Communicated by Shun-ichi Amari

Dynamics and Formation of Self-Organizing Maps

Jun Zhang
Neurobiology Group, 3210 Tolman Hall, University of California, Berkeley, CA 94720 USA

Amari (1983, 1989) proposed a mathematical formulation on the self-organization of synaptic efficacies and neural response fields under the influence of external stimuli. The dynamics as well as the equilibrium properties of the cortical map were obtained analytically for neurons with binary input–output transfer functions. Here we extend this approach to neurons with arbitrary sigmoidal transfer function. Under the assumption that both the intracortical connection and the stimulus-driven thalamic activity are well localized, we are able to derive expressions for the cortical magnification factor, the point-spread resolution, and the bandwidth resolution of the map. As a highlight, we show analytically that the receptive field size of a cortical neuron in the map is inversely proportional to the cortical magnification factor at that map location, the experimentally well-established rule of inverse magnification in retinotopic and somatotopic maps.

1 Introduction

The self-organization of the nervous system and the consequential formation of cortical maps have been studied quite extensively (von der Malsburg 1973; Swindale 1980; Kohonen 1982; Linsker 1986; Miller *et al.* 1989). A cortical map, or more generally, a computational map refers to the neural structure of representing a continuous stimulus parameter by a place-coded populational response, whose peak location reflects the mapped parameter (Knudsen *et al.* 1987). The cortical neurons in the map, each with a slightly different range of stimulus selectivity established during developmental course, operate as preset parallel filters on the afferent stimulus almost simultaneously. The stimulus parameter, now coded as the location of the most active neuron(s), can be accessed by higher processing centers via relatively simple neural connections.

The network models for such cortical maps are usually composed of several layers of neurons from sensory receptors to cortical units, with feedforward excitations between the layers and lateral (or recurrent) con-

 © 1991 Massachusetts Institute of Technology

nection within the layer. Standard techniques include (1) Hebbian rule and its variations for modifying synaptic efficacies, (2) lateral inhibition (in the general sense) for establishing topographical organization of the cortex as well as sharpening the cells' tuning properties, and (3) adiabatic approximation in decoupling the dynamics of relaxation (which is on the fast time scale) and the dynamics of learning (which is on the slow time scale) of the network. However, in most cases, only computer simulation results were obtained and therefore provided limited mathematical understanding of the self-organizing neural response fields.

In Takeuchi and Amari (1979) and Amari (1983, 1989), a general mathematical formulation was presented to study analytically the existence conditions, the resolution and magnification properties, as well as the dynamic stability of cortical maps. This rather rigorous approach yielded very illuminating results. In particular, they suggested by perturbation analysis that, in the presence of periodic boundary conditions of the mapping, the relative values of the afferent spread size and the receptive field size will determine the emergence of a block-like, columnar structure as opposed to a continuous, topographic organization. Since their analysis was restricted to binary-valued neurons only, that is, neurons with step-function as their input–output transfer function, it is certainly desirable to extend this approach to the more general case of neurons with arbitrary sigmoidal transfer functions.

2 Dynamics of Self-Organization Revisited

The network that Amari and colleagues investigated consists of three layers, a sensory receptor layer, a thalamic layer, and a cortical layer, with feedforward connections between the layers and lateral inhibition within the cortical layer only (Fig. 1). Following Takeuchi and Amari (1979), the activity of the cortical neuron at location $\mathbf{x}$ (a 2D vector in general) and time t may be described by its net input $u(\mathbf{x}, t)$ (postsynaptic membrane potential with respect to the resting state) and output $v(\mathbf{x}, t)$ (average firing rate of the spike train) interrelated via some monotone increasing (sigmoidal) input–output function: $v = f(u)$, $u \in (-\infty, \infty)$, $v \in (0, 1)$. To further indicate that these variables are functions of stimulus parameter $\mathbf{y}$ (a vector) and time parameter of the learning dynamics τ, we shall write in this article $u(\mathbf{x}, \mathbf{y}, t, \tau)$ and $v(\mathbf{x}, \mathbf{y}, t, \tau)$ instead. The receptors have delta-function tuning to the stimulus parameter, and they feed into the thalamic layer with localized afferent spread. Notice that $\mathbf{y}$ is used to denote the stimulus variable *as well as* to index cells in the thalamic layer according to their optimal stimulus parameter (i.e., according to their connections from the receptor layer). The cortical neurons in the model receive both thalamocortical afferent projections as well as intracortical

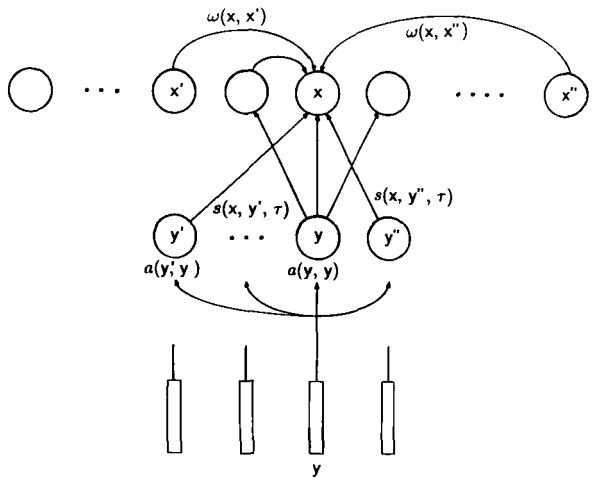

Figure 1: The three layered network model for the self-organizing cortical maps. The receptors, having delta-function tuning for the stimulus $\mathbf{y}$, feed into thalamic neurons, with the stimulus-driven thalamic activity denoted by $a(\mathbf{y}', \mathbf{y})$. The cortical neurons receive both the intracortical interaction characterized by the weighting function $\omega(\mathbf{x}, \mathbf{x}')$ and the thalamocortical input with synaptic connection $s(\mathbf{x}, \mathbf{y}, \tau)$ modifiable (based on Hebbian learning rule) during development. Note that each thalamic neuron is indexed by its optimal driving stimulus (according to the connections from the receptor layer).

lateral connections. The relaxation of the system is dictated, on the fast time scale t, by the equation

$$\begin{aligned} \frac{\partial}{\partial t} u(\mathbf{x}, \mathbf{y}, t, \tau) &= -u(\mathbf{x}, \mathbf{y}, t, \tau) + \int \omega(\mathbf{x}, \mathbf{x}')\, f[u(\mathbf{x}', \mathbf{y}, t, \tau)]\, \mathrm{d}\mathbf{x}' \\ &\quad + \int s(\mathbf{x}, \mathbf{y}', \tau)\, a(\mathbf{y}', \mathbf{y})\, \mathrm{d}\mathbf{y}' \end{aligned} \tag{2.1}$$

where $\omega(\mathbf{x}, \mathbf{x}')$ characterizes the weighting of lateral connections within the cortex from location $\mathbf{x}'$ to location $\mathbf{x}$, assumed to be unchanged with time; $a(\mathbf{y}', \mathbf{y})$ represents the thalamocortical afferent activity at $\mathbf{y}'$ [the first argument in the function $a(\cdot, \cdot)$] on the presentation of the stimulus $\mathbf{y}$ [the second argument in $a(\cdot, \cdot)$]; and $s(\mathbf{x}, \mathbf{y}, \tau)$ is the synaptic efficacy from the thalamocortical afferent $\mathbf{y}$ to the cortical neuron $\mathbf{x}$, which varies on a slow time scale τ and is thus treated as constant on the fast time scale t. This "adiabatic approximation" allows Amari (1989) to construct a global Lyapunov function $L[u]$ that is a function of $\mathbf{y}, t, \tau$ ($\mathbf{x}$ having been integrated) and thus represents the overall pattern of cortical

activity. It was proved that, on the stimulus presentation $\mathbf{y}$ at time τ, the value of $L[u]$ monotonously decreases, on the fast time scale t, as $u(\mathbf{x}, \mathbf{y}, t, \tau)$ evolves according to equation 2.1 until $L[u] = L(\mathbf{y}, t, \tau)$ reaches a minimum value $L_{\min}(\mathbf{y}, \tau)$ while $u(\mathbf{x}, \mathbf{y}, t, \tau)$ reaches its "equilibrium" solution $\bar{u}[\mathbf{x}, \mathbf{y}, \tau, s(\cdot), a(\cdot)]$ [$\bar{u}$ is a functional of $s(\mathbf{x}, \mathbf{y}, \tau)$ and $a(\mathbf{y}', \mathbf{y})$, and the bar denotes the equilibrium of the relaxation phase]. This establishes a cortical response field $\bar{u} = \bar{u}(\mathbf{x}, \mathbf{y}, \tau)$ relating the external stimulus $\mathbf{y}$ to the cortical activity at $\mathbf{x}$ at time τ. To study the self-organization of this mapping, the synaptic efficacy $s(\mathbf{x}, \mathbf{y}, \tau)$ is assumed to be modifiable, on the slow time scale τ, according to the following equation of learning (Hebbian rule):

$$\frac{\partial}{\partial \tau} s(\mathbf{x}, \mathbf{y}, \tau) = -s(\mathbf{x}, \mathbf{y}, \tau) + \eta \int a(\mathbf{y}, \mathbf{y}')\, f[\bar{u}(\mathbf{x}, \mathbf{y}', \tau)]\, p(\mathbf{y}')\, \mathrm{d}\mathbf{y}' \tag{2.2}$$

Note that, for the dynamics of learning, stimuli presentations are considered to be stochastically independent at each time τ with some prescribed probability distribution $p(\mathbf{y}')$. Here we set $p(\mathbf{y}') = \textit{constant}$ (and thus let absorbed into the constant η) to indicate the normal developmental course. Note that the integration is with respect to the second slot of $a(\cdot, \cdot)$, the argument representing stimulus parameter. At the end of the network learning, synapses get "matured" so that $s(\mathbf{x}, \mathbf{y}, \tau)$ becomes the time-independent $S(\mathbf{x}, \mathbf{y})$:

$$S(\mathbf{x}, \mathbf{y}) = \eta \int a(\mathbf{y}, \mathbf{y}')\, f[U(\mathbf{x}, \mathbf{y}')]\, \mathrm{d}\mathbf{y}' \tag{2.3}$$

whereas $\bar{u}(\mathbf{x}, \mathbf{y}, \tau)$ becomes the time-independent $U(\mathbf{x}, \mathbf{y})$:

$$\begin{aligned} U(\mathbf{x}, \mathbf{y}) &= \int \omega(\mathbf{x}, \mathbf{x}')\, f[U(\mathbf{x}', \mathbf{y})]\, \mathrm{d}\mathbf{x}' + \int S(\mathbf{x}, \mathbf{y}')\, a(\mathbf{y}', \mathbf{y})\, \mathrm{d}\mathbf{y}' \\ &= \int \omega(\mathbf{x}, \mathbf{x}') f[U(\mathbf{x}', \mathbf{y})] \mathrm{d}\mathbf{x}' + \int \kappa(\mathbf{y}, \mathbf{y}') f[U(\mathbf{x}, \mathbf{y}')] \mathrm{d}\mathbf{y}' \end{aligned} \tag{2.4}$$

Here $\kappa(\mathbf{y}, \mathbf{y}')$ is the autocorrelation between the thalamocortical afferents defined as

$$\kappa(\mathbf{y}, \mathbf{y}') = \eta \int a(\mathbf{y}'', \mathbf{y})\, a(\mathbf{y}'', \mathbf{y}')\, \mathrm{d}\mathbf{y}'' \tag{2.5}$$

Equivalently, we may write

$$f^{-1}[V(\mathbf{x}, \mathbf{y})] = \int \omega(\mathbf{x}, \mathbf{x}')\, V(\mathbf{x}', \mathbf{y})\, \mathrm{d}\mathbf{x}' + \int \kappa(\mathbf{y}, \mathbf{y}')\, V(\mathbf{x}, \mathbf{y}')\, \mathrm{d}\mathbf{y}' \tag{2.6}$$

3 Reduced Equation of Cortical Map Formation

The master equation (equation 2.4 or 2.6) developed by Amari and colleagues describes the formation of cortical maps as equilibrium solutions to the dynamic self-organization of layered neural systems. In Amari

(1989), the resolution and magnification properties were studied for neurons with binary input–output transfer function, where $f(\cdot)$ assumes the value of either 1 or 0 and thus the integral in equation 2.4 can be explicitly evaluated. In Chernjavsky and Moody (1990), the case of linear transfer function $f(u) = au + b$ was studied. Here we relax these restrictions and consider arbitrary transfer function $f(u)$. We will derive approximations of equation 2.6 for well-localized, translation-invariant functions of the intracortical connection $\omega(\mathbf{x}, \mathbf{x}')$ and the stimulus-driven thalamic activity $a(\mathbf{y}, \mathbf{y}')$:

$$\omega(\mathbf{x}, \mathbf{x}') = \omega(\mathbf{x} - \mathbf{x}') \tag{3.1}$$

$$a(\mathbf{y}, \mathbf{y}') = a(\mathbf{y} - \mathbf{y}') \tag{3.2}$$

It follows from equations 2.5 and 3.2 that the afferent autocorrelation $\kappa(\mathbf{y}, \mathbf{y}')$ must also be translation invariant:

$$\kappa(\mathbf{y}, \mathbf{y}') = \kappa(\mathbf{y} - \mathbf{y}') \tag{3.3}$$

Now we consider the first integral term in equation 2.6. For simplicity, $\mathbf{x}$, $\mathbf{y}$ are taken as real numbers x, y here (i.e., the mapping is one-dimensional). Expanding $V(x', y)$ into the Taylor series around point (x, y)

$$\begin{aligned} V(x', y) &= V(x, y) + \frac{\partial V(x, y)}{\partial x}(x' - x) + \cdots \\ &\quad + \frac{1}{n!}\frac{\partial^n V(x, y)}{\partial x^n}(x' - x)^n + \cdots \end{aligned} \tag{3.4}$$

we have

$$\int \omega(x - x') V(x', y)\, \mathrm{d}x' = a_0 V(x, y) + a_1 \frac{\partial V}{\partial x} + \cdots + a_n \frac{\partial^n V}{\partial x^n} + \cdots \tag{3.5}$$

where

$$\begin{aligned} a_k &= \frac{1}{k!} \int \omega(x - x')\, (x' - x)^k\, \mathrm{d}x' \\ &= \frac{(-1)^k}{k!} \int \omega(t)\, t^k\, \mathrm{d}t \end{aligned} \tag{3.6}$$

Similarly,

$$\int \kappa(y - y')\, V(x, y')\, \mathrm{d}y' = b_0\, V(x, y) + b_1 \frac{\partial V}{\partial y} + \cdots + b_n \frac{\partial^n V}{\partial y^n} + \cdots \tag{3.7}$$

with

$$b_k = \frac{(-1)^k}{k!} \int \kappa(t)\, t^k\, \mathrm{d}t \tag{3.8}$$

Therefore, the master equation 2.6 is transformed into[1]

$$\begin{aligned} f^{-1}(x,y) &= (a_0+b_0)\,V + \left(a_1\frac{\partial V}{\partial x} + b_1\frac{\partial V}{\partial y}\right) + \cdots \\ &+ \left(a_n\frac{\partial^n V}{\partial x^n} + b_n\frac{\partial^n V}{\partial y^n}\right) + \cdots \end{aligned} \tag{3.9}$$

By assuming that $\omega(t)$ and $a(t)$ are well localized, we imply a_k, b_k converge rapidly to 0 as $k \to \infty$. Taking only a few leading terms in the expansion, and further assuming that $\omega(x,x') = \omega(|x-x'|)$ and $a(y,y') = a(|y-y'|)$ are both even functions of their arguments, therefore making $a_1 = 0$ and $b_1 = 0$, we obtain

$$f^{-1}[V(x,y)] = (a_0+b_0)\,V + \left(a_2\frac{\partial^2 V}{\partial x^2} + b_2\frac{\partial^2 V}{\partial y^2}\right) \tag{3.10}$$

or

$$G(V) = a_2\frac{\partial^2 V}{\partial x^2} + b_2\frac{\partial^2 V}{\partial y^2} \tag{3.11}$$

with

$$G(V) = f^{-1}(V) - (a_0+b_0)\,V \tag{3.12}$$

If the cortical lateral connection is balanced in its total excitation and total inhibition, $a_0 = 0$. If the afferent autocorrelation is normalized, $b_0 = \eta > 0$. Equation 3.11 is a semilinear second-order partial differential equation. When $a_2b_2 > 0$, it is of elliptic type; when $a_2b_2 < 0$, it is of hyperbolic type. The standard techniques of solving equation 3.11 can be found in mathematical textbooks, such as Chester (1971). In particular, 3.11 may be linearized and transformed into the canonical forms of (when $a_2b_2 < 0$) $\partial_{\mathbf{XY}}V + cV = 0$, known as the telegraph equation, or (when $a_2b_2 > 0$) $\nabla^2 V + cV = 0$, known as the Helmholtz equation. These linear second-order partial differential equations have closed-form solutions when given appropriate boundary conditions.

4 Resolution and Magnification of the Map

The solution of equation 3.11 $V = V(x,y)$ represents the response of the neuron at location x due to stimulus parameter y after the cortical map matures. When x is fixed, that is, at a particular cortical location x_0, the neuron's response is a function of stimulus parameter y. Maximal

[1]This simplified derivation of equation 3.9 is suggested to the author by Dr. S. Amari. This equation was earlier obtained by expanding $\omega(x-x')$ into the sum of $\delta(x-x')$ and its derivatives $\delta^{(n)}(x-x')$, where the delta-function is envisioned as the limiting case (i.e., with zero width) of a normalized gaussian function and its successive derivatives represent derivatives of gaussian that become less and less localized (Zhang 1990).

response is achieved for some optimal stimulus y_0, which is determined by

$$\left.\frac{\partial V(x,y)}{\partial y}\right|_{x=x_0} = 0 \tag{4.1}$$

Obviously the optimal stimulus thus obtained is different for each location x_0. The optimal stimulus parameter y as a function of cortical locations x may be written as

$$y = \mu(x) \tag{4.2}$$

so that equation 4.1 holds identically for all x

$$V_2[x, \mu(x)] = 0 \tag{4.3}$$

Here and in the sequel, we use the subscript(s) 1, 2 of V to denote partial derivative(s) with respect to the first and/or second argument(s) in $V(\cdot,\cdot)$. Upon the presentation of an optimal stimulus, the maximal response of the neuron at x is

$$V_{\max}(x) = V[x, \mu(x)] \tag{4.4}$$

Suppose that this maximal response is everywhere the same (i.e., cortical neurons are indistinguishable)

$$V_{\max}(x) = \textit{constant} \tag{4.5}$$

or

$$\frac{\mathrm{d}}{\mathrm{d}x} V[x, \mu(x)] = V_1[x, \mu(x)] + V_2[x, \mu(x)]\mu'(x) = 0 \tag{4.6}$$

It follows from equations 4.6 and 4.3 that

$$V_1[x, \mu(x)] = 0 \tag{4.7}$$

or equivalently

$$\left.\frac{\partial V(x,y)}{\partial x}\right|_{y=\mu(x)} = 0 \tag{4.8}$$

Hence $x = \mu^{-1}(y)$ defines the location of maximal response (i.e., center of the cortical map) as a function of the stimulus variable. Differentiating equations 4.7 and 4.3 yields, respectively,

$$\frac{\mathrm{d}}{\mathrm{d}x} V_1[x, \mu(x)] = V_{11}[x, \mu(x)] + V_{12}[x, \mu(x)]\,\mu'(x) = 0 \tag{4.9}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} V_2[x, \mu(x)] = V_{21}[x, \mu(x)] + V_{22}[x, \mu(x)]\,\mu'(x) = 0 \tag{4.10}$$

Remembering that the order of partial differentiations is interchangeable $V_{12} = V_{21}$, we immediately have

$$V_{11}[x, \mu(x)] - V_{22}[x, \mu(x)]\,[\mu'(x)]^2 = 0\ . \tag{4.11}$$

On the other hand, equation 3.11 should always be satisfied:

$$G(V_{\max}) = a_2 V_{11}[x, \mu(x)] + b_2 V_{22}[x, \mu(x)] \tag{4.12}$$

From 4.11 and 4.12, we finally obtain

$$V_{11}[x, \mu(x)] = \frac{G(V_{\max})\ [\mu'(x)]^2}{a_2\,[\mu'(x)]^2 + b_2} \tag{4.13}$$

$$V_{22}[x, \mu(x)] = \frac{G(V_{\max})}{a_2\,[\mu'(x)]^2 + b_2} \tag{4.14}$$

The above results can be understood intuitively. Recall that the cortical magnification factor (CMF) is defined as the ratio of a resulting shift of the mapped location in the cortex over a change in the stimulus parameter. In the present context, it is simply equal to $[\mu'(x)]^{-1} = [d\mu(x)/dx]^{-1}$, the reciprocal of the derivative of the function $y = \mu(x)$, which is solvable from equation 4.3 or 4.7. The cortical magnification factor is apparently a function of cortical location x.

The resolution of the map can be described by two related measures. For a fixed stimulus parameter, the extent of cortical regions being excited is a measure of the stimulus localization in a populational response ("point-spread resolution"). At a particular cortical location, the range of effective stimuli is a measure of the stimulus selectivity of a single cell ("bandwidth resolution"). To get the intuitive picture, we draw a family of "isoclines" of $V(x, y)$ in the (x, y)-coordinates whereby $V(x, y) = constant$ along each curve (Fig. 2a). The variation of $V(x, y)$ in the vertical direction indicates to look at cells' response at a fixed cortical location ($x = x_0$) while changing the stimulus parameter — the vertical bar measures, in reciprocal, the bandwidth resolution. The variation of $V(x, y)$ in the horizontal direction indicates to fix the stimulus parameter while looking at responses of cells at different cortical locations — the horizontal bar measures, in reciprocal, the point-spread resolution. In both cases, of course, one needs to specify a criterion (in terms of percentage of maximal response, for instance) to discuss the magnitude of each resolution measure. From the graph, it is obvious that these two measures are interrelated.

If we take a slice (cross-section) along the vertical direction, the value of $V(x_0, y)$ may be schematically plotted (Fig. 2b). Note that this is a plot of response amplitude V versus the stimulus parameter y, with the cortical location x_0 fixed. The peak location of this curve corresponds to $V_{\max} = V(x_0, y_0)$, with $y_0 = \mu(x_0)$ representing the optimal stimulus for the cell located at x_0. The "width" of this tuning curve represents *the extent*

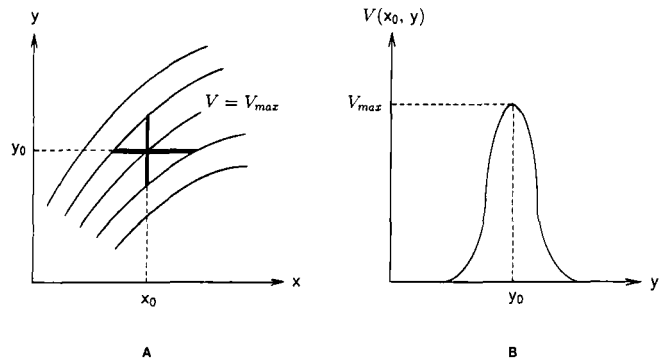

Figure 2: (a) The "isoclines" of V are plotted on the (x, y)-coordinates, whereby along each curve $V(x, y) = constant$. In particular, the curve $y = \mu(x)$ corresponding to $V(x, y) = V_{\max}$ was labeled. The vertical bar is a measure of the bandwidth resolution of the cortical map, while the horizontal bar is a measure of the point-spread resolution of the map (see text for details). (b) Taking a slice (cross-section) along the vertical direction in (a), the value of $V(x_0, y)$ is plotted as a function of the stimulus parameter y, representing the tuning curve of the cell at a particular cortical location x_0.

of stimulus selectivity, or the reciprocal of the bandwidth resolution of the map. If the cell's tuning curve is symmetric about its peak, it may be approximated by a quadratic function (at least near the peak location $y_0 = \mu(x_0)$, where $\partial V(x_0, y_0)/\partial y$ according to equation 4.3)

$$V(x_0, y) \simeq V(x_0, y_0) + \frac{1}{2}\frac{\partial^2 V(x_0, y_0)}{\partial y^2}(y - y_0)^2 + \cdots \tag{4.15}$$

The "width" of this parabola is inversely related to the quadratic coefficient $\partial^2 V(x_0, y_0)/\partial y^2$, or simply $V_{22}[x_0, \mu(x_0)]$. We may replace x_0 with x to indicate that this analysis applies to all cortical locations. Therefore $V_{22}[x, \mu(x)]$ as calculated in 4.14 is nothing but the bandwidth resolution of the map. Similarly, $V_{11}[x, \mu(x)]$ can be viewed as the point-spread resolution of the cortical map. These two resolutions are linked to the cortical magnification factor via equation 4.11.

It is interesting to note that equations 4.9 and 4.10 yield

$$V_{11}[x, \mu(x)]\, V_{22}[x, \mu(x)] - V_{12}[x, \mu(x)] V_{21}[x, \mu(x)] = 0 \tag{4.16}$$

or that the graph $z = V(x, y)$ is parabolic at its peak points $[x, \mu(x)]$. This, along with the restrictions that $V_{11}[x, \mu(x)]$ and $V_{22}[x, \mu(x)]$ are both negative, constitutes the conditions of a continuous, homogeneous map.

Similarly, the occurrence of a discontinuous, columnar-like structure of the map corresponds to having isolated peak points at which the elliptic graph $z = V(x, y)$ attains the maximal value:

$$V_{11}(x_0, y_0)\, V_{22}(x_0, y_0) - V_{12}^2\,(x_0, y_0) < 0 \tag{4.17}$$

5 Rule of Inverse Magnification

The continuous and homogeneous cortical map as discussed in the previous section is a topographic map that uniformly associates a cortical location with each stimulus. The retinotopic map is an example where the stimulus parameter being mapped is the position in the frontal-parallel visual field. The somatotopic map is another example where the stimulus parameter is the location in the skin surface. In both cases, the receptive field size (RF) is a synonym of our previously defined "bandwidth" of a cortical neuron, be it an effective area of the visual space or an effective patch of the skin surfaces. In terms of the square-root of areal measurement, RF is simply $(-V_{22})^{-1/2}$, or

$$\mathrm{RF}(x) = \sqrt{\frac{a_2\,[\mu'(x)]^2 + b_2}{-G(V_{\max})}} \tag{5.1}$$

If $b_2 = 0$ or b_2 is very small (for a discussion, see Appendix), then 5.1 becomes

$$\mathrm{RF}(x) \propto \mu'(x) = \mathrm{CMF}^{-1}(x) \tag{5.2}$$

or finally

$$\mathrm{RF}(x) \cdot \mathrm{CMF}(x) = \textit{constant} \tag{5.3}$$

The product of the receptive field size and the cortical magnification factor is nothing but the size (in terms of cortical distance) that a cell receives its total input and would be activated. From equation 4.11, this product also equals $(-V_{11})^{-1/2}$, the "point-image" of a stimulus (McIlwain 1986). That the total cortical distance to influence (drive) a cell and the overall size of cortical point-image are constant imply that the cortex is uniform in neuronal connections to implement its computations.

The physiological uniformity of the cortex has long been observed experimentally. In monkey striate visual cortex, Hubel and Weisel (1974) reported that, despite the large scattering of cells' receptive field sizes of cells at each eccentricity (now believed to correspond to functionally different cell groups), the average size (in square-root of areal measurement) is roughly proportional to the inverse of the cortical magnification factor. This inverse magnification rule was also revealed in monkey somatosensory cortex (Sur *et al.* 1980), and was demonstrated most convincingly in the studies of reorganization of the hand-digit representation under

various surgical and behavioral manipulations (Jenkins *et al.* 1990). This remarkable relationship $\mathrm{RF}(x) \cdot \mathrm{CMF}(x) = \textit{constant}$ is compatible with the anatomical uniformity of the cortex, in particular the uniform dendritic field sizes (which is the anatomical substrate of receptive field) of each cell type across the cortex.

In Grajski and Merzenich (1990), the self-organization of somatotopic map was simulated using a three-layered network essentially the same as the one being discussed here. These authors demonstrated that the general principles of Hebbian rule, lateral inhibition, and adiabatic approximation are sufficient to account for the inverse relationship between the receptive field size and the cortical magnification factor. A similar result was also obtained by a probabilistic analysis of the Kohonen-type network (Obermayer *et al.* 1990). These empirical and computer studies are all consistent with our analytical result, and therefore nicely complement each other in helping us understand the principles as well as properties such as the inverse magnification rule of self-organizing cortical maps.

6 Conclusions

The analytic power of this approach toward a unified description of self-organization of cortical maps, as developed by Amari and colleagues and extended here, greatly facilitates mathematical appreciations of the dynamics as well as the equilibrium behavior of the neural system. The present formulation embodies the general scheme of layered neural networks with feedforward (thalamocortical) excitations and lateral (intracortical) connections, and takes into account features such as the autocorrelation in the stimulus-driven activities of the thalamic afferents and the Hebbian rule of synaptic modification. The magnification and the resolution of the map are derived analytically to allow comparisons with experimental data. In particular, the linear relationship between the receptive field size and the inverse cortical magnification factor (namely the inverse magnification rule) as derived under this formulation is consistent with both experimental observations and results from computer simulations.

Appendix A

We discuss the condition $b_2 = 0$ in this appendix. From equation 3.8,

$$b_2 = \frac{1}{2} \int_{-\infty}^{\infty} \kappa(t)\, t^2 \, \mathrm{d}t \tag{A.1}$$

According to equations 2.5 and 3.2,

$$\begin{aligned}\kappa(y-y') &= \int a(y''-y)\,a(y''-y')\,\mathrm{d}y'' \\ &= \int a(y'')\,a[y''+(y-y')]\,\mathrm{d}y'' \end{aligned} \tag{A.2}$$

which is simply the autocorrelation operation

$$\kappa(t) = \int_{-\infty}^{\infty} a(t')\,a(t'+t)\,\mathrm{d}t' \tag{A.3}$$

So,[2]

$$\begin{aligned} b_2 &= \frac{1}{2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} t^2\,a(t')\,a(t+t')\,\mathrm{d}t'\,\mathrm{d}t \\ &= \frac{1}{2}\int_{-\infty}^{\infty} a(t')\,\mathrm{d}t' \int_{-\infty}^{\infty} (y-t')^2\,a(y)\,\mathrm{d}y \end{aligned} \tag{A.4}$$

where we put $y = t + t'$. Denoting

$$A_k = \int_{-\infty}^{\infty} a(y)\,y^k\,\mathrm{d}y\,, \qquad k = 0,1,2 \tag{A.5}$$

we have

$$\begin{aligned} b_2 &= \frac{1}{2}\int_{-\infty}^{\infty} a(t')\,\mathrm{d}t' \int_{-\infty}^{\infty} (\,y^2 - 2\,y\,t' + t'^2\,)\,a(y)\,\mathrm{d}y \\ &= \frac{1}{2}\left(A_0\,A_2 - 2\,A_1^2 + A_2\,A_0\right) \\ &= A_0\,A_2 - A_1^2 \end{aligned} \tag{A.6}$$

For an even-symmetric $a(t)$, $A_1 = 0$. We finally obtain

$$b_2 = \left[\int_{-\infty}^{\infty} a(t)\,\mathrm{d}t\right]\left[\int_{-\infty}^{\infty} t^2\,a(t)\,\mathrm{d}t\right] \tag{A.7}$$

Therefore, the condition $b_2 = 0$ implies that the integral of $a(t)$, either weighted by t^2 or not, should be zero. Physiologically, the ON/OFF regions in the response fields of the thalamic (geniculate) neurons must be balanced in its total excitation and total inhibition.

Acknowledgments

This work was supported by PHS Grant EY–00014. The author especially thanks Dr. S. Amari for his helpful comments and for simplifying proofs that have enhanced this manuscript. Thanks are also extended to Drs. K. K. De Valois and R. L. De Valois for their generous support and constant encouragement.

[2]The following simplified proof is provided by Dr. S. Amari, and replaces a *previous* proof using Fourier transform techniques.

References

Amari, S. 1983. Field theory of self-organizing neural nets. *IEEE Trans. SMC* **SMC-13**, 741–748.

Amari, S. 1989. Dynamical study of formation of cortical maps. In *Dynamic Interactions in Neural Networks: Models and Data*, M. A. Arbib and S. Amari, eds., pp. 15–34. Springer-Verlag, New York.

Chernjavsky, A., and Moody, J. 1990. Spontaneous development of modularity in simple cortical models. *Neural Comp.* **2**, 334–350.

Chester, C. R. 1971. *Techniques in Partial Differential Equations.* McGraw-Hill, New York.

Grajski, K. A., and Merzenich, M. M. 1990. Hebb-type dynamics is sufficient to account for the inverse magnification rule in cortical somatotopy. *Neural Comp.* **2**, 71–84.

Hubel, D. H., and Wiesel, T. N. 1974. Uniformity of monkey striate cortex: A parallel relationship between field size, scatter, and magnification factor. *J. Comp. Neurol.* **158**, 295–306.

Jenkins, W. M., Merzenich, M. M., Ochs, M. T., Allard, T., and Guíc-Robles, E. 1990. Functional reorganization of primary somatosensory cortex in adult owl monkeys after behaviorally controlled tactile stimulation. *J. Neurophys.* **63**, 82–104.

Kohonen, T. 1982. Self-organized formation of topologically correct feature maps. *Biol. Cybern.* **43**, 59–69.

Knudsen, E. I., du Lac, S., and Esterly, S. D. 1987. Computational maps in the brain. *Annu. Rev. Neurosci.* **10**, 41–65.

Linsker, R. 1986. From basic network principles to neural architecture. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7508–7512, 8390–8394, 8779–8783.

Malsburg, Ch. von der 1973. Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik* **14**, 85–100.

McIlwain, J. T. 1986. Point images in the visual system: New interest in an old idea. *Trends Neurosci.* **9**, 354–358.

Miller, K. D., Keller, J. B., and Stryker, M. P. 1989. Ocular dominance column development: Analysis and simulation. *Science* **245**, 605–615.

Obermayer, K., Ritter, H., and Schulten, K. 1990. A neural network model for the formation of topographic maps in the CNS: Development of receptive fields. In *Proc. Int. Joint Conf. Neural Networks (IJCNN'90), San Diego*, **II**, 423–429.

Sur, M., Merzenich, M. M., and Kaas, J. H. 1980. Magnification, receptive-field area, and "hypercolumn" size in area 3b and 1 of somatosensory cortex in owl monkeys. *J. Neurophys.* **44**, 295–311.

Swindale, N. V. 1980. A model for the formation of ocular dominance stripes. *Proc. R. Soc. London Ser. B.* **208** 243–264.

Takeuchi, A., and Amari, S. 1979. Formation of topographic maps and columnar microstructures in nerve fields. *Biol. Cybern.* **35**, 63–72.

Zhang, J. 1990. Dynamical self-organization and formation of cortical maps. In *Proc. Int. Joint Conf. Neural Networks (IJCNN'90), San Diego*, **III**, 487–492.

Received 23 July 90; accepted 12 November 90.