Topology Preservation in Self-Organizing Feature Maps: Exact Definition and Measurement

Thomas Villmann, Ralf Der, Michael Herrmann, and Thomas M. Martinetz

Abstract— **The neighborhood preservation of self-organizing feature maps like the Kohonen map is an important property which is exploited in many applications. However, if a dimensional conflict arises this property is lost. Various qualitative and quantitative approaches are known for measuring the degree of topology preservation. They are based on using the locations of the synaptic weight vectors. These approaches, however, may fail in case of nonlinear data manifolds. To overcome this problem, in this paper we present an approach which uses what we call the induced receptive fields for determining the degree of topology preservation. We first introduce a precise definition of topology preservation and then propose a tool for measuring it, the topographic function. The topographic function vanishes if and only if the map is topology preserving. We demonstrate the power of this tool for various examples of data manifolds.**

Index Terms—**Feature maps, Kohonen map, neural networks, topology.**

I. Introduction

Based on a lattice A of N neural units $i \in A$, Kohonen's self-organizing feature map algorithm (SOFM) is able to form a topology preserving map $\mathcal{M}_A$ of a data manifold $M \subseteq \Re^d$ [1]. This property can be employed in a variety of information processing tasks, ranging from speech and image processing over robotics to data reduction and knowledge processing [2]–[10]. To each neural unit $i \in A$ a reference or synaptic weight vector $w_i \in \Re^d$ is assigned. The map $\mathcal{M}_A = (\Psi_{A\to M}, \Psi_{M\to A})$ of M formed by A is then defined by the mapping $\Psi_{M\to A}$ from M to A and the inverse mapping $\Psi_{A\to M}$ from A to M. These two mappings are determined by

$$\mathcal{M}_A = \begin{cases} \Psi_{M\to A} : M \to A; & v \in M \mapsto i^*(v) \in A \\ \Psi_{A\to M} : A \to M; & i \in A \mapsto w_i \in M \end{cases} \quad (1)$$

with $i^*(v)$ as the neural unit with its synaptic weight vector $w_{i^*(v)}$ closest to v, i.e., with

$$\|w_{i^*(v)} - v\| \leq \|w_j - v\| \quad \forall j \in A. \quad (2)$$

Manuscript received September 3, 1994; revised October 29, 1995 and October 20, 1996. This work was sponsored by the the project "LADY" of the German Federal Ministry of Research and Technology (BMFT) under Grant 01 IN 106B/3 and by the project "NERES" sponsored by the German Federal Ministry of Research and Technology (BMFT) under Grant 01 IN 102 A7.

T. Villmann and R. Der are with the Institute of Informatics, University of Leipzig, D-04109 Leipzig, Germany.

M. Herrmann is with Nordita, DK-2000 København, Denmark.

T. A. Martinetz is with Siemens AG, Corporate Research and Development, D-81730 München, Germany.

Publisher Item Identifier S 1045-9227(97)01749-9.

Starting from a fixed lattice structure A, Kohonen's self-organizing feature map algorithm distributes the synaptic weight vectors w_i such, that the map $\mathcal{M}_A$ of M formed by A is as topology preserving as possible. The reference vectors w_i are adapted in a learning step according to

$$\Delta w_i = \epsilon h_{i^*,i}(v - w_i) \quad \text{for all } i \in A \quad (3)$$

where $v \in M$ is the presented stimulus vector, $i^*(v)$ is defined again by (2), and the neighborhood function

$$h_{i^*,i} = \exp\left(-\frac{\|i^* - i\|_A^2}{2\sigma^2}\right) \quad (4)$$

determines the neighborhood range in A by the choice of the radius σ. $\|\cdot\|_A$ denotes the Euclidean distance in A. ϵ is the learning parameter.

To what degree the topology is preserved depends on the choice of the lattice structure A. Depending on the form of the manifold M, a one-dimensional, two-dimensional, etc. lattice has to be chosen to obtain the best result. In most applications, however, the form of the manifold M is not known and, hence, it is not clear *a priori* which lattice structure one should choose. One has to try different lattice structures and determine somehow which lattice yields the highest degree of topology preservation.

Various qualitative and quantitative methods for characterizing the degree of topology preservation have been proposed [11]–[17]. All these approaches are based on the evaluations of the position of the neurons in the lattice and, on the other hand, on the evaluations of the position of their weight vectors only. However, none of these approaches take the form of the data manifold into the measurement. This can not provide correct results in the case of nonlinear submanifolds $M \subseteq \Re^d$. Figs. 1 and 2 show examples of a linear and a nonlinear data manifold, respectively. In both the linear and the nonlinear case the positions of the reference vectors are identical. Hence, none of the above approaches can distinguish a correct folding due to the folded nonlinear data manifold from a folding due to a topological mismatch between M and A as in the linear case, because only the weight vectors are considered. In particular, when using the SOFM for nonlinear principle component analysis (PCA) one has to have a means to differentiate between these two cases.

In our paper we give a new approach for quantifying topology preservation which explicitly takes the structure of the data manifold into account. This approach, which employs what we call the *topographic function*, can be applied to linear

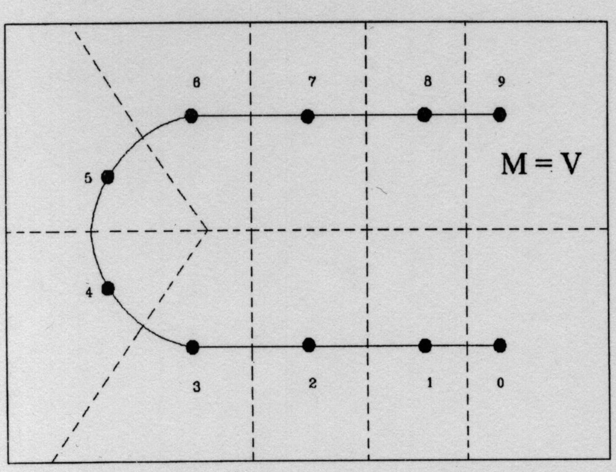


Fig. 1. Example of a linear ($M = V$) data manifold together with the hypothetical positions of the images of the neural units. The induced receptive fields are drawn by dashed lines.

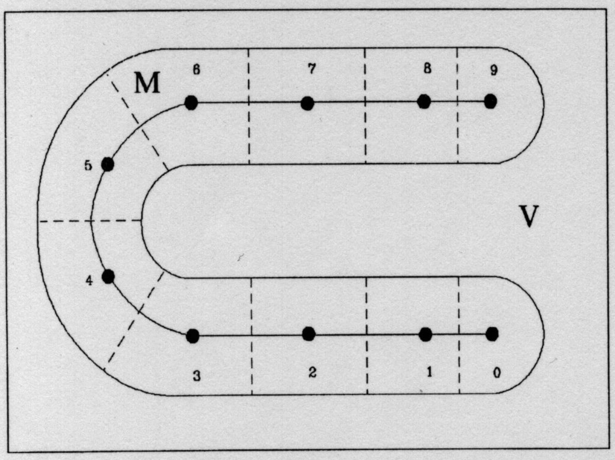


Fig. 2. Example of a nonlinear ($M \subset V$) data manifold together with the hypothetical positions of the images of the neural units. The induced receptive fields are drawn by dashed lines.

and nonlinear data manifolds M and, further, allows us to quantify the range of the folds. This paper is structured as follows.

In Section II we introduce a mathematical definition of topology preservation for rectangular lattices A, which in Section III leads to the so-called topographic function as a measure for the degree of topology preservation of a map $\mathcal{M}_A$ of M. It is shown how the topographic function can be evaluated by a simple mechanism based on the "competitive Hebbian rule," which was introduced in [18], [19]. In Section IV we demonstrate via examples, ranging from the logistic map over speech data to satellite images, the potential of the topographic function not only as a measure for the degree of topology preservation but as a general means for obtaining information about the dimensionality and structure of the data manifold M. The results are compared with the results provided by the so-called topographic product, which was proposed by [11] as an alternative method for measuring the degree of topology preservation. In the last section, we show how the mathematical definition of topology preservation for rectangular lattices can be extended to more general lattice structures.

II. A Definition of Topology Preservation for Rectangular Lattices

We want to call a map $\mathcal{M}_A = (\Psi_{A\to M}, \Psi_{M\to A})$ of M "topology preserving," if both the mapping $\Psi_{M\to A}$ from M to A as well as the inverse mapping $\Psi_{A\to M}$ from A to M is neighborhood preserving. Hence, to determine whether a SOFM is topology preserving we have to measure these two neighborhood preservations. Yet, it is not clear what in general neighborhood preservation means [12]. Therefore, to be able to measure these two neighborhood preservations we first have to define them. Per definition we regard the mapping $\Psi_{M\to A}$ from M to A as being neighborhood preserving if reference vectors w_i, w_j, which are adjacent on M, belong to vertices i, j, which are neighbors in A. On the other hand, the inverse mapping $\Psi_{A\to M}$ from A to M is neighborhood preserving if adjacent vertices i, j are mapped onto locations w_i, w_j which are neighbors on M. How can we define *neighborhood of vertices* i, j *in* A and *neighborhood of reference vectors* w_i, w_j *on* M in a way that the intuitive understanding of topology preservation of a SOFM is captured?

A reasonable definition for *neighborhood of reference vectors* w_i, w_j *on* M was given in [18] and [19] based on the masked Voronoi polyhedra of w_i and w_j. Two *synaptic weight vectors* w_i, w_j *are adjacent on* M if and only if their receptive fields R_i, R_j on M, determined by the masked Voronoi polyhedra $R_i = \tilde{V}_i, R_j = \tilde{V}_j$ with

$$\tilde{V}_i = \{v \in M \mid \|v - w_i\| \le \|v - w_j\| \quad \forall j \in A\} \tag{5}$$

are adjacent, i.e., if and only if $R_i \cap R_j \neq \emptyset$. We notice that the definition of adjacency by the nonvanishing intersections makes sense, because the $\tilde{V}_i$ are defined as closed sets.

Two vertices $i = (i_1, \cdots, i_{d_A}), j = (j_1, \cdots, j_{d_A})$ of a rectangular d_A-dimensional lattice are adjacent in A if and only if they are nearest neighbors in the lattice A. Obviously, this demands vertices which are adjacent in the lattice according to the Euclidean norm $\|\cdot\|_E$ or the summation-norm

$$\|\cdot\|_\Sigma \overset{\text{def}}{=} \sum_{j=1}^{d_A} |(\cdot)_j| \tag{6}$$

to be assigned to neighboring locations w_i. However, a proper definition of the two neighborhood preservations requires us to take into account a second neighborhood. This is illustrated in Figs. 3 and 4, which show a two-dimensional rectangular lattice representing a square in an ideal state and with small distortions, respectively. To be able to discern the adjacency of reference vectors w_i their receptive fields, i.e., their masked Voronoi polygons $\tilde{V}_i$, are depicted. Of course, the maps in both Figs. 3 and 4 are topology preserving. This means, adjacent locations w_i have to belong to adjacent vertices. However this is not the case, at least not according to the Euclidean norm $\|\cdot\|_E$, as we can discern from Fig. 4. However, also in the ideal case of Fig. 3 one can find violations of topology preservation in the above sense: the receptive fields of vertices which are diagonal neighbors in A have a nonvanishing intersection, i.e.,

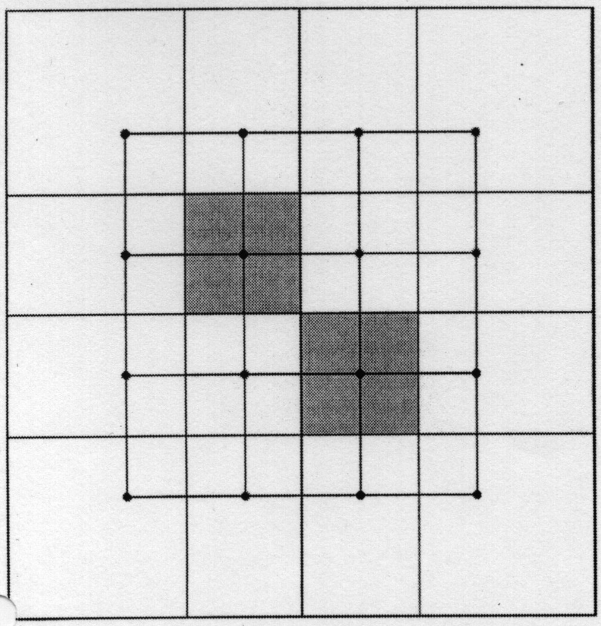


Fig. 3. Ideal mapping of a squared input space onto a squared lattice of neural units. The locations w_i together with their receptive fields are shown. Two receptive fields are especially depicted, the vertices of which lie diagonally in the lattice A. These receptive fields have a nonvanishing intersection of one point and, hence, are adjacent.

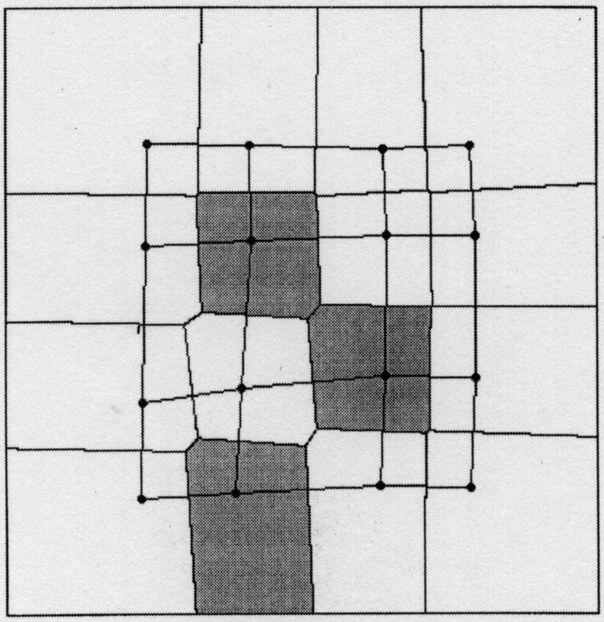


Fig. 4. Map with small distortions of a squared input space onto a squared lattice of neural units. The locations w_i together with their receptive fields are shown. Three receptive fields are depicted gray scaled, the vertices belonging to the locations of which lie diagonally in the lattice A (pair wise).

their locations are adjacent. Hence, this definition would be too strict. To overcome this problem we introduce an additional neighborhood in A, a neighborhood based on the maximum-norm

$$\|\cdot\|_{\max} \stackrel{\text{def}}{=} \max_{j=1}^{d_A} |(\cdot)_j| \tag{7}$$

of the vertices. Now, in Fig. 4 all adjacent reference vectors w_i are mapped onto vertices which are neighbors according to the new neighborhood defined by $\|\cdot\|_{\max}$. On the other hand, if one tries to use this distance measure (7) also to determine the degree of topology preservation of the map $\Psi_{A\to M}$, it will fail, as we can see in Fig. 4. Therefore, here we have to take the Euclidean distance in A.

This leads us to the following definition of topology preservation of SOFM's with rectangular lattices.

Definition 1: Let A be a d_A-dimensional rectangular lattice and M be a data manifold $M \subseteq \Re^d$. A map $\mathcal{M}_A = (\Psi_{A\to M}, \Psi_{M\to A})$ of M is topology preserving if both the mapping $\Psi_{M\to A}$ from M to A and the inverse mapping $\Psi_{A\to M}$ from A to M is neighborhood preserving.

1) The mapping $\Psi_{M\to A}$ is neighborhood preserving if and only if locations w_i, w_j which are adjacent on M belong to vertices i, j which are adjacent in A according to the **maximum-norm** $\|\cdot\|_{\max}$ on A.
2) The mapping $\Psi_{A\to M}$ is neighborhood preserving if and only if vertices i, j which are adjacent in A according to the **Euclidean norm** $\|\cdot\|_E$ or the **summation-norm** $\|\cdot\|_\Sigma$ on A are assigned to neighboring locations $w_i, w_j \in M$.

These definitions of neighborhood preservation are valid for the special but most widespread case of SOFM's based on rectangular lattices. How they can be enhanced to more general lattice structures is shown in Section V.

III. The Topographic Function Φ_A^M

A. Definition of the Topographic Function Φ_A^M

Let A be a $N_1 \times N_2 \times \cdots \times N_{d_A}$ rectangular lattice of dimension d_A. The lattice consists of $N = N_1 \times N_2 \times \cdots \times N_{d_A}$ neural units, and each unit i is indicated by $i = (i_1, \cdots, i_{d_A})$. To each i a reference vector is assigned which maps i onto a location w_i on the given data manifold M. As has been outlined in [18] and [19], the masked Voronoi polyhedra in (5) define the so-called induced Delaunay triangulation $\mathcal{D}_M$ of the set of w_i. The induced Delaunay triangulation is the graph which connects those and only those w_i, w_j which have adjacent masked Voronoi polyhedra $\tilde{V}_i, \tilde{V}_j$, i.e., which have adjacent receptive fields R_i, R_j on M. The induced Delaunay triangulation $\mathcal{D}_M$ defines a distance metric $\|\cdot\|_{\mathcal{D}_M}$ between the w_i. The distance

$$d_{\mathcal{D}_M}(i,j) \stackrel{\text{def}}{=} \|w_i - w_j\|_{\mathcal{D}_M} \tag{8}$$

between two reference vectors is then determined by their shortest distance within the graph $\mathcal{D}_M$. Hence, two reference vectors w_i, w_j are adjacent on M according to our definition in Section II if and only if they are nearest neighbors in $\mathcal{D}_M$, i.e., if and only if $d_{\mathcal{D}_M}(i,j) = 1$. We can now define the *topographic function* Φ_A^M which is able to measure the topology preservation of a SOFM according to our definition.

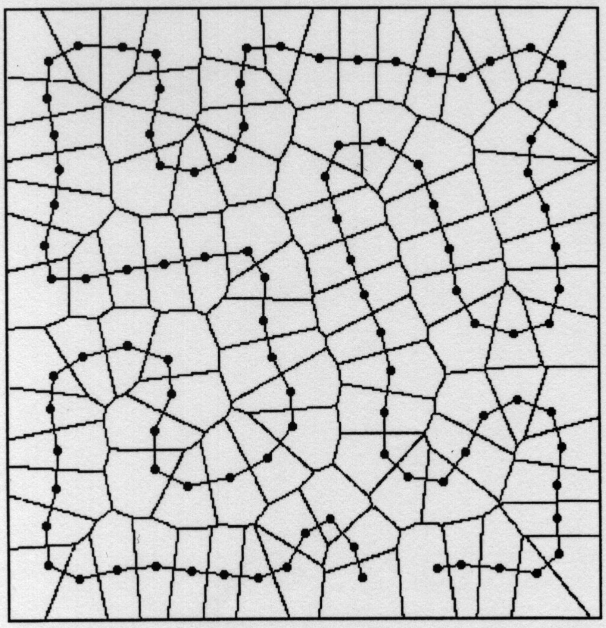


Fig. 5. Plot of a mapping of a squared input space onto a chain of 100 neural units. The receptive fields of the units are shown.

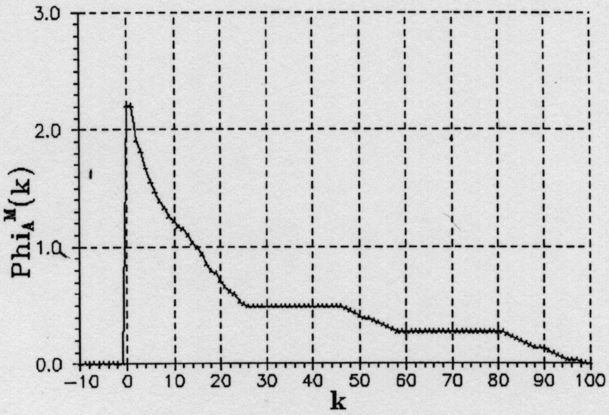


Fig. 6. The topographic function of a mapping of a squared input space onto a chain of 100 neural units.

For each neural unit i we define

$$\begin{aligned} f_i(k) &\stackrel{\text{def}}{=} \#\{j \mid \|i-j\|_{\max} > k;\ d_{\mathcal{D}_M}(i,j) = 1\} \\ f_i(-k) &\stackrel{\text{def}}{=} \#\{j \mid \|i-j\|_E = 1;\ d_{\mathcal{D}_M}(i,j) > k\} \end{aligned} \tag{9}$$

with $k = 1, \cdots, N-1$. $\#\{\cdot\}$ denotes the cardinality of a set. Looking at a neural unit $i, f_i(k)$ measures the neighborhood preservation of $\Psi_{M\to A}$, and $f_i(-k)$ measures the neighborhood preservation of $\Psi_{A\to M}$, as they have been defined in Section II. The topographic function of the map $\mathcal{M}_A = (\Psi_{A\to M}, \Psi_{M\to A})$ is then defined by

$$\Phi_A^M(k) \stackrel{\text{def}}{=} \begin{cases} \frac{1}{N}\sum_{j\in A} f_j(k) & k > 0 \\ \Phi_A^M(1) + \Phi_A^M(-1) & k = 0 \\ \frac{1}{N}\sum_{j\in A} f_j(k) & k < 0. \end{cases} \tag{10}$$

We obtain $\Phi_A^M \equiv 0$ and, particularly, $\Phi_A^M(0) = 0$ if and only if the SOFM is perfectly topology preserving.

The largest $k^+ > 0$ for which $\Phi_A^M(k^+) \neq 0$ holds yields the range of the largest fold if the effective dimension of the data manifold M is larger then the dimension d_A of the lattice A. This is depicted in Figs. 5 and 6. Fig. 5 shows a map of a squared data manifold onto a chain of 100 neural units, together with their receptive fields. The folds involve the whole chain and, hence, the topographic function vanishes only for k-values larger then $k^+ = 98$, as can be seen in Fig. 6. On the other hand, the smallest $k^- < 0$ for which $\Phi_A^M(k) \neq 0$ holds yields the range of violations of topology preservation if the effective dimension of the data manifold M is smaller than the dimension d_A of the lattice A. In this way the values k^+ and k^- give information about the degree of the dimensional conflict. Small values of k^+ and k^- indicate that there are only local dimensional conflicts, whereas large values indicate the global character of the dimensional conflict.

To compare lattices with different structures and number of neural units it is useful to introduce a normalization of the k-values so that

$$k \mapsto k_A \in [-1, 1] \tag{11}$$

holds.

B. Evaluating the Topographic Function Φ_A^M

Calculating Φ_A^M requires to determine the induced Delaunay triangulation $\mathcal{D}_M$. A way to determine $\mathcal{D}_M$ has been proposed in [18] and [19]. Let $\mathbf{C}$ be a connectivity matrix which determines connections between units $i, j \in A$ (in addition to the connectivity matrix defined by the fixed lattice structure). Initially, the elements $\mathbf{C}_{ij} \in \{0, 1\}$ of $\mathbf{C}$ are set to zero. Furthermore, we assume that the set $\mathcal{W} = \{w_1, \cdots, w_N\}$ of all the reference vectors is dense in M, i.e., if the reference vectors w_{i^*}, w_{j^*} are closest and second closest to an arbitrary given $v \in M$ then the triangle $\Delta(v, w_{i^*}, w_{j^*})$ lies completely in M. Then it can be shown that simply by sequentially presenting input vectors $v \in M$ and each time connecting those two units i^*, j^* (setting $\mathbf{C}_{i^*j^*} = 1$) the reference vectors w_{i^*}, w_{j^*} of which are closest and second closest to v, the connectivity matrix $\mathbf{C}$ converges to

$$\lim_{t\to\infty} \mathbf{C}_{ij} = 1 \quad \Leftrightarrow \quad R_i \cap R_j \neq \emptyset \tag{12}$$

[18], [19]. After a sufficient number of input vectors v have been presented, the connectivity matrix $\mathbf{C}$ connects units and only units i, j the receptive fields $R_i = \tilde{V}_i$, $R_j = \tilde{V}_j$ of which are adjacent and, hence, defines the induced Delaunay triangulation $\mathcal{D}_M$. With this algorithm we obtain the following scheme for determining the graph structure $\mathbf{C}$ of the Delaunay triangulation.

1) present an input vector $v \in M$;
2) determine the nearest reference vector w_{i^*} and the second nearest reference vector w_{j^*};
3) connect the units i^*, j^*, i.e., set $\mathbf{C}_{i^*j^*} := 1$;
4) go to step 1.

With this connectivity matrix $\mathbf{C}$ we are able to determine the distances $d_{\mathcal{D}_M}(i,j)$ between the reference vectors [20], which then allows us to calculate Φ_A^M according to (9) and (10).

In practice it is possible to determine the matrix $\mathbf{C}$ in a parallel way to the learning algorithm for the weights. Therefore we introduce in addition to the usual algorithm a suitable chosen maximal age $a_{\max}$ of a connection $\mathbf{C}_{ij}$. For instance, we choose $a_{\max} = N/p$ where N is the number of neural units and p is $p = \int_{v \in M} vP(v)dv$ for all $v \in V$. In every time step the age a_{i^*j} of all existing connections of the best matching unit i^* is increased. The age $a_{i^*j^*}$ of the new connection $\mathbf{C}_{i^*j^*}$ of the two best matching units is set to zero. Connections $\mathbf{C}_{i^*j}$ with an age higher than the maximal age $a_{\max}$ will be removed. This approach was first used in [18] and [19] to compute the Delaunay triangulation of a "neural-gas-network" parallel to the evaluation of the net. With this algorithm we obtain the following scheme for determining Φ_A^M:

1) present an input vector $v \in M$;
2) determine the nearest reference vector w_{i^*} and the second nearest reference vector w_{j^*};
3) increase the age a_{i^*j} for all j for which $\mathbf{C}_{i^*j} = 1$ holds;
4) connect the units i^*, j^*, i.e., set $\mathbf{C}_{i^*j^*} := 1$ and $a_{i^*j^*} := 0$;
5) set $\mathbf{C}_{i^*j} := 0$ if $a_{i^*j} > a_{\max}$;
6) go to step 1.

At least we have to discuss what a sufficient number of input vectors for the computation of the connectivity matrix $\mathbf{C}$ is. For obtaining a rough lower bound we investigate a rectangular network A of N neurons with the dimension d_A and a homogen data distribution in $M \subseteq \Re^d$. Then each (inner) neuron possess $3^{d_A} - 1$ lattice neighbors. If we are only interested in to determine whether topology defects occur or not, at least $c_{\min} \cdot N \cdot (3^{d_A} - 1)$ inputs are necessary with $c_{\min}$ has to be range greater the 10^1. If the scale of the topology defects is also of interest all possible connections have to be considered. Then this lower band increases to $c_{\min} \cdot N^2 \cdot (3^{d_A} - 1)$. Using the above parallel approach for determining the connectivity matrix these lower bounds in the reality are not restrictive.

IV. Applications of the Topographic Function and Comparison with the Topographic Product

Kohonen's algorithm defines a SOFM from a data manifold M embedded in a d-dimensional input space $\Re^d$ onto a d_A-dimensional lattice A of neural units. A method for quantifying the topology preservation of a SOFM is the topographic product P which was introduced by Bauer and Pawelzik [11]. It measures the neighborhood preservation of the mapping from the neural units i in A onto their reference vectors w_i. Thereby however, the topographic product does not take into account the shape of M, but considers only the neighborhood relations of the reference vectors within the embedding space V. Hence, an approach based on the topographic product is not able to differentiate between correct foldings arising from a nonlinear data manifold M and incorrect foldings which may result from a dimensional conflict between M and A or an incorrect formation of the map (topological defects, twists, kinks). An example which illustrates the problem is shown in Figs. 1 and 2. For both the linear and nonlinear M the topographic product has the same value indicating a loss of topology preservation. However, in the nonlinear case the map has been formed correctly.

As mentioned in the introduction all other known measures also take only the position of the weight vectors into account. Hence, the above described problem is not specific for the topographic product. However, because the topographic product is most widely spread we compare the topographic function only with this measure.

For a better understanding we briefly introduce the topographic product. For a detailed description see [11]. The topographic product considers all orders of neighborhood. For each neuron i the sequences $n_k^A(i)$ and $n_k^M(i)$ have to be determined, where $n_k^A(i)$ denotes the kth neighbor of i, with Euclidean distance measured in A, and $n_k^M(i)$ denotes the kth neighbor of i, with Euclidean distance evaluated in M between w_i and $w_{n_k^M(i)}$. From these sequences, the intermediate quantity

$$P_3(i,k) = \left(\prod_{l=1}^{k} \frac{d^M\left(w_i, w_{n_l^A(i)}\right)}{d^M\left(w_i, w_{n_l^M(i)}\right)} \cdot \frac{d^A\left(i, n_l^A(i)\right)}{d^A\left(i, n_l^M(i)\right)} \right)^{\frac{1}{2k}} \tag{13}$$

is computed. Averaging over all neighborhood orders k and neurons i leads to the topographic product

$$P = \frac{1}{N(N-1)} \sum_i \sum_{k=1}^{N-1} \log(P_3(i,k)). \tag{14}$$

The topographic product P can take on positive or negative values, which have to be interpreted as follows: In the linear case we get the following values for P, depending on whether d_A is smaller, equal to or larger then d [11]

$$\begin{aligned} P &< 0 \quad \text{for } d_A < d \\ P &\approx 0 \quad \text{for } d_A = d \\ P &> 0 \quad \text{for } d_A > d. \end{aligned} \tag{15}$$

For instance, if we map a squared input space onto a lattice of 16×16 neural units, we obtain $P = 0.0005$.

We applied both the topographic function Φ_A^M and the topographic product P to various examples of linear and nonlinear data manifolds. In the linear cases both approaches gave the same results. In the nonlinear cases, however, only the topographic function provided correct results. If a real dimensional conflict occurs, i.e., the map "folds" itself into the input space, the topographic function indicates this situation. Further, it is even able to measure the scale of existing folds. As an example we consider again the mapping from a squared input space onto a chain of 100 neural units. In this case the chain folds itself into the input space like a Peano curve, as shown in Fig. 5. The topographic function reveals the various length-scales of the folds. The highest k-value K^* for which $\Phi_A^M(k) \neq 0$ indicates the longest range. For our example we find $K^* = 98$ (see Fig. 6), i.e., the range includes nearly the whole chain. The topographic product yields $P = -0.107$ which also indicates the dimensional conflict.

To demonstrate the difference between the topographic product and the topographic function in the case of nonlinear

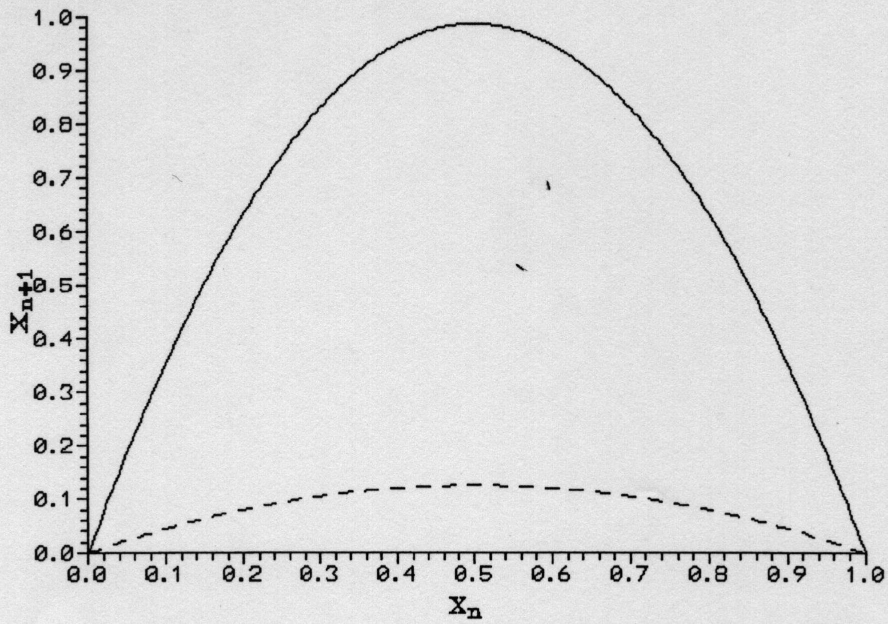


Fig. 7. Two examples of a plot of pairs $(x_n, x_{n+1}) \in M_\lambda$ of the logistic mapping with different values for the nonlinearity-parameter $\lambda = 0.5$ and $\lambda = 3.95$. The maximum of the x_{n+1}-values is $\frac{\lambda}{4}$.

data manifolds we first investigate the logistic map

$$x_{n+1} = x_n\lambda(1 - x_n). \tag{16}$$

The states (x_n, x_{n+1}) of the system form a nearly linear submanifold M_λ for small values of λ, but a nonlinear one otherwise (see Fig. 7). For various cases of λ we trained a chain of 64 neural units to represent M_λ. We computed both the topographic product and the topographic function and obtained for each λ the result $\Phi_A^M \equiv 0$. The topographic product P, however, decreased with increasing λ

$$\begin{aligned} \lambda &= 0.50 : P = 0.0009 \\ \lambda &= 3.00 : P = -0.002 \\ \lambda &= 3.95 : P = -0.015. \end{aligned} \tag{17}$$

The negative values of P indicate an increasing dimensional conflict [see (15)]. The submanifold M_λ, however, is always one-dimensional; i.e., there is actually no dimensional conflict.

As second example we take the twice iterated logistic map, i.e.,

$$x_{n+2} = \lambda^2 x_n(1 - x_n)(1 - \lambda x_n(1 - x_n)). \tag{18}$$

The submanifold is now generated by the states (x_n, x_{n+2}) of the system (see Fig. 8). For $\lambda = 3.00$ and $\lambda = 3.95$ we obtain $P = -0.007$ and $P = -0.06$, respectively. The topographic function, however, vanishes in both cases, as it should.

Now we investigate a more realistic example. The satellites of LANDSAT-TM type produce pictures of the earth in seven different spectral bands. The ground resolution in meter is 30 × 30 for the bands 1–5 and band 7. Band 6 has a resolution of 60 × 60 only. The spectral bands represent useful domains of the whole spectrum in order to detect and discriminate vegetation, water, rock formations and cultural features [21], [22]. The spectral information, i.e., the intensity of the bands associated with each pixel of a LANDSAT scene is represented by a vector $v \in \Re^d$ with $d = 7$, the number of spectral bands. Because of the rougher resolution of band 6 (thermal band) this channel is often dropped. Hence, the LANDSAT data may be represented as clouds of data points in a six-dimensional space. The aim of any classification algorithm is to subdivide this data space into subsets of data points which belong to a certain category corresponding to a specific feature like wood, industrial region, etc., each feature being specified by a certain prototype data vector. An approach with self-organizing feature maps has been successfully applied in meteorology (cloud detection) [23] and earth surface clustering of Kuwait [24].

One way to get good results for visualization is to use a SOFM dimension $d_A = 3$. Then we are able to interpret the positions of the neurons in the *three*-dimensional neuron lattice A as a vector $c = (r, g, b)$ in the color space $\mathcal{C}$, where r, g, b are the intensity of the colors red, green and blue. This assigns colors to categories (winner neurons) so that we end up immediately with the pseudo color version of the original picture for visual interpretation [25]. However, since we are mapping the data clouds from a six-dimensional input space onto a *three*-dimensional color space there may arise dimensional conflicts and the visual interpretation may fail. Therefore, we tested lattices with the dimensions $d_A = 1, 2, 3$ with 256, 16 × 16 and 7 × 6 × 6 neural units, respectively, to see whether they are topology preserving according to both the topographic product and the topographic function. The values for the topographic products are

$$\begin{aligned} d_A &= 1 : P = 0.0449 \\ d_A &= 2 : P = 0.0066 \\ d_A &= 3 : P = -0.0569 \end{aligned} \tag{19}$$

which suggests to prefer the two-dimensional configuration. However, the topographic function still indicates mismatches,

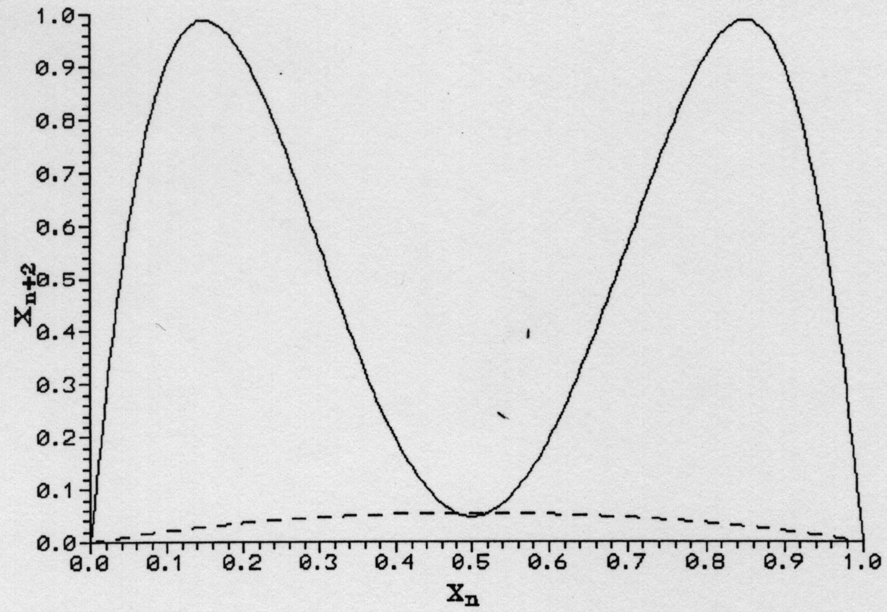


Fig. 8. Two examples of a plot of pairs $(x_n, x_{n+2}) \in M_\lambda$ of the twice-iterated logistic mapping with different values for the nonlinearity-parameter $\lambda = 0.5$ and $\lambda = 3.95$. The maximum of the x_{n+2}-values is $\frac{\lambda}{4}$ andthe local minimum for $x_n = 0.5$ has the value $\frac{\lambda^2(\lambda-4)}{16}$.

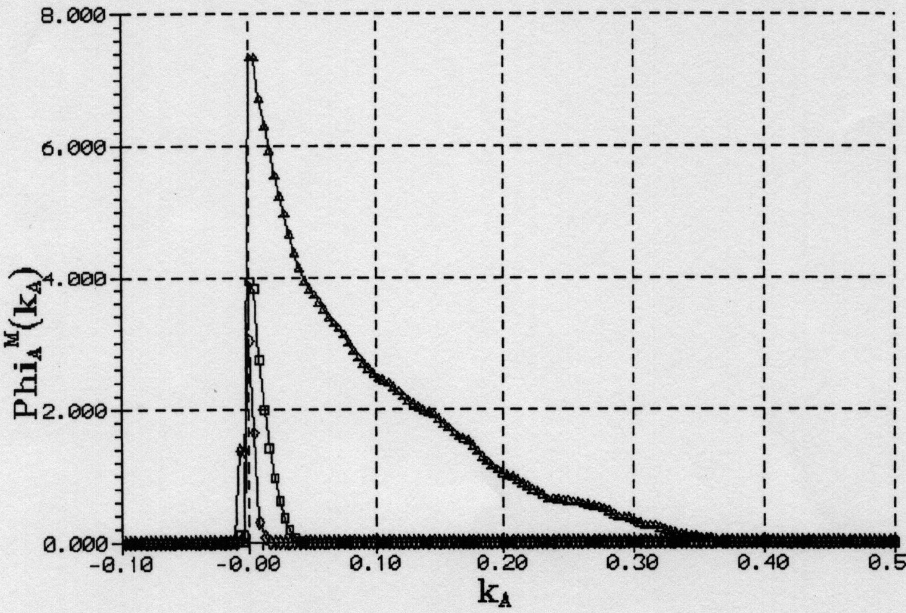


Fig. 9. The topographic functions obtained from a mapping of a six-dimensional LANDSAT TM satellite image onto a chain of 256 neural units (points as △), onto a squared lattice of 16 × 16 neural units (points as □) and onto a three-dimensional lattice of 7 × 6 × 6 neural units (points as ◇). The normalization (11) of the k-values was used.

as shown in Fig. 9 where the normalization (11) of the k-values was used. These results demonstrate that a visual interpretation of the results without a detailed consideration of the topology preserving property of the SOFM is misleading.

Finally, we discuss, again in comparison to the topographic product, the application of our approach to a set of speech data from the DPI-database of the III. Physikalisches Institut, Universität Göttingen, Germany. The data preprocessing is described in [26]. Here we remark only that the 4500 feature vectors represent a data submanifold lying in a 19-dimensional input space. We applied various SOFM's of lattice dimensions $d_A = 2, \cdots, 4$ to the data manifold. The topographic function for all cases is shown in Fig. 10. In agreement with [11] we obtain that in all cases the topology preservation of the map is not perfect because the high $\Phi_A^M(0)$-values. The $\Phi_A^M(0)$-values of the three-dimensional and the two-dimensional lattices are suggesting to take these, again in anology to the results obtained from the topographic product, for which the values

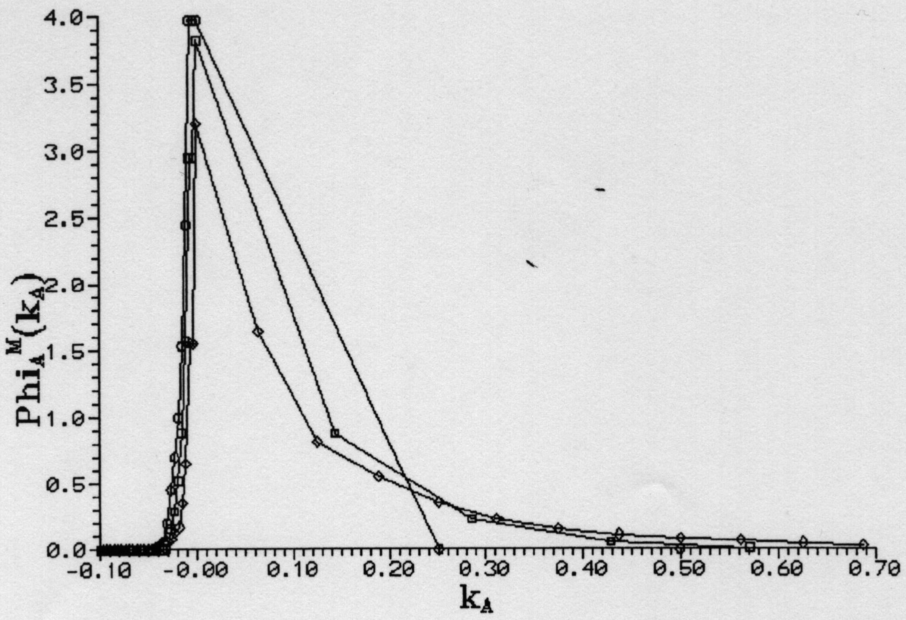


Fig. 10. The topographic functions obtained from a mapping of a set of speech data onto various lattices of neural units (two-dimensional—□, three-dimensional—◇, four-dimensional—○). The normalization (11) of the k-values was used.

$P_{2\,\mathrm{dim}} = -0.0282, P_{3\,\mathrm{dim}} = 0.0191, P_{4\,\mathrm{dim}} = 0.0367$ were obtained [27]. Other investigations of these speech data yield the effective dimension $d \approx 2.34$ [27], i.e., the modified topographic function obtains similar results.

V. A Definition of Topology Preservation for More General Lattices

In this section we extend our Definition 1 of topology preservation to the case of more general lattice structures. We now only assume A to be a network of N neurons which are situated at points $i = (i_1, \cdots, i_{d_A}) \in \Re^{d_A}$. The basic idea is to describe the property of topology preservation in terms of mathematical topology. The property of topology preservation of a map may then be based on the continuity of this map between topological spaces.

The induced Voronoi diagram $\mathcal{V}_M$ of a subset $M \subseteq \Re^d$ and its dual, the induced Delaunay graph (Voronoi graph) $\mathcal{D}_M$ with respect to a set $S = \{w_1, \cdots, w_N\}$ of points $w_i \in M \subseteq \Re^d$, is given by the masked Voronoi polyhedra

$$\tilde{V}_i = \{x \in M \mid \|x - w_i\| \le \|x - w_j\| \quad j = 1, \cdots, N,\ j \ne i\} \tag{20}$$

as shown in [18] and [19]. We remark that the Voronoi polyhedra are closed sets. The cells form a complete partitioning of M in the sense that $M = \bigcup_{i=1}^N \tilde{V}_i$. The induced Voronoi diagram $\mathcal{V}_M$ uniquely corresponds to its induced Delaunay graph $\mathcal{D}_M$ [28]. Two Voronoi cells $\tilde{V}_i, \tilde{V}_j$ are connected in $\mathcal{D}_M$ if and only if the intersection of it is nonvanishing, i.e., $\tilde{V}_i \cap \tilde{V}_j \ne \emptyset$ [18], [19], [28], [29]. Now we can define in $\mathcal{D}_M$ a graph metric as the minimal path length in the graph. In the general case the Voronoi diagram $\mathcal{V}$ of $\Re^d$ with respect to S is given by the Voronoi cells defined by

$$V_i = \{x \in \Re^d \mid \|x - w_i\| \le \|x - w_j\|\ j = 1, \cdots, N,\ j \ne i\}. \tag{21}$$

Using the concepts introduced above we are now able to define in general terms what topology preservation for arbitrary lattices A with the connectivity graph $\mathbf{C}^A$ means.

In analogy to Section III-A we define two kinds of neighborhood in the lattice A, but now as abstract topological definitions:

Definition 2: Suppose A to be a network of N neurons which are situated at points $i = (i_1, \cdots, i_{d_A}) \in \Re^{d_A}$ with reference or synaptic weight vectors $w_i \in M \subseteq \Re^d$. The connectivity graph $\mathbf{C}^A$ of A defines the structure of A. Let furthermore $\mathbf{C}^A(i)$ denote $\mathbf{C}^A$ where the neural unit i was taken as root. A (discrete) topology $\mathcal{T}_A^+(i)$ is induced by the graph metric in $\mathbf{C}^A(i)$. $\mathcal{T}_A^+(i)$ is said to be the **strong neighborhood topology in A with respect to** i, and $(A, \mathcal{T}_A^+(i))$ is a topological space.

Definition 3: Consider for the moment A to be a set of points in $\Re^{d_A}$. Let $\mathcal{V}$ be the Voronoi diagram of $\Re^{d_A}$ with respect to A and $\mathcal{D}_A$ be its dual, the Delaunay graph. Let furthermore $\mathcal{D}_A(i)$ denote $\mathcal{D}_A$ where the neural unit i was taken as root. $\mathcal{D}_A(i)$ is equipped with the graph metric that in turn induces the (discrete) topology $\mathcal{T}_A^-(i)_M$ in $\mathcal{V}$ and, hence, also in A. $\mathcal{T}_A^-(i)$ is said to be the **weak neighborhood topology in A with respect to** i, and $(A, \mathcal{T}_A^-(i))$ is a further topological space defined on the set A.

Remark 1: In the case of a rectangular lattice the weak neighborhood topology $\mathcal{T}_A^-(i)$ is weaker then the strong neighborhood topology $\mathcal{T}_A^+(i)$ also in the sense of mathematical topology [30], [31].

In the next step we introduce a topology on the set of the synaptic weight vectors on the basis of their receptive fields, which again allows us to describe the neighborhood relationships between two vectors.

Definition 4: Let $\Psi_{A \to M} : A \to M^A \subset M \subseteq \Re^d$ be a map attributing to each neuron i a specific vector $w_i \in M^A$

with $M^A = \{w_i \in \Re^d \mid i \in A\}$. Furthermore, let $\mathcal{V}_M$ be the induced Voronoi diagram of M with respect to M^A. Let $\mathcal{G}_M$ be the dual Delaunay graph of $\mathcal{V}_M$. Let furthermore $\mathcal{G}_M(i)$ denote $\mathcal{G}_M$ where the neural unit i was taken as root. A (discrete) topology $\mathcal{T}_A^+(i)$ with respect to $i \in A$ is induced by the graph metric in $\mathcal{G}_M(i)$. $\mathcal{G}_M(i)$ is equipped with the graph metric that in turn induces the local (discrete) topology $\mathcal{T}_{M^A}(i)$ in $\mathcal{G}_M(i)$ and, hence, also in M^A. $\mathcal{T}_{M^A}(i)$ is said to be the $\Psi_{A\to M}$-**induced neighborhood topology with respect to** i in M^A and $(M^A, \mathcal{T}_{M^A}(i))$ is a topological space.

Now topology preservation of a map can be expressed by the following definition.

Definition 5: The map $\mathcal{M}_A = (\Psi_{A\to M}, \Psi_{M\to A})$ is said to be **topology preserving** if both $\Psi_{M\to A} : (M^A, \mathcal{T}_{M^A}(i)) \to (A, \mathcal{T}_A^-(i))$ and $\Psi_{A\to M} : (A, \mathcal{T}_A^+(i)) \to (M^A, \mathcal{T}_{M^A}(i))$ are continuous maps of the respective topological spaces for all neural units $i \in A$ where $\Psi_{M\to A} : \Re^d \supseteq M \to A$ is defined by $i(v) = \arg(\min_{i\in A} \|v - w_i\|)$. Irrespective of the different topologies $\Psi_{M\to A}$ is the inverse mapping of $\Psi_{A\to M}$.

We have immediately the following two corollaries for the most important cases of rectangular and hexagonal (triangular) lattices.

Corollary 1: In the case of a rectangular d_A-dimensional lattice A of neurons the strong topology is induced by the Euclidean norm $\|\cdot\|_E$ in A or the summation-norm $\|\cdot\|_\Sigma = \sum_{j=1}^{d_A} |(\cdot)_j|$, and the weak topology is induced by the maximum-norm $\|\cdot\|_{\max} = \max_{j=1}^{d_A} |(\cdot)_j|$. The systems of open sets $\mathcal{S}_\Sigma(i)$ and $\mathcal{S}_{\max}(i)$ defining the topologies $\mathcal{T}_A^+(i) = \mathcal{T}_A^\Sigma(i)$ and $\mathcal{T}_A^-(i) = \mathcal{T}_A^{\max}(i)$ are determined by

$$\mathcal{S}_\Sigma(i) = \{s_k \mid s_k = \{l \in A \mid \|i - l\|_\Sigma = k \geq 1\}\} \quad (22)$$

and

$$\mathcal{S}_{\max}(i) = \{s_k \mid s_k = \{l \in A \mid \|i - l\|_{\max} = k \geq 1\}\} \quad (23)$$

respectively.

Corollary 2: In the special case of A being a hexagonal (triangular) lattice the weak and strong topology coincide. Hence, the definition of topology preservation relies on a single topology in the net which corresponds to the strong neighborhood topology.

The conclusion in Corollary 2 is in agreement with the definition of neighborhood given in [19]. By means of the Definitions 2, 3, 4, and 5 we can now generalize the definition of the topographic function given in the (9) and (10). For each unit i we define

$$\begin{aligned} f_i(k) &\stackrel{\text{def}}{=} \#\left\{j \mid d_{\mathcal{T}_A^-(i)}(i,j) > k;\ d_{\mathcal{T}_{M^A}(i)}(i,j) = 1\right\} \\ f_i(-k) &\stackrel{\text{def}}{=} \#\left\{j \mid d_{\mathcal{T}_A^+(i)}(i,j) = 1;\ d_{\mathcal{T}_{M^A}(i)}(i,j) > k\right\} \end{aligned} \quad (24)$$

with $k = 1, \cdots, N-1$. $\#\{\cdot\}$ denotes the cardinality of the set and

$$d_{\mathcal{T}(i)}(i,j) \stackrel{\text{def}}{=} \|w_i - w_j\|_{\mathcal{T}^\circ(i)}. \quad (25)$$

is a distance measure based on the topology $\mathcal{T}^\circ(i)$. Looking at a neural unit i, $f_i(k)$ with $k > 0$ determines the continuity of $\Psi_{M\to A}$ and $f_i(k)$ with $k < 0$ determines the continuity of $\Psi_{A\to M}$ as defined above. The topographic function of the neural lattice A with respect to the input manifold M is then, in analogy to Definition (10), again defined as

$$\Phi_A^M(k) \stackrel{\text{def}}{=} \begin{cases} \frac{1}{N}\sum_{j\in A} f_j(k) & k > 0 \\ \Phi_A^M(1) + \Phi_A^M(-1) & k = 0 \\ \frac{1}{N}\sum_{j\in A} f_j(k) & k < 0 \end{cases} \quad (26)$$

in analogy to Section III-A we can make the following remark.

Remark 2: We obtain $\Phi_A^M \equiv 0$ and, particularly, $\Phi_A^M(0) = 0$ if and only if the SOFM is perfectly topology preserving. The largest $k^+ > 0$ for which $\Phi_A^M(k^+) \neq 0$ holds yields the range of the largest fold if the effective dimension of the data manifold M is larger than the dimension d_A of the lattice A. The smallest $k^- < 0$ for which $\Phi_A^M(k) \neq 0$ holds yields the range of the largest fold if the effective dimension of the data manifold M is smaller then the dimension d_A of the lattice A. Small values of k^+ and k^- indicate that there are only local conflicts, whereas large values indicate a global dimensional conflict.

Kohonen gave a definition for what it means for a *one-dimensional* topographic map to be *ordered*, which finally should be discussed in the light of the definition studied in this paper. The following definition, only valid for the one-dimensional case, was introduced in [32].

Definition 6: We consider a chain of N neural units i with weight vectors w_i and receptive fields $R_i = \tilde{V}_i$ as defined in 5. Let $\eta_i(v)$ be an activity function of the ith neuron with respect to a stimuli vector $v \in M \subseteq \Re^d$, for instance the negative distance $-\|w_i - v\|$ or the inverse distance $(\frac{1}{\|w_i - v\|})$. Let $X = \{x_j \in M \subseteq \Re^d \mid j = 1, \cdots, N\}$ be a set of points such, that $x_1 \stackrel{\text{Rel}}{\circ} x_2 \stackrel{\text{Rel}}{\circ} x_3 \stackrel{\text{Rel}}{\circ} \cdots \stackrel{\text{Rel}}{\circ} x_N$ holds and for each neural unit i_j exists a $x_j \in X$ for which $x_j \in R_{i_j}$ holds. The $\stackrel{\text{Rel}}{\circ}$ is an arbitrary suitably chosen relation, not necessarily transitive. The system is said to implement a *one-dimensional ordered mapping* if for $i_1 > i_2 > i_3 > \cdots > i_N$

$$\begin{aligned} \eta_{i_1}(x_1) &= \max_{i=1,\cdots,N} \eta_i(x_1) \\ \eta_{i_2}(x_2) &= \max_{i=1,\cdots,N} \eta_i(x_2) \\ \eta_{i_3}(x_3) &= \max_{i=1,\cdots,N} \eta_i(x_3) \\ &\vdots \\ \eta_{i_N}(x_N) &= \max_{i=1,\cdots,N} \eta_i(x_N) \end{aligned} \quad (27)$$

holds.

In this paper we have given an explicit order relation $\stackrel{\text{Rel}}{\circ}$ based on an underlying topology, in contrast to the requirement of the existence of such a relation. The proposed straightforward generalization of definition (27) to higher dimensions is (as pointed out in [33]) by no means trivial. In fact, one needs to make the relation $\stackrel{\text{Rel}}{\circ}$ more explicit in order to find whether definition (27) is applicable to higher-dimension, too. In this sense we gave the justification for Kohonen's early framework and solved the problem of general definition of topology preservation what he carried out in [12], although the starting point of our investigation was the desire to improve the methods proposed in [11].

A further issue is the topology preservation of the map $\Psi_{M \rightarrow A}$ which has been pointed out to be crucial as well for a strict definition of the intuitive notion of topology preservation.

VI. Conclusion

We presented a novel approach to the problem of measuring the topology preservation of a SOFM. The approach is based on the neighborhood relations between receptive fields. The introduced topographic function is an improvement over the topographic product suggested in [11] since it determines the degree of topology preservation by considering the shape of the given input manifold M. This was demonstrated for various examples of nonlinear input manifolds. Furthermore, we developed a general definition of topology preservation for SOFM, based on topological spaces. It turns out that this definition is a generalization of a definition which has been proposed by Kohonen for the one-dimensional case.

Acknowledgment

The authors thank H.-U. Bauer and K. Pawelzik of the University of Frankfurt/Main, Germany, for helpful discussions and critical remarks.

References

[1] T. Kohonen, *Self-Organization and Associative Memory*, Springer Series in Information Science 8. Berlin: Springer, 1984.
[2] W. D. Brandt, H. Behme, and H. W. Strube, "Bildung von Merkmalen zur Spracherkennung mittels phonotopischer Karten," in *Fortschritte zur Akustik—DAGA'91*, DPG-GmbH, Bad Honnef, Germany, pp. 1057–1060, 1991.
[3] T. Kohonen, K. Mäkisara, and T. Saramäki, "Phonotopic maps—Insightful representation of phonological features for speech recognition," in *Proc. 7th Int. Conf. Pattern Recognition*, Montreal, Canada, 1984, pp. 182–185.
[4] T. Kohonen, "The self-organizing map," in *Proc. IEEE*, vol. 78, 1464–1480, 1990.
[5] T. M. Martinetz, H. Ritter, and K. Schulten, "Three-dimensional neural net for learning visuomotor-coordination of a robot arm," *IEEE Trans. Neural Networks*, vol. 1, pp. 131–136, 1990.
[6] N. M. Nasrabadi and Y. Feng, "Vector quantization of images based upon Kohonen self-organizing feature maps," in *Proc. IEEE Int. Conf. Neural Networks*, San Diego, CA, 1988, pp. 1101–1108.
[7] J. Naylor and K. P. Li, "Analysis of a neural network algorithm for vector quantization of speech parameters," in *Proc. 1st Annu. INNS Mtg.*. New York: Pergamon, 1988, p. 310.
[8] H. Ritter and K. Schulten, "Topology conserving mappings for learning motor tasks," in *Neural Networks for Computing, AIP Conf. Proc.*, J. S. Denker, Ed, Snowbird, UT, vol. 151, 1986, pp. 376–380.
[9] A. Ultsch, "Knowledge acquisition with self-organizing neural networks," in *Proc. Int. Conf. Artificial Neural Networks 1993*, I. Aleksander and J. Taylor, Eds. Amsterdam, The Netherlands: North-Holland, 1992, pp. 735–739.
[10] A. Ultsch, "Self organized feature maps for monitoring and knowledge acquisition of a chemical process," in *Proc. Int. Conf. Artificial Neural Networks 1993*, S. Gielen and B. Kappen, Eds. Berlin: Springer-Verlag, pp. 864–867.
[11] H.-U. Bauer and K. Pawelzik, "Quantifying the neighborhood preservation of self-organizing feature maps," *IEEE Trans. Neural Networks*, vol. 3, pp. 570–579, 1992.
[12] T. Kohonen, *Self-Organizing Maps*, Springer Series in Information Science 30. Berlin: Springer-Verlag, 1994.
[13] H. Ritter, T. M. Martinetz, and K. Schulten, *Neural Computation and Self-Organizing Maps*. Reading, MA: Addison Wesley, 1992.
[14] R. Der and T. Villmann, "Dynamics of self-organized feature mapping," in *New Trends in Neural Computation*, Lecture Notes in Computer Science 686, J. Mira, J. Cabestany, and A. Prieto, Eds. Berlin: Springer-Verlag, 1993, pp. 312–315.
[15] R. Der, M. Herrmann, and T. Villmann, "Time behavior of topological ordering in self-organized feature mapping," to appear in *Biol. Cybern.*, 1994.
[16] P. Demartines and F. Blayo, "Kohonen self-organizing maps: Is the normalization necessary?" *Complex Syst.*, vol. 6, pp. 105–123, 1992.
[17] S. Zrehen, "Analyzing Kohonen maps with geometry," in *Proc. Int. Conf. Artificial Neural Networks 1993*, S. Gielen and B. Kappen, Eds. Berlin: Springer-Verlag, 1993, pp. 609–612.
[18] T. Martinetz, "Competitive Hebbian learning rule forms perfectly topology preserving maps," in *Proc. Int. Conf. Artificial Neural Networks 1993*, S. Gielen and B. Kappen, Eds. Berlin: Springer-Verlag, 1993, pp. 427–434.
[19] T. Martinetz and K. Schulten, "Topology representing networks," *Neural Networks*, vol. 7, no. 3, pp. 507–522, 1994.
[20] G. Biess, *Graphentheorie*. Leipzig, Germany: BSB B. G. Teubner Verlagsgesellschaft, 1988.
[21] H.-P. Bähr and Th. Vögtle, *Digitale Bildverarbeitung: Anwendung in Photogrammetrie, Kartographie und Fernerkundung*. Karlsruhe, Germany: Wichmann-Verlag, 1991.
[22] *ERDAS: Geophysical Software System for Satellite Image Processing, Manual*, Geosystems GmbH, Germany.
[23] A. Visa, K. Valkealahti, and O. Simula, "Cloud detection based on texture segmentation by neural network methods," in *Proc. IEEE Int. Joint Conf. Neural Networks*, vol. 2, Singapore, 1991, pp. 1001–1006.
[24] M. Groß and F. Seibert, "Neural network image analysis for environmental protection," *Zentrum für Graphische Datenverarbeitung*, Darmstadt, Germany, 1993.
[25] UMWELT-FORSCHUNGSZENTRUM Halle-Leipzig, Arbeitsgruppe "Angewandte Landschaftsökologie."
[26] T. Gramss and H. W. Strube, "Recognition of isolated words based on psychoacoustics and neurobiology," *Speech Commun.*, vol. 9, pp. 35–40, 1990.
[27] H.-U. Bauer and K. Pawelzik, "Topologieerhaltung und dimensionsbestimmung: Methoden der nichtlinearen dynamik in neuroinformatik und spracherkennung," vol. 306, *Informatik Fachberichte*, D. König and M. Lang, Eds. Berlin: Springer-Verlag, pp. 146–152, 1992.
[28] B. Delaunay, "Sur la Spère Vide," *Bull. Acad. Sci. USSR (VII)*, Classe Sci. Mat. Nat., pp. 793–800, 1934.
[29] G. Voronoi, "Nouvelles Aoolications des Parameters à la Theorie des Formes Quadratiques," Deuxième Mémorie: Recherches sur les Paralléloèdres Primitifs, J. reine angew. Math., vol. 134, pp. 198–287, 1908.
[30] L. W. Kantorowich and G. P. Akilov, *Funktionalanalysis in normierten Räumen*. Berlin: Akademie-Verlag, 1978.
[31] A. N. Kolmogorov and S. V. Fomin, *Reelle Funktionen und Funktionalanalysis*. Berlin: Deutscher Verlag der Wissenschaften, 1975.
[32] T. Kohonen, "A simple paradigm for the self-organized formation of structured feature maps," in Competition and Cooperation in Neural Nets, *Lecture Notes in Biomathematics*, S. Amari and M. A. Arbib, Eds. Berlin: Springer-Verlag, 1982, pp. 248–266.
[33] M. Cottrell, J. C. Fort, and G. Pages, "Two or three things that we know about the Kohonen algorithm," in *Proc. European Symp. Artificial Neural Networks (ESANN)*, Brussels, Belgium, 1944, pp. 235–244.

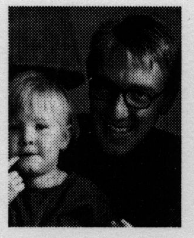
Thomas Villmann, for a photography and biography, please see this issue, pp. 226.

Ralf Der was born Breitenbrunn, Germany. He received the Ph.D. degree in nuclear physics at Leipzig University, Germany, in 1971 and the Dr.sc.nat. (Habilitation) in 1984 from the Academy of Sciences of the former GDR.

He worked as Researcher, later on as the Scientific Leader, of a small research group at the Academy of Sciences of the GDR from 1971 until 1989. During that time his research concentrated on nonequilibrium statistical mechanics, in particular kinetic theory, the quantum statistical theory of chemical reactions, irreversibility, and the self-organization of matter. As a member of the 24th Soviet Antarctic Expedition he carried out glaciological research in 1979 to 1980. Since 1989 he has been working as a Senior Researcher (Privatdozent) at the University of Leipzig with research interests concentrating on the statistical mechanics of learning and self-organization in neural networks and the theory and application of reinforcement learning in control and optimization problems. He is the head of several research projects ranging from both the applications of neural networks to the principles of cortical computation.

Michael Herrmann received the diploma degree in mathematics in 1988 and the Ph.D. degree in computer science in 1993, both from Leipzig University, Germany.

He has been a Postdoctoral Researcher in the Institute of Informatics at Leipzig University and subsequently at NORDITA, Copenhagen. Presently, he is with RIKEN's Laboratory for Information Representation in Japan. His research interests include the theory of artifical neural networks, in particular self-organizing maps, neural modeling, and dynamic programming.



Thomas M. Martinetz was born in 1962. He studied physics and mathematics in München and Cologne, Germany, and received the diploma degree in physics and the Ph.D. degree in theoretical physics from the Technical University of München in 1988 and 1991, respectively.

From 1989 to 1991 he was with the Beckman Institute and the Department of Physics of the University of Illinois at Urbana-Champaign. In 1991 he joined the Neural Networks Group of the Corporate Research and Development Department of Siemens in München, and from 1994 to 1996 he was Assistant to the Senior Vice President there. In 1996 he became Professor at the "Institut für Neuroinformatik" of the University of Bochum. Since 1987 his research has been in the area of neural computation. His main interests include the design, application, and theoretical analysis of neural-network architectures for adaptive signal processing, motor coordination, interacting agents, and process control. He is author of several papers and coauthor of a textbook on neural computation.