Self-organization property of Kohonen's map with general type of stimuli distribution

Ali A. Sadeghi
Arbeitsgruppe Technomathematik
Universität Kaiserslautern
67653 Kaiserslautern,Germany.

Abstract

Here the self-organization property of one-dimensional Kohonen's algorithm in its $\mathbf{2k}-$neighbor setting with a general type of stimuli distribution and non-increasing learning rate is considered.We prove that the probability of self-organization for all initial values of neurons is uniformly positive. For the special case of a constant learning rate , it implies that the algorithm self-organizes with probability one.

Keywords Neural networks, Kohonen's SOM , One-dimensional self-Organization, Winner definition.

1 Introduction

Among several models which have been proposed for self-organizing feature maps , the Kohonen algorithm is the most popular one. The analytical behavior of the algorithm and also its very extended areas of application are widely investigated ,see [3, 5, 7, 10, 11] and references therein . Although the algorithm is computationally simple, analytical investigation of its properties turned out to be very complicated. The existing results , including this paper , deal basically with the one-dimensional Kohonen algorithm.

The Kohonen net consists of a layer of neurons which are labelled from 1 to N. The set of neurons , denoted here by I, has a fixed neighborhood relationship, that is, to every neuron i a set of neighboring neurons $V_i \subset I$ is assigned. To connect the set of neurons with data space Q, which is usually a subset of $\mathbb{R}^m$, each neuron i is mapped to a weight vector $X_i \in Q$.

Every $v \in Q$ corresponds with the winner neurons $i^*(v)$ which satisfy

$$\| X_{i^*(v)} - v \| \leq \| X_i - v \| \qquad \forall \ i \in I, \tag{1}$$

where $\| \, . \, \|$ denotes the Euclidean norm on $\mathbb{R}^m$. To avoid confusions we occasionally denote the winner $i^*(v)$ corresponding to the weight vector X by $i^*(X, v)$.

The adaptation of weight vectors will be carried out by an unsupervised learn process . At the end of the learn process the map $v \in Q \mapsto i^*(v) \in \{1, \cdots, N\}$ preserves (in some sense) the topology of the data space. The learning starts with an initial weight vector $X^0 = (X_1^0, \cdots, X_N^0)^T$, which will be updated as follows

$$X_i^{n+1} = X_i^n + \epsilon_n f_{i^*i}(v_{n+1} - X_i^n) \qquad \forall \ i \in I, \quad n = 0, 1, \cdots \tag{2}$$

where $v_{n+1} \in Q$ is an independent random variable distributed identically by some probability distribution P, ϵ_n is the so-called learning rate and $f : (i, i^*) \mapsto f_{ii^*} \in [0, 1]$ is the neighborhood function . The initial weight vector X^0 may be chosen at random from Q^N.

In this paper we consider the Kohonen algorithm in its one-dimensional case, where $Q \subset \mathbb{R}$ and $V_i = \{i-1, i, i+1\} \cap I$. Moreover

$$f_{i^*i} = \begin{cases} \gamma_0 & \text{if} \quad i = i^*, \\ \gamma_1 \quad , & \text{if} \quad i = i^* - 1 \quad \text{or} \quad i = i^* + 1, \\ \vdots & \qquad \vdots \\ \gamma_{\mathbf{k}} & \text{if} \quad i = i^* - \mathbf{k} \quad \text{or} \quad i = i^* + \mathbf{k}, \\ 0 & \text{otherwise}, \end{cases} \tag{3}$$

in which $\gamma_0 = 1$, $0 < \gamma_{\mathbf{k}} \leq \cdots \leq \gamma_1 \leq 1$ and $\mathbf{k} > 0$. The probability distribution P is assumed to be diffuse, that is , if A has Lebesgue measure zero , then $P(A) = 0$.

The evolution of the weights in the one-dimensional Kohonen algorithm can be decomposed into two phases:

1- Self-organization phase, in which the weights of neurons become organized.The set of organized states for this case reduces to the set of vectors in $\mathbb{R}^N$ which have ascending or descending components. An important fact distinguishing the one-dimensional algorithm is ,that if the weights become ordered (organized) they retain their ordering for ever.

The existing results concerning the self-organization property of the algorithm are limited to some special cases of the one-dimensional algorithm ,see [1, 2, 3, 4, 5, 6].

To provide a proof of the self-organization , Cottrell and Fort [3] and also Bouton and Pagès [1, 2] considered the case that ϵ_n is a constant and the algorithm is a homogeneous Markov process . Using the properties of these kinds of processes, they proved that for a width one neighborhood function ($\mathbf{k} = 1$ with $\gamma_1 = 1$) the algorithm self-organizes to a topology preserving map with probability one (for both uniform and non-uniform stimuli distributions).

Erwin et al. [4] studied the case of $\mathbf{k} \geq N/2$ and a uniform stimuli distribution. In the most recent papers Flanagan [5, 6] extended the results in [4] to the case of a more general probability distribution of stimuli with the restriction that $\epsilon_t \geq \alpha_m > 0$ and the neighborhood function is restrictly decreasing.

For $\mathbf{k} = 0$ it is known that the one-dimensional algorithm fails to make up a topology preserving map.

2- Convergence phase, in which the weight vectors converge to their final values. In this phase learn process (2) reduces to a special case of the Robbins-Monro algorithm and its behavior is governed by the so-called mean (or average) differential equation (m.d.e.). This phase has been studied in [1, 3, 7, 11]. Bouton and Pagés [1] considered the one dimensional algorithm in its width one setting with $\gamma_1 = 1$. For uniformly distributed stimuli it was shown that the algorithm converges to the fixed point of the m.d.e. with probability one . Later on Fort and Pagés extended this result to some more general class of neighborhood functions [7].

The almost sure convergence of the algorithm for a quite general class of distributions and neighborhood functions was studied by Sadeghi in [11], where it is shown that the m.d.e. is cooperative and irreducible and the set of organized states is positively invariant. This together with a stability result due to Hirsch implied the almost sure convergence of the algorithm under the condition that the m.d.e. owns a unique fixed point with ordered components.

In this paper we want to extend the knowledge about the algorithm in its self-organization phase. Section 2 includes a generalization of the winner definition , which allows us to investigate the behavior of the algorithm on whole Q^N. Section 3 includes the main results of the paper, Theorems 1 and 2, where it is shown that regardless of the initial weights of the neurons the probability of self-organization is uniformly positive . We assume no positive lower bound for ϵ_n. For the case of a constant learning rate , we show that the self-organization occurs with probability one. The conditions imposed on the neighborhood function are weak enough to ensure the convergence of the algorithm in its quantization phase[11], particularly a big class of non-increasing neighborhood functions is included. Some final remarks are made in section 4.

2 About the winner

Without loss of generality in the remainder of the paper we always set $Q = [0, 1]$. The following conventions will be adopted :

$$D \;:=\; \{x \in [0,1]^N | x_k = x_l \Rightarrow k = l \quad \forall \quad k, l \in I\}, \quad D' := [0,1]^N \setminus D,$$

$$
\begin{aligned}
F^+ &:= \{x \in [0,1]^N | \quad 0 < x_1 < x_2 < \ldots < x_N < 1\}, \\
F^- &:= \{x \in [0,1]^N | \quad 1 > x_1 > x_2 > \ldots > x_N > 0\} \quad \text{and} \quad F := F^+ \cup F^-, \\
\Psi_i^n &:= \{j \in I \quad | \quad X_i^n = X_j^n\}, \\
S(x,\eta) &:= \{y \in [0,1]^N \quad | \quad |x_i - y_i| < \eta \quad \forall\, i \in I\}.
\end{aligned}
$$

We denote by $|\Psi_i^n|$ the number of elements of Ψ_i^n.

The definition (1) assigns a unique winner neuron for almost all $x \in D$. Moreover, if $x, x^k \in D$ and $x^k \to x$, then the Lebesgue measure of the set of stimuli, for which $i^*(x,v) \neq i^*(x^k,v)$ tends to zero.This assures the stability of the winner on D.

To treat the algorithm theoretically, one needs to extend the winner definition, in a way that the uniqueness of the winner neuron on whole $[0,1]^N$ is guaranteed. Principally such an extension is easy to do ; many authors have chosen the neuron with the smallest index satisfying (1) as winner for the case of conflict.However, such a definition causes instabilities of winner; the Lebesgue measure of the set of stimuli ,for which $i^*(x,v) \neq i^*(x^k,v)$ does not tend to zero as $k \to \infty$. This is source of many difficulties in a mathematical study of the algorithm, see [1]. Sequelly we introduce a new definition of the winner which enables us to overcome the stability problem.

Let Σ_N be the set of functions $o^k(.)$,

$$o^k(.) : \{1, \cdots, N\} \to I, \quad k = 1, ..., N!.$$

Σ_N will be termed in this paper as the set of (inverse) ordering functions ; for $x \in \mathbb{R}^N$ let $\Sigma_N^x \subset \Sigma_N$ be the set of those ordering functions $o^k(.)$ which satisfy

$$i > j \Rightarrow x_{o^k(i)} \geq x_{o^k(j)} \quad \forall i,j \in \{1, \cdots, N\}.$$

If $x \in D$, then Σ_N^x consists of a unique ordering function which will be called the ordering of the point x. i will be referred as the ordering of the neuron j according to $o^k(.)$ if $o^k(i) = j$.

Set

$$
\begin{aligned}
W_x^0(v) &:= \{i \in I \quad | \quad |x_i - v| \leq |x_j - v| \quad \forall \quad j \in I\}; \\
W_x(v) &:= \{i \in W_x^0(v) \quad | \quad x_i \leq x_j \quad \forall \quad j \in W_x^0(v)\}.
\end{aligned}
$$

For any $x \in [0,1]^N$ and any given ordering function $o^k(.) \in \Sigma_N$ we define the $k-$winner $i_k^*(v)$, or more precisely $i_k^*(x,v)$, which is in fact a generalization of $i^*(v)$ to $[0,1]^N$.

DEFINITION. For $i \in W_x(v)$ if $v > x_i$ $(v < x_i)$, then $i_k^*(v)$ is the neuron which has the greatest (smallest) ordering in $W_x(v)$ according to $o^k(.)$.

As a simple example, let $N=3$ and $x=(0.1,0.2,0.2)$. Then

$$\begin{array}{ll} o^1(.)=\{(1,1)(2,2)(3,3)\}, & o^2(.)=\{(1,1)(2,3)(3,2)\}, \\ o^3(.)=\{(1,2)(2,1)(3,3)\}, & o^4(.)=\{(1,2)(2,3)(3,1)\}, \\ o^5(.)=\{(1,3)(2,2)(3,1)\}, & o^6(.)=\{(1,3)(2,1)(3,2)\}, \end{array}$$

$\Sigma_3=\{o^k(.) \quad k=1,\cdots,6\}$ and $\Sigma_3^x=\{o^1(.),o^2(.)\}$.

If $v=0.17$ then $W_x^0(v)=W_x(v)=\{2,3\}$, $i_1^*(v)=i_3^*(v)=i_4^*(v)=2$ and $i_2^*(v)=i_5^*(v)=i_6^*(v)=3$.

If $v=0.15$ then $W_x^0(v)=\{1,2,3\}$, $W_x(v)=\{1\}$, $i_k^*(v)=1$ for all k.

It is clear that if $x\in D$, then $W_x(v)$ consists of a unique neuron and therefore $i_k^*(v)=i^*(v)$ for almost all $v\in[0,1]$ and all $o^k(.)\in\Sigma_N$. But if the algorithm starts from a point $x\in D'$, then , as soon as $X^n\in D'$, there exists stimuli v for which $W_x(v)$ consists of more than one neuron and consequently the k-winner depends on the ordering function $o^k(.)$. This means , for $x\in D'$ the probability that F be reached within t steps depends on the ordering function which we choose. This is in accordance with the fact that in any neighborhood of a point $x\in D'$ there exists points of D with all different possible orderings.

Next let us introduce the following function :

$$P_x(\tau_\pi\le t):=\min_{o^k(.)\in\Sigma_N^x}\mathbf{P}(X^t\in\pi \mid X^0=x \quad ; \quad i^*(v):=i_k^*(v)).$$

Here $\mathbf{P}(.|.)$ denotes the conditional probability function and π is any subset of $[0,1]^N$. Note that $P_x(\tau_\pi\le t)$ is also the minimum of $\mathbf{P}(X^t\in\pi \mid X^0=x \quad ; \quad i^*(v):=i_k^*(v))$ over Σ_N.

The rest of this paper deals with the treatment of the analytical properties of $P_x(\tau_F\le t)$.

3 Self-Organization

In this section we present a proof of the self-organization property of Kohonen's algorithm. This proof has not only the advantage that it is valid for general $2\mathbf{k}$-neighbor setting with quite reasonable restrictions on the input distribution and the learning rate, but also it is shorter than the proofs which have been appeared up to now[1, 2, 3, 4, 5].

If there exists a $T\in\mathbb{N}$, such that $X^T\in F$ with probability one, then the algorithm is said to have the self-organization property.It is known that if $X^T\in F$ for some T, then $X^{T+n}\in F$ for all n, [10].

The method we use to establish the self-organization property is to show that the probability of making up an organized configuration within a finite number of steps is uniformly positive for all initial weights . As a first step , in Lemmas 1 - 3 , we show that $P_x(\tau_F \leq t)$ has a positive lower bound if any initial state x can be taken to the organized configuration with a positive probability for any $i^*_k(v) \in \Sigma^x_N$. Then , in Lemma 4 and Theorem 1 , we show that this is actually the case.

For $x \in [0,1]^N$ suppose Σ^x_N consists of m different ordering functions $o^i(.)$, $i = 1 \cdots, m$, $m \leq N!$. Let $p \in \{1, \cdots, m\}$ and define

$$S^p(x,\eta) := S(x,\eta) \cap \{y \in D \quad | \quad \Sigma^y_N = \{o^p(.)\}\}, \qquad S^{m+1}(x,\eta) := S(x,\eta) \cap D',$$
$$S^{m+1,p}(x,\eta) := \{y \in S^{m+1}(x,\eta) \quad | \quad P_y(\tau_F \leq t) = \mathbf{P}(Y^t \in F \quad | \quad Y^0 = y, i^*(v) := i^*_p(v))\}.$$

For sufficiently small η it is easy to verify that $S(x,\eta) = \cup_{p=1}^m (S^p(x,\eta) \cup S^{m+1,p}(x,\eta))$.

$\mathcal{V}_p(x)$ is the set of all events $\nu = (v_1, \cdots, v_t)$ which take $X^0 = x$ to F with $i^*(v) := i^*_p(v)$, where $o^p(.) \in \Sigma^x_N$.

Lemma 1. *For any $x \in [0,1]^N$ and $\nu \in \mathcal{V}_p(x)$ there exists a $\eta > 0$ such that ν takes all $y \in S^p(x,\eta) \cup S^{m+1,p}(x,\eta)$ to F by putting $i^*(v) := i^*_p(v)$.*

Proof. The proof amounts to the study of two different cases.

Case 1, $x \in D$. In this case the winner is uniquely defined and we do not mention explicitly about the ordering function.

Let $\nu = (v_1, \cdots, v_n) \in \mathcal{V}_p(x)$ be any chain of stimuli which takes x to F and $X^1, \cdots, X^t$ the corresponding weight vectors. k denotes the ordering of i^* at time n, that is , $o(k) = i^*(v_n, X^n)$ and $\bar{X}^n_{i^*} := 0.5(X^n_{i^*} + X^n_{o(k-1)})$, $\bar{X}^n_{i^*+1} := 0.5(X^n_{i^*} + X^n_{o(k+1)})$.

For $x \in D$ there exista a $\eta > 0$ such that

$$\eta < 0.5 \min(|\bar{X}^n_{i^*+1} - v_n|, |\bar{X}^n_{i^*} - v_n|), \quad \forall \quad n = 0, \cdots, t-1.$$

Note that the set of all stimuli for which η dose not exist has Lebesgue measure zero. Moreover it is independent of the ordering function $o^p(.)$.

We show, by induction, that if $Y^0 = y \in S(x,\eta)$, then $Y^t \in S(X^t,\eta)$.

Let $Y^{n-1} \in S(X^{n-1},\eta)$ for $n \in \{1, \cdots, t\}$. Because of $\eta < 0.5 \min(|\bar{X}^n_{i^*+1} - v_n|, |\bar{X}^n_{i^*} - v_n|)$, we have $i^*(v_n, Y^{n-1}) = i^*(v_n, X^{n-1})$ and therefore

$$Y^n_i = (1 - \epsilon_n \gamma_{|i-i^*|}) Y^{n-1}_i + \epsilon_n \gamma_{|i-i^*|} v_n \quad \forall \quad i \in I, \tag{4}$$

which implies,

$$X_i^n - (1 - \epsilon_n \gamma_{|i-i^*|})\eta < Y_i^n < X_i^n + (1 - \epsilon_n \gamma_{|i-i^*|})\eta \quad \forall \quad i \in I,$$

that is, $Y^n \in S(X^n, \eta)$.

Recall that F is an open subset of $[0,1]^N$, for sufficiently small η we have $Y^t \in F$.

Case 2, $x \in D'$. Consider again any event $\nu = (v_1, \cdots, v_n) \in \mathcal{V}_p(x)$ and the corresponding weight vectors $X^1, \cdots, X^t$. Let k be the ordering of i_p^* at time n, that is , $o^p(k) = i_p^*(v_n, X^n)$ and $\bar{X}_{i_p^*}^n := 0.5(X_{i_p^*}^n + X_{o^p(k-1)}^n)$, $\bar{X}_{i_p^*+1}^n := 0.5(X_{i_p^*}^n + X_{o^p(k+1)}^n)$.

Choose a $\eta > 0$ which satisfies

$$\eta < 0.5 \min(|\bar{X}_{i_p^*+1}^n - v_n|, |\bar{X}_{i_p^*}^n - v_n|), \quad \forall \quad n = 0, \cdots, t-1.$$

Suppose $Y^0 = y \in S^p(x, \eta) \cup S^{m+1,p}(x, \eta)$, $p \in \{1, \cdots, m\}$. We use induction to show that, after $\nu = (v_1, \cdots, v_n)$ with $i^*(v) := i_p^*(v)$, we have $Y^t \in S(X^t, \eta)$.

Assume that $Y^n \in S(X^n, \eta)$ for all $n \in \{0, \cdots, l-1\}$, where $l \in \{1, \cdots, t-1\}$. We show this is also true for $n = l$.

As a first step let us show that $i_p^*(v_l, Y^{l-1}) = i_p^*(v_l, X^{l-1})$.

It is clear that $i_p^*(v_1, Y^0) = i_p^*(v_1, X^0)$. Now let us assume $i_p^*(v_n, Y^{n-1}) = i_p^*(v_n, X^{n-1})$ for $n = 1, \cdots, s-1$, $s \le l$. It is enough to show this is also true for $n = s$.

Suppose $|\Psi_{i_p^*(v_s, X^{s-1})}^{s-1}| > 1$. With probability one for all $i, j \in \Psi_{i_p^*(v_s, X^{s-1})}^{s-1}$ we have $X_i^n = X_j^n$, $n = 0, 1, \cdots, s-1$. This can be the case only if $Y_i^n = Y_j^n$, which implies that for both X^{s-1} and Y^{s-1} $\Psi_{i_p^*(v_s, X^{s-1})}^{s-1}$ is the same, that is , for $j \in \Psi_{i_p^*(v_s, X^{s-1})}^{s-1}$ we have $Y_i^{s-1} = Y_j^{s-1}$ iff $i \in \Psi_{i_p^*(v_s, X^{s-1})}^{s-1}$. This together with $Y^{s-1} \in S(X^{s-1}, \eta)$ and $\min_{k,l \in I, X_k^{s-1} \neq X_l^{s-1}} |X_k^{s-1} - X_l^{s-1}| > 6\eta$ implies that $i_p^*(v_n, Y^{n-1}) = i_p^*(v_n, X^{n-1})$ for $n = s$.

For $|\Psi_{i_p^*(v_s, X^{s-1})}^{s-1}| = 1$, the definition of $i_p^*(v)$ implies $i_p^*(v_s, X^{s-1}) = i_p^*(v_s, Y^{s-1})$.

Now equation (4) holds, that is , after the event v_l we have $Y^l \in S(X^l, \eta)$.

We come to the conclusion that if $Y^0 = y \in S^p(x, \eta) \cup S^{m+1,p}(x, \eta)$ and $\eta < 0.5 \min(|\bar{X}_{i_p^*+1}^n - v_n|, |\bar{X}_{i_p^*}^n - v_n|)$, then after any event $(v_1, \cdots, v_t) \in \mathcal{V}_p(x)$, by putting $i^*(v) := i_p^*(v)$, we have $Y^t \in S(X^t, \eta)$, that is , for sufficiently small η, $Y^t \in F$.

□

Remark 1. Let $\mathcal{V}_p^\alpha(x) \subset \mathcal{V}_p(x)$ be the set of all events $(v_1, \cdots, v_t)$ taking any $y \in S^p(\alpha, x) \cup S^{m+1,p}(\alpha, x)$ to F. We have $\mathcal{V}_p^\beta(x) \subset \mathcal{V}_p^\alpha(x)$ if $\beta > \alpha$. Furthermore from the proof of the Lemma 1 it follows that if P is diffuse, then for any $\epsilon > 0$ there exists an $\alpha > 0$ such that $\mathbf{P}(\mathcal{V}_x \setminus \mathcal{V}_x^\alpha) < \epsilon$.

Lemma 2. *If* P *is diffuse , then ,for all* $t \in \mathbb{N}$, $P_x(\tau_F \le t)$ *is lower semi-continuous on* $[0,1]^N$.

Proof. To establish the claim of Lemma we need to prove for any arbitrary sequence x^k if $x^k, x \in [0,1]^N$ and $x^k \to x$ as $k \to \infty$, then $\underline{\lim}_{k\to\infty} P_{x^k}(\tau_F \le t) \ge P_x(\tau_F \le t)$.

Let m denotes the number of ordering functions in Σ_N^x. There exists an integer K such that for all $k \ge K$ we have $\Sigma_N^{x^k} \subset \Sigma_N^x$. This implies that we can split up the sequence x^k, $k = K, \cdots$ into m disjoint subsequences $x^{k,1}, \cdots, x^{k,m}$ such that $x^{k,p} \in S^p(x,\eta) \cup S^{p,m+1}(x,\eta)$ for $p = 1, \cdots, m$ if η is sufficiently small .

Now Lemma 1 implies $\underline{\lim}_{k\to\infty} P_{x^{k,p}}(\tau_F \le t) \ge \mathbf{P}(\mathcal{V}_p^\alpha(x))$ for any sufficiently small $\alpha > 0$, which implies $\underline{\lim}_{k\to\infty} P_{x^{k,p}}(\tau_F \le t) \ge \mathbf{P}(\mathcal{V}_p(x))$. On the other hand , we have $\mathbf{P}(\mathcal{V}_p(x)) \ge P_x(\tau_F \le t)$, which implies

$$\underline{\lim}_{k\to\infty} P_{x^k}(\tau_F \le t) \ge P_x(\tau_F \le t) \quad \forall \quad x \in [0,1].$$

□

In the proof of Lemma 3 we use the known fact that for a decreasing sequence of closed sets F_t, if $\bigcap_{t\ge 0} F_t = \emptyset$, then there exists a T such that $F_T = \emptyset$, see e.g. [8] .

Lemma 3. *Suppose* P *is diffuse and* $supp\ P = [0,1]$. *If* $P_x(\tau_F < \infty) > 0$ *for all* $x \in [0,1]^N$, *then*

$$\exists \quad \theta > 0, \qquad \exists \quad T \in \mathbb{N} \quad \textit{such that} \quad \forall \quad x \in [0,1]^N, \quad P_x(\tau_F \le T) > \theta.$$

Proof. Let be $F_t = \{x \in [0,1]^N \quad | \quad P_x(\tau_F \le t) = 0\}$. The assumption $P_x(\tau_F < \infty) > 0$ implies that $\bigcap_{t\ge 0} F_t = \emptyset$. Moreover $F_{t+1} \subset F_t$. Using the lemma 1 we know that F_t is closed for all t and this implies the existence of $T \in \mathbb{N}$ with $F_T = \emptyset$. Now $P_x(\tau_F \le T)$ is a positive and lower semi-continuous function of x on $[0,1]^N$ which implies the existence of a lower bound $\theta > 0$.

□

It is remarkable that the lemmas 1-3 remain true if we replace F with any other open subset of $[0,1]^N$.

Lemma 4. *Suppose* P *is diffuse ,* $suppP = [0,1]$, $\mathbf{k} \ge 1$, $\gamma_{\mathbf{j}+1} < \gamma_{\mathbf{j}}$ *for some* $\mathbf{j}$, $0 \le \mathbf{j} \le \mathbf{k}$ *and* $N \ge 2\mathbf{j}+1$. *For any* $x \in D'$ *there exists a finite integer* T' *such that* $P_x(\tau_D < T') > 0$.

Proof. We show for any arbitrary $o^s(.) \in \Sigma_N^x$ there exists some positive T' such that $\mathbf{P}(X^{T'} \in D \quad | \quad X^0 = x \quad , i^*(v) := i^*_s(v)) > 0$. Remind that $\omega_i := \{i - \mathbf{j}, \cdots, i, \cdots, i + \mathbf{j}\} \cap I$.

Suppose $|\Psi_i^0| \leq l, \quad l > 1$, for all $i \in I$. We show that $|\Psi_i^R| \leq l - 1$ for all $i \in I$ with a positive probability , where $R \leq 3|A|/l$ and $A := \{i \in I| \quad |\Psi_i^0| = l\}$.

Let $|\Psi_i^0| = l$ and p, q be the smallest and greatest orderings in Ψ_i^0 according to $o^s(.) \in \Sigma_N^x$, respectively.

If either $\Psi_i^0 \setminus \omega_{o^s(q)} \neq \emptyset$ or $\Psi_i^0 \setminus \omega_{o^s(p)} \neq \emptyset$, then the event $i^*(v_1) = o^s(q)$ or $i^*(v_1) = o^s(p)$ leads to $\Psi_i^1 \leq l - 1$, respectively. Both of these events may happen with a positive probability.

Now suppose $\Psi_i^0 \subset \omega_{o^s(q)}$ and $\Psi_i^0 \subset \omega_{o^s(p)}$. This can happen only if $l \leq 2\mathbf{j}$. There exists a neuron u such that $u \notin \Psi_i^0$ and either $u \in \omega_{o^s(q)} \setminus \omega_{o^s(p)}$ or $u \in \omega_{o^s(p)} \setminus \omega_{o^s(q)}$. Without loss of generality ,we suppose $u \in \omega_{o^s(q)} \setminus \omega_{o^s(p)}$.

If the event $i^*(v_1) = u$ happens with a positive probability , then the probability of $|\Psi_i^1| \leq l - 1$ is positive,too. In otherwise , after the events $i^*(v_1) = o^s(q)$ and $i^*(v_2) = o^s(p)$ we have $\Psi_u^1 \subset \omega_{o^s(q)}$ and $\Psi_u^2 \cap \omega_{o^s(p)} = \emptyset$. Let q' be the greatest ordering acording to $o^s(.)$ in Ψ_u^2. Then the event $i^*(v_3) = o^s(q')$ moves $o^s(q)$ while $o^s(p)$ remains fixed , i.e., with a positive probability we have $|\Psi_i^3| \leq l - 1$.

Repeating these events for all different set of neurons with $|\Psi_i^0| = l$, after R steps we have $|\Psi_i^R| \leq l - 1$ for all i.

□

Theorem 1. *Suppose P is diffuse , supp $P = [0, 1]$, $\Sigma_n \epsilon_n = \infty$, $\mathbf{k} \geq 1$, $\gamma_{\mathbf{j}+1} < \gamma_{\mathbf{j}}$ for some $\mathbf{j}$, $0 \leq \mathbf{j} \leq \mathbf{k}$ and $N \geq 2\mathbf{j} + 1$. Then*

$$\exists \quad \theta > 0 \qquad \exists \quad T \in \mathbb{N} \quad \text{such that} \quad \forall \quad x \in [0, 1]^N, \quad P_x(\tau_F \leq T) > \theta.$$

Proof. To establish the assertion using Lemmas 3 and 4 and the Markov property it is enough to show that for any $x \in D$ there exists a $T'' \in N$ such that $\mathbf{P}(\tau_F \leq T''|X^0 = x) > 0$. We will introduce a finite chain of events which takes $X^0 = x \in D$ to F with a positive probability. We emphasize that there are other chain of events which do the job as well.

Note that if $X^n \in D$, then every neuron may win at time n with a positive probability.

Let $F_k^+ = \{x \in [0, 1]^N \quad | \quad 0 < x_1 < \cdots < x_k, \quad x_i > x_k \quad \forall \quad i > k\}$.

If $x \notin F_1^+$ but there exists a neuron r, with $r > \mathbf{j} + 1$ such that $x_r > x_1$ then after the event

$$i^*(v_n) = r, \qquad 1 \leq n < T_1,$$

for sufficiently big T_1 we have $x_i > x_1$ for $i \in \omega_r$. With an analogous argument the following chain of events takes x to F_1^+ :

$$\begin{aligned}
i^*(v_n) &= r + \mathbf{j}, & T_1 \leq n < T_2,\\
&\vdots \\
i^*(v_n) &= r + l\mathbf{j}, & T_{l-1} \leq n < T_l,\\
i^*(v_n) &= r - \mathbf{j}, & T_l \leq n < T_{l+1},\\
&\vdots \\
i^*(v_n) &= r - p\mathbf{j}, & T_{l+p-1} \leq n < T_{l+p},\\
i^*(v_n) &= \mathbf{j} + 2, & T_{l+p} \leq n < T_{l+p+1}.
\end{aligned}$$

Here $r + l\mathbf{j} \leq N < r + (l+1)\mathbf{j}$ and $r - (p+1)\mathbf{j} < \mathbf{j} + 2 \leq r - p\mathbf{j}$. The times T_s, $1 \leq s \leq l + p + 1$, are a numbers of iterations after which we have $x_i > x_1$ for all $i \in \omega_{i^*(v_{T_{s-1}})}$. The condition $\Sigma_n \epsilon_n = \infty$ implies that T_s is finite for all $1 \leq s \leq l+p+1$.

If $x \in F_k^+$ for some $k \in \{1, \cdots, \mathbf{j} - 1\}$, then after the chain of events

$$i^*(v_n) = k, \qquad 1 \leq n < T_1',$$

we have $x \in F_k^+$ and moreover $x_k < x_i < x_l$ for all $i \in \{k+1, \cdots, k+\mathbf{j}\}$ and $l \geq k + \mathbf{j} + 1$.

Now the following chain of events takes x to F_{k+1}^+ :

$$i^*(v_n) = k + \mathbf{j} + 2, \qquad T_1' \leq n < T_2',$$

where T_2' is the smallest time after which we have $x_{k+1} < x_i$ for all $k+1 < i \leq N$. Again $\Sigma_n \epsilon_n = \infty$ implies that T_1' and T_1' are finite.

If $x \in F_k^+$ and $k \geq \mathbf{j}$ then the event $i^*(v_n) = k - \mathbf{j} + 1$ moves the neuron $k+1$ to the left while the neurons $k+2$ up to N remain fixed. Repeating this event ,after some finite steps $T_1'' \in N$ we have $x_{k+1} < x_i$ for all $i > k+1$, that is , $X^{T_1''} \in F_{k+1}^+$.

The only case which remains is $x \notin F_1$ and there does not exists some neuron $r > \mathbf{j} + 1$, such that $x_r > x_1$. In this case with exactly the same argument , this time for $F_k^- := \{x \in [0,1]^N \mid 1 > x_1 > \cdots > x_k, \quad x_i < x_k \quad \forall \quad i > k\}$, it can be shown that after a finite number of iterations with a positive probability $X^{\mathcal{T}} \in F_1^-$. Moreover if $x \in F_k^-$ $1 \leq k < N$, then with a positive probability after some finite time we have $X^{\mathcal{T}'} \in F_{k+1}^-$.

□

It should be mentioned that , the constant θ generally is time dependent, so that the Theorem 1 can not be considered as a with probability one self-organization result. Nevertheless, in case of a constant learning rate, θ is time independent and the following extension of the Theorem 1 holds.

Theorem 2. *Suppose* P *is diffuse ,* *supp* $P = [0,1]$, ϵ_n *is constant,* $\mathbf{k} \geq 1$, $\gamma_{\mathbf{j}+1} < \gamma_{\mathbf{j}}$ *for some* $\mathbf{j}$, $0 \leq \mathbf{j} \leq \mathbf{k}$ *and* $N \geq 2\mathbf{j}+1$. *Then there exists a finite integer* T *such that for all* $x \in [0,1]^N$ *and* $o^k(.) \in \Sigma_N$ *we have*

$$\mathbf{P}(\tau_F \leq T \mid X^0 = x, i^*(v) := i^*_k(v)) = 1.$$

Remark 2. To establish the Theorems 1 and 2 we did not used the symmetry of the neighborhood function f. Therefore these results are also valid for asymmetric non-increasing neighborhood functions. In such a cases $\mathbf{j}$ is the smallest index which satisfies $\gamma_{\mathbf{j}+1} < 1$ and $\gamma_{-\mathbf{j}-1} < 1$.

4 Conclusion

Here we proved the almost sure self-organization of one-dimensional Kohonen algorithm, under quite reasonable restrictions on the stimuli distribution , neighborhood function and the learning rate. A natural question which arises is how to generalize the proof to the case of a more general data space and network structure? There are two major difficulties which prevent such a generalization.

Firstly , there does not exists an absorbing set like F in multi-dimensional cases. Nevertheless , the set of organized states may be defined , see [5, 7, 12] for different definitions. A possible approach to tackle the multi-dimensional case could be to show a weak self-organization property , in the sense that , the neuron weights go to an organized configuration in finite time with probability one , although later on they may leave this configuration.

Secondly , no generalization of winner definition is known, which possesses the properties of the $k-$winner as was defined in section 2 of this paper. It was the only difficulty which prevented the establishment of a result similar to Lemmas 1-3 for multi-dimensional algorithm.

NOMENCLATURE

F	Set of organized configurations
F^+	Set of ascending configurations
F^-	Set of descending configurations
f_{ii^*}	Neighborhood function
I	Set of neurons
$\mathbf{j}$	The smallest index with $\gamma_{\mathbf{j}} < 1$.
$\mathbf{k}$	Width of neighborhood function
$\underline{\lim}$	Lower limit point of a sequence
n	Discrete time index, $n \in \mathbb{N}$
N	Total number of neurons
$o^k(.)$	Ordering function
P	Data distribution
$\mathbf{P}(.\vert.)$	The conditional probability function
Q	Data space
$\mathbb{R}$	Set of real numbers
$\mathbb{N}$	Set of positive integers numbers
X^n	Weight vector at time n
$\vert\Psi\vert$	Number of the elements of a set Ψ
τ_π	The time to reach a set π
ϵ_n	Learning rate
γ_i	Neighborhood function for the ith neuron
$\emptyset$	Empty set
ω_i	$\{i - \mathbf{j}, \cdots, i, \cdots, i + \mathbf{j}\} \cap I$

References

[1] Bouton, C., & Pagés, G. (1993) Self-organization and a.s. convergence of the one-dimensional Kohonen algorithm with non-uniformly distributed stimuli.*Stochastic Procces and their Applications* , **47** , 249-274 .

[2] Bouton, C., & Pagés, G. (1994) Convergence in distribution of the one-dimensional Kohonen algorithm when the stimuli are not uniform . *Advances in Applied Probability*, **26**, 80-103 .

[3] Cottrell,M., & Fort,J.C. (1987). Étude d'un processus d'auto-organisation . *Annales de l' Institut Henri Poincaré*,**23**(1) , 1-20 (in French).

[4] Erwin, E., Obermeyer,K. & Schulten ,K.(1992). Self-organizing maps: ordering, convergence properties and energy functions.*Biological Cybernetic* ,**67** , 47-55 .

[5] Flanagan,J.A. (1996). Self-organisation in Kohonen's SOM,*Neural Networks* ,**7** , 1185-1197.

[6] Flanagan,J.A. (1997). Self-organisation in the one -dimensional SOM with a reduced width neighbourhood,*Proc. SOM Workshop, Helsinki.*

[7] Fort, J.C., Pagés, G. (1996) . About the Kohonen Algorithm: Strong or Weak Self-Organization ? *Neural Networks* , **5**, 773-785.

[8] Hewitt, E. & Stromberg, K. (1969).*Real and abstract analysis.* Springer-Verlag .

[9] Kohonen, T. (1982). Self-organized formation of topologically correct feature maps.*Biological Cybernetics*,**43**, 59-69.

[10] Kohonen,T. (1995). *Self-organizing maps* , Springer-Verlag .

[11] Sadeghi, A.A. . Asymptotic behavior of self-organizing maps with non-uniform stimuli distribution, preprint 166, Arbeitsgruppe Technomathematik, Universität Kaiserslautern, Juli 1996, (to appear in The Annals of Applied Probability).

[12] Villmann,T., Der, R. , M. Herrmann, M. & Martinetz,T. .(1997). Topology preservation in self-organizing feature maps: Exact definition and measurement.*IEEE Transaction on Neural Networks*, **2**, , 256-266.