Relation between the "SOM" algorithm and the distorsion measure

Joseph Rynkiewicz
SAMOS-MATISSE, University of Paris I
Paris. France
rynkiewi@univ-paris1.fr

Abstract - *We study the relation between the equilibrium point of the SOM algorithm and the minimum of the distorsion measure. After calculating the derivatives of the distorsion measure, we show that these points are well separated in general. We illustrate, with a simple example, how it occurs.*

Key words - distorsion measure, asymptotic convergence

1 Introduction

The extended variance or distorsion measure, is certainly the most popular criteria for assessing the quality of the classification of a Kohonen map (see Kohonen [6]). This measure yields us a assesment of model properties with respect to the data and overcome the absence of cost function in the SOM algorithm. Moreover it has been shown that the SOM algorithm is an approximation for the gradient of the distortion measure (see Graepel et al.[4]).
However, if it is proven that the Kohonen converges in some cases when the number of observations tends to the infiny, it is also known that the limit doesn't minimize the theoretical distorsion measure (see for example Erwin et al. [2]). This property seems to be paradoxal, in one hand SOM seems minimizing the distorsion for a finite number of observations, but this behaviour is no more true for the limit, i.e. an infinity of observations.
In this paper we will investigate the relationship between SOM and distorsion measure, it is organized as follow : first we recall the mathematical definition of the distorsion measure in the discrete and continuous case, then we calculate the derivatives of the distorsion in the continuous case, we deduce from this calculation that the point minimizing the limit distorsion can be very far from the equilibrium point of the SOM and finally we illustrate, with a simple example, why the apparent contradiction between the discrete and the continuous case occurs.

2 The distorsion measure

We adopt in this section the notation of Cottrell et al. [1]. We consider a set of units indexed by $I \subset \mathbb{Z}^d$ with the neighborhood function Λ from $I - I := \{i - j,\ i, j \in I\}$ to [0, 1] satisfying $\Lambda(k) = \Lambda(-k)$ and $\Lambda(0) = 1$, note that such neighborhood function can be discrete or continuous. These units will be called in the sequel "centroïds". Let $x := (x_i)_{i \in I}$ be the set of units, the Vorono tessellation $(C_i(x))_{i\in I}$ is defined by

$$C_i(x) := \left\{\omega \in [0,1]^d \,|\, \|x_i - \omega\| < \|x_k - \omega\| \text{ si } k \neq i\right\}$$

In case of equality we assign $\omega \in C_i(x)$ thanks the lexicographical order.

2.1 Distorsion for the discrete case

We assume that the observations are in a finite set $\Omega = \{\omega_1, \cdots, \omega_n\}$ and are uniformly distributed on this set. Then, the distorsion measure or extended variance is

$$V_n(x) = \frac{1}{2n} \sum_{i\in I} \sum_{\omega \in C_i(x)} \left(\sum_{j\in I} \Lambda(i-j) \|x_j - \omega\|^2 \right)$$

It is well know that this function is not continuous with respect to the centroïds $(x_i)_{i\in I}$.

2.2 Distorsion for the continuous case

Let us assume that P is the distribution function of the observations. the theoretical distorsion measure is

$$V(x) = \frac{1}{2} \sum_{i,j\in I} \Lambda(i-j) \int_{C_i(x)} \|x_j - \omega\|^2 \, dP$$

In the sequel, we suppose that the distribution P has a density with respect to the Lebesgue measure bounded by a constant $B > 0$.

3 Derivability of $V(x)$

Let us now write

$$D_I := \left\{ \left(x_i = \left(x_i^1, \cdots, x_i^d \right) \right)_{i\in I} \in \left([0,1]^d \right)^I \middle| \forall k \in \{1, \cdots, d\} \left\| x_i^k - x_j^k \right\| > 0 \text{ si } i \neq j \right\}$$

For i and $j \in I$, let us note $\overrightarrow{n}_x^{ij}$ the vector $\frac{x_j - x_i}{\|x_j - x_i\|}$ and let

$$M_x^{ij} :=: \left\{ u \in \mathbb{R}^d / \left\langle u - \frac{x_i - x_j}{2}, x_i - x_j \right\rangle = 0 \right\}$$

be the mediator hyperplan. Let us note $\lambda_x^{ij}(\omega)$ the Lebesgue measure on M_x^{ij}. Fort and Pagès [3], show the following lemma :

Lemma 1 *Let ϕ be a $\mathbb{R}$ valued continuous function on $[0,1]^d$. For $x \in D_I$, let be $\Phi_i(x) := \int_{C_i(x)} \phi(\omega)\, d\omega$. We note also $(e_1, \cdots, e_d)$ the canonical base of $\mathbb{R}^d$. The function Φ_i is continuously derivable on D_I and $\forall i \neq j$, $l \in \{1, \cdots, d\}$*

$$\frac{\partial \Phi_i}{\partial x_j^l}(x) = \int_{\bar{C}_i(x) \cap \bar{C}_j(x)} \phi(\omega) \left\{ \frac{1}{2} \left\langle \overrightarrow{n}_x^{ij}, e_l \right\rangle + \frac{1}{\|x_j - x_i\|} \times \left\langle \left(\frac{x_i + x_j}{2} - \omega \right), e_l \right\rangle \right\} \lambda_x^{ij}(\omega)\, d\omega$$

Moreover, if we note $\frac{\partial\Phi_i}{\partial x_i}(x) := \begin{pmatrix} \frac{\partial\Phi_i}{\partial x_j^1}(x) \\ \vdots \\ \frac{\partial\Phi_i}{\partial x_j^d}(x) \end{pmatrix}$

$$\frac{\partial\Phi_i}{\partial x_i}(x) = -\sum_{j\in I, j\neq i} \frac{\partial\Phi_i}{\partial x_j}(x)$$

we then deduce the theorem

Theorem 1 *If* $P(d\omega) = f(\omega)\, d\omega$*, where* f *is continuous on* $[0;1]^d$*, then* V *is continuously derivable on* D_I *and we have*

$$\begin{aligned} \frac{\partial V}{\partial x_i}(x) &= \sum_{k\in I} \Lambda(i-k) \int_{C_k(x)} (x_i - \omega)\, P(d\omega) \\ &+\tfrac{1}{2} \sum_{j\in I} \sum_{k\in I, k\neq i} (\Lambda(k-j) - \Lambda(i-j)) \\ &\times \int_{\bar{C}_k(x)\cap\bar{C}_i(x)} \|x_j - \omega\|^2 \left(\tfrac{1}{2} \overrightarrow{n}_x^{ki} + \tfrac{1}{\|x_k - x_i\|} \times \left(\tfrac{x_i + x_k}{2} - \omega\right)\right) \\ &f(\omega)\, \lambda_x^{ki} d\omega \end{aligned}$$

where $\frac{\partial V}{\partial x_i}(x) = \begin{pmatrix} \frac{\partial V}{\partial x_i^1}(x) \\ \vdots \\ \frac{\partial V}{\partial x_i^d}(x) \end{pmatrix}$

Proof As the function $V(x)$ is continuous on D_I, we only have to show that the partial derivatives exist and are continuous. We note $h_i^l \in \mathbb{R}^{|I|\times d}$ the vector with all components null except the component corresponding to x_i^l , which is $h > 0$. Then

$$\begin{aligned} &\frac{V\left(x+h_i^l\right)-V(x)}{h} = \\ &\frac{\frac{1}{2}\sum_{k,j\in I,\, k,j\neq i} \Lambda(k-j) \int_{C_k\left(x+h_i^l\right)} \|x_j-\omega\|^2 P(d\omega) - \frac{1}{2}\sum_{k,j\in I,\, k,j\neq i} \Lambda(k-j) \int_{C_k(x)} \|x_j-\omega\|^2 P(d\omega)}{h} \\ &+\frac{\frac{1}{2}\sum_{j\in I,\, j\neq i} \Lambda(i-j) \int_{C_i\left(x+h_i^l\right)} \|x_j-\omega\|^2 P(d\omega) - \frac{1}{2}\sum_{j\in I, j\neq i} \Lambda(i-j) \int_{C_i(x)} \|x_j-\omega\|^2 P(d\omega)}{h} \\ &+\frac{\frac{1}{2}\sum_{k\in I, k\neq i} \Lambda(k-i) \int_{C_k\left(x+h_i^l\right)} \left\|x_i+h_i^l-\omega\right\|^2 P(d\omega) - \int_{C_k(x)} \|x_i-\omega\|^2 P(d\omega)}{h} \\ &+\frac{\frac{1}{2}\left(\int_{C_i\left(x+h_i^l\right)} \left\|x_i+h_i^l-\omega\right\|^2 P(d\omega) - \int_{C_i(x)} \|x_i-\omega\|^2 P(d\omega)\right)}{h} \end{aligned}$$

so

$$\begin{aligned} &\lim_{h\to 0} \frac{V\left(x+h_i^l\right)-V(x)}{h} = \tfrac{1}{2}\sum_{k,j\in I, k,j\neq i} \Lambda(k-j) \\ &\int_{\bar{C}_k(x)\cap\bar{C}_i(x)} \|x_j-\omega\|^2 \left\{\tfrac{1}{2}\left\langle \overrightarrow{n}_x^{ki}, e_l\right\rangle + \tfrac{1}{\|x_i - x_k\|} \times \left\langle \left(\tfrac{x_k+x_i}{2} - \omega\right), e_l\right\rangle\right\} \lambda_x^{ki}(\omega)\, d\omega \\ &-\tfrac{1}{2}\sum_{k,j\in I, k,j\neq i} \Lambda(i-j) \\ &\int_{\bar{C}_k(x)\cap\bar{C}_i(x)} \|x_j-\omega\|^2 \left\{\tfrac{1}{2}\left\langle \overrightarrow{n}_x^{ki}, e_l\right\rangle + \tfrac{1}{\|x_i - x_k\|} \times \left\langle \left(\tfrac{x_k+x_i}{2} - \omega\right), e_l\right\rangle\right\} \lambda_x^{ki}(\omega)\, d\omega \\ &+\lim_{h\to 0} \frac{\frac{1}{2}\sum_{k\in I, k\neq i} \Lambda(k-i) \int_{C_k\left(x+h_i^l\right)} \|x_i-\omega\|^2 + 2h(x_i^l - w^l) + o(h) P(d\omega) - \int_{C_k(x)} \|x_i-\omega\|^2 P(d\omega)}{h} \\ &+\lim_{h\to 0} \frac{\frac{1}{2}\left(\int_{C_i\left(x+h_i^l\right)} \|x_i-\omega\|^2 + 2h(x_i^l - w^l) + o(h) P(d\omega) - \int_{C_i(x)} \|x_i-\omega\|^2 P(d\omega)\right)}{h} \end{aligned}$$

and

$$\begin{array}{l}\lim_{h\to 0}\frac{V\left(x+h_i^l\right)-V(x)}{h}=\frac{1}{2}\sum_{k,j\in I,k,j\neq i}\left(\Lambda\left(k-j\right)-\Lambda\left(i-j\right)\right)\\ \int_{\bar{C}_k(x)\cap\bar{C}_i(x)}\left\|x_j-\omega\right\|^2\left\{\frac{1}{2}\left\langle\overrightarrow{n}_x^{ki},e_l\right\rangle+\frac{1}{\|x_i-x_k\|}\times\left\langle\left(\frac{x_k+x_i}{2}-\omega\right),e_l\right\rangle\right\}\lambda_x^{ki}\left(\omega\right)d\omega\\ +\frac{1}{2}\sum_{k\in I,k\neq i}\Lambda\left(k-i\right)\\ \int_{\bar{C}_k(x)\cap\bar{C}_i(x)}\left\|x_i-\omega\right\|^2\left\{\frac{1}{2}\left\langle\overrightarrow{n}_x^{ki},e_l\right\rangle+\frac{1}{\|x_i-x_k\|}\times\left\langle\left(\frac{x_k+x_i}{2}-\omega\right),e_l\right\rangle\right\}\lambda_x^{ki}\left(\omega\right)d\omega\\ -\frac{1}{2}\sum_{k\in I,k\neq i}\int_{\bar{C}_k(x)\cap\bar{C}_i(x)}\left\|x_i-\omega\right\|^2\left\{\frac{1}{2}\left\langle\overrightarrow{n}_x^{ki},e_l\right\rangle+\frac{1}{\|x_i-x_k\|}\times\left\langle\left(\frac{x_k+x_i}{2}-\omega\right),e_l\right\rangle\right\}\lambda_x^{ki}\left(\omega\right)d\omega\\ +\sum_{k\in I}\Lambda\left(k-i\right)\int_{C_k(x)}(x_i^l-w^l)P(d\omega)\end{array}$$

finally

$$\begin{array}{l}\lim_{h\to 0}\frac{V\left(x+h_i^l\right)-V(x)}{h}=\frac{\partial V}{\partial x_i^l}\left(x\right)=\frac{1}{2}\sum_{k,j\in I,k\neq i}\left(\Lambda\left(k-j\right)-\Lambda\left(i-j\right)\right)\\ \int_{\bar{C}_k(x)\cap\bar{C}_i(x)}\left\|x_j-\omega\right\|^2\left\{\frac{1}{2}\left\langle\overrightarrow{n}_x^{ki},e_l\right\rangle+\frac{1}{\|x_i-x_k\|}\times\left\langle\left(\frac{x_k+x_i}{2}-\omega\right),e_l\right\rangle\right\}\lambda_x^{ki}\left(\omega\right)d\omega\\ +\sum_{k\in I}\Lambda\left(k-i\right)\int_{C_k(x)}(x_i^l-w^l)P(d\omega)\ \blacksquare\end{array}$$

If we assume that the minimum of the extended variance, is reached in the interior of D_I (i.e. that no centroïds collapse), we deduce from the previous results that it doesn't match the equilibrium of the Kohonen algorithm. Indeed, a point $x^* := (x_i^*)_{i\in I}$ asymptotically stable for the Kohonen algorithm will verify for all $i \in I$:

$$\sum_{k\in I}\Lambda\left(i-k\right)\int_{C_k(x)}\left(x_i-\omega\right)P\left(d\omega\right)=0$$

and it can match with a minimum of the limit distorsion only if

$$\begin{array}{l}\frac{1}{2}\sum_{j\in I}\sum_{k\in I,k\neq i}\left(\Lambda\left(k-j\right)-\Lambda\left(i-j\right)\right)\\ \times\int_{\bar{C}_k(x)\cap\bar{C}_i(x)}\left\|x_j-\omega\right\|^2\left(\frac{1}{2}\overrightarrow{n}_x^{ki}+\frac{1}{\|x_k-x_i\|}\times\left(\frac{x_i+x_k}{2}-\omega\right)\right)f\left(\omega\right)\lambda_x^{ki}d\omega\end{array}=0$$

but, in general, this term is not null.

4 Example of a Kohonen string with 3 centroïds

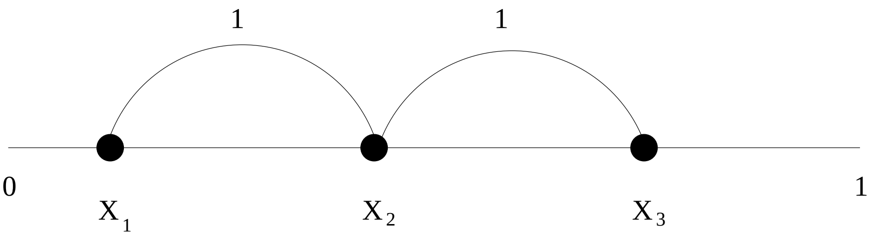
The previous section has shown that the minimum of extended variance doesn't match the equilibrium of the Kohonen algorithm. We will illustrate this with a simple example. The classical explanation (see Kohonen [5]) of local potential minimization by the Kohonen algorithm is far from being satisfactory. Actually it seems that the minimum of the distorsion measure always occurs on a discontinuity point, where the function is not derivable.
To illustrate this, let a Kohonen string be on segment $[0, 1]$ (see figure 1), with a discrete neighborhood

4.1 The theoretical difference

The equilibrium of SOM algorithm is reached on points x verifying

$$\begin{array}{l}\frac{\partial V}{\partial x_1}\left(x\right)=\int_{C_1(x)}\left(x_1-\omega\right)P\left(d\omega\right)+\int_{C_2(x)}\left(x_1-\omega\right)P\left(d\omega\right)=0\\ \frac{\partial V}{\partial x_2}\left(x\right)=\int_{C_1(x)}\left(x_2-\omega\right)P\left(d\omega\right)+\int_{C_2(x)}\left(x_2-\omega\right)P\left(d\omega\right)+\int_{C_3(x)}\left(x_2-\omega\right)P\left(d\omega\right)=0\\ \frac{\partial V}{\partial x_3}\left(x\right)=\int_{C_2(x)}\left(x_3-\omega\right)P\left(d\omega\right)+\int_{C_3(x)}\left(x_3-\omega\right)P\left(d\omega\right)=0\end{array}$$

Figure 1: Kohonen string



but the minima of the extended variance are reached on points x verifying

$$\begin{array}{l}
\frac{\partial V}{\partial x_1}(x) = \int_{C_1(x)} (x_1-\omega) P(d\omega) + \int_{C_2(x)} (x_1-\omega) P(d\omega) - \frac{1}{4}\left\|x_3 - \frac{x_1+x_2}{2}\right\|^2 f\left(\frac{x_1+x_2}{2}\right) = 0 \\
\frac{\partial V}{\partial x_2}(x) = \int_{C_1(x)} (x_2-\omega) P(d\omega) + \int_{C_2(x)} (x_2-\omega) P(d\omega) + \int_{C_3(x)} (x_2-\omega) P(d\omega) \\
-\frac{1}{4}\left\|x_3 - \frac{x_1+x_2}{2}\right\|^2 f\left(\frac{x_1+x_2}{2}\right) + \frac{1}{4}\left\|x_1 - \frac{x_3+x_2}{2}\right\|^2 f\left(\frac{x_3+x_2}{2}\right) = 0 \\
\frac{\partial V}{\partial x_3}(x) = \int_{C_2(x)} (x_3-\omega) P(d\omega) + \int_{C_3(x)} (x_3-\omega) P(d\omega) + \frac{1}{4}\left\|x_1 - \frac{x_2+x_3}{2}\right\|^2 f\left(\frac{x_2+x_3}{2}\right) = 0
\end{array}$$

If we assume, for example, that the density of observations is uniform $\mathcal{U}_{[0;1]}$ then these two sets of points have no point in common. Indeed, if the two sets are equal then

$$\left\{\begin{array}{l} x_3 - \frac{x_1+x_2}{2} = 0 \\ x_1 - \frac{x_2+x_3}{2} = 0 \end{array}\right.$$

Therefore, $x_1 = x_2 = x_3$, but this point is clearly not an equilibrium of the Kohonen map.

4.2 Illustration of the behavior of the distorsion measure

We will see that if you draw data with a uniform distribution on the segment [0, 1] and then you compute the minimum of the extended variance, then these minimum is always on a discontinuity point. The more observations you have, the more discontinuities you have but the global function looks more and more regular. This is not surprising since we know that the limit is derivable.

4.2.1 The method of simulation

Since we have no numerical algorithm to compute the exact minimum of variance, we proceed by exhaustive research based on a discretization of the space of the centroïds. To avoid too much computation, the step of discretization is chosen as 0.001. The following figures are obtained in the following way :

1. Simulate n "data" $(\omega_1, \cdots, \omega_n)$, chosen with an uniform law on $[0, 1]$.
2. Search exhaustively, on the discretization of D_I, the string minimizing the extended variance.
3. For the best string (x_1^*, x_2^*, x_3^*) the graphical representations are obtained in the following way :

- 3D Representation : we keep one centroïd in the triplet (x_1^*, x_2^*, x_3^*), then we move the other around a small neighborhood of its optimal position. The level z is the extended variance multiplied by the number of observations n.
- 2D Representation : we keep two centroïds in the triplet (x_1^*, x_2^*, x_3^*), then we move the last one around a small neighborhood of it optimal position. The level z is the extended variance multiplied by the number of observations n.

The following figures show the results obtained for a number of observations n varying from 10, 100 and 1000. We notice that, even for a small number of observations, the minima are always on discontinuity points.

Figure 2: Distorsion measure for 10 observations

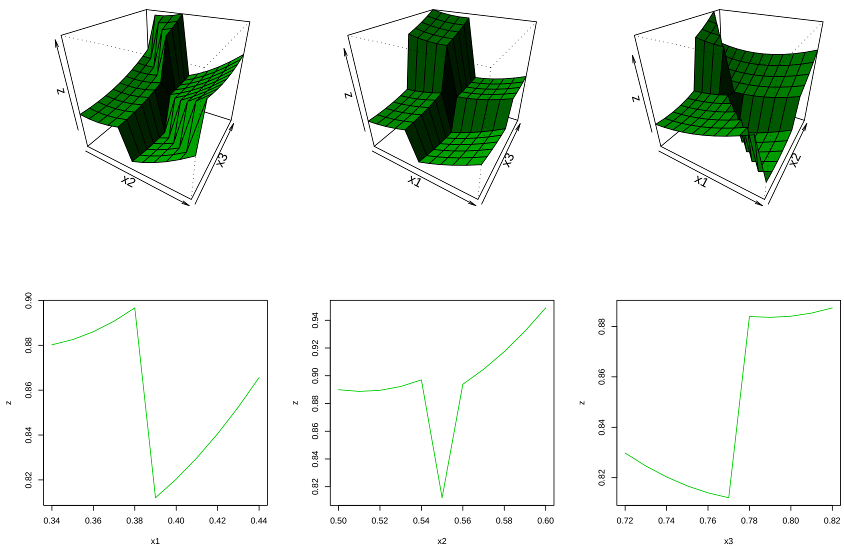


Figure 3: Distorsion measure for 100 observations

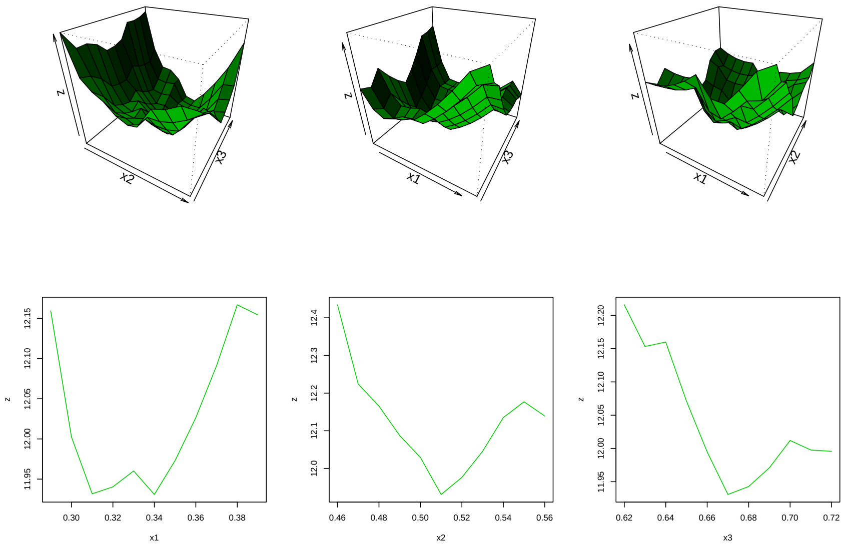
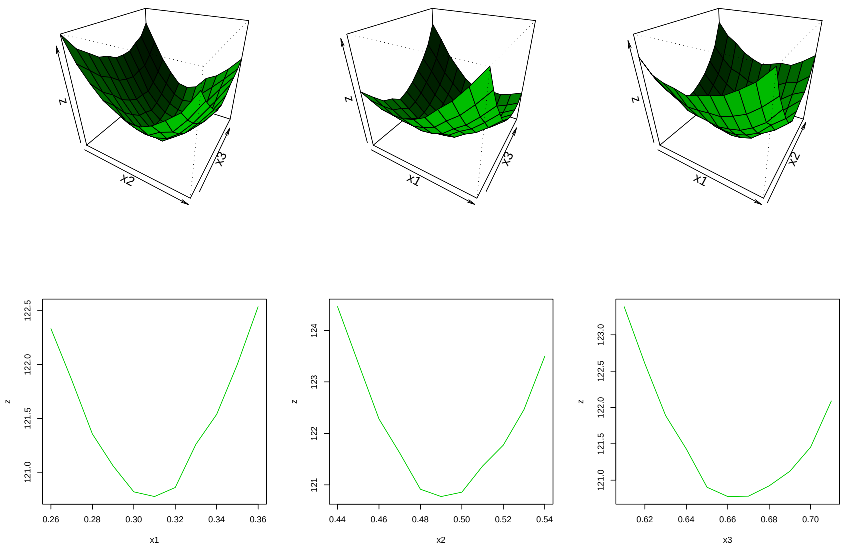


Figure 4: Distorsion measure for 1000 observations



5 Conclusion

The Kohonen algorithm was supposed to give an approximation of the minimum of the distorsion measure, but if it was the case, then why can the points of equilibrium of the algorithm be different from the theoretical minimum of distorsion? We have shown on an exemple that in discrete cases the minimum is reached on discontinuity points, so the local derivability of the distorsion measure does not seem to be an important property and is not a satisfactory explanation for the behavior of the Kohonen algorithm when the number of observations is finite.

References

[1] M. Cottrell, J.C. Fort, and G. Pags. Theoretical aspects of the SOM algorithm. *Neurocomputing*, 21:119–138, 1998.

[2] E. Erwin, K. Obermayer, and K. Schulten. Self-Organizing Maps : Ordering, Convergence properties and Energy Functions. *Biol. Cyb.*, 67:47–55, 1992.

[3] Jean-Claude Fort and Gilles Pagès. On the A.S. convergence of the Kohonen algorithm with a general neighborhood function. *The Annals of Applied Probability*, 5:4:1177–1216, 1995.

[4] T. Graepel, M. Burger, and K. Obermayer. Phase transition in stochastic self organizing maps. *Physical Review*, E(56):3876–3890, 1997.

[5] T. Kohonen. *Self-Organizing Maps*, volume 30 of *Springer Series in Information Sciences.* Springer, 1995.

[6] T. et. al. Kohonen. *Artificial neural networks,* **Vol. 2**. North Holland, 1991.