ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 257-262

A non linear Kohonen algorithm

Jean-Claude FORT * Gilles PAGÈS †

Introduction and background material

It is well known (see [5],[6]) that, wether the stimuli distribution μ on $\mathbb{R}^d$ is continuous[1] or not, the *d-dim* Kohonen algorithm with 0 neighbour[2] and n points (or *units*) is a stochastic gradient that derives from a potential defined

by $$\forall\, x=(x_1,\cdots,x_n)\in(\mathbb{R}^d)^n \quad E_n^{2,\mu}(x):=\int \min_{1\leq i\leq n}\|x_i-\omega\|^2\mu(d\omega) \qquad (1)$$

provided that μ has a compact support. In Neural Networks terminology x_i is for the weight of *unit i* and ω is a generic stimulus. For obvious reasons, this potential may be considered as a quadratic one. In [6] was introduced a non quadratic generalization of this potential. For every $\alpha>0$, the potential $E_n^{\alpha,\mu}$ was simply defined on the space $(\mathbb{R}^d)^n$ by

$$\forall\, x=(x_1,\cdots,x_n)\in(\mathbb{R}^d)^n \quad E_n^{\alpha,\mu}(x_1,\cdots,x_n):=\int \min_{1\leq i\leq n}\|x_i-\omega\|^\alpha\mu(d\omega). \qquad (2)$$

The Kohonen potential corresponds to $\alpha=2$. Actually, if μ is continuous,

$$E_n^{\alpha,\mu}(x)=\sum_{i=1}^n\int_{C_i(x)}\|x_i-\omega\|^\alpha\mu(d\omega) \qquad (3)$$

where $(C_i(x))_{1\leq i\leq n}$ denotes the so-called Euclidean Voronoï tesselation of the space $\mathbb{R}^d$ related to x. In fact, this tesselation is only defined on $D_n:=\{x\in(\mathbb{R}^d)^n\,/\,x_i\neq x_j \iff i\neq j\}$, by

$$C_i(x):=\{u\in\mathbb{R}^d\,/\,\|x_i-u\|<\|x_k-u\|\},\ \text{if } k\neq i\},\ 1\leq i\leq n.$$

*Univ. Nancy I, Fac. Sciences, Dpt de Math., B.P. 239, F-54506 Vandœuvre-Lès-Nancy Cedex & SAMOS, Univ. Paris I, 75634 Paris Cedex 13. Mail: fortjc@iecn.u-nancy.fr.

†Labo. de Probabilités, URA 224, Univ. P. & M. Curie, Tour 56, F-75252 Paris Cedex 05 & Univ. Paris 12, UFR Sciences etTechnologie, Dpt Math. 61, av. du Gal de Gaulle, F-94010 Crteil Cedex. Mail: gpa@ccr.jussieu.fr.

[1] A measure is continuous iff no hyperplan is weighed by μ. All the distributions that have a density are continuous.

[2] also known as a "space quantization algorithm".

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 257-262

Following decomposition (3), $E_n^{\alpha,\mu}(x)$ was called the "(α,μ)-magnitude of the Voronoï tesselation at $x=(x_1,\cdots,x_n)\in(\mathbb{R}^d)^n$".

When $\alpha\in(0,2]$, the (α,μ)-magnitude function $E_n^{\alpha,\mu}$ was originally introduced in [6] as an upper-bounding modulus in a new method for high dimensional numerical integration of α-hölder[3] functions or functions having a $(\alpha-1)$-hölder first derivative. This new numerical integration technique is based on the Voronoï tesselation of any n-uplet x^* that minimizes $E_n^{\alpha,\mu}$ (see [6]).

The proposed numerical method for the computation of x^* is of course the gradient descent algorithm related to the potential $E_n^{\alpha,\mu}$

$$X^{t+1}=X^t-\varepsilon_{t+1}H_n^\alpha(X^t,\omega^{t+1}),\ X^0\in D_n \tag{4}$$

where, for every $x\in D_n$ and $\omega\in C$, $H_n^\alpha(x,\omega):=\left(\frac{x_i-\omega}{\|x_i-\omega\|}\|x_i-\omega\|^{\alpha-1}\mathbf{1}_{C_i(x)}(\omega)\right)_{1\le i\le n}$, $(\varepsilon_t)_{t\ge1}$ is $]0,1[$-valued sequence of steps and ω^t is an i.i.d. sequence of random variables with distribution μ. This formula straightforwardly derives from the integral representation on D_n of $\nabla E_n^{\alpha,\mu}(x)=\alpha\left(\int_{C_i(x)}\|x_i-\omega\|^{\alpha-1}\frac{x_i-\omega}{\|x_i-\omega\|}\mu(d\omega)\right)_{1\le i\le n}$.

1 Design of the non linear Kohonen algorithm

Let us consider now a general unit set $I\subset\mathbb{Z}^d$ endowed with a topological structure provided by a neigbourhood function σ defined on $I\times I$. In most practical cases $\sigma(i,j):=v(i-j)$ with $v(-x)=v(x)$ and we will often denote $\sigma(i-j)$ instead of $\sigma(i,j)$. Then the algorithm displays as

(i) Computation of the winning unit $i^{t+1}:=i(\omega^{t+1},X^t)=\text{argmin}_k||\omega^{t+1}-X_k^t||$. In case of conflict, one takes the lexicographic minimum,

$$(ii)\ \forall j\in I,\ X_j^{t+1}=X_j^t-\varepsilon_{t+1}\sigma(i^{t+1}-j)\frac{(X_j^t-\omega^{t+1})}{||X_j^t-\omega^{t+1}||}||X_j^t-\omega^{t+1}||^{\alpha-1}, \tag{5}$$

where $(\varepsilon_t)_{t\ge1}$ is still a sequence of $(0,1)$-valued real numbers.

Where does this extension come from? Assume for some time that μ has a discrete support, namely $supp(\mu):=\{\omega_1,\cdots,\omega_p,\cdots\}$ and set

$$\forall x\in(\mathbb{R}^d)^n\quad E_I^{\alpha,\mu,\sigma}(x):=\sum_{i,j\in I}\sigma(i-j)\int_{C_j(x)}||x_i-\omega||^\alpha\mu(d\omega).$$

Then, following [7] or [8], one checks that, whenever no ω_p lies in the borders $\partial C_i(x)$ of the tessels $C_i(x)$ *i.e.* $supp(\mu)\cap(\cup_{i\in I}\partial C_i(x))=\emptyset$, $E_I^{\alpha,\mu,\sigma}$ is differentiable at $x=(x_i)_{i\in I}$ with a gradient $\nabla E_I^{\alpha,\mu,\sigma}$ given by

[3] $f:E\mapsto F$ is α-hölder iff $\forall x,y\in E$, $||f(x)-f(y)||_E<C||x-y||_F^\alpha$ where E and F are normed vector space.

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 257-262

$$\nabla E_I^{\alpha,\mu,\sigma}(x) = \left(\sum_{j\in I}\sigma(i-j)\int_{C_j(x)}\frac{x_i-\omega}{\|x_i-\omega\|}\|x_i-\omega\|^{\alpha-1}\mu(d\omega)\right)_{i\in I}. \quad (6)$$

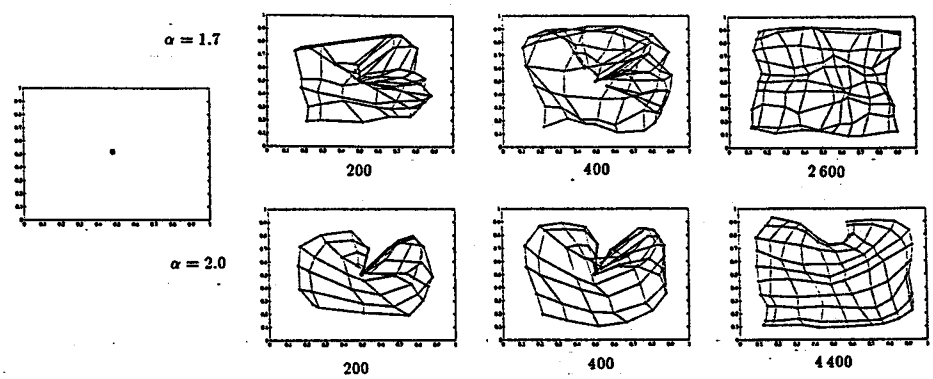
A contrario when μ has a density f *i.e.* $\mu = f(\omega)d\omega$, $\frac{\partial}{\partial x_k}\mu(C_i(x))$ is generally not 0 whenever $\overline{C_i}(x)\cap\overline{C_j}(x)\neq\emptyset$. Actually, following [4], it reads

$$\frac{\partial}{\partial x_k}\mu(C_i(x)) = \int_{\overline{C_i}(x)\cap\overline{C_k}(x)} f(\omega)\left(\frac{1}{2}\vec{n}_x^{ik} + \frac{1}{\|x_i-x_k\|}(\frac{x_i+x_k}{2}-\omega)\right)\lambda_x^{ik}(d\omega)$$

where $\vec{n}_x^{ik} := \frac{x_k-x_i}{\|x_k-x_i\|}$ and λ_x^{ik} denotes the (d–1)-dimensional Lebesgue measure on $\overline{C_i}(x)\cap\overline{C_k}(x)$ (when not reduced to a (d–2)-dimensional affine subspace). So, in this case $\nabla E_I^{\alpha,\mu,\sigma}$ is still differentiable at any point of D_n, but has no longer any integral representation with respect to μ. More generally, the equation (6) holds as soon as $\mu(C_i(x+h))-\mu(C_i(x))=o(h)$ as $h\to 0$.

2 Application to the accelereted self-organization

Rather unexpectedly, this generalization of the algorithm turned out to have some interesting self-organizing feature. Actually, this observation is the main motivation for writing this contribution. Many 2-dimensional simulations implemented with various distributions showed that self-organization is carried much faster with α less than 2, at least when ε is small. For instance with a unit set $I:=\{1,\cdots,7\}\times\{1,\cdots,7\}$, an 8-neighbour σ function[4], $\mu:=U([0,1]^2)$ and $\varepsilon:=0.1$, we observed an obvious self-organization after 2 000 trials with $\alpha=1.7$ instead of 4 400 trials with the usual $\alpha=2$ parameter. Fig. 1 below shows the main self-organizing steps with these two values of α.



Nevertheless, if definitely "obvious" on simulations, self-organization is not

[4] $\sigma(i-j):=\mathbf{1}_{\{|i_1-j_1|\le 1\}}\mathbf{1}_{\{|i_2-j_2|\le 1\}}$ where $i=(i_1,i_2)$, $j=(j_1,j_2)$.

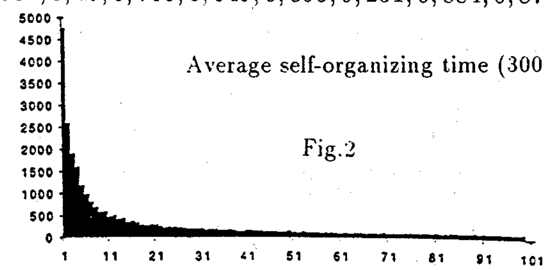
ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 257-262

a rigorously defined notion in multi-dimensional settings so far. So, in order to carry out a relevant study of the self-organization time as a function of the parameter α, we had to restrict to the one dimensional setting where self-organization amounts to monotonicity of the $i \mapsto x_i$. We based our comparison on a former empirical study mentionned in [2] and recalled below in a few words.

2.1 Self-organization of the linear 1-dim Kohonen with two neighbours

According to the theoretical results the self-organizing time of the 2 neighbour 1-dimensional Kohonen algorithm is *a.s.* finite and actually has an exponential moment ([3], [1]) for example when the stimuli distribution is locally continuous.

The simulations already discussed in [2] emphasize that the self-organizing time is quite "reasonnable" when ε is not too small. These simulations, implemented with $n = 10$ points and $\mu = U([0,1])$, were carried out on 300 independant trials of (independant) stimuli for each of the 99 selected values of the step $\varepsilon \in \{k/100,\ 1 \leq k \leq 99\}$. The same initial value was chosen at random for all the simulations. It contains 7 breaks of monotonocity : $x := (0,102; 0,901; 0,49; 0,700; 0,049; 0,895; 0,251; 0,884; 0,875; 0,692)$.



Average self-organizing time (300 trials)

Fig.2

These results obviously show that the the self-organization time, at least in the 1-dimensional setting, is a steep decreasing function of the step ε. If it looks natural for small values of ε ($\varepsilon \approx 0$), this is much more unexpected for the great values of ε ($\varepsilon \approx 1$). This phenomenon is strongly related to the number of neighbours, actually 2. Thus, if one considers a (linear) Kohonen with 4 or 6 neighbours, the self-organizing time has a minimum value and then increases again. If $\alpha < 2$ we will see below that the same phenomenon occurs.

2.2 Simulations with the non linear Kohonen algorithm

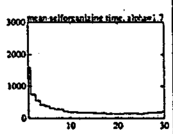
The most striking feature of the above study is that, at least with 2 neighbours, the closer to 1 ε is, the faster self-organization occurs. However, $\varepsilon \approx 1$ is a totally irrealistic choice when taking into account the quantization phase of the process which requires to let ε go to 0. So, it seems quite valuable to cut down the self-organization time for some ε close to 0.

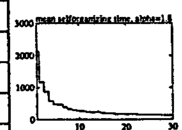
Notice that *a contrario*, whenever $\alpha < 2$, self-organization is no longer stable,

ESANN'1994 proceedings - European Symposium on Artificial Neural Networks
Brussels (Belgium), 20-21-22 April 1994, D-Facto public., ISBN 2-9600049-1-4, pp. 257-262

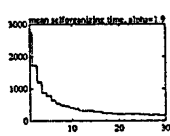
at least on a theoretical point of view [5]. However, as far as actual simulations are concerned, self-organization is quite robust for small enough ε. Of course some important degradations is observed when ε increases. That is why we processed our testing bench only for small $\varepsilon \in (0, 0.45]$, namely $\varepsilon \in \{\frac{k}{100}, 1 \leq k \leq 10, 0.12, 0,14, 0.16, 0.18, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45\}$.

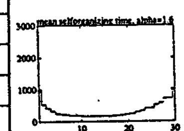
The non linear Kohonen algorithm was tested with 4 values of the parameter α, namely $\alpha \in \{1.6, 1.7, 1.8, 1.9\}$, to which the linear case $\alpha = 2$ was added in the below tables. The initial value and the number of trials – 300 – for each couple (α, ε) is the same as for $\alpha = 2$. The results are displayed in the below table (rounded up to the nearest integer) and frequency histograms.





ε \ α	1.6	1.7	1.8	1.9	2.0
0.01	956*	1 589	2 132	2 756	4 791
0.02	530*	743	1 178	1 727	2 735
0.03	410*	552	882	1 215	1 764
0.04	320*	420	587	872	1 349
0.05	252*	377	474	781	1 134
0.06	252*	314	477	638	983
0.07	224*	286	394	547	908
0.08	218*	280	330	486	724
0.09	190*	216	293	486	601
0.10	193*	193	267	461	523
0.12	171*	175	240	378	499
0.14	173*	175	245	312	463
0.16	177	150*	199	286	385
0.18	185	141*	170	240	349
0.20	209	142*	173	222	322
0.25	407	131*	149	211	284
0.30	719	216	134*	180	224
0.35	2 610	369	132*	148	198
0.40	7 114	660	157	143*	178
0.45	21 170	1 554	1 942	126*	174





The * denotes the minimum self-organizing time observed for a given ε.

2.3 Conclusions

In our opinion, the following remarks are quite interesting for future simulations and tests:

- If $\varepsilon < 0.15$ then the self-organizing time is always increasing with α.
- If $\varepsilon > 0.15$ then the self-organizing time goes through a minimum as a function of α. For example, if $\varepsilon = 0.16$, it is achieved at $\alpha = 1.7$; for $\varepsilon = 0.30$, it is achieved at $\alpha = 1.8$. We think that, for every $\varepsilon \in (0.15, 1)$, there is an optimal value $\alpha_{\min}(\varepsilon)$ of the parameter α that probably grows up with ε.

[5] It means that $i \mapsto X_i^t$ may loose its monotinicity after self-organization occured which is impossible if $\alpha = 2$.

• Whatsoever, the most interesting feature of these simulations is that when ε is close to 0, the self-organization goes faster with smaller α. Thus, if $\varepsilon = 0.01$, the self-organization time is more than 5 times shorter with $\alpha = 1.6$ than with $\alpha = 2$. Consequently, using $\alpha = 1.6$ and small values of ε will both achieve an accelerated self-organization and a good space quantization.

Conclusion

In this paper, we give some evidences about the superiority of this "non linear" Kohonen algorithm to achieve both self-organization and space quantization, provided that one works with some small constant gain parameter ε. Although these first empirical results are quite promising, some further investigations must be carried out in two directions:

• A wide testing bench to confirm the efficiency of the method and to develop some "know how" concerning the optimal parameters (α, ε).

• Some theoretical study in order to provide some analytic knowledge of the function $\varepsilon \mapsto \alpha_{\min}(\varepsilon)$.

References

[1] C. Bouton, G. Pagès, Self-organization and convergence of the one-dimensional Kohonen algorithm with non uniformly distributed stimuli, *Stochastic Processes and their Applications*, **47**, 1993, p.249-274.

[2] C. Bouton, G. Pagès, Self-organization of the 1-dimensional Kohonen algorithm, Proceedings of *Aspects théoriques des réseaux de neurones*, Congrès satellite du Congrès Européen de Mathématiques, Paris, 2-3 Juillet 1993.

[3] M. Cottrell, J.C. Fort, Etude d'un algorithme d'auto-organisation, *Annales de l'Institut Henri Poincaré*, **23**, 1987, n°1, p.1-20.

[4] J.C. Fort, G. Pagès, Sur la convergence *p.s.* de l'algorithme de Kohonen généralisé, note aux CRAS, série I, 1993, **317**, p.389-394.

[5] T. Kohonen, Self-organizing maps: optimization approaches, *Artificial Neural Networks*, T. Kohonen et *al.* ed., Elsevier Science Publishers B.V., 1991, p.981-990.

[6] G. Pagès, Mosaïques de Voronoï, algorithmes de quantification de l'espace et intégration numérique, pré-publication n°138 du Labo. Proba. de l'Univ. Paris 6, 1992 (soumis). Version (anglaise) abrégée dans *Proceedings of the ESANN 93*, M. Verleysen éd., Editions Quorum (ISBN 2-9600049-0-6), Bruxelles, 1993, p.221-228.

[7] H. Ritter, T. Martinetz, K. Schulten, *Neural Computation and Self-organizing Maps*, Addison-Wesley, 1992, New-York, 306p.

[8] P. Ruzicka, On convergence of learning algorithm for topological maps, *Neural Networks World*, **4**, 1993, p.413-424.