

## ASYMPTOTIC BEHAVIOR OF SELF-ORGANIZING MAPS WITH NONUNIFORM STIMULI DISTRIBUTION

BY ALI A. SADEGHI

*Universität Kaiserslautern*

Here the almost sure convergence of one-dimensional Kohonen's algorithm in its general form, namely, the  $2k$ -neighbor setting with a nonuniform stimuli distribution, is proved. We show that the asymptotic behavior of the algorithm is governed by a cooperative system of differential equations which is irreducible. The system of differential equations possesses an asymptotically stable equilibrium, a compact subset of whose domain of attraction will be visited by the state variable  $X^n$  infinitely often. The assumptions on the stimuli distribution and the neighborhood functions are weakened, too.

**1. Introduction.** The Kohonen self-organizing map (SOM) was originally devised to model a process of self-organization between different sensory inputs, such as visual and acoustic inputs, and different areas of the cortex. The algorithm has the ability to construct a structured representation of data from an input space, possibly with a reduction of the dimensionality. Meanwhile, it has found widespread use in many artificial intelligence applications. The areas of application include, for example, speech recognition [13], robotics [19], computer vision [16], and so forth.

The maps generated by SOM algorithm are characterized by the fact that similar inputs are mapped onto the neighboring neurons. Due to this characteristic, the map is said to be topology preserving. However, an exact preservation of the neighborhood relations is possible only if the input space and the network have the same dimensionalities.

The adaptation of the weights in this 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 quantization of the input space. Depending on the nature of the application, each of these phases may become more or less important. For instance, in numerical integration, [17], the asymptotic behavior of the algorithm plays a more important role than its other features.

Let  $I$  be a finite set of neurons labelled from 1 to  $N$ . The Kohonen net defined on  $I$  is a triple  $\tau = (V, Q, F)$  where  $V = (V_i; i \in I)$  is a graph on  $I$  in which  $V_i \subset I$  is the set of all neurons connected to neuron  $i$  (its neighbors) such that we have the following:

1.  $i \in V_i$  for all  $i \in I$ ;
2.  $j \in V_i \Rightarrow i \in V_j$  for all  $i, j \in I$ .

---

Received February 1997; revised May 1997.

AMS 1991 subject classifications. Primary 60J05; secondary 93D20, 92B20.

Key words and phrases. Neural networks, stochastic approximation, theory of differential equations.

Then  $Q$  is the set of states of neurons, which usually is a subset of  $\mathbb{R}^m$ . Every neuron  $i$  takes a weight vector  $X_i \in Q$ .

$F = \{f_{ij}: i, j \in I\}$  is the set of neighborhood functions;  $f_{ij}: I \times I \rightarrow [0, 1]$ .

This network is used for building up a projection from  $\mathbb{R}^m$  to the set of neurons, which is usually arranged as a  $d$ -dimensional net. Every  $v \in Q$  corresponds with the winner neuron  $i^*(v)$ , which satisfy

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

where  $\|\cdot\|$  denotes the Euclidean norm on  $\mathbb{R}^m$ . In the case where there is more than one possible choice of winner, a predefined rule is used to choose it. Here we choose the neuron with the smallest index as the winner.

The weights  $X_i^n$  are adapted in a learning phase according to

$$(2) \quad X_i^{n+1} = X_i^n + \varepsilon_n f_{i^*i}(v - X_i^n) \quad \forall i \in I, \quad n = 0, 1, \dots,$$

where  $v \in Q$  is chosen at random (according to some probability distribution  $P$ ),  $\varepsilon_n \in (0, 1)$  is the learning parameter and  $f_{i^*i}$  are the neighborhood functions for  $i$  and  $i^*$ . The initial weight vector  $X^0 = (X_1^0, \dots, X_N^0)^T$  can be chosen at random from  $Q^N$ .

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

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

where  $\gamma_0 = 1$ ,  $0 < \gamma_k \leq \dots \leq \gamma_1 \leq 1$  and  $k \geq 0$ . Throughout the paper we use the convention  $\gamma_i = 0$  for  $i > k$ . The neighborhood function  $f_{i^*i}$ , as defined by (3), includes all the neighborhood functions used in applications (see, e.g., [14, 19]).

Although an exact definition of topology preservation for a  $d$ -dimensional network involves a lot of technicalities (cf. [8, 20]), it can be easily defined in the one-dimensional case. A one-dimensional map is topology preserving iff it is ordered, that is, either  $X_i < X_j \Leftrightarrow i < j$  (ascending) or  $X_i < X_j \Leftrightarrow i > j$  (descending), for all  $i, j$ . Let  $F^+ := \{x \in [0, 1]^N | 0 < x_1 < \dots < x_N < 1\}$  and  $F^- := \{x \in [0, 1]^N | 0 < x_N < \dots < x_1 < 1\}$ , then  $F := F^+ \cup F^-$  is the set of all ordered states. It is well known that once the one-dimensional Kohonen map becomes ordered, it retains its ordering forever [14].

The adaptation process (2), in the one-dimensional case, may be considered as a stochastic dynamical system in  $\mathbb{R}^N$ . Such time-discrete systems have been treated by many authors in the stochastic approximation context [1, 15].

A common method used to study the long-time behavior of these systems is to compare them with the so-called mean differential equation, (m.d.e), which under certain conditions behaves asymptotically the same as the original discrete system. This method will be applied in this paper to establish almost sure (a.s.) convergence of the algorithm.

The behavior of the one-dimensional Kohonen's algorithm after self organization has been investigated by Cottrell and Fort [4], Bouton and Pagès [2, 3] and Fort and Pagès [7]. Their results confirm a.s. convergence of the algorithm to an equilibrium point of the corresponding m.d.e., if the stimuli is distributed uniformly and  $\varepsilon_n$  converges to zero slowly enough.

In the nonuniform case it is shown that, under suitable assumptions, the m.d.e. owns an asymptotically stable equilibrium point (cf. [7]). However, this result is not enough to ensure a.s. convergence of the algorithm and no conclusion has been drawn for the case where stimuli is distributed non-uniformly.

The major difficulty, which prevents a generalization of this result to the nonuniform case, is that no Liapunov function is known for the m.d.e., in this case. This strongly suggests a study of the m.d.e. using new stability results published during the last decade; see [11, 12] and references therein. These results take advantage of the general properties, which a system possibly enjoys, to determine the attracting area of asymptotically stable solutions.

In this paper we use the cooperative and irreducible character of the m.d.e. to prove that  $X^n$  converges almost surely (or with probability one), if the m.d.e. owns a unique equilibrium in  $\bar{F}^+$ . While retaining the restrictions on  $\varepsilon_n$ , we weaken the conditions imposed on  $f_{ij}$ .

For the case that m.d.e. owns a countable number of equilibria, we show, under certain conditions, almost all  $x \in \bar{F}^+$  belong to the attracting area of the asymptotically stable equilibria. This is an extension of the convergence results formerly established [2, 3, 7].

The outline of the paper is as follows. Section 2 begins with the formulation of the problem as a Robbins–Monro algorithm, and then we introduce conditions under which the m.d.e. governs long-time behavior of the algorithm. Section 3 is devoted to the properties of the m.d.e, where it is shown that the m.d.e. is irreducibly cooperative in  $F^+$  and  $\bar{F}^+$  is positively invariant. This enables us to establish the main result of the paper, namely, the a.s. convergence of the algorithm. In Section 4 we use the m.d.e. to investigate effects of the neighborhood function  $f_{ij}$  on final distribution of neurons. Some concluding remarks are also contained in Section 4.

**2. Robbins–Monro formulation of the algorithm.** In the rest of this paper, we always assume that the initial value  $X^0$  of the algorithm lies in  $F^+$ . Two possible orderings, ascending ( $X^0 \in F^+$ ) or descending ( $X^0 \in F^-$ ), are mathematically equivalent, and all results which we present here are valid for the descending case, as well. Moreover, we set  $Q = [0, 1]$ . The probability distribution  $P$  is always assumed to be diffuse, that is, for  $A \subset [0, 1]$  if  $A$  has Lebesgue measure zero, then  $P(A) = 0$  and  $\text{supp } P = [0, 1]$ .

The adaptation rule (2) may be rewritten in a more general form:

$$(4) \quad X^{n+1} = X^n - \varepsilon_n \eta_n X^n + \varepsilon_n \eta_n I_{N \times 1} v_n,$$

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

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

Here  $i^*$  denotes the so-called winner unit which was defined by (1).

REMARK 1. Although  $\eta_n$  is not an i.i.d. random variable, it satisfies the relation

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

where  $\mathbf{P}(\cdot | \cdot)$  is the conditional probability function.

Let us adopt the following notations:

$$(5) \quad \begin{aligned} \bar{X}_i^n &:= 0.5(X_i^n + X_{i-1}^n) \quad \text{for } 1 < i \leq N \text{ and} \\ \bar{X}_1^n &:= 0, \quad \bar{X}_{N+1}^n := 1 \quad \forall n \geq 1, \\ P_i(X^n) &:= P([\bar{X}_i^n, \bar{X}_{i+1}^n]), \\ Q_i(X^n) &:= \int_{\bar{X}_i^n}^{\bar{X}_{i+1}^n} v P(dv) \quad \forall 1 \leq i \leq N, \\ R_i(X^n) &:= P_i(X^n) + (P_{i-1}(X^n) + P_{i+1}(X^n))\gamma_1 \\ &\quad + \dots + (P_{i-k}(X^n) + P_{i+k}(X^n))\gamma_k, \\ S_i(X^n) &:= Q_i(X^n) + (Q_{i-1}(X^n) + Q_{i+1}(X^n))\gamma_1 \\ &\quad + \dots + (Q_{i-k}(X^n) + Q_{i+k}(X^n))\gamma_k, \\ P_i(X^n) &:= Q_i(X^n) := 0 \quad \text{if } i \leq 0 \text{ or } i > N. \end{aligned}$$

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

$$(6) \quad h_n(X^n) := E(-\eta_n X^n + \eta_n I_{N \times 1} v_n) = \int_0^1 (-\eta_n X^n + \eta_n I_{N \times 1} v) P(dv);$$

then a straightforward computation yields

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

Since for any finite  $N$  the function  $h_n(\cdot)$  is independent of  $n$ , we remove the index  $n$  and set  $h(\cdot) := h_n(\cdot)$ . The diffuseness of  $P$  implies that  $h(\cdot)$  is continuous on  $F^+$  and has a continuous extension on  $\bar{F}^+$  (the closure of  $F^+$ ).

For  $x \in \mathbb{R}^N \setminus F^+$ , we define  $h(x)$  by the right-hand side of (7). So  $h(\cdot): \mathbb{R}^N \mapsto \mathbb{R}^N$  is well defined and continuous. We emphasize that the function  $h(\cdot)$  as defined here coincides with the mean value of the algorithm only on  $\bar{F}^+$ . In fact the mean value function is not continuous on  $[0, 1]^N$ .

Now define

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

$\xi_n$  is a random variable with  $E \xi_n = 0$ .

The recursive algorithm (4) can now be rewritten as

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

This is a Robbins–Monro-like algorithm, which was originally proposed for finding the roots of a continuous function  $h(x)$ . It is remarkable that the strongest kind of convergence which can be expected for algorithms like (9) is, in fact, a.s. convergence.

The asymptotic behavior of such algorithms has been studied by Kushner and Clark [15]. Their well-known result provides conditions for a.s. convergence of the algorithm. For reference let us quote the following conditions:

- A.1.  $h(\cdot)$  is a continuous  $\mathbb{R}^N$ -valued function on  $\mathbb{R}^N$ ;
- A.2.  $\{\varepsilon_n\}$  is a sequence of positive real numbers such that  $\varepsilon_n \rightarrow 0$  and  $\sum \varepsilon_n = \infty$ ;
- A.3.  $\xi_n$  is a sequence of  $\mathbb{R}^N$ -valued random variables such that for some  $T > 0$  and each  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \mathbf{P} \left( \sup_{j \geq n} \max_{t \leq T} \left| \sum_{i=m(jT)}^{m(jT+t)-1} \varepsilon_i \xi_i \right| \geq \varepsilon \right) = 0.$$

Here  $\mathbf{P}(\cdot)$  is probability function and  $m(\cdot)$  is defined by

$$t_n := \sum_{i=0}^{n-1} \varepsilon_i,$$

$$m(t) := \begin{cases} \max\{n: t_n \leq t\}, & t \geq 0, \\ 0, & t < 0. \end{cases}$$

Define the function  $X_n(\cdot)$  by

$$(10) \quad \begin{aligned} X_0(t_n) &:= X^n, \\ X_0(t) &:= \frac{t_{n+1}-t}{\varepsilon_n} X^n + \frac{t-t_n}{\varepsilon_n} X^{n+1} \quad \text{for } t_n \leq t < t_{n+1}, \\ X_n(t) &:= \begin{cases} X_0(t+t_n), & t \geq -t_n, \\ X^0, & t \leq -t_n. \end{cases} \end{aligned}$$

Now we are in a position to state Theorem 1, which is a special case of Theorem (2.3.1) in [15] and later on will be used to establish a.s. convergence of  $X^n$ .

**THEOREM 1** (Kushner and Clark [15]). *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  $\Omega_0$  such that  $v \notin \Omega_0$  implies that  $\{X_n(\cdot)\}$  is equicontinuous and also that the limit  $X(\cdot)$  of any convergent subsequence of  $\{X_n(\cdot)\}$  is bounded and satisfies the system of differential equations*

$$(11) \quad \dot{x} = h(x)$$

*on the time interval  $(-\infty, \infty)$ . 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  $v \notin \Omega_0$  and there is a compact set  $A \subset DA(x_0)$  such that  $X^n \in A$  infinitely often, we have  $X^n \rightarrow x_0$  as  $n \rightarrow \infty$ .*

**SCHEME OF PROOF.** Define the functions  $\bar{X}_0(\cdot)$  and  $M_n(\cdot)$  by

$$\begin{aligned} \bar{X}_0(t) &:= X^n \quad \text{for } t_n \leq t < t_{n+1}, \\ M_0(t_n) &:= \sum_{i=0}^{n-1} \varepsilon_i \xi_n, \\ M_0(t) &:= \frac{t_{n+1}-t}{\varepsilon_n} M_0(t_n) + \frac{t-t_n}{\varepsilon_n} M_0(t_{n+1}) \quad \text{for } t_{n-1} \leq t < t_n, \\ M_n(t) &:= \begin{cases} M_0(t+t_n) - M_0(t_n), & t \geq -t_n, \\ -M_0(t_n), & t \leq -t_n. \end{cases} \end{aligned}$$

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  $n \rightarrow \infty$ , then conditions A.1 to A.3 imply  $\delta^n(t) \rightarrow 0$  and  $M_n(t) \rightarrow 0$ .  $\{X_n(t)\}$  is equicontinuous and bounded; hence, 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  $n \rightarrow \infty$ ,  $X_{n_i}(t)$  converges to a solution of  $\dot{x} = h(x)$ .  $\square$

In the rest of this section we show that the Kohonen algorithm, as formulated by (9), satisfies conditions A.1 and A.3.

LEMMA 1. Suppose  $\text{supp } P = [0, 1]$ ,  $P$  is diffuse,  $\varepsilon_n > 0$  for all  $n$ ,  $\sum_n \varepsilon_n = \infty$  and  $\sum_n \varepsilon_n^2 < \infty$ . If  $X^0 \in F^+$ , then, with probability 1, any convergent subsequence of  $\{X_n\}$ , as defined by (1), (2) and (10), converges to a solution of the system of differential equations

$$(12) \quad \dot{x} = \begin{bmatrix} -R_1(x)x_1 + S_1(x) \\ \vdots \\ -R_i(x)x_i + S_i(x) \\ \vdots \\ -R_N(x)x_N + S_N(x) \end{bmatrix}.$$

Let  $x_0$  be a locally asymptotically stable (in the sense of Liapunov) solution to (12), 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  $X^n \rightarrow x_0$  as  $n \rightarrow \infty$ .

PROOF. For  $X^0 \in F^+$ , the algorithm (1) and (2) coincides with (8) and (9), so it suffices to prove the lemma for (8) and (9).

The continuity of  $h(\cdot)$  on  $\mathbb{R}^N$  is clear by its definition.

For A.3, it suffices to note that the variance of  $\xi_n$  is uniformly bounded and moreover its conditional expectation value satisfies

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

Thus  $\{\sum_{i=0}^n \varepsilon_i \xi_i\}$  is a martingale sequence and if  $\sum_{i=0}^{\infty} \varepsilon_i^2 < \infty$ , then A.3. is fulfilled; see [15], pages 26 and 27 for more details.

The boundedness condition is fulfilled automatically by the algorithm (1) and (2).  $\square$

With regard to Lemma 1, the key point for the investigation of the asymptotic behavior of one-dimensional Kohonen's algorithm is to find out about the asymptotically stable solutions of (12) in  $\bar{F}^+$  and their domains of attraction. This is the point which we are going to address in the rest of the paper.

**3. Stability analysis of the mean differential equation.** In this section, our ultimate goal is to show that under certain conditions, the m.d.e. (12) owns an asymptotically stable equilibrium, whose domain of attraction includes  $\bar{F}^+$ . But before going into technicalities, let us say that the approach

basically consists of showing that the m.d.e. (12) is cooperative and irreducible and  $\bar{F}^+$  is positively invariant for the flow of (12). An extended treatment of these ideas can be found in [11, 12] and references therein. However, in order to use the results of the theory of irreducible cooperative differential equations, we have to present the following short introduction to the most important ideas of this theory.

Consider a system of differential equations

$$(13) \quad \dot{x} = f(x), \quad x \in \Omega \subset \mathbb{R}^m, \quad f: \Omega \mapsto \mathbb{R}^m.$$

The system (13) is cooperative if  $f$  is continuously differentiable and

$$\frac{\partial f_i}{\partial x_j} \geq 0 \quad \text{for all } j \neq i \text{ and } x \in \Omega.$$

An  $m \times m$  matrix  $A = [a_{ij}]$  is said to be irreducible if it does not map any nonzero proper linear subspace of  $\mathbb{R}^m$  into itself.

There are two different criteria in the literature for actually showing that the matrix  $A = [a_{ij}]$  is irreducible:  $A = [a_{ij}]$  is irreducible if and only if the directed graph with vertices  $\{1, \dots, m\}$  and directed edges  $(i, j)$  for  $a_{ij} \neq 0$  is connected by directed paths (it is the criterion used in [11, 12], the results of which will be used here), or  $A = [a_{ij}]$  is irreducible if and only if there does not exist a permutation matrix  $P$  such that

$$(14) \quad PAP^T = \begin{bmatrix} B & C \\ 0 & D \end{bmatrix},$$

with  $B_{r \times r}$  and  $D_{N-r \times N-r}$  for  $1 \leq r \leq N-1$ . A permutation matrix is a matrix with exactly one nonzero entry, namely 1, in each row and each column.

It is a classical problem to show that the two criteria are equivalent (see, e.g., [9, 10]).

The system (13) is irreducible if the Jacobian matrix  $J(x) = [\partial f_i / \partial x_j(x)]_{1 \leq i, j \leq m}$  is irreducible for all  $x \in \Omega$ .

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

Let  $\Phi_x(t)$  and  $D\Phi_x(t)$  denote the solution to (13) with  $\Phi_x(0) = x$  and its Jacobian at time  $t$ , respectively. We say the flow  $\Phi$  has positive derivatives if  $D\Phi_x(t)$  has only positive entries for all  $t > 0$  and  $x \in \Omega$ .  $\Phi$  is strongly monotone in  $\Omega$  provided  $x \leq y$  implies  $\Phi_x(t) < \Phi_y(t)$  for all  $t > 0$  and  $x, y \in \Omega$  when  $x \neq y$ . For  $x \in \Omega$  we also write  $x(t) := (x_1(t), \dots, x_N(t))$  for  $\Phi_x(t)$ .

A set  $F \subset \mathbb{R}^m$  is called positively invariant if for all  $x \in F$  and all  $t \geq 0$ , for which  $\Phi_x(t)$  is defined,  $\Phi_x(t) \in F$ . By  $\bar{F}$  and  $\partial F$  we denote the closure and the boundary of  $F$  ( $\bar{F} \setminus F$  if  $F$  is open), respectively.

If any neighborhood of a point  $x_0$  contains a point  $x_\varepsilon \in \{\Phi_x(t)\}$ , then  $x_0$  is said to be a limit point of  $\Phi_x(t)$ .



An equilibrium is a point  $x^*$  for which  $f(x^*) = 0$ ;  $E$  is the set of equilibria in  $\Omega$ . The equilibrium  $x^*$  is simple if the Jacobian  $J(x^*)$  is invertible. It is a sink if all eigenvalues of the Jacobian have negative real parts.

The asymptotic behavior of cooperative irreducible systems of differential equations has been investigated by Hirsch [11, 12]. The advantage of Hirsch's method is that in order to find the attracting area of an equilibrium, no Liapunov function is needed. Moreover, the existence of an asymptotically stable equilibrium is guaranteed if  $E$  includes a unique point or if  $E$  is countable and all equilibria are simple. For the convenience of the reader, we recall the following important results. In the sequel  $\Omega$  is an open  $p$ -convex subset of  $\mathbb{R}^m$ .

LEMMA 2 (Hirsch [11]). *Let  $f$  be a cooperative irreducible vector field on the open  $p$ -convex set  $\Omega \subset \mathbb{R}^m$ . Then we have the following:*

- (i)  $\Phi$  has positive derivatives;
- (ii)  $\Phi$  is strongly monotone.

Theorem 2 establishes conditions under which  $\Phi_x(t)$  converges to a sink for almost all  $x \in \Omega^c$ . The term  $\Omega^c \subset \Omega$  is a set of points  $x \in \Omega$  whose corresponding solution  $\Phi_x(t)$  ( $t \geq 0$ ) has a compact closure in  $\Omega$ .

THEOREM 2 (Hirsch [11]). *Assume  $\Phi$  has positive derivatives,  $E$  is countable and all the equilibria are simple. Then  $\Phi_x(t)$  converges to a sink as  $t \rightarrow \infty$ , for almost all  $x \in \Omega^c$ .*

In Theorem 3, which extends this result to the whole domain  $\Omega^c$ , it is assumed that  $\Omega^c$  is open.

THEOREM 3 (Hirsch [12]). *Assume that  $\Phi$  is strongly monotone and  $f(x)$  has a unique equilibrium  $p \in \Omega^c$ . Then  $\Phi_x(t) \rightarrow p$  for all  $x \in \Omega^c$ .*

For the proofs see [11], Theorem 4.4, and [12], Theorem 10.3.

Next we apply the above-mentioned results to the m.d.e. (12).

As a first step let us consider the existence of an equilibrium. A result similar to Lemma 3 was first established by Bouton and Pagès [2] for  $k = 0$  and  $k = 1$ ,  $\gamma_1 = 1$ . Here we modify the argument to generalize it to  $0 \leq k \leq N$  and  $\gamma_k \leq \dots \leq \gamma_1 \leq 1$ . In the rest of this paper,  $\gamma_i = 0$  for  $i > k$ ,  $x_0 := 0$ ,  $x_{N+1} := 1$ .

LEMMA 3. *Consider the set  $F^+ := \{x \in [0, 1]^N \mid 0 < x_1 < x_2 < \dots < x_N < 1\}$ .*

- (i) *If  $P$  is diffuse, then there exists a  $x^* \in \bar{F}^+$  such that  $h(x^*) = 0$ .*
- (ii) *If  $P$  is diffuse,  $\text{supp } P = [0, 1]$ ,  $\gamma_{j+1} < \gamma_j$  and  $N \geq 2j + 1$  for some  $j$ ,  $0 \leq j \leq k$ ; then  $x^* \in F^+$ .*

PROOF. (i)  $\bar{F}^+$  is a compact subset of  $\mathbb{R}^N$ . So using the Brouwer theorem (see, e.g., [18]), 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  $\bar{F}^+$ . This ensures that its

mean value, that is  $x + h(x)$ , maps  $\bar{F}^+$  into  $\bar{F}^+$ , as well. The continuity of  $h(x)$  was shown in Section 2.

(ii) For notational convenience let us define

$$Z_{i,j} := Q_i(x^*) - x_j^* P_i(x^*) := \int_{\bar{x}_i^*}^{\bar{x}_{i+1}^*} x P(dx) - x_j^* \int_{\bar{x}_i^*}^{\bar{x}_{i+1}^*} P(dx).$$

It is clear that  $Z_{i,j} \geq 0$  for  $i > j$  and  $Z_{i,j} \leq 0$  for  $i < j$ .

If  $x_1^* = 0$ , then  $h_1(x^*) = S_1(x^*) = 0$ . Now (5) yields  $Q_1(x^*) = \dots = Q_{k+1}(x^*) = 0$  and thus  $x_2^* = \dots = x_{k+2}^* = 0$ . With the same argument  $x_i^* = 0$ , for all  $1 \leq i \leq N$ , which implies  $h_N(x^*) = \int_0^1 x P(dx) = 0$  and this contradicts the assumptions.

Suppose  $x_p^* = x_{p-1}^*$ ,  $x_p^* < x_{p+1}^* < \dots < x_N^* < 1$  for some  $p$ ,  $1 < p \leq N$ . Now consider

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

the equality  $x_p^* = x_{p-1}^*$  implies  $Z_{i,p} = Z_{i,p-1}$  and

$$(15) \quad h_p(x^*) - h_{p-1}(x^*) = \sum_{i=0}^k (\gamma_i - \gamma_{i+1}) Z_{p+i,p} + \sum_{i=1}^{k+1} (\gamma_i - \gamma_{i-1}) Z_{p-i,p} = 0.$$

Under the assumptions, each of the contributions is nonnegative, so all of them vanish.

For  $p \leq N - j$ , we have  $p + j \in \{1, \dots, N\}$  and  $(\gamma_j - \gamma_{j+1}) Z_{p+j,p} = 0$ , which contradicts the assumption  $x_p^* < x_{p+1}^* < \dots < x_N^* < 1$ .

Now let  $p > N - j$ . Then  $p - j - 1 \in \{1, \dots, N\}$ ; this implies  $Z_{p-j-1,p} = 0$ , which yields  $x_{p-j-1}^* = x_{p-j}^*$ . This means (15) with  $p - j$  instead of  $p$  is valid and thus  $Z_{p,p-j} = 0$ . For  $p < N$  this assures  $x_p^* = x_{p+1}^*$  and for  $p = N$  we have  $x_N^* = 1$ ; both of which contradict the assumption  $x_p^* < x_{p+1}^* < \dots < x_N^* < 1$ .

The only remaining case is  $x_{p-1}^* < x_p^* = x_{p+1}^* = \dots = x_N^* = 1$ ,  $1 \leq p \leq N$ . In this case we have  $Z_{p,p} < 0$ . Now  $h_p(x^*) = 0$  implies

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

None of the contributions is positive, so this case is also impossible.  $\square$

It is a known result that  $F^+$  is an absorbing set for the Kohonen algorithm (cf. [14]). Lemma 4 shows that this property is preserved by the m.d.e. (12). Later on we will use this result to establish the a.s. convergence of  $X^n$ . First, let us recall that  $h(\cdot)$  is a continuous function defined by the right-hand side of (12) on  $\mathbb{R}^N$  and coincides with the mean function (6) on  $\bar{F}^+$ .

LEMMA 4. Suppose  $P$  is diffuse,  $\text{supp } P = [0, 1]$  and  $\gamma_k \leq \gamma_{k-1} \leq \dots \leq \gamma_1 \leq 1$ . The following statements are valid:

- (i) If  $\gamma_{j+1} < \gamma_j$  and  $N \geq 2j + 1$  for some  $0 \leq j \leq k$ , then  $F^+$  is positively invariant.
- (ii) 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  $F^+$ .
- (iii) For all  $x \in \partial F^+$  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. The diffuseness and boundedness of  $P$  implies that  $h(\cdot)$  is Lipschitz continuous on  $\mathbb{R}^N$ . This, in turn, implies the existence of a unique solution  $\Phi_x(t)$  starting from any point  $x \in \mathbb{R}^N$  at time  $t = 0$ . Moreover  $\Phi_x(t)$  is a continuous function of  $t$ . Consequently, for  $x \in F^+$  and  $t > 0$  if  $\Phi_x(t)$  leaves  $F^+$ , then it has a limit point on  $\partial F^+$ . So for both (i) and (ii) it suffices to show that if  $x \in F^+$ ,  $\Phi_x(t)$  has no limit point on  $\partial F^+$ ; that is, for any arbitrary sequence  $t_n$ ,  $\Phi_x(t_n) \not\rightarrow \partial F^+$  as  $n \rightarrow \infty$ , provided that  $\Phi_x(t_n)$  is defined.

For  $x \in F^+$  and  $\varepsilon > 0$ , we introduce

$$S(x, \varepsilon) = \{y \in F^+ \mid x_i - \varepsilon < y_i < x_i + \varepsilon \ \forall i\}.$$

Suppose  $x^* = (0, \dots, 0, x_{p+1}^*, \dots, x_N^*)$  is a limit point of  $\Phi_x(t)$ ,  $0 < x_{p+1}^* \leq \dots \leq x_N^* \leq 1$  and  $1 \leq p \leq N$ .

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

$$(16) \quad \frac{dx_p(t)}{dt} = -R_p(x(t))x_p(t) + S_p(x(t)) > 0.$$

Now let  $\Phi_x(t_k) \notin S(x^*, \varepsilon/2)$ . The continuity of  $\Phi_x(t)$  together with the inequality (16) imply that  $\Phi_x(t) \notin S(x^*, \varepsilon/2)$  for all  $t > t_k$  [as soon as  $\Phi_x(t) \in S(x^*, \varepsilon) \setminus S(x^*, \varepsilon/2)$ ,  $x_p(t)$  increases and hence  $\Phi_x(t)$  never reaches  $S(x^*, \varepsilon/2)$ ]. On the other hand, if  $\Phi_x(t_k) \in S(x^*, \varepsilon/2)$ , then there exists a time interval  $[t_k, t_k + \delta t]$  in which  $x_p(t)$  is increasing and  $\Phi_x(t_k + \delta t) \notin S(x^*, \varepsilon/2)$ . This assures that there does not exist any sequence  $t_n$  such that  $\Phi_x(t_n) \rightarrow x^*$  (this includes also the case  $t_n - t_{n-1} \rightarrow \infty$ ); that is,  $\Phi_x(t)$  has no limit point on the hyperplane  $x_1 = x_2 = \dots = x_p = 0$  for any  $1 \leq p \leq N$ .

Next let  $x^* = (x_1^*, \dots, x_{p-1}^*, x_p^*, \dots, x_N^*)$  be a limit point of  $\Phi_x(t)$  for some  $1 < p \leq N$ ,  $x_{p-1}^* = x_p^*$  and  $x_p^* < \dots < x_N^* < 1$ . We have

$$(17) \quad \begin{aligned} \frac{d(x_p(t) - x_{p-1}(t))}{dt} &= (x_{p-1}(t) - x_{p-1}^*)R_{p-1}(x(t)) - (x_p(t) - x_p^*)R_p(x(t)) \\ &\quad + \sum_{i=1}^{k+1} (\gamma_{i-1} - \gamma_i)(x_p^*P_{p-i}(x(t)) - Q_{p-i}(x(t))) \\ &\quad + \sum_{i=0}^k (\gamma_{i+1} - \gamma_i)(x_p^*P_{p+i}(x(t)) - Q_{p+i}(x(t))). \end{aligned}$$

For sufficiently small  $\varepsilon$ , if  $\Phi_x(t)$  enters  $S(x^*, \varepsilon)$ , then apart from  $(x_{p-1}(t) - x_{p-1}^*)R_{p-1}(x(t))$  and  $(x_p(t) - x_p^*)R_p(x(t))$ , all the terms are nonnegatives. Since  $\varepsilon$  can be chosen arbitrarily small, either for some  $\varepsilon > 0$  we have  $(d/dt)(x_p(t) - x_{p-1}(t)) > 0$  for  $x(t) \in S(x^*, \varepsilon)$  or all these terms vanish as  $\varepsilon \rightarrow 0$ .

If  $1 \leq p \leq N - j$ , then  $p + j \in \{1, \dots, N\}$ ; hence  $x_p^* < \dots < x_N^* < 1$  implies that there exists a  $\varepsilon > 0$  and a  $\eta \in \mathbb{R}_+$  such that  $(\gamma_{j+1} - \gamma_j)(x_p^* P_{p+j}(x(t)) - Q_{p+j}(x(t))) > \eta > 0$  for  $x(t) \in S(x^*, \varepsilon)$ .

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

$$(\gamma_j - \gamma_{j+1})(x_p^* P_{p-j-1}(x(t)) - Q_{p-j-1}(x(t))) \rightarrow 0,$$

then  $x_{p-j-1}^* = x_{p-j}^*$ . Now equation (17) with  $p - j$  instead of  $p$  is valid and the assumption  $x_p^* < \dots < x_N^* < 1$  implies  $(\gamma_{j+1} - \gamma_j)(x_{p-j}^* P_p(x(t)) - Q_p(x(t))) > \eta > 0$  for some  $\varepsilon > 0$  and  $\eta \in \mathbb{R}_+$ , which in turn assures  $(d/dt)(x_{p-j}(t) - x_{p-j-1}(t)) > 0$ . So for the case  $p > N - j$  we have either  $(d/dt)(x_p(t) - x_{p-1}(t)) > 0$  or  $x_{p-j-1}^* = x_{p-j}^*$  and  $(d/dt)(x_{p-j}(t) - x_{p-j-1}(t)) > 0$  on  $S(\varepsilon, x^*)$ .

It remains to show that if  $x_i^* = x_{i-1}^*$  for some  $i \in \{2, \dots, N\}$ , then the inequality  $(d/dt)(x_i - x_{i-1}) > 0$  in  $S(\varepsilon, x^*)$  implies  $\Phi_x(t_n) \not\rightarrow x^*$  for any arbitrary sequence  $t_n$ . This can be done in exactly the same way as the case  $x^* = (0, \dots, 0, x_{p+1}, \dots, x_N)$ ; if  $\Phi_x(t_n) \notin S(x^*, \varepsilon/2)$  then  $\Phi_x(t) \notin S(x^*, \varepsilon/2)$  for any  $t > t_n$  and if  $\Phi_x(t_n) \in S(x^*, \varepsilon/2)$  then, in a time interval  $[t_n, t_n + \delta t]$ ,  $(x_p - x_{p-1})$  increases and  $\Phi_x(t_n + \delta t) \notin S(x^*, \varepsilon/2)$ .

Finally let  $x^* = (x_1^*, \dots, x_{p-1}^*, x_p^*, \dots, x_N^*)$  be a limit point of  $\Phi_x(t)$ ,  $x_p^* = \dots = x_N^* = 1$ ,  $x_p^* > x_{p-1}^*$  and  $1 \leq p \leq N$ . If  $\Phi_x(t) \in S(x^*, \varepsilon)$  for some  $t > 0$  and  $\varepsilon$  is sufficiently small, then we have

$$\frac{dx_p}{dt} = \sum_{i=0}^k \gamma_i (Q_{p-i}(x(t)) - x_p P_{p-i}(x(t))) < 0,$$

with the same argument as in the last cases; this implies that  $x_p$  does not reach 1 and thus  $x^*$  can not be a limit point of  $\Phi_x(t)$ .

For assertion (iii) let  $x \in \partial F^+$ . Then by an argument analogous to (i) and (ii), it follows that, for some  $\delta > 0$  and  $0 < t < \delta$ ,  $\Phi_x(t) \in F^+$ .  $\square$

Lemma 5 shows that there is a strong relation between the m.d.e (12) and the theory of cooperative irreducible differential equations.

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

- (i) *The m.d.e. (12) 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. (12) is irreducible on  $F^+$ .*

PROOF. For  $x \in F^+$ , the Jacobian of  $h(x)$  reads as follows:

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

where

$$\begin{aligned} (18) \quad a_{i,i} &= -(P_i(x) + \gamma_1(P_{i-1}(x) + P_{i+1}(x)) \\ &\quad + \cdots + \gamma_k(P_{i-k}(x) + P_{i+k}(x))) \\ &\quad + 0.25(1 - \gamma_1)(x_{i+1} - x_i)p(\bar{x}_{i+1})\mathbf{1}_{\{i+1 \leq N\}} \\ &\quad + 0.25(1 - \gamma_1)(x_i - x_{i-1})p(\bar{x}_i), \\ a_{i,i+q} &= 0.5(\gamma_{|q|} - \gamma_{|q-1|})(x_i - \bar{x}_{i+q})p(\bar{x}_{i+q})\mathbf{1}_{\{1 < i+q \leq N\}} \\ &\quad + 0.5(\gamma_{|q+1|} - \gamma_{|q|})(x_i - \bar{x}_{i+q+1})p(\bar{x}_{i+q+1})\mathbf{1}_{\{i+q+1 \leq N\}}. \end{aligned}$$

Here  $1 \leq i + q \leq N$ ,  $q \neq 0$ . Clearly all the off-diagonal elements are nonnegative for  $x \in F^+$ .

To prove that  $J(x)$  is irreducible for all  $x \in F^+$ , let us first consider the case  $j > 0$  ( $\gamma_1 = 1$ ). In this case  $a_{i,i} < 0$  for all  $i$ .

Now consider any set  $\tau_s \subset \{1, \dots, N\}$ , where  $s$  denotes the number of indices in  $\tau_s$  and let  $\nu_i$  be the  $i$ th row of  $J(x)$ .

Since  $\gamma_{j+1} < \gamma_j$ , we have

$$\text{if } 1 \leq i \leq N - j - 1, \text{ then } a_{i,i+j} > 0 \text{ and } a_{i,i+j+1} > 0,$$

$$\text{if } j + 1 < i \leq N, \text{ then } a_{i,i-j} > 0 \text{ and } a_{i,i-j-1} > 0.$$

The condition  $N > 2j + 1$  implies that in every row there exists at least two nonzero off-diagonal entries and for any  $\tau_s$  the  $s \times N$  matrix  $[\nu_i]_{i \in \tau_s}$  has at most  $N - s - 1$  columns with all entries equal zero.

Now assume there exists a permutation matrix  $P = [P_{kl}]$ , which transforms  $J(x)$  into the form (14) for some  $1 \leq r \leq N$  and let  $PJP^T = [b_{kl}]_{1 \leq k, l \leq N}$ . Consider the function  $s(\cdot): \{1, \dots, N\} \rightarrow \{1, \dots, N\}$ , which satisfies  $P_{ks(k)} = 1$  [for any given  $P$  the function  $s(\cdot)$  is uniquely defined]. Then we have  $b_{kl} = a_{s(k)s(l)}$  and consequently  $a_{s(k)s(l)} = 0$  for  $r < k \leq N$  and  $1 \leq l \leq r$ . This means that there must be a set  $\tau_{N-r}$  such that  $[\nu_i]_{i \in \tau_{N-r}}$  has  $r$  columns with all entries equal zero, which is impossible.

For the case  $j = 0$ , we have  $a_{i,i+1} > 0$  and  $a_{i,i-1} > 0$  provided they are defined. Thus for any  $k, l \in \{1, \dots, N\}$  if  $l > k$ , then  $a_{k,k+1} > 0, a_{k+1,k+2} > 0, \dots, a_{l-1,l} > 0$  and  $a_{l,l-1} > 0, a_{l-1,l-2} > 0, \dots, a_{k+1,k} > 0$ . Furthermore for any  $l \in \{1, \dots, N\}$  we have either  $a_{l,l-1} > 0, a_{l-1,l} > 0$  or  $a_{l,l+1} > 0, a_{l+1,l} > 0$ , which completes the proof.  $\square$

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

**THEOREM 4.** *Assume the following conditions hold:*

- (i)  $\varepsilon_n > 0 \forall n, \sum_n \varepsilon_n^2 < \infty$  and  $\sum_n \varepsilon_n = \infty$ ;
- (ii)  $\text{supp } P = [0, 1]$ ,  $P$  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. (12) owns a unique equilibrium in  $F^+$ .

*Then, with probability one,  $\{X^n\}$  converges to the equilibrium of the m.d.e. (12) in  $F^+$ .*

**REMARK 2.** If  $X^n$  is ordered in a descending manner, then Theorem 4 will be valid while the m.d.e. (12) changes its form as follows:

$$(19) \quad P_i(X^n) = \int_{\bar{X}_{i+1}^n}^{\bar{X}_i^n} P(dv), \quad Q_i(X^n) = \int_{\bar{X}_{i+1}^n}^{\bar{X}_i^n} v P(dv) \quad \forall 1 \leq i \leq N.$$

**PROOF OF THEOREM 4.** Lemmas 4(ii) and 5 show that Theorem 3 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(iii) implies that the same is true for  $\bar{F}^+$  and finally Lemma 1 implies a.s. convergence of  $X^n$  to the equilibrium.  $\square$

While the uniqueness of the equilibrium in  $\bar{F}^+$  is sufficient for a.s. convergence of  $X^n$ , the case of nonunique equilibrium cannot be excluded from an applications point of view. Corollary 1 establishes a weaker version of Theorem 4, which does not require uniqueness of equilibrium.

**COROLLARY 1.** *Assume conditions (i) and (iii) of Theorem 4 and (ii) and (iv) following.*

(ii)  $\text{supp } P = [0, 1]$ ,  $P$  has a density  $P(dv) = p(v)dv$ , which is continuous on  $[0, 1]$ ,  $p > 0$  on  $(0, 1)$  and one of the following conditions hold.

- (ii)(a)  $\log p$  is concave on  $[0, 1]$  and  $p(0) + p(1) > 0$ ;
- (ii)(b)  $\log p$  is strictly concave on  $[0, 1]$ ;
- (iv) The m.d.e. (12) owns a countable number of equilibria in  $F^+$ .

Then  $\Phi_x(t)$  approaches an asymptotically stable equilibrium for almost all  $x \in \bar{F}^+$ .

PROOF. It suffices to show that all the equilibria in  $F^+$  are simple; then the claim of Corollary 1 follows from Lemmas 4 and 5 and Theorem 2.

According to Lemma 3 there exists  $x^* \in F^+$  such that  $h(x^*) = 0$ . Using the Gershgorin theorem for irreducible matrices, it is enough to show that  $\partial h_i(x^*)/\partial x_i < 0$ ,  $\sum_{k=1}^N \partial h_i(x^*)/\partial x_k \leq 0$  for all  $i \in \{1, \dots, N\}$  and  $\sum_{k=1}^N \partial h_i(x^*)/\partial x_k < 0$  for at least one  $i$ ; see [9].

Equation (12) yields

$$x_i^* = \frac{S_i(x^*)}{R_i(x^*)};$$

combining this expression with (18) we find

$$\sum_{k=1}^N \frac{\partial h_i(x^*)}{\partial x_k} = -\frac{D_i(x^*)}{R_i(x^*)},$$

where

$$D_i(x^*) := R_i^2(x^*) - \sum_{k=1}^{N-1} (\gamma_{k+1-i} - \gamma_{k-1}) p(\bar{x}_{k+1}^*) (S_i(x^*) - x_{k+1}^* R_i(x^*)).$$

Following Fort and Pagés [7] for  $u \in [0, 1]^{N+1}$ , we introduce

$$\begin{aligned} \varphi_i^N(u, \gamma) := & \left( \sum_{k=1}^N \gamma_{|k-i|} \int_{u_k}^{u_{k+1}} p(v) dv \right)^2 \\ & - \sum_{k=1}^{N+1} \tau(k, i) p(u_k) \sum_{r=1}^N \gamma_{|r-i|} \int_{u_r}^{u_{r+1}} (v - u_k) p(v) dv, \end{aligned}$$

where  $\gamma = (\gamma_0, \dots, \gamma_N)$  and

$$\tau(k, i) := \gamma_{|k-i|} \mathbf{1}_{\{k \leq N\}} - \gamma_{|k-i-1|} \mathbf{1}_{\{k \geq 2\}}.$$

Then it is easy to see that

$$D_i(x^*) = \varphi_i^N(\bar{x}^*, \gamma) + p(0) \gamma_{|i-1|} S_i(x^*) + p(1) \gamma_{|N-i|} (R_i(x^*) - S_i(x^*)).$$

In [7] it is shown that under condition (ii)(a)  $\varphi_i^N(\bar{x}, \gamma) \geq 0$  for all  $x \in \bar{F}^+$  and  $i \in \{1, \dots, N\}$ . Moreover, if  $\gamma_1 < 1$  and  $\bar{x}_i < \bar{x}_{i+1}$ , then (ii)(b) implies  $\varphi_i^N(\bar{x}, \gamma) > 0$  for any  $N$  and  $i \in \{1, \dots, N\}$ . For the sake of brevity we do not repeat the proofs of these results.

Now suppose (ii)(a) is fulfilled. Then  $D_i(x^*) \geq 0$  for all  $i$ ; moreover, either  $D_1(x^*) > 0$  or  $D_N(x^*) > 0$ .

For the case that condition (ii)(b) is fulfilled, assume  $\gamma_{j+1} < \gamma_j = 1$  and set  $\bar{\gamma}_i := \gamma_{i+j}$ . Then  $\bar{\gamma}_1 < 1$  and

$$\varphi_l^N(\bar{x}, \gamma) = \varphi_1^{N-j-l+1}(\bar{x}_1, \bar{x}_{j+l+1}, \dots, \bar{x}_{N+1}, \bar{\gamma}) > 0 \quad \text{for } 1 \leq l \leq j,$$

$$\varphi_l^N(\bar{x}, \gamma) = \varphi_{l-j}^{N-2j}(\bar{x}_1, \dots, \bar{x}_{l-j}, \bar{x}_{l+j+1}, \dots, \bar{x}_{N+1}, \bar{\gamma}) > 0 \quad \text{for } j \leq l \leq N-j,$$

$$\varphi_l^N(\bar{x}, \gamma) = \varphi_{l-j}^{l-j}(\bar{x}_1, \dots, \bar{x}_{l-j}, \bar{x}_{N+1}, \bar{\gamma}) > 0 \quad \text{for } N-j \leq l \leq N,$$

which imply that  $D_i(x^*) > 0$  for all  $i$ .

$R_i(x)$  is positive on  $F^+$ , so  $\sum_{k=1}^N \partial h_i(x^*) / \partial x_k < 0$  if  $D_i(x^*) > 0$  and  $\sum_{k=1}^N \partial h_i(x^*) / \partial x_k \leq 0$  if  $D_i(x^*) \geq 0$ . Moreover for any  $i$  there exists  $k \neq i$  such that  $\partial h_i(x^*) / \partial x_k > 0$ ; this together with  $\partial h_i(x^*) / \partial x_k \geq 0$  for  $i \neq k$  and  $\sum_{k=1}^N \partial h_i(x^*) / \partial x_k \leq 0$  imply that  $\partial h_i(x^*) / \partial x_i < 0$  for all  $i \in \{1, \dots, N\}$ .  $\square$

**REMARK 3.** The proof of Corollary 1 implies actually more than simplicity of equilibria. We have shown that all the equilibria have eigenvalues with negative real parts.

**4. Conclusion.** Here we have established a.s. convergence of the one-dimensional Kohonen algorithm for a reasonably large class of stimuli distributions and neighborhood functions. The term a.s. convergence is used here in the same sense as the unconditionally a.s. convergence in [2, 7]. Corollary 1 establishes a convergence result in a weaker sense, which is, however, stronger than the “conditionally a.s. convergence” or the “convergence in the Kushner–Clark sense”; while the conditional convergence amounts to the existence of some neighborhood, in which the fixed point is asymptotically stable, Corollary 1 assures that the neighborhood includes almost all  $x \in \bar{F}^+$ .

The existence and uniqueness of equilibrium in  $F^+$  is an important issue which possibly plays a key role in the asymptotic behavior of Kohonen’s algorithm. Lemma 3 establishes an existence result under weaker restrictions than the known criteria, [7]. These restrictions are used in Theorem 4 and Corollary 1 to establish the convergence results and include all the assumptions which are known to be sufficient for establishment of ordering [5, 6].

The restrictions imposed on the learning parameters are necessary to ensure a.s. convergence of the algorithm. In fact, many numerical experiments show that if  $\varepsilon_n$  decreases faster than what condition (i) of Theorem 4 allows, then  $X^n$  may get stuck in a nonoptimum equilibrium. It can be shown that a better rate of convergence will be reached if  $\sum_n \varepsilon_n^2$  converges as fast as possible without a violation of condition (i).

Although the neighborhood functions used in most applications are symmetric, nonincreasing asymmetric functions may be utilized, too. To establish Theorem 4 we did not make any use of the symmetry of neighborhood functions and hence the results are also valid for asymmetric nonincreasing functions. Of course, in such a case, equations (5) and (12) change their forms. Moreover,



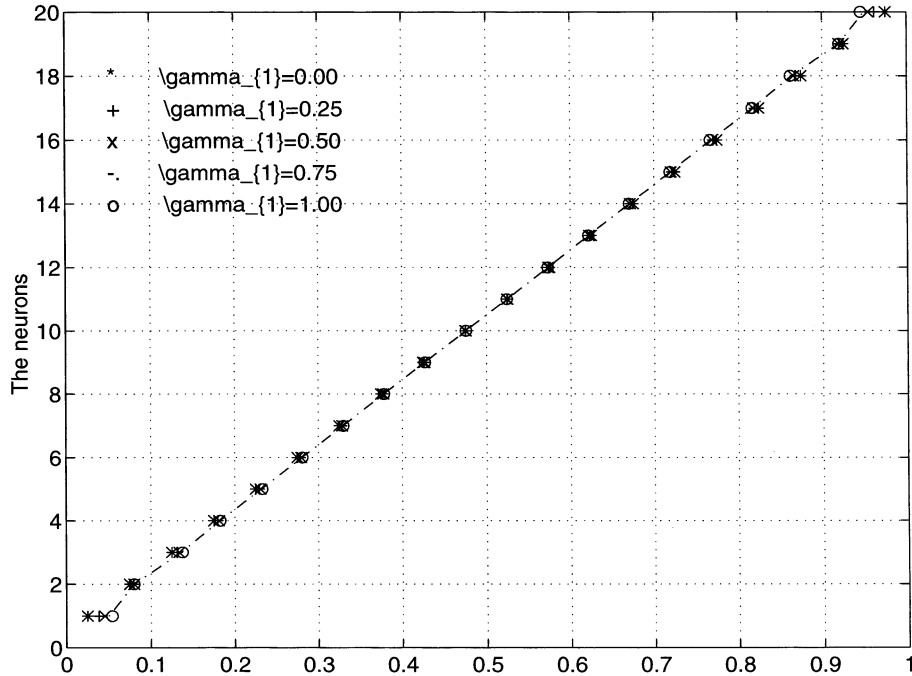


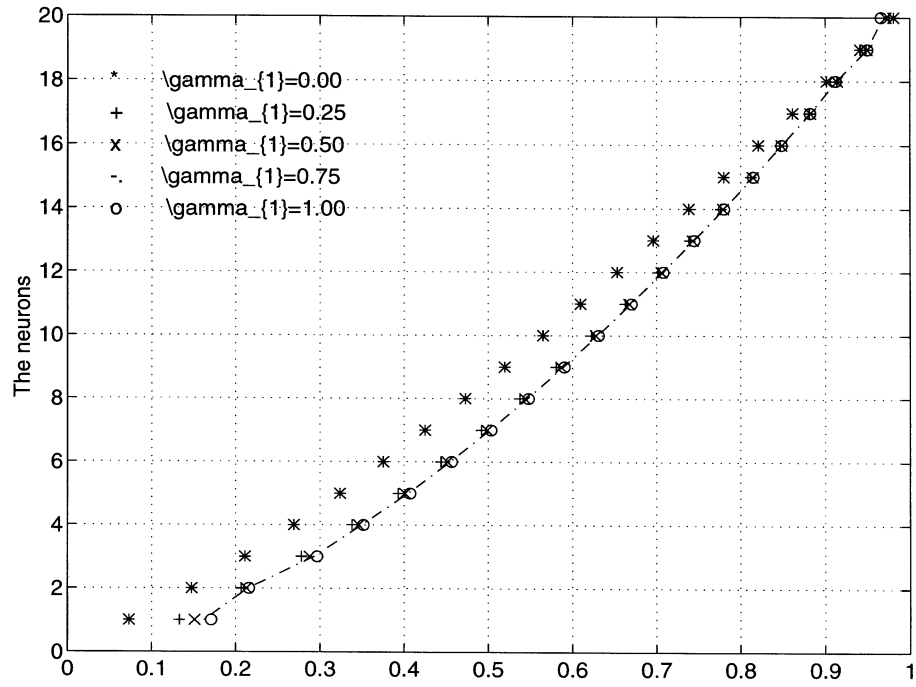
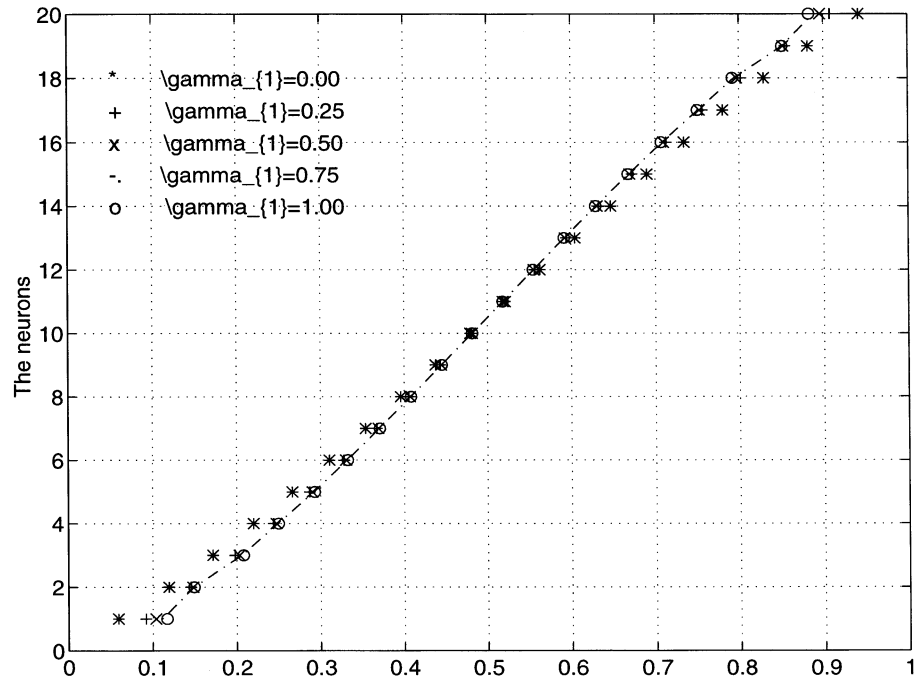
FIG. 1. The final values of neurons for  $P(x) = 1$  in  $[0, 1]$ .

the index  $j$  in Theorem 4 and Lemmas 3–5 will be the smallest index for which  $\gamma_{j+1} < \gamma_j$  and  $\gamma_{-j-1} < \gamma_{-j}$ .

Effects of the stimuli density distribution  $p$  and the neighborhood parameters  $\gamma_1, \dots, \gamma_k$  on the final position of neurons in  $Q$  are interesting issues, which can be studied by stochastic approximation methods. Using the m.d.e. (12) and the results of Section 3, we are able to find out the possible final values of neurons. For  $N = 20$ ,  $k = 1$ ,  $p(v) = 1, 2v, 6v(v-1)$  on  $[0, 1]$  and  $\gamma_1 = 0.00, 0.25, 0.50, 0.75, 1.00$ , the final values are depicted in Figures 1–3. The results show that there exist two kinds of final maps made up of the equilibrium of the m.d.e., corresponding to  $\gamma_1 = 0$  or  $\gamma_1 > 0$ .

For  $\gamma_1 > 0$ , the final map is not sensitive to the different values of  $\gamma_1$ . This suggests that the rate of convergence can be improved by choosing a suitable positive  $\gamma_1$ . Another aspect of these simulations is the investigation of the number of zeros which  $h(x)$  owns. Computations show that for these cases there is always a unique zero in  $\bar{F}^+$ . Hence Figures 1–3 present the final values, to which, according to Theorem 4, the algorithm converges with probability 1.

**Acknowledgment.** The author thanks Professor D. Prätzel-Wolters for his support, which made this work possible.

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] BENVENISTE, A., MÉTIVIER, M. and PRIOURET, P. (1990). *Adaptive Algorithms and Stochastic Approximations*. Springer, New York.
- [2] BOUTON, C. and PAGÉS, G. (1993). Self-organization and a.s. convergence of the one-dimensional Kohonen algorithm with non-uniformly distributed stimuli. *Stochastic Process. Appl.* **47** 249–274.
- [3] BOUTON, C. and PAGÉS, G. (1994). Convergence in distribution of the one-dimensional Kohonen algorithm when the stimuli are not uniform. *Adv. in Appl. Probab.* **26** 80–103.
- [4] COTTRELL, M. and FORT, J. C. (1987). Étude d'un processus d'auto-organisation. *Ann. Inst. H. Poincaré* **23** 1–20.
- [5] ERWIN, E., OBERMEYER, K. and SCHULTEN, K. (1992). Self-organizing maps: ordering, convergence properties and energy functions. *Biol. Cybern.* **67** 47–55.
- [6] FLANAGAN, J. A. (1996). Self-organisation in Kohonen's SOM. *Neural Networks* **7** 1185–1197.
- [7] FORT, J. C. and PAGÉS, G. (1995). On the a.s. convergence of the Kohonen algorithm with a general neighborhood function. *Ann. Appl. Probab.* **4** 1177–1216.
- [8] FORT, J. C. and PAGÉS, G. (1996). About the Kohonen algorithm: strong or weak self-organization? *Neural Networks* **9** 773–785.
- [9] HACKBUSCH, W. (1992). *Elliptic Differential Equations: Theory and Numerical Treatment*. Springer, Heidelberg.
- [10] NOBLE, B. and DANIEL, J. W. (1988). *Applied Linear Algebra*. Prentice-Hall, Englewood Cliffs, NJ.
- [11] HIRSCH, M. W. (1985). Systems of differential equations that are competitive or cooperative 2: convergence almost everywhere. *Siam J. Math. Anal.* **16** 423–439.
- [12] HIRSCH, M. W. (1988). Stability and convergence in strongly monotone dynamical systems. *J. Reine Angew. Math.* **383** 1–53.
- [13] KOHONEN, T. (1989). Speech recognition based on topology-preserving neural maps. In *Neural Computing Architecture* (I. Alexander, ed.) 26–40. Kogan Page, London.
- [14] KOHONEN, T. (1995). *Self-Organizing Maps*. Springer, Berlin.
- [15] KUSHNER, H. J. and CLARK, D. S. (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Springer, New York.
- [16] OJA, E. (1992). Self-organizing maps and computer vision. In *Neural Networks for Perception* **1** (H. Wechsler, ed.) 368–385. Academic Press, Boston.
- [17] PAGÉS, G. (1993). Voronoi tessellation, space quantization algorithms and numerical integration. *Proceedings of the ESANN93 Conference, Brussels* 221–228. D Facto.
- [18] REED, M. and SIMON, B. (1980). *Methods of Modern Mathematical Physics 1. Functional Analysis*. Academic Press, New York.
- [19] RITTER, H., MARTINETZ, T. and SCHULTEN, K. (1989). Topology-conserving maps for motor control. In *Neural Networks, from Models to Applications* (L. Personnaz and G. Dreyfus, eds.) 579–591. IDSET, Paris.
- [20] VILLMANN, T., DER, R., HERRMANN, M. and MARTINETZ, T. (1997). Topology preservation in self-organizing feature maps: exact definition and measurement. *IEEE Trans. Neural Networks* **8** 256–266.

ARBEITSGRUPPE TECHNOMATHEMATIK  
 UNIVERSITÄT KAISERSLAUTERN  
 67653 KAISERSLAUTERN  
 GERMANY  
 E-MAIL: sadeghi@mathematik.uni-kl.de