

Analyzing Phase Transitions in High-Dimensional Self-Organizing Maps

M. Riesenhuber*, H.-U. Bauer^{†‡}, T. Geisel[†]

Institut für Theoretische Physik and SFB Nichtlineare Dynamik,

Universität Frankfurt, Robert-Mayer-Str. 8-10,

60054 Frankfurt/Main, Fed. Rep. of Germany

email: bauer@chaos.uni-frankfurt.de, Fax: 069 / 798 28354

Abstract

The Self-Organizing Map (SOM), a widely used algorithm for the unsupervised learning of neural maps, can be formulated in a low-dimensional “feature map” variant which requires prespecified parameters (“features”) for the description of receptive fields, or in a more general high-dimensional variant which allows to self-organize the structure of individual receptive fields as well as their arrangement in a map. We present here a new analytical method to derive conditions for the emergence of structure in SOMs which is particularly suited for the as yet inaccessible high-dimensional SOM variant. Our approach is based on an evaluation of a map distortion function. It involves only an ansatz for the way stimuli are distributed among map neurons; the receptive fields of the map need not be known explicitly. Using this method we first calculate regions of stability for four possible states of SOMs projecting from a rectangular input space to a ring of neurons. We then analyze the transition from non-oriented to oriented receptive fields in a SOM-based model for the development of orientation maps. In both cases, the analytical results are well corroborated by the results of computer simulations.

submitted to Biological Cybernetics, December 14, 1995 revised version, July 14, 1996

*Present address: Department of Brain and Cognitive Sciences and Center for Biological and Computational Learning, Massachusetts Institute of Technology E25-221, Cambridge, MA 02139, USA

†Present address: Max-Planck-Institut für Strömungsforschung, Postfach 28 53, 37018 Göttingen, FRG

‡Corresponding author

1. Introduction

The activity-driven self-organization of neural maps is a central paradigm in the ontogenetic development of the brain. Many models for such processes were introduced and analyzed, often in the context of specific developmental phenomena like the formation of ocular dominance and orientation columns in the visual cortex (examples include von der Malsburg's (1973) early work on the formation of orientation columns, or Miller et al.'s model (1989) on the formation of ocular dominance bands; for a recent review see Erwin et al., 1995). Other map formation algorithms were formulated in rather general terms and found applications not only in the biological realm, but also as part of technical systems. This latter class of more abstract map algorithms includes Kohonen's self-organizing map algorithm (SOM, for a general discussion see Kohonen, 1995, Ritter et al., 1992). The SOM not only formed the basis of several specific models for the development of biological maps (Ritter, 1986, Obermayer et al., 1990a, Obermayer et al., 1990b, Obermayer et al., 1992, Goodhill, 1993, Wolf et al., 1994, Bauer, 1995), but was also used as a neighborhood preserving vector quantizer in signal processing tasks.

The SOM-algorithm can be formulated in two different ways, a high-dimensional and a low-dimensional variant. The difference between the two variants is best illustrated in the context of a sensory map. Here, stimuli in a sensory layer are projected in a topographic fashion onto the neurons of the map. In the high-dimensional SOM, the stimuli are formally described by an activity distribution over many sensory channels which make up the sensory layer (Obermayer et al., 1990a, Obermayer et al., 1990b, Goodhill, 1993). Correspondingly, the receptive fields of the neurons are described as synaptic weight distributions in the high-dimensional space of sensory channels. An alternative, low-dimensional description can be used, if particular "features" of the stimulus and receptive fields can be expected to represent the nature of the map sufficiently well. Examples where this reduction to a so-called self-organizing feature map (SOFM) has been used include models for the formation of orientation and ocular dominance maps (Obermayer et al., 1992, Bauer, 1995) as well as the modeling of retinotopic maps (see, e.g., Wolf et al., 1994). In the latter case, the relevant features are the retinal positions of stimuli and receptive fields, given by their resp. centers of gravity only. Simulations of SOFMs require

a substantially smaller numerical effort; in addition more theoretical results are available for this kind of map. However, the feature space ansatz requires a priori assumptions about the structure of receptive fields, whereas the high-dimensional SOMs can develop both the arrangement of receptive fields in the map and the receptive field structure itself. Furthermore, the necessity to extract a small number of parameters from a stimulus in the SOFM case precludes the investigation of the effects of more complicated stimulus shapes on the self-organization process.

Theoretical investigations of SOMs or SOFMs often aim at relating the emergence of a particular structure in the map to parameters of the adaptation process or the stimulus ensemble. An archetypical problem in this context is the transition from a flat map to a distorted map as a consequence of a dimension mismatch between input space and output space, if the stimulus amplitude along the additional dimensions exceeds a critical value. For the low-dimensional SOFMs Ritter and Schulten (1988) analytically derived critical values for the parameters such that the dimension mismatch would induce the distorted structure in the map. Their method was subsequently adapted by Obermayer et al. (1992) to analyze the emergence of orientation and ocular dominance structure and by Bauer (1995) to investigate the impact of areal geometry on the preferred orientation of ocular dominance bands in SOFM-based models for the development of visual maps.

Considering the potentially richer repertoire of self-organization phenomena of the less restricted high-dimensional SOMs, a comparable method for their analysis would be desirable. Here we present a new and simple approach to compare different states of SOMs, and in this way to calculate critical parameter values for phase transitions of these maps. After a short introduction to SOMs and SOFMs in the next section, our method is described in detail in the third section. In the fourth and fifth section we then apply the method first to a test problem of abstract nature, then to an analysis of the formation of oriented receptive fields in a SOM-model for orientation maps in the visual cortex. As we will see, in both cases the high-dimensional SOM behaves in a qualitatively different way from the low-dimensional SOFM. A discussion of our approach and its results concludes the paper.

2. High-dimensional SOMs and low-dimensional SOFMs

In a Self-Organizing Map (SOM) stimuli \mathbf{v} are projected onto neurons \mathbf{r} , located at the vertices of a (typically two-dimensional) output space lattice. The stimuli \mathbf{v} are given as an M -dimensional activity distribution over input channels i , normalized to a constant sum S ,

$$\sum_i^M v_i = S . \quad (2.1)$$

Each neuron \mathbf{r} has associated with it a receptive field which is also formalized as an M -dimensional vector $\mathbf{w}_{\mathbf{r}}$. A stimulus \mathbf{v} is mapped onto that neuron \mathbf{s} whose receptive field $\mathbf{w}_{\mathbf{s}}$ has the largest overlap with the stimulus,

$$\mathbf{s} = \arg \max_{\mathbf{r}} (\mathbf{w}_{\mathbf{r}} \cdot \mathbf{v}) . \quad (2.2)$$

This amounts to a winner-take-all rule, a strong nonlinearity which in a biological context is explained as a consequence of a strong nonlinear lateral cortical interaction (Kohonen, 1993, 1995). The SOM is adapted to a specific projection task by appropriate changes of the receptive field vectors $\mathbf{w}_{\mathbf{r}}$. To this purpose, a sequence of stimuli is presented to the map, for each \mathbf{v} the current best-matching neuron \mathbf{s} is determined and a learning step

$$\Delta \mathbf{w}_{\mathbf{r}} = \epsilon h_{\mathbf{r}\mathbf{s}} (\mathbf{v} - \mathbf{w}_{\mathbf{r}}) \quad (2.3)$$

is carried out. Here, ϵ is the size of learning steps, $h_{\mathbf{r}\mathbf{s}}$ denotes a neighborhood function, usually chosen to be of Gaussian shape,

$$h_{\mathbf{r}\mathbf{s}} = e^{-\frac{\|\mathbf{r}-\mathbf{s}\|^2}{2\sigma^2}} , \quad (2.4)$$

which enforces neighboring neurons to align their receptive fields. Via the neighborhood function h the property of topography is imposed on the SOM. Summing Eq. (2.3) over the input channels i shows that in the course of learning, the normalization (2.1) also induces a (linear) normalization of the weight vectors $\mathbf{w}_{\mathbf{r}}$. The region in stimulus space which will be mapped onto a particular neuron \mathbf{r} , is called the Voronoi cell $\Omega_{\mathbf{r}}$ of this neuron.

Instead of using the full stimulus distribution \mathbf{v} and the receptive field distributions $\mathbf{w}_{\mathbf{r}}$ a Self-Organizing Map can also be formulated in terms of features (eg., the centers of

gravity) $\tilde{\mathbf{v}}$ and $\tilde{\mathbf{w}}_{\mathbf{r}}$ which are extracted from \mathbf{v} and $\mathbf{w}_{\mathbf{r}}$ by application of a linear operator R ,

$$\tilde{\mathbf{v}} = R(\mathbf{v}), \quad \tilde{\mathbf{w}}_{\mathbf{r}} = R(\mathbf{w}_{\mathbf{r}}) . \quad (2.5)$$

Application of R to (2.3) reveals (cf. Ritter et al., 1992) that the $\tilde{\mathbf{w}}_{\mathbf{r}}$ obey dynamics of the same form as that of the $\mathbf{w}_{\mathbf{r}}$,

$$\Delta \tilde{\mathbf{w}}_{\mathbf{r}} = \epsilon h_{rs}(\tilde{\mathbf{v}} - \tilde{\mathbf{w}}_{\mathbf{r}}) . \quad (2.6)$$

This identical form suggests that the low-dimensional feature map, now called a SOFM, should yield very similar, if not identical, results as the full SOM as far as the representation of the features is concerned. However, the mapping rule (2.2) cannot be transformed in a similarly stringent fashion, as the dot product is not a useful measure in the feature space (Sutton and Reggia, 1994). Instead, the SOFM-algorithm employs the minimization of the Euclidean distance between $\tilde{\mathbf{w}}_{\mathbf{r}}$ and $\tilde{\mathbf{v}}$,

$$\mathbf{s} = \arg \min_{\mathbf{r}} \|\tilde{\mathbf{w}}_{\mathbf{r}} - \tilde{\mathbf{v}}\| . \quad (2.7)$$

Equation (2.7) deviates from (2.2), because only for square-normalized vectors the minimum Euclidean distance corresponds to the maximum dot product. Even more important, the distance measures in (2.2) and (2.7) operate on vectors in different spaces which can yield different best-matching neurons even if the normalization issue is ignored (see Fig. 1 for an illustration; for a more detailed discussion of the relation between SOMs and SOFMs for the special case of ocular dominance maps, see Miller, 1995). Hence, despite the similarity of the learning eqs. (2.3) and (2.6), the different mapping rules (2.2) and (2.7) prohibit a transfer of results from one variant of the map algorithm to the other.

3. Evaluating Tesselations of the Stimulus Space

One goal of theoretical investigations of map formation algorithms is to relate the classes of states emerging in a map to parameters of the stimulus ensemble or the map formation algorithm (like, e.g., the width σ of the neighborhood function). Of particular interest are those values of the parameters where the resulting structure of the map undergoes a qualitative change, a phase transition. For SOFMs with mismatching input and output space dimensions, consideration of the fluctuations of the map about a state with

trivial structure led to conditions on the parameters when the trivial state ceases to be stable (Ritter and Schulten, 1988). For parameter values beyond the critical ones, the map evolves to a different state (the structure of which, however, is not revealed by this analysis).

A similar method to determine instabilities in high-dimensional SOMs has as yet not been available. Part of the reason may be that for high-dimensional maps it is difficult (if not impossible) to directly make an ansatz for the state of the map, i.e. for the receptive field vectors $\mathbf{w}_{\mathbf{r}}$. If the $\mathbf{w}_{\mathbf{r}}$ were available, different states of the map could be compared using the “naive energy function”

$$E_{\mathbf{w}} = \sum_{\mathbf{r}} \sum_{\mathbf{r}'} \sum_{\mathbf{v}' \in \Omega_{\mathbf{r}'}} (\mathbf{v}' - \mathbf{w}_{\mathbf{r}})^2 e^{\left(-\frac{\|\mathbf{r}-\mathbf{r}'\|^2}{2\sigma^2}\right)}, \quad (3.1)$$

which can be used to analyze the map formation in the case of a discrete set of stimuli (Ritter et al., 1992); $\Omega_{\mathbf{r}}$ denotes the subset of stimuli which are mapped onto \mathbf{r} . Even though the SOM learning dynamics does not proceed along the gradient of this function (or any other energy function) in the case of a continuous stimulus ensemble (Tolat, 1990) the deviations become small in the limit of an ordered map with large values for σ (Erwin et al., 1992). It is also known that modification of the SOM winner rule lead to a map formation algorithm following exactly the gradient of an energy function (Heskes and Kappen, 1993, Luttrell, 1994). Therefore, a sensible strategy to determine the final state of a SOM is to compare distortion functions for different map states.

However, evaluation of $E_{\mathbf{w}}$ requires explicit knowledge of the receptive field vectors $\mathbf{w}_{\mathbf{r}}$ for which a direct ansatz is in general difficult to make. Therefore, we here propose to focus not on the $\mathbf{w}_{\mathbf{r}}$ themselves, but rather on the tessellation they induce in the stimulus space, i.e. on the sets $\Omega_{\mathbf{r}}$. This tessellation can be of quite simple form, even if the actual shape of the receptive fields $\mathbf{w}_{\mathbf{r}}$ is not (see examples in sections 4 and 5). Furthermore, in SOMs, most continuous changes of map or stimulus parameters leave the tessellation unaltered, even though the $\mathbf{w}_{\mathbf{r}}$ are also changed continuously, whereas qualitative changes of map states usually correspond to changes of the tessellation.

Once an ansatz for the $\Omega_{\mathbf{r}}$ has been made, the stability of the resp. map states can be compared using either of two, rather similar, distortion functions. First, one might think of computing the $\mathbf{w}_{\mathbf{r}}$ as superpositions of stimuli,

$$\mathbf{w}_{\mathbf{r}} = c^{-1} \sum_{\mathbf{r}'} \sum_{\mathbf{v}' \in \Omega_{\mathbf{r}'}} \mathbf{v}' e\left(-\frac{\|\mathbf{r}-\mathbf{r}'\|^2}{2\sigma^2}\right), \quad (3.2)$$

$$c = \sum_{\mathbf{r}'} \sum_{\mathbf{v}' \in \Omega_{\mathbf{r}'}} e\left(-\frac{\|\mathbf{r}-\mathbf{r}'\|^2}{2\sigma^2}\right), \quad (3.3)$$

and put this result into the energy function (3.1),

$$E_{\mathbf{w}} = \sum_{\mathbf{r}} \sum_{\mathbf{r}'} \sum_{\mathbf{v}' \in \Omega_{\mathbf{r}'}} \left(\mathbf{v}' - c^{-1} \sum_{\mathbf{r}''} \sum_{\mathbf{v}'' \in \Omega_{\mathbf{r}''}} \mathbf{v}'' e\left(-\frac{\|\mathbf{r}-\mathbf{r}''\|^2}{2\sigma^2}\right) \right)^2 e\left(-\frac{\|\mathbf{r}-\mathbf{r}'\|^2}{2\sigma^2}\right). \quad (3.4)$$

However, Eq. (3.4) turns out to be prohibitively complicated for analytical evaluation and, therefore, is not considered in the remainder of this article.

Instead, we make a further simplifying step, disregard the precise shape of receptive fields altogether, and consider the distortion function

$$E_{\mathbf{v}} = \sum_{\mathbf{r}} \sum_{\mathbf{r}'} \sum_{\mathbf{v}' \in \Omega_{\mathbf{r}'}} \sum_{\mathbf{v} \in \Omega_{\mathbf{r}}} (\mathbf{v}' - \mathbf{v})^2 e\left(-\frac{\|\mathbf{r}-\mathbf{r}'\|^2}{2\sigma^2}\right), \quad (3.5)$$

which evaluates the variance of the stimuli not with regard to the $\mathbf{w}_{\mathbf{r}}$, but with regard to the set of stimuli within the Voronoi cell $\Omega_{\mathbf{r}}$. Assuming an equal number of stimuli per Voronoi cell, the distortion measures (3.4) and (3.5) are minimized for the same tessellations in the limit of $\sigma \rightarrow 0$. Assuming further a mirror-symmetry of the Voronoi cells with respect to the neuron output space positions, the deviations are small for non-vanishing σ . Both assumptions are not critical and allow us to use an evaluation of $E_{\mathbf{v}}$ with regard to various sensible tessellations of the stimulus space to compare the relative stability of different map states, and to compute state diagrams of SOMs. As we will see in the next two chapters, $E_{\mathbf{v}}$ can be calculated for high-dimensional SOMs in a sometimes surprisingly simple fashion.

4. Analyzing a phase transition in a 4-neuron map

As a first example we consider SOMs which map a rectangular input space onto a chain of 4 neurons. The input space is discretized as a $4M \times 2sM$ lattice, where s is a parameter.

Periodic boundary conditions are applied in the x -direction of the input rectangle and the output chain of neurons. The stimuli are assumed to be Gaussians of width $\sigma_{stim} \ll M$, centered at randomly varying positions (x_0, y_0) in the input rectangle.

This example constitutes a $4M \times 2sM$ -dimensional problem and is treated as such in the following. However, since the input channels are coupled due to the finite extension of the stimuli, and since the stimuli are parametrized in a two-dimensional space, we can identify the Voronoi cells of the neurons with regions in this two-dimensional space. In particular, we can discuss the tessellations of the high-dimensional stimulus space in terms of the shape of regions (called “tessellation regions”) in the parameter space. Several tessellations are imaginable:

- All stimuli map onto just one neuron. This constellation is unstable and is not further discussed.
- All stimuli are mapped onto either of two neurons (not neighboring in the output space), with the two intermediate neurons never winning. For symmetry each of the possible winner-neurons is best-matching for half the stimuli. For distortion minimization this group of stimuli forms a connected region in the two-dimensional parameter space (see Fig. 2a). The never-winning intermediate neurons adjust their receptive fields as the average of their neighbors’ receptive fields. Such a constellation is numerically observed (Fig. 2f) and stable; it is analogous to a low-dimensional SOFM which tries to cover a strongly bent input space and has to spend a few neurons in “empty space” in order to do so (see Fig. 3.7 in Kohonen, 1995, as an example).
- The stimuli can also be split among all four neurons, with each neuron being best-matching for a quarter of the stimuli. Several shapes of the corresponding tessellation regions are imaginable, some of which are observed (Figs. 2b-e,g-j).

The question now arises, which of the above states is attained for each combination of parameters s and σ . Let us consider first the two-neuron type solution (Fig. 2a) vs. the group of four-neuron type solutions (Figs. 2b-e). In the limit of a small spatial stimulus

extension, most pairs \mathbf{v}, \mathbf{v}' have negligible overlap, and their squared difference takes on a value $(\mathbf{v} - \mathbf{v}')^2 \approx 2\mathbf{v}^2 = c_0$. Neglecting the impact of stimulus pairs with appreciable overlap for the moment, the $E_{\mathbf{v}}$ can be evaluated approximately, leading to

$$\begin{aligned} E_{\mathbf{v},2} &= \sum_{\mathbf{r}=1,3} \sum_{\mathbf{v} \in \Omega_{\mathbf{r}}} \sum_{\mathbf{r}'=1,3} \sum_{\mathbf{v}' \in \Omega_{\mathbf{r}'}} (\mathbf{v} - \mathbf{v}')^2 e^{-\frac{\|\mathbf{r}-\mathbf{r}'\|^2}{2\sigma^2}} \\ &= 2(2M \cdot 2sM)c_0 \left((2M \cdot 2sM)e^0 + (2M \cdot 2sM)e^{\frac{-4}{2\sigma^2}} \right), \end{aligned} \quad (4.1)$$

$$\begin{aligned} E_{\mathbf{v},4} &= \sum_{\mathbf{r}=1}^4 \sum_{\mathbf{v} \in \Omega_{\mathbf{r}}} \sum_{\mathbf{r}'=1}^4 \sum_{\mathbf{v}' \in \Omega_{\mathbf{r}'}} (\mathbf{v} - \mathbf{v}')^2 e^{-\frac{\|\mathbf{r}-\mathbf{r}'\|^2}{2\sigma^2}} \\ &= 4(M \cdot 2sM)c_0 \left((M \cdot 2sM)e^0 + 2(M \cdot 2sM)e^{\frac{-1}{2\sigma^2}} + (M \cdot 2sM)e^{\frac{-4}{2\sigma^2}} \right). \end{aligned} \quad (4.2)$$

Here, the leading factor of 2 (resp. 4) results from the sum over \mathbf{r} , the following factor of $2M \cdot 2sM$ (resp. $(M \cdot 2sM)$) from the sum over $\mathbf{v} \in \Omega_{\mathbf{r}}$, and the further factors of $2M \cdot 2sM$ (resp. $(M \cdot 2sM)$) result from the sums over $\mathbf{v}' \in \Omega_{\mathbf{r}'}$. Equating $E_{\mathbf{v},2}$ and $E_{\mathbf{v},4}$ yields the transition between the corresponding two map solutions, at

$$e^0 + e^{\frac{-4}{2\sigma^2}} = 2e^{\frac{-1}{2\sigma^2}} \quad \Leftrightarrow \quad \sigma = 0.91. \quad (4.3)$$

So the map will attain a two-neuron type solution at $\sigma > 0.91$, and a four-neuron type solution at $\sigma < 0.91$, independent of the elongation parameter s .

In a second step we now consider the different possible geometries for the four-neuron type solutions. Since the respective tessellation regions are of the same size, the different four-neuron solutions yield the same values for $E_{\mathbf{v}}$ if stimulus overlap is neglected. Instead, we now have to consider the number of overlapping stimulus pairs, and the fraction of stimulus pairs overlapping across the boundaries of tessellation regions. To cast these considerations into mathematical terms we need to introduce further variables. For a stimulus \mathbf{v} , let $N_{\mathbf{v}}$ denote the total number of other stimuli, $N_{\mathbf{v},0}$ the number of other stimuli mapping to the same neuron, $N_{\mathbf{v},1}$ the number of other stimuli mapping to nearest-neighbor neurons, and $N_{\mathbf{v},2}$ the number of other stimuli mapping to next-nearest neighbor neurons. Furthermore let $\tilde{N}_{\mathbf{v}}$ denote the total number of other stimuli which overlap with \mathbf{v} , let $\tilde{N}_{\mathbf{v},0}$ denote the number of other stimuli which overlap with \mathbf{v} and map to the same neuron, let $\tilde{N}_{\mathbf{v},1}$ denote the number of other stimuli which overlap with \mathbf{v} and map to a nearest-neighbor neuron, and let $\tilde{N}_{\mathbf{v},2}$ denote the number of other stimuli which overlap with \mathbf{v} and map

to a next-nearest neighbor neuron. Finally we denote the average contribution to $E_{\mathbf{v}}$ of a pair of stimuli with appreciable overlap by c_1 ,

$$c_1 = \frac{\sum_{\mathbf{v}' \in \Lambda_{\mathbf{v}}} (\mathbf{v}' - \mathbf{v})^2}{\sum_{\mathbf{v}' \in \Lambda_{\mathbf{v}}} 1},$$

where $\Lambda_{\mathbf{v}}$ denotes the set of stimuli which have centers of gravity closer to the center of gravity of \mathbf{v} than the stimulus width (σ_{stim}).

Using these quantities, the contribution $s_{\mathbf{v}}$ of stimulus \mathbf{v} to $E_{\mathbf{v}}$ amounts to

$$\begin{aligned} s_{\mathbf{v}} = & \tilde{N}_{\mathbf{v},0} c_1 e^0 + (N_{\mathbf{v},0} - \tilde{N}_{\mathbf{v},0}) c_0 e^0 + \\ & \tilde{N}_{\mathbf{v},1} c_1 e^{-\frac{1}{2\sigma^2}} + (N_{\mathbf{v},1} - \tilde{N}_{\mathbf{v},1}) c_0 e^{-\frac{1}{2\sigma^2}} + \\ & \tilde{N}_{\mathbf{v},2} c_1 e^{-\frac{4}{2\sigma^2}} + (N_{\mathbf{v},2} - \tilde{N}_{\mathbf{v},2}) c_0 e^{-\frac{4}{2\sigma^2}}. \end{aligned} \quad (4.4)$$

The stimuli \mathbf{v} differ in their position relative to a tessellation region boundary. L_1 stimuli are located close to a boundary to a nearest-neighbor neuron, i.e. have their centers of gravity closer to the boundary than the stimulus width. For these we have $\tilde{N}_{\mathbf{v},0} = \tilde{N}_{\mathbf{v}}/2$, $\tilde{N}_{\mathbf{v},1} = \tilde{N}_{\mathbf{v}}/2$, $\tilde{N}_{\mathbf{v},2} = 0$. Further L_2 stimuli are located close to a boundary to a next-nearest-neighbor neuron ($\tilde{N}_{\mathbf{v},0} = \tilde{N}_{\mathbf{v}}/2$, $\tilde{N}_{\mathbf{v},1} = 0$, $\tilde{N}_{\mathbf{v},2} = \tilde{N}_{\mathbf{v}}/2$ for these). The rest of $N - L_1 - L_2$ stimuli have $\tilde{N}_{\mathbf{v},0} = \tilde{N}_{\mathbf{v}}$, $\tilde{N}_{\mathbf{v},1} = 0$, $\tilde{N}_{\mathbf{v},2} = 0$. Using the sum relation $\tilde{N}_{\mathbf{v}} = \tilde{N}_{\mathbf{v},0} + \tilde{N}_{\mathbf{v},1} + \tilde{N}_{\mathbf{v},2}$, and collecting terms independent of the lengths of the boundaries L_1, L_2 , we finally obtain

$$\begin{aligned} E_{\mathbf{v}} = & \sum_{\mathbf{v}} s_{\mathbf{v}} \\ \approx & const + \left(\frac{L_1}{2}(e^0 - e^{-\frac{1}{2\sigma^2}}) + \frac{L_2}{2}(e^0 - e^{-\frac{4}{2\sigma^2}})\right) \tilde{N}(c_0 - c_1). \end{aligned} \quad (4.5)$$

Using Eq. (4.5), and

$$L_1 = 16s, \quad L_2 = 0, \quad (\text{line-like}) \quad (4.6)$$

$$L_1 = 8 + 8s, \quad L_2 = 0, \quad (\text{box-like}) \quad (4.7)$$

$$L_1 = \frac{8}{\cos \alpha}, \quad L_2 = 8 - 4 \tan \alpha, \quad (\text{house-like}) \quad (4.8)$$

$$L_1 = \frac{16s}{\cos \alpha}, \quad L_2 = 0, \quad (\text{wedge-like}) \quad (4.9)$$

the different four-neuron type solutions depicted in Figs. 2b-e can now be compared (α denotes the angle between the oblique part of the tessellation region boundary and the

horizontal). Obviously, the wedge-like solution has always a longer boundary L_1 than the line-like solution, and, hence, has always a larger value of $E_{\mathbf{v}}$. This corresponds well to the fact that we never observed such solutions in simulations. A transition between the box-like and the line-like solution takes place at $s = 1$. The evaluation of the house-like solution requires first an optimization with regard to the angle α . Performing this optimization, and comparing the resulting $E_{\mathbf{v}}$ with the corresponding values for the box-like and the line-like solutions numerically, we obtain the phase diagram depicted in Fig. 3a. Numerical simulations of maps for various values of σ and s resulted in the phase diagram depicted in Fig. 3b. All different states $a - d$ were observed, and the transitions between these states take place at approximately those values of σ and s determined from our approximate calculation of $E_{\mathbf{v}}$. Considering that the differences between the four-neuron type states are quite subtle, and that consistent numerical results require long convergence times at small values of the learning step size, the coincidence between analytical and numerical phase diagrams can be regarded as very good.

An interesting further aspect of this test problem is the comparison with the corresponding low-dimensional SOFMs. Here, the problem comes down to a crudely discretized variant of the dimension mismatch problem, solved by Ritter and Schulten (1988). Even without a detailed adaptation of their analysis to the present geometry, a roughly linear relation between s_{crit} , the value of s for fixed σ for which the linear solution becomes unstable, and σ should be expected. Evaluation of our distortion measure $E_{\mathbf{v}}$ for SOFMs shows that a box-like solution b' has the lowest energy among the possible non-linear solutions. The transition from the linear solution c' to b' should take place at (see Appendix 1)

$$s_{crit}^2 = \frac{1 + 2 \exp(-\frac{1}{2\sigma^2}) - \exp(-\frac{4}{2\sigma^2})}{1 - \exp(-\frac{4}{2\sigma^2})}. \quad (4.10)$$

Our numerical results for simulations of 4-neuron SOFMs, receiving input from an rectangular input space of width 4 and height $2s$, agree quite well with this analytical result (Fig. 4). Specifically, s_{crit} does not depend on σ for $\sigma < 0.5$ and increases linearly with σ for $\sigma > 0.5$ up to $\sigma = 2$, the maximum value for this geometry. Two factors contribute to the remaining deviations between the constant $s_{crit}(\sigma < 0.5)$ and the slope of $s_{crit}(\sigma > 0.5)$. For small values of σ , the map often failed to converge to classifiable states, resulting in an error margin for the numerical transition values. Secondly, the error involved in the

replacement of equation (3.4) by (3.5) may be responsible for the deviation of about 10% in the regime of large σ . Note that for $\sigma > 1.5$ only the linear solution results analytically as well as numerically. So, the high-dimensional SOM exhibits a two-neuron state, as well as a σ -independent transition from state b to c , whereas the low-dimensional SOFM has no two-neuron states for principal reasons and a σ -dependent transition from b' to c' .

5. Phase Transitions in a SOM-Model for the Development of Orientation Maps

A more interesting application of the SOM-algorithm has first been discussed by Obermayer et al. (1990a). Using elongated ellipses as stimuli, these authors showed that the SOM-algorithm is able to generate orientation maps as observed in the visual cortex. Maps which were driven with stimuli of small or no elongation failed to develop orientation-sensitive cells and, consequently, an orientation map (Obermayer, 1992). Here we set out to calculate the properties of the transition between a map of orientation-insensitive cells, resulting from stimuli with weak orientation, and a map of orientation-sensitive cells, resulting from more strongly oriented stimuli. As in the previous example, our calculation will be based on an approximate comparison of the distortion (3.5) for the two tessellations of stimulus space which correspond to the two map solutions. In the following two subsections we demonstrate two different methods of evaluating Eq. (3.5), an analytic summation using a reduced stimulus set, and a numerical procedure, using the full stimulus set.

5.1 Analytic Results with Reduced Stimulus Set

In order to simplify an analytic evaluation of the sums in Eq. (3.5) we reduce the model to its essentials. In the present context this means using a stimulus ensemble which is as small as possible yet rich enough to generate orientation maps. One such ensemble consists of Gaussian stimuli, with half-widths σ_1 and $\sigma_2 > \sigma_1$, which come in two orientations only (vertical and horizontal), and which are centered at the positions of the input channels. A “vertical stimulus” here means that the width of the (elliptic) stimulus in the y -direction is σ_2 , larger than the width σ_1 in the x -direction. For a “horizontal stimulus”, σ_1 and σ_2

are exchanged. We further assume that the input and output spaces are discretized such that there is a one-to-one relation between input channels and map neurons.

In this reduced model, a non-oriented solution would have receptive fields such that both, the vertically and horizontally oriented stimuli centered at one input channel, would be mapped onto the corresponding map neuron. With this mapping taking place at each input channel we can conclude for symmetry reasons that the receptive fields are non-oriented, i.e., do not have a single preferred orientation. Denoting by $\Delta_{ij||}$ the squared difference between two vertical stimuli separated by a distance i in the x -direction, and by j in the y -direction, and by $\Delta_{ij|-}$, Δ_{ij--} , $\Delta_{ij-|}$ the analogous differences, the distortion $E_{\mathbf{v},non-ori}$ for a map with non-oriented receptive fields is

$$E_{\mathbf{v},non-ori} = N^2 \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} e^{-\frac{i^2+j^2}{2\sigma^2}} \left(\Delta_{ij||} + \Delta_{ij|-} + \Delta_{ij-|} + \Delta_{ij--} \right). \quad (5.1)$$

Using the results for $\Delta_{...}$ derived in App. 2, and replacing the sums of Eq. (3.5) by integrals (in the limit of $\sigma_1, \sigma_2, \sigma \gg 1$), this leads to

$$E_{\mathbf{v},non-ori} = N^2 4\pi^2 \sigma_1 \sigma_2 \sigma^2 \left(4 - \frac{2\sigma_1 \sigma_2}{\sqrt{(\frac{1}{2}\sigma^2 + \sigma_1^2)(\frac{1}{2}\sigma^2 + \sigma_2^2)}} - \frac{4}{\sigma_1^2 + \sigma_2^2 + \sigma^2} \right). \quad (5.2)$$

In contrast, the stimulus ensemble could also be tessellated such that the vertical stimuli at two neighboring positions are mapped onto one neuron, and the resp. horizontal stimuli to a different neuron (see Fig. 5). Obviously, the receptive fields of the neurons would have to be oriented for this tessellation. To take the neighborhood effects into account, we assume that isorientation domains are broad compared to the width of σ . Averaging over the appropriate orientation combinations we find

$$\begin{aligned} E_{\mathbf{v},ori} &= N^2 \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} e^{-\frac{i^2+j^2}{2\sigma^2}} (\Delta_{2ij||} + \Delta_{2ij--} + \\ &\quad \frac{1}{2}\Delta_{2i+1j||} + \frac{1}{2}\Delta_{2i-1j||} + \frac{1}{2}\Delta_{2i+1j--} + \frac{1}{2}\Delta_{2i-1j--}) \\ &= N^2 4\pi^2 \sigma_1 \sigma_2 \sigma^2 \left(4 - \left(1 + e^{-\frac{1}{2\sigma_1^2} \frac{\sigma_1^2 + \sigma^2}{\sigma_1^2 + 2\sigma^2}} \right) \frac{\sigma_1 \sigma_2}{\sqrt{(2\sigma^2 + \sigma_1^2)(\frac{1}{2}\sigma^2 + \sigma_2^2)}} - \right. \\ &\quad \left. \left(1 + e^{-\frac{1}{2\sigma_2^2} \frac{\sigma_2^2 + \sigma^2}{\sigma_2^2 + 2\sigma^2}} \right) \frac{\sigma_1 \sigma_2}{\sqrt{(2\sigma^2 + \sigma_1^2)(\frac{1}{2}\sigma^2 + \sigma_2^2)}} \right). \quad (5.3) \end{aligned}$$

Equating (5.2) and (5.3) now yields the transition point, i.e. the critical value for the long half-axis σ_2 of the elliptic stimuli, as a function of the width σ_1 of the short half-axis

and of the width σ of the neighborhood function. Fig. 6 shows the results of a numerical evaluation of the transition point for general values of σ_1, σ . A very interesting aspect of the results depicted in Fig. 6 is that the critical σ_2 is related to σ_1 by an additive constant.

This behaviour can be made more explicit in at least one limiting case, the case of large stimuli as compared to the neighborhood width σ ($\sigma_1, \sigma_2 \gg \sigma$). Then, we obtain from equating (5.2) and (5.3) after cutting common factors and developing to first order in $(\sigma/\sigma_{1,2})^2$,

$$(\sigma_2 - \sigma_1)^2 \approx \frac{3}{2} \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1 \sigma_2} \sigma^2. \quad (5.4)$$

Further approximation of (5.4) yields

$$\sigma_2 \approx \sigma_1 + \sqrt{3}\sigma. \quad (5.5)$$

This further analysis not only explicitly reiterated the above-mentioned linear dependence between σ_2, σ_1 . We also find the additive constant to be proportional to the neighborhood width σ .

In the corresponding low-dimensional SOFM-model, the comparable stimulus parameter is the orientation specificity, which for elliptic stimuli is best identified with the eccentricity σ_2/σ_1 . For orientation columns to emerge, a critical eccentricity has to be exceeded, with the critical value of σ_2/σ_1 being proportional to the neighborhood width σ (Obermayer et al., 1992). So for the SOFM, σ_1 and the critical σ_2 are related by a multiplicative constant, whereas in the case of the SOM it is an additive constant, a somewhat surprising difference. Both constants depend linearly on the neighborhood width σ .

5.2 Numerical Results with Reduced Stimulus Set

How do the analytical results obtained in the last section compare to simulations of orientation maps? In this and the next subsection we address this question in two steps. First we simulate a map model which is as close as possible to the simplified map model analyzed in the last section. Comparison of the numerical results to Eq. (5.5) will allow us to assess the quality of our analytical methods. In the next subsection we then proceed to present numerical (plus some analytical) results for the full SOM orientation map model.

This second step will allow us to assess the consequences of the simplification of the orientation map model considered in sections 5.1 and 5.2.

We first simulated orientation maps of extension 15×15 , stimulated by ellipses with Gaussian profile, width σ_1 and $\sigma_2 \geq \sigma_1$, centered randomly at one of the 15×15 input channels, with either a horizontal or a vertical orientation. For each value σ of the SOM neighborhood width and σ_1 of the shorter of the two widths of the ellipse, we varied the elongation σ_2 of the longer width of the stimuli. For the resulting maps, the average degree of orientation $O(\mathbf{x})$ of the receptive fields was determined by using the same measure as Miller (1994),

$$O(\mathbf{x}) = \frac{\|v_R(\mathbf{x})\|}{\sqrt{\sum_{n=0}^{N-1} R(\mathbf{x}, n)^2}} \quad . \quad (5.6)$$

where $v_R(\mathbf{x}) = \sum_{n=0}^{N-1} R(\mathbf{x}, n) \exp(2\pi i n/N)$ is a complex number equal to the sum of the best responses $R(\mathbf{x}, n)$ of the receptive field to stimuli with an orientation in the interval $[n\pi/N, (n+1)\pi/N]$ multiplied by a complex vector on the unit circle with the corresponding orientation.

With this measure, an orientation specificity of $O = 0$ corresponds to a receptive field that responds equally well to stimuli with arbitrary orientation. Oriented receptive fields \mathbf{x} yield $O(\mathbf{x}) > 0$. A typical result of the orientation specificity as a function of the stimulus elongation σ_2 is depicted in Fig. 7. The prominent feature is the steep increase of $O(\mathbf{x})$ beyond a critical value of σ_2 . In a second stage of the simulations, we computed the critical values for σ_2 for various combinations of σ and σ_1 (see Fig. 8). For each of the four exemplary values of σ , $\sigma_{2,crit}$ depends linearly on σ_1 ,

$$\sigma_{2,crit} = a\sigma_1 + b \quad . \quad (5.7)$$

Doing an autoregression analysis on the $\sigma_{2,crit}(\sigma_1)$ -curves yields the parameters $a(\sigma), b(\sigma)$ (with correlation coefficients $r > 0.997$). As shown in Fig. 9, the slope parameter has a value of $a \approx 1$, and does not depend on σ . The offset parameter b depends linearly on σ . Doing a second autoregression analysis, we finally obtain

$$\sigma_{2,crit} = 1.04\sigma_1 + 0.95\sigma + 0.45 \quad . \quad (5.8)$$

These numerical results coincide quite well with the analytically derived relation (5.5). The linear dependence of $\sigma_{2,crit}$ on σ_1 is confirmed, with a σ -independent slope parameter

$a \approx 1$, and a σ -proportional offset parameter b . For an assessment of the deviation between the numerically observed $b = 0.95\sigma + 0.45$ and the analytically derived $b = \sqrt{3}\sigma$, one has to take into account the complicated three-step extraction process leading to the numerical value, as well as the various approximations used to derive Eq. (5.5). Furthermore, the rather small system size might also lead to some quantitative deviations. In summary we regard the agreement between calculation and simulation as very good.

5.3 Results with Full Stimulus Set

In order to carry out the analytic evaluation of $E_{\mathbf{v}}$ for the orientation map, we considered a reduced stimulus set. Now, we lift this restriction and investigate orientation maps with full stimulus ensembles, i.e. with ensembles containing a large number of possible center positions and orientations.

Two complications with regard to an evaluation of the $E_{\mathbf{v}}$ can result. If we still know the tessellation of stimulus space, but cannot analytically evaluate the sums in Eq. (3.5) anymore, we can simply take resort to a numerical summation. In the present example, with many possible stimulus orientations and locations per neuron, this is the case for the trivially structured map with non-oriented receptive fields. A more difficult case arises if we cannot make simple assumptions about the tessellation anymore (as is the case for the map solution with oriented receptive fields). Here the following approximation scheme for $E_{\mathbf{v}}$ can be applied: One assumes that each neuron is best-matching for an equal number ν of stimuli. One then picks one test stimulus \mathbf{v}_{test} , and determines the $\nu - 1$ further stimuli which lie closest to \mathbf{v}_{test} . The mutual squared distances evaluated for this group of stimuli then yield an approximation to $E_{\mathbf{v}}$ in the limit of $\sigma \rightarrow 0$. Using this approximation, we find for the full stimulus set again a linear relation $\sigma_{2,crit} = \sigma_1 + b$ (see Fig. 10). Simulations of SOM-orientation maps with full stimulus ensemble finally yield results very similar to those obtained with the restricted stimulus ensemble (Fig. 10), in this way justifying a posteriori the approximation used in subsections 5.1 and 5.2.

6. Discussion und Conclusions

In the present paper we introduced a new method to evaluate structures emerging in high-dimensional SOMs. This method opens a first analytical access to the phase tran-

sition properties of this rather general algorithm for neural map self-organization. Our method does not require the evaluation of the receptive field profiles which emerge for the individual neurons. Instead, it relies on an evaluation of the way the stimulus space is tessellated, i.e. distributed among the different neurons. This is advantageous, because the receptive field profiles change with varying parameters of the stimulus distribution or the learning algorithm even within one qualitative type of map solution whereas the tessellation remains identical.

The analysis of two exemplary problems showed that the phase transition properties of high-dimensional SOM-maps can differ from those of low-dimensional SOFMs. In the case of the 4-neuron example, we found a state of the high-dimensional map, the 2-neuron type solution, which cannot occur in the low-dimensional map at all. In addition, while the transition from the trivial line-like solution to the non-trivial box-like and house-like solutions showed a strong dependence on the neighborhood width σ in the SOFM model, the transition in the high-dimensional problem was well determined by the geometry of the problem alone.

Our second example is the analysis of a SOM-based model for the development of orientation maps. The high degree of abstraction of the SOM provokes questions with regard to the biological plausibility of such a model. Here, one of the central issues is that of cortical “competition”, causing a localized cortical response to a single stimulus, with a strongly peaked point spread function. This assumption is realized in a particularly straightforward way by the SOM-winner-take-all rule, plus the Gaussian neighborhood function. In more explicit form, cortical competition by lateral inhibition was already incorporated in other, less abstracted models (Malsburg, 1973) and was recently observed in a highly detailed simulation of cortical response properties to oriented stimuli (Somers et al., 1995). Second, high-dimensional SOM-models as investigated in this paper describe and self-organize receptive fields in terms of synaptic weight distributions. This level of description avoids the often used feature map approximation and thus is easier to interpret in biological terms. It allows to investigate maps with receptive fields which cannot be parametrized by linear operators. Third, the present model assumes a single input layer with oriented stimuli. Such stimuli would occur in patterned vision, the role of which for

the development of orientation maps has been underlined in a study by Chapman and Stryker (1993). Our study shows that under these circumstances a critical orientation of the stimuli has to be exceeded for oriented receptive fields to develop.

This result is particularly interesting in comparison to other high-dimensional orientation map models. In so-called correlation-based models, it was found that oriented receptive fields can emerge as a consequence of competition between two layers of ON-center and OFF-center cells even if the correlation function of activity between the two layers is circularly symmetric (Miller, 1992, 1994, Miyashita and Tanaka, 1992). There, orientation specificity results from a segregation between ON-center and OFF-center inputs within a single receptive field. Even though these models are formulated in terms of correlation functions, and not of individual stimuli, their results should also apply to a set of circular symmetric stimuli, i.e. to a vanishing difference between elongations σ_2 and σ_1 . While the present study proves that a corresponding break of symmetry cannot occur with stimuli in a single input layer, the question has to be addressed whether or not the SOM framework allows a break of rotational symmetry if the projection assumption of the correlation-based models is adopted. In case the ON-center and OFF-center segregation mechanism can also occur in principle in a SOM, then the present antagonism between this scenario and the assumption of oriented stimuli will give way to a more combined point of view. Quantitative aspects of orientation map development, like those raised by the Chapman and Stryker experiment, can be addressed by adjusting the balance between the two development mechanisms. We expect that our method will prove to be a valuable tool in the investigation of these questions.

Appendix 1: Calculation of the critical box-width for the 4-neuron map

To calculate the phase transition for the SOFM-version of the box-to-4-neuron example of section 4 we need to evaluate Eq. (3.5) for the two tessellations of the input space induced by the line-like and box-like map solutions. Since we have a continuum of stimuli now, we have to replace the sums by integrals. Further, the periodic boundary conditions have

to be taken into account. Exploiting the symmetries of the geometry and the symmetry between the 4 neurons, we arrive at the following distortion formulas:

1. Line-like solution:

$$\begin{aligned}
E_{\mathbf{v}} &= 4 \left(\int_0^1 dx \int_0^1 dx' \int_{-s}^s dy \int_{-s}^s dy' \{ (x-x')^2 + (y-y')^2 \} + \right. \\
&\quad 2e^{-\frac{1}{2\sigma^2}} \int_1^2 dx \int_0^1 dx' \int_{-s}^s dy \int_{-s}^s dy' \{ (x-x')^2 + (y-y')^2 \} + \\
&\quad \left. 2e^{-\frac{4}{2\sigma^2}} \int_2^{5/2} dx \int_0^1 dx' \int_{-s}^s dy \int_{-s}^s dy' \{ (x-x')^2 + (y-y')^2 \} \right) \\
&= 4 \left(\frac{8s^4}{3} (1 + 2e^{-\frac{1}{2\sigma^2}} + e^{-\frac{4}{2\sigma^2}}) + \frac{2s^2}{3} (1 + 14e^{-\frac{1}{2\sigma^2}} + 19e^{-\frac{4}{2\sigma^2}}) \right). \tag{7.1}
\end{aligned}$$

2. Box-like solution:

$$\begin{aligned}
E_{\mathbf{v}} &= 4 \left(\int_0^2 dx \int_0^2 dx' \int_0^s dy \int_0^s dy' \{ (x-x')^2 + (y-y')^2 \} + \right. \\
&\quad e^{-\frac{1}{2\sigma^2}} \int_0^2 dx \int_0^2 dx' \int_0^s dy \int_{-s}^0 dy' \{ (x-x')^2 + (y-y')^2 \} + \\
&\quad 2e^{-\frac{1}{2\sigma^2}} \int_0^2 dx \int_{-1}^0 dx' \int_0^s dy \int_0^s dy' \{ (x-x')^2 + (y-y')^2 \} + \\
&\quad \left. 2e^{-\frac{4}{2\sigma^2}} \int_0^2 dx \int_{-1}^0 dx' \int_0^s dy \int_{-s}^0 dy' \{ (x-x')^2 + (y-y')^2 \} \right) \\
&= 4 \left(\frac{2s^4}{3} (1 + 8e^{-\frac{1}{2\sigma^2}} + 7e^{-\frac{4}{2\sigma^2}}) + \frac{8s^2}{3} (1 + 5e^{-\frac{1}{2\sigma^2}} + 4e^{-\frac{4}{2\sigma^2}}) \right). \tag{7.2}
\end{aligned}$$

Equating the two distortions to obtain the critical value for s^2 then yields Eq. (4.10).

Appendix 2: Calculation of Stimulus Overlap Integrals

For the calculation of the tessellation distortions in subsection 5.1, some stimulus overlap integrals $\Delta_{ij..}$ were needed, where i denotes the horizontal displacement between stimuli, j their vertical displacement, and $..$ indicates vertical ($|$) or horizontal ($-$) orientation of the stimuli. Assuming large stimuli as compared to the lattice constant of the input space ($\sigma_1, \sigma_2 \gg 1$), we can approximate the sums occurring in Eq. (3.5) by integrals and find for the first such overlap

$$\Delta_{ij||} = \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \left(e^{-\frac{x^2}{2\sigma_1^2}} e^{-\frac{y^2}{2\sigma_2^2}} - e^{-\frac{(x-i)^2}{2\sigma_1^2}} e^{-\frac{(y-j)^2}{2\sigma_2^2}} \right)^2$$

$$= 2\pi\sigma_1\sigma_2 \left(1 - e^{-\frac{i^2}{2\sigma_1^2}} e^{-\frac{j^2}{2\sigma_2^2}} \right). \quad (8.1)$$

For Δ_{ij--} we can use the relation

$$\Delta_{ij--} = \Delta_{ji||}. \quad (8.2)$$

Finally, combined horizontal and vertical orientations yield

$$\Delta_{ij|-} = \Delta_{ji-|} \quad (8.3)$$

$$= \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \left(e^{-\frac{x^2}{2\sigma_1^2}} e^{-\frac{y^2}{2\sigma_2^2}} - e^{-\frac{(x-i)^2}{2\sigma_2^2}} e^{-\frac{(y-j)^2}{2\sigma_1^2}} \right)^2$$

$$= 2\pi\sigma_1\sigma_2 \left(1 - \frac{2\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2} e^{-\frac{i^2+j^2}{2(\sigma_1^2+\sigma_2^2)}} \right). \quad (8.4)$$

Acknowledgements

This work has been supported by the Deutsche Forschungsgemeinschaft through SFB 185 “Nichtlineare Dynamik”, TP E3. One of the authors (H.-U. B.) would also like to acknowledge the kind hospitality and support experienced during a stay at the International Computer Science Institute (ICSI), Berkeley, USA.

References

- H.-U. Bauer, 1995. Development of oriented ocular dominance bands as a consequence of areal geometry, *Neural Computation* **7**, 36-50.
- B. Chapman, M. P. Stryker, 1993. Development of orientation selectivity in ferret visual cortex and effects of deprivation, *J. Neurosci.* **13**, 5251-5262.
- E. Erwin, K. Obermayer, K. Schulten, 1992. Self-organizing maps: ordering, convergence properties and energy functions, *Biol. Cyb.* **67**, 47-55.
- E. Erwin, K. Obermayer, K. Schulten, 1995. Models of orientation and ocular dominance columns in the visual cortex: A critical comparison, *Neur. Comp.* **7**, 425-468.
- G. J. Goodhill, 1993. Topography and ocular dominance: a model exploring positive correlations, *Biol. Cyb.* **69**, 109-118.

- T. M. Heskes, B. Kappen, 1993. Error potentials for self-organization, Proc. IEEE ICNN 1993, IEEE Press, 1219.
- S. Luttrell, 1994. A Bayesian analysis of Self-Organizing Maps, *Neur. Comp.* **6**, 767.
- T. Kohonen, 1993. Physiological interpretation of the self-organizing map algorithm, *Neur. Netw.* **6**, 895-906.
- T. Kohonen, 1995. The self-organizing map, Springer, Berlin.
- C. von der Malsburg 1973. Self-organization of orientation sensitive cells in the striate cortex, *Kybernetik* **14**, 81-100.
- K. D. Miller, J. B. Keller, M. P. Stryker, 1989. Ocular dominance column development: Analysis and simulation, *Science* **245**, 605-615.
- K. D. Miller, 1992. Development of orientation columns via competition between ON- and OFF-center inputs, *NeuroRep.* **3**, 73-79.
- K. D. Miller, 1994. A model for the development of simple-cell receptive fields and the ordered arrangement of orientation columns through activity dependent competition between On- and Off-center inputs, *J. Neurosci.* **14**, 409-441.
- K. D. Miller, 1995. Derivation of low-dimensional feature map representation of ocular dominance from a high-dimensional synaptic representation, submitted to *Neur. Comp.*
- M. Miyashita, S. Tanaka, 1992. A mathematical model for the self-organization of orientation columns in visual cortex, *NeuroRep.* **3**, 69-72.
- K. Obermayer, H. Ritter, K. Schulten, 1990a. A principle for the formation of the spatial structure of cortical feature maps, *Proc. Nat. Acad. Sci.* **87**, 8345-8349.
- K. Obermayer, H. Ritter, K. Schulten, 1990b. Large-scale simulations of self-organizing neural networks on parallel computers: application to biological modelling, *Parallel Comp.* **14**, 381-404.

K. Obermayer, 1992. Adaptive Neuronale Netze und ihre Anwendung als Modelle der Entwicklung kortikaler Karten, infix Verlag, Sankt Augustin.

K. Obermayer, G. G. Blasdel, K. Schulten, 1992. Statistical-mechanical analysis of self-organization and pattern formation during the development of visual maps, Phys. Rev. A **45**, 7568-7589.

H. Ritter, K. Schulten, 1986. On the stationary state of Kohonen's self-organizing sensory mapping, Biol. Cybern. **54**, 99-106.

H. Ritter, K. Schulten, 1988. Convergence properties of Kohonen's topology conserving maps: Fluctuations, stability and dimension selection, Biol. Cybern. **60**, 59-71.

H. Ritter, T. Martinetz, K. Schulten, 1992. Neural computation and self-organizing maps. Addison Wesley, Reading, Mass.

D. C. Somers, S. B. Nelson, M. Sur, 1995. An emergent model of orientation selectivity in cat visual cortical simple cells, J. Neurosci. **5**, 5448-5465.

G. G. Sutton and J. A. Reggia, 1994. Effects of normalization constraints on competitive learning, IEEE Trans. Neur. Netw. **5**, 502-504.

V. V. Tolat, 1990. An analysis of Kohonen's self-organizing maps using a system of energy functions, Biol. Cyb. **64**, 155-164.

F. Wolf, H.-U. Bauer, T. Geisel, 1994. Formation of field discontinuities and islands in visual cortical maps, Biol. Cyb. **70**, 525-531.

Figure Captions

Fig. 1: Illustration of differences in the distance metric between high-dimensional (**a,b**) and low-dimensional (**c,d**) maps. In each case, the input space is a square. For the high-dimensional case, the input space is discretized as a square lattice, the double bar stimulus is indicated as the (combined) gray regions, and the two exemplary receptive fields in **a** and **b** are the (combined) black regions. The stimulus-receptive field overlap is larger in **b**, so the respective neuron will be the winner in the high-dimensional case. In the low-dimensional variant of this example, stimuli and receptive fields are represented by their centers of gravity (crosses and triangles, resp.). Now, the neuron depicted in **c** (and **a**) is the winner, i.e. the high- and low-dimensional variants of the map yield different winners.

Fig. 2: Caricatures (**a-e**) and results of simulations (**f-j**) for receptive fields of neurons in the high-dimensional variant of the 4-neuron problem (see text). An input space of $4M \times 2sM$, $M = 10$, input channels, arranged as a rectangle, is mapped onto a ring of 4 neurons. Parts **a-e** indicate regions of input channels, such that stimuli centered within these regions are mapped to the same neuron. Either all stimuli are mapped to only two of the neurons (**a**), or to all four neurons, with the respective regions being arranged in a line-like (**b**), a box-like (**c**), a house-like (**d**) or a wedge-like (**e**) fashion. Depending on s and σ , the first four of these solutions are found numerically. Corresponding synaptic weight distributions \mathbf{w}_r for one exemplary neuron are depicted in (**f-i**) as gray value images (f: $s = 1.3$, $\sigma = 1.13$, g: $s = 0.8$, $\sigma = 0.78$, h: $s = 1.2$, $\sigma = 0.78$, i: $s = 1.2$, $\sigma = 0.35$, j: not observed; parameters for all simulations: 10^6 iterations, learning rate ϵ reduced exponentially from 0.1 to 0.0005).

Fig. 3: Phase diagrams for high-dimensional SOMs for the four-neuron example. (**a**): analytically determined phase diagram in the σ - s plane. The letters indicate the type of solution in each of the regions (notation analogous to Fig. 2). (**b**): numerical results superimposed on the analytical phase diagram. The letters indicate the numerically obtained type of map at the respective position in the σ - s plane, averaged over at least three realizations of the map at each parameter combination. The letter x

denotes conflicting numerical results. Number of iterations and learning schedule same as in Fig. 2.

Fig. 4: Phase diagram for low-dimensional SOFMs version of the four-neuron example. Here we evaluate only the transition between a line-like solution (indicated as \mathbf{b}') and a box-like solution (\mathbf{c}' , see text). Solid line: analytical result (4.10), dashed line: numerical result (1.5×10^7 iterations, ϵ reduced exponentially from 0.02 to 0.0001, virtually identical results were obtained with $\epsilon : 0.1 \rightarrow 0.0001$).

Fig. 5: Illustration of possible tessellations of a reduced stimulus space in a SOM-model for the development of orientation maps. Stimuli are indicated as ellipses in the lower square (retinal space), map neurons as crosses in the upper square (orientation map). Case (a): two stimuli of different orientation but located at the same position in retinal space are mapped to the same neuron (corresponding to a non-oriented receptive field). Case (b): two stimuli with the same orientation but centered at different locations are combined (corresponding to oriented receptive fields). In the latter case we assume for the arrangement of receptive fields in the map that neighboring neurons receive input from stimuli of the same orientation.

Fig. 6: Critical length $\sigma_{2,crit}$ as a function of σ and σ_1 , obtained by equating $E_{\mathbf{v},non-ori}$ and $E_{\mathbf{v},ori}$ as given by Eqs. (5.2) and (5.3). Note the linear increase of $\sigma_{2,crit}$ as a function of σ as well as of σ_1 .

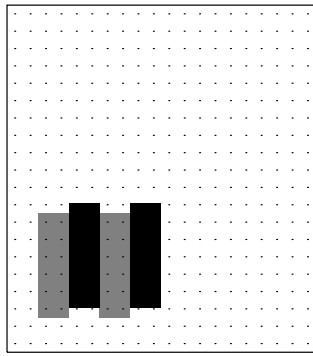
Fig. 7: Degree of orientation preference O in high-dimensional SOMs as a function of σ_2 , the longer half-axis of the elliptic stimuli. The maps were simulated using the reduced stimulus set described in subsection 5.1, simulation parameters were $\sigma = 1.06$, $\sigma_1 = 2.12$, $\epsilon : 0.1 \rightarrow 0.001$ in 2×10^5 steps. The error bars result from averaging O over the receptive fields.

Fig. 8: Critical elongations $\sigma_{2,crit}$ for the non-oriented—oriented transition, as a function of the elongation σ_1 at different values of σ . The symbols indicate the results of simulations, the solid lines are autoregression fits to these points. The dotted lines show the extrapolation back to $\sigma_1 = 0$. Other simulation parameters same as in Fig. 7.

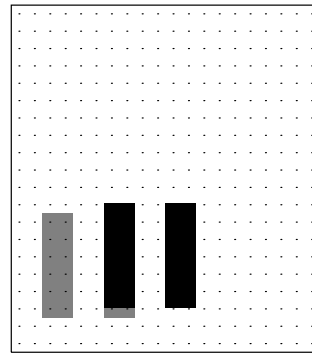
Fig. 9: Slope $a(\sigma)$ (stars) and offset $b(\sigma)$ (crosses) of the lines $\sigma_{2,crit} = a\sigma_1 + b$ in Fig. 8. The lines are the results of an autoregression analysis of the $a(\sigma), b(\sigma)$ -points.

Fig. 10: Analogous to Fig. 8, but for the full stimulus ensemble.

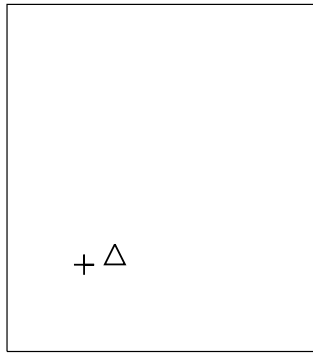
Fig. 1:



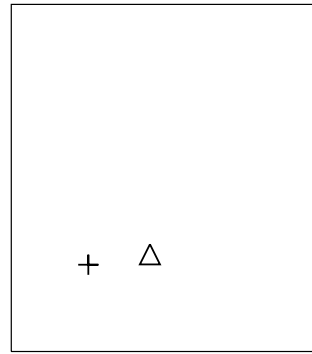
a



b

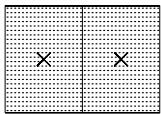


c

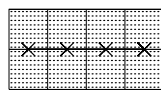


d

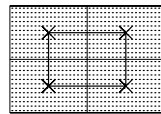
Fig. 2:



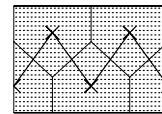
a



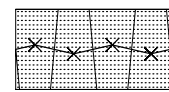
b



c



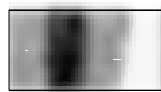
d



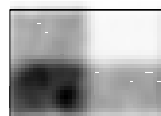
e



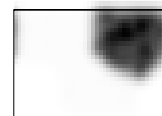
f



g



h



i

Not observed

j

Fig. 3:

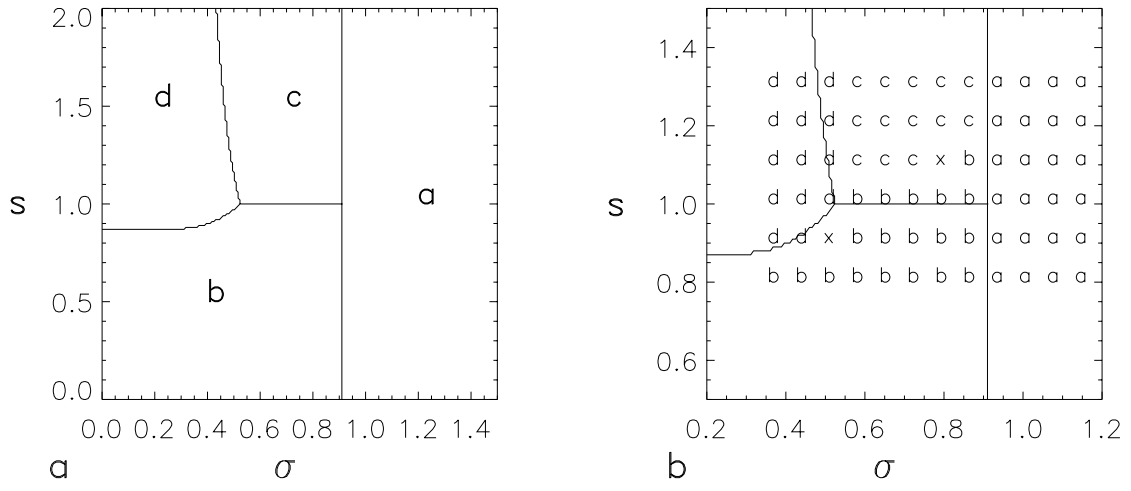


Fig. 4:

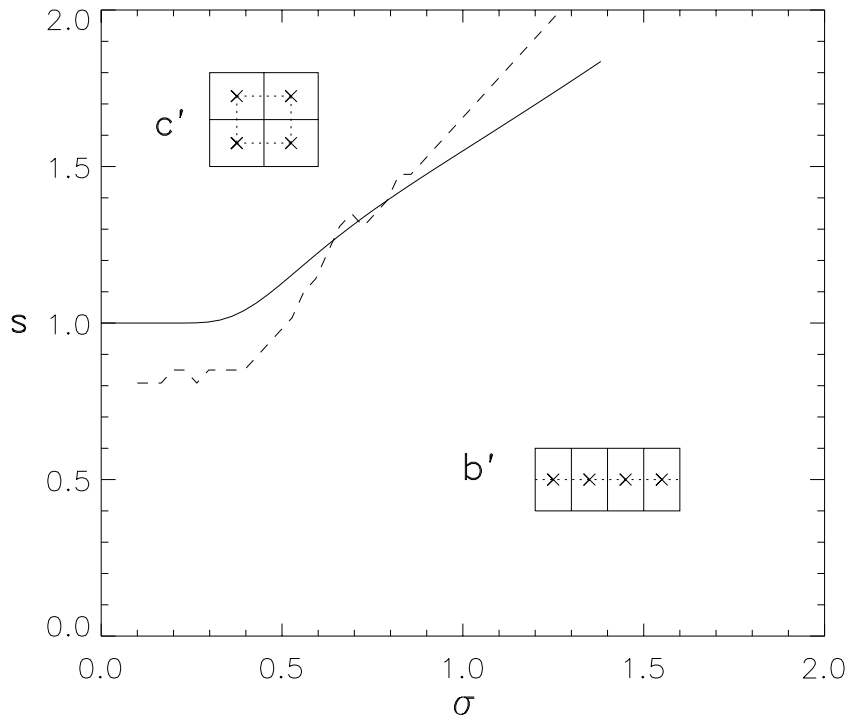


Fig. 5:

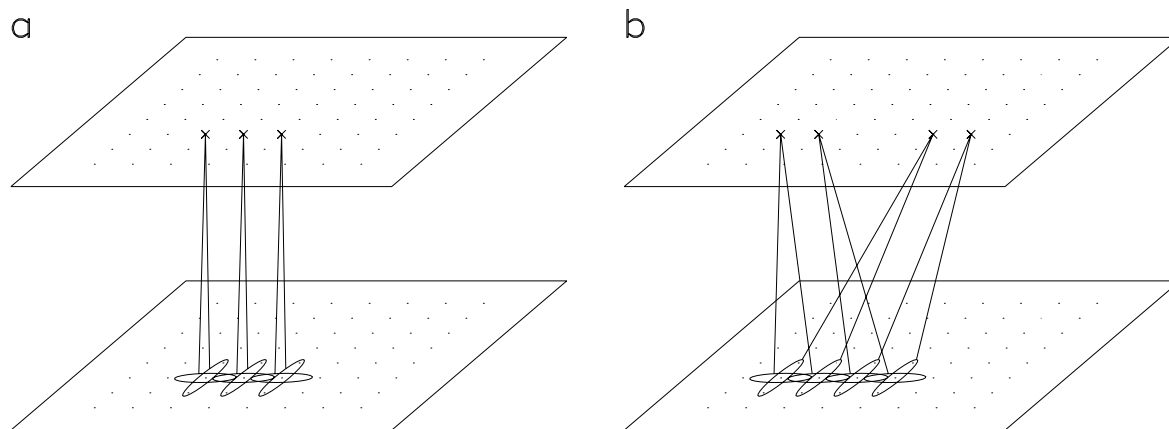


Fig. 6:

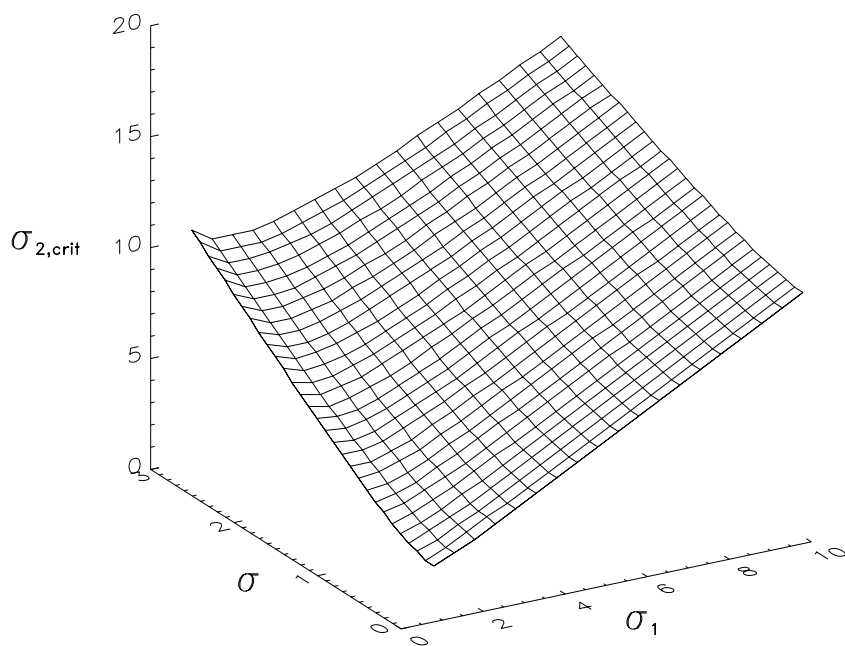


Fig. 7:

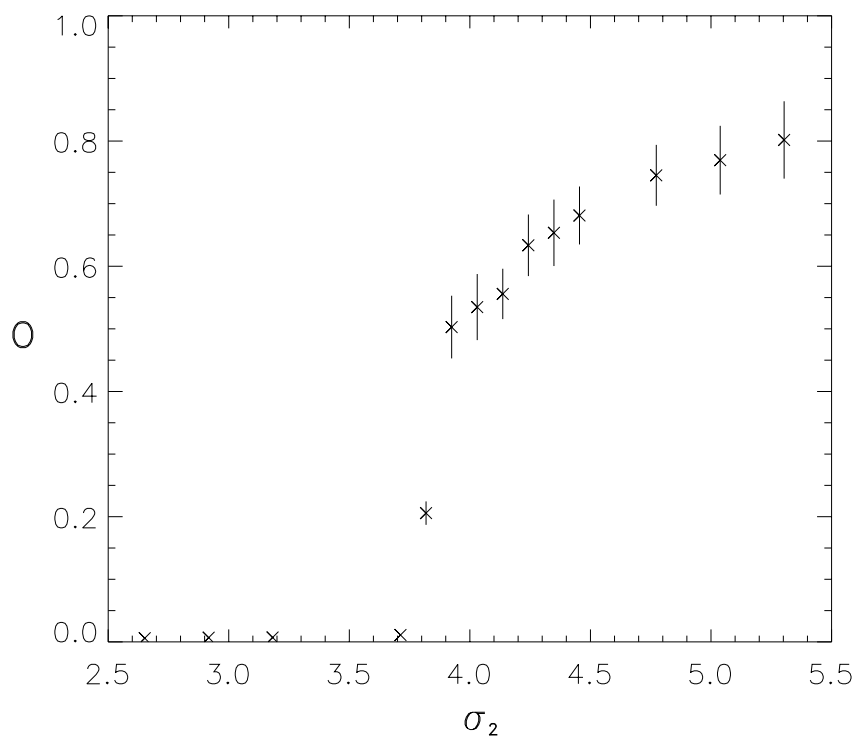


Fig. 8:

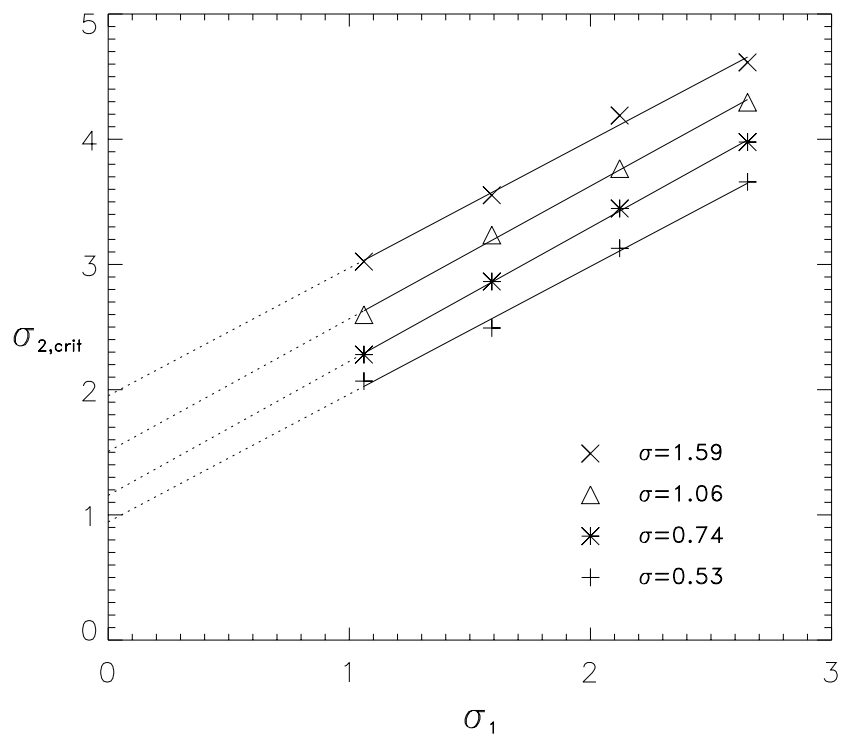


Fig. 9:

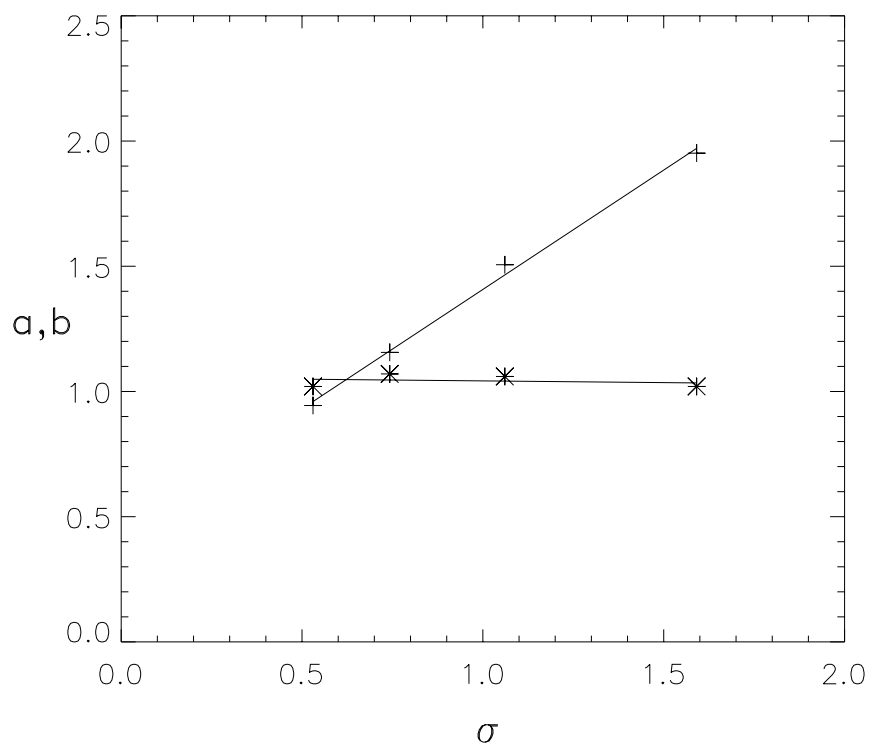


Fig. 10:

