$
\begin{aligned}
\operatorname{Tr}(\gamma^{(l)}) &= \frac{1}{2}\sum_{i=1}^{n_l} (\sigma_l(1) - \sigma_l(0)) \frac{f_l(\tilde{x}^{l*})}{x_i^{l*} - \tilde{x}_i^{l*}} - (\sigma_l(0) - \sigma_l(1)) \frac{f_l(\tilde{x}_{i+1}^{l*})}{x_{i+1}^{l*} - \tilde{x}_{i+1}^{l*}} \\
&= -(1 - \sigma_l(1)) \sum_{i=1}^{n_l} \frac{f_l(\tilde{x}_i^{l*})}{x_i^{l*} - \tilde{x}_i^{l*}}.
\end{aligned}
$$

These yield

$$
\begin{aligned}
\operatorname{Tr}(\nabla h^{\sigma}(x^*)) = \sum_{l=1}^{2} \Bigg[&\left\{ \sum_{k=1}^{n_l} \mu_l(C_k^l(x^{l*})) \sum_{i=1}^{n_l} \sigma_l(|i-k|) \right\} \\
&\times \left\{ \sum_{k=1}^{n_{l'}} \mu_{l'}(C_k^{l'}(x^{l'*})) \sum_{i=1}^{n_{l'}} \sigma_{l'}(|i-k|) \right\} \\
&-(1 - \sigma_l(1)) \Bigg(E_{l'}(x^{l'*}) \sum_{i=1}^{n_l} \frac{f_l(\tilde{x}_i^{l*})}{x_i^{l*} - \tilde{x}_i^{l*}} \\
&\qquad + \sum_{i=1}^{n_l} f_l(\tilde{x}_i^{l*})(x_i^{l*} - \tilde{x}_i^{l*}) \Bigg) \Bigg],
\end{aligned}
$$

where we set

$$
E_l(x) := \sum_{i=1}^{n_l} \int_{C_i(x)} (x_i - \omega^l)^2 \mu_l(d\omega^l).
$$

We will now find a lower bound for

$$
\Lambda := \sum_{i=1}^{n_l} f_l(\tilde{x}_i^{l*}) \left(x_i^{l*} - \tilde{x}_i^{l*} + \frac{E_{l'}(x^{l'*})}{x_i^{l*} - \tilde{x}_i^{l*}} \right).
$$

Clearly,

$$
\Lambda \ge a_l \sum_{i=1}^{n_l} x_i^{l*} - \tilde{x}_i^{l*} + \frac{E_{l'}(x^{l'*})}{x_i^{l*} - \tilde{x}_i^{l*}} \quad \text{where } a_l := \inf_{u \in (0,1)} f_l(u).
$$

Now

$$\begin{aligned} E_l(x) &\geq a_l \sum_{i=1}^{n_l} \int_{\tilde{x}_i}^{\tilde{x}_{i+1}} (x_i - \omega^l)^2 \, d\omega^l = \frac{a_l}{3} \sum_{i=1}^{n_l} \left[(x_i - \tilde{x}_i)^3 - (x_i - \tilde{x}_{i+1})^3 \right] \\ &\geq \frac{a_l}{3} \sum_{i=1}^{n_l} (\tilde{x}_{i+1} - \tilde{x}_i)\big((x_i - \tilde{x}_i)^2 + (x_i - \tilde{x}_{i+1})^2 + (x_i - \tilde{x}_i)(x_i - \tilde{x}_{i+1}) \big) \\ &\geq \frac{a_l}{6} \sum_{i=1}^{n_l} (\tilde{x}_{i+1} - \tilde{x}_i)\left[(x_i - \tilde{x}_i)^2 + (x_i - \tilde{x}_{i+1})^2 \right] \\ &\geq \frac{a_l}{6} \frac{1}{2n_l^2} = \frac{a_l}{12 n_l^2}, \end{aligned}$$

since we have $0 \leq \tilde{x}_{i+1} - \tilde{x}_i = \tilde{x}_{i+1} - x_i + x_i - \tilde{x}_i$ and $\sum_{i=1}^{n_l} \tilde{x}_{i+1} - \tilde{x}_i = 1$. Hence

$$\begin{aligned} \Lambda &\geq a_l \left(\frac{1}{2} + \frac{a_{l'}}{12 n_{l'}^2} \sum_{i=1}^{n_l} \frac{1}{x_i^{l*} - \tilde{x}_i^{l*}} \right) \\ &= a_l \left(\frac{1}{2} + \frac{a_{l'}}{24 n_{l'}^2} \sum_{i=1}^{n_l} \frac{1}{\tilde{x}_{i+1}^{l*} - \tilde{x}_i^{l*}} \right) \\ &\geq a_l \left(\frac{1}{2} + \frac{a_{l'}}{24} \left(\frac{n_l}{n_{l'}} \right)^2 \right), \end{aligned}$$

using once again that $\sum_{i=1}^{n_l} (\tilde{x}_{i+1}^{l*} - \tilde{x}_i^{l*}) = 1$ and $\tilde{x}_{i+1}^{l*} - \tilde{x}_i^{l*} > 0$. Finally, one gets

$$\begin{aligned} \operatorname{Tr}(\nabla h^\sigma(x^*)) &\leq 2 \max_{k,l} \left(\sum_{i=1}^{n_l} \sigma_l(|k-i|) \right)^2 \\ &\quad - \sum_{l=1,2} (1 - \sigma_l(1)) a_l \left(\frac{1}{2} + \frac{a_{l'}}{24} \left(\frac{n_l}{n_{l'}} \right)^2 \right) \qquad (21) \\ &\leq 8 \max_{l=1,2} \left(\sum_{j=0}^{n_l - 1} \sigma_l(j) \right)^2 - \frac{a_l a_{l'}}{24} \sum_{l=1,2} (1 - \sigma_l(1)) \left(\frac{n_l}{n_{l'}} \right)^2. \end{aligned}$$

This last inequality straightforwardly yields the expected result.

(b) In this particular setting all the eigenvalues can be calculated. First, for notational convenience, we introduce the tensor product $A \otimes B$ of an $n_1 \times n_1$ matrix A by an $n_2 \times n_2$ matrix B as an $n_1 n_2 \times n_1 n_2$ matrix whose coefficients are defined by $(A \otimes B)_{(i_2 - 1)n_1 + i_1, (j_2 - 1)n_2 + j_1} := A_{i_1 j_1} B_{i_2 j_2}$. Some straightforward algebra shows that if λ and μ are, respectively, eigenvalues of A and B, then $\lambda\mu$ is an eigenvalue of $A \otimes B$.

Then notice that $\chi^{(l)} = \theta^{(l)} = 0$, so that $\nabla h^\sigma(x^*)$ is a block diagonal matrix with two blocks: $\nabla h^{\mu_1, \sigma_1} \otimes \rho^{(2)} + \nu^{(1)} \otimes \gamma^{(2)}$ and $\rho^{(1)} \otimes \nabla h^{\mu_2, \sigma_2} + \gamma^{(1)} \otimes \nu^{(2)}$. Furthermore, $\rho^{(l)} = \mathrm{Id}_{n_l}/n_l$ and $\nu^{(l)} = \mathrm{Id}_{n_l}/12 n_l^3$. Thus, if $\lambda^{(l)}$ and $\mu^{(l)}$ are eigenvalues of $\nabla h^{\mu_l, \sigma_l}$ and $\gamma^{(l)}$, then $(1/n_l)\lambda^{(l')} + (1/12n_{l'}^3)\mu^{(l)}$ is an eigenvalue of $\nabla h^{\sigma}(x^*)$, $l = 1, 2$. We want to calculate

$$\min_{l \neq l', 1 \leq l, l' \leq 2} \left(\frac{\lambda_{\min}^{(l)}}{n_{l'}} + \frac{\mu_{\min}^{(l')}}{12 n_l^3} \right).$$

We have

$$\gamma^{(l)} = -n_l \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ -1 & 2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 2 & -1 \\ 0 & \cdots & 0 & -1 & 1 \end{bmatrix}$$

and

$$\nabla h^{\mu_l, \sigma_l} = \frac{1}{4n_l} \begin{bmatrix} 3 & -1 & 0 & \cdots & 0 \\ -1 & 2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 2 & -1 \\ 0 & \cdots & 0 & -1 & 3 \end{bmatrix}.$$

A little algebra yields the eigenvalues of these matrices:

$$\mu_k^{(l)} := -4n_l \sin^2\left(\frac{k\pi}{2n_l}\right), \qquad 0 \leq k \leq n_l - 1,$$

and

$$\lambda_k^{(l)} = \frac{1}{n_l} \sin^2\left(\frac{k\pi}{2n_l}\right), \qquad 1 \leq k \leq n_l.$$

Therefore

$$\frac{\lambda_{\min}^{(l)}}{n_{l'}} + \frac{\mu_{\min}^{(l')}}{12n_l^3} = \frac{1}{n_{l'}^2} \sin^2\left(\frac{\pi}{2n_l}\right) - \frac{1}{3n_l} \cos^2\left(\frac{\pi}{2n_{l'}}\right).$$

So the lowest eigenvalue of $\nabla h(x^*)$ is positive iff

$$\text{(22)} \qquad \frac{\sin(\pi/(2n_1))}{\cos(\pi/(2n_2))} > \frac{1}{\sqrt{3}} \frac{n_2}{n_1} \quad \text{and} \quad \frac{\sin(\pi/(2n_2))}{\cos(\pi/(2n_1))} > \frac{1}{\sqrt{3}} \frac{n_1}{n_2}.$$

Indeed, as $\sin(x) \leq x$, (22) implies first that

$$\frac{2n_l}{\pi} \cos\left(\frac{\pi}{2n_l}\right) < \sqrt{3},$$

which, in turn, implies, using that $\cos(x) \geq 1 - x^2/2$,

$$n_l < \frac{\pi}{2(\sqrt{5} - \sqrt{3})} \approx 3.116.$$

Straightforward checking yields the claimed result.

Statement (c) is a straightforward consequence of the computation (see the proof of Theorem 6)

$$\nabla h^{\mu,\sigma}(x^*) = \begin{bmatrix} \nabla h^{\mu_1,\sigma_1} & 0 \\ 0 & \zeta^{(1)} + \mathrm{var}(\mu_2)\gamma^{(1)} \end{bmatrix} \qquad (\text{a } 2n_1n_2 \times 2n_1n_2 \text{ matrix})$$

and the fact that $\zeta^{(1)} := \mathrm{Diag}(\zeta_1^{(1)}, \dots, \zeta_{n_1}^{(1)})$ is diagonal and positive. □

A NUMERICAL APPLICATION OF (c). If $\mu_1 := U([0,1])$, $n_1 = 30$, $\sigma_1(0) := 1$, $\sigma_1(1) := 0.6$, $\sigma_1(2) := 0.3$ and $\sigma_1(k) := 0$ otherwise and if μ_2 is any distribution on $[0,1]$ with expectation $\frac{1}{2}$, then $x^{1*} \otimes \frac{1}{2}$ is stable whenever $\mathrm{std}(\mu_2) < \mathrm{std}_c(\mu_2) = 0.037$.

3.4. *A general result about dimensional stability.* The aim of this section is to show that a stable equilibrium remains stable when *plunged* into a higher-dimensional space, provided that the perpendicular component of the stimuli distribution has a small enough variance.

THEOREM 6. *Assume that x^{1*} is a stable equilibrium point of a d_1-dimensional Kohonen algorithm $K_{d_1}(I_1, \mu_1, \sigma_1)$ and let*

$$x^{2*} := \int_{[0,1]^{d_2}} \omega^{(d_2)} \mu_2(d\omega^{(d_2)}) \in [0,1]^{d_2},$$

where μ_1 and μ_2 have continuous densities. Then there exists $\eta > 0$, such that $\|\Sigma^2_{\mu_2}\| < \eta$ implies that $x^* = x^{1*} \otimes x^{2*}$ is a stable equilibrium of $K_{d_1+d_2}(I, \mu, \sigma)$, where, as usual, $\mu := \mu_1 \otimes \mu_2$, $\sigma := \sigma_1 \otimes \sigma_2 \equiv \sigma_1$ and $\Sigma^2_{\mu_2}$ denotes the variance matrix of μ_2.

PROOF. Our first task is to calculate, for every $\mathbf{i}, \mathbf{j} \in I$, the $(d_1 + d_2) \times (d_1 + d_2)$ matrix

$$\frac{\partial h_{\mathbf{j}}^{\mu,\sigma}}{\partial x_{\mathbf{j}}}(x^*) = \left[\left(\frac{\partial h_{\mathbf{j}}^{\mu,\sigma,l}}{\partial x_{\mathbf{j}}} \right)^{l'} (x^*) \right]_{1 \leq l, l' \leq d_1 + d_2}$$

in the above setting. The result relies on formula (18) in Theorem 4. Notice first that $I := I_1 \times I_2 = I_1 \times \{1\} \equiv I_1$ and that $C_{\mathbf{i}}(x^{1*} \otimes x^{2*}) = C_{i_1}^{(d_1)}(x^{1*}) \times$

$C_{\mathbf{i}}^{(d_2)}(x^{2*}) = C_{i_1}^{(d_1)}(x^{1*}) \times [0,1]^{d_2}$. Then, setting $\omega := (\omega^{(d_1)}, \omega^{(d_2)})$, one gets

$$\begin{aligned}\left(\frac{\partial h_{\mathbf{i}}^{\mu,\sigma,l}}{\partial x_{\mathbf{j}}}\right)^{l'}(x^*) = {} & \sum_{k_1 \in I_1} \sigma_1(|i_1 - k_1|)\mu_1\big(C_{k_1}^{(d_1)}(x^{1*})\big)\delta_{i_1 j_1}\delta_{ll'} \\ & + \sum_{k_1 \neq j_1} (\sigma_1(|i_1 - k_1|) - \sigma_1(|i_1 - j_1|)) \\ & \times \int_{\overline{C}_{k_1}^{(d_1)}(x^{1*}) \cap \overline{C}_{j_1}^{(d_1)}(x^{1*})} \lambda_{x^{1*}}^{\mathbf{kj}}(d\omega^{(d_1)}) f_1(\omega^{(d_1)}) \\ & \times \int_{[0,1]^{d_2}} d\omega^{(d_2)} f_2(d\omega^{(d_2)})\big((x_{\mathbf{i}}^*)^l - \omega^l\big) \\ & \times \left(\frac{1}{2}\big(\vec{n}_{x^*}^{\mathbf{kj}}\big)^{l'} + \frac{1}{\|x_{\mathbf{k}}^* - x_{\mathbf{j}}^*\|}\left(\frac{(x_{\mathbf{k}}^*)^{l'} + (x_{\mathbf{j}}^*)^{l'}}{2} - \omega^{l'}\right)\right).\end{aligned}$$

The product form of x^* yields that $(\vec{n}_{x^*}^{\mathbf{kj}})^l = 0$ and

$$\frac{(x_{\mathbf{k}}^*)^l + (x_{\mathbf{j}}^*)^l}{2} = (x_{\mathbf{i}}^*)^l = \int_{[0,1]^{d_2}} \omega^l \mu_2(d\omega^{(2)}) = (x^{2*})^l$$

for every $l \in \{d_1 + 1, \dots, d_1 + d_2\}$ since the $x_{\mathbf{k}}^*$'s live in $[0,1]^{d_1} \times \{x^{2*}\}$.

Then a straightforward application of Fubini's theorem leads to

$$\frac{\partial h_{\mathbf{i}}^{\mu,\sigma}}{\partial x_{\mathbf{j}}}(x^*) = \begin{bmatrix} \dfrac{\partial h_{i_1}^{\mu_1,\sigma_1}}{\partial x_{j_1}} & 0 \\ 0 & \zeta_{i_1}^{(\mu_1)}\delta_{i_1 j_1}\,\mathrm{Id}_{d_2} + \gamma_{i_1 j_1}^{(\mu_1)}\Sigma_{\mu_2} \end{bmatrix},$$

where $\mathbf{i} = (i_1, 1)$, $\mathbf{j} = (j_1, 1)$, $\zeta_i^{(\mu_1)}$ is defined as in Proposition 4 (so it is positive) and $\gamma_{ij}^{(\mu_1)}$ is given by

$$\gamma_{ij}^{(\mu_1)} := \sum_{k \neq j} (\sigma_1(|i-k|) - \sigma_1(|i-j|)) \frac{\mu_1\big(\overline{C}_k^{(d_1)}(x^{1*}) \cap \overline{C}_j^{(d_1)}(x^{1*})\big)}{\|x_k^{1*} - x_j^{1*}\|}.$$

When $\|\Sigma_{\mu_2}^2\| \to 0$, the eigenvalues of the limiting matrix of $\nabla h^{\mu,\sigma}(x^*)$ obviously are $\zeta_i^{(\mu_1)}$, $1 \le i \le d_1$, and the eigenvalues of $\nabla h^{\mu_1,\sigma_1}(x^{1*})$. □

REMARKS. Case (c) of Theorem 5 is contained in the above result. The result still holds when μ_2 is no longer absolutely continuous.

INTERPRETATION. This result points out the ability of the Kohonen algorithm to detect the significant dimension of some possibly noisy inputs. It constitutes a generalization of a result established by physical methods in the d_1 = two-dimensional and d_2 = one-dimensional setting for uniformly distributed stimuli (see [16]).

4. Some simulation results.

4.1. *The results.* As a first conclusion to this paper, we display some simulations related to the multidimensional theoretical results contained in Theorem 5.

Figure 4 deals with case (a) of Theorem 5. It shows that whenever $n_1 = n_2$, $\mu = U([0,1]^2)$ and $\sigma_l(k)$ is "far enough" from $\mathbf{1}_{\{k=0\}}$ (zero-neighbor setting), the square equilibrium grid made up of the one-dimensional equilibrium is definitely numerically stable. On the contrary, as claimed in the theorem, when $|n_1 - n_2|$ becomes large enough the grid becomes more and more unstable. The simulations were processed with $\sigma_l(0) = 1$, $\sigma_l(1) = 0.6$, $\sigma_l(2) := 0.3$, $\sigma_l(k) := 0$, $k \geq 3$. The tested grids are $(n_1; n_2) = (20;20), (10;20), (3;20)$ and the number of trials is $N = 5 \times 10^4$.

Whenever the distribution μ_1 is not symmetric, even the square grids are no longer stable. For instance, Figure 5 shows a simulation processed with the product of two independent exponentials with $\mu_1 = \mu_2 :=$ Exp(3), truncated on [0, 1]. We set $n_1 = n_2 = 10$, $\sigma_l(1) := 0.75$, $\sigma_l(2) := 0.3$, $\sigma_l(3) := 0.066$, $\sigma_l(4) := 0.01$, $l = 1,2$ and $\varepsilon_t := 0.05/(1 + t/50{,}000)$. The scalar equilibrium points were numerically computed using a one million trial simulation. Obviously, the square grid is unstable.

Some simulations related to the numerical application of claim (c) in Theorem 5 are displayed in Figure 6. The simulations were processed setting μ_2 to a (truncated) Gaussian distribution with standard deviations std := 0.038, std := 0.037, std := 0.036 on the y axis; n_1 was set at 30. Three values for the number of trials N were selected: $N = 125 \times 10^3, 3 \times 10^6, 5 \times 10^6$. It confirms that the critical value for stability of the straight line is $\text{std}_c := 0.037$.

Figure 7 illustrates the results of Theorem 6 when $d_1 = 2$, $d_2 = 1$, I_1 is a 20×20 square grid with the same neighborhood function as in Figure 4, $\mu_1 := U([0,1]^2)$ and μ_2 is a truncated Gaussian distribution with successive std 0.07 and 0.06. The critical value obviously lies between these bounds.

4.2. *A few provisional conclusions.* In the one-dimensional setting, it seems that any further progress about the Kohonen algorithm depends on some new theoretical results on the (unconditional) a.s. convergence of the stochastic algorithms. One may reasonably imagine that they will require a deeper study of the ordinary differential equation related to the system.

As far as the neighborhood function σ is concerned, assumption $(\mathscr{S})$ may perhaps be relaxed in some way: some less stringent constraints on the monotonicity of σ could still work.

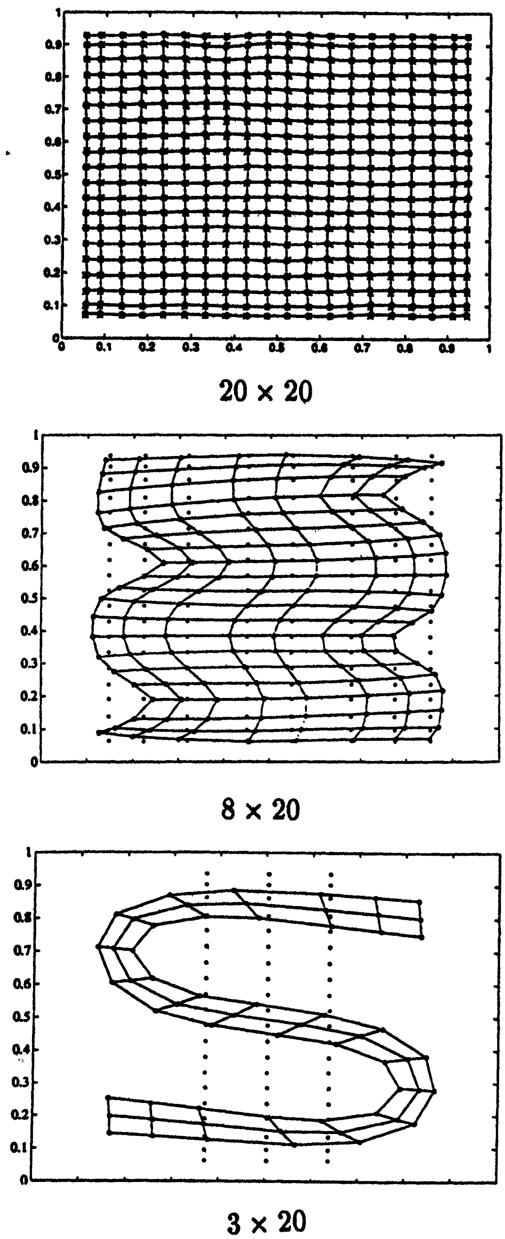
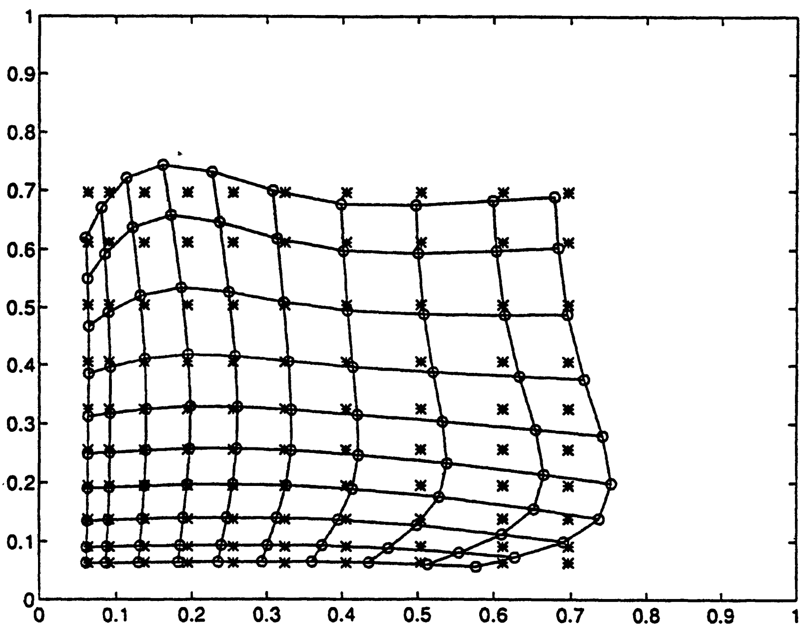
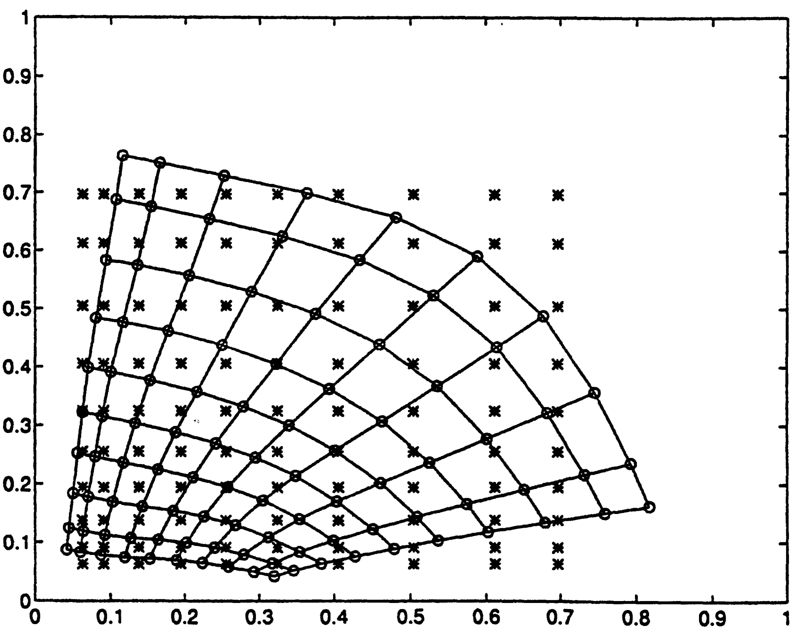


FIG. 4.



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

FIG. 5.

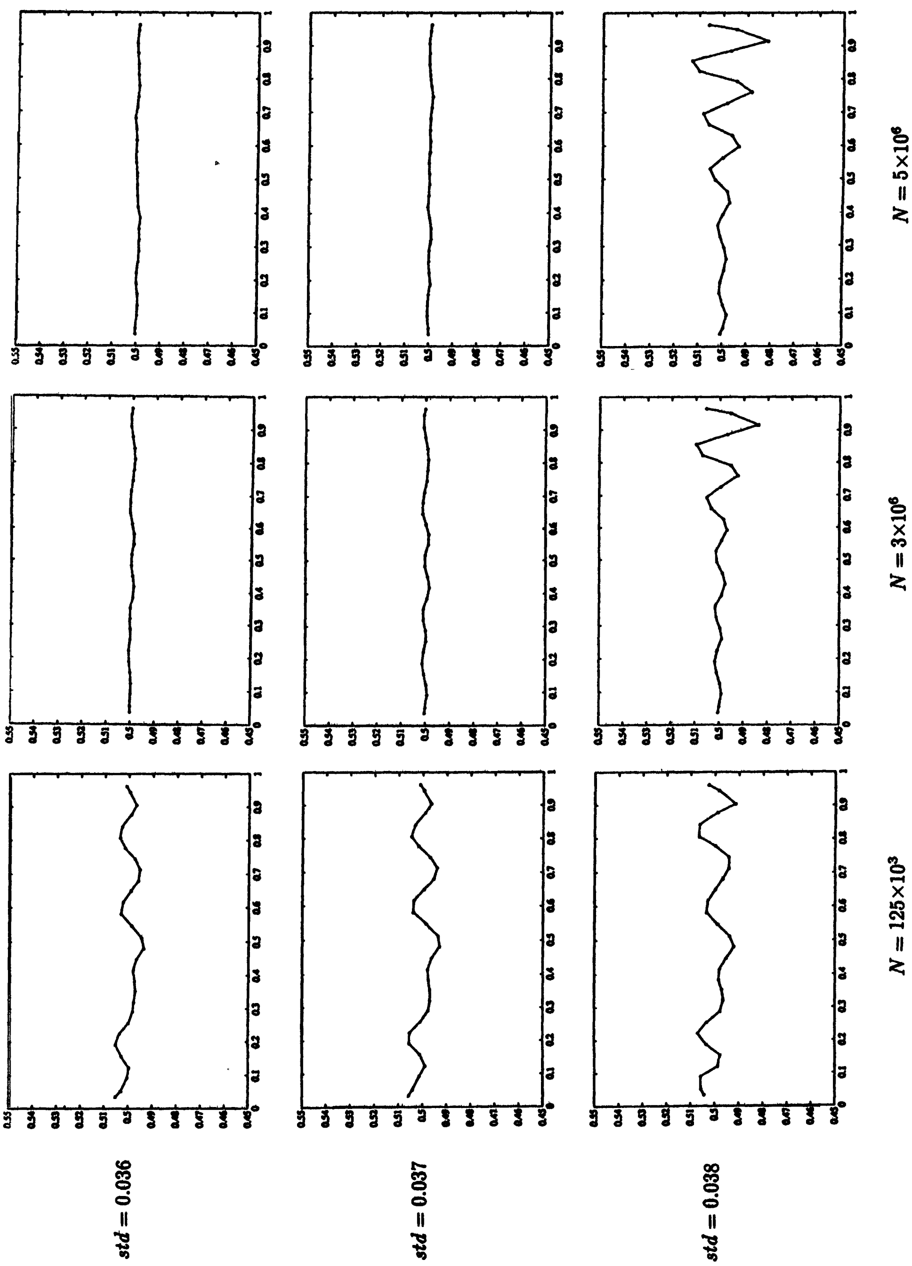
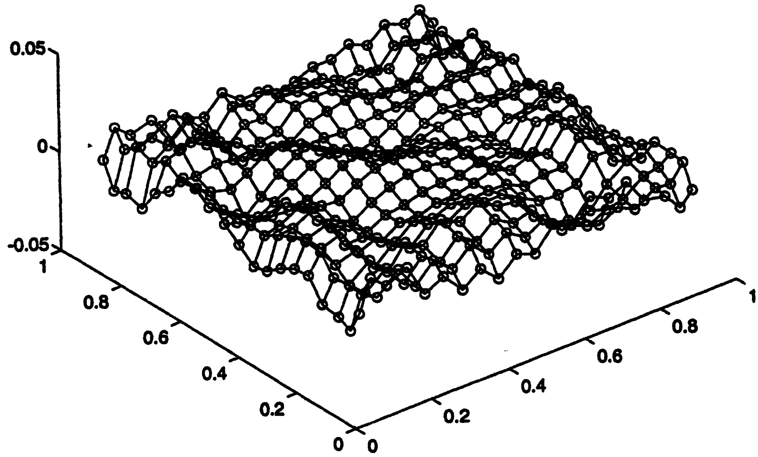
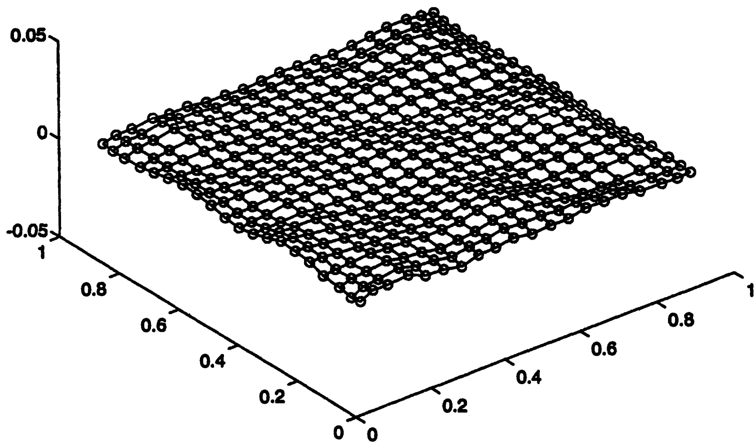

std = 0.036
std = 0.037
std = 0.038
$N = 125\times10^3$
$N = 3\times10^6$
$N = 5\times10^6$

FIG. 6.



$std = 0.07 \qquad N = 4.10^6$



$std = 0.06 \qquad N = 4.10^6$

FIG. 7.

In the multidimensional setting, there is no doubt that further results could be derived from the general formula established for ∇h^{σ}. The most intuitive—but in no case the easiest—one could be to establish the stability of the square grids when the stimuli distribution $\mu = U([0,1]^2)$ and $\sigma_l(k_l) = \mathbf{1}_{\{|k_l| \leq 1\}}$ (two-dimensional Kohonen algorithm with eight neighbors). On a more theoretical level, it would be valuable to prove the existence of at least one stable equilibrium point for the multidimensional generalized Kohonen algorithm, under some general assumptions on σ and μ.

APPENDIX

PROOF OF PROPOSITION 12. First of all notice that $\lambda_x^{\mathbf{kj}}(\overline{C}_{\mathbf{k}}(x) \cap \overline{C}_{\mathbf{j}}(x)) \neq 0$ iff

$$\mathbf{k} \in \{(j_1, j_2); (j_1 - 1, j_2); (j_1 + 1, j_2); (j_1, j_2 - 1); (j_1, j_2 + 1)\},$$

so that

$$\begin{aligned}
\frac{\partial h_{\mathbf{i}}^{\sigma,1}}{\partial x_{\mathbf{j}}}(x) = & \sum_{k_1, k_2} \sigma_1(|i_1 - k_1|)\sigma_2(|i_2 - k_2|)\mu_1\big(C_{k_1}(x^1)\big)\mu_2\big(C_{k_2}(x^2)\big)\delta_{\mathbf{ij}}\vec{e}^{\,1} \\
& + \frac{1}{2}\sigma_2(|i_2 - j_2|)\big(\sigma_1(|i_1 - j_1 + 1|) - \sigma_1(|i_1 - j_1|)\big) \\
& \times f_1\big(\tilde{x}_{j_1}^1\big)\int_{\tilde{x}_{j_2}^2}^{\tilde{x}_{j_2+1}^2}\big(x_{i_1}^1 - \tilde{x}_{j_1}^1\big)\left(\vec{e}^{\,1} + \frac{2\big(x_{j_2}^2 - \omega^2\big)}{\big(x_{j_1}^1 - x_{j_1-1}^1\big)}\vec{e}^{\,2}\right) f_2(\omega^2)\, d\omega^2 \\
& + \frac{1}{2}\sigma_2(|i_2 - j_2|)\big(\sigma_1(|i_1 - j_1 - 1|) - \sigma_1(|i_1 - j_1|)\big)f_1\big(\tilde{x}_{j_1+1}^1\big) \\
& \times \int_{\tilde{x}_{j_2}^2}^{\tilde{x}_{j_2+1}^2}\big(x_{i_1}^1 - \tilde{x}_{j_1+1}^1\big)\left((-\vec{e}^{\,1}) + 2\frac{\big(x_{j_2}^2 - \omega^2\big)}{\big(x_{j_1+1}^1 - x_{j_1}^1\big)}\vec{e}^{\,2}\right) f_2(\omega^2)\, d\omega^2 \\
& + \frac{1}{2}\sigma_1(|i_1 - j_1|)\big(\sigma_2(|i_2 - j_2 + 1|) - \sigma_2(|i_2 - j_2|)\big) \\
& \times f_2\big(\tilde{x}_{j_2}^2\big)\int_{\tilde{x}_{j_1}^1}^{\tilde{x}_{j_1+1}^1}\big(x_{i_1}^1 - \omega^1\big)\left(\vec{e}^{\,2} + \frac{2\big(x_{j_1} - \omega^1\big)}{\big(x_{j_2}^2 - x_{j_2-1}^2\big)}\vec{e}^{\,1}\right) f_1(\omega^1)\, d\omega^1 \\
& + \frac{1}{2}\sigma_1(|i_1 - j_1|)\big(\sigma_2(|i_2 - j_2 - 1|) - \sigma_2(|i_2 - j_2|)\big)f_2\big(\tilde{x}_{j_2+1}^2\big) \\
& \times \int_{\tilde{x}_{j_1}^1}^{\tilde{x}_{j_1+1}^1}\big(x_{i_1}^1 - \omega^1\big)\left((-\vec{e}^{\,2}) + \frac{2\big(x_{j_1}^1 - \omega^1\big)}{\big(x_{j_2+1}^2 - x_{j_2}^2\big)}\vec{e}^{\,1}\right) f_1(\omega^1)\, d\omega^1.
\end{aligned}$$

There are two components: the first one is

$$\begin{aligned}
& \left(\sum_{k_1=1}^{n_1} \sigma_1(|i_1 - k_1|)\mu_1\big(C_{k_1}(x^1)\big)\right)\left(\sum_{k_2=1}^{n_2} \sigma_2(|i_2 - k_2|)\mu_2\big(C_{k_2}(x^2)\big)\right)\delta_{\mathbf{ij}} \\
& \quad + \frac{1}{2}\Big[\big(\sigma_1(|i_1 - j_1 + 1|) - \sigma_1(|i_1 - j_1|)\big)f_1\big(\tilde{x}_{j_1}^1\big)\big(x_{i_1}^1 - \tilde{x}_{j_1}^1\big) \\
& \qquad - \big(\sigma_1(|i_1 - j_1 + 1|) - \sigma_1(|i_1 - j_1|)\big)f_1\big(\tilde{x}_{j_1+1}^1\big)\big(x_{i_1}^1 - \tilde{x}_{j_1+1}^1\big)\Big]
\end{aligned}$$

$$\times \sigma_2(|i_2 - j_2|)\mu_2\big(C_{j_2}(x^2)\big)$$
$$+\frac{1}{2}\left[\big(\sigma_2(|i_2 - j_2 + 1|) - \sigma_2(|i_2 - j_2|)\big)\frac{f_2\big(\tilde{x}^2_{j_2}\big)}{x^2_{j_2} - x^2_{j_2-1}}\right.$$
$$\left. -\big(\sigma_2(|i_2 - j_2 - 1|) - \sigma_2(|i_2 - j_2|)\big)\frac{f_2\big(\tilde{x}^2_{j_2+1}\big)}{x^2_{j_2+1} - x^2_{j_2}}\right]$$
$$\times 2\sigma_1(|i_1 - j_1|)\int_{\tilde{x}^1_{j_1}}^{\tilde{x}^1_{j_1+1}}\big(x^1_{i_1} - \omega^1\big)\big(x^1_{j_1} - \omega^1\big)f_1(\omega^1)\,d\omega^1;$$

the second one is

$$\frac{1}{2}\left[\big(\sigma_1(|i_1 - j_1 + 1|) - \sigma_1(|i_1 - j_1|)\big)f_1(\tilde{x}_{j_1})\frac{x^1_{i_1} - x^1_{j_1}}{x^1_{j_1} - x^1_{j_1-1}}\right.$$
$$\left. +\big(\sigma_1(|i_1 - j_1 - 1|) - \sigma_1(|i_1 - j_1|)\big)f_1\big(\tilde{x}^1_{j_1+1}\big)\frac{x^1_{i_1} - \tilde{x}^1_{j_1}}{x^1_{j_1+1} - x^1_{j_1}}\right]$$
$$\times 2\sigma_2(|i_2 - j_2|)\int_{\tilde{x}^2_{j_2}}^{\tilde{x}^2_{j_2+1}}\big(x^2_{j_2} - \omega^2\big)f_2(\omega^2)\,d\omega^2$$
$$+\frac{1}{2}\Big[\big(\sigma_2(|i_2 - j_2 + 1|) - \sigma_2(|i_2 - j_2|)\big)f_2\big(\tilde{x}^2_{j_2}\big)$$
$$-\big(\sigma_2(|i_2 - j_2 - 1|) - \sigma_2(|i_2 - j_2|)\big)f_2\big(\tilde{x}^2_{j_2+1}\big)\Big]$$
$$\times 2\sigma_1(|i_1 - j_1|)\int_{\tilde{x}^1_{j_1}}^{\tilde{x}^1_{j_1+1}}\big(x^1_{i_1} - \omega^1\big)f_1(\omega^1)\,d\omega^1.$$

We recognize the first term to be $\zeta^{(1)}_{i_1j_1}\zeta^{(2)}_{i_2j_2} + \alpha^{(1)}_{i_1j_1}\rho^{(2)}_{i_2j_2} + \nu^{(1)}_{i_1j_1}\gamma^{(2)}_{i_2j_2}$ and the second one to be $\eta^{(1)}_{i_1j_1}\chi^{(2)}_{i_2j_2} + \theta^{(1)}_{i_1j_1}\beta^{(2)}_{i_2j_2}$. The computation of

$$\frac{\partial h^{J,2}_i(x)}{\partial x_j}$$

follows the same pattern. □

REFERENCES

[1] BOUTON, C. (1988). Approximation gaussienne d'algorithmes à dynamique markovienne. *Ann. Inst. H. Poincaré Probab. Statist.* **24** 131–155.

[2] BOUTON, C. and PAGÈS, G. (1993). Self-organization and convergence of the one-dimensional Kohonen algorithm with non uniformly distributed stimuli. *Stochastic Process. Appl.* **47** 249–274.

[3] BOUTON, C. and PAGÈS, G. (1994). Convergence in distribution of the one-dimensional Kohonen algorithms when the stimuli are not uniform. *Adv. in Appl. Probab.* **26** 80–103.

[4] COTTRELL, M. and FORT, J.-C. (1987). Etude d'un algorithme d'auto-organisation. *Ann. Inst. H. Poincaré Probab. Statist.* **23** 1–20.

[5] COTTRELL, M., LETRÉMY, P. and ROY, E. (1993). Analyse d'une table de contingence au moyen d'une carte de Kohonen: une analyse des correspondances. In *Actes des XXV^e journées de l'A.S.U.* Vannes (France).
[6] DUFLO, M. (1991). *Méthodes récursives aléatoires*. Masson, Paris.
[7] GANTMACHER, F. R. (1959). *The Theory of Matrices* **1–2**. Chelsea, New York.
[8] JACOD, J. and SHIRYAEV, A. N. (1990). *Limit Theorems for Stochastic Processes*. Springer, Berlin.
[9] KOHONEN, T. (1982). Analysis of a simple self-organizing process. *Biol. Cybernet.* **44** 135–140.
[10] KOHONEN, T. (1989). *Self-Organization and Associative Memory*, 3rd ed. Springer, Berlin.
[11] KOHONEN, T. (1991). Self-organizing maps: optimization approaches. In *Artificial Neural Networks* (T. Kohonen et al., eds.) 981–990. North-Holland, Amsterdam.
[12] KUSHNER, H. J. (1984). *Approximation and Weak Convergence Methods for Random Processes with Applications to Stochastic System Theory*. MIT Press.
[13] KUSHNER, H. J. and CLARK, D. S. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer, New York.
[14] PAGÈS, G. (1993). Mosaïques de Voronoï, algorithmes de quantification de l'espace et intégration numérique. Abridged English version in *Proc. ESANN'93* (M. Verleysen, ed.) 221–228. Editions D Facto, Brussels.
[15] RITTER, H., MARTINETZ, T. and SCHULTEN, K. (1992). *Neural Computation and Self-Organizing Maps*. Addison-Wesley, New York.
[16] RITTER, H. and SCHULTEN, K. (1988). Convergence properties of Kohonen's topology conserving maps: fluctuation, stability and dimension selection. *Biol. Cybernet.* **60** 59–71.
[17] VARFIS, A. and VERSINO, C. (1992). Clustering of socio-economic data with Kohonen maps. *Neural Network World* **2** 813–834.

UNIVERSITÉ NANCY I
INSTITUT ELIE CARTAN
B.P. 239
F-54506 VANDOEUVRE-LÈS-NANCY CEDEX
FRANCE

LABORATOIRE DE PROBABILITÉS
URA 224
UNIVERSITÉ PIERRE ET MARIE CURIE
4, PLACE JUSSIEU
75252 PARIS CEDEX 05
FRANCE