Biol. Cybern. 54, 99–106 (1986)

Biological Cybernetics

On the Stationary State of Kohonen's Self-Organizing Sensory Mapping

H. Ritter and K. Schulten

Department of Physics, Technical University of Munich, D-8046 Garching, Federal Republic of Germany

Abstract. The stationary state of the self-organizing sensory mapping of Kohonen is investigated. For this purpose the equation for the stationary state is derived for the case of one-dimensional and two-dimensional mappings. The equation can be solved for special cases, including the general one-dimensional case, to yield an explicit expression for the local magnification factor of the map.

1 Introduction

Self-organizing sensory mappings play a crucial role in the development and maintenance of many functions of the nervous system and especially the brain. Different sensory inputs, such as tactile (Kaas 1983; Merzenich 1983), visual (Whitteridge 1973) and acoustic (Suga 1979; Pickles 1982) inputs, are known to be mapped onto different areas of the cerebral cortex in an orderly, topology-preserving fashion, i.e., similar inputs are mapped onto neighbouring places in the cortex. These mappings are not genetically prespecified in a detailed manner but instead self-organize during the early stages of the formation of the nervous system. To some extent the mappings can remain plastic even later and adapt to subsequent changes in the environment or the sensors themselves. The degree of plasticity varies for different cortical mappings. For instance, the mapping from retina to cortex after its formation remains plastic only for a relatively short period of time, whereas for the somatosensory map considerable plasticity has been found even in adult animals (Kaas 1983; Merzenich 1983). In addition, different types of reorganisation after partial damage to afferent inputs have been observed (Kaas 1983).

Several algorithms for the formation of such mappings have been suggested (Edelman 1985; Takeuchi 1979; Willshaw 1976, 1979). In the following we will consider a proposal due to Kohonen (Kohonen 1982a, b). This proposal is not meant to model biological details but rather tries to capture the most essential features of such mappings for the benefit of remaining computationally tractable. The formation of the map is driven by a random sequence of sensory input signals whose probability distribution imprints on the final map in such a way that regions of the input signal space corresponding to frequent signal occurrences are mapped onto larger areas than regions corresponding to rarer input signals. Therefore the map magnifies more important sensory regions at the expense of less important sensory regions.

Below we shall illustrate the algorithm, obtain an equation for the final (stationary) map in terms of the signal probability distribution and derive the local magnification factor for special cases, including the general one-dimensional case.

2 The Model

As in (Kohonen 1982a, b) we consider a map $\phi: A \to B$ where B represents a lattice of neuronal units labelled by r and A a spatially continuous sensory source with elements v. A may represent, for example, the coordinate set of somatosensory receptors distributed densely over the body surface and B the set of those neuronal units of a layer in the cerebral cortex to which the somatosensory receptors are linked. The lattice B receives a sequence of input signals drawn randomly from A, the t-th signal $[t = 1, 2, 3 \ldots]$ being represented by $\mathbf{v}(t)$ $[\mathbf{v}(t) \varepsilon A]$. Each $\mathbf{v}(t)$ is received by all elements r of B simultaneously. To each unit belongs a vector $\mathbf{w}(\mathbf{r}, t) \varepsilon A$, which determines the response of unit r upon arrival of a signal $\mathbf{v}(t)$. The response shall be given by $f(\|\mathbf{v}(t) - \mathbf{w}(\mathbf{r}, t)\|)$, where $f(x)$ is a smooth real function peaked at $x = 0$ and of Gaussian type. Calling the union of all those points of A, which are closer to $\mathbf{w}(\mathbf{r}, t)$ than to any other $\mathbf{w}(\mathbf{s}, t)$, $\mathbf{s} \neq \mathbf{r}$, the "receptive field" A_r

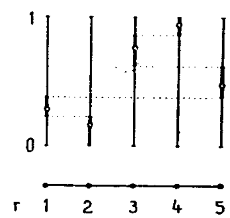

Fig. 1. Units and their receptive fields for the case of $A = [0, 1]$ and a linear array of 5 units (*full circles, bottom*). Above each unit r a copy of A is shown with the hollow circle denoting the value of $\mathbf{w}(\mathbf{r})$. The subset of A consisting of all those points, which are closer to $\mathbf{w}(\mathbf{r})$ than to any $\mathbf{w}(\mathbf{s})$, $\mathbf{s} \neq \mathbf{r}$, is shown bold and constitutes the receptive field of unit r

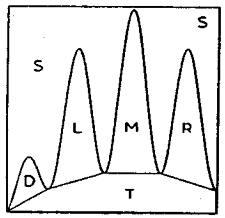

Fig. 2. Input space A: the hand surface is represented by the subset H consisting of the union of the areas D, L, M, R and T corresponding to thumb (D), left, middle, right finger (L, M, R) and palm (T). The remaining area S surrounding H does not yield inputs to B, i.e. $\mathbf{v}(t)$ never lies in this area

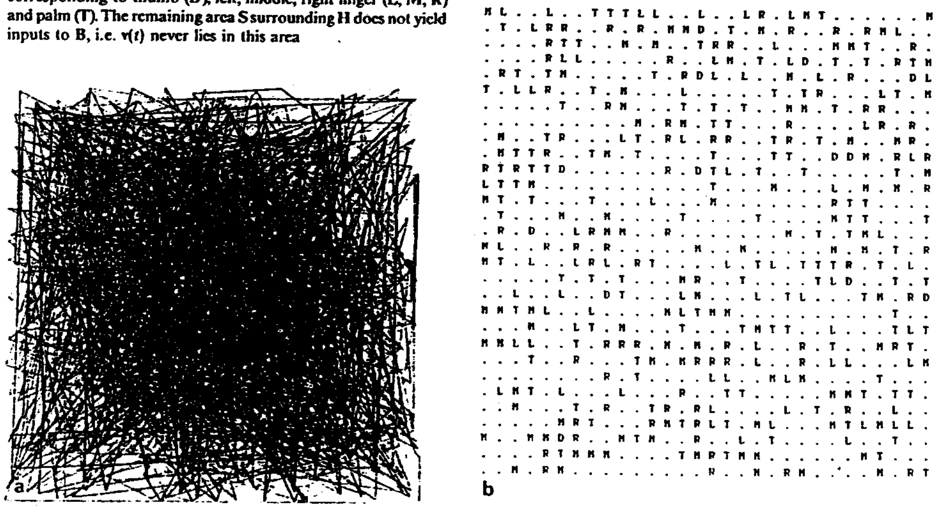

Fig. 3. a Initial configuration of the $\mathbf{w}(\mathbf{r}, 0)$ in input space: each $\mathbf{w}(\mathbf{r}, 0)$ takes on a value corresponding to a point in the $x-y$-plane and belongs to a unit at the mesh-point r of a 30×30 square mesh. Each mesh-point r is drawn at the location $\mathbf{w}(\mathbf{r}, 0)$. **b** Initial "cortical" configuration of the $\mathbf{w}(\mathbf{r}, 0)$: shown is a top view of the array of the 30×30 units. Each character position stands for one unit and characters D, L, M, R, T denote the region containing the receptive field center $\mathbf{w}(\mathbf{r}, 0)$ of the respective unit. *Dots* mark units which have not yet a receptive field within the hand surface H

of unit r (Fig. 1), we always have the maximal response at that unit r, for which $\mathbf{v}(t) \varepsilon A_r$. The mapping $\phi: A \to B$ we are seeking is then specified as follows: the image of a vector $\mathbf{v} \varepsilon A$ is the particular unit $u \varepsilon B$, which maximally responds to the signal v.

Initially the vectors $\mathbf{w}(\mathbf{r}, 0)$ and therefore the receptive fields of the individual units $\mathbf{r} \varepsilon B$ are distributed arbitrarily (e.g. randomly) in the input space A. Each incoming signal $\mathbf{v}(t) \varepsilon A$, $t = 1, 2, 3 \ldots$, causes the following adaptation step to take place:

1) Selection of the unit r with maximal response upon $\mathbf{v}(t)$

2) Modification of the receptive fields of unit r and all neighbouring units s according to

$$\mathbf{w}(\mathbf{s}, t+1) = \mathbf{w}(\mathbf{s}, t) + h(\mathbf{r} - \mathbf{s}, t) \cdot (\mathbf{v}(t) - \mathbf{w}(\mathbf{s}, t)) .$$

For each t, $h(\mathbf{x}, t)$ is peaked at $\mathbf{x} = 0$ and again of Gaussian type (either in each component if B is a high dimensional lattice or in the modulus of x), whose width $d(t)$ is a slowly decreasing function of t. All units s at a distance to unit r exceeding $d(t)$ receive only very little modification through 1), whereas all closer units are modified notably so as to improve their response to signal $\mathbf{v}(t)$. In the spirit of Edelman's group selection theory (Edelman 1985) a unit might be interpreted as a

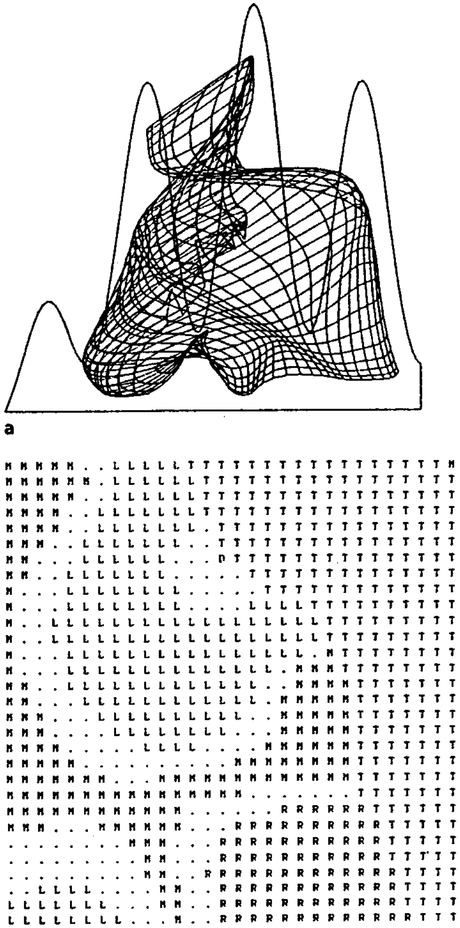

Fig. 4. a As Fig. 3a, but after 500 iterations. Superimposed is the hand area H, from which input signals v are originating. **b** As Fig. 3b, but after 500 iterations

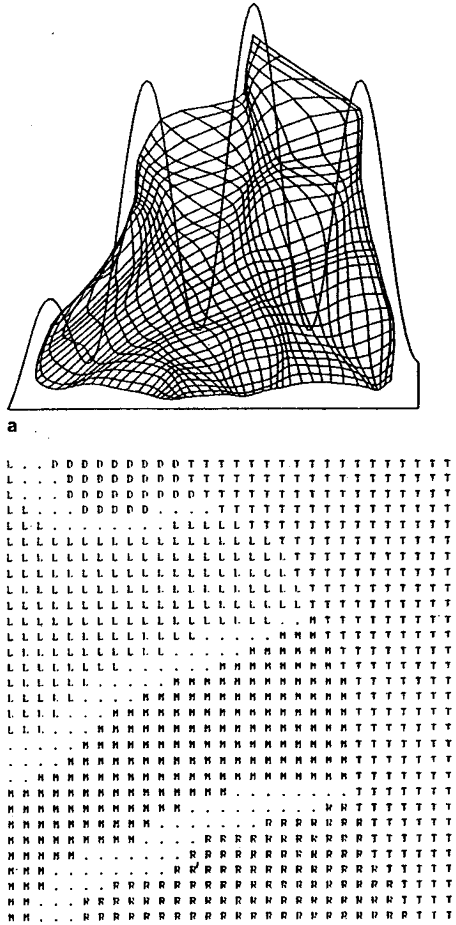

Fig. 5. a As Fig. 4a, but after 3000 iterations. **b** As Fig. 4b, but after 3000 iterations

group of neurons. At each input the most responsive group is selected and competes with neighbouring groups for a yet better response.

As an illustration of this algorithm we show the process of formation of the somatosensory mapping from the surface of a hand to the somatosensory region of the cortex. The map is chosen initially random and is shown to develop to a final ordered map. In this example the target area B on the cortex has been chosen to be a square array of 30×30 units and the input space A is the two-dimensional square depicted in Fig. 2. However, inputs v are only offered from the union H of the areas D, L, M, R and T in Fig. 2, which together represent the hand surface in our model. The initial points $\mathbf{w}(\mathbf{r}, 0)$ were equiprobably distributed over the whole input space A, i.e., the square enclosing H. This is represented in Fig. 3 which shows the initial positions $\mathbf{w}(\mathbf{r}, 0)$ together with straight line connections

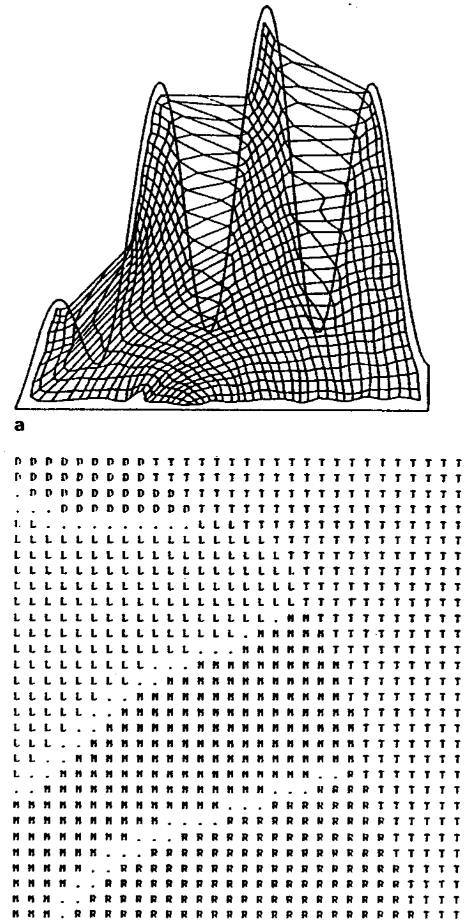

Fig. 6. a As Fig. 4a, but after 20000 iterations. **b** As Fig. 4b, but after 20000 iterations

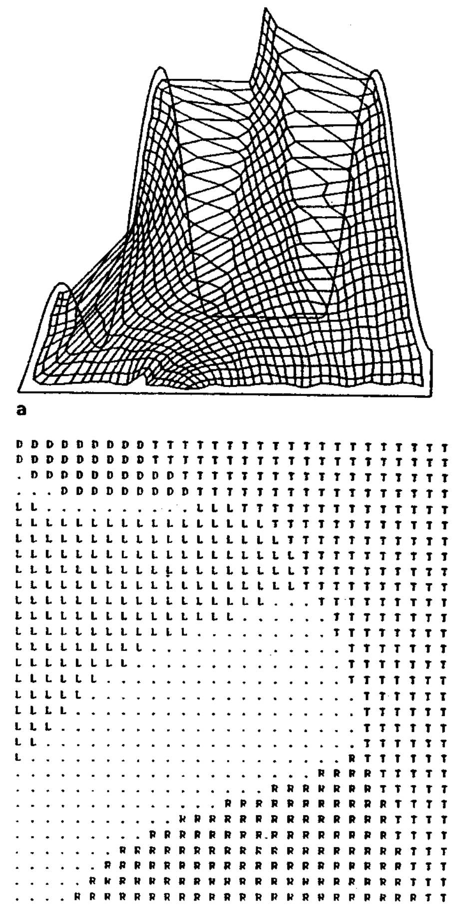

Fig. 7. a The same as the preceeding figure, but now with region M excluded from contributing input signals v; this situation corresponds to the removal of a finger. **b** Receptive field centers after removal of region M: units formerly responsive to finger *M* are now deprived of their inputs

between those pairs $\mathbf{w}(\mathbf{r}_1, t)$ and $\mathbf{w}(\mathbf{r}_2, t)$ for which $\mathbf{r}_1$ and $\mathbf{r}_2$ are neighbours in the lattice B. Obviously neighbourhood relationships are not conserved by the initial map $\mathbf{w}(\mathbf{r}, 0)$. When we carried out the algorithm described above, the input signals $\mathbf{v}(t)$ were selected randomly from H, and their probability density was chosen to increase towards the regions corresponding to the fingertips of the hand region in order to account for the higher density of tactile sensors there.

The function $h(\mathbf{x}, t)$ chosen in our simulation was a slowly decaying amplitude $a(t)$ times a Gaussian with initial width $d(t)$ of 5 lattice spacings, slowly decreasing to a final value of 2 lattice spacings after 5000 iterations and then remaining there for the rest of the simulation. The initial value of $a(t)$ was 0.5, exponentially decaying to 0.1 during the first 5000 iterations and constant

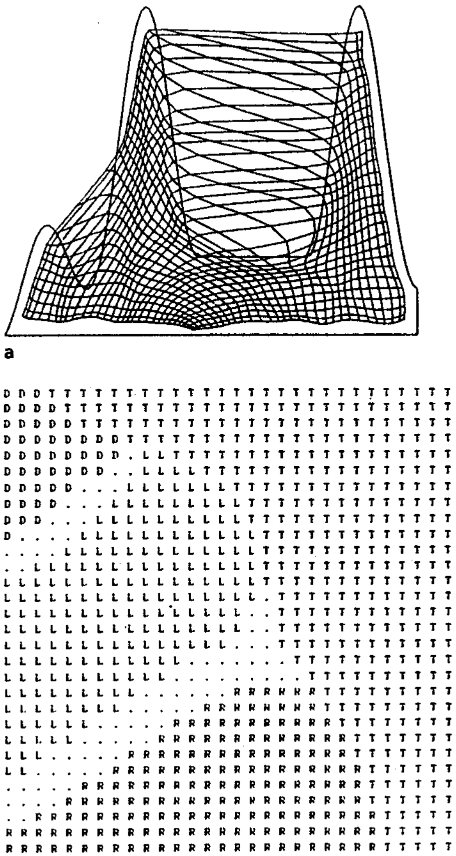

Fig. 8. a Readapted map after 50000 iterations subsequent to dissection of M. Formerly deprived units are re-employed by adjacent regions L, R, T thus allowing a finer representation there. b The same in "cortical view" as in Fig. 3b: formerly deprived units have developed receptive field centers located in neighbouring regions L, R, T

thereafter. Figures 4–6 show different stages in the formation of the map. After 20000 iterations the map has reached a rather orderly state. Following an experiment of Merzenich and Jenkins (Merzenich 1983) we "remove" at this stage the middle finger by envoking in the continuation of the algorithm no further inputs from the region M of this finger (see Fig. 7). The algorithm with a $d(t)$ value of 2 lattice spacings still exhibited enough plasticity to slowly adapt in the course of 50000 further iterations to this removal. Figure 8a shows the final distribution of the values $w(r, t)$ over the input space and Fig. 8b depicts the array B with its units marked by the location of their center of maximal sensitivity. The "cortical region" which in Fig. 7b immediately after the "amputation" is seen to be deprived of inputs has now been "invaded" by sensory input mainly from the adjacent regions L, R and T respectively, whereas more distant parts have changed only slightly. This plasticity is very similiar to that found for the somatosensory map in the experiment referred to above. The rearrangement of the map is accompanied by an increase of the map's local magnification factor for the adjacent parts of regions L, R and T, which results in a higher spatial sensory resolution there. This is also discernible from Fig. 8a, where an increase in the local density of the mesh-points $w(r, t)$ in the surround of the "amputation" can be seen. This latter effect is also in good qualitative agreement with experimental observations (Merzenich 1983).

3 Equation for the Final (Stationary) Mapping

As is shown in Kohonen (1982a, c), repeating the above steps 1) and 2) and decreasing $d(t)$ sufficiently slowly yields an ordered mapping from A onto the array of units such that neighbouring units are sensitive to neighbouring regions of A, irrespective of the initial values $w(r,0)$. The important dependence of the final mapping upon the probability distribution of the input signal $v(t)$ was discussed only qualitatively in (Kohonen 1982a) and shall be supplemented here by a more quantitative treatment.

As long as $d(t)$ is nonzero, $w(r, t)$ undergoes a usually nonzero change at each time step. Given a configuration $w(s, t)$ at time t, the expectation value of its change up to time $t+1$ is

$$\langle w(s, t+1)-w(s, t)\rangle=\langle h(s-r, d(t))\cdot(v(t)-w(s, t))\rangle \qquad (1)$$

where $\langle\ldots\rangle$ denotes the average over all possible values of $v(t)$ and $h(r, d(t))$ stands for the former $h(r, t)$ to make the d-dependence explicit.

Keeping $d(t)=d$ fixed for the moment, we shall call a configuration $w_d(s)$ an equilibrium configuration, if $w(s, t)=w_d(s)$ yields a vanishing expectation value in (1). We want to consider the equilibrium configuration in the limit of vanishing fluctuations, i.e.

$$w_0(s):=\lim_{d\to 0} w_d(s).$$

The following analysis will be restricted to the case of A and B being of the same dimension n (although the algorithm is capable of establishing a map between

different dimensional A and B either, see (Kohonen 1984) and the validity of two main assumptions:

i) We assume that for sufficiently many units and all sufficiently small d the equilibrium configurations $w_d(r)$ are sufficiently slowly varying with r to allow replacing them by corresponding smooth functions over a continuum of r-values and consequently setting $\phi^{-1}=w_0$. This basically assumes that the topological ordering of the final state has already occurred.

ii) We will assume bijective equilibrium configurations w_d. This is a reasonable assumption, since the discrete algorithm has the tendency to avoid mapping the same subregion of the signal space A to different parts of B.

In additional we require $h(x, d)$ to be of Gaussian type with width of order d and with vanishing first and isotropic second moments for all, d, i.e.

$$\int h(x, d)x_i x_j d^n x=\delta_{ij}M_d. \qquad (2)$$

We are now going to derive a necessary and sufficient differential equation for w_0 to be stationary. We start with the equilibrium condition for w_d

$$\langle h(r-s, d)\cdot(v(t)-w_d(s))\rangle=0 \qquad (3)$$

for all s. The location r of the maximally responding unit in B is in our continuum approximation determined through the implicit equation

$$w_d(r)=v(t). \qquad (4)$$

As we now proceed to average over $v(t)$, we will drop all references to t as $P(v)$ is independent of t, so that the only remaining time dependence is via $d=d(t)$. We then obtain

$$\begin{aligned}0&=\langle h(r(v)-s, d)\cdot(v-w_d(s))\rangle\\&=\langle h(r(v)-s, d)\cdot(w_d(r(v))-w_d(s))\rangle\\&=\int h(r(v)-s, d)\cdot(w_d(r(v))-w_d(s))\\&\quad\cdot P(v)d^n v.\end{aligned}$$

We introduce $q=r-s$ instead of v as the integration variable, write $Q(r)$ instead of $P(v(r))$ and denote by $D(r)$ the absolute value of the determinant of the Jacobian $J(r):=\partial v/\partial r$ i.e.

$$D(r)=|\det(\partial_i w_{d,j})| \qquad (5)$$

where we have made use of (4) to replace $v(t)$ by $w_d:=(w_{d,1}\ldots w_{d,n})^T$. These steps yield

$$0=\int h(q, d)\cdot(w_d(s+q)-w_d(s))\cdot Q(s+q)D(s+q)d^n q. \qquad (6)$$

For small values of d $h(q,d)$ is sharply peaked at $q=0$, so that we may expand in q and retain only the contribution due to the lowest nonvanishing moments of h (double indices are to be summed over)

$$\begin{aligned}0&=\int h(q, d)\,(q_i\partial_i w_d+\tfrac12 q_i q_j\partial_i\partial_j w_d+\ldots)\\&\quad\cdot(Q+q_k\partial_k Q+\ldots)\cdot(D+q_l\partial_l D+\ldots)d^n q\\&=\int h(q, d)q_i q_j d^n q\\&\quad\cdot((\partial_i w_d)\partial_j(QD)+\tfrac12 QD\cdot\partial_i\partial_j w_d)(s)+0(d^4)\\&=M_d\cdot[(\partial_i w_d)\partial_i(QD)\\&\quad+\tfrac12 QD\cdot\partial_i^2 w_d](s)+0(d^4).\end{aligned}$$

A necessary and sufficient condition for this equation to hold in the limit $d\to 0$ is

$$\partial_i w_0\left(\frac{\partial_i Q}{Q}+\frac{\partial_i D}{D}\right)=-\partial_i\partial_i w_0/2 \qquad (7)$$

or, introducing the Jacobian $J_{ij}=\partial_j w_{0,i}$:

$$J\cdot\nabla\ln(Q\cdot D)=-\tfrac12\Delta w_0. \qquad (8)$$

As we are only interested in the limit $d=0$, we shall henceforth denote w_0 by w solely. An alternative form of (8) is obtained via

$$\begin{aligned}\nabla\ln(Q\cdot D^{3/2})&=-\tfrac12 J^{-1}\Delta w+\tfrac12\nabla\ln(D)\\&=-\frac{1}{2\cdot D}(D\cdot J^{-1}\Delta w-\nabla D),\end{aligned}$$

or

$$\nabla\ln(Q\cdot D^{3/2})=-\frac{1}{2\cdot D}\cdot u\cdot\mathrm{sgn}(\det J), \qquad (9)$$

where u is given by

$$u=\det(J)\cdot J^{-1}\Delta w-\nabla\det(J).$$

In two dimensions with $w(r)=(a(r), b(r))^T$ this can be written more symmetrically as

$$u=\begin{pmatrix}(\nabla b)^T\partial_2(\nabla a)-(\nabla a)^T\partial_2(\nabla b)\\(\nabla a)^T\partial_1(\nabla b)-(\nabla b)^T\partial_1(\nabla a)\end{pmatrix}. \qquad (10)$$

Equations (8) or (9), together with suitable boundary conditions, determine the equilibrium configuration w(r), which in turn represents the inverse of the original map $A\to B$, since w(r) is the center in A of maximal sensitivity of unit $r\varepsilon B$.

4 Discussion

Although the nonlinearity of (8) and (9) makes a general discussion unfeasible, in one and two dimensions some consequences concerning the relationship between the local magnification factor and the driving probability distribution P may be drawn immediately.

As w(r) represents the inverse of the map $A\to B$, the local magnification factor M of the latter is given by

$M=1/D$ (cf. (4)). It has a simple dependence on the density $P(\mathbf{v}(t))$ of inputs in at least two cases.

The first case arises for w such that u vanishes. Then $Q\cdot D^{3/2}=\text{const.}$ and, therefore, (employing the identity $P(\mathbf{w}(\mathbf{r}))=Q(\mathbf{r})$)

$$M(\mathbf{w})=D^{-1}\propto P(\mathbf{w})^{2/3}. \tag{11}$$

u vanishes whenever A, B are either (i) both one-dimensional or (ii) of rectangular shape and P is a product $P(\mathbf{w})=P_A(a)\cdot P_B(b)$ with $\mathbf{w}=(a,b)^T$. In the latter case the choice $a=a(x)$, $b=b(y)$ splits (9) into two first order equations with x and y decoupled, yielding

$$x=c_1\cdot\int_{a_0}^{a} P_A(\alpha)^{2/3}d\alpha \tag{12}$$

$$y=c_2\cdot\int_{b_0}^{b} P_B(\beta)^{2/3}d\beta\,. \tag{13}$$

The four integration constants c_1, c_2, a_0, b_0 are fixed by a particular choice for the (arbitrary) starting point and the (arbitrary) scale for the labelling of the units in the x- and y-directions, respectively.

The second case in which a relationship between $P(\mathbf{v})$ and $\mathbf{w}(\mathbf{r})$ can be established is when w can be represented by a complex function

$$\mathbf{w}=(\mathrm{Re}\,\omega,\ \mathrm{Im}\,\omega)^T \tag{14}$$

with ω analytic in $z=x+iy$. This yields $QD=\text{const.}$ and therefore

$$M(\mathbf{w})\propto P(\mathbf{w})\,. \tag{15}$$

An example is given by $P(\mathbf{w})=\text{const.}/\|\mathbf{w}\|^2$ and the spaces

$$\mathrm{A}=\{\mathbf{w}\,|\,e^{-\pi}<\|\mathbf{w}\|<1\ \&\ w_2>0\}\,,$$

$$\mathrm{B}=[0,N]x[0,N]\,.$$

For this choice follows for $\omega(z)$ defined through (14)

$$\omega(z)=\exp\frac{\pi\cdot i\cdot z}{N}\,. \tag{16}$$

This yields a map ϕ from the semi-annulus A onto the square B. Such a kind of map connects, for example, the retina with the visual cortex.

In general, in the case of a two-dimensional mapping the magnification factor $M(\mathbf{w})$ of the stationary map is not expressible as a simple function of the local probability density $P(\mathbf{w})$ of the driving input as is implied in (Kohonen 1982c, 1984). Only in the one-dimensional case such relationship can be derived. The derivation yields $M(w)\propto P(w)^{2/3}$, a result which may be in contrast to the intuitive, but incorrect expectation $M(w)\propto P(w)$ suggested in (Kohonen 1984).

To test our findings we simulated Kohonen's map for the case of a one-dimensional lattice B of 1000 units

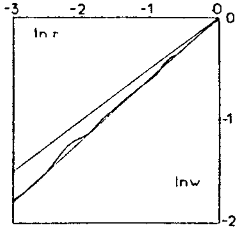

Fig. 9. Linear map after 100000 Iterations. Shown is $\ln(w(r))$ versus $\ln r$ for a probability density $P(v)=2v$, $v\in[0,1]$. Superimposed are the starting configuration $w(r,t=0)=\sqrt{r}$ (upper line segment) and the theoretical steady state map $w(r)=r^{3/5}$ (lower line segment). The apparent leftward increase of remnant fluctuations is due to the logarithmic axes

and an interval $\mathrm{A}=[0,1]$. The probability density of inputs from A was chosen linearly, i.e. $P(w)=2w$. Figure 9 represents the result of this simulation. Initially we chose the map $w(r)=\sqrt{r}$, whose magnification factor is proportional to P. After 100000 iterations the map has developed away from its initial configuration and reached the equilibrium curve $w(r)=r^{3/5}$, corresponding to a magnification factor $M(w)\propto w^{2/3}$ as predicted by (11).

5 Conclusion

We have derived an equation for the equilibrium state of a self-organizing topographic mapping due to Kohonen and for some special cases derived analytical expressions of the local magnification factor in terms of the probability density of the driving input. It is shown, that the local magnification factor in the one-dimensional case is proportional to $P^{2/3}$, whereas in two dimensions no general local expression in terms of the probability density can be given.

References

Edelman GM, Finkel LH (1985) Neuronal group selection in the cerebral cortex. In: 1st Symposium of the Neurosciences Institute, La Lolla, California, October 3–8 1982 (in press)

Kaas JH, Merzenich MM, Killackey HP (1983) The reorganization of somatosensory cortex following peripheral nerve damage in adult and developing mammals. Annu Rev Neurosci 6:325–56

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

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

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

Kohonen T (1984) Self-organization and associative memory. Springer, Berlin Heidelberg New York Tokyo

Merzenich MM, Jenkins WM (1983) Dynamic maintenance and alterability of cortical maps in adults; some implications. In: Klinke R, Hartmann R (eds) Hearing – Physiological Bases and Psychophysics. Springer, Berlin Heidelberg New York Tokyo

Pickles JO (1982) An introduction to the physiology of hearing. Academic Press, New York

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

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

Whitteridge D (1973) Projection of optic pathways to the visual cortex. In: Jung R (ed) Visual centers in the brain. Springer, Berlin Heidelberg New York (Handbook of sensory physiology, vol. VII/3B, pp 247–268)

Willshaw DJ, Malsburg C von der (1976) How patterned neural connections can be set up by self-organization. Proc R Soc Lond Ser 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 Lond Ser B 287:203–243

Received: December 7, 1985

Helge Ritter und Klaus Schulten
Dept. of Physics
Technical University of Munich
James-Franck-Strasse
D-8046 Garching
Federal Republic of Germany