Biol. Cybern. 60, 59–71 (1988)

Biological Cybernetics
© Springer-Verlag 1988

Convergence Properties of Kohonen's Topology Conserving Maps: Fluctuations, Stability, and Dimension Selection

H. Ritter and K. Schulten

Physic-Department, Technische Universität München, James-Franck-Strasse,
D-8046 Garching, Federal Republic of Germany

Abstract. We analyse a Markovian algorithm for the formation of topologically correct feature maps proposed earlier by Kohonen. The maps from a space of input signals onto an array of formal neurons are generated by a learning scheme driven by a random sequence of input samples. The learning is described by an equivalent Fokker-Planck equation. Convergence to an equilibrium map can be ensured by a criterion for the time dependence of the learning step size. We investigate the stability of the equilibrium map and calculate the fluctuations around it. We also study an instability responsible for a phenomenon termed by Kohonen "automatic selection of feature dimensions".

1 Introduction

Pattern recognition and signal processing tasks can be facilitated considerably by appropriate encodings of the relevant input signals. Very often the inputs are elements in a high-dimensional space and encodings are highly desirable which capture the essential data interrelationships in a subspace of only few dimensions. Such encoding schemes correspond to maps which project onto lower-dimensional spaces such that, to the degree possible, topological (neighborhood) relationships are conserved. Brains of many higher animals appear to achieve such maps through a stream of sensory inputs, the target space being two-dimensional sheets of neural networks (Kaas et al. 1983; Knudsen et al. 1987; Suga and O'Neill 1979). Several algorithms have been suggested with the objective to account for the underlying neural processes. The algorithms possess the capability of establishing topology conserving maps from a random sequence of input samples by learning (Grossberg 1976a, b; Willshaw and v. d. Malsburg 1976, 1979; v. d. Malsburg 1979; Takeuchi and Amari 1979; Kohonen 1982a–c; Overton and Arbib 1982; Erdi and Barna 1984).

In this paper we want to focus on a particular algorithm proposed earlier by Kohonen (Kohonen 1982a–c). The benefit of Kohonen's algorithm lies in its simple computational form, which besides providing a plausible neural model allows its efficient application to pattern recognition and control tasks, such as speech recognition (Kohonen 1984a; Kohonen et al. 1984), image processing (Bertsch and Dengler 1987) and motor learning for robots (Ritter and Schulten 1986b, 1987).

The algorithm's aim is to generate a mapping of a higher dimensional space V spanned by the inputs onto an, usually two-dimensional, array of formal neurons. The map is generated by establishing a correspondence between inputs from V and neurons in the array such that the topological (neighborhood) relationships among the inputs are reflected as faithfully as possible in the arrangement of the corresponding neurons in the array. The correspondence is obtained iteratively by a sequence of training steps. Each training step requires the presentation of an input randomly chosen from the space V. The input activates a localized subset of neurons in the array, whose synaptic weights are then adjusted such as to improve their response to a subsequent reoccurrence of the activating input.

The above procedure is represented mathematically as a Markov process whose states are the synaptic weights of the formal neurons and whose transitions are triggered by the inputs. Several mathematical properties of this process have already been investigated (Kohonen 1982b, 1984, 1986; Ritter and Schulten 1986a; Cottrell and Fort 1986), most notably among them the dependence of the final weights upon the statistical distribution of the inputs and proofs of convergence to a stationary map under different conditions. Further interesting properties, which so far have received less attention and which are the subject of this paper, are *(i)* a criterion for the choice of a suitable decrease of the learning step size with time guaranteeing convergence to an equilibrium map during the final phase of the algorithm, *(ii)* the statistical fluctuations of the evolving map brought about by statistical occurrence of inputs and *(iii)* a mathematical analysis of the phenomenon called "automatic selection of feature dimensions" and studied earlier in simulations by Kohonen (Kohonen 1984a).

Our approach starts with a derivation of a Fokker-Planck equation describing the learning process in the vicinity of equilibrium maps and valid in the limit of small learning step size. This allows to provide a necessary and sufficient condition which guarantees convergence of the learning scheme to an asymptotic equilibrium map in the final convergence phase of the algorithm. The condition concerns the choice of a proper time sequence for the learning step sizes. For the case of inputs chosen uniformly from a space with a shape of a multi-dimensional parallelepiped, we calculate the statistical fluctuations around the asymptotic equilibrium map. This shows that for the map to differ from the asymptotic equilibrium map by the order of the lattice spacing of the neurons or less, it is necessary to scale the final learning step size inversely proportional to the number of neurons in the array. We also analyse the phenomenon of "automatic selection of feature dimensions" and show that it can be understood as occurrence of an instability, which arises if the variance of the inputs along one of the dimensions not well represented in the map exceeds a critical value. The instability is preceeded by the occurrence of large fluctuations of a characteristic wavelength. Both, critical variance and characteristic wavelength, are calculated for the case of a multi-dimensional parallelepiped as input space.

2 The Algorithm

This section gives a brief account of Kohonen's algorithm. For more details see, for example, (Kohonen 1982a, c, 1984).

The algorithm employs an array A of formal neurons receiving a random sequence of input samples from a space V to be mapped onto A. Each input $\mathbf{v} \in V$ is represented as a vector of activities $v_1, v_2, \ldots, v_d$ on d input lines, where d is the dimension of the space V. Each neuron is labelled by its position $\mathbf{r} \in A$ and connected to all input lines $l = 1 \ldots d$ via "synaptic strengths" w_{rl}, $l = 1 \ldots d$. To refer to all synaptic strengths of a neuron $\mathbf{r}$ simultaneously, we use the vector $\mathbf{w}_r := (w_{r1}, w_{r2}, \ldots, w_{rd})^T$. An input $\mathbf{v}$ induces an "excitation" $\alpha(\mathbf{v}) = h(\mathbf{v} \cdot \mathbf{w}_r)$ of neuron $\mathbf{r}$, where $h(\cdot)$ is some "sigmoidal" function between 0 and 1. This excitation leads to a modification of the neuron's synaptic strengths. Assuming a Hebb-like rule with an additional memory decay term of equal strength, the coincidence of presynaptic input $\mathbf{v}$ and postsynaptic excitation $h(\mathbf{v} \cdot \mathbf{w}_r)$ alters $\mathbf{w}_r$ by

$$\delta \mathbf{w}_r = h(\mathbf{w}_r \cdot \mathbf{v})(\mathbf{v} - \mathbf{w}_r). \tag{1}$$

In order to incorporate the strong lateral inhibition acting among real neurons which are some distance apart, Kohonen suggested to apply (1) only in the vicinity of the neuron $\mathbf{s}$ most vigorously excited by $\mathbf{v}$. This neuron is determined by

$$\|\mathbf{w}_s - \mathbf{v}\| = \min_{\mathbf{r} \in A} \|\mathbf{w}_r - \mathbf{v}\|. \tag{2}$$

In Kohonen's algorithm (1) is then replaced by

$$\delta \mathbf{w}_r = \varepsilon \cdot h^0_{rs} \cdot (\mathbf{v} - \mathbf{w}_r) \quad \text{for all } \mathbf{r} \in A, \tag{3}$$

where $0 \leqq h^0_{rs} \leqq 1$ now is a prespecified adjustment function of the distance $\mathbf{r} - \mathbf{s}$, which, together with the "step size" ε, determines how much a weight vector $\mathbf{w}_r$ of neuron $\mathbf{r}$ in the vicinity of $\mathbf{s}$ is modified. The function h^0_{rs} has its maximum at $\mathbf{r} = \mathbf{s}$ and decays to zero as $\|\mathbf{r} - \mathbf{s}\|$ increases.

The algorithm is summarized as follows:

0. Assign suitable initial values to the weights $\mathbf{w}_r \in V$. If no a priori information is available assign random values.

1. Select a vector ("sensory input") $\mathbf{v} \in V$ according to some prespecified probability distribution $P(\mathbf{v})$.

2. Determine the location $\mathbf{s}$ for which $\|\mathbf{v} - \mathbf{w}_s\|$ is minimal, i.e.

$$\|\mathbf{v} - \mathbf{w}_s\| \leqq \|\mathbf{v} - \mathbf{w}_r\| \quad \text{for all } \mathbf{r} \in A. \tag{4}$$

3. Perform a learning step affecting all neurons $\mathbf{r}$ in the neighborhood of $\mathbf{s}$ (with $\mathbf{s}$ included)

$$\mathbf{w}_r^{\text{new}} = \mathbf{w}_r^{\text{old}} + \varepsilon h^0_{rs}(\mathbf{v} - \mathbf{w}_r^{\text{old}}) \tag{5}$$

and continue with 1.

Mathematically, these steps establish a Markov process the states of which are the tuples of the weight vectors $\mathbf{w}_r$ of the array A and the transitions of which, described by (5), are determined by the probability distribution $P(\mathbf{v})$ of inputs $\mathbf{v} \in V$. By (4) each $\mathbf{v} \in V$ gets "mapped" to a location $\mathbf{s}$ in the array and this (discrete) mapping of V onto A, being specified by the set of weight vectors $\mathbf{w}_r$ of the neurons, gradually evolves under this Markov process. Kohonen showed that by slowly diminishing both the step size ε and the width of h^0_{rs} the vectors $\mathbf{w}_r$ asymptotically settle to equilibrium values which for many tasks represent a useful two-dimensional mapping of the multi-dimensional space V onto A (Kohonen 1982a–c, 1984, 1986;

Kohonen et al. 1984). The resulting map has several remarkable properties: First, it represents most faithfully those dimensions of V along which the variance in the sequence of inputs v is most pronounced. These will often correspond to the most important features of the inputs. Second, it tries to preserve continuity, i.e. map similar inputs v to neighboring locations in A, thus preserving neighborhood relationships in the space V. Finally, it reflects differences in the sampling density $P(\mathbf{v})$ of the space V in a natural way: regions in V from which inputs have occurred frequently are mapped cnto larger domains of A and, therefore, with better resolution than regions in V from which only few inputs have emerged.

3 Derivation of the Fokker-Planck Equation

In this section we shall derive a Fokker-Planck equation for the time development of the map valid in the limit of small step size and in the vicinity of its stationary state.

Let us assume an array A of N formal neurons, labelled by their discrete positions $\mathbf{r}\in A$ and let $\mathbf{w}=(\mathbf{w}_{\mathbf{r}_1},\ldots,\mathbf{w}_{\mathbf{r}_N})$ denote a state of the array (i.e. $\mathbf{w}$ is a tuple of vectors $\mathbf{w}_\mathbf{r}$, one for each location $\mathbf{r}\in A$). For any given state $\mathbf{w}$ and neuron $\mathbf{r}\in A$ we call the set

$$F_\mathbf{r}(\mathbf{w})=\{\mathbf{v}\in \mathrm{V}\mid \|\mathbf{v}-\mathbf{w}_\mathbf{r}\|\leqq\|\mathbf{v}-\mathbf{w}_\mathbf{s}\| \quad \text{for all } \mathbf{s}\in A\} \tag{6}$$

the *feature set* of neuron $\mathbf{r}$. This renders a partition of V into mutually disjoint feature sets $F_\mathbf{r}$. The set $F_\mathbf{r}$ is precisely the subset of V, which is mapped to the location $\mathbf{r}\in A$ under the map specified by the state $\mathbf{w}$.

Under the algorithm each selection of an input vector $\mathbf{v}$ transforms the current state of a map, denoted $\mathbf{w}'$, into a new state $\mathbf{w}$ according to the rule

$$\mathbf{w}_\mathbf{r}=(1-\varepsilon h^0_{\mathbf{rs}})\mathbf{w}'_\mathbf{r}+\varepsilon h^0_{\mathbf{rs}}\mathbf{v}\,, \tag{7}$$

where $\mathbf{s}$ is the location of the feature set $F_\mathbf{s}(\mathbf{w}')$ containing the stimulus $\mathbf{v}$, i.e. $\mathbf{v}\in F_\mathbf{s}(\mathbf{w}')$. The parameter ε determines the size of one adaptation step. We will tacitly assume that it may depend on the step number or "iteration time" t. For independent random choices of the successive input vectors $\mathbf{v}$ (7) represents a Markov process. We want to investigate the time evolution of $\mathbf{w}$ in the limit of small ε.

To this end we could either take the differential equation obtained from (7) by taking the limit $\varepsilon\to 0$ and averaging over the random variable $\mathbf{v}$ and use techniques from the theory of stochastic approximation to compare $\mathbf{w}_\mathbf{r}$ with the trajectory of this differential equation (see e.g. Kushner and Clark 1978). Alternatively, we can study the time development of the distribution of an ensemble of processes (7). Here we will adopt the latter approach. The convergency conditions thus obtained [(46) and (47)] turn out to coincide with the conditions required to prove convergency by the first approach.

For a more compact notation we define a transformation $\mathbf{T}$ such that (7) is represented as

$$\mathbf{w}=\mathbf{T}(\mathbf{w}',\mathbf{v},\varepsilon)\,. \tag{8}$$

The transition probability $Q(\mathbf{w},\mathbf{w}')$ for a transition from state $\mathbf{w}'$ to state $\mathbf{w}$ can be written

$$Q(\mathbf{w},\mathbf{w}')=\sum_\mathbf{r}\int_{F_\mathbf{r}(\mathbf{w}')}d\mathbf{v}\,\delta(\mathbf{w}-\mathbf{T}(\mathbf{w}',\mathbf{v},\varepsilon))P(\mathbf{v})\,. \tag{9}$$

Instead of considering the states of individual systems, one rather describes an ensemble of arrays whose states $\mathbf{w}$ at iteration time t are distributed according to a distribution function $\tilde{S}(\mathbf{w},t)$. At each time step the transition probability transforms the distribution function $\tilde{S}(\mathbf{w},t)$ according to the usual Chapman-Kolmogoroff equation (see e.g. van Kampen 1981; Gardiner 1985)

$$\begin{aligned}\tilde{S}&(\mathbf{w},t+1)\\ &=\int d^N\mathbf{w}'Q(\mathbf{w},\mathbf{w}')\tilde{S}(\mathbf{w}',t)\\ &=\sum_\mathbf{r}\int d^N\mathbf{w}'\int_{F_\mathbf{r}(\mathbf{w}')}d\mathbf{v}P(\mathbf{v})\delta(\mathbf{w}-\mathbf{T}(\mathbf{w}',\mathbf{v},\varepsilon))\tilde{S}(\mathbf{w}',t)\,.\end{aligned} \tag{10}$$

To perform the $\mathbf{w}'$-integration, which runs over all N weight vector variables $\mathbf{w}'_\mathbf{r}$, $\mathbf{r}\in A$, we need the Jacobian

$$J(\varepsilon)=\left[\det\frac{\partial\mathbf{T}}{\partial\mathbf{w}}\right]^{-1}. \tag{11}$$

Assuming $\mathbf{v}\in F_\mathbf{s}(\mathbf{w}')$ for the moment, we obtain

$$J(\varepsilon)=\left[\prod_\mathbf{r}(1-\varepsilon h^0_{\mathbf{rs}})\right]^{-d}. \tag{12}$$

Here d is the dimension of the input vectors $\mathbf{v}$. Since $h^0_{\mathbf{rs}}$ is assumed to depend only on the difference $\mathbf{r}-\mathbf{s}$, J actually is independent of $\mathbf{s}$ and, therefore, depends only on ε.

Performing the $\mathbf{w}'$-integration one obtains

$$\begin{aligned}\tilde{S}(\mathbf{w},t+1)=J(\varepsilon)&\sum_\mathbf{r}\int d\mathbf{v}\chi_\mathbf{r}(\mathbf{T}^{-1}(\mathbf{w},\mathbf{v},\varepsilon),\mathbf{v})\\ &\times P(\mathbf{v})\tilde{S}(\mathbf{T}^{-1}(\mathbf{w},\mathbf{v},\varepsilon),t)\,.\end{aligned} \tag{13}$$

Here $\chi_\mathbf{r}(\mathbf{w},\mathbf{v})$ denotes the characteristic function for the feature set $F_\mathbf{r}(\mathbf{w})$, i.e.

$$\chi_\mathbf{r}(\mathbf{w},\mathbf{v})=\begin{cases}1\,, & \text{if } \mathbf{v}\in F_\mathbf{r}(\mathbf{w});\\ 0\,, & \text{else}\,.\end{cases} \tag{14}$$

$\mathbf{T}^{-1}$ stands for the inverse of the transformation $\mathbf{T}$. If $\mathbf{v}\in F_\mathbf{s}(\mathbf{w})$, $\mathbf{T}^{-1}(\mathbf{w},\mathbf{v},\varepsilon)$ is given by

$$[\mathbf{T}^{-1}(\mathbf{w},\mathbf{v},\varepsilon)]_\mathbf{r}=\mathbf{w}_\mathbf{r}+\varepsilon h_{\mathbf{rs}}(\mathbf{w}_\mathbf{r}-\mathbf{v})\,, \tag{15}$$

where we have introduced the new function $h_{\mathbf{rs}}:=h^0_{\mathbf{rs}}/(1-\varepsilon h^0_{\mathbf{rs}})$. The difference between $h_{\mathbf{rs}}$ and $h^0_{\mathbf{rs}}$ is only of order ε.

For $\varepsilon \ll 1$ and $\mathbf{v} \in F_s(\mathbf{w})$ we can expand $\tilde{S}(\mathbf{T}^{-1}(\mathbf{w}, \mathbf{v}, \varepsilon), t)$ as

$$\begin{aligned}&\tilde{S}(\mathbf{T}^{-1}(\mathbf{w}, \mathbf{v}, \varepsilon), t)\\ &= \tilde{S}(\mathbf{w}, t) + \varepsilon \sum_{rm} h_{rs}(\mathrm{w}_{rm} - \mathrm{v}_m) \frac{\partial \tilde{S}}{\partial \mathrm{w}_{rm}}\\ &+ \frac{1}{2} \varepsilon^2 \sum_{rm} \sum_{rn} h_{rs} h_{r's}(\mathrm{w}_{rm} - \mathrm{v}_m)(\mathrm{w}_{r'n} - \mathrm{v}_n) \frac{\partial^2 \tilde{S}}{\partial \mathrm{w}_{rm} \partial \mathrm{w}_{r'n}}\\ &+ O(\varepsilon^3).\end{aligned} \tag{16}$$

Likewise, $J(\varepsilon)$ can be expanded as

$$J(\varepsilon) = 1 + \varepsilon J_1 + \tfrac{1}{2} \varepsilon^2 J_2 + \ldots, \tag{17}$$

where

$$J_1 = d \cdot \sum_r h_{rs} = d \cdot \sum_r h_{r0} \tag{18}$$

is independent of s. Substituting (16), (17) into (13), keeping only derivatives up to second order and of these only the leading order in ε, we obtain

$$\begin{aligned}&\frac{\tilde{S}(\mathbf{w}, t+1) - \tilde{S}(\mathbf{w}, t)}{\varepsilon}\\ &= J_1 \tilde{S}(\mathbf{w}, t)\\ &+ \sum_s \int_{F_s(\mathbf{w})} \mathrm{d}\mathbf{v} P(\mathbf{v}) \sum_{rm} h_{rs}(\mathrm{w}_{rm} - \mathrm{v}_m) \frac{\partial \tilde{S}}{\partial \mathrm{w}_{rm}}\\ &+ \frac{\varepsilon}{2} \sum_s \int_{F_s(\mathbf{w})} \mathrm{d}\mathbf{v} P(\mathbf{v})\\ &\times \sum_{rm} \sum_{rn} h_{rs} h_{r's}(\mathrm{w}_{rm} - \mathrm{v}_m)(\mathrm{w}_{r'n} - \mathrm{v}_n) \frac{\partial^2 \tilde{S}}{\partial \mathrm{w}_{rm} \partial \mathrm{w}_{r'n}}.\end{aligned} \tag{19}$$

In the vicinity of the stationary state we expect $\tilde{S}(\mathbf{w}, t)$ to be peaked around the stationary expectation value $\bar{\mathbf{w}}$ of $\mathbf{w}$, which obeys

$$\int \mathrm{d}\mathbf{v} P(\mathbf{v}) \mathbf{T}(\bar{\mathbf{w}}, \mathbf{v}, \varepsilon) - \bar{\mathbf{w}} = 0. \tag{20}$$

Therefore, we make a shift of coordinates and define

$$S(\mathbf{u}, t) := \tilde{S}(\bar{\mathbf{w}} + \mathbf{u}, t), \tag{21}$$

i.e. $S(\mathbf{u}, t)$ is the distribution of the deviations $\mathbf{u}$ from the stationary expectation value $\bar{\mathbf{w}}$. For the following it is convenient to introduce the quantities

$$\hat{P}_r(\mathbf{w}) := \int_{F_r(\mathbf{w})} \mathrm{d}\mathbf{v} P(\mathbf{v}), \tag{22}$$

$$\bar{\mathbf{v}}_r := \frac{1}{\hat{P}_r(\mathbf{w})} \int_{F_r(\mathbf{w})} \mathrm{d}\mathbf{v} P(\mathbf{v}) \mathbf{v}, \tag{23}$$

$$\mathrm{V}_{rm}(\mathbf{w}) := \sum_s (\mathrm{w}_{rm} - \bar{\mathrm{v}}_{sm}) h_{rs} \hat{P}_s(\mathbf{w}), \tag{24}$$

$$\begin{aligned}\mathrm{D}_{rmr'n}(\mathbf{w}) := \sum_s h_{rs} h_{r's} \Big[&(\mathrm{w}_{rm} - \bar{\mathrm{v}}_{sm})(\mathrm{w}_{r'n} - \bar{\mathrm{v}}_{sn}) \hat{\mathrm{P}}_s(\mathbf{w})\\ &+ \int_{F_s(\mathbf{w})} (\mathrm{v}_m \mathrm{v}_n - \bar{\mathrm{v}}_{sm} \bar{\mathrm{v}}_{sn}) P(\mathbf{v}) \mathrm{d}\mathbf{v}\Big].\end{aligned} \tag{25}$$

$\hat{P}_r(\mathbf{w})$ is the probability for an input to belong to feature set F_r, $\bar{\mathbf{v}}_r$ is the centroid of the input distribution, restricted to feature set F_r.

In the limit of small ε we may evaluate the $O(\varepsilon)$-term in (19) directly at $\mathbf{w} = \bar{\mathbf{w}}$ and replace $S(\mathbf{u}, t+1) - S(\mathbf{u}, t)$ *by* $\partial_t S(\mathbf{u}, t)$. This results in the Fokker-Planck equation

$$\begin{aligned}\frac{1}{\varepsilon} \partial_t S(\mathbf{u}, t) = {} & J_1 S(\mathbf{u}, t)\\ &+ \sum_{rm} \mathrm{V}_{rm}(\bar{\mathbf{w}} + \mathbf{u}) \frac{\partial S(\mathbf{u}, t)}{\partial u_{rm}}\\ &+ \frac{\varepsilon}{2} \sum_{rmr'n} \mathrm{D}_{rmr'n}(\bar{\mathbf{w}}) \frac{\partial^2 S(\mathbf{u}, t)}{\partial \mathrm{u}_{rm} \partial \mathrm{u}_{r'n}}.\end{aligned} \tag{26}$$

The first order term represents the restoring force. It vanishes at $\mathbf{u} = 0$ and, therefore, must be retained to linear order in $\mathbf{u}$. This yields

$$\begin{aligned}&\sum_{rm} \mathrm{V}_{rm}(\bar{\mathbf{w}} + \mathbf{u}) \frac{\partial S(\mathbf{u}, t)}{\partial \mathrm{u}_{rm}}\\ &\quad = -\sum_{rm} \frac{\partial \mathrm{V}_{rm}}{\partial \mathrm{w}_{rm}} S\\ &\quad + \sum_{rmr'n} \frac{\partial}{\partial \mathrm{u}_{rm}} \left(\frac{\partial \mathrm{V}_{rm}}{\partial \mathrm{w}_{r'n}} (\bar{\mathbf{w}}) \mathrm{u}_{r'n} S \right).\end{aligned} \tag{27}$$

To obtain a convenient expression for $\sum_{rm} \partial \mathrm{V}_{rm} / \partial \mathrm{w}_{rm}$, we start from

$$\begin{aligned}\mathbf{V}_r(\mathbf{w}) &= \sum_s h_{rs} \int_{F_s(\mathbf{w})} \mathrm{d}\mathbf{v} P(\mathbf{v})(\mathbf{w}_r - \mathbf{v})\\ &= \frac{1}{\varepsilon} \int \mathrm{d}\mathbf{v} P(\mathbf{v})(\mathbf{w}_r - \mathbf{T}(\mathbf{w}, \mathbf{v}, \varepsilon)_r)\end{aligned} \tag{28}$$

and obtain

$$\sum_{rm} \frac{\partial \mathrm{V}_{rm}}{\partial \mathrm{w}_{rm}} = \frac{1}{\varepsilon} \int \mathrm{d}\mathbf{v} P(\mathbf{v}) \operatorname{Tr} \left(1 - \frac{\partial \mathbf{T}}{\partial \mathbf{w}} \right). \tag{29}$$

The deviation of the Jacobian $\partial \mathbf{T} / \partial \mathbf{w}$ from the identity matrix is of order ε. Writing $\dfrac{\partial \mathbf{T}}{\partial \mathbf{w}} = 1 + \varepsilon \mathrm{A}$, we conclude from (11)

$$J(\varepsilon) = \det(1 - \varepsilon \mathrm{A}) + O(\varepsilon^2) = 1 - \varepsilon \cdot \operatorname{Tr} \mathrm{A} + O(\varepsilon^2). \tag{30}$$

Comparison with (17) yields

$$J_1 = -\operatorname{Tr} \mathrm{A} = \frac{1}{\varepsilon} \operatorname{Tr} \left(1 - \frac{\partial \mathbf{T}}{\partial \mathbf{w}} \right). \tag{31}$$

Inserting this into (29) results in the relation

$$\sum_{rm} \frac{\partial \mathrm{V}_{rm}}{\partial \mathrm{w}_{rm}} = J_1. \tag{32}$$

This brings us to the final form of our Eq. for $S(\mathbf{u}, t)$

$$\frac{1}{\varepsilon}\,\partial_t S(\mathbf{u}, t) = \sum_{\mathbf{r}m\mathbf{r}'n} \frac{\partial}{\partial \mathrm{u}_{\mathbf{r}m}}\, \mathrm{B}_{\mathbf{r}m\mathbf{r}'n} \mathrm{u}_{\mathbf{r}'n} S(\mathbf{u}, t) + \frac{\varepsilon}{2} \sum_{\mathbf{r}m\mathbf{r}'n} \mathrm{D}_{\mathbf{r}m\mathbf{r}'n} \frac{\partial^2 S(\mathbf{u}, t)}{\partial \mathrm{u}_{\mathbf{r}m} \partial \mathrm{u}_{\mathbf{r}'n}}, \tag{33}$$

with the constant matrix $\mathbf{B}$ given by

$$\mathrm{B}_{\mathbf{r}m\mathbf{r}'n} := \left(\frac{\partial \mathrm{V}_{\mathbf{r}m}(\mathbf{w})}{\partial \mathrm{w}_{\mathbf{r}'n}}\right)_{\mathbf{w}=\bar{\mathbf{w}}}. \tag{34}$$

Equation (33) is the desired Fokker-Planck equation for the Markov process (7).

For the expectation value $\bar{\mathrm{u}}_{\mathbf{r}m}(t) = \langle \mathrm{u}_{\mathbf{r}m} \rangle_S$ and for the correlation matrix

$$\mathrm{C}_{\mathbf{r}m\mathbf{s}n}(t) = \langle (\mathrm{u}_{\mathbf{r}m} - \bar{\mathrm{u}}_{\mathbf{r}m})(\mathrm{u}_{\mathbf{s}n} - \bar{\mathrm{u}}_{\mathbf{s}n}) \rangle_S$$

of the distribution S obeying (33) explicit expressions can be derived (for details see e.g. van Kampen 1981; Gardiner 1985). Defining the matrix

$$\mathbf{Y}(t) = \exp\left(-\mathbf{B} \int_0^t \varepsilon(\tau)\,\mathrm{d}\tau\right), \tag{35}$$

$\bar{\mathbf{u}}(t)$ is given by

$$\bar{\mathbf{u}}(t) = \mathbf{Y}(t)\bar{\mathbf{u}}(0), \tag{36}$$

where $\bar{\mathbf{u}}(0)$ is the expectation value at $t=0$, and $\mathbf{C}(t)$ by

$$\mathbf{C}(t) = \mathbf{Y}(t)\left[\mathbf{C}(0) + \int_0^t \varepsilon(\tau)^2 \mathbf{Y}(\tau)^{-1} \mathbf{D}(\mathbf{Y}(\tau)^{-1})^T\,\mathrm{d}\tau\right] \mathbf{Y}(t)^T. \tag{37}$$

If the initial condition is a δ-distribution, i.e. $S(\mathbf{u}, 0) = \prod_{\mathbf{r}m} \delta(\mathrm{u}_{\mathbf{r}m} - \mathrm{u}(0)_{\mathbf{r}m})$, $S(\mathbf{u}, t)$ is a Gaussian

$$S(\mathbf{u}, t) = \det(2\pi\mathbf{C})^{-1/2} \exp(-\tfrac{1}{2}(\mathbf{u} - \bar{\mathbf{u}})^T \mathbf{C}^{-1}(\mathbf{u} - \bar{\mathbf{u}})). \tag{38}$$

If $\varepsilon(t)$ is chosen such that in the limit $t \to \infty$ the initial conditions become irrelevant, e.g. for constant ε, then replacing $\mathbf{C}$ and $\bar{\mathbf{u}}$ in (38) by their asymptotic limits yields the stationary solution. A further simplification arises, if $\mathbf{B}$ and $\mathbf{D}$ commute. In this case the integral in (37) can be explicitly evaluated for $\varepsilon = \text{constant}$ and the stationary distribution is the Gaussian (38) with

$$\mathbf{C} = \varepsilon(\mathbf{B} + \mathbf{B}^T)^{-1}\mathbf{D}. \tag{39}$$

4 Convergence to Equilibrium State

In this section we consider the question under which conditions the Markov process (7) converges with probability 1 to a stationary expectation value $\bar{\mathbf{w}}$. To achieve such convergence both the variance of the state distribution and the expectation value $\bar{\mathbf{u}}(t)$ of the deviation from $\bar{\mathbf{w}}$ must vanish in the limit $t \to \infty$.

As a consequence of (37), the correlation matrix $\mathbf{C}(t)$ obeys (see e.g. van Kampen 1981)

$$\dot{\mathbf{C}} = -\varepsilon(t)(\mathbf{BC} + \mathbf{CB}^T) + \varepsilon(t)^2\mathbf{D}. \tag{40}$$

The time derivative of the Euclidean matrix norm $\|\mathbf{C}\|^2 := \sum_{\mathbf{r}m\mathbf{r}'n} \mathrm{C}^2_{\mathbf{r}m\mathbf{r}'n}$ is then

$$\tfrac{1}{2}\,\partial_t \|\mathbf{C}\|^2 = -\varepsilon(t)\,\mathrm{Tr}\,\mathbf{C}(\mathbf{B} + \mathbf{B}^T)\mathbf{C} + \varepsilon(t)^2\,\mathrm{Tr}\,\mathbf{DC}. \tag{41}$$

We will require that $\mathbf{C}$ remains bounded if $\varepsilon(t)$ is constant and the initial correlation matrix $\mathbf{C}(0)$ is sufficiently small but else arbitrary. This is a stability condition for the equilibrium state $\bar{\mathbf{w}}$. Since both $\mathbf{C}$ and $\mathbf{D}$ are symmetric and non-negative, $\mathrm{Tr}\,\mathbf{DC} \geqq 0$ holds. Therefore, the stability condition requires that $(\mathbf{B} + \mathbf{B}^T)$ be positive, i.e. there exists a constant $\beta > 0$ such that

$$\mathrm{Tr}\,\mathbf{C}[\mathbf{B}(\bar{\mathbf{w}}) + \mathbf{B}(\bar{\mathbf{w}})^T]\mathbf{C} > \beta\|\mathbf{C}\|^2/2. \tag{42}$$

Consequently, there is another constant $\gamma > 0$ such that

$$\partial_t \|\mathbf{C}\|^2 \leqq -\varepsilon(t)\beta\|\mathbf{C}\|^2 + \varepsilon(t)^2\gamma. \tag{43}$$

Integrating (43), we obtain

$$\|\mathbf{C}(t)\|^2 \leqq \gamma \int_0^t \varepsilon(t')^2 \exp\left(-\beta\int_{t'}^t \varepsilon(t'')\,\mathrm{d}t''\right)\mathrm{d}t'. \tag{44}$$

Any positive function $\varepsilon(t)$, for which the right hand side of (44) vanishes asymptotically guarantees the desired convergence of $\mathbf{C}$ towards zero. In the Appendix we will show that this condition is equivalent to requiring $\lim_{t\to\infty} \varepsilon(t) = 0$ together with

$$\lim_{t\to\infty} \int_0^t \varepsilon(t')\,\mathrm{d}t' = \infty. \tag{45}$$

Besides $\lim_{t\to\infty} \mathbf{C}(t) = 0$ this also ensures $\lim_{t\to\infty} \bar{\mathbf{u}}(t) = 0$ [cf. (35), (36)], and, therefore, guarantees the convergence to the equilibrium expectation $\bar{\mathbf{w}}$ with probability 1. The criterion cannot be weakened: according to (40) $\lim_{t\to\infty} \varepsilon(t) = 0$ is necessary for the variance to decay to zero and due to (35), (34), (45) is necessary to ensure Eq. $\lim_{t\to\infty} \bar{\mathbf{u}}(t) = 0$. Hence, for convergence to an equilibrium state $\bar{\mathbf{w}}$ obeying the stability condition, we have shown:

Let $\varepsilon(t) > 0$ be any positive function sufficiently small for the Fokker-Planck-equation (33) to describe the original Markov process (7) accurately. Then the two conditions

$$\lim_{t\to\infty} \int_0^t \varepsilon(t')\,\mathrm{d}t' = \infty, \tag{46}$$

$$\lim_{t\to\infty} \varepsilon(t) = 0. \tag{47}$$

are necessary and sufficient to guarantee with probability 1 convergence to $\bar{\mathbf{w}}$ for all initial states sufficiently close to the equilibrium expectation $\bar{\mathbf{w}}$.

Condition (46) is identical to the first convergence condition obtained by Cottrell and Fort (1986) for a closely related process. However, their second condition, namely that $\int_0^\infty \varepsilon(t)^2\,dt<\infty$, is not required in our case and is replaced by the milder requirement (47). In particular these conditions are fullfilled for all decay laws $\varepsilon(t)\propto t^{-\alpha}$ with $0<\alpha\leqq 1$. For laws with $\alpha>1$ or exponential decay laws (46) is not fullfilled, i.e. some residual error remains even in the $t\rightarrow\infty$-limit.

5 Analysis for Spatially Uniform Input Vector Density

5.1 General Considerations

In the following we shall discuss the case of a spatially uniform probability density $P(\mathbf{v})$ of the input vectors $\mathbf{v}$ more closely. We shall assume that any change in the gain factor ε occurs sufficiently slowly to replace $S(\mathbf{u}, t)$ by the stationary solution for the corresponding constant value of ε. In general, the input vectors $\mathbf{v}$ will be drawn from a volume whose dimension d is much larger than two. The above Markov process will try to detect the two most significant dimensions of the volume and map these across the array, thus providing a two-dimensional map which, despite of being necessarily a many-to-one projection, is as faithful as possible. This is illustrated in Fig. 1a for the case of V being a three-dimensional parallelepiped of size $40\times 40\times 10$ units, for which a map on an array of 40×40 neurons is sought. Figure 1a shows a "snapshot" of the resulting map from a Monte Carlo simulation of the algorithm with a finite learning step size of $\varepsilon=0.05$ and random input vectors $\mathbf{v}$ drawn uniformly from the parallelepiped. For each neuron the location $\mathbf{w}_\mathbf{r}\in \mathrm{V}$, (i.e. the center of its feature set in V) is drawn, and locations corresponding to neighboring neurons in the array are connected by lines. This visualizes the map as an "embedding" of the array in the parallelepiped V. Since the parallelepiped is fairly "flat", the resulting map is essentially symmetric and represents a projection onto the subspace spanned by the two largest dimensions of the parallelepiped.

For non-zero step size ε, the map usually will exhibit statistical fluctuations around its equilibrium configuration. These fluctuations can be seen in Fig. 1a both as shallow "bumps" in the embedded surface and as weak tangential distortions of the mesh and will be calculated below.

However, if inputs scatter too much along some or all of the additional $d-2$ dimensions, restriction only to the two main dimensions would yield a poor representation for many inputs $\mathbf{v}$. In this situation, the map described above looses its stability and shifts to a more complicated stable equilibrium map. This new map usually is of lower symmetry and corresponds to an embedding of the array in V which is strongly folded in direction of the additionally needed dimensions. This behaviour, termed "automatic selection of feature dimensions" and mentioned already above, is illustrated in Fig. 1b, where the vertical extension of the parallelepiped has been increased from 10 to 14 units. As a result, the symmetric map is no longer stable and the associated embedding switches to a new configuration with large folds protruding into the vertical direction, indicating that the new stable map has broken the symmetry of the underlying uniform input distribution in favor of a better representation of the vertical dimension of the parallelepiped. Below we will show that this change to a new equilibrium configuration occurs at a critical value $2s^*$ of the height of the parallelepiped and that the maps, approaching this value from below, exhibit increasing fluctuations of a typical wavelength λ^*. Both s^* and λ^* will be calculated in the sequel.

The use of a three-dimensional volume is not a serious restriction of generality, as each of the additional $d-2$ dimensions contributes in the same

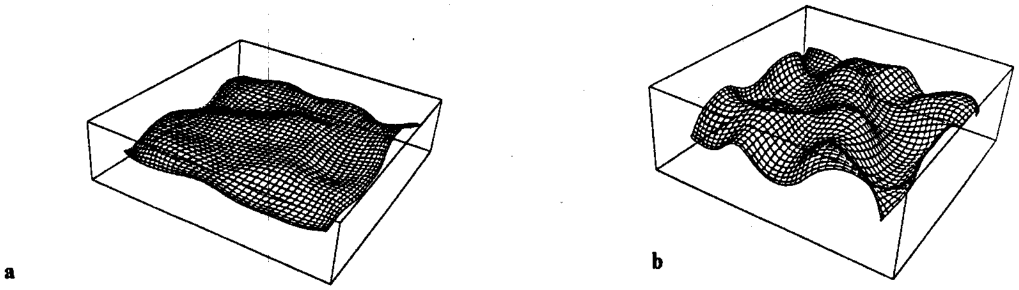

Fig. 1. a Snapshot of Monte Carlo simulation of a map between an array of 40 × 40 neurons and a three-dimensional parallelepiped of size 40 × 40 × 10 units. Since the height (10 units) of the parallelepiped is sufficiently small, the map is essentially a projection onto the two main dimensions of the parallelepiped. The shallow "bumbs" indicate weak distortions due to equilibrium fluctuations. **b** Same simulation as in Fig. 1a, but for a height of 14 units. The map of Fig. 1a is no longer stable and has changed to a less symmetric map given by an embedding with large folds protruding into the vertical direction, thus, furnishing a better representation of this dimension

manner and independently of the others to the instability and the equilibrium fluctuations. Taking an array A of $N \times N$ neurons[1] and a three-dimensional parallelepiped V given by $0 \leqq x, y \leqq N, -s \leqq z \leqq s$, the uniform probability distribution is $P(\mathbf{v}) = [2sN^2]^{-1}$. To avoid edge effects we impose periodic boundary conditions along the x- and y-directions. Then, by symmetry, $\bar{\mathbf{w}}_{\mathbf{r}} = \mathbf{r}$, $\mathbf{r} = m\mathbf{e}_x + n\mathbf{e}_y$ must define an equilibrium state $\bar{\mathbf{w}}$. To show this we calculate the L.H.S. of (20)

$$\begin{aligned}&\int \mathrm{d}\mathbf{v}\, P(\mathbf{v})\, [T(\bar{\mathbf{w}}, \mathbf{v}, \varepsilon)]_{\mathbf{r}} - \bar{\mathbf{w}}_{\mathbf{r}} \\ &\quad = \varepsilon \sum_{\mathbf{s}} h_{\mathbf{rs}} \int_{F_{\mathbf{s}}(\bar{\mathbf{w}})} \mathrm{d}\mathbf{v}\, P(\mathbf{v})\, (\mathbf{v} - \bar{\mathbf{w}}_{\mathbf{r}}) \\ &\quad = \varepsilon \sum_{\mathbf{s}} h_{\mathbf{rs}} N^{-2} (\bar{\mathbf{w}}_s - \bar{\mathbf{w}}_r) = 0 . \end{aligned} \tag{48}$$

Here we have made use of $\int_{F_{\mathbf{s}}(\bar{\mathbf{w}})} \mathbf{v}\, \mathrm{d}\mathbf{v} = \bar{\mathbf{w}}_{\mathbf{s}}$ for the special configuration $\bar{\mathbf{w}}$ and the symmetry of $h_{\mathbf{rs}}$. As $P(\mathbf{v})$ is constant and we have periodic boundary conditions, neither any origin nor any direction is preferred. Hence, any configuration obtained by translating or rotating $\bar{\mathbf{w}}$ is an equilibrium configuration as well. Our special choice thus amounts to a convenient selection of origin and orientation of our coordinate system.

For non-uniform input distribution $P(\mathbf{v})$ an equilibrium configuration can be calculated only in special cases (Ritter and Schulten 1986a). Generally, for non-uniform $P(\mathbf{v})$ we expect the analysis of this chapter to remain valid "locally", i.e. up to distances over which $P(\mathbf{v})$ does not vary significantly. This means, that fluctuations with sufficiently short wave length will still behave as calculated below. The behaviour at long wave lengths, however, may be very different. In particular, the divergence of the equilibrium deviations at long wave lengths calculated below is due to the translational invariance of the system and will be absent in the general case.

If a distribution S with finite variance and expectation $\bar{\mathbf{w}}$ exists, $\bar{\mathbf{w}}$ is stable and we can calculate any fluctuations about $\bar{\mathbf{w}}$ from S, or, more directly, from (39).

Let $S(\mathbf{u}) = \lim_{t \to \infty} S(\mathbf{u}, t)$ denote the stationary distribution function of the equilibrium deviations $\mathbf{u} = \mathbf{w} - \bar{\mathbf{w}}$ for a constant step size ε. Due to the translational invariance both $\mathrm{D}_{\mathbf{r}m,\,'}$ and $\mathrm{B}_{\mathbf{r}m\mathbf{r}'n}$ depend only on the difference $\mathbf{r} - \mathbf{r}'$ and on m, n. Therefore, we can decouple (33) if we represent $S(\mathbf{u})$ in terms of the Fourier mode amplitudes

$$\hat{\mathbf{u}}_{\mathbf{k}} = \frac{1}{N} \sum_{\mathbf{r}} \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}} \mathbf{u}_{\mathbf{r}} \tag{49}$$

of $\mathbf{u}$. Each mode amplitude turns out to be distributed independently, i.e. by separation of variables

$$S(\mathbf{u}) = \prod_{\mathbf{k}} \hat{S}_{\mathbf{k}}(\hat{\mathbf{u}}_{\mathbf{k}}), \tag{50}$$

we obtain a set of mutually independent stationary Fokker-Planck equations for the individual mode distributions $\hat{S}_{\mathbf{k}}$

$$\begin{aligned}&\sum_{mn} \hat{\mathrm{B}}(\mathbf{k})_{mn} \frac{\partial}{\partial \mathbf{u}_m} \mathbf{u}_m \hat{S}_{\mathbf{k}}(\mathbf{u}) \\ &\quad + \frac{\varepsilon}{2} \sum_{mn} \hat{\mathrm{D}}(\mathbf{k})_{mn} \frac{\partial^2}{\partial \mathbf{u}_m \partial \mathbf{u}_n} \hat{S}_{\mathbf{k}}(\mathbf{u}) = 0 . \end{aligned} \tag{51}$$

Here $\hat{\mathbf{D}}(\mathbf{k})$ and $\hat{\mathbf{B}}(\mathbf{k})$ are $d \times d$-matrices given by

$$\begin{aligned}\hat{\mathbf{D}}(\mathbf{k}) &= \sum_{\mathbf{r}} \mathrm{e}^{i\mathbf{k}(\mathbf{r}-\mathbf{r}')} \mathbf{D}_{\mathbf{rr}'} \\ &= \frac{1}{N^2} [(\nabla_k \hat{h}(k))(\nabla_k \hat{h}(k))^T + \mathbf{M}\hat{h}(\mathbf{k})^2] , \end{aligned} \tag{52}$$

and, similarly,

$$\hat{\mathbf{B}}(\mathbf{k}) = \frac{\hat{h}(0)}{N^2} \left[\mathbf{1} - \frac{\hat{h}(\mathbf{k})}{\hat{h}(0)} \hat{\mathbf{a}}(\mathbf{k}) \right] - \frac{1}{N^2} (i\nabla_k \hat{h}(\mathbf{k})) \hat{\mathbf{b}}(\mathbf{k})^T , \tag{53}$$

where we have defined $\mathbf{k} := (k_x, k_y, 0)^T$ to facilitate the notation. $\mathbf{M}$ is given by

$$\begin{aligned}\mathbf{M} &= \frac{1}{2s} \int_{F_{\mathbf{r}}(\bar{\mathbf{w}})} \mathrm{d}\mathbf{v} (\mathbf{v}\mathbf{v}^T - \bar{\mathbf{v}}_{\mathbf{r}} \bar{\mathbf{v}}_{\mathbf{r}}^T) \\ &= \begin{pmatrix} 1/12 & 0 & 0 \\ 0 & 1/12 & 0 \\ 0 & 0 & s^2/3 \end{pmatrix} , \end{aligned} \tag{54}$$

i.e. $\mathbf{M}$ is the correlation matrix of the input vectors $\mathbf{v}$ over a feature set $F_{\mathbf{r}}$ [since all $F_{\mathbf{r}}$ are identical and $P(\mathbf{v})$ is constant, $\mathbf{M}$ does not depend on the choice of $\mathbf{r}$]. The function $\hat{h}(\mathbf{k})$ is the discrete Fourier transform of the neighborhood function $h_{\mathbf{rs}}$, i.e.

$$\hat{h}(\mathbf{k}) = \sum_{\mathbf{r}} \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}} h_{\mathbf{rs}} , \tag{55}$$

and matrix $\hat{\mathbf{a}}(\mathbf{k})$ and vector $\hat{\mathbf{b}}(\mathbf{k})$ are the Fourier transforms of the functions

$$\mathbf{a}_{\mathbf{rr}'} := \left. \frac{\partial \bar{\mathbf{v}}_{\mathbf{r}}(\mathbf{w})}{\partial \mathbf{w}_{\mathbf{r}'}} \right|_{\bar{\mathbf{w}}} , \tag{56}$$

$$\mathbf{b}_{\mathbf{rr}'} := \frac{1}{\hat{P}_{\mathbf{r}}} \left. \frac{\partial \hat{P}_{\mathbf{r}}(\mathbf{w})}{\partial \mathbf{w}_{\mathbf{r}'}} \right|_{\bar{\mathbf{w}}} . \tag{57}$$

Hence, both $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ depend only on the geometry of the equilibrium configuration and not on the neighborhood function h. Matrix $\mathbf{a}$ measures the shift of the centroid of a feature set under small deformations of the equilibrium state and $\mathbf{b}$ essentially measures the

[1] The reader should note that the number of neurons is now N^2 rather than N

corresponding volume change of a feature set. In our case, the feature set $F_\mathbf{r}(\mathbf{w})$ is given by the volume bounded by the four planes orthogonally bisecting the four distances between $\mathbf{w}_\mathbf{r}$ and its four nearest lattice neighbors, together with the two planes $z=\pm s$. Elementary geometry yields

$$\mathbf{a}_{\mathbf{r}\mathbf{r}'}=\delta_{\mathbf{r}\mathbf{r}'}\begin{pmatrix}2/3 & 0 & 0\\ 0 & 2/3 & 0\\ 0 & 0 & 4s^2/3\end{pmatrix}-(\delta_{\mathbf{r}+\mathbf{e}_x,\mathbf{r}'}+\delta_{\mathbf{r}-\mathbf{e}_x,\mathbf{r}'})\begin{pmatrix}-1/4 & 0 & 0\\ 0 & 1/12 & 0\\ 0 & 0 & s^2/3\end{pmatrix}-(\delta_{\mathbf{r}+\mathbf{e}_y,\mathbf{r}'}+\delta_{\mathbf{r}-\mathbf{e}_y,\mathbf{r}'})\begin{pmatrix}1/12 & 0 & 0\\ 0 & -1/4 & 0\\ 0 & 0 & s^2/3\end{pmatrix} \tag{58}$$

and

$$\mathbf{b}_{\mathbf{r}\mathbf{r}'}=\frac{1}{2}\sum_{\mathbf{n}=\pm\mathbf{e}_x,\mathbf{e}_y}\mathbf{n}(\delta_{\mathbf{r}+\mathbf{n},\mathbf{r}'}-\delta_{\mathbf{r}\mathbf{r}'}). \tag{59}$$

The corresponding Fourier transforms are

$$\begin{aligned}\hat{\mathbf{a}}(\mathbf{k})=&\frac{1}{6}(4+3\cos k_x-\cos k_y)\mathbf{e}_x\mathbf{e}_x^T\\&+\frac{1}{6}(4-\cos k_x+3\cos k_y)\mathbf{e}_y\mathbf{e}_y^T\\&+\frac{2s^2}{3}(2-\cos k_x-\cos k_y)\mathbf{e}_z\mathbf{e}_z^T,\end{aligned} \tag{60}$$

$$\hat{\mathbf{b}}(\mathbf{k})=-i\cdot(\mathbf{e}_x\sin k_x+\mathbf{e}_y\sin k_y). \tag{61}$$

We are now ready to discuss the behavior of the system in the vicinity of the equilibrium state $\bar{\mathbf{w}}$. From $\lim\limits_{k\to\infty}\hat{\mathbf{b}}(\mathbf{k})=0$ and $\lim\limits_{k\to\infty}\hat{\mathbf{a}}_{mn}(\mathbf{k})=\delta_{mn}(1-\delta_{m,3})$, we see that equilibrium deviations along the 1- and 2-direction sense a vanishing restoring force in the long wavelength limit, corresponding to the two zero eigenvalues developed by $\hat{\mathbf{B}}(\mathbf{k})$ in this limit. Correspondingly, the equilibrium fluctuations associated with these modes can become very large with increasing wavelength. This is due to the translational invariance of the system along the 1- and 2-direction. In contrast, the remaining u_3-modes experience a finite restoring force and, therefore, finite fluctuations even at $\mathbf{k}=0$.

However, these modes are subject to a different source of instability. As $\hat{\mathbf{a}}_{33}(\mathbf{k})$ is proportional to s^2, $\hat{\mathbf{B}}(\mathbf{k})$ may develop a negative eigenvalue for the 3-direction, if s grows too large. As a result, a subset or even all of these modes may become unstable if s starts to exceed a critical value. Hence, when the input vector distribution along the transversal dimensions becomes too broad the symmetry underlying the distribution $P(\mathbf{v})$ is broken and the array A folds into a new nonsymmetric equilibrium configuration. This symmetry breaking is preceeded by a strong increase of fluctuations of modes with a characteristic wavelength λ^*.

For a more detailed analysis and to calculate λ^*, we shall consider the limiting cases of long-ranged and short-ranged adjustment function $h_{\mathbf{r}\mathbf{s}}$.

5.2 Long-Ranged $h_{\mathbf{r}\mathbf{s}}$

We consider the adjustment function Gaussian shaped, i.e.

$$h_{\mathbf{r}\mathbf{r}'}=\sum_{\mathbf{s}}\delta_{\mathbf{r}+\mathbf{s},\mathbf{r}'}\exp\left(-\frac{s^2}{2\sigma^2}\right), \tag{62}$$

with lateral width σ, for which we will require $1\ll\sigma\ll N$. In this case it is a good approximation to replace the finite discrete Fourier transform by an infinite continuous one, yielding

$$\hat{h}(\mathbf{k})=2\pi\sigma^2\exp(-\sigma^2k^2/2). \tag{63}$$

Substituting (63) into (52) we obtain

$$\hat{\mathbf{D}}(\mathbf{k})=\frac{4\pi^2\sigma^4}{N^2}[\mathbf{k}\mathbf{k}^T\sigma^4+\mathbf{M}]\exp(-k^2\sigma^2). \tag{64}$$

The non-vanishing elements of $\hat{\mathbf{B}}(\mathbf{k})$ are

$$\hat{\mathbf{B}}_{11}=\frac{2\pi\sigma^2}{N^2}\left(1-\frac{1}{6}(4+3\cos k_x-6k_x\sigma^2\sin k_x-\cos k_y)\exp(-k^2\sigma^2/2)\right), \tag{65}$$

$$\hat{\mathbf{B}}_{22}=\frac{2\pi\sigma^2}{N^2}\left(1-\frac{1}{6}(4-\cos k_x-6k_y\sigma^2\sin k_y+3\cos k_y)\exp(-k^2\sigma^2/2)\right), \tag{66}$$

$$\hat{\mathbf{B}}_{33}=\frac{2\pi\sigma^2}{N^2}\left(1-\frac{2s^2}{3}(2-\cos k_x-2\cos k_y)\times\exp(-k^2\sigma^2/2)\right), \tag{67}$$

$$\hat{\mathbf{B}}_{12}=\frac{2\pi\sigma^4}{N^2}\cdot k_x\sin k_y\exp(-k^2\sigma^2/2), \tag{68}$$

$$\hat{\mathbf{B}}_{21}=\frac{2\pi\sigma^4}{N^2}\cdot k_y\sin k_x\exp(-k^2\sigma^2/2). \tag{69}$$

To simplify these expressions, we observe that for $\sigma \gg 1$ either $e^{-\sigma^2 k^2}$ is very small or k_x and k_y are sufficiently small to expand the sines and cosines to leading order. Further neglecting k^2-terms relative to $k^2\sigma^2$-terms we obtain for $\hat{\mathbf{B}}$ the simpler expression

$$\hat{\mathbf{B}}(\mathbf{k}) \approx \frac{2\pi\sigma^2}{N^2}\left[1-\left(1-\sigma^2\mathbf{k}\mathbf{k}^T+\frac{s^2k^2}{3}\,\mathbf{e}_z\mathbf{e}_z^T\right)\right.$$
$$\left.\times\exp(-k^2\sigma^2/2)\right]. \tag{70}$$

In this approximation $\hat{\mathbf{B}}(\mathbf{k})$ and $\hat{\mathbf{D}}(\mathbf{k})$ commute and both have the same eigenvectors, namely $\boldsymbol{\xi}_3=\mathbf{e}_z$, $\boldsymbol{\xi}_2=\mathbf{k}$ and the vector $\boldsymbol{\xi}_1=\mathbf{k}^\perp$ perpendicular to these two. The corresponding eigenvalues λ_n^B and λ_n^D for $\hat{\mathbf{B}}(\mathbf{k})$ and $\hat{\mathbf{D}}(\mathbf{k})$ are

$$\begin{aligned}\lambda_1^B(\mathbf{k}) &= \frac{2\pi\sigma^2}{N^2}\,(1-e^{-k^2\sigma^2/2});\\ \lambda_1^D(\mathbf{k}) &= \frac{\pi^2\sigma^4}{3N^2}\,e^{-k^2\sigma^2};\end{aligned} \tag{71}$$

$$\begin{aligned}\lambda_2^B(\mathbf{k}) &= \frac{2\pi\sigma^2}{N^2}\,(1-(1-k^2\sigma^2)\,e^{-k^2\sigma^2/2});\\ \lambda_2^D(\mathbf{k}) &= \frac{\pi^2\sigma^4}{3N^2}\,(12k^2\sigma^4+1)\,e^{-k^2\sigma^2};\end{aligned} \tag{72}$$

$$\begin{aligned}\lambda_3^B(\mathbf{k}) &= \frac{2\pi\sigma^2}{N^2}\left(1-\frac{s^2k^2}{3}\,e^{-k^2\sigma^2/2}\right);\\ \lambda_3^D(\mathbf{k}) &= \frac{4\pi^2\sigma^4}{3N^2}\,s^2\,e^{-k^2\sigma^2}.\end{aligned} \tag{73}$$

$\hat{\mathbf{B}}$ represents the strength of the drift term responsible for driving the expectation value of the distribution towards its equilibrium value. Hence (71), (72) imply that the system is "stiffer" for deviations along the $\boldsymbol{\xi}_2$-mode, for which deviations and wave vector $\mathbf{k}$ are parallel, than it is for deviations along the $\boldsymbol{\xi}_1$-mode, for which the deviations are perpendicular to $\mathbf{k}$. For wavelengths which are large compared to the lateral extension σ of the adjustment function we asymptotically have $\lambda_2^B(\mathbf{k})=3\lambda_1^B(\mathbf{k})=O(k^2)$, i.e. the $\boldsymbol{\xi}_2$-mode is three times stiffer than the $\boldsymbol{\xi}_1$-mode and both stiffnesses vanish as $\mathbf{k}$ tends to zero. The stiffness of the $\boldsymbol{\xi}_3$-mode does not vanish at $\mathbf{k}=0$. This mode is subject to a different kind of instability: if s is large enough, $\lambda_3^B(\mathbf{k})$ can be negative for a whole band of $\mathbf{k}$-values. In this case the associated modes, and with them the symmetric equilibrium configuration chosen for our expansion, are unstable and the system will develop towards a different equilibrium state. To see this more clearly, we consider the fluctuations of the associated eigenmode amplitudes u_n. From (39) follows

$$\langle u_n(\mathbf{k})^2\rangle = \frac{\varepsilon\lambda_n^D(\mathbf{k})}{2\lambda_n^B(\mathbf{k})}, \qquad n=1,2,3. \tag{74}$$

All other correlations vanish. Hence, we have

$$\langle u_1(\mathbf{k})^2\rangle = \varepsilon\pi\sigma^2\,\frac{\exp(-k^2\sigma^2)}{12(1-\exp(-k^2\sigma^2/2))}, \tag{75}$$

$$\langle u_2(\mathbf{k})^2\rangle = \varepsilon\pi\sigma^2\,\frac{(12k^2\sigma^4+1)\exp(-k^2\sigma^2)}{12-12(1-k^2\sigma^2)\exp(-k^2\sigma^2/2)}, \tag{76}$$

$$\langle u_3(\mathbf{k})^2\rangle = \varepsilon\pi\sigma^2\,\frac{s^2\exp(-k^2\sigma^2)}{3-s^2k^2\exp(-k^2\sigma^2/2)}. \tag{77}$$

For the fluctuations of u_1 and u_2 the deviation of $\mathbf{w}_r$ from its equilibrium value $\bar{\mathbf{w}}_r$ lies along one of the two main directions of the mapping. These fluctuations affect the positions $\mathbf{r}$ in the array to which the feature sets F_r are mapped and, therefore, will be called "longitudinal" fluctuations. As we can see from (75) and (76), all fluctuations with wavelengths significantly below σ are practically absent. Consequently, the main contribution to any statistical distortions of the mapping comes from fluctuations with long wavelengths which exhibit a $1/k^2$-singularity. To estimate the effect of these fluctuations on the final mapping, we expand (75) for the lowest possible wave number $k=2\pi/N$, assuming $k\sigma=2\pi\sigma/N \ll 1$. This yields

$$\langle u_1^2\rangle^{1/2} \approx N\sqrt{\varepsilon/24\pi} \approx 0.12N\varepsilon^{1/2}. \tag{78}$$

For this expression to be of the order of one lattice spacing or less, ε must be chosen inversely proportional to N^2, i.e. inversely proportional to the number of neurons in the array. However, it should be noted that for practical applications these smooth, fluctuating distortions over a large spatial scale are usually not very disturbing, as one is interested primarily in preserving the correct neighborhood relationships along the most important feature dimensions. Many applications will tolerate, therefore, much larger values of ε in the final convergence phase.

The u_3-mode affects the deviation of each $\mathbf{w}_r$ in the direction perpendicular to the array and, hence, the specification of the associated feature set F_r along this dimension. From (77) we see that in contrast to u_1 and u_2 the transverse fluctuations remain finite at $\mathbf{k}=0$ but, as already indicated above, their stability now depends sensitively on the value of s. Instability arises when s reaches a value s^* for which the denominator in (77) is no longer positive for all $\mathbf{k}$. The smallest value of s for which this happens is $s^*=\sigma\sqrt{3e/2}\approx 2.02\sigma$. The wavelength of the associated mode is $\lambda^*=\sigma\pi\sqrt{2}\approx 4.44\sigma$. One can conclude that the system tolerates inputs with a maximum transverse variance which is proportional to the width σ of the lateral adjustment function h_{rs}. By varying σ one can control the tolerance of the map against variance of inputs along the remaining orthogonal directions. When s approaches s^* from below, the system exhibits fluctuations which become increasingly intense and especially pronounced around the wavelength $\lambda\approx 4.44\sigma$. When s exceeds s^*, the symmetric equilibrium configuration becomes unstable and the system settles into a new equilibrium state.

5.3 Short-Ranged h_{rs}

In the short range limit h_{rs} includes only the nearest neighbors, i.e.

$$h_{rs}=\delta_{rs}+\sum_{n=\pm e_x,e_y}\delta_{r+n,s}. \tag{79}$$

In this case

$$\hat{h}(\mathbf{k})=1+2\cos k_x+2\cos k_y. \tag{80}$$

Focusing the representative case $k_2=0$, $k:=k_1$, we obtain

$$\langle u_1(\mathbf{k})^2\rangle=\frac{\varepsilon\cdot(3+2\cos k)^2}{4(1-\cos k)(9-2\cos k)}, \tag{81}$$

$$\langle u_2(\mathbf{k})^2\rangle=\frac{\varepsilon\cdot(44\sin^2 k+12\cos k+13)}{12(1-\cos k)(11+6\cos k)}, \tag{82}$$

$$\langle u_3(\mathbf{k})^2\rangle=\frac{\varepsilon s^2\cdot(1+2\kappa)^2}{2(4s^2\kappa^2-6s^2\kappa+15-4s^2)}, \tag{83}$$

where $\kappa:=\cos k_x+\cos k_y$. Equation (83) is also valid for $k_2\neq 0$. We observe again the $1/k^2$-singularity of longitudinal equilibrium fluctuations as one approaches long wavelengths. As before, we find $\hat{\mathbf{B}}_{11}(\mathbf{k})>\hat{\mathbf{B}}_{22}(\mathbf{k})$, i.e. u_2 is stiffer than u_1. As $\hat{\mathbf{D}}_{11}(\mathbf{k})=\hat{\mathbf{D}}_{22}(\mathbf{k})$ this situation is also reflected in the fluctuations of the two modes, which are smaller for the stiffer mode u_2. A similar analysis as in 5.2 yields $\langle u_1(\mathbf{k})^2\rangle_{\max}^{1/2}\approx 0.2\varepsilon^{1/2}N$. An overall distortion of the order of one lattice spacing or less again requires ε to be scaled inversely proportional to the number N^2 of neurons in the lattice. The critical value for the onset of the transverse instability in this case is $s^*:=\sqrt{12/5}\approx 1.549$ and the associated first unstable modes are characterized by $\kappa^*=3/4$. If $k_y=0$, this corresponds to a rather short wavelength of about 3.45 lattice spacings which is again comparable to the lateral width of the adjustment function.

6 Monte Carlo Simulations

To test the analytical results of the preceeding sections, we have carried out Monte Carlo simulations of the Markov process (7) on a two-dimensional square array of 32 × 32 units, i.e. $N=32$, with a constant step size of $\varepsilon=0.01$ and adjustment function (79). The resulting correlation functions $\langle u_n(\mathbf{k})^2\rangle$ are compared with the analytical expressions (81)–(83) in Figs. 2–4. We have

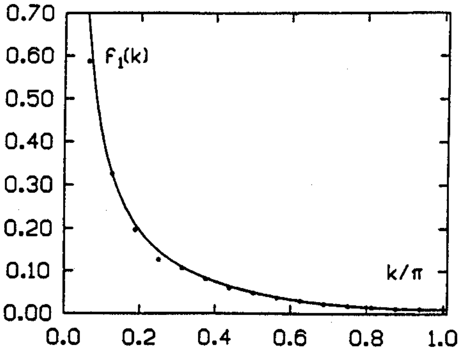

Fig. 2. Dependence of fluctuations of "soft mode" u_1 for the short-ranged adjustment function of (79) on the wave number k. The data points were obtained from a Monte Carlo simulation with 20000 samples of the Markov process (7) for fixed $\varepsilon=0.01$ and $s=0.0001$. Superimposed is the analytical result according to (81)

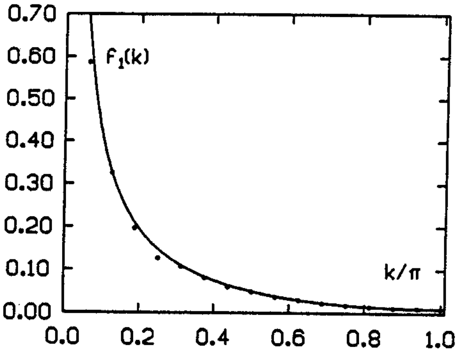

Fig. 3. Fluctuations of the "hard mode" u_2 of the same simulation as in Fig. 2 above [analytical result according to (82)]. For small wave numbers the fluctuations are smaller than for u_1. For larger wave numbers this distinction looses its significance

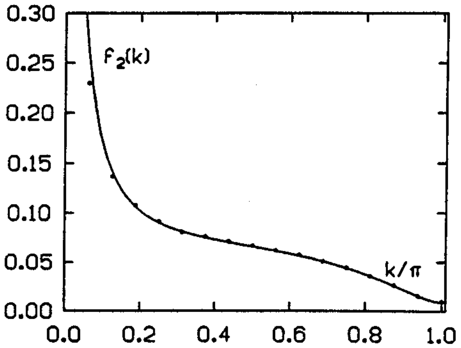

Fig. 4. Fluctuations of the "transverse mode" u_3 [analytical results according to (83)] for three different values of the thickness parameter s: for $s=0.0001$, i.e. essentially a two-dimensional input distribution, only small transverse fluctuations arise. For $s=1.3$ fluctuations begin to exhibit a broad maximum at about $k=0.58\pi$, which becomes very pronounced for $s=1.5$, which is closely below the critical value $s^*\approx 1.54$

also studied map formation for a long-ranged adjustment function. This case is discussed further below.

We generated an ensemble of states by taking 20000 successive "snapshots" of the Markov process (7), starting from the equilibrium configuration $\bar{\mathbf{w}}_{\mathbf{r}} = m\mathbf{e}_x + n\mathbf{e}_y$, $m, n = 1, 2, \ldots, 32$. Successive snapshots were 2000 Markov steps apart. From the generated ensemble, we evaluated the correlation functions $f_n(\mathbf{k}) := \langle u_n(\mathbf{k})^2 \rangle^{1/2}$, $n = 1, 2, 3$ at the discrete wave vectors $\mathbf{k} = \mathbf{e}_x \cdot 2\pi l/N$, $l = 1 \ldots 32$. The resulting data points for the "hard" mode u_1 and the "soft" mode u_2 at $s = 10^{-4}$ are shown in Figs. 2 and 3, respectively. Superimposed are the predictions due to (81), (82), which are in very good agreement with the simulation data. Figure 4 shows the data points of correlation function $f_3(\mathbf{k})$ (in units of s) for the transverse mode u_3 for the three parameter values $s = 10^{-4}$, 1.3 and 1.5 together with graphs of the theoretical predictions according to (83). At $s = 10^{-4}$ fluctuations decrease monotonously with decreasing wavelength. However, as s begins to approach the critical value $s^* \approx 1.54$ the fluctuations of modes in the vicinity of $k^* \approx 0.58\pi$ start to increase markedly. At $s = 1.5$, i.e. only little below the transition at $s^* \approx 1.54$, the fluctuations are already sufficiently strong to protrude into a significant part of the vertical extension of the parallelepiped. This signals the onset of the instability point, at and above which a discussion in terms of equilibrium deviations $\mathbf{u} = \mathbf{w} - \bar{\mathbf{w}}$ is no longer possible. For all three parameter values, the agreement with the theoretical predictions is very good.

A similar Monte Carlo simulation for the long-ranged case would require prohibitively much computation time. We have restricted, therefore, the simulation of the long-ranged case to a one-dimensional array consisting of a chain of $N = 128$ neurons. The parallelepiped was a two-dimensional strip of vertical extension $2s$ and length $N = 128$. For the step size we assumed again the value $\varepsilon = 0.01$. We generated 10000 snapshots, which were 1000 Markov steps apart. The analysis of Sect. 5 carries over to this case in a straightforward manner with very similar results. For the one-dimensional analog of the adjustment function (62), we obtain for the equilibrium fluctuations of the (only) longitudinal (u_1) and the transverse (u_2) mode:

$$\langle u_1(\mathbf{k})^2 \rangle = \frac{\varepsilon\sigma\sqrt{2\pi}(12k^2\sigma^4 + 1)\exp(-k^2\sigma^2)}{12(2 - [1 + \cos k - 2\sigma^2 k \sin k]\exp(-k^2\sigma^2/2))} \tag{84}$$

$$\langle u_2(\mathbf{k})^2 \rangle = \frac{\varepsilon\sigma\sqrt{2\pi}s^2\exp(-k^2\sigma^2)}{6 - 4s^2(1 - \cos k)\exp(-k^2\sigma^2/2)}. \tag{85}$$

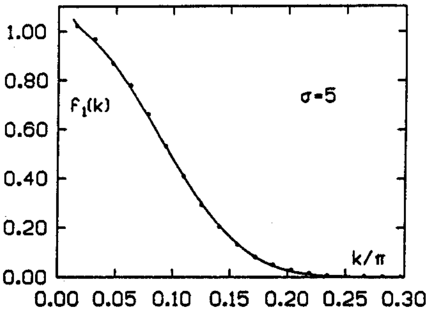

Fig. 5. Dependence of longitudinal fluctuations for a Gaussian adjustment function of width $\sigma = 5$ on wave number k. Data points are from a Monte Carlo simulation of a chain of $N = 128$ neurons. Superimposed is the theoretical result according to (84). The predicted exponential decay towards increasing wave numbers is well reproduced

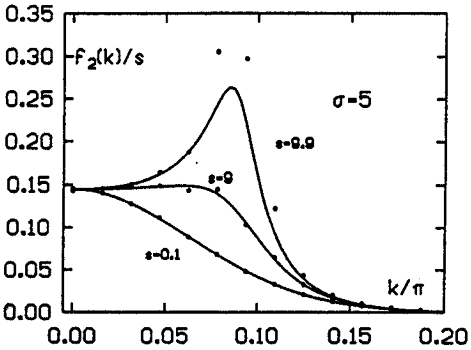

Fig. 6. The corresponding transverse fluctuations for three different values of s [analytical results according to (85)]. In comparison to Fig. 4 the critical value is $s^* \approx 10.1$ and the fluctuations show an exponential decline for larger k-values. The maximum associated with the transversal instability occurs at lower k-values due to the longer-ranged adjustment function

Apart from an additional prefactor of $(\sigma\sqrt{2\pi})^{-1}$ the $k \to 0$-limit of these expressions is identical to (76) and (77) of the two-dimensional case and symmetry breaking by the transverse modes occurs at precisely the same value s^* and wavelength λ^* as for the case studied in Sect. 5.2. In Figs. 5 and 6 we show a comparison of the graphs of the theoretical correlation functions (84), (85) and the results from a Monte Carlo simulation for $\sigma = 5$. Figure 5 shows the data points of the Monte Carlo simulation for the longitudinal fluctuations $f_1(k)$ at $s = 0.1$. The expected exponential decrease for $k^2\sigma^2 > 1$ is clearly reproduced. On the other hand, the expected $1/k$-singularity for $f_1(k)$ does not show up, as the required very small k-values are attained only on longer chains. Figure 6 presents the transverse fluctuations $f_2(k)$ for the three cases $s = 0.1$, i.e. an essentially one-dimensional input vector distribution, $s = 9.0$, i.e. markedly below the critical value $s^* \approx 2.02\sigma \approx 10.1$, and at $s=9.9$, i.e. just below s^*. Most significant differences to the short-ranged case discussed above and well confirmed by the data points of the simulation are the shift of the instability to lower wave numbers reflecting the longer scale $\sigma=5$ set by the long-ranged adjustment function, and the exponential decline of the fluctuations for $k\sigma \gg 1$.

7 Conclusion

Many pattern recognition and signal processing tasks can profit from a class of neurally inspired adaptive algorithms with the capability of forming low-dimensional reduced representations or maps of the most essential features of their input signals by learning. Convergence to these maps is a stochastic process driven by a sequence of input samples and besides a consideration of its average behaviour requires also a study of statistical fluctuations. We have carried out such study for one prominent representative of these algorithms which is due to Kohonen and which can be described by a Markov process. For this purpose we have derived a Fokker-Planck equation which describes the time evolution of the distribution function of this process in the final convergence phase. We could derive a criterion for the time dependence of the step size which guarantees convergence. For finite times, we find the presence of two different types of fluctuations. The first type affects the positions of the features in the mapping and consists of waves with large amplitude for long wavelengths and practically vanishing amplitudes for wavelengths significantly below the diameter of the adjustment function used in the algorithm. The second type affects the mapped features themselves. These "transverse" fluctuations increase with the variance of the inputs along the dimensions orthogonal to the map. If this variance exceeds some critical value, the associated modes become unstable and a reconfiguration of the mapping occurs. This behaviour has been observed earlier in computer simulations as an "automatic selection of feature dimensions". For the simple geometry of a parallelepiped as the support of the input distribution, the reconfiguration results in a breaking of the underlying symmetry of the input distribution. For this situation we can calculate the critical value of the variance and the typical wavelength of the associated unstable modes.

The approach of this paper may prove useful for the investigation of related self-organizing mapping algorithms. In addition, the theoretical results concerning convergence, fluctuations, and automatic selection of feature dimensions can facilitate the practical design of the algorithm for pattern recognition and related applications. Further questions especially interesting in this respect and likely to be accessible by this approach are the optimal choice of the step size $\varepsilon(t)$ for finite times and the extension of the analysis to more general input distributions.

Acknowledgement. This work has been supported by the German Ministry of Science and Technology (ITR-8800-G9).

Appendix

We want to show that for any positive function $\varepsilon(t)$ the conditions

$$\lim_{t\to\infty} \int_0^t \varepsilon(\tau)\,d\tau = \infty, \quad \lim_{t\to\infty} \varepsilon(t) = 0, \qquad (i)$$

and

$$\lim_{t\to\infty} \int_0^t \varepsilon(t')^2 \exp\left(-\beta \int_{t'}^t \varepsilon(t'')\,dt''\right) dt' = 0, \qquad (ii)$$

are equivalent for any $\beta > 0$.

Proof. *(ii)*→*(i)* is obvious for $\varepsilon > 0$. To show *(i)*→*(ii)*, choose $\delta > 0$ arbitrarily small and $a > 0$ such that $\varepsilon(t) < \beta\delta$ for all $t > a$. Let $\varepsilon_{\max} := \max_t \varepsilon(t)$. Then choose $b > a$ such, that $\exp\left(-\beta \int_a^t \varepsilon(\tau)\,d\tau\right) < \beta\delta/\varepsilon_{\max}$ for all $t > b$. Then for all $t > b$

$$\begin{aligned}
&\int_0^t \varepsilon(t')^2 \exp\left(-\beta \int_{t'}^t \varepsilon(t'')\,dt''\right) dt' \\
&\quad = \frac{1}{\beta}\left(\int_0^a + \int_a^t\right)\left[\varepsilon(t')\frac{\partial}{\partial t'}\exp\left(-\beta \int_{t'}^t \varepsilon(t'')\,dt''\right)\right] dt' \\
&\quad \leqq \frac{\varepsilon_{\max}}{\beta}\left[\exp\left(-\beta \int_{t'}^t \varepsilon(t'')\,dt''\right)\right]_{t'=0}^{t'=a} \\
&\quad + \delta \cdot \left[\exp\left(-\beta \int_{t'}^t \varepsilon(t'')\,dt''\right)\right]_{t'=a}^{t'=t} \\
&\quad \leqq \frac{\varepsilon_{\max}}{\beta} \cdot \frac{2\beta\delta}{\varepsilon_{\max}} + \delta = 3\delta.
\end{aligned}$$

As δ may be made arbitrarily small, *(ii)* must hold.

References

Bertsch H, Dengler J (1987) Klassifizierung und Segmentierung medizinischer Bilder mit Hilfe der selbstlernenden topologischen Karte. In: Paulus E (ed) 9. DAGM-Symposium Mustererkennung. Springer Informatik Fachberichte 149. Springer, Berlin Heidelberg New York, pp 166–170

Cottrell M, Fort JC (1986) A stochastic model of retinotopy: a self-organizing process. Biol Cybern 53:405–411

Erdi P, Barna G (1984) Self-organizing mechanism for the formation of ordered neural mappings. Biol Cybern 51:93–101

Gardiner CW (1985) Handbook of stochastic methods, 2nd edn. Springer, Berlin Heidelberg New York

Grossberg S (1976a) On the development of feature detectors in the visual cortex with applications to learning and reaction-diffusion systems. Biol Cybern 21:145–159

Grossberg S (1976b) Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors. Biol Cybern 23:121–134

Kaas JH, Merzenich MM, Killackey HP (1983) The reorganization of somatosensory cortex following peripheral nerve damage in adult and developing mammals. Ann Rev Neurosci 6:325–356

Kampen NG van (1981) Stochastic processes in physics and chemistry. North Holland, Amsterdam

Knudsen EI, du Lac S, Esterly SD (1987) Computational maps in the brain. Ann Rev Neurosci 10:41–65

Kohonen T (1982a) Self-organized formation of topologically correct feature maps. Biol Cybern 43:59–69

Kohonen T (1982b) Analysis of a simple self-organizing process. Biol Cybern 44:135–140

Kohonen T (1982c) Clustering, taxonomy and topological maps of patterns. Proceedings of the 6th International Conference on Pattern Recognition, Munich., pp 114–128

Kohonen T (1984) Self-organization and associative memory. Springer Series in Information Sciences 8. Springer, Berlin Heidelberg New York

Kohonen T (1986) Learning vector quantization for pattern recognition. Helsinki University of Technology, Report TKK-F-A601

Kohonen T, Mäkisara K, Saramäki T (1984) Phonotopic maps, – insightful representation of phonological features for speech recognition. Proceedings of the 7th International Conference on Pattern Recognition, Montreal., pp 182–185

Kushner HJ, Clark DS (1978) Stochastic approximation methods for constrained and unconstrained systems. Springer, Berlin Heidelberg New York

Malsburg C von der (1979) Development of ocularity domains and growth behaviour of axon terminals. Biol Cybern 32:49–62

Overton KJ, Arbib MA (1982) The branch arrow model of the formation of retino-tectal connections. Biol Cybern 45:157–175

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

Ritter H, Schulten K (1986b) Topology conserving mappings for learning motor tasks. In: Denker JS (ed) Neural networks of computing. AIP Conference Proceedings 151. Snowbird, Utah, pp 376–380

Ritter H, Schulten K (1987) Extending Kohonen's self-organizing mapping algorithm to learn ballistic movements. In: Eckmiller R, von der Malsburg C (eds) Neural computers. Springer, Berlin Heidelberg New York, pp 393–406

Suga N, O'Neill WE (1979) Neural axis representing target range in the auditory cortex of the mustache Bat Sci 206:351–353

Takeuchi A, Amari S (1979) Formation of topographic maps and columnar microstructures. Biol Cybern 35:63–72

Willshaw DJ, Malsburg C von der (1976) How patterned neural connections can be set up by self-organization. Proc R Soc London B 194:431–445

Willshaw DJ, Malsburg C von der (1979) A marker induction mechanism for the establishment of ordered neural mappings: its application to the retinotectal problem. Proc R Soc London B 287:203–243

Received: January 18, 1988

Dr. Helge Ritter
Physik-Department
Technische Universität München
James-Franck-Strasse
D-8046 Garching
Federal Republic of Germany