TA 81.

Asymptotic Behaviour of Self-Organizing Maps With Non-Uniform Stimuli Distribution

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

Abstract

Here the almost sure convergence of one dimensional Kohonen's algorithm in its general form, namely, $2k$ point neighbour setting with a non-uniform stimuli distribution is proved.We show that the asymptotic behaviour of the algorithm is governed by a cooperative system of differential equations which in general is irreducible. The system of differential equation has an asymptotically stable fixed point which a compact subset of its domain of attraction will be visited by the state variable X^n infinitely often.

Key words neural networks, stochastic approximation, theory of differential equations.

AMS subject classification 34D45, 82C32, 92B20, 93D20.

1 Introduction

Self-organizing maps play a crucial role in many functions of the nervous system as well as artificial intelligent tasks . Different sensory inputs , such as visual and acoustic inputs , are known to be mapped onto different areas of the cortex in an orderly, topology preserving manner, i.e., similar inputs are mapped onto neighbouring places in the cortex. These mappings are not genetically prespecified in detail but instead self-organize during the early stages of the formation of nervous system. For more details see [9] and references therein.

Among a number of algorithms which have been suggested for the formation of such mappings, Kohonen's algorithm is the most popular one. This algorithm has been employed successfully in a wide range of applications including speech recognition [8], robotics [14],

*This work is supported by German Academic Exchange Service .

computer vision [11] , etc. The adaptation of the weights in the Kohonen algorithm can be decomposed into two phases . In the first phase it self organizes a topology preserving map and then it converges to the final weights which are supposed to make a better representation of the input space. Depending on the nature of the application each of these phases may become more or less important. For example in numerical integration [12] the asymptotic behaviour of the algorithm is more important than its other features.

Let I be a set of n neurons labeled from 1 to N. The Kohonen net defined on I is a triple $\tau = (G, Q, F)$ where

- $G = (I, (V_i : i \in I))$ is a graph on I , in which $V_i \subset I$ is the set of all neurons connected to neuron i (its neighbours) such that
  1) $j \in V_i \Rightarrow i \in V_j$ for all $i, j \in I$,
  2) $i \in V_i$ for all $i \in I$.
- Q is the set of states of neurons which usually is a subset of R^m.Every neuron takes a weight vector $X_i \in Q$.
- $F = \{f_{ij} : i, j \in I\}$ is the set of neighbourhood functions, $f_{ij} : I \times I \rightarrow R$.

This network is used to build a mapping from R^m to the set of neurons which is usually arranged as a d-dimensional net. Every $v \in Q$ corresponds with the neurons $i^*(v)$ which satisfy

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

To be more accurate, this is a mapping from Q to the power set of I.

The weights X_i^n will be adapted in the learning phase according to

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

where $v \in Q$ is chosen randomly (according to some probability distribution $P(v)$) ,ϵ_n is the learning parameter and f_{i^*i} are the neighbourhood functions for i and i^*.

In this paper we consider the one dimensional Kohonen net. In this case every neuron i takes a value $X_i \in Q \subset R$ and $V_i = \{i-1, i, i+1\} \cap \{1, \cdots, N\}$. Moreover

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

in which $\gamma_0 = 1$, $0 < \gamma_k \leq \cdots \leq \gamma_1 \leq 1$ and $k \geq 0$. The initial weight vector $X^0 = \{X_1^0, \cdots, X_N^0\}$ can be chosen randomly from Q^N.

Although the exact definition of topology preservation involves much , cf. [15], it can be easily defined for a one dimensional map. A map is topology preserving iff it is ordered,i.e., either $X_i < X_j \Leftrightarrow i < j$ (ascending) or $X_i < X_j \Leftrightarrow i > j$ (descending), for all i, j. It is well known that once the one dimensional Kohonen map becomes ordered, it will preserve its ordering for ever,[9] .

Two special cases of this algorithm, namely $k = 0$ and $k = 1$, $\gamma_1 = 1$, have been investigated by Cottrell and Fort , [4], and Bouton and Pagès , [2, 3]. The results of these papers confirm the almost sure (a.s.) convergence of the algorithm for both cases, if the stimuli is distributed uniformly . In the non-uniform case it is shown that, if $Log(P(v))$ is strictly concave then the corresponding mean differential equation has an asymptotically stable equilibrium state, cf. [2] . However, it is not enough to ensure the a.s. convergence of the algorithm and no conclusion concerning the convergence to an equilibrium state has been achieved for stimulies which are distributed non-uniformly.

The ability of the algorithm to self-organize a topology preserving map has been proved for 2-point neighbour setting case, [4, 2]. Due to very many implementations of the algorithm it is believed that this result is valid generally for 2k-point neighbour setting.In this paper the asymptotic behaviour of the algorithm will be investigated, with the assumption that the ordering has been already established . We will prove the a.s. convergence of the 2k point neighbour algorithm for all continuous stimuli distributions $P(v)$ whose support set is [0,1].

In one dimensional case the adaptation process (2) may be considered as a stochastic dynamical system in R^N . Such time discrete systems have been treated by many authors in the stochastic approximation context, [1, 10]. A usual method to study the long time behaviour of these systems is to compare them with the so called mean differential equation , (m.d.e), which under certain conditions has the same asymptotic behaviour as the original discrete system. In Section 2 of this paper we formulate the problem as a Robbins-Monro algorithm and introduce the conditions under which the m.d.e. governs the long time behaviour of the algorithm. Section 3 is devoted to the properties of the m.d.e , in which it is shown that the stochastic variable X^n evolves in the domain of attraction of the equilibrium state of the m.d.e. . This enables us to establish the main result of this paper , namely a.s. convergence of the algorithm (Theorem 2) . In Section 4 we use the m.d.e. to investigate the effect of neighbourhood function f_{ij} on the final distribution of neuron .The concluding remarks are also contained in Section 4.

2 Robbins-Monro formulation of the algorithm

Throughout this paper we suppose that the order has been established as ascending.Two possible orderings, ascending or descending, are mathematically equivalent and all the results are valid for descending case as well. Moreover, we let $Q = [0, 1]$ and $X^n \in$

$[0,1]^N$ the vector of all weights at step n. The probability distribution $P(v)$ is always assumed to be regular enough such that $\mathbf{P}(0 \leq v \leq x)$, the probability of $v \in [0,x]$, is continuous in $[0,1]$. This is the case for all functions which are Lebesgue integrable in $[0,1]$.

The adaptation algorithm (2) may be rewritten in a more general form ,

$$X^{n+1} = X^n - \epsilon_n \eta_n X^n + \epsilon_n \eta_n I_{N\times 1} v_n, \tag{4}$$

where $\epsilon_n \in R$, $I_{N\times 1} = (1, \cdots, 1)^T$, $v_n \in [0,1]$ is an identically independent distributed (i.i.d.) random variable with a distribution $P(v)$, and $\eta_n = \eta(X^n, v) := R^N \times R \mapsto R^{N\times N}$ is a piecewise continuous function which associates with any pair $(X, v) \in R^N \times R$ the matrix $\eta_n = [\eta_n^{ij}]$,

$$\eta_n^{ij} = \begin{cases} 1, & \text{if} \quad i = j = i^*, \\ \gamma_1, & \text{if} \quad i = j = i^* - 1 \quad \text{or} \quad i = j = i^* + 1, \\ \vdots & \vdots \\ \gamma_k, & \text{if} \quad i = j = i^* - k \quad \text{or} \quad i = j = i^* + k, \\ 0, & \text{otherwise.} \end{cases}$$

Here i^* denotes the so called winner unit, which is defined by (1).

Remark 1. Although η_n is not an i.i.d. random variable, it still satisfies the relation

$$\mathbf{P}(\eta_n | \eta_{n-1}, \eta_{n-2}, \cdots, \eta_0; X^n, X^{n-1} \cdots X^0) = \mathbf{P}(\eta_n | X^n),$$

in which $\mathbf{P}(.|.)$ is the conditional probability function.

Let us adopt the following notations

$$\bar{X}_i^n := 0.5(X_i^n + X_{i-1}^n) \quad \text{for} \quad 1 < i \leq N \quad \text{and} \quad \bar{X}_1^n := 0, \bar{X}_{N+1}^n := 1 \quad \forall n \geq 1,$$

$$P_i(X^n) := \int_{\bar{X}_i^n}^{\bar{X}_{i+1}^n} P(v)dv, \quad Q_i(X^n) := \int_{\bar{X}_i^n}^{\bar{X}_{i+1}^n} vP(v)dv, \quad \forall \quad 1 \leq i \leq N,$$

$$R_i(X^n) := P_i(X^n) + (P_{i-1}(X^n) + P_{i+1}(X^n))\gamma_1 + \cdots + (P_{i-k}(X^n) + P_{i+k}(X^n))\gamma_k,$$

$$S_i(X^n) := Q_i(X^n) + (Q_{i-1}(X^n) + Q_{i+1}(X^n))\gamma_1 + \cdots + (Q_{i-k}(X^n) + Q_{i+k}(X^n))\gamma_k,$$

$$P_i(X^n) := Q_i(X^n) := 0, \qquad \text{if} \qquad i \leq 0 \quad \text{or} \quad i > N. \tag{5}$$

Introduce $h_n(X^n)$ for the expectation value of $(-\eta_n X^n + \eta_n I_{N\times 1} v_n)$,

$$h_n(X^n) = E(-\eta_n X^n + \eta_n I_{N\times 1} v_n) = \int_0^1 (-\eta_n X^n + \eta_n I_{n\times 1} v) p(v) dv, \qquad (6)$$

then

$$h_n(X^n) = \begin{bmatrix} -R_1(X^n)X_1^n + S_1(X^n) \\ \vdots \\ -R_i(X^n)X_i^n + S_i(X^n) \\ \vdots \\ -R_N(X^n)X_N^n + S_N(X^n) \end{bmatrix}. \qquad (7)$$

Since $h_n(.)$ is independent of n, we remove the index n and set $h(.) = h_n(.)$. Now define

$$\xi_n = -\eta_n X^n + \eta_n I_{N\times 1} v_n - h(X^n). \qquad (8)$$

ξ_n is a random variable with expectation value zero.

The recursive algorithm (4) can now be rewritten as

$$X^{n+1} = X^n + \epsilon_n h(X^n) + \epsilon_n \xi_n. \qquad (9)$$

This is a Robbins-Monro like algorithm which was originally suggested to find the roots of a function $h(x)$. The asymptotic behaviour of such algorithms has been studied by Kushner and Clark [10], from which we quote the following conditions and theorem which is a special case of theorem (2.3.1) in [10].

A.1. $h(.)$ is a continuous R^N valued function on R^N.

A.2. $\{\epsilon_n\}$ is a sequence of positive real numbers such that $\epsilon_n \to 0$ and $\Sigma\epsilon_n = \infty$.

A.3. ξ_n is a sequence of R^N valued random variables such that for some $T > 0$ and each $\epsilon > 0$

$$\lim_{x\to 0} \mathbf{P}\{\sup_{j\geq n} \max_{t\leq T} |\sigma_{i=m(jT)}^{m(jT+t)-1} \epsilon_i \xi_i| \geq \epsilon\} = 0.$$

Here $P(.)$ is the probability function and $m(.)$ is defined as

$$\begin{aligned} t_n &= \Sigma_{i=0}^{n-1} \epsilon_i, \\ m(t) &= \begin{cases} \max\{n : t_n \leq t\}, & t \geq 0, \\ 0, & t < 0. \end{cases} \end{aligned}$$

Define the function $X_n(.)$ by

$$\begin{aligned} X_0(t_n) &= X^n, \\ X_0(t) &= \frac{t_{n+1}-t}{\epsilon_n} X^n + \frac{t-t_n}{\epsilon_n} X^{n+1} \quad \text{for} \quad t_n \le t < t_{n+1}, \\ X_n(t) &= \begin{cases} X_0(t+t_n), & t \ge -t_n, \\ X^0, & t \le -t_n, \end{cases} \end{aligned}$$

now we are in a position to state the following theorem

Theorem 1. *(Kushner and Clark 1978) Let X^n be given by (9). Assume A.1 to A.3, and let X^n be bounded with probability 1. Then there is a null set Ω_0 such that $v \notin \Omega_0$ implies that $\{X_n(.)\}$ is equicontinuous, and also that the limit $X(.)$ of any convergent subsequence of $\{X_n(.)\}$ is bounded and satisfies the system of differential equations*

$$\dot{x} = h(x) \tag{10}$$

on the time interval $(-\infty, \infty)$. Let x_0 be a locally asymptotically stable (in the sense of Liapunov) solution to (10) ,with domain of attraction $DA(x_0)$. Then if $v \notin \Omega_0$ and there is a compact set $A \subset DA(x_0)$ such that $X^n \in A$ infinitely often, we have $X^n \to x_0$ as $n \to \infty$.

Scheme of proof. Define the functions $\bar{X}_0(.)$ and $M_0(.)$ by

$$\begin{aligned} \bar{X}_0(t) &= X^n \quad \text{for} \quad t_n \le t < t_{n+1}, \\ M_0(t_n) &= \sum_{i=0}^{n-1} \epsilon_i \xi_n, \\ M_0(t) &= \frac{t_{n+1}-t}{\epsilon_n} M_0(t_n) + \frac{t-t_n}{\epsilon_n} M_0(t_{n+1}), \quad \text{for} \quad t_{n-1} \le t < t_n, \\ M_n(t) &= \begin{cases} M_0(t+t_n) - M_0(t_n), & t \ge -t_n, \\ -M_0(t_n), & t \le -t_n, \end{cases} \end{aligned}$$

we may now write

$$X_n(t) = X_n(0) + \int_0^t h(\bar{X}_0(t_n+s))ds + M_n(t).$$

Set

$$\int_0^t h(\bar{X}_0(t_n+s))ds = \int_0^t h(X_n(s))ds + \delta^n(t),$$

then we have

$$X_n(t) = X_n(0) + \int_0^t h(X_n(s))ds + \delta^n(t) + M_n(t).$$

If $n \to \infty$ then conditions A-1 to A-3 imply $\delta^n(t) \to 0$ and $M_n(t) \to 0$. $\{X_n(t)\}$ is equicontinuous and bounded , and using the Arzelà-Ascoli theorem it contains converging subsequences . Now let $X_{n_i}(t)$ be a converging subsequence of $X_n(t)$. Then as $n \to \infty$, $X_{n_i}(t)$ converges to a solution of $\dot{x} = h(x)$.

□

In the rest of this section ,we show that the Kohonen algorithm ,as formulated in (9) ,satisfies the conditions A.1 and A.3 , provided that $X_1^n \leq X_2^n \leq \cdots \leq X_N^n$.

Lemma 1. *Suppose* $\{\epsilon_n\}$ *is a sequence of positive real numbers such that* $\Sigma_n \epsilon_n = \infty$ *and* $\Sigma_n \epsilon_n^2 < \infty$. *Then ,with probability 1 , any convergent subsequence of* $\{X^n\}$ *converges to a solution of the system of differential equations*

$$\dot{x} = \begin{bmatrix} -R_1(x)x_1 + S_1(x) \\ \vdots \\ -R_i(x)x_i + S_i(x) \\ \vdots \\ -R_N(x)x_N + S_N(x) \end{bmatrix}. \qquad (11)$$

Let x_0 *be a locally asymptotically stable (in the sense of Liapunov) solution to (11) ,with domain of attraction* $DA(x_0)$. *Then if there is a compact set* $A \subset DA(x_0)$ *such that* $X^n \in A$ *infinitely often,with probability 1 we have* $X^n \to x_0$ *as* $n \to \infty$.

Proof. The functions $P_i(X^n)$ and $Q_i(X^n)$, $1 \leq i \leq N$, are continuous , i.e. , $h(.)$ is a continuous function of its argument .

For A.3., it suffices to note that the variance of ξ_n is uniformly bounded and moreover its conditional expectation value satisfy

$$E[\xi_n | X^0, X^1, \cdots, X^n] = E[\xi_n | X^n] = 0.$$

Then, $\{\sum_{i=0}^n \epsilon_i \xi_i\}$ is a matriangle sequence and if $\sum_{i=0}^\infty \epsilon_i^2 < \infty$, A.3. is fulfiled, cf. [10, pp. 26-27].

The boundedness condition on $\{X^n\}$ is fulfiled automatically by Kohonen's algorithm.

□

3 Stability analysis of the mean differential equation

In this section our ultimate aim is to show that the system (11) has an asymptotically stable equilibrium, which a compact subset of its domain of attraction will be visited by the stochastic process X^n infinitely often.

Consider the system of differential equations

$$\dot{x} = f(x), \qquad x \in \Omega \subset R^m, \qquad f : \Omega \mapsto R^m. \tag{12}$$

The following terminology will be used throughout this section: The system (12) is cooperative if f is continuously differentiable and

$$\frac{\partial f_i}{\partial x_j} \geq 0 \qquad \text{for all} \qquad j \neq i.$$

An $m \times m$ matrix $A = [a_{ij}]$ is irreducible if it does not map any nonzero proper linear subspace of R^m into itself. A necessary and sufficient condition for A to be irreducible is that for any $\alpha, \beta \in \mathcal{N}$, $1 \leq \alpha, \beta \leq m$ there exists a chain of integers $\alpha = \alpha_0, \alpha_1, \cdots, \alpha_k = \beta$ such that $a_{\alpha_{i-1}\alpha_i} \neq 0$ for all $0 < i \leq k$, see e.g. [5]. The system (12) is irreducible if the Jacobian matrix $J(x) = [\partial f_i / \partial x_j(x)]$ is irreducible for all $x \in \Omega$.

For vectors $x, y \in R^m$ we write $x \leq y$ if $x_i \leq y_i$ for all i. A set $\Omega \subset R^m$ is p-convex whenever $x, y \in \Omega$ and $x \leq y$, then Ω contains the entire line segment joining x and y.

Let $\Phi_x(t)$ denotes the solution of (12) which satisfies $\Phi_x(0) = x$. The matrix $D\Phi_x(t)$ is the spatial derivative of $\Phi_x(t)$. We say the flow Φ has positive derivatives if $D\Phi_x(t) > 0$ for all $t > 0$ and all $x \in \Omega$. It has eventually positive derivatives if there exists a $t_0 > 0$ such that $D\Phi_x(t) > 0$ for all $t \geq t_0$ and all $x \in \Omega$. Clearly if Φ has positive derivatives , it has also eventually positive derivatives. Φ is strongly monotone provided $x \leq y$ implies $\Phi_x(t) < \Phi_y(t)$ for all $t > 0$ and $x, y \in \Omega$ such that $x \neq y$.

A set $F \subset R^m$ is called positive invariant, if for all $x \in F$ and all $t \geq 0$ for which $\Phi_x(t)$ is defined, $\Phi_x(t) \subset F$. $\bar{F}$ denotes the closure of F.

If any neighbourhood of a point x_0 contains a point $x_\epsilon \in \Phi_x(t)$ then x_0 is said to be a limit point of $\Phi_x(t)$. The ω-limit set of a solution $\Phi_x(t)$ is the set of points $p \in \Omega$ such that $\Phi_x(t_k) \to p$ for some sequence $t_k \to \infty$

An equilibrium is a point x^* for which $f(x^*) = 0$. E is the set of all equilibria in Ω. The equilibrium x^* is simple if zero is not an eigenvalue of the jacobian $J(x^*)$. It is a sink if all eigenvalues of the jacobian have negative real parts. If there exists an open set Γ, not necessarily containing x^*, such that for all $x \in \Gamma$, $\Phi_t(x)$ converges to x^* uniformly as $t \to \infty$, then x^* is called a trap. A simple trap is known to be a sink.

The asymptotic behaviour of cooperative irreducible systems of differential equations has been investigated by Hirsch [6, 7]. The advantage of Hirsch's method is that to find the domain of attraction of an equilibrium no Liapunov function is needed. Moreover the asymptotic stability of the equilibrium is guaranteed if there is a unique equilibrium or if E is countable and all equilibria are simple. In fact the simplicity of all equilibria assures the existence of some $x^* \in E$ such that the jacobian $J(x^*)$ has all its eigenvalues in the negative complex half-plane. For the convenience of the reader we recall the following important results. In the sequel Ω is an open p-convex subset of R^m.

Lemma 2. *(Hirsch 1985) Let f be a cooperative irreducible vector field on the open p-convex set $\Omega \subset R^m$. Then*

a) Φ has positive derivatives ,

b) Φ is strongly monotone .

The lemmas 3 and 4 are valid if $\Phi_x(t)$ has eventually positive derivatives. $\Omega^c \subset \Omega$ is a set of points x whose corresponding flow $\Phi_x(t)$ has a compact closure in Ω .

Lemma 3. *(Hirsch 1985) There is a set $Q \subset \Omega^c$ having Lebesgue measure zero, such that $\Phi_x(t)$ approaches the equilibrium set E as $t \to \infty$, for all $x \in \Omega^c \backslash Q$.*

Lemma 4. *(Hirsch 1985)*

a) Assume E is countable . Then $\Phi_x(t)$ converges to a trap as $t \to \infty$. for almost all $x \in \Omega^c$.

b) Assume all equilibria are simple . Then $\Phi_x(t)$ converges to a sink as $t \to \infty$, for almost all $x \in \Omega^c$.

Lemma 5. *(Hirsch 1988) Assume that Φ is strongly monotone and $f(x)$ has a unique equilibrium $p \in \Omega^c$. Then $\Phi_x(t) \to p$ for all $x \in \Omega$.*

For the proofs see [6, 7].

Now we apply the above mentioned lemmas to the mean differential equation (11). As a first step let us consider the existence of an equilibrium. A result similar to the next lemma was first established by Bouton and Pagès for $k = 0$ and $k = 1$, $\gamma_1 = 1$ in [2]. Here we modify the argument to generalize it to $1 \leq k \leq N$ and $\gamma_k \leq \cdots \leq \gamma_1 \leq 1$. In the rest of this paper $\gamma_j = 0$ for $j > k$.

Lemma 6. *Consider the set* $F^+ = \{x \in [0,1]^N | \quad 0 < x_1 < x_2 < \ldots < x_N < 1\}$,

a) if $P(v)$ *is Lebesgue integrable on [0,1] ,then there exists a* $x^* \in \bar{F}^+$ *such that* $h(x^*) = 0$.

b) If supp $P(v) = [0,1]$, $\gamma_{j+1} < \gamma_j$, $N \geq 2j+1$ *for some* j, $0 \leq j \leq k$, *then* $x^* \in F^+$.

Proof. a) $\bar{F}^+$ is a compact subset of R^N. So using the Brower theorem, see e.g. [13], it is sufficient to show that $x + h(x)$ maps $\bar{F}^+$ into itself continuously. For all $v \in [0,1]$, $x - \eta_n x + \eta_n I_{N\times 1} v$ maps $\bar{F}^+$ into $\bar{F}^+$. It ensures that its mean value, i.e. $x + h(x)$, maps $\bar{F}^+$ into $\bar{F}^+$ as well. The continuity condition was shown in Section 2.

b) For notational convenience we define $Z_{i,j} = Q_i(x^*) - x_j^* P_i(x^*)$, then

$$Z_{i,j} \geq 0 \quad \text{for} \quad i > j \quad \text{and} \quad Z_{i,j} \leq 0 \quad \text{for} \quad i < j.$$

If $x_1^* = 0$ then $h_1(x^*) = 0$ implies $S_1(x^*) = 0$, but this means that $Q_1(x^*) = \cdots = Q_{k+1}(x^*) = 0$, i.e., $x_2^* = \cdots = x_{k+2}^* = 0$. With the same argument $x_i^* = 0$, for all $1 \leq i \leq N$, which implies $h_N(x^*) = \int_0^1 xP(x)dx = 0$ and this contradicts the diffusity of $P(x)$.

Suppose $x_p^* = x_{p-1}^*$, $x_p^* < x_{p+1}^* < \cdots < x_N^* < 1$ for some p, $1 < p \leq N$. Then we have $Z_{i,p} = Z_{i,p-1}$,

$$\begin{aligned} h_p(x^*) &= Z_{p,p} + \gamma_1(Z_{p-1,p} + Z_{p+1,p}) + \cdots + \gamma_k(Z_{p-k,p} + Z_{p+k,p}) = 0, \\ h_{p-1}(x^*) &= Z_{p-1,p-1} + \gamma_1(Z_{p-2,p-1} + Z_{p,p-1}) + \cdots + \gamma_k(Z_{p+k-1,p-1} + Z_{p-k-1,p-1}) = 0, \end{aligned}$$

and

$$h_p(x^*) - h_{p-1}(x^*) = \Sigma_{i=0}^k(\gamma_i - \gamma_{i+1})Z_{p+i,p} + \Sigma_{i=1}^{k+1}(\gamma_i - \gamma_{i-1})Z_{p-i,p} = 0.$$

Under the assumptions, each of the contributions in this equation is non-negative , so all of them have to vanish.

If $p \leq N-j$ then $p+j \in \{i, \cdots, N\}$ and $(\gamma_j - \gamma_{j+1})Z_{p+j,p} > 0$ which is a contradiction.

Now let $p > N-j$. Then $p-j-1 \in \{1, \cdots, N\}$ and this implies $Z_{p-j-1,p} = 0$, i.e. , $x_{p-j-1}^* = x_{p-j}^*$. Now $p-j \leq N-j$, which implies $Z_{p,p-j} = 0$. If $p < N$ this implies $x_p^* = x_{p+1}^*$ which is a contradiction and if $p = N$, then it has to be $x_N^* = 1$, which again is a contradiction.

The only case which remains is $x_{p-1}^* < x_p^* = x_{p+1}^* = \cdots = x_N^* = 1$, $1 \leq p \leq N$. $h_p(x^*) = 0$ implies

$$Z_{p,p} + \gamma_1 Z_{p-1,p} + \cdots + \gamma_k Z_{p-k,p} = 0. \tag{13}$$

All the contributions in (13) are non-positive , so all of them ought to vanish.But $Z_{p,p} < 0$ which is again a contradiction.

□

It is a known result that F^+ is an absorbing set for the Kohonen algorithm, cf. [9].The following lemma shows that this property is preserved by the m.d.e. (11). We will use this result later on to establish the a.s. convergence of the algorithm.

Lemma 7. *Suppose* $1 \geq \gamma_1 \cdots \geq \gamma_k > 0$. *Then the following statements are valid,*

a) F^+ *is a positive invariant set for (11),*

b) For all $x \in F^+$, *if* $\gamma_{j+1} < \gamma_j$ *and* $N \geq 2j+1$ *for some* $0 \leq j \leq k$, *then* $\Phi_x(t)$ *has a compact closure in* F^+,

c) For all $x \in \bar{F}^+ \backslash F^+$ *if* $\gamma_{j+1} < \gamma_j$ *and* $N \geq 2j+1$ *for some* $0 \leq j \leq k$, *then* $\Phi_x(t) \in F^+$ *for* $t > 0$.

Proof. For both *a)* and *b)* It suffices to show that if $x \in F^+$ then $\Phi_x(t)$ has no limit point on $\bar{F}^+ \backslash F^+$.

Suppose $x^* = (0, \cdots, 0, x^*_{p+1}, \cdots, x^*_N)$ is a limit point of $\Phi_x(t)$ for some p , $1 \leq p \leq N$, and $0 < x^*_{p+1} \leq \cdots \leq x^*_N \leq 1$. Set

$$S(x^*, \epsilon) = \{x \in F^+ | \, x^* - \epsilon < x < x^* + \epsilon\}. \tag{14}$$

For any sufficiently small ϵ , if $\Phi_x(t)$ enters $S(x^*, \epsilon)$ then we have

$$\frac{dx_p}{dt} = -R_p(x)x_p + S_p(x) > 0. \tag{15}$$

This implies that $x_p(t)$ can not approach zero, i.e. $\Phi_x(t)$ has no limit point on the hyperplane $x_1 = x_2 = \cdots = x_p = 0$ for any $1 \leq p \leq N$.

Now let $x^* = (x^*_1, \cdots, x^*_{p-1}, x^*_p, \cdots, x^*_N)$, $1 < p \leq N$, $x^*_{p-1} = x^*_p$, and $x^*_p < \cdots < x^*_N < 1$, be a limit point of $\Phi_x(t)$. We have

$$\begin{aligned} \frac{d(x_p - x_{p-1})}{dt} &= (x_{p-1} - x^*_{p-1})R_{p-1}(x) - (x_p - x^*_p)R_p(x) \\ &+ \Sigma_{i=1}^{k+1}(\gamma_{i-1} - \gamma_i)(x^*_p P_{p-i}(x) - Q_{p-i}(x)) + \Sigma_{i=0}^{k}(\gamma_{i+1} - \gamma_i)(x^*_p P_{p+i}(x) - Q_{p+i}(x)). \end{aligned} \tag{16}$$

For sufficiently small ϵ, if $\Phi_x(t)$ enters $S(x^*,\epsilon)$, then apart from $(x_{p-1}-x^*_{p-1})R_{p-1}(x)$ and $(x_p - x^*_p)R_p(x)$, all the contributions in (16) are non-negative. Because ϵ can be chosen arbitrarily small , either $\frac{d(x_p-x_{p-1})}{dt} > 0$ or all these terms ought to vanish. So we have to show that there exists always non vanishing terms.

If $1 \le p \le N-j$ then $p+j \in \{1,\cdots,N\}$ and hence $(\gamma_{j+1}-\gamma_j)(x^*_pP_{p+j} - Q_{p+j}) > 0$.

For $p > N-j$ we have $p-j-1 \in \{1,\cdots,N\}$. If

$$(\gamma_j - \gamma_{j+1})(x^*_pP_{p-j-1} - Q_{p-j-1}) = 0,$$

then it has to hold $x^*_{p-j-1} = x^*_{p-j}$. Now $p-j \le N-j$ which implies $(\gamma_{j+1}-\gamma_j)(x^*_{p-j}P_p - Q_p) = 0$. If $p < N$ this implies $x^*_p = x^*_{p+1}$ which contradicts the assumption $x^*_p < x^*_{p+1}$. For $p = N$ this implies $x^*_p = 1$ which contradicts the assumption $x^*_p < 1$.

Finally let $x^* \in (x^*_1,\cdots,x^*_{p-1},x^*_p,\cdots,x^*_N)$, $x^*_p = \cdots = x^*_N = 1$, $x^*_p > x^*_{p-1}, 1 \le p \le N$. If for some $t > 0$, $\Phi_t(x) \in S(x^*,\epsilon)$, then for sufficiently small ϵ we have

$$\frac{dx_p}{dt} = \Sigma^k_{i=0}\gamma_i(Q_{p-i}(x) - x_pP_{p-i}(x)) < 0, \tag{17}$$

which implies that x_p dose not reach 1 and x^* can not be a limit point of $\Phi_t(x)$. The assertion c) can be proved by letting $x^* \in \bar{F}^+\backslash F^+$ at time $t=0$. Then by an analogous argument as *b*) it comes that for some $\delta > 0$ and $0 < t < \delta$, $\Phi_t(x) \in F^+$.

The assertion c) can be proved by letting $x^* \in \bar{F}^+\backslash F^+$ at time $t=0$. Then by an analogous argument as *b*) it comes that for some $\delta > 0$ and $0 < t < \delta$, $\Phi_t(x) \in F^+$.

□

The following lemma shows that there is a strong relation between the system (11) and the theory of cooperative irreducible differential equations.

Lemma 8. *Suppose* $P(v)$ *is continuous in* $[0,1]$, $supp\,P(v) = [0,1]$ *and* $\gamma_k \le \gamma_{k-1} \le \cdots \le \gamma_1 \le 1$, *then we have*

a) (11) is cooperative in $\bar{F}^+$.

b) If $\gamma_{j+1} < \gamma_j$, *for some* $0 \le j \le k$ *and* $N \ge 2j+1$, *then (11) is irreducible and all its equilibria are simple in* F^+,

Proof. The Jacobian of $h(x)$ is

$$J(x) = \begin{bmatrix}
a_{1,1} & \cdots & a_{1,k+1} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \cdots & \cdots & \vdots \\
a_{k+2,1} & \cdots & a_{k+2,k+2} & \cdots & a_{k+2,2k+3} & \ddots & & & \vdots \\
0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 & & \vdots \\
\vdots & \ddots & a_{i,i-k-1} & \cdots & a_{i,i} & \cdots & a_{i,i+k+1} & \ddots & \vdots \\
\vdots & & 0 & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\
\vdots & & & \ddots & a_{N-k-1,N-2k-2} & \cdots & a_{N-k-1,N-k-1} & \cdots & a_{N-k-1,N} \\
\vdots & \cdots & 0 & & \ddots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & \cdots & \cdots & \cdots & 0 & a_{N-k-1,N} & \cdots & a_{N,N}
\end{bmatrix},$$

$$\begin{aligned}
a_{i,i} &= -(P_i(X) + \gamma_1(P_{i-1}(X) + P_{i+1}(X)) + \cdots + \gamma_k(P_{i-k}(X) + P_{i+k}(X))) \\
&\quad +0.25(1-\gamma_1)(x_{i+1} - x_i)P(\bar{x}_{i+1}) + 0.25(1-\gamma_1)(x_i - x_{i-1})P(\bar{x}_i), \\
a_{i,i+p} &= 0.5(\gamma_{|p|} - \gamma_{|p-1|})(x_i - \bar{x}_{i+p})P(\bar{x}_{i+p}) + 0.5(\gamma_{|p+1|} - \gamma_{|p|})(x_i - \bar{x}_{i+p+1})P(\bar{x}_{i+p+1}).
\end{aligned}$$

Here $1 \leq i+p \leq N$, $p \neq 0$. Clearly all the off-diagonal elements of the jacobian matrix are non-negative in $\bar{F}^+$.

Inside F^+ the condition $\gamma_{j+1} < \gamma_j$ implies that for any i , $1 \leq i \leq N$, if

$$i+j, i+j+1,\ i-j, i-j-1 \in \{1, \cdots, N\}$$

then

$$a_{i,i+j} > 0,\ a_{i,i+j+1} > 0,\ a_{i,i-j} > 0,\ a_{i,i-j-1} > 0\ .$$

The condition $N \geq 2j+1$ implies that at least for one i it holds $\{i+j, i-j\} \subset \{1, \cdots, N\}$. This assures the irreducibility of $J(x)$.

Moreover $rank\ J(x) = N$, which in turn ensures zero is no eigenvalue of this matrix.

□

Now we are in the position to state the main result of this paper. As before, here it is supposed that the order has been already established as ascending .

Theorem 2. *Assume the following conditions hold :*

a) $\{\epsilon_n\}$ *is a sequence of positive real numbers such that* $\Sigma_n \epsilon_n^2 < \infty$ *and* $\Sigma_n \epsilon_n = \infty$,

b) $P(v)$ *is continuous in* $[0,1]$ *and supp* $P(v) = [0,1]$,

c) (11) has a unique equilibrium in F^+,

d) $\gamma_{j+1} < \gamma_j$ *for some* $0 \le j \le k$ *and* $N \ge 2j+1$,

then ,with probability 1 , $\{X^n\}$ *converges to an asymptotically stable equilibrium of (11) in* F^+.

Remark 2. If X^n is ordered in a descending manner then theorem 2 is valid, but the system (11) changes its form as following:

$$P_i(X^n) = \int_{\bar{X}_{i+1}^n}^{\bar{X}_i^n} P(v)dv, \qquad Q_i(X^n) = \int_{\bar{X}_{i+1}^n}^{\bar{X}_i^n} vP(v)dv, \qquad \forall \quad 1 \le i \le N. \tag{18}$$

Proof. Lemmas 7 and 8 show that the Lemma 5 is applicable.Here $\Omega = \Omega^c = F^+$. We conclude that $\bar{F}^+$ is a subset of the domain of attraction of the eqilibrium and the lemma 1 implies the almost sure convergence of the algorithm to it.

□

Remark 3. If equation (11) has more than one equilibrium, then the lemmas 2-4 imply that there exists a set $Q \subset F^+$ with Lebesgue measure zero such that $\Phi_x(t)$ approaches an asymptotically stable equilibrium for all $x \in \bar{F}^+ \backslash Q$. This together with lemma 1 implies that if $\{X^n\}$ has a limit point $x_0 \in \bar{F}^+ \backslash Q$, then x_0 is a sink of (11) , which in turn implies that a compact subset of $DA(x_0)$, the domain of attraction of x_0 , is visited infinitely often by $\{X^n\}$, i.e. $\{X^n\}$ converges to x_0 with probability one.

4 Conclusion

In this paper we have established the a.s. convergence of one dimensional kohonen's algorithm for a fairly large class of stimuli distributions and neighbourhood functions.

The conditions imposed on the learn parameters are necessary to ensure the a.s. convergence. In fact many numerical experiments show that if ϵ_n decrease too fast, such that it violates the condition a) in theorem 2 , then the algorithm may get stock in a non-optimum equilibrium. It can be shown that a better rate of convergence can be reached by letting $\sum_n \epsilon_n^2$ converge as fast as possible. On the other hand the conditions imposed on the stimuli distribution cover the applications of the Kohonen algorithm.

The symmetric neighbourhood function (3) is very often used in applications. Unsymmetric neighbourhood functions ,although practically seem to be not interesting, also could be introduced. In this paper we did not use the symmetric property of this function and hence the results are valid for nonsymmetric functions, as well.

An interesting issue concerning the final mappings produced by Kohonen's algorithm is the effect of the stimuli distribution $P(v)$ and the neighborhood parameters $\gamma_1, \cdots, \gamma_k$ on the final position of neurons in Q. Using the differential equation (11) we are able to find the possible final values of neurons. The final values for $N = 20$, $k = 1$, $P(v) = 1,\ 2v,\ 6v(v-1)$ and $\gamma_1 = 0.00,\ 0.25,\ 0.50,\ 0.75,\ 1.00$ are depicted in figures 1-3. The results show that there exists two kinds of final mappings produced by the algorithm corresponding to $\gamma_1 = 0$ or $\gamma_1 > 0$.

For $\gamma_1 > 0$, the final mapping is not sensitive to different values of γ_1. By the way, γ_1 may be used for a better rate of convergence. In this case the final mapping makes a better representation of stimuli distribution $P(v)$ than $\gamma_1 = 0$, although even for $\gamma_1 = 0$ the final values clearly change their behaviour respective to different choices of $P(v)$.

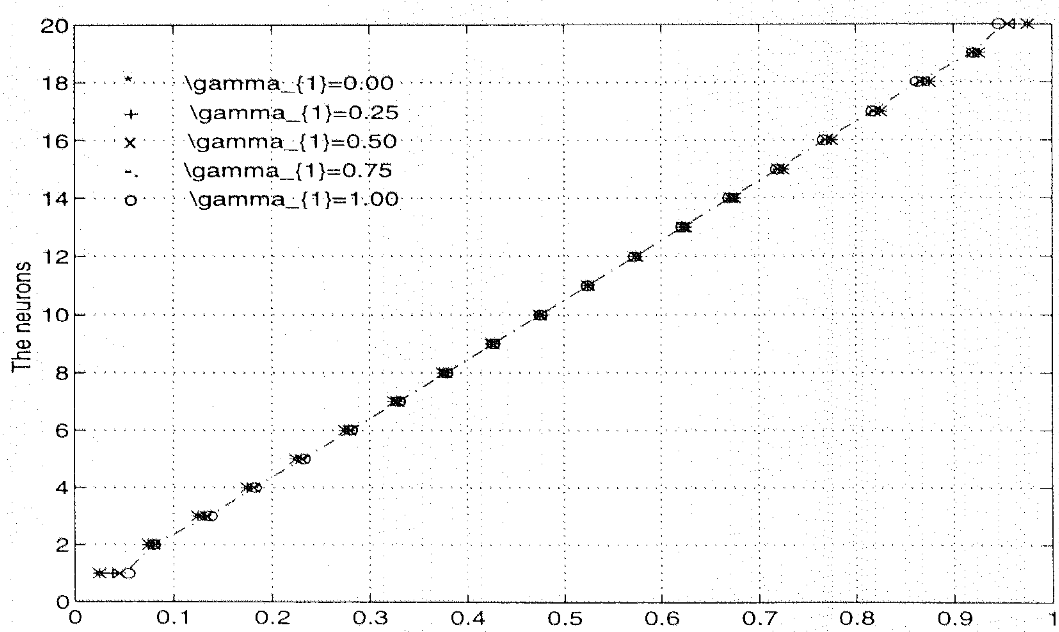

Fig. 1. The final values of neurons for P(x)=1 in [0,1]

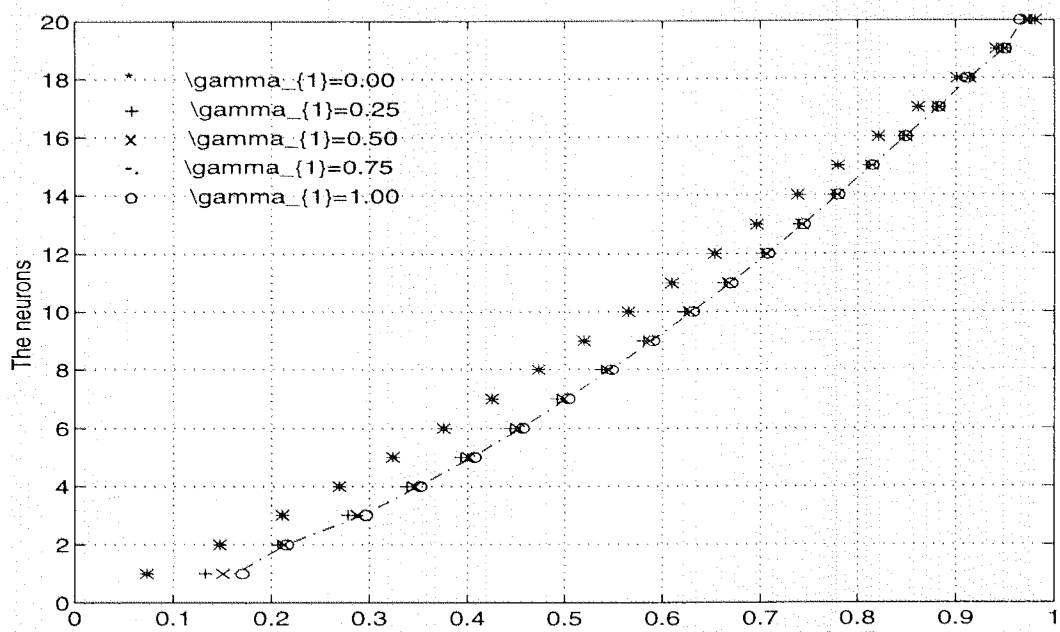

Fig. 2. The final values of neurons for P(x)=2x in [0,1]

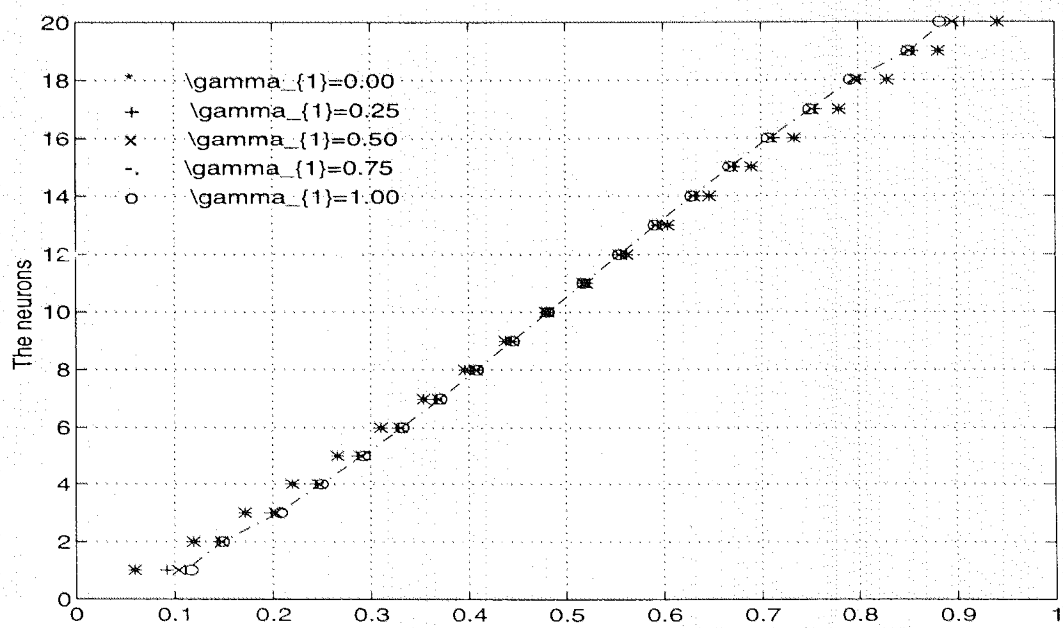

Fig. 3. The final values of neurons for P(x)=6x(x-1) in [0,1]

References

[1] A. Benveniste and M. Métivier and P. Priouret, *Adaptive algorithms and stochastic approximations* , Springer-Verlag, 1990.

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

[3] C. Bouton and G. Pagés,*Convergence in distribution of the one-dimensional Kohonen algorithm when the stimuli are not uniform*,Adv. Appl. Prob. 26(1994),pp. 80-103 .

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

[5] W. Hackbusch,*Theorie und numerik elliptischer differentialgleichungen*, Teubner , Stuttgart, 1986.

[6] M. W. Hirsch ,*Systems of differential equations that are competitive or cooperative 2: convergence almost everywhere* , Siam J. Math. Anal. , 16 (1985),pp. 423-439 .

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

[8] T. Kohonen, *Speech recognition based on topology preserving neural maps* ,in Neural Computation, I.Alexander ,ed., Kogan Page,London,1989.

[9] T. Kohonen , *Self-organizing maps* , Springer-Verlag ,1995.

[10] H.J.Kushner and D.S.Clark, *Stochastic approximation methods for constrained and unconstrained systems*, Springer-Verlag , 1978.

[11] E. Oja ,*Self-organizing maps and computer vision*, in Neural Networks for Perception, H.Wechsler ,ed. ,vol. 1, Academic Press, Boston.

[12] G. Pagés, *Voronoi tesselation ,space quantization algorithms and numerical integration* , Proc. of the ESANN93 Conference , Brussels ,D Facto ed. ,1993,pp. 221-228.

[13] M. Reed and B. Simon ,*Methods of modern mathematical physics,vol. 1, Functional analysis*, Academic Press,1980.

[14] H. Ritter and T. Martinetz and K. Schulten, *Topology conserving maps for motor control*,Neural Networks , from Models to Applications, IDSET,Paris, 1989.

[15] T. Villmann and R. Der and M. Herrmann and T. Martinetz, *Topology preservation in SOFMs : general definition and efficient measurement.*,Informatik aktuell Fuzzy-Logik,Springer,1994,pp. 159-166.