ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 193-198

Almost Sure Convergence of the One-dimensional Kohonen Algorithm

Michel BENAÏM
Univ. TOULOUSE (France)

Jean-Claude FORT
Univ. Nancy I & SAMOS-Paris I (France)

Gilles PAGÈS
Univ. Paris 12 & Paris 6, URA 224 (France)

Abstract

We show in a very general framework the a.s convergence of the Kohonen algorithm in dimension 1 (units and stimuli) after self-organization when the learning rate decreases to 0 in a suitable way. The main assumption is a log-concavity of the stimuli distribution, but this includes all the usual probability distributions (uniform, exponential, gamma distribution with parameter ≥ 1, ...)

1 Introduction.

Since 1982, when T. Kohonen presented his self-organizing algorithm (see [9],[10]), the first rigorous proof of *a.s.* convergence has been obtained in the uniformly distributed case and a 1-dimensional array of units with the two nearest neighbors (see [2],[3](hum!!)). In [5], this result was extended to a more general class of decreasing enough neighborhood functions, but still when the stimuli are uniformly distributed. Furthermore, a proof of conditional convergence (in the Kushner& Clark sense) was obtained for the same class of neighborhood functions under a ln-concavity assumption on the stimuli distribution. In the present paper, gathering all the previous results and calling upon a strong result by M. Hirsch on cooperative dynamical systems we establish the *a.s.* convergence toward a *unique* equilibrium point under a the same assumption on the distribution and a more general one on the neighborhood functions.

First we shortly recall the basic definitions and known results about the Kohonen one-dimensional algorithm.

The units are identified to the set $\{1, 2, \cdots, n\}$. σ denotes the neighborhood function. It satisfies : $\sigma(0) := 1$, $\sigma(k) = \sigma(-k)$, σ non increasing. The stimuli ω^t, $t \geq 1$ are i.i.d., $[0,1]$-valued and have a *continuous* distribution μ. A weight is associated to each unit and $X^t := (X_i^t)_{1 \leq i \leq n}$ denotes the weight vector at time t. The X_i^t's are supposed $[0,1]$-valued too. Let $X^0 := (x_i)_{1 \leq i \leq n}$ the initial weights. At time $t+1$ the algorithm is recursively defined in two phases by:

(i) Competitive phase:

$$\text{computation of the winning unit } i^{t+1} := i(\omega^{t+1}, X^t) = \underset{k \in I}{\operatorname{argmin}} |\omega^{t+1} - X_k^t|.$$

(ii) Cooperative phase:

$$\forall j \in \{1, 2, \cdots, n\}, \quad X_j^{t+1} = X_j^t - \epsilon_{t+1}\sigma(i^{t+1} - j)(X_j^t - \omega^{t+1})$$

where $(\epsilon_t)_{t \geq 1}$ is a sequence of $]0,1[$-valued real numbers. ε_t is the *learning rate* at time t.

$(X^t)_{t \in \mathbb{N}}$ is a Markov chain – homogeneous if $\epsilon_t = \epsilon > 0$ – and, if $x \in D := \{x \in [0,1]^n / x_i \neq x_j \text{ if } i \neq j\}$ then, $\mathbb{P}_x$-*a.s.*, $X^t \in D$ for every $t \in \mathbb{N}$. Thus, as soon as $X^0 \in D$ *a.s.*, the algorithm is *a.s.* well defined.

2 Previous results

Let $F_n^+ := \{x \in [0,1]^n, 0 < x_1 < x_2 < \cdots < x_n < 1\}$ and $F_n^- = \{x \in [0,1]^n, 0 < x_n < x_{n+1} < \cdots < x_1 < 1\}$.

- *Self-organization.*

- if $\sigma(k) := \mathbf{1}_{\{|k| \leq 1\}}$ (see [3]), or if $k \mapsto \sigma(k)$ is non decreasing decreasing (and non negative), then F_n^+ and F_n^- are absorbing sets (see [4], [5]). When the learning rate is constant, $\varepsilon_t = \varepsilon$, the entering time of X^t in $F_n := F_n^+ \cup F_n^-$ is $\mathbb{P}_x$-*a.s.* finite and has an exponential moment, uniformly in $x \in [0,1]^n$ (see [3] when $\mu := U([0,1])$, [2] for more general distributions μ).

- *Convergence.*

- if $\sum_t \varepsilon_t = +\infty$ and $\sum_t \varepsilon_t^2 < +\infty$ (decreasing learning rate) and if σ satisfies

$$(H_\sigma) \equiv \text{there exists } k_0 \leq \tfrac{n-1}{2} \text{ s.t. } \sigma(k_0 + 1) < \sigma(k_0)$$

then :

(a) the mean function of the algorithm, $-h$ (see (??) below), can be extended to a continuous function on the closure $\overline{F_n^+}$ of F_n^+ whenever μ weights no single point. If μ has a positive density f, there is at least one equilibrium point x^* inside F_n^+ and any equilibrium points actually lies inside F_n^+.

ESANN'1997 proceedings - European Symposium on Artificial Neural Networks
Bruges (Belgium), 16-17-18 April 1997, D-Facto public., ISBN 2-9600049-7-3, pp. 193-198

(b) if μ fulfills $\mathcal{L}\equiv\begin{cases}\bullet & \text{either a strictly log-concave density } f \text{ on }]0,1[\\ \bullet & \text{or a log-concave density } f \text{ on }]0,1[\text{ s.t. } f(0_+)+f(1_-)>0\end{cases}$
then h is Lipschitz and all the equilibrium points x^* are stable (*i.e.* have a stable attracting area).

(c) if $\mu = U([0,1])$, h has a unique equilibrium point x^* in F_n^+ and $X^t \to x^*$ $\mathbb{P}_x$-*a.s.*

Claim (a) is established under the optimal assumption (H_σ) in [13]. Claims (b) and (c) can be found in [5] under a (slightly) more restrictive assumption than (H_σ). Nevertheless it can be straightforwardly extended whenever (a) is established under (H_σ).

3 *A.s.*-convergence toward a unique equilibrium point.

We proceed in three steps:

– first we prove that there is a unique equilibrium point x^* of the O.D.E $\dot{x} = -h(x,\sigma)$,

– then we verify the assumptions of Hirsch's Theorem about the strongly monotone dynamical systems,

– and finally we apply a slightly improved version of the Kushner & Clark Theorem to conclude.

We begin by writing the O.D.E.. It reads :

$$\begin{aligned} \dot{x} &= -h(x,\sigma) \\ \text{with} \qquad h_i(x,\sigma) &= \sum_{k=1}^{n} \sigma(|k-i|) \int_{]\tilde{x}_k,\tilde{x}_{k+1}]} (x_i - \omega)\mu(d\omega) \end{aligned}$$

$$\text{where we set :} \qquad \tilde{x}_1 = 0_-, \quad \tilde{x}_k = \frac{x_k + x_{k-1}}{2} \quad 2 \le k \le n, \quad \tilde{x}_{n+1} = 1^+.$$

We have the following results, that where partially (items (i), (ii)) "guessed" in [13].

Proposition 1 *(i) The dynamical system $\dot{x} = -h(x)$ is cooperative on F_n^+ (i.e. the non diagonal elements of $\nabla h(x)$ are non positive).*

(ii) The matrices $\nabla h(x)$ are irreducible on F_n^+.

(iii) There is a unique equilibrium point in F_n^+.

(iv) The set of the limiting values of a trajectory starting from $x_0 \in \overline{F}_n^+$ is a compact connected set of F_n^+.

We use the Theorem 0.5 of [8] that says that all the trajectories of a strongly monotone dynamical system on a set X, with compact orbit closures in the interior of X and a unique equilibrium point, converge to it. From this we have :

Corollary 2 *All the trajectories of the O.D.E.* $\dot{x} = -h(x)$ *starting in* $\overline{F}_n^+$ *converge to* x^*.

We now state a simplified version of an improved Kushner & Clark's result (see *e.g.* [11],[1],[7]).

Theorem 3 *Let* $X^{t+1} = X^t + \varepsilon_{t+1}[-h(X^t) + \Delta M^{t+1}]$ *be a stochastic algorithm taking its values in a compact set* K *of* $\mathbb{R}^n$. *Assume that* h *is Lipshitz and that* ΔM^{t+1} *is the sequence of* L^q*-bounded increments of a martingale for* $q \geq 2$. *If* $\sum_{t\geq 1} \varepsilon_t = +\infty$, $\sum_{t\geq 1} \varepsilon_t^{(1+q/2)} < +\infty$ *and if the flow of the O.D.E.* $\dot{x} = -h(x)$ *converges (on* K*) to the unique equilibrium point* x^* *of* h, *assumed to be* stable *in the K& C sense, then*

$$X^t \text{ converges } \mathbb{P}\text{-}x \text{ a.s. to } x^*.$$

Then, it derives from the above Corollary and Theorem 3, the result

Theorem 4 *If* μ *satisfies condition* $\mathcal{L}$, *if* σ *satisfies* (H_σ), *if* $X^0 \in \overline{F}_n^+$ *then* X^t *converges* $\mathbb{P}_x$ *a.s. to the unique equilibrium point* x^*, *unique zero of* h *in* $\overline{F}_n^+$.

4 Sketch of proofs

We mention here how to prove the four items of Proposition 1.
(i) In [5] it is shown that, setting $f(0_-) = f(1_+) := 0$:

$$\begin{aligned} \forall x \in F_n^+,\ \forall i \neq j\ \frac{\partial h_i}{\partial x_j}(x) &= \frac{\sigma(|i+1-j|) - \sigma(|i-j|)}{2}(x_i - \tilde{x}_j) f(\tilde{x}_j) \\ &+ \frac{\sigma(|i-j|) - \sigma(|i-1-j|)}{2}(x_i - \tilde{x}_{j+1}) f(\tilde{x}_{j+1}) \end{aligned}$$

which is clearly non positive.

(ii) We assume $\sigma(k_0) < \sigma(k_0 + 1)$ for some $k_0 \leq \frac{n-1}{2}$. Let A the matrix

$$A := [a_{ij}]_{1\leq i,j\leq n},\ a_{ij} := |\sigma(|i+1-j|) - \sigma(|i-j|)| \mathbf{1}_{2\leq j\leq n}.$$

It is obvious that the irreducibility of $\nabla h(x)$ and of the matrix $B := [b_{ij}]_{1\leq i,j\leq n}$, $b_{ij} = a_{ij} + a_{i,j+1}$ $(a_{i,n+1} := 0)$, are equivalent. We just note that the "diagonal" $a_{i,j}$, $k + \ell =$ of A is made of positive elements. This prove the result since now B has then two consecutive positive "diagonals".

(iii) We apply the well known result that follows(see [12]): let V be a C^0 manifold with boundary ∂V and let f be a C^0 vector field on V pointing outside V on ∂V and having a finite set of zeros inside V. Then the sum of all the Morse indices of the zeros of f is equal to the Euler characteristics of V.

$\overline{F}_n^+$ is homeomorphic to the unit closed disk D_n so its Euler characteristics is 1. But all the zeros of h have a stable attracting area, hence they all have an index 1. This proves uniqueness as soon as the vector field $-h$ is pointing outside ∂F_n^+. A straightforward computation this is the case when σ is decreasing. In the general case, uniqueness derives from the implicit function theorem which shows that locally the equation $h(x^*, \sigma) = 0$ defines a function $x^* := \varphi(\sigma)$.

(iv) We prove that the trajectories of $\dot{x} = -h(x)$ started in $\overline{F}_n^+$ have no limiting point on $\partial\overline{F}_n^+$. To this end, we define for every $x \in \partial\overline{F}_n^+$, the function $\mathcal{E}(x)$ equal to the number of sets of packed components of x. We show that $\mathcal{E}(x - \frac{1}{2}h(x)) < \mathcal{E}(x)$. It follows that the solution $x(t)$ of the O.D.E lives in F_n^+ for all $t > 0$. It remains to show that there is no limiting value of the O.D.E on $\partial\overline{F}_n^+$. By carefully inspecting the behaviour of the algorithm we prove that it always separates (at least) 2 stuck components at each iteration when the n-tuple $x \in \partial\overline{F}_n^+$. Thus we deduce that for the O.D.E the speed of separation is non zero on the compact set $\partial\overline{F}_n^+$. Thus the O.D.E eventually leaves $\partial\overline{F}_n^+$ with a bounded below speed which does not allow any limiting value on $\partial\overline{F}_n^+$.

5 Conclusion.

The result of this paper almost ends the study of the *a.s.* convergence of the 1-dimensional Kohonen algorithm (*i.e.* one-dimensional units and stimuli): most usual distribution fulfill the *log*-concavity assumption. The case of higher dimension (even the simplest *i.e.* the string in the unit square) turns out to be much more difficult since we cannot find some organized absorbing set (see [6]) and thus the monotonicity of the O.D.E certainly fails.

Michel BENAÏM,
Univ. Paul Sabatier,
Toulouse.

Jean-Claude FORT,
Univ. Nancy I, Institut Elie Cartan,
B.P. 239, F-54506 Vandœuvre-Lès-Nancy Cedex
Mail: fortjc@iecn.u-nancy.fr

G. Pagès, Laboratoire de Probabilités,
URA 224, Univ. P.&M. Curie,

4, Pl. Jussieu, F-75252 Paris Cedex 05.
Mail: gpa@ccr.jussieu.fr

[1] M. Benam, A Dynamical System Approach to Stochastic Approximations, *SIAM J. Control And Optimization*, **34**, n° 2, 1996, pp.437-472.

[2] C. Bouton, G. Pags, Self-organization and convergence of the one-dimensional Kohonen algorithm with non uniformly distributed stimuli, *Stoch. Proc. & Appl.*, **47**, 1993, pp.249-274.

[3] M. Cottrell, J.C. Fort, tude d'un algorithme d'auto-organisation, *Ann. Inst. H. Poincar*, **23**, n°1, 1987, pp.1-20.

[4] A. Flanagan, Self-organizing Neural Networks, Thse 1306, E.P.F.L., Lausanne, 1994.

[5] J.C. Fort, G. Pags, On the A.S. Convergence of the Kohonen Algorithm With a General Neighborhood Function, *The Ann. of Applied Proba.*, **5**, n° 4, 1995, pp.1177-1216.

[6] J.C. Fort, G. Pags, About the Kohonen algorithm: strong or weak self-organization?, *Neural Networks*, **9**, n° 5, 1996, pp.773-785.

[7] J.C. Fort, G. Pags, Convergence of Stochastic Algorithms : from the Kushner & Clark Theorem to the Lyapounov Functional Method, *Adv. in Applied Proba.*, 1996.

[8] M. Hirsch, Stability and convergence in strongly monotone dynamical systems, *J. Reine Angew. Math.*, **383**, 1988, pp.1-53.

[9] T. Kohonen, Analysis of a simple self-organizing process, *Biol. Cybern.*, **44**, 1982, p135-140.

[10] T. Kohonen, *Self-organization and associative memory*, 3^{rd} edition Springer, Berlin, 1989.

[11] H.J. Kushner, D.S. Clark, *Stochastic Approximation for Constraint and Unconstraint Systems*, Appl. Math. Sci. Series, 26, Springer, 1978.

[12] J.W. Milnor, *Topology from the differential viewpoint*, 2^{rd} edition the University Press of Virginia, 1969.

[13] A. Sadeghi, Asymptotic Behaviour of Self-Organizing Maps with Non-Uniform Stimuli Distribution, forthcoming in *I.E.E.E. Neural Networks*, 1997.