ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 167-172

Self-organization and convergence of the one-dimensional Kohonen algorithm.

Ali A. Sadeghi

Fachbereich Mathematik , Universität Kaiserslautern
67653 Kaiserslautern , Germany

Abstract. Here the self-organization and a.s. convergence of one-dimensional Kohonen's algorithm in its **2k**−neighbor setting with general type of stimuli distribution and non-increasing learning rate is considered.We show that the probability of self-organization for all initial values of neurons is uniformly positive. Moreover, in the convergence phase the asymptotic behavior of the algorithm is governed by a cooperative and irreducible differential equation. This implies the a.s. convergence of algorithm if the differential equation has a unique fixed point.

1. Introduction

We start with a short definition of the Kohonen algorithm.

Let I be a finite set of neurons labelled from 1 to N. There is a neighborhood function $f_{ij} : I \times I \to [0,1]$ such that $f_{i,i-l} = f_{i,i+l} := \gamma_l \in [0,1]$, $\gamma_l \geq \gamma_{l+1}$, $\gamma_0 = 1$. A weight X_i^n is associated to each neuron i at time n and $X^n := (X_i^n)_{1 \leq i \leq N}$ denotes the weight vector. Every $v \in [0,1]$ corresponds with the winner neuron $i^*(v)$ which satisfy

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

The algorithm starts usually with a randomly chosen X^0. The weights X_i^n are then adapted in the learning phase according to

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

where $v_n \in [0,1]$ are i.i.d. random variables with probability distribution P,$\epsilon_n \in (0,1)$ is the learning parameter.

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

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

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 167-172

Robbins-Monro algorithm and its behavior is governed by the so-called mean (or average) differential equation. This phase has been studied in [2, 3, 7, 9].

In section 2 we show that under quite general conditions , the probability of self-organization is uniformly positive, regardless of the initial weights of the neurons. In section 3 it is shown that the mean differential equation is cooperative and irreducible in F^+ and $\bar{F}^+$ is positively invariant. This enables us to establish the a.s. convergence of the algorithm in its convergence phase.

2. Self-Organization

2.1. The Winner Definition

We adopt the following conventions :

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

The definition (1) assigns a unique winner neuron for almost all $(x, v) \in D \times [0,1]$. 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. Hier we introduce a new definition of the winner which possesses both uniqueness and stability properties on whole $[0,1]^N$.

Let Σ_N be the set of functions $o^k(.) : \{1, \cdots, N\} \to I$, $k = 1, ..., N!$. For $x \in \mathbb{R}^N$, $\Sigma_N^x \subset \Sigma_N$ denotes the set of those functions in $o^k(.)$ 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 function. i will be referred as the ordering of the neuron j according to $o^k(.)$ if $o^k(i) = j$. Now set

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

For any $x \in [0,1]^N$ and any given 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 the winner definition 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 $o^k(.)$.

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.

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 167-172

Next let us introduce the following function :

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

Here $\mathbf{P}(.|.)$ denotes the conditional probability function. The rest of this section deals with the treatment of the analytical properties of $P_x(\tau_F \le t)$.

2.2. Proof of Self-Organization

The proof presented here is valid for general **2k**-neighbor setting with quite reasonable restrictions on the input distribution and the learning rate.The method we use to establish the self-organization property is to show that $P_x(\tau_F \le t) \ge \theta > 0$ uniformly if $P_x(\tau_F \le t) > 0$ for all x.

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

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

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

Sequelly $\mathcal{V}_p(x)$ denotes 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 $o^p(.) \in \Sigma_N^x$.

Lemma 1. *For any $x \in [0,1]^N$ and almost all $\nu \in \mathcal{V}_p(x)$ 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. Consider any event $\nu = (v_1,\cdots,v_t) \in \mathcal{V}_p(x)$ and the corresponding weight vectors $X^1,\cdots,X^t$. Let k be the ordering of i_p^* at time n, that is , $o^p(k) = i_k^*(v_n, X^n)$ and define $\bar{X}_{i_p^*}^n := 0.5(X_{i_p^*}^n + X_{o^p(k-1)}^n)$, $\bar{X}_{i_p^*+1}^n := 0.5(X_{i_p^*}^n + X_{o^p(k+1)}^n)$.

Choose a $\eta > 0$ which satisfies

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

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

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 $\nu = (v_1,\cdots,v_t)$ 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 $|\Psi_{i_p^*(v_s,X^{s-1})}^{s-1}| > 1$. ($|\Psi|$ denotes the number of the elements of a set Ψ). With probability one for all $i,j \in \Psi_{i_p^*(v_s,X^{s-1})}^{s-1}$ we have $X_i^n =$

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 167-172

$X_j^n, \quad n = 0, 1, \cdots, s-1$. This can be the case only if $Y_i^n = Y_j^n$, which implies $\Psi^{s-1}_{i_p^*(v_s, X^{s-1})} = \Psi^{s-1}_{i_p^*(v_s, Y^{s-1})}$, that is, for $j \in \Psi^{s-1}_{i_p^*(v_s, Y^{s-1})}$ we have $Y_i^{s-1} = Y_j^{s-1}$ iff $i \in \Psi^{s-1}_{i_p^*(v_s, X^{s-1})}$. This together with $Y^{s-1} \in S(X^{s-1}, \eta)$ and $\eta < 0.5 \min(|\bar{X}^n_{i^*+1} - v_n|, |\bar{X}^n_{i^*} - v_n|)$ implies that $i_p^*(v_n, Y^{n-1}) = i_p^*(v_n, X^{n-1})$ for $n = s$.

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

Now we have

$$Y_i^n = (1 - \epsilon_n \gamma_{|i - i_p^*|}) Y_i^{n-1} + \epsilon_n \gamma_{|i - i_p^*|} v_n \quad \forall \quad i \in I,$$

which implies,

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

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 $Y^t \in F$.

□

The Lemmas 2 and 3 and Theorem 1 are direct implications of Lemma 1. Their exact proofs are presented in [10].

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

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

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

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

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

3. Almost Sure Convergence

In this section we review the most recent results concerning the a.s. convergence of the one-dimensional Kohonen algorithm and some open questions which still need to be worked.For the winner definition we return to the original definition (1). In the case that there is more than one possible choice for the winner, we choose the neuron with smallest index as the winner.

It is known that after self-organization the following differential equation governs the asymptotic behavior of the Kohonen algorithm.

$$\dot{x} = \begin{bmatrix} -R_1(x)x_1 + S_1(x) \\ \vdots \\ -R_N(x)x_N + S_N(x) \end{bmatrix}, \tag{3}$$

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 167-172

where $x \in \mathbb{R}^N$ (N is the number of neurons) and

$$R_i(x^n) := P_i(x^n) + (P_{i-1}(x^n) + P_{i+1}(x^n))\gamma_1 + \cdots + (P_{i-k}(x^n) + P_{i+k}(x^n))\gamma_k,$$
$$S_i(x^n) := Q_i(x^n) + (Q_{i-1}(x^n) + Q_{i+1}(x^n))\gamma_1 + \cdots + (Q_{i-k}(x^n) + Q_{i+k}(x^n))\gamma_k,$$
$$P_i(x^n) := P([\bar{x}_i^n, \bar{x}_{i+1}^n]), \quad Q_i(x^n) := \int_{\bar{x}_i^n}^{\bar{x}_{i+1}^n} vP(dv), \quad \bar{x}_i^n := 0.5(x_i^n + x(i-1)^n).$$

In the remainder of this paper $h(x)$ denotes the right hand side of (3).

The a.s. convergence of one-dimensional Kohonen's algorithm has been investigated in [2, 3, 7] . The results established in these papers confirm a.s. convergence of the algorithm to the zero of $h(x)$, if the stimuli is distributed uniformly and ϵ_n converges to zero slowly enough. The major difficulty, which prevented a generalization of this result to the non-uniform case , was that no Liapunov function is known for (3) in general case.

In [9] the cooperative and irreducible characters of (3) are employed to prove that X^n converges almost surely (or with probability one), if (3) possesses an unique equilibrium in $\bar{F}^+$. This result is proved in following steps.

Lemma 4. *Consider the set* $F^+ := \{x \in [0,1]^N \mid 0 < x_1 < x_2 < \ldots < x_N < 1\}$,

- *(i) If* P *is diffuse ,then there exists a* $x^* \in \bar{F}^+$ *such that* $h(x^*) = 0$.
- *(ii) If* P *is diffuse ,* supp $P = [0,1]$, $\gamma_{j+1} < \gamma_j$ *and* $N \geq 2j+1$ *for some* j, $0 \leq j \leq k$, *then* $x^* \in F^+$.
- *(iii) Under the same assumptions as in (ii)* F^+ *is positively invariant and any solution* $x(t)$ *of (3) with* $x(0) \in F^+$ *has a compact closure in* F^+.
- *(iv) Under the same assumptions as in (ii) any solution* $x(t)$ *of (3) with* $x(0) \in \partial F^+$ *satisfies* $x(t) \in F^+$ *for* $t > 0$.

Lemma 5. *Suppose* supp $P = [0,1]$, P *has a density* $P(dv) = p(v)dv$ *which is continuous on* $[0,1]$, $p > 0$ *on* $(0,1)$ *and* $\gamma_k \leq \gamma_{k-1} \leq \cdots \leq \gamma_1 \leq 1$, *then we have*

- *(i) The m.d.e. (3) is cooperative on* F^+.
- *(ii) If* $\gamma_{j+1} < \gamma_j$ *for some* $0 \leq j \leq k$ *and* $N > 2j+1$, *then the m.d.e. (3) is irreducible on* F^+ .

Lemmas 4 and 5 are proved in [9].

Theorem 2. *Assume the following conditions hold :*

- *(i)* $\epsilon_n > 0 \quad \forall \quad n$, $\sum_n \epsilon_n^2 < \infty$ *and* $\sum_n \epsilon_n = \infty$;
- *(ii)* supp $P = [0,1]$, P *has a density* $P(dv) = p(v)dv$ *which is continuous on* $[0,1]$ *and* $p > 0$ *on (0,1);*
- *(iii)* $\gamma_{j+1} < \gamma_j$ *for some* $0 \leq j \leq k$ *and* $N > 2j+1$;
- *(iv) the m.d.e. (3) possesses a unique equilibrium in* F^+;

then ,with probability 1, $\{X^n\}$ *converges to the equilibrium of the m.d.e. (3) in* F^+.

Proof. For the proof we use the theorem 4.4 of [8]. Lemmas 4 – (iii) and 5 show that this theorem is applicable.Here $\Omega = \Omega^c = F^+$. We come to the conclusion that the equilibrium is asymptotically stable and F^+ is a subset of its domain of attraction . Now Lemma 4 – (iv) implies that the same is true for

ESANN'1998 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 22-23-24 April 1998, D-Facto public., ISBN 2-9600049-8-1, pp. 167-172

$\bar{F}^{+}$ and finally the Kushner-Clark theorem implies a.s. convergence of X^n to the equilibrium.

□

A natural question which now arises is that under which conditions the uniqueness of the fixed point is guaranteed. In [1] it is proved that each of the following conditions are sufficient for the uniqueness of the fixed point of $h(x)$ in F^{+}.

- $\log p(v)$ is strictly concave on (0,1),
- $\log p(v)$ is concave on (0,1) and $p(0_{+}) + p(1_{-}) > 0$.

In spite of these results, there are still unanswered questions. Firstly the class of stimuli, for which the uniqueness is proved, seems not to be the biggest one. Secondly, in many applications one deals with non-continuous density functions. In such a case the system (3) is not cooperative any more and hence its asymptotic behavior remains unclear. A possible approach to tackle this problem and even the multi-dimensional algorithm is to consider the convergence in distribution of the algorithm, instead of its almost sure convergence.

References

[1] Benaim M., Fort J. C., Pagés G., Convergence of the One-dimensional Kohonen algorithm. To appear in *Adv. in Appl. Prob.*

[2] Bouton C., Pagés G., Self-organization and a.s. convergence of the one-dimensional Kohonen algorithm with non-uniformly distributed stimuli. *Stochastic Process. Appl.* **47**, 1993, 249-274.

[3] Cottrell M., Fort J. C., Étude d'un processus d'auto-organisation. *Ann. Inst. Henri Poincaré* **23** ,1987, 1-20.

[4] Erwin E., Obermeyer K., Schulten K., Self-organizing maps: ordering, convergence properties and energy functions. *Biol. Cybern.* **67**,1992,47-55 .

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

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

[7] Fort J. C., Pagés G.,On the a.s. convergence of the Kohonen algorithm with a general neighborhood function.*The Ann. of Appl. Prob.* **4**,1995,1177-1216.

[8] Hirsch M. W., Stability and convergence in strongly monotone dynamical systems. *J. reine angew. Math.* **383**,1988, 1-53.

[9] 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).

[10] Sadeghi A.A., Self-Organization property of Kohonen's map with general type of stimuli distribution , preprint 181, Arbeitsgruppe Technomathematik, Universität Kaiserslautern, September 1997, Available on http://www.mathematik.uni-kl.de/