Title Page

Title: Mathematical Analysis of a Correlation-Based Model for Orientation Map Formation

Author: Tadashi Yamazaki
Affiliation:
c/o Dr. Osamu Watanabe,
Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology.
2-12-1 Meguro, Tokyo 152-8552, JAPAN.

Acknowledgements:
The author would like to thank Dr. Osamu Watanabe at Tokyo Institute of Technology for supervising this work and for many helpful discussions. He also thanks Dr. Shigeru Tanaka at the Institute of Physical and Chemical Research (RIKEN) and all members of his laboratory for their valuable discussions and suggestions.

Correspondence Address:
Tadashi Yamazaki,
c/o Dr. Osamu Watanabe,
Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology.
2-12-1 Meguro, Tokyo 152-8552, JAPAN.
phone: +81-3-5734-2688(voice), +81-3-5734-3204(FAX)
email: tyam@is.titech.ac.jp

Running Title: Mathematical Analysis of a Correlation-Based Model
Preferred Section: 3. Mathematical and Computational Analysis

Abstract and Keywords

Abstract. We consider a correlation-based model for orientation map formation proposed by Miller (1994) and study the formation mathematically. We perform the Fourier transform and compute the principal component of the model. With our analysis, the roles of functions are clarified and the result indicates that the developed orientation maps have the following properties. (1) It develops oriented receptive fields; (2) Preferred orientations smoothly change on the cortical surface; (3) The periodicity of preferred orientations does not appear; (4) The periodicity of phases appears; (5) Singular points appear irregularly on the cortical surface. Our analytical results are justified by computer simulations.

Keywords. Miller's model, self-organization, development, orientation columns, oriented receptive field, mathematical analysis, correlation-based model

Text

1 Introduction

Since von der Malsburg (1973) developed a mathematical model for the orientation map formation,a large number of models have been proposed, and some of them have been widely reviewed and discussed in depth (Erwin, Obermayer, & Shulten, 1995; Swindale, 1996). These discussions, however, are based on numerical simulations. There have been few attempts to consider the formation mathematically. With mathematical analyses, the roles of functions are clarified and then the properties of developed orientation maps by the model are derived. In this paper, we consider a correlation-based model proposed by Miller (1994) and analyze it mathematically.

Although several mathematical analyses of a correlation-based model have been performed to date, the formation has not been investigated sufficiently. Shouval and Cooper (1996) studied a similar correlation based model. They defined an energy function of the model and discussed the structures of receptive fields and orientation columns corresponding to the lowest energy. However, the retinotopy has not been considered in their analysis, that is, in their model, all cells in the visual cortex have the same receptive field. Wimbauer, Gerstner, & van Hemmen (1998) computed the shapes of receptive fields possibly developed by the principal component analysis; however, they have not mentioned how these receptive fields are arranged on a two-dimensional cortical surface. That is, the development of orientation columns has not been analyzed. To clarify these points, we take another approach and give some explanations for the above problems.

We consider some simplifications in a way similar to those reported in Wimbauer, Gerstner, & van Hemmen (1998), and then we perform the Fourier transform and compute the principal component (i.e., the fastest growing pattern) of the simplified model. By our analysis, we clarify the roles of functions as follows. (1) A Mexican-hat-type correlation function determines the spatial frequency of receptive fields; (2) The combination of a Gaussian arbor function and a Gaussian intracortical function develops oriented receptive fields; (3) The intracortical function develops stripes of ON/OFF dominance similar to ocular dominance columns on a layer on which each cell has its receptive field if the range is sufficiently wide; (4) Preferred orientations and phases of receptive fields are determined by the stripes within the receptive fields. These observations indicate that this model develops orientation maps with the following properties. (1) It develops oriented receptive fields; (2) Preferred orientations smoothly change on the cortical surface; (3) The periodicity of orientations on the cortical surface does not appear; (4) The periodicity of phases appears; (5) Singular points appear irregularly on the cortical surface. Finally, we confirm our results by computer simulations.

This paper is organized as follows. In section 2, we review Miller's model and consider some simplifications. Section 3 contains our main results. Section 4 is devoted to computer simulations.

2 Miller's Model and Mathematical Framework

The model consists of two neuronal layers (Fig. 1). The pre- and postsynaptic layers correspond to the LGN and the visual cortex, respectively, and cells in each layer are arranged on a two-dimensional grid. Two types of synaptic connections, 'ON' type and 'OFF' type, are assumed, and each corresponds to the synaptic connection from an ON-center LGN cell and an OFF-center LGN cell. Both layers conserve *the retinotopy*, that is, neighboring cells in the cortex and in the LGN correspond to neighboring positions on the retina. Thus, the neighborhood relationships of the visual input are conserved. The initial synaptic weights are assigned uniformly at random.

The modification rule of synaptic weights at the t th step is

$$\Delta w^{\mathrm{ON}(t)}(\mathbf{x},\alpha) = \lambda A(\mathbf{x}-\alpha)\sum_{\mathbf{y},\beta} I(\mathbf{x}-\mathbf{y}) \left[C^{\text{ON-ON}}(\alpha-\beta)w^{\mathrm{ON}(t)}(\mathbf{y},\beta) + C^{\text{OFF-OFF}}(\alpha-\beta)w^{\mathrm{OFF}(t)}(\mathbf{y},\beta)\right]. \tag{1}$$

A, I, C, and w are respectively the arbor function, the intracortical interaction function, the correlation functions, and the synaptic weights. The vectors $\mathbf{x}, \mathbf{y}$ denote two-dimensional postsynaptic positions, and α, β indicate two-dimensional presynaptic positions. The equation for OFF synaptic weights is identical after exchange of 'ON' and 'OFF' everywhere. A, I, and C are given by the following equations.

$$\begin{aligned} A(\mathbf{x}) &\stackrel{\text{def}}{=} \exp\left(-|\mathbf{x}|^2/2\sigma_A^2\right), \quad I(\mathbf{x}) \stackrel{\text{def}}{=} \exp\left(-|\mathbf{x}|^2/2\sigma_I^2\right), \\ C(\alpha) &\stackrel{\text{def}}{=} C_1 \exp\left(-|\alpha|^2/2\sigma_{C1}^2\right) - C_2 \exp\left(-|\alpha|^2/2\sigma_{C2}^2\right), \end{aligned} \tag{2}$$

$$\begin{aligned} C^{\text{ON-ON}}(\alpha) &\stackrel{\text{def}}{=} C(\alpha) \quad (\text{resp., } C^{\text{OFF-OFF}}), \\ C^{\text{ON-OFF}}(\alpha) &\stackrel{\text{def}}{=} -\varepsilon C(\alpha) \quad (\text{resp., } C^{\text{OFF-ON}}). \end{aligned}$$

In addition to these, this model includes an upper and a lower bound for each synaptic weight and the normalization to keep the sum of synaptic weights for each postsynaptic cell constant.

We consider some simplifications. First, we consider the continuous version of the modification rule. Second, we omit the lower and upper bounds and the normalization. We assume that the dynamics of the model is determined in the early stages of the self-organization, i.e., before synaptic weights do not reach the saturation limits. On the other hand, Miller and MacKay (1994) have shown that the subtractive normalization does not change the principal component. Thus, without the normalization, the dynamics of the model may not change. Third, we introduce $w^{(t)}(\mathbf{x},\alpha)$ as the difference between $w^{\mathrm{ON}(t)}(\mathbf{x},\alpha)$ and $w^{\mathrm{OFF}(t)}(\mathbf{x},\alpha)$ as $w^{(t)}(\mathbf{x},\alpha) \stackrel{\text{def}}{=} w^{\mathrm{ON}(t)}(\mathbf{x},\alpha) - w^{\mathrm{OFF}(t)}(\mathbf{x},\alpha)$. Fourth, for each $w^{(t)}(\mathbf{x},\alpha)$, we consider the position α as the relative position $\Delta\alpha$ from the position $\mathbf{x}$. We define $w^{(t)}(\mathbf{x},\Delta\alpha)$ as $w^{(t)}(\mathbf{x},\Delta\alpha) \stackrel{\text{def}}{=} w^{(t)}(\mathbf{x},\alpha-\mathbf{x})$. From the above simplifications, it follows that

$$\Delta w^{(t)}(\mathbf{x},\Delta\alpha) = A(\Delta\alpha)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(\mathbf{x}-\mathbf{y})C((\mathbf{x}-\mathbf{y})+(\Delta\alpha-\Delta\beta))w^{(t)}(\mathbf{y},\Delta\beta)d\Delta\beta d\mathbf{y}. \tag{3}$$

In this paper, we analyze the dynamics of the modification rule defined by Eqn.(3).

Mathematical Framework

Since Eqn. (3) is a linear equation, it can be solved by expanding $w^{(t)}(\mathbf{x}, \Delta\alpha)$ into some independent bases. One obtains

$$w^{(t)}(\mathbf{x}, \Delta\alpha) = \sum_{k=0}^{\infty} a_k \exp(f_k t) p_k(\mathbf{x}, \Delta\alpha),$$

where a_k and f_k are constants while $p_k(\mathbf{x}, \Delta\alpha)$ is a basis. Since $\lim_{t\to\infty} |\exp(f_k t) - \exp(f_{k'} t)| = \infty$ for some $f_k \neq f_{k'}$, $p_{k_*}(\mathbf{x}, \Delta\alpha)$ such that $k_* = \arg\max_k f_k$ will dominate $w^{(t)}(\mathbf{x}, \Delta\alpha)$. Therefore, our goal is to find such decomposition and $p_{k_*}(\mathbf{x}, \Delta\alpha)$ which we call *the dominant component.*

In addition, we assume that $\Delta w^{(t)}(\mathbf{x}, \Delta\alpha)$ decomposes into coefficients c_k of t and bases $p_k(\mathbf{x}, \Delta\alpha)$.

$$\Delta w^{(t)}(\mathbf{x}, \Delta\alpha) = \sum_{k=0}^{\infty} c_k(t) p_k(\mathbf{x}, \Delta\alpha).$$

Then, $w^{(t+1)}(\mathbf{x}, \Delta\alpha)$ can be computed as follows.

$$w^{(t+1)}(\mathbf{x}, \Delta\alpha) = \Delta w^{(t)}(\mathbf{x}, \Delta\alpha) + w^{(t)}(\mathbf{x}, \Delta\alpha) = \cdots = \sum_{k=0}^{\infty} \sum_{s=0}^{t} c_k(s) p_k(\mathbf{x}, \Delta\alpha) + w^{(0)}(\mathbf{x}, \Delta\alpha).$$

In a similar way, let $\Delta^2 w^{(t)}(\mathbf{x}, \Delta\alpha)$ be $\Delta w^{(t+1)}(\mathbf{x}, \Delta\alpha) - \Delta w^{(t)}(\mathbf{x}, \Delta\alpha)$ and assume that $\Delta^2 w^{(t)}(\mathbf{x}, \Delta\alpha)$ decomposes into coefficients d_k and bases $q_k(\mathbf{x}, \Delta\alpha)$.

$$\Delta^2 w^{(t)}(\mathbf{x}, \Delta\alpha) = \sum_{k=0}^{\infty} d_k(t) q_k(\mathbf{x}, \Delta\alpha).$$

It follows that

$$\Delta w^{(t+1)}(\mathbf{x}, \Delta\alpha) = \sum_{k=0}^{\infty} \sum_{s=0}^{t} d_k(s) q_k(\mathbf{x}, \Delta\alpha) + \Delta w^{(0)}(\mathbf{x}, \Delta\alpha).$$

Then,

$$\begin{aligned}
w^{(t+1)}(\mathbf{x}, \Delta\alpha) &= \Delta w^{(t)}(\mathbf{x}, \Delta\alpha) + w^{(t)}(\mathbf{x}, \Delta\alpha) \\
&= \sum_{k=0}^{\infty} \sum_{s=0}^{t-1} d_k(s) q_k(\mathbf{x}, \Delta\alpha) + \Delta w^{(0)}(\mathbf{x}, \Delta\alpha) + w^{(t)}(\mathbf{x}, \Delta\alpha) \\
&= \sum_{k=0}^{\infty} \sum_{s=0}^{t-1} d_k(s) q_k(\mathbf{x}, \Delta\alpha) + \Delta w^{(0)}(\mathbf{x}, \Delta\alpha) \\
&\quad + \sum_{k=0}^{\infty} \sum_{s=0}^{t-2} d_k(s) q_k(\mathbf{x}, \Delta\alpha) + \Delta w^{(0)}(\mathbf{x}, \Delta\alpha) + w^{(t-1)}(\mathbf{x}, \Delta\alpha) \\
&\;\vdots \\
&= \sum_{k=0}^{\infty} \sum_{i=1}^{t} \sum_{s=0}^{t-i} d_k(s) q_k(\mathbf{x}, \Delta\alpha) + t \Delta w^{(0)}(\mathbf{x}, \Delta\alpha).
\end{aligned}$$

Thus, if $d_{k_0}(t) > d_{k_1}(t) > d_{k_2}(t) > \cdots$ holds for some $k_0, k_1, \cdots$ and for any t, and if each $d_{k_i}(t)$ increases exponentially as t increases, it follows that the structure of $w^{(t+1)}(\mathbf{x}, \Delta\alpha)$ is conquered by $q_{k_0}(\mathbf{x}, \Delta\alpha)$. Therefore, our goal now is to find such decomposition of $\Delta^2 w^{(t)}(\mathbf{x}, \Delta\alpha)$ and $q_{k_0}(\mathbf{x}, \Delta\alpha)$ which we also call the dominant component.

3 Analytical Results

This section contains our main results. First we perform the Fourier transform of Eqn. (3) and compute the decomposition of $\Delta^2 w^{(t)}(\mathbf{x}, \Delta\alpha)$ in the Fourier space, that is, the decomposition of $\Delta^2 \widehat{w}^{(t)}(\mathbf{u}, \mathbf{n})$, where the symbol $\widehat{\ }$ denotes the Fourier transform. Then we obtain the dominant component of $\Delta^2 \widehat{w}^{(t)}(\mathbf{u}, \mathbf{n})$. After that, by computing the Fourier inverse transform, we obtain the dominant component of $\Delta^2 w^{(t)}(\mathbf{x}, \Delta\alpha)$, that is, the dominant component of $w^{(t)}(\mathbf{x}, \Delta\alpha)$. Finally, we give an interpretation of how the dominant component describes the formation of orientation maps.

The Fourier transform of Eqn. (3) is as follows.

$$\Delta \widehat{w}^{(t)}(\mathbf{u}, \mathbf{n}) = \widehat{I}(\mathbf{u} - \mathbf{n}) \int_{-\infty}^{\infty} \widehat{A}(\mathbf{n} - \mathbf{m}) \widehat{C}(\mathbf{m}) \widehat{w}^{(t)}(\mathbf{u}, \mathbf{m}) d\mathbf{m}, \tag{4}$$

where

$$\widehat{A}(\mathbf{l}) = \sigma_A^2 e^{-\sigma_A^2 |\mathbf{l}|^2/2}, \quad \widehat{I}(\mathbf{l}) = \sigma_I^2 e^{-\sigma_I^2 |\mathbf{l}|^2/2},$$
$$\widehat{C}(\mathbf{l}) = C_1 \sigma_{C1}^2 e^{-\sigma_{C1}^2 |\mathbf{l}|^2/2} - C_2 \sigma_{C2}^2 e^{-\sigma_{C2}^2 \mathbf{l}^2/2}.$$

It is known that the Fourier transform of a DOG function takes the form of bi-modal (Fig. 2) depending on the choice of parameters. Since $C(\alpha)$ is a DOG function, we assume $\widehat{C}(\mathbf{l})$ is a bi-modal function which shows its peaks at $\sigma \stackrel{\text{def}}{=} |\mathbf{l}|$. Then, the integral domain could be restricted to the circle of radius σ because only $\widehat{w}^{(t)}(\mathbf{u}, \mathbf{n})$ such that $|\mathbf{n}| = \sigma$ are effective. Thus, we have

$$\Delta \widehat{w}^{(t)}(\mathbf{u}, \mathbf{n}) \approx \widehat{I}(\mathbf{u} - \mathbf{n}) \int_{|\mathbf{m}|=\sigma} \widehat{A}(\mathbf{n} - \mathbf{m}) \widehat{w}^{(t)}(\mathbf{u}, \mathbf{m}) d\mathbf{m}. \tag{5}$$

Since $|\mathbf{n}|$ and $|\mathbf{m}|$ are σ, we represent Eqn. (5) in polar coordinates. We denote $\mathbf{u}$,$\mathbf{n}$,$\mathbf{m}$ as

$$\mathbf{u} \stackrel{\text{def}}{=} u e^{i\theta_u}, \quad \mathbf{n} \stackrel{\text{def}}{=} \sigma e^{i\theta_n}, \quad \text{and} \quad \mathbf{m} \stackrel{\text{def}}{=} \sigma e^{i\theta_m},$$

then

$$|\mathbf{u} - \mathbf{n}| = \sqrt{u^2 + \sigma^2 - 2u\sigma\cos(\theta_u - \theta_n)}, \quad \text{and}$$
$$|\mathbf{n} - \mathbf{m}| = \sigma\sqrt{2(1 - \cos(\theta_n - \theta_m))}.$$

Let $\widetilde{I}(u, \theta)$, $\widetilde{A}(\theta)$, and $\widetilde{w}^{(t)}(u, \theta_u, \theta_n)$ be as follows.

$$\widetilde{I}(u, \theta) \stackrel{\text{def}}{=} \widehat{I}\left(\sqrt{u^2 + \sigma^2 - 2u\sigma\cos\theta}\right),$$
$$\widetilde{A}(\theta) \stackrel{\text{def}}{=} \widehat{A}\left(\sigma\sqrt{2(1 - \cos\theta)}\right), \quad \text{and}$$
$$\widetilde{w}^{(t)}(u, \theta_u, \theta_n) \stackrel{\text{def}}{=} \widehat{w}^{(t)}(\mathbf{u}, \mathbf{n}).$$

Thus, Eqn. (5) is written as

$$\Delta\widetilde{w}^{(t)}(u,\theta_u,\theta_n) = \widetilde{I}(u,\theta_u-\theta_n)\int_0^{2\pi}\widetilde{A}(\theta_n-\theta_m)\widetilde{w}^{(t)}(u,\theta_u,\theta_m)d\theta_m. \tag{6}$$

Defining $\Delta\theta_n$ and $\Delta\theta_m$ respectively as $\Delta\theta_n \overset{\text{def}}{=} \theta_u-\theta_n$ and $\Delta\theta_m \overset{\text{def}}{=} \theta_u-\theta_m$, and rewriting $\Delta\widetilde{w}^{(t)}(u,\theta_u,\theta_u-\Delta\theta_n)$ as $\Delta\widetilde{w}_*^{(t)}(u,\theta_u,\Delta\theta_n)$, we obtain

$$\Delta\widetilde{w}_*^{(t)}(u,\theta_u,\Delta\theta_n) = \widetilde{I}(u,\Delta\theta_n)\int_0^{2\pi}\widetilde{A}(\Delta\theta_n-\Delta\theta_m)\widetilde{w}_*^{(t)}(u,\theta_u,\Delta\theta_m)d\Delta\theta_m.$$

By the Fourier series expansion, we can restate it as follows.

$$\Delta\widetilde{w}_*^{(t)}(u,\theta_u,\Delta\theta_n) = \widetilde{I}(u,\Delta\theta_n)\sum_{k=0}^{\infty}\widehat{\widetilde{A}}(k)\widehat{\widetilde{w}_*}^{(t)}(u,\theta_u,k)e^{ik\Delta\theta_n}, \tag{7}$$

where

$$\widehat{\widetilde{A}}(k) = 4\pi^2\sigma_A^2 e^{-\sigma_A^2\sigma^2} I_k(\sigma_A^2\sigma^2),$$

and I_k is the modified Bessel function of the k th order.

We compute $\Delta^2\widetilde{w}_*^{(t)}(u,\theta_u,\Delta\theta_n)$ and decompose it into a series of $\widetilde{I}(u,\Delta\theta_n)e^{ik\Delta\theta_n}$ such that $k=0,1,2,\cdots$. We first prove that its 0th component, i.e., $\widetilde{I}(u,\Delta\theta_n)$, grows much faster than the others.

Since $\Delta\widetilde{w}^{(t)}(u,\theta_u,\Delta\theta_n)$ is decomposed as Eqn.(7), we can state $\Delta\widetilde{w}_*^{(t+1)}(u,\theta_u,\Delta\theta_n)$ as

$$\begin{aligned}
\Delta\widetilde{w}_*^{(t+1)}(u,\theta_u,\Delta\theta_n) &= \widetilde{I}(u,\Delta\theta_n)\sum_{k=0}^{\infty}\widehat{\widetilde{A}}(k)\widehat{\widetilde{w}}_*^{(t+1)}(u,\theta_u,k)e^{ik\Delta\theta_n} \\
&= \widetilde{I}(u,\Delta\theta_n)\sum_{k=0}^{\infty}\widehat{\widetilde{A}}(k)\left(\widehat{\widetilde{w}}_*^{(t)}(u,\theta_u,k)+\Delta\widehat{\widetilde{w}}_*^{(t)}(u,\theta_u,k)\right)e^{ik\Delta\theta_n} \\
&= \Delta\widetilde{w}_*^{(t)}(u,\theta_u,\Delta\theta_n)+\widetilde{I}(u,\Delta\theta_n)\sum_{k=0}^{\infty}\widehat{\widetilde{A}}(k)\Delta\widehat{\widetilde{w}}_*^{(t)}(u,\theta_u,k)e^{ik\Delta\theta_n}.
\end{aligned}$$

Since

$$\begin{aligned}
\Delta\widehat{\widetilde{w}}_*^{(t)}(u,\theta_u,k) &= \int_0^{2\pi}\Delta\widetilde{w}_*^{(t)}(u,\theta_u,\theta)e^{-ik\theta}d\theta \\
&= \int_0^{2\pi}\widetilde{I}(u,\theta)\sum_{l=0}^{\infty}\widehat{\widetilde{A}}(l)\widehat{\widetilde{w}}_*^{(t)}(u,\theta_u,l)e^{il\theta}e^{-ik\theta}d\theta \\
&= \int_0^{2\pi}\widetilde{I}(u,\theta)\sum_{l=0}^{\infty}\overline{\widetilde{A}}(l)\overline{\widetilde{w}}_*^{(t)}(u,\theta_u,l)e^{-il\theta}e^{-ik\theta}d\theta \\
&= \sum_{l=0}^{\infty}\overline{\widetilde{A}}(l)\overline{\widetilde{w}}_*^{(t)}(u,\theta_u,l)\int_0^{2\pi}\widetilde{I}(u,\theta)e^{-il\theta}e^{-ik\theta}d\theta \\
&= \sum_{l=0}^{\infty}\overline{\widetilde{A}}(l)\overline{\widetilde{w}}_*^{(t)}(u,\theta_u,l)\widehat{\widetilde{I}}(u,k+l),
\end{aligned}$$

where $\overline{f}(l) \stackrel{\text{def}}{=} \int_0^{2\pi} f(x)e^{ilx}dx$ and

$$\begin{aligned}\overline{\widehat{\widehat{A}}}(k) = \widehat{\widehat{A}}(-k) = \widehat{\widehat{A}}(k) = 4\pi^2\sigma_A^2 e^{-\sigma_A^2\sigma^2} I_k(\sigma_A^2\sigma^2),\\ \widehat{\widehat{I}}(u,k) = 4\pi^2\sigma_I^2 e^{-\sigma_I^2\sigma^2} I_k(\sigma_I^2 u\sigma),\end{aligned} \tag{8}$$

it follows that

$$\begin{aligned}\Delta^2\widetilde{w}_*^{(t)}(u,\theta_u,\Delta\theta_n) &\stackrel{\text{def}}{=} \Delta\widetilde{w}_*^{(t+1)}(u,\theta_u,\Delta\theta_n) - \Delta\widetilde{w}_*^{(t)}(u,\theta_u,\Delta\theta_n)\\ &= \widetilde{I}(u,\Delta\theta_n)\sum_{k=0}^{\infty}\widehat{\widehat{A}}(k)\sum_{l=0}^{\infty}\overline{\widehat{A}}(l)\overline{\widetilde{w}}_*^{(t)}(u,\theta_u,l)\widehat{\widehat{I}}(u,k+l)e^{ik\Delta\theta_n}\\ &= \sum_{k=0}^{\infty}\left[\widehat{\widehat{A}}(k)\sum_{l=0}^{\infty}\overline{\widehat{A}}(l)\overline{\widetilde{w}}_*^{(t)}(u,\theta_u,l)\widehat{\widehat{I}}(u,k+l)\right]\cdot\widetilde{I}(u,\Delta\theta_n)e^{ik\Delta\theta_n}.\end{aligned}$$

The coefficient of the kth component is $\widehat{\widehat{A}}(k)\sum_{l=0}^{\infty}\overline{\widehat{A}}(l)\overline{\widetilde{w}}_*^{(t)}(u,\theta_u,l)\widehat{\widehat{I}}(u,k+l)$. We can numerically verify that $\widehat{\widehat{A}}(0) > \widehat{\widehat{A}}(k) > 0$ and $\widehat{\widehat{I}}(u,l) > \widehat{\widehat{I}}(u,k+l) > 0$ for any $k,l>0$ and for feasible ranges of $\sigma_A^2\sigma^2$ and $\sigma_I^2 u\sigma$ of Eqn. (8) (e.g., less than 100000000). Thus we have

$$\left|\widehat{\widehat{A}}(k)\sum_{l=0}^{\infty}\overline{\widehat{A}}(l)\overline{\widetilde{w}}_*^{(t)}(u,\theta_u,l)\widehat{\widehat{I}}(u,k+l)\right| \leq \left|\widehat{\widehat{A}}(0)\sum_{l=0}^{\infty}\overline{\widehat{A}}(l)\overline{\widetilde{w}}_*^{(t)}(u,\theta_u,l)\widehat{\widehat{I}}(u,l)\right|.$$

Therefore the 0th component grows fastest.

We next show that the coefficient of the 0th component grows exponentially. This makes sure that the 0th component is the dominant component of $\Delta^2\widetilde{w}_*^{(t)}(u,\theta_u,\Delta\theta_n)$. Let $C(t)$ be the coefficient at the t th step. This can be computed as follows.

$$\begin{aligned}C(t) &= \widehat{\widehat{A}}(0)\sum_{k_t=0}^{\infty}\cdots\sum_{k_0=0}^{\infty}\widehat{\widehat{A}}(k_t)\widehat{\widehat{I}}(u,k_t+k_{t-1})\cdots\widehat{\widehat{A}}(k_1)\widehat{\widehat{I}}(u,k_1+k_0)\widehat{\widehat{A}}(k_0)\widehat{\widehat{w}}_*^{(0)}(u,\theta_u,k_0)\\ &> \widehat{\widehat{A}}(0)^t\widehat{\widehat{I}}(u,0)^t\widehat{\widehat{w}}_*^{(0)}(u,\theta_u,k_0).\end{aligned}$$

Hence the 0th component is the dominant component of $\Delta^2\widetilde{w}_*^{(t)}(u,\theta_u,\Delta\theta_n)$ because the 0th component grows faster than the others and the growing speed is exponential with respect to t. As mentioned in the previous section, the 0th component is also the dominant component of $\widetilde{w}_*^{(t)}(u,\theta_u,\Delta\theta_n)$. Therefore we obtain

$$\widetilde{w}_*^{(t)}(u,\theta_u,\Delta\theta_n) \approx C(t)\widetilde{I}(u,\Delta\theta_n). \tag{9}$$

Next we consider $C(t)$ a bit more carefully.

$$C(t) = \sum_{k_t=0}^{\infty}\cdots\sum_{k_0=0}^{\infty}\widehat{\widehat{A}}(k_t)\widehat{\widehat{I}}(u,k_t+k_{t-1})\cdots\widehat{\widehat{A}}(k_1)\widehat{\widehat{I}}(u,k_1+k_0)\widehat{\widehat{A}}(k_0)\widehat{\widehat{w}}_*^{(0)}(u,\theta_u,k_0).$$

Because $\widehat{\widetilde{A}}(k)$ and $\widehat{\widetilde{I}}(u,k)$ take the maximum values when $k=0$, we may think that $C(t)$ is dominated by $\widehat{\widetilde{A}}(0)^{t+1}\widehat{\widetilde{I}}(u,0)^t\widehat{\widetilde{w}}_*^{(0)}(u,\theta_u,0)$. Then from Eqn. (9), we obtain

$$\begin{aligned}\widetilde{w}_*^{(t)}(u,\theta_u,\Delta\theta_n) &\approx \widehat{\widetilde{A}}(0)^{t+1}\widehat{\widetilde{I}}(u,0)^t\widehat{\widetilde{w}}_*^{(0)}(u,\theta_u,0)\widetilde{I}(u,\Delta\theta_n)\\ &= \widehat{\widetilde{A}}(0)^{t+1}\widehat{\widetilde{I}}(u,0)^t\int_0^{2\pi}\widetilde{w_*}^{(0)}(u,\theta_u,\theta)d\theta\cdot\widetilde{I}(u,\Delta\theta_n)\\ &= \widehat{\widetilde{A}}(0)^{t+1}\widehat{\widetilde{I}}(u,0)^t\int_0^{2\pi}\widetilde{w}^{(0)}(u,\theta_u,\theta_u-\theta)d\theta\cdot\widetilde{I}(u,\Delta\theta_n)\\ &= \widehat{\widetilde{A}}(0)^{t+1}\widehat{\widetilde{I}}(u,0)^t\int_0^{2\pi}\widetilde{w}^{(0)}(u,\theta_u,\theta)d\theta\cdot\widetilde{I}(u,\Delta\theta_n)\\ &= \widehat{\widetilde{A}}(0)^{t+1}\widehat{\widetilde{I}}(u,0)^t\widetilde{w}^{(0)}(u,\theta_u)\widetilde{I}(u,\Delta\theta_n),\end{aligned} \tag{10}$$

where

$$\widetilde{w}^{(0)}(u,\theta_u)\stackrel{\text{def}}{=}\int_0^{2\pi}\widetilde{w}^{(0)}(u,\theta_u,\theta)d\theta.$$

Since $\widetilde{w_*}^{(t)}(u,\theta_u,\Delta\theta_n)=\widetilde{w}^{(t)}(u,\theta_u,\theta_u-\Delta\theta_n)$, we can restate Eqn.(10) as

$$\widetilde{w}^{(t)}(u,\theta_u,\theta_n)=\widehat{\widetilde{A}}(0)^{t+1}\widehat{\widetilde{I}}(u,0)^t\widetilde{w}^{(0)}(u,\theta_u)\widetilde{I}(u,\theta_u-\theta_n).$$

Therefore representing it in rectangular coordinates again, we obtain

$$\widehat{w}^{(t)}(\mathbf{u},\mathbf{n})=\widehat{\widetilde{A}}(0)^{t+1}\widehat{\widetilde{I}}(|\mathbf{u}|,0)^t\widehat{w}^{(0)}(\mathbf{u})\widehat{I}(\mathbf{u}-\mathbf{n}),$$

where

$$\widehat{w}^{(0)}(\mathbf{u})\stackrel{\text{def}}{=}\int_{|\mathbf{m}|=\sigma}\widehat{w}^{(0)}(\mathbf{u},\mathbf{m})d\mathbf{m}. \tag{11}$$

By computing the Fourier inverse transform, we have the dominant component of $w^{(t)}(\mathbf{x},\Delta\alpha)$ as follows.

$$w^{(t)}(\mathbf{x},\Delta\alpha)\approx\int_{-\infty}^{\infty}\int_{|\mathbf{n}|=\sigma}\widehat{\widetilde{A}}(0)^{t+1}\widehat{\widetilde{I}}(|\mathbf{u}|,0)^t\widehat{w}^{(0)}(\mathbf{u})\widehat{I}(\mathbf{u}-\mathbf{n})e^{i\mathbf{ux}}e^{i\mathbf{n}\Delta\alpha}d\mathbf{n}d\mathbf{u}. \tag{12}$$

Since $\widehat{\widetilde{I}}(|\mathbf{u}|,0)^t$ grows exponentially as t increases, this function extracts a certain $\mathbf{u}$ such that $|\mathbf{u}|=u_{\max}$, where

$$u_{\max}=\arg\max_u\widehat{\widetilde{I}}(u,0)=\arg\max_u e^{-\sigma_I^2(u^2+\sigma^2)/2}I_0(\sigma_I^2u\sigma).$$

This implies that we can restrict the integral domain of $\mathbf{u}$ from infinite to $|\mathbf{u}|=u_{\max}$ only. We can numerically verify that when σ_I is small $u_{\max}=0$ while as σ_I increases $u_{\max}$ converges to σ. Therefore, when σ_I is small, the integral domain could be restricted only to $|\mathbf{u}|=0$. On the other hand, when σ_I is large enough, we could restrict the integral domain only to $|\mathbf{u}|=\sigma$.

In the following we assume that σ_I is large enough. Then it follows from the above discussion,

$$
\begin{aligned}
w^{(t)}(\mathbf{x},\Delta\alpha) &\approx \int_{|\mathbf{u}|=\sigma}\int_{|\mathbf{n}|=\sigma} \widehat{\widehat{A}}(0)^{t+1}\widehat{\widehat{I}}(|\mathbf{u}|,0)^t\widehat{w}^{(0)}(\mathbf{u})\widehat{I}(\mathbf{u}-\mathbf{n})e^{i\mathbf{ux}}e^{i\mathbf{n}\Delta\alpha}d\mathbf{n}d\mathbf{u}\\
&\propto \int_{|\mathbf{u}|=\sigma}\int_{|\mathbf{n}|=\sigma} \widehat{w}^{(0)}(\mathbf{u})\widehat{I}(\mathbf{u}-\mathbf{n})e^{i\mathbf{ux}}e^{i\mathbf{n}\Delta\alpha}d\mathbf{n}d\mathbf{u}\\
&= \int_{|\mathbf{u}|=\sigma} \widehat{w}^{(0)}_{\cos}(\mathbf{u})\cos\mathbf{ux}\int_{|\mathbf{n}|=\sigma}\widehat{I}_{\cos}(\mathbf{u},\mathbf{n})\cos\mathbf{n}\Delta\alpha d\mathbf{n}d\mathbf{u}\\
&\quad+\int_{|\mathbf{u}|=\sigma} \widehat{w}^{(0)}_{\sin}(\mathbf{u})\sin\mathbf{ux}\int_{|\mathbf{n}|=\sigma}\widehat{I}_{\sin}(\mathbf{u},\mathbf{n})\sin\mathbf{n}\Delta\alpha d\mathbf{n}d\mathbf{u},
\end{aligned}
\tag{13}
$$

where

$$
\begin{aligned}
\widehat{I}_{\cos}(\mathbf{u},\mathbf{n}) &\stackrel{\text{def}}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(\mathbf{x})\cos\mathbf{ux}\cos\mathbf{nx}d\mathbf{x},\\
\widehat{I}_{\sin}(\mathbf{u},\mathbf{n}) &\stackrel{\text{def}}{=} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(\mathbf{x})\sin\mathbf{ux}\sin\mathbf{nx}d\mathbf{x}.
\end{aligned}
$$

Now, Eqn. (13) explains the detail of the formation of orientation maps, in particular, (1) how the shape of a receptive field for a cell is determined, (2) how the preferred orientation for a cell is determined, and (3) how preferred orientations and phases change on the cortical surface. We examine the structure of Eqn. (13) and discuss these points. During the discussion, the roles of functions are clarified and we observe that how those functions contribute to the formation, in other words, how the formation is constrained. Due to the constraints, orientation maps developed by this model obtain some properties. We observe those properties.

First, we consider parts of Eqn. (13) defined as follows.

$$
\mathrm{rf}_{\cos}(\mathbf{u},\Delta\alpha) \stackrel{\text{def}}{=} \int_{|\mathbf{n}|=\sigma}\widehat{I}_{\cos}(\mathbf{u},\mathbf{n})\cos\mathbf{n}\Delta\alpha d\mathbf{n},\quad \text{and}\quad \mathrm{rf}_{\sin}(\mathbf{u},\Delta\alpha) \stackrel{\text{def}}{=} \int_{|\mathbf{n}|=\sigma}\widehat{I}_{\sin}(\mathbf{u},\mathbf{n})\sin\mathbf{n}\Delta\alpha d\mathbf{n}.
\tag{14}
$$

When $\mathbf{u}$ is given, these become functions of $\Delta\alpha$. Thus these function determine the shape of receptive fields. $\mathrm{rf}_{\cos}(\mathbf{u},\Delta\alpha)$ computes a weighted sum of two-dimensional cosine waves of 180 degrees whose spatial frequency is σ, because the integral domain is limited to $|\mathbf{n}|=\sigma$. The coefficient $\widehat{I}_{\cos}(\mathbf{u},\mathbf{n})$ takes the maximum value when $\mathbf{n}=\mathbf{u}$ and decreases as $\mathbf{n}$ takes apart from $\mathbf{u}$. Thus, the shape of $\mathrm{rf}_{\cos}(\mathbf{u},\Delta\alpha)$ is tri-lobed with the orientation $\theta=\tan^{-1}(\mathbf{u})$ like top four figures in Fig. 3. In a similar way, the shape of $\mathrm{rf}_{\sin}(\mathbf{u},\Delta\alpha)$ is bi-lobed with the same orientation like bottom four figures in Fig. 3. The sharpness of the orientation tuning is determined by the range of I. Wider I makes the tuning sharper because $\widehat{I}$ becomes narrower. Thus, the preferred orientation is determined by the given $\mathbf{u}$. We then explore how $\mathbf{u}$ is given.

Since σ_I is assumed large enough, $\widehat{I}(\mathbf{u}-\mathbf{n})$ is as a very sharp function, which we may regard as the delta function. From Eqn. (13), it follows that

$$\begin{aligned} w^{(t)}(\mathbf{x}, \Delta\alpha) &\propto \int_{|\mathbf{u}|=\sigma} \widehat{w}^{(0)}(\mathbf{u}) e^{i\mathbf{u}(\mathbf{x}+\Delta\alpha)} d\mathbf{u} \\ &= \int_{|\mathbf{u}|=\sigma} \widehat{w}^{(0)}_{\cos}(\mathbf{u}) \cos \mathbf{u}(\mathbf{x}+\Delta\alpha) + \widehat{w}^{(0)}_{\sin}(\mathbf{u}) \sin \mathbf{u}(\mathbf{x}+\Delta\alpha) d\mathbf{u}. \end{aligned} \tag{15}$$

This equation is identical to the modification rule of some spectral models of ocular dominance columns (Swindale, 1980; Roger & Schwartz, 1990), and it develops stripes or blobs of the sign of $w^{(t)}(\mathbf{x}, \mathbf{0})$ on the cortical surface, where the sign of $w^{(t)}(\mathbf{x}, \mathbf{0})$ corresponds to the dominance between ON and OFF (Fig. 4). That is, for any cortical point, the dominant stripe is associated around that point. The orientation of the stripe corresponds to the preferred orientation of the receptive field of the cell on that point, namely, the above $\mathbf{u}$ is given by this stripe.

Thus for any cortical point $\mathbf{x}$, the preferred orientation of the cell on $\mathbf{x}$ becomes parallel to the orientation of the dominant stripe around $\mathbf{x}$. Moreover, on the cortical surface, the change of the stripes, i.e., the change of $w^{(t)}(\mathbf{x}, \Delta\alpha)$, moving $\mathbf{x}$ and that moving $\Delta\alpha$ in the same direction are identical, so the phase of the receptive field of a cell on $\mathbf{x}$ also becomes identical to the phase of the dominant stripe around $\mathbf{x}$ (Fig. 5).

Here we summarize the formation of orientation maps and clarify the roles of functions A, I, C to the formation. First, on the cortical surface, stripes are associated. This stripes are developed by function I if σ_I is large enough, and elongate in irregular directions because the initial weights are assigned uniformly at random. Next, for any cortical point, the cell on that point develops its receptive field whose preferred orientation and phase obeys the dominant stripe around that point, as implied by Eqn. (15). Finally, the combination of A and I works to develop oriented receptive fields. Both functions extract the DC component of $w^{(0)}(\mathbf{x}, \Delta\alpha)$ with respect of $\Delta\alpha$, what we called the 0th component, and I modifies the DC component to the direction of given $\mathbf{u}$ in the way of Eqn.(14). The spatial frequency of receptive fields is approximated by σ, which is the peak frequency of $\widehat{C}$.

These observations indicate that developed orientation maps obtain the following properties. Since stripes elongate in irregular directions, the preferred orientations smoothly change on the cortical surface, however the change is not periodic. It follows that singular points appear irregularly on the cortical surface. On the other hand, the change of phases becomes periodic.

Finally, it should be noted that this development appears similar to that obtained by Miyashita and Tanaka's thermodynamic model (Miyashita & Tanaka, 1992). In their model, all cells on the visual cortex share a layer called "synaptic terminal point", and synaptic terminals on the layer are arranged similar to ocular dominance columns or blobs. Cortical cells are assumed to exist over the layer while keeping the topography and each cell has its "receptive field" on the layer defined by clipping a small region by an isotropic function, e.g., a Gaussian function.

4 Computer Simulations

We justify our mathematical analysis by some computer simulations.

We consider 25×25 presynaptic cells and the same number of postsynaptic cells. Each postsynaptic cell receives synaptic connections from 13×13 presynaptic cells with keeping the retinotopy. We make use of Eqn. (3) as the modification rule, where parameters are taken from Erwin, Obermayer, & Shulten (1995) as follows: $\sigma_A^2 = 4.5$, $\sigma_I^2 = 16.0$, $C_1 = 1.00$, $C_2 = 1/9$, $\sigma_{C1}^2 = 0.61$, and $\sigma_{C2}^2 = 5.45$. On the other hand, we take σ_I^2 large enough as discussed in the previous section. Fig. 2 shows the shapes of $C(\alpha)$ and $\widehat{C}(\mathbf{l})$, and with this setting, σ, the peak frequency of $\widehat{C}$, is approximately 0.953. Initial synaptic weights are assigned uniformly at random from the range $[-0.4, 0.4]$, and the modification rule is applied 80 times. To avoid finite size effects, the periodic boundary condition is adopted.

Figure 6 shows the developed orientation map. As can be seen, receptive fields become oriented as shown in Fig. 3 and the change of phases of receptive fields is similar to Fig. 5.

One can calculate the spatial frequency of receptive fields. A calculation shows that the mean and the variance are respectively 0.962 and 0.03. This is a good approximation of $\sigma \approx 0.953$ (Recall that σ is the peak frequency of $\widehat{C}$).

Figure 7 shows the stripes of $w^{(t)}(\mathbf{x}, \mathbf{0})$ computed by Eqn.(3) and the preferred orientations extracted from Fig. 6. As can be observed, the preferred orientations become parallel to the stripes, and preferred orientations smoothly change on the cortical surface while the change is not periodic.

5 Conclusion

We considered a correlation-based model for the orientation map formation proposed by Miller (1994). We performed the Fourier transform and computed the principal component of the model mathematically. Then using the principal component, we studied how the oriented receptive fields and orientation columns were developed. With our analysis, we clarified the roles of functions as follows. (1) A Mexican-hat-type correlation function determines the spatial frequencies of receptive fields; (2) The combination of a Gaussian arbor function and a Gaussian intracortical function develops oriented receptive fields; (3) The intracortical function develops stripes similar to ocular dominance columns on a layer on which each cell has its receptive field if the range is sufficiently wide; (4) The preferred orientations and the phases of receptive fields for each cell are determined by the stripes within the receptive fields. These observations inditicated that this model develops orientation maps with the following properties. (1) It develops oriented receptive fields; (2) Preferred orientations smoothly change on the cortical surface; (3) The periodicity of preferred orientations does not appear; (4) The periodicity of phases appears; (5) Singular points appear irregularly on the cortical surface. Finally our analytical results were justified by computer simulations.

References

Erwin, E., Obermayer, K., Shulten, K. (1995). Models of orientation and ocular dominance columns in the visual cortex: a critical comparison. *Neural Computation*, 7, 425–468.

Malsburg, C. von der. (1973). Self-organization of orientation sensitivity cells in the striate cortex. *Kybernetik*, 14, 85–100.

Miller, K.D. (1994). A model for the development of simple cell receptive fields and the ordered arrangement of orientation columns through activity-dependent competition between on- and off-center inputs. *Journal of Neuroscience*, 14, 409–441.

Miller, K.D., MacKay, D.J.C. (1994). The role of constraints in Hebbian learning. *Neural Computation*, 6, 100-126.

Miyashita, M., Tanaka, S. (1992). A mathematical model for the self-organization of orientation columns in visual cortex. *NeuroReport*, 3, 69–72.

Rojer, A.S., Schwartz, E.L. (1990). Cat and monkey cortical columnar patters modeled by bandpass-filtered 2d white noise. *Biological Cybernetics*, 62, 381–391.

Shouval, H., Cooper, L.N. (1996). Organization of receptive fields in networks with Hebbian learning: the connection between synaptic and phenomenological models. *Biological Cybernetics*, 74, 439–447.

Swindale, N.V. (1980). A model for the formation of ocular dominance stripes. *Proceedings of the Royal Society of London*, B.208, 243–264.

Swindale, N.V. (1996). A development of topography in the visual cortex: a review of models. *Network*, 7, 161–247.

Wimbauer, S., Gerstner, W., van Hemmen, J.L. (1998). Analysis of a correlation-based model for the development of orientation-selective receptive fields in the visual cortex. *Network*, 9, 449-466.

Figures

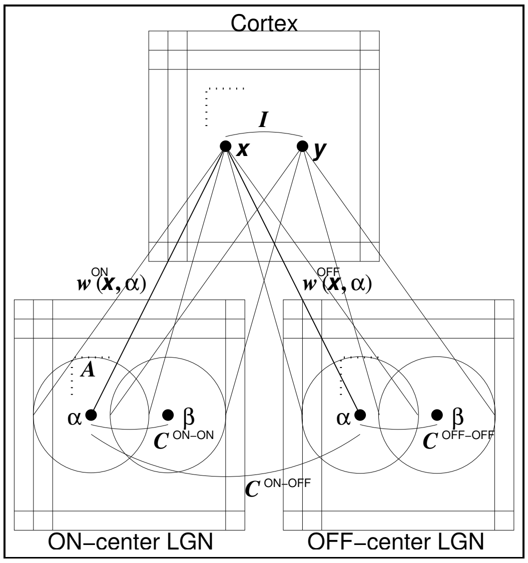

Figure 1.

Mathematical Analysis of a Correlation-Based Model for Orientation Map Formation

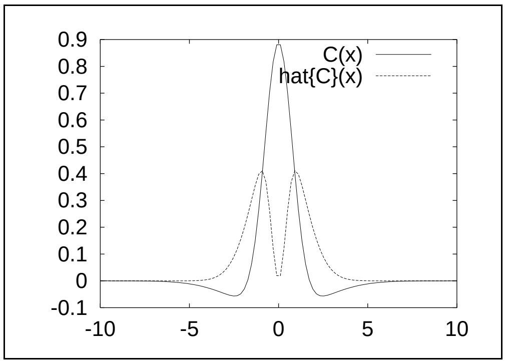

Figure 2.

Mathematical Analysis of a Correlation-Based Model for Orientation Map Formation

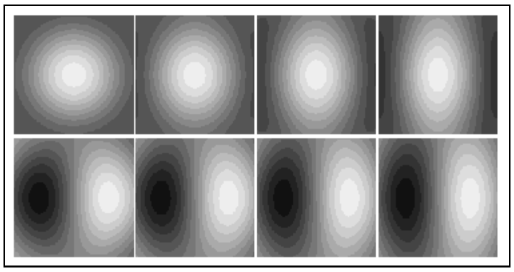

Figure 3.

Mathematical Analysis of a Correlation-Based Model for Orientation Map Formation

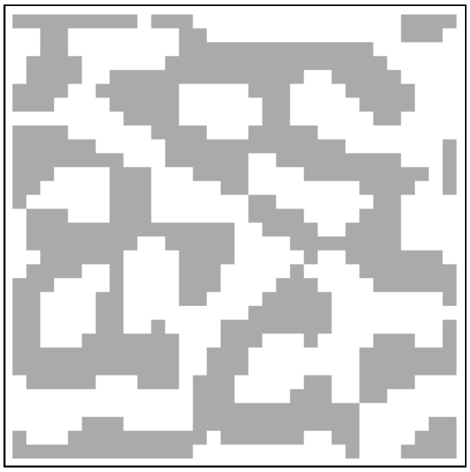

Figure 4.

Mathematical Analysis of a Correlation-Based Model for Orientation Map Formation

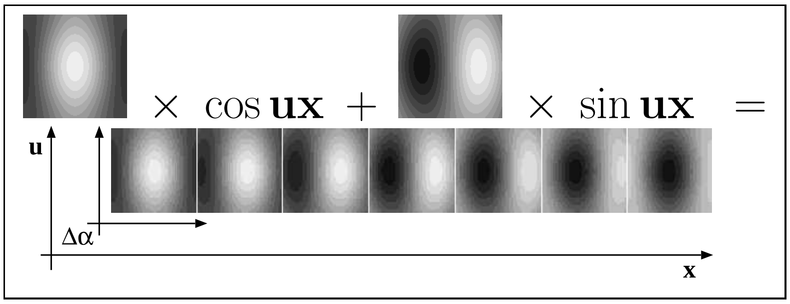

Figure 5.

Mathematical Analysis of a Correlation-Based Model for Orientation Map Formation

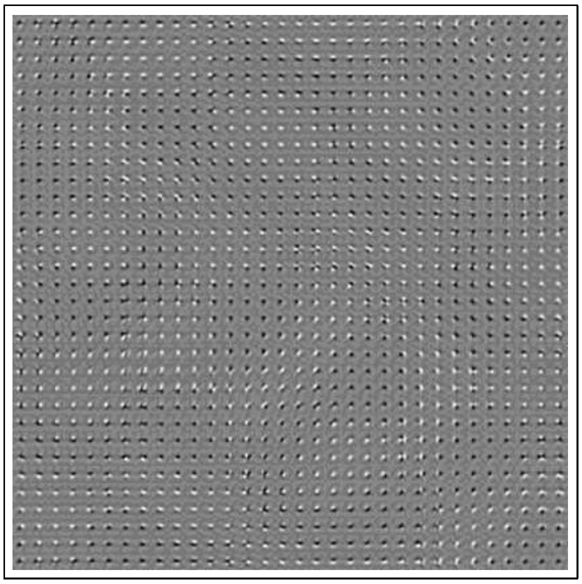

Figure 6.

Mathematical Analysis of a Correlation-Based Model for Orientation Map Formation

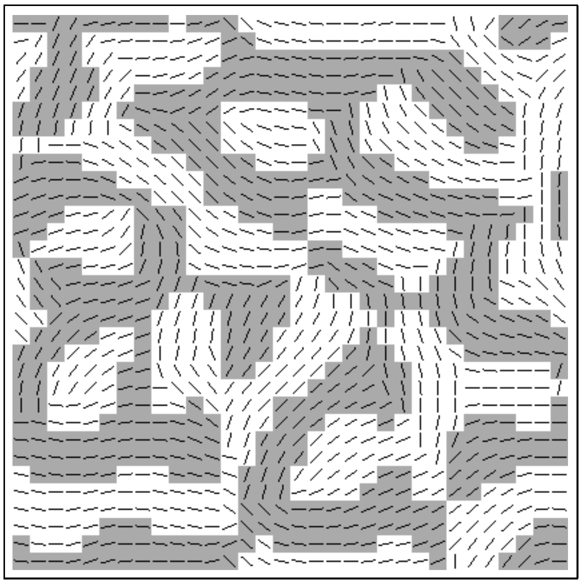

Figure 7.

Mathematical Analysis of a Correlation-Based Model for Orientation Map Formation

Figure Legends

Fig. 1. Neural network structure of Miller's model.

Fig. 2. Shapes of a DOG function $C(\alpha)$ in Eqn. (2) (real line) and its Fourier transform (dashed line) under the following parameters. $C_1 = 1.00$, $C_2 = 1/9$, $\sigma_{C1}^2 = 0.61$, and $\sigma_{C2}^2 = 5.45$.

Fig. 3. Shapes of (a) and (b) in Eqn. (14). The upper four figures are those of (a), and from left to right, σ_I^2 values are respectively 0, 10.0, 20.0, 30.0. The lower figures are those of (b) with the same parameters. $\mathbf{u}$ is taken as a vertical vector, so these figures show the orientation to the vertical direction.

Fig. 4. Signs of $w(\mathbf{x}, \mathbf{0})$ in Eqn. (15). Black and white regions represent negative and positive, respectively, which correspond to the dominance between ON and OFF.

Fig. 5. The change of the shapes of receptive fields. For a vertical unit vector $\mathbf{u}$ (i.e., stripes run vertically), the shapes of receptive fields arranged in parallel to the x-axis change, which depend on the position on the x-axis. Phases of these receptive fields also depend on the position.

Fig. 6. Orientation map obtained from a computer simulation. Receptive fields corresponding to each cell are arranged on the cortical surface.

Fig. 7. Stripes of $w^{(t)}(\mathbf{x}, \mathbf{0})$ computed by Eqn.(3) and the preferred orientations extracted from Fig. 6. A square is associated for each cell position, and an oriented line in each square represents the preferred orientation of the cell at that position while the background color corresponds to the sign of stripes at that position. Grey and white respectively corresponds to negative and positive.