Research Report No.: C-144
Title: A Mathematical Analysis of Development of Oriented Receptive Fields in Linsker's Model (revised version)
Author: YAMAZAKI Tadashi
Affiliation:
c/o Prof. Osamu Watanabe,
Dept. of Mathematical and Computing Sciences, Tokyo Institute of Technology.
email: tyam@is.titech.ac.jp
Date: July 14, 2000

Abstract. Linsker's multi-layer neural network model is considered, in particular, development of an oriented receptive field of a cell on layer G is studied mathematically. We attempt to investigate which features in his model are essential for developing oriented receptive fields. Here we consider a slightly simplified version of Linsker's model and analyze its process of synaptic weights modification by computing the Fourier transform of the weight modification rule. Through our mathematical analysis, we obtain the following observations: (1) a Mexican-hat shape function as the correlation function extracts the Fourier components of a certain frequency, (2) the shape of the Fourier cosine component is circular symmetric while that of the Fourier sine component is bi-modal, and (3) the arbor function determines either the Fourier sine component or the Fourier cosine component dominates the other. These observations are also justified by our computer simulations. Finally we discuss the role of Linsker's parameters k_1 and k_2.
Notice: This research report supercedes C-135.

1 Introduction

In a series of papers, Linsker (1986) proposed a multi-layer feed-forward neural network model and studied the self-organization process of the model. He found that cells on a lower layer obtains center-surround structured receptive fields while cells on a higher layer, e.g., layer G, obtains bi-lobed or tri-lobed structured receptive fields through computer simulations[1]. His model contained parameters k_1 and k_2, and he investigated the shapes of receptive fields obtained after a number of iterations on self-organization processes changing those parameters. Linsker (1990) also analyzed his model mathematically following an information theoretic approach.

Inspired by Linsker's works, researchers have investigated Linsker's model from several different view points[2]. Kammen and Yuille (1988) introduced an energy function to measure the stability of synaptic structures, and they analyzed the properties of synaptic structures with the minimum energy, under the assumption that the energy function takes the minimum in the finally obtained synaptic structure. Since their analysis is based on their internal parameters, it is not clear which properties of Linsker's model is crucial for the development of oriented receptive fields. Mackay and Miller (1990) focused on development

[1]Linsker called receptive fields with a central excitatory stripe and two inhibitory flanking stripes "bi-lobed". On the other hand, we call them *tri-lobed* throughout this paper, and we call a bi-lobed or a tri-lobed receptive field an *oriented* receptive field.

[2]The original Linsker's model was not usually studied in these investigations. From various technical reasons, researchers investigated somewhat simplified versions of Linsker's model.

of a receptive field of a cell on a lower layer, e.g., layer C where circular symmetric receptive fields are developed, and studied how shapes of receptive fields are changed depending on parameters k_1 and k_2. However the analysis of higher layers, in particular, the layer G has been left open. Feng, Pan and Roychowdhury (1997) gave a rigorous numerical analysis of the parameter space for which oriented receptive fields are developed. But its theoretical analysis has not been fully investigated. Thus, the question of which properties of Linsker's model are crucial and how these properties work to develop oriented receptive fields has not been understood well. In this paper, for investigating this question, we focus on the layer G of Linsker's model and analyze the modification of synaptic structure mathematically.

For our analysis, we also consider a simplified version of Linsker's model. Our model consists of two neuronal layers that correspond to layer F and layer G in Linsker's model. The lower layer (i.e., layer F) contains a number of cells that project synaptic connections to the upper layer cell, on the other hand, the upper layer (i.e., layer G) consists of a single cell that receives connections from cells on the lower layer. The self-organization process of this model starts by assigning initial synaptic weights between cells in the lower layer and and the cell in the upper layer uniformly at random. These weights are updated repeatedly by applying a modification rule, that is, a rule that changes the amount of synaptic weights by computing convolution between the distribution of synaptic weights and a *correlation function* that describes correlations of activities among cells in the lower layer. In our modification rule, we use a Mexican-hat shape function as the correlation function since in Linsker's model the shape of correlations among cells in layer F was Mexican-hat. Thus, our model is for studying the synaptic weight modification between cells in layer F and a single cell in layer G of Linsker's model. Here we introduce an *arbor function* that determines the range of cells in the lower layer to which the modification rule applies. This function was implicit in Linsker's model, however we consider it explicitly.

By using this model, we study which properties and parameter settings are crucial for developing oriented receptive fields by investigating the self-organization process mathematically. We consider the Fourier transform of our weight modification rule. Firstly we focus on the Fourier transform of a typical correlation function under which oriented receptive fields are developed, and see that the Fourier transform is bi-modal with a clear peak value and that the distance of the peak from the center is π/R, where R is the arbor radius, that is, the radius of "effective" cell area. Further we study our modification rule under this assumption, namely, this type of bi-modal Fourier transform of the correlation function. The Fourier transform of the correlation function extracts a certain frequency $\mathbf{n}$ such that $|\mathbf{n}| = \pi/R$ on a circle. This means that the Fourier coefficients are distributed on the circle, so we can think of the distribution as an one-dimension function. From the property of the arbor function, we see that the one-dimension function, that is, the distribution of those coefficients, does not oscillate. Then, we separate those coefficients into the cosine part and the sine part, namely, the circle of the Fourier cosine coefficients and that of the sine coefficients. From the property of the Fourier transform, these two circles show different slants in two-dimension. The cosine circle becomes flat, that is, all coefficients take more or less the same value, and the shape of the Fourier cosine component becomes something circular symmetric. On the other hand, the sine circle slants and shows a peak at a certain frequency $\mathbf{n}_*$. Thus the shape of the Fourier sine component becomes something bi-lobed and its orientation is determined by $\mathbf{n}_*$. Finally, we see that the growth speeds of the Fourier cosine component and the sine component are different and are determined by the arbor

function. These observations are also justified by our computer simulations.

For discussing the role of our assumption, we investigate some cases under slightly different settings. We show that if our assumption on the shape of the Fourier transform of a correlation function fails to hold, then oriented receptive fields are not developed in those cases. We also discuss the role of parameters k_1 and k_2 in Linsker's model. We point out that these parameters, if they are chosen appropriately, might give the same effect as our assumption on the shape of the Fourier transform of a correlation function.

This paper is organized as follows. In Section 2 we make a brief overview of Linsker's model and introduce our simplified model. Section 3 explains our main mathematical results. Here we study the self-organization process of our model analytically. Section 4 is devoted for computer simulations to give justifications for our analysis. In Section 5 we discuss the role of our assumption and that of k_1 and k_2.

Finally, all proofs of lemmas can be found in Appendix.

2 Linsker's model and our simplified model

In this section, firstly we will make a brief overview of Linsker's model. Since our interest is development of an oriented receptive field, we focus on the afferent synaptic connections between layer F and layer G, where the activity of cells on Layer F follows a Mexican-hat covariance function, and a cell on Layer G obtains an oriented receptive field. Secondly, we consider a correlation based model which is a simplified version of his model and in the following we study instead. Relations between the simplified model and his original model will be discussed in the last section.

Linsker's model

Neural network structure

We consider a multi-layer network model (Fig. 1). The model consists of seven neuronal layers called layer $A, B, \ldots, G$, respectively. Each cell of a layer (here we denote as layer $\mathcal{L}$) projects connections approximating a Gaussian distribution centered on a corresponding point in the successive layer (we denote as layer $\mathcal{M}$). It should be noted that Linsker himself introduced lateral connections between cells in layer G. Here however, we omit those connections since the emergence of oriented receptive fields is independent on existing such connections.

Modification rule

The activity level $X^{\mathcal{M}}$ of a cell in layer $\mathcal{M}$ can be computed by the linear sum of its synaptic weights w_i and the activity level of cells in layer L, plus a constant a_1; i.e.

$$X^{\mathcal{M}} = \sum_{i \in \mathcal{L}} w_i X_i^{\mathcal{L}} + a_1.$$

Here it should be noted that $X^{\mathcal{M}}$ might be negative as well as positive. Synaptic weights are modified according to a linear Hebbian rule;

$$\Delta w_i = c_1 \left(X^{\mathcal{M}} - c_2\right)\left(X_i^{\mathcal{L}} - c_3\right) + c_4,$$

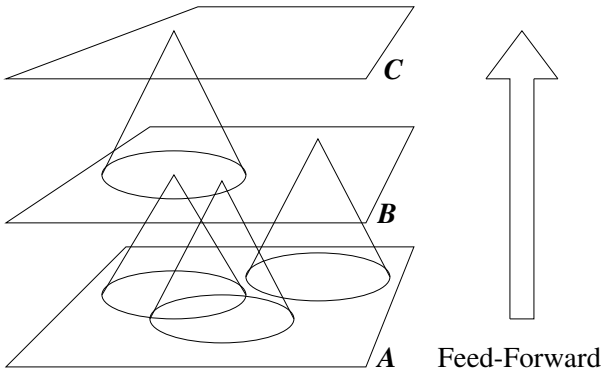

Figure 1: The first three layers of Linsker's model.

where $c_1, \ldots, c_4$ are constants. Synaptic weights are modified following each random presentation of an activity pattern in layer $\mathcal{L}$ to layer $\mathcal{M}$, however assuming that w_i changes on a much longer time scale than the presentations, we can take an average over the ensemble of patterns $X_i^{\mathcal{L}}$ in layer $\mathcal{L}$. Then, the mean rate of change of w_i can be computed as follows.

$$\Delta w_i = k_1 + \sum_{j \in \mathcal{L}} \left(Q_{ij}^{\mathcal{L}} + k_2 \right) w_j, \tag{1}$$

where k_1, k_2 are constants and $Q_{ij}^{\mathcal{L}} = \left\langle (X_i^{\mathcal{L}} - x^{\mathcal{L}})(X_j^{\mathcal{L}} - x^{\mathcal{L}}) \right\rangle$ is the covariance matrix of points i and j in layer $\mathcal{L}$, and $x^{\mathcal{L}}$ is an average activity of cells in layer $\mathcal{L}$ over the ensemble of presentations. Synaptic weights are constrained by an upper bound and a lower bound, so that $-w_{\min} \leq w_i \leq w_{\max}$ holds.

Computer simulations

Linsker trained this network sequentially, first training the connections from A to B, using random and uncorrelated noise as the inputs from A to B; i.e., Q_{ij}^{A} is the identity matrix. Once the connections had stabilized, the covariance matrix Q_{ij}^{B} was computed and used to modify connections between layer B and C. This procedure was repeated for subsequent layers. During this process, the shape of a receptive field showed an ordered structure, in particular, cells in layer G obtained bi-lobed or tri-lobed receptive fields, where the covariance matrix in layer F; i.e., Q_{ij}^{F} can be expressed as a Mexican-hat function.

Our simplified model

We consider the continuous form of the modification rule, that is, for any cell located at position $\mathbf{x}$ on layer F, we use $w(\mathbf{x})$ instead of w_j to denote its synaptic weight to the cell on layer G. Similarly we denote the covariance between cells at position $\mathbf{x}$ and $\mathbf{y}$ on layer F as $C(\mathbf{x} - \mathbf{y})$. Thus, the covariance between two cells is computed by only the distance between those cells. On the other hand, we trace the self-organization process in steps so that we see how the self-organization process develops an oriented receptive field.

In summary, for our step wise analysis, we express the modification rule as follows.

$$\Delta w^{(t)}(\mathbf{x}) = A(\mathbf{x}) \left(k_1 + \int_{-\infty}^{\infty} \left(C(\mathbf{x} - \mathbf{y}) + k_2 \right) w^{(t)}(\mathbf{y}) d\mathbf{y} \right). \tag{2}$$

$w^{(t)}(\mathbf{x})$ is the synaptic weight of the cell on $\mathbf{x}$ at t th step while $\Delta w^{(t)}(\mathbf{x})$ is the growth rate at this step. $A(\mathbf{x})$ is an *arbor function* that determines the area on layer $F$of "effective cells", that have connections to the cell on layer G. We consider the following two functions as arbor functions.

$$A(\mathbf{x}) = \begin{cases} 1 & |x| \leq R, \\ 0 & \text{otherwise.} \end{cases}, \qquad \text{and} \qquad A(\mathbf{x}) = e^{-\frac{|\mathbf{x}|^2}{2\sigma_A^2}},$$

where R and σ_A are positive constants that determine the radius of effective cell area, which we call *arbor radius*. In a similar way, we denote the diameter of effective cell area *arbor diameter*.) Moreover, we call those functions *Step* and *Gaussian*, respectively.

On the other hand, since Q_{ij}^L can be expressed as a Mexican-hat function, we use the following DOG function as our correlation function.

$$C(\mathbf{x}) = K_1 \exp\left(-\frac{|\mathbf{x}|^2}{2\sigma_1^2}\right) - K_2 \exp\left(-\frac{|\mathbf{x}|^2}{2\sigma_2^2}\right), \tag{3}$$

where $\sigma_1 < \sigma_2$. DOG functions of this type are used to express Mexican-hat functions (Miyashita and Tanaka (1992); Miller (1994)).

Finally, initial synaptic weights $w^{(0)}(\mathbf{x})$ are assigned uniformly at random only within the arbor radius, that is,

$$w^{(0)}(\mathbf{x}) = A(\mathbf{x}) \cdot (\text{uniformly at random})$$

With this setting, the following proposition holds.

Proposition 1.
the following equation is identical to the original equation (2).

$$\Delta w_*^{(t)}(\mathbf{x}) = k_1 + \int_{-\infty}^{\infty} \left(C(\mathbf{x}-\mathbf{y}) + k_2\right) A(\mathbf{y}) w_*^{(t)}(\mathbf{y}) d\mathbf{y},$$

by taking $w_*^{(0)}(\mathbf{x})$ *as*

$$w_*^{(0)}(\mathbf{x}) = (\textit{uniformly at random}),$$

We consider the equation for $w_*(\mathbf{x})$ as $w(\mathbf{x})$, instead of Eqn. (2).

From technical reasons, we further simplify this equation by omitting k_1, k_2 and consider the following updating rule.

$$\Delta w^{(t)}(\mathbf{x}) = \int_{-\infty}^{\infty} C(\mathbf{x}-\mathbf{y}) A(\mathbf{y}) w^{(t)}(\mathbf{y}) dy. \tag{4}$$

Thus we will use Eqn. (4) in the following discussion. However, later we consider the model including k_1, k_2 and discuss the role of those constants.

3 Analytical Results

This section contains our main results. We focus on the growth rate of synaptic weights, because if the shape of the growth rate becomes bi-lobed, then the shape of the receptive field that is obtained finally is also bi-lobed. Thus, we do not focus on the synaptic weights itself. Here it should be noted that we assume an upper bound and a lower bound are large enough and during the process synaptic weights never reach those bounds, therefore we do not take into account the role of those constraints.

Before starting the analysis, we give a brief overview of our analysis to make its outline clear to understand.

- We consider the Fourier transform of Eqn. (4).
- The Fourier transform of the correlation function extracts a certain frequency on a circle so that the shape of the Fourier transform is bi-modal (Eqn. (5)).
- Since the Fourier coefficients are distributed on the circle, we can think of the distribution as an one-dimensional function (Eqn. (9)).
- From the property of the arbor function, we see that the one-dimensional function, that is, the distribution of those coefficients, does not oscillate.
- Then, we separate those coefficients into the cosine part and the sine part, namely, the circle of the Fourier cosine coefficients and that of the sine coefficients (Eqn. (10)). From the property of the Fourier transform, these two circles show different slants in two-dimension. The cosine circle becomes flat and the shape of the Fourier cosine component becomes something circular symmetric. On the other hand, the sine circle slants and shows a peak at a certain frequency $\mathbf{n}_*$ as shown by Eqn. (12). Thus the shape of the Fourier sine component becomes something bi-lobed and its orientation is determined by $\mathbf{n}_*$.
- Finally, we see that the growth speeds of the Fourier cosine component and the sine component are different and are determined by the arbor function (Eqn. (13)).

We start our analysis from calculating the Fourier transform of our correlation function $C(\mathbf{x})$ defined as Eqn.(3) in the previous section.

Lemma 1.
The Fourier transform $\widehat{C}(\mathbf{n})$ of the correlation function $C(\mathbf{x})$ defined as Eqn.(3) is the following DOG function.

$$\widehat{C}(\mathbf{n}) = 2\pi\sigma_1^2 K_1 \exp\left(-\frac{\sigma_1^2|\mathbf{n}|^2}{2}\right) - 2\pi\sigma_2^2 K_2 \exp\left(-\frac{\sigma_2^2|\mathbf{n}|^2}{2}\right). \tag{5}$$

As can be seen, the Fourier transform $\widehat{C}(\mathbf{n})$ of $C(\mathbf{x})$ is also a DOG function. Here an important fact is that depending on the choice of K_1, K_2, σ_1, and σ_2, the shape of $\widehat{C}(\mathbf{n})$ is either uni-modal or bi-modal (see e.g., Fig. 2).

Furthermore, computer simulation results (see, e.g., (Miyashita and Tanaka (1992); Miller (1994)) show that the shape of $\widehat{C}(\mathbf{n})$ needs to be bi-modal to develop an oriented

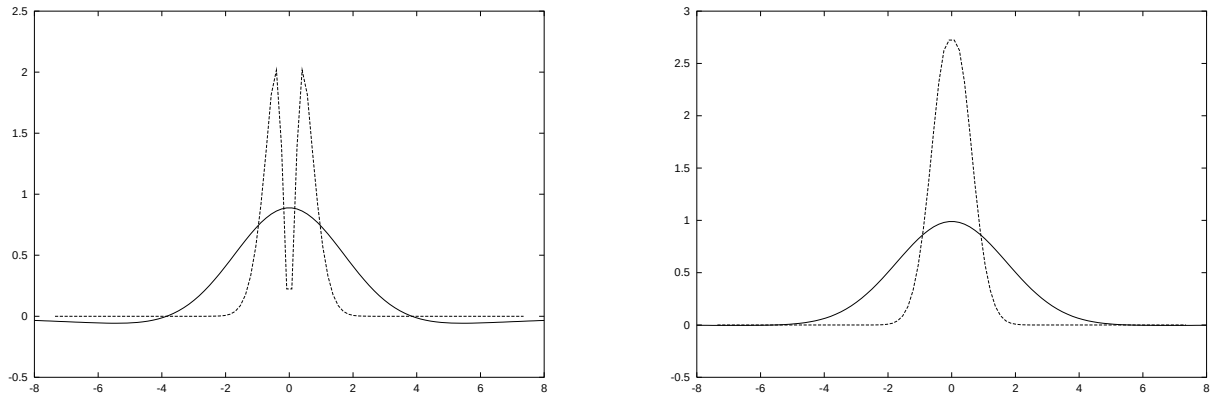

Figure 2: A correlation function (real line) and its Fourier transform divided by 2π (dashed line). (Left: those parameters are the following. $\sigma_1^2 = 8.0$, $\sigma_2^2 = 27.0$,$K_1 = 1.0$, $K_2 = (\sigma_1/\sigma_2)^2$.) (Right: the parameters are the same as the left except $K_2 = 0.001$.)

receptive field. Moreover, these results also indicate that the distance of the peak from the center is essential and it should be close to π/R, where R is the arbor radius. Based on these observations we propose here to assume that the shape of $\widehat{C}(\mathbf{n})$ is bi-modal with a clear peak value and that the distance of the peak from the center is π/R.

Technically, in order to simplify our analysis, we make use of the following assumption in our analysis.

Assumption 1.
For some constant $\alpha > 1$ and some sufficiently small constant $\varepsilon > 0$, $\widehat{C}(\mathbf{n})$ is defined as follows.

$$\widehat{C}(\mathbf{n}) = \begin{cases} C_*, & \sigma - \varepsilon \leq |\mathbf{n}| \leq \sigma + \varepsilon, \\ 0, & \textit{otherwise}, \end{cases}$$

where σ is a parameter determined by $K_1, K_2, \sigma_1, \sigma_2$. We assume $\sigma = \pi/R$.

From this assumption, it follows

$$\Delta w^{(t)}(\mathbf{x}) = \left(\frac{C_*}{4\pi^2}\right) \int_{|\mathbf{n}_t| \approx \sigma} \widehat{Aw}^{(t)}(\mathbf{n}_t) e^{i\mathbf{n}_t \cdot \mathbf{x}} d\mathbf{n}_t,$$

where $\widehat{Aw}^{(t)}(\mathbf{n}_t)$ is the Fourier transform of $A(\mathbf{x})w^{(t)}(\mathbf{x})$, and $|\mathbf{n}_t| \approx \sigma$ of the integral domain stands for $\sigma - \varepsilon \leq |\mathbf{n}_t| \leq \sigma + \varepsilon$. We make use of this notation for the sake of simplicity. This equation says that the convolution between $C(\mathbf{x})$ and $A(\mathbf{x})w^{(t)}(\mathbf{x})$ in Eqn. (4) extracts components of $A(\mathbf{x})w^{(t)}(\mathbf{x})$ that have a certain frequency.

This formula is for the tth step. Then we expand it from the first step as follows.

Lemma 2.

$$\Delta w^{(t)}(\mathbf{x}) = \left(\frac{C_*}{4\pi^2}\right) \int_{|\mathbf{n}| \approx \sigma} \widehat{Aw}^{(0)} * \left(1 + \left(\frac{C_*}{4\pi^2}\right) \widetilde{A}\right)^{(t)_*} (\mathbf{n}) e^{i\mathbf{n} \cdot \mathbf{x}} d\mathbf{n}, \tag{6}$$

where $*$ stands for a convolution, $(f)^{(t)*}$ is t times self-convolutions of a function f, and

$$\widetilde{A}(\mathbf{n}) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} A(\mathbf{x}) e^{i\mathbf{n}\cdot\mathbf{x}} d\mathbf{x}.$$

The term 1 in $\left(1 + \left(\frac{C_*}{4\pi^2}\right)\right)$ extracts all frequencies of $|\mathbf{n}| \approx \sigma$ with the same level. Thus this 1 gives the effect to grow all Fourier coefficients of $|\mathbf{n}| \approx \sigma$ while this does not give any effects to change the shape of the oriented receptive field in a self-organization process. Since we focus on the shape of receptive fields, we may omit this 1. Hence, Eqn. (6) could be simplified as follows.

$$\Delta w^{(t)}(\mathbf{x}) = \left(\frac{C_*}{4\pi^2}\right)^{t+1} \int_{|\mathbf{n}_t|\approx\sigma} \cdots \int_{|\mathbf{n}_0|\approx\sigma} \widehat{Aw}^{(0)}(\mathbf{n}_0) \widetilde{A}(\mathbf{n}_0 - \mathbf{n}_1) \cdots \widetilde{A}(\mathbf{n}_{t-1} - \mathbf{n}_t) e^{i\mathbf{n}_t\cdot\mathbf{x}} d\mathbf{n}_0 \cdots d\mathbf{n}_t. \quad (7)$$

The closed form of $\widetilde{A}(\mathbf{n})$ can be obtained as follows.

Lemma 3.
A is Gaussian.

$$\widetilde{A}(\mathbf{n}) = (2\pi\sigma_A^2) e^{-\frac{\sigma_A^2 |\mathbf{n}|^2}{2}}.$$

A is Step.

$$\widetilde{A}(\mathbf{n}) = 2\pi R \frac{J_1(R|\mathbf{n}|)}{|\mathbf{n}|},$$

where $J_1(x)$ is the Bessel function of the first kind of order 1.

Next, we focus on how the Fourier coefficients are distributed in steps. For any i, let $f_i(\mathbf{n})$ be the following function;

$$f_i(\mathbf{n}) \stackrel{\text{def}}{=} \begin{cases} \int_{|\mathbf{n}_{i-1}|\approx\sigma} \cdots \int_{|\mathbf{n}_0|\approx\sigma} \widehat{Aw}^{(0)}(\mathbf{n}_0) \\ \qquad \widetilde{A}(\mathbf{n}_0 - \mathbf{n}_1)\widetilde{A}(\mathbf{n}_1 - \mathbf{n}_2) \cdots \widetilde{A}(\mathbf{n}_{i-1} - \mathbf{n}) d\mathbf{n}_0 \cdots d\mathbf{n}_{i-1}, & |\mathbf{n}| = \sigma, \\ 0, & \text{otherwise.} \end{cases} \quad (8)$$

On the circle such that $|\mathbf{n}| = \sigma$, for any i, $f_i(\mathbf{n})$ can be considered as an one-dimensional function of θ, which we denote $f_i(\sigma\theta)$. Thus, we can perform the Fourier transform to $f_i(\sigma\theta)$ to decompose the convolution into a simple multiplication as follows.

$$\widehat{f}_i(u) = \widehat{\widehat{Aw}}^{(0)}(u) \widehat{\widetilde{A}}^i(u).$$

Since $\widetilde{A}(u)$ is Gaussian (if $A(x)$ is Gaussian), $\widehat{\widetilde{A}}(u)^i$ shows its maximum value on $u = 0$. Hence $\widehat{\widetilde{A}}(u)^i$ extracts only frequencies around $u = 0$. (This property also holds in the case of which $A(x)$ is Step.) By computing the inverse Fourier transform, we obtain,

$$f_i(\sigma\theta) = \int_{u\approx 0} \widehat{\widehat{Aw}}^{(0)}(u) e^{iu(\sigma\theta)} du. \quad (9)$$

Since $u \approx 0$, $f_i(\sigma\theta)$ does not oscillate in $\theta \in [0, 2\pi]$. Later we will confirm it by computer simulations.

Let us decompose Eqn. (7) into the Fourier cosine component and the Fourier sine component. The result is as follows.

$$\begin{aligned}\Delta w^{(t)}(\mathbf{x}) = &\left(\frac{2C_*}{4\pi^2}\right)^{t+1} \int_{|\mathbf{n}_t^+|\approx\sigma} f_t^{\cos}(\mathbf{n}_t)\cos(\mathbf{n}_t\cdot\mathbf{x})d\mathbf{n}_t \\ &+ \left(\frac{2C_*}{4\pi^2}\right)^{t+1} \int_{|\mathbf{n}_t^+|\approx\sigma} f_t^{\sin}(\mathbf{n}_t)\sin(\mathbf{n}_t\cdot\mathbf{x})d\mathbf{n}_t,\end{aligned} \tag{10}$$

where

$$\begin{aligned}f_t^{\cos}(\mathbf{n}_t) &\stackrel{\text{def}}{=} \int_{|\mathbf{n}_{t-1}|^+\approx\sigma} \cdots \int_{|\mathbf{n}_0|^+\approx\sigma} \widehat{Aw}^{(0)\cos}(\mathbf{n}_0)\widetilde{A}^{\cos}(\mathbf{n}_0,\mathbf{n}_1)\cdots\widetilde{A}^{\cos}(\mathbf{n}_{t-1},\mathbf{n}_t)d\mathbf{n}_0\cdots d\mathbf{n}_{t-1}, \\ f_t^{\sin}(\mathbf{n}_t) &\stackrel{\text{def}}{=} \int_{|\mathbf{n}_{t-1}|^+\approx\sigma} \cdots \int_{|\mathbf{n}_0|^+\approx\sigma} \widehat{Aw}^{(0)\sin}(\mathbf{n}_0)\widetilde{A}^{\sin}(\mathbf{n}_0,\mathbf{n}_1)\cdots\widetilde{A}^{\sin}(\mathbf{n}_{t-1},\mathbf{n}_t)d\mathbf{n}_0\cdots d\mathbf{n}_{t-1},\end{aligned} \tag{11}$$

$$\begin{aligned}\widetilde{A}^{\cos}(\mathbf{n},\mathbf{m}) &\stackrel{\text{def}}{=} \int_{-\infty}^{\infty} A(\mathbf{x})\cos(\mathbf{n}\cdot\mathbf{x})\cos(\mathbf{m}\cdot\mathbf{x})d\mathbf{x}, \\ \widetilde{A}^{\sin}(\mathbf{n},\mathbf{m}) &\stackrel{\text{def}}{=} \int_{-\infty}^{\infty} A(\mathbf{x})\sin(\mathbf{n}\cdot\mathbf{x})\sin(\mathbf{m}\cdot\mathbf{x})d\mathbf{x},\end{aligned}$$

where due to the symmetry of the Fourier coefficients by the definition of the Fourier transform, it is sufficient to consider only the first and second quadrants as the integral domain, which we denote as $|\mathbf{n}|^+ = \sigma$ by using + symbol.

Due to the symmetry of the sign of the Fourier cosine coefficients, we have

$$f_t^{\cos}(\mathbf{n}) = f_t^{\cos}(-\mathbf{n}).$$

Furthermore, $f_t(\sigma\theta)$ does not oscillate, we may think that $f_t^{\cos}(\mathbf{n})$ takes the almost the same value for all $\mathbf{n}$. This means that the shape of the Fourier cosine component becomes a concentric circle. On the other hand, since the sign of the Fourier sine coefficients are changed, namely,

$$f_t^{\sin}(\mathbf{n}) = -f_t^{\sin}(-\mathbf{n}),$$

there is the frequency $\mathbf{n}_*$ such that

$$f_t^{\sin}(\mathbf{n}_*) = \max_{\mathbf{n}}\{f_t^{\sin}(\mathbf{n})\}. \tag{12}$$

Thus, the sine functions in the both sides of $\mathbf{n}_*$ are canceled out and $\sin(\mathbf{n}_*\cdot\mathbf{x})$ remains, and hence the shape of the Fourier sine component becomes bi-lobed with the orientation $\theta_{\mathbf{n}_*} = \tan^{-1}(n_{*1}/n_{*0})$, where $(n_{*0}, n_{*1}) \stackrel{\text{def}}{=} \mathbf{n}_*$.

Finally, how does one of the Fourier cosine and sine components dominate the other? In following, we will observe that the growth rate of $f_t^{\cos}(\mathbf{n})$ with respect to t and that of $f_t^{\sin}(\mathbf{n})$ are different and the difference comes from the shape of the arbor function. Let

$g_t^{\cos}(\mathbf{n})$ and $g_t^{\sin}(\mathbf{n})$ be functions that are as the same as $f_t^{\cos}(\mathbf{n})$ and $f_t^{\sin}(\mathbf{n})$ except excluding $\widehat{Aw}^{(0)\cos}(\mathbf{n}_0)$ and $\widehat{Aw}^{(0)\sin}(\mathbf{n}_0)$.

$$\begin{aligned} g_t^{\cos}(\mathbf{n}) &\stackrel{\text{def}}{=} \int_{|\mathbf{n}_{t-1}|^+\approx\sigma} \cdots \int_{|\mathbf{n}_0|^+\approx\sigma} \widetilde{A}^{\cos}(\mathbf{n}_0,\mathbf{n}_1)\cdots\widetilde{A}^{\cos}(\mathbf{n}_{t-1},\mathbf{n})d\mathbf{n}_0\cdots d\mathbf{n}_{t-1}, \\ g_t^{\sin}(\mathbf{n}) &\stackrel{\text{def}}{=} \int_{|\mathbf{n}_{t-1}|^+\approx\sigma} \cdots \int_{|\mathbf{n}_0|^+\approx\sigma} \widetilde{A}^{\sin}(\mathbf{n}_0,\mathbf{n}_1)\cdots\widetilde{A}^{\sin}(\mathbf{n}_{t-1},\mathbf{n})d\mathbf{n}_0\cdots d\mathbf{n}_{t-1}. \end{aligned} \tag{13}$$

Here we consider portions of those functions which we denote $\widetilde{A}^{\cos}(\mathbf{n})$ and $\widetilde{A}^{\sin}(\mathbf{n})$ respectively.

$$\begin{aligned} \widetilde{A}^{\cos}(\mathbf{n}) &\stackrel{\text{def}}{=} \int_{|\mathbf{m}|^+\approx\sigma} \widetilde{A}^{\cos}(\mathbf{n},\mathbf{m})d\mathbf{m}, \quad &= \int_{|\mathbf{m}|^+\approx\sigma}\int_{-\infty}^{\infty} A(\mathbf{x})\cos(\mathbf{n}\cdot\mathbf{x})\cos(\mathbf{m}\cdot\mathbf{x})d\mathbf{x}d\mathbf{m}, \\ \widetilde{A}^{\sin}(\mathbf{n}) &\stackrel{\text{def}}{=} \int_{|\mathbf{m}|^+\approx\sigma} \widetilde{A}^{\sin}(\mathbf{n},\mathbf{m})d\mathbf{m}, \quad &= \int_{|\mathbf{m}|^+\approx\sigma}\int_{-\infty}^{\infty} A(\mathbf{x})\sin(\mathbf{n}\cdot\mathbf{x})\sin(\mathbf{m}\cdot\mathbf{x})d\mathbf{x}d\mathbf{m}, \end{aligned}$$

Since $\widetilde{A}^{\cos}(\mathbf{n})$ and $\widetilde{A}^{\sin}(\mathbf{n})$ are circular symmetric with respect to $\mathbf{n}$ such that $|\mathbf{n}| \approx \sigma$, for any $\mathbf{n}$, those functions can be considered as some constants by taking $\mathbf{n}$ as, for instance, $(\sigma, 0)$, namely,

$$\widetilde{A}^{\cos} \stackrel{\text{def}}{=} \int_{|\mathbf{m}|^+\approx\sigma} \widetilde{A}^{\cos}(\mathbf{n},\mathbf{m})d\mathbf{m}\Big|_{\mathbf{n}\approx(\sigma,0)}, \quad \text{and} \quad \widetilde{A}^{\sin} \stackrel{\text{def}}{=} \int_{|\mathbf{m}|^+\approx\sigma} \widetilde{A}^{\sin}(\mathbf{n},\mathbf{m})d\mathbf{m}\Big|_{\mathbf{n}\approx(\sigma,0)}.$$

Then $g_t^{\cos}(\mathbf{n})$ and $g_t^{\sin}(\mathbf{n})$ can be simplified as follows.

$$\begin{aligned} g_t^{\cos}(\mathbf{n}) &= \widetilde{A}_{\cos}^t \int_{|\mathbf{n}_{t-1}|^+\approx\sigma} \cdots \int_{|\mathbf{n}_0|^+\approx\sigma} d\mathbf{n}_0\cdots d\mathbf{n}_{t-1}, = \widetilde{A}_{\cos}^t (2\pi\sigma)^t, \\ g_t^{\sin}(\mathbf{n}) &= \widetilde{A}_{\sin}^t (2\pi\sigma)^t. \end{aligned}$$

Since $g_t^{\cos}(\mathbf{n})$ and $g_t^{\sin}(\mathbf{n})$ are only defined at $\mathbf{n}$ such that $|\mathbf{n}| \approx \sigma$, again those are constants. Thus, we can see that the values of $\widetilde{A}_{\cos}$ and $\widetilde{A}_{\sin}$ determine which component dominates between the Fourier cosine component and the Fourier sine component. In other words, by calculating $\widetilde{A}_{\cos}$ and $\widetilde{A}_{\sin}$, we can compute shapes of receptive fields without doing computer simulations.

In the case of which the arbor function is Gaussian, we can compute $\widetilde{A}_{\cos}$ and $\widetilde{A}_{\sin}$ rigorously as follows.

$$\begin{aligned} \widetilde{A}^{\cos} &= \int_{|\mathbf{m}|^+\approx\sigma}\int_{-\infty}^{\infty} A(\mathbf{x})\cos(\mathbf{n}\cdot\mathbf{x})\cos(\mathbf{m}\cdot\mathbf{x})d\mathbf{x}d\mathbf{m}\Big|_{\mathbf{n}\approx(\sigma,0)}, \\ &= \int_{|\mathbf{m}|^+\approx\sigma} \pi\sigma_A^2\left(e^{-\frac{\sigma_A^2}{4}|\mathbf{n}+\mathbf{m}|^2} + e^{-\frac{\sigma_A^2}{4}|\mathbf{n}-\mathbf{m}|^2}\right)d\mathbf{x}d\mathbf{m}\Big|_{\mathbf{n}\approx(\sigma,0)}, \\ \widetilde{A}^{\sin} &= \int_{|\mathbf{m}|^+\approx\sigma} \pi\sigma_A^2\left(e^{-\frac{\sigma_A^2}{4}|\mathbf{n}+\mathbf{m}|^2} - e^{-\frac{\sigma_A^2}{4}|\mathbf{n}-\mathbf{m}|^2}\right)d\mathbf{x}d\mathbf{m}\Big|_{\mathbf{n}\approx(\sigma,0)}, \end{aligned}$$

Since both of $e^{-\frac{\sigma_A^2}{4}|\mathbf{n}+\mathbf{m}|^2}$ and $e^{-\frac{\sigma_A^2}{4}|\mathbf{n}-\mathbf{m}|^2}$ are positive, in this case the Fourier cosine component will dominate the Fourier sine component.

On the other hand, suppose that the arbor function is Step. Unfortunately we can not compute $A^{\cos}$ and $A^{\sin}$. Thus we consider an approximation. Since $\widetilde{A}^{\cos}(\mathbf{n},\mathbf{m})$ and $\widetilde{A}^{\sin}(\mathbf{n},\mathbf{m})$ take the maximum values when $\mathbf{n}=\mathbf{m}$, we consider only the case of $\mathbf{n}=\mathbf{m}$. Then it follows,

$$\widetilde{A}^{\cos}(\mathbf{n},\mathbf{n}) = \int_{-\infty}^{\infty} A(\mathbf{x})\cos^2(\mathbf{n}\cdot\mathbf{x})d\mathbf{x}, \quad \text{and} \quad \widetilde{A}^{\sin}(\mathbf{n},\mathbf{n}) = \int_{-\infty}^{\infty} A(\mathbf{x})\sin^2(\mathbf{n}\cdot\mathbf{x})d\mathbf{x}.$$

Here we can observe that $\widetilde{A}^{\cos}(\mathbf{n},\mathbf{n})$ is less than $\widetilde{A}^{\sin}(\mathbf{n},\mathbf{n})$ (Fig. 3). Hence, in this case the Fourier sine component might dominate and a bi-lobed receptive field would be developed.

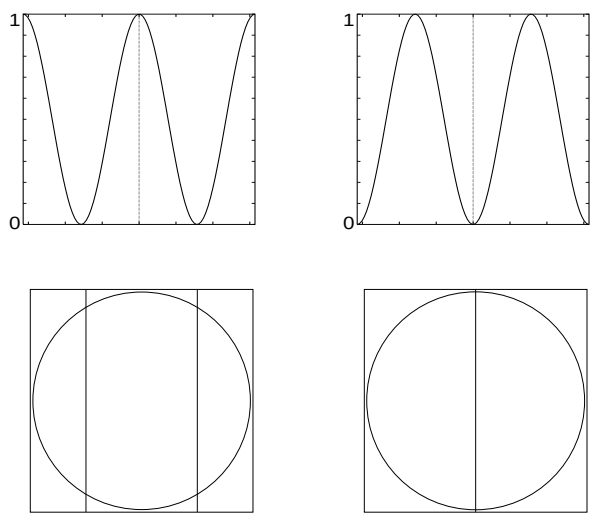

Figure 3: One-dimensional representation of $\cos^2(\mathbf{n}\cdot\mathbf{x})$ of the fundamental frequency with respect to the horizontal axis (top left) and that of $\sin^2(\mathbf{n}\cdot\mathbf{x})$ (top right). The peak value of $\cos^2(\mathbf{n}\cdot\mathbf{x})$ appears at the origin and the both borders of the horizontal axis, while that of $\sin^2(\mathbf{n}\cdot\mathbf{x})$ appears at the midpoint between the origin and the border. Applying $A(\mathbf{x})$ to $\cos^2(\mathbf{n}\cdot\mathbf{x})$ eliminates the value at the border, so $\widetilde{A}^{\cos}(\mathbf{n},\mathbf{n})$ becomes much smaller than that before applying $A(\mathbf{x})$. On the other hand, $A(\mathbf{x})$ does not affect so much to $\widetilde{A}^{\sin}(\mathbf{n},\mathbf{n})$ since at the border the value of $\sin^2(\mathbf{n}\cdot\mathbf{x})$ is small enough.

In summary, we observe that if the arbor function is Step, then the Fourier sine component dominates and the shape of the receptive field becomes bi-lobed. On the other hand, if the arbor function is Gaussian, then the Fourier cosine component dominates and a circular symmetric receptive field is developed.

4 Computer Simulations

In this section, we perform computer simulations to give justifications to our analytical results. Firstly we show the setting of our computer simulations and then we see the results.

Neural network structure

We consider two layer neural network. The lower layer is represented by 13 by 13 grids of cells while the upper layer consists of one cell. The cell on the upper layer connects to a circle of diameter 13 grid intervals of cortical cells, in other words, the arbor radius is 6.5.

Modification rule

According to Eqn. (4), synaptic weights are modified by the following rule.

$$\Delta w^{(t)}(\mathbf{x}) = \int_{-\infty}^{\infty} C(\mathbf{x}-\mathbf{y})A(\mathbf{y})w^{(t)}(\mathbf{y})d\mathbf{y}.$$

where $C(\mathbf{x})$ and $A(\mathbf{x})$ are

$$C(\mathbf{x}) = K_1 \exp\left(-\frac{|\mathbf{x}|^2}{2\sigma_1^2}\right) - K_2 \exp\left(-\frac{|\mathbf{x}|^2}{2\sigma_2^2}\right),$$

$$A(\mathbf{x}) = \begin{cases} 1 & |x| \leq R, \\ 0 & \text{otherwise.} \end{cases}, \qquad \text{and} \qquad A(\mathbf{x}) = e^{-\frac{|\mathbf{x}|^2}{2\sigma_A^2}},$$

Here we make use of the following parameters: $\sigma_1^2 = 3.0$,$\sigma_2^2 = 27.0$, $K_1 = 1.00$, $K_2 = (\sigma_1/\sigma_2)^2 = 0.111$, $R = 6.5$ and $\sigma_A^2 = 8.0$. With this setting the shape of the correlation function is represented as Fig. 2 left while that of Arbor functions appear in Fig. 4.

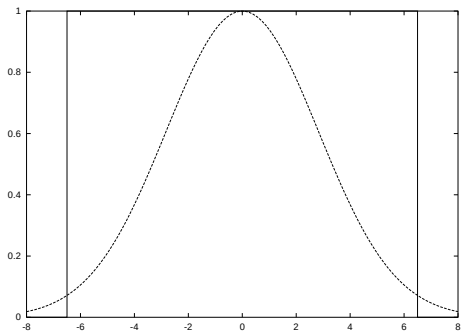

Figure 4: Arbor functions. Parameters are $R = 6.5$ and $\sigma_A^2 = 8.0$.

Self-organization process

The initial synaptic weights are assigned from $[-0.4, 0.4]$ uniformly at random. As the pseudo random generator to assign initial synaptic weights, we use MT (Matsumoto and Nishimura (1998)). Then the above modification rule is applied in steps. After 80 steps, we stop the self-organization process and compute the shape of the receptive field.

In following, we investigate two computer simulations for each arbor function.

The case of which A is Step

Firstly we compute $A^{\cos}$ and $A^{\sin}$ to predict what kind of receptive fields are obtained with this setting. Since a calculation shows that $A^{\cos}/A^{\sin} \approx 0.70$, we would expect that with this setting, a bi-lobed receptive field will be obtained.

The following is the result of a computer simulation. Figure 5 consists of from left to right the distribution of initial synaptic weights, the final shape of the receptive field, the Fourier cosine component and the sine component of the final shape, respectively. As can

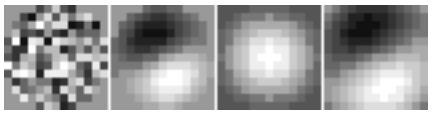

Figure 5: (From left to right) the distribution of initial synaptic weights, the final shape of the receptive field, the Fourier cosine component and the sine component of the final shape. In each figure, its weights are normalized.

be seen, although the initial synaptic weights are assigned uniformly at random, after a number of steps, a bi-lobed receptive field is obtained. Moreover, the final shape of the receptive field is as the same as that of its Fourier sine component while its Fourier cosine component is a concentric circle.

Fig. 6 shows the values of the Fourier cosine and sine coefficients, namely, $f_t^{\cos}(\mathbf{n})$ and $f_t^{\sin}(\mathbf{n})$ in Eqn. (11), in the first five steps. We can see that, at the beginning of the simulation, both of the Fourier coefficients oscillate, however after that, those converges to slanted circles. At last, the circle of the Fourier cosine coefficients are arranged to be flat, on the other hand, in the case of the Fourier sine coefficients, the circle slants and only one coefficient shows the maximum value.

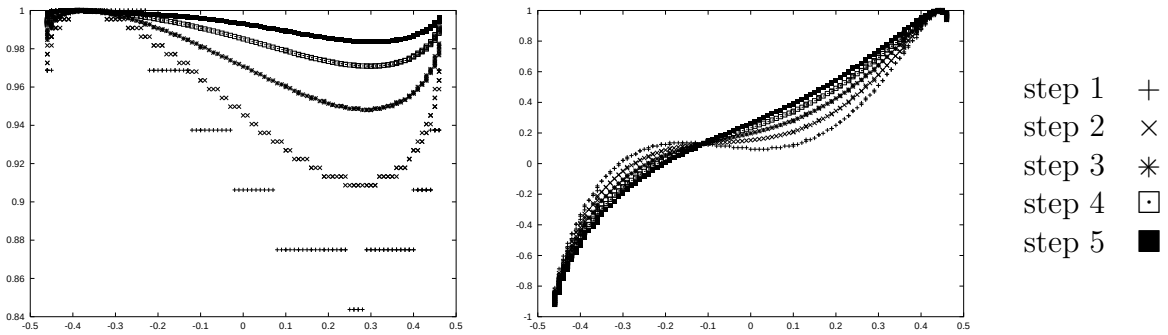

Figure 6: The growth of the Fourier cosine coefficients (left) and the sine coefficients (right) in the first five steps. Values of coefficients in each step are normalized and plotted. As can be seen, at the first step, the arrangement of those coefficients oscillates, on the other hand, as the self-organization proceeds by, those are arranged into circles.

The case of which A is Gaussian

In this case, a calculation shows $A^{\cos}/A^{\sin} \approx 1.44$. Thus, we expect that circular symmetric receptive fields are developed. Figure 7 shows from left to right, the final shape of the receptive field, its Fourier cosine and sine components. As can be seen, a circular symmetric receptive field is developed as we expected.

Those results support our analytical results. The correlation function extracts the Fourier components of a certain frequency, where the peak of $\widehat{C}$ appears at points that the distance from the center is $0.46 \approx \pi/6.5 = 0.48$. Synaptic weights grows exponentially with respect to the number of steps. Final shapes of receptive fields are described as one of its Fourier cosine component and its Fourier sine component. The shape of the Fourier

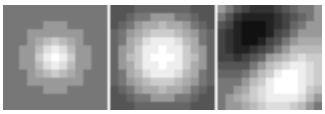

Figure 7: (From left to right) the final shape of the receptive field, the Fourier cosine component and the sine component of the final shape.

cosine component changes into a concentric circle while that of the Fourier sine component becomes bi-lobed. Finally, the arbor function determines which one of the Fourier cosine component and the sine component dominates the other.

5 Discussions

5.1 The role of Assumption 1

Here we discuss the justification of Assumption 1 in Section 3. We will see what happens if this assumption does not hold.

Firstly, suppose that the shape of $\widehat{C}(\mathbf{n})$ is uni-modal. Since $\widehat{C}(\mathbf{n})$ shows the peak value at the origin, it extracts the frequency around $\mathbf{n} = \mathbf{0}$. It follows

$$\widehat{C}(\mathbf{0}) = \int_{-\infty}^{\infty} C(\mathbf{x}) e^{-i\mathbf{0}\cdot\mathbf{x}} d\mathbf{x} = \int_{-\infty}^{\infty} C(\mathbf{x}) d\mathbf{x},$$
$$\widehat{Aw}^{(t)}(\mathbf{0}) = \int_{-\infty}^{\infty} A(\mathbf{x}) w^{(t)}(\mathbf{x}) d\mathbf{x}.$$

Since these do not oscillate, all synaptic weights take the same sign. Thus, any structure of receptive fields is not developed. On the other hand, $\int_{-\infty}^{\infty} C(\mathbf{x}) d\mathbf{x} = 0$ is a sufficient condition to make the shape of $\widehat{C}(\mathbf{n})$ bi-modal.

Next suppose that the peak is moved from the origin. The distance of the peak from the origin represents the peak frequencies of $e^{i\mathbf{n}\cdot\mathbf{x}}$; i.e., of triangle functions. In our assumption the distance is $\sigma = \pi/R$, this means that those triangle functions have the fundamental periodicity within the arbor diameter. Thus, if the distance is wider, the width of stripes of positive and negative subregions in the receptive field becomes narrower, and vice versa. The distance is determined by the range of the correlation function. It follows from lemma 2, a wider correlation function has a narrower Fourier transform and vice versa. Thus, the width of stripes in the receptive field is determined by the range of the correlation function.

5.2 The role of k_1, k_2

Here we will see that k_1, k_2 work to make the shape of $\widehat{C}(\mathbf{x})$ bi-modal even if original $\widehat{C}(\mathbf{x})$ is uni-modal. Mackay and Miller (1990) also obtained the similar result, They showed that by taking k_1, k_2 appropriately (mainly by taking k_2 as a large negative), Linsker's model developed bi-lobed receptive fields even if the correlation function was a Gaussian.

Firstly we discuss the role of k_1. As we saw, our modification rule (Eqn. (4)) can be

expressed in terms of the Fourier transform as follows.

$$\Delta w^{(t)}(\mathbf{x}) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty} \widehat{C}(\mathbf{n})\widehat{Aw}^{(t)}(\mathbf{n})e^{i(\mathbf{n}\cdot\mathbf{x})}d\mathbf{n}.$$

We separate the right hand side into the component that the frequency is zero and the others. Then the rule turns into the following one.

$$\Delta w^{(t)}(\mathbf{x}) = \frac{1}{4\pi^2}\widehat{C}(\mathbf{0})\widehat{Aw}^{(t)}(\mathbf{0})e^{i(\mathbf{0}\cdot\mathbf{x})} + \frac{1}{4\pi^2}\int_{|\mathbf{n}|\neq\mathbf{0}} \widehat{C}(\mathbf{n})\widehat{Aw}^{(t)}(\mathbf{n})e^{i(\mathbf{n}\cdot\mathbf{x})}d\mathbf{n}.$$

Since

$$\widehat{C}(\mathbf{0}) = \int_{-\infty}^{\infty} C(\mathbf{x})dxdy, \quad \text{and} \quad \widehat{Aw}^{(t)}(\mathbf{0}) = \int_{-\infty}^{\infty} A(\mathbf{x})w^{(t)}(\mathbf{x})d\mathbf{x},$$

the first term in the right hand side is some constant, and if $\widehat{C}(\mathbf{x})$ is uni-modal, this term dominates. Therefore, if we define k_1 as

$$k_1 \stackrel{\text{def}}{=} -\frac{1}{4\pi^2}\int_{-\infty}^{\infty} C(\mathbf{x})d\mathbf{x}\int_{-\infty}^{\infty} A(\mathbf{y})w^{(t)}(\mathbf{y})d\mathbf{y},$$

the effect of this term could be suppressed. At least, we can make computer simulations and estimate the value of the constant.

Next we consider the role of k_2. Here we suppose that k_2 is the following function $k_2(\mathbf{x})$.

$$k_2(x,y) = \begin{cases} k_2 & |\mathbf{x}| \leq R, \\ 0 & \text{otherwise.} \end{cases}$$

Thus k_2 is effective only within the arbor radius.

Let $\widehat{K}(\mathbf{n})$ be the Fourier transform of $k_2(\mathbf{x})$, then $\widehat{K}(\mathbf{n})$ can be computed as follows.

$$\widehat{K}(\mathbf{n}) = \int_{-\infty}^{\infty} k_2(\mathbf{x})e^{-i\mathbf{n}\mathbf{x}}d\mathbf{x} = 2\pi k_2 R\frac{J_1(|\mathbf{n}|R)}{|\mathbf{n}|}.$$

This function takes the maximum value $2\pi k_2 R$ at the center and decreases to 0 toward to points $\mathbf{n}$ where $|\mathbf{n}| \approx 3.8/R \approx \pi/R$. After that this function oscillates and finally converges to 0. Thus, we could regard $\widehat{K}(\mathbf{n})$ as a sharp uni-modal function and that is "effective" only within the range $|\mathbf{n}| \leq \sigma$ (Fig. 8 left). Hence even if the $\widehat{C}(\mathbf{n})$ is uni-modal, by taking k_2 appropriately, we can make $\widehat{(C+K)}(\mathbf{n})$ bi-modal (Fig. 8 right). Moreover, the peak of $\widehat{(C+K)}(\mathbf{n})$ appears around $|\mathbf{n}| \approx \sigma$.

6 Conclusion

We considered Linsker's feed-forward neural network model and studied mathematically how and why this model develops an oriented receptive field. We simplified this model and performed the Fourier transform to convert the convolution into a simple multiplication.

Through our mathematical analysis, we made the following observations: (1) the correlation function; that is a DOG function, extracts Fourier components of a certain frequency,

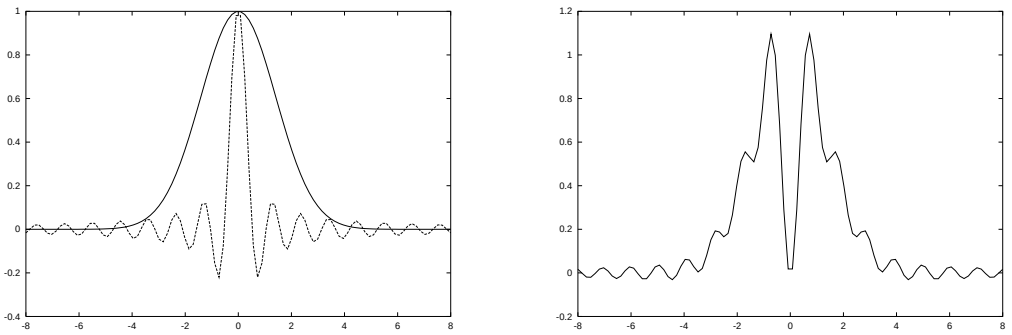

Figure 8: Left: The shape of $\widehat{C}(\mathbf{n})$ (real line) and that of $\widehat{K}(\mathbf{n})$ (dashed line). We assume that $\widehat{C}(\mathbf{n})$ is uni-modal. Right: the shape of $(\widehat{C - K})(\mathbf{n})$ for an appropriate k_2. As can be seen, the shape is bi-modal.

(2) the arrangement of Fourier coefficients converges to a circle, and the circle has a certain slant. (3) (2) implies the shape of the Fourier cosine component changes into a concentric circle while that of the Fourier sine component is bi-lobed of which the orientation is determined by the maximum Fourier coefficient. (4) the arbor function determines either the Fourier sine component or the Fourier cosine component dominates the other. Thus, (3) and (4) suggests that if the Fourier cosine component is dominant, a circular symmetric receptive field is developed, on the other hand, if the Fourier sine component is dominant, in this case a bi-lobed receptive field is obtained.

These observations were also justified by our computer simulations, and finally we discussed the role of Linsker's parameters k_1 and k_2.

Acknowledgments

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

References

Feng, J., Pan, H., Roychowdhury, V. (1997). Linsker-type Hebbian Learning: A Qualitative Analysis on the Parameter Space. *Neural Network*, 10, 705–720.

Kammen, D., Yuille, A. (1988). Spontaneous Symmetry-Breaking Energy Functions and the Emergence of Orientation Selective Cortical Cells. *Biological Cybernetics*, 59, 23-31.

Linsker, R. (1986). From basic network principles to neural architecture (series). *Proc. Natl. Acad. Sci. USA*, 83, 7508–12, 8390–4, 8779–83.

Linsker, R. (1990). Self-Organization in a Perceptual System: How Network Models and Information Theory May Shed Light on Neural Organization. In S. Hanson and C. Olson, eds., *Connectionist Modeling and Brain Function*, MIT Press, 351-392.

Mackay, D., Miller, K. (1990). Analysis of Linsker's application of Hebbian rules to linear networks. *Network*, 1, 257–297.

Matsumoto, M., Nishimura, T. (1998). Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Trans. on Modeling and Computer Simulation*, 8, 3-30. (The source code can be retrieved from `http://www.math.keio.ac.jp/~matumoto/emt.html`)

Miller, K. (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.

Miyashita, M., Tanaka, S. (1992). A Mathematical Model for the Self-Organization of Orientation Columns in Visual Cortex. *NeuroReport*, 3 , 69–72.

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

Yamazaki, T. (2000). A Mathematical Analysis of Development of Oriented Receptive Fields in Linsker's Model. *Research Report C-144*, Department of Mathematical and Computing Sciences, Tokyo Institute of Technology, 2000;
`http://www.is.titech.ac.jp/research/research-report/C/index.html`

Appendix

Proof of Proposision 1.

By induction of t, we will show that

$$w^{(t)}(\mathbf{x}) = A(\mathbf{x})w_*^{(t)}(\mathbf{x}).$$

Firstly, from the assignment of the initial weights, it follows that

$$w^{(0)}(\mathbf{x}) = A(\mathbf{x})w_*^{(0)}(\mathbf{x}).$$

Next, suppose that $w^{(t-1)}(\mathbf{x}) = A(\mathbf{x})w_*^{(t-1)}(\mathbf{x})$ holds. Then

$$\begin{aligned}
w^{(t)}(\mathbf{x}) &= w^{(t-1)}(\mathbf{x}) + \Delta w^{(t-1)}(\mathbf{x}) \\
&= A(\mathbf{x})w_*^{(t-1)}(\mathbf{x}) \\
&\quad + A(\mathbf{x})\left(k_1 + \int_{-\infty}^{\infty} \left(C(\mathbf{x}-\mathbf{y}) + k_2\right) w^{(t-1)}(\mathbf{y})d\mathbf{y},\right) \\
&= A(\mathbf{x})w_*^{(t-1)}(\mathbf{x}) \\
&\quad + A(\mathbf{x})\left(k_1 + \int_{-\infty}^{\infty} \left(C(\mathbf{x}-\mathbf{y}) + k_2\right) A(\mathbf{x})w_*^{(t-1)}(\mathbf{y})d\mathbf{y}\right), \\
&= A(\mathbf{x})w_*^{(t-1)}(\mathbf{x}) + A(\mathbf{x})\Delta w_*^{(t-1)}(\mathbf{x}) \\
&= A(\mathbf{x})w_*^{(t)}(\mathbf{x}).
\end{aligned}$$

□

Proof of Lemma 1.

$$\begin{aligned}
\widehat{C}(\mathbf{n}) &= \int_{-\infty}^{\infty} \left(K_1 e^{-\frac{|\mathbf{x}|^2}{2\sigma_1^2}} - K_2 e^{-\frac{|\mathbf{x}|^2}{2\sigma_2^2}}\right) e^{-i\mathbf{n}\cdot\mathbf{x}} d\mathbf{x} \\
&= K_1 \int_{-\infty}^{\infty} e^{-\frac{|\mathbf{x}|^2 - 2\sigma_1^2 i\mathbf{n}\cdot\mathbf{x}}{2\sigma_1^2}} d\mathbf{x} - K_2 \int_{-\infty}^{\infty} e^{-\frac{|\mathbf{x}|^2 - 2\sigma_2^2 i\mathbf{n}\cdot\mathbf{x}}{2\sigma_2^2}} d\mathbf{x} \\
&= K_1 \int_{-\infty}^{\infty} e^{-\frac{(\mathbf{x}-i\sigma_1^2\mathbf{n})^2 - (i\sigma_1^2\mathbf{n})^2}{2\sigma_1^2}} d\mathbf{x} - K_2 \int_{-\infty}^{\infty} e^{-\frac{(\mathbf{x}-i\sigma_2^2\mathbf{n})^2 - (i\sigma_2^2\mathbf{n})^2}{2\sigma_2^2}} d\mathbf{x} \\
&= K_1 e^{-\frac{\sigma_1^2|\mathbf{n}|^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(\mathbf{x}-i\sigma_1^2\mathbf{n})^2}{2\sigma_1^2}} d\mathbf{x} - K_2 e^{-\frac{\sigma_2^2|\mathbf{n}|^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{(\mathbf{x}-i\sigma_2^2\mathbf{n})^2}{2\sigma_2^2}} d\mathbf{x},
\end{aligned}$$

where

$$\begin{aligned}
\int_{-\infty}^{\infty} e^{-\frac{(\mathbf{x}-i\sigma_1^2\mathbf{n})^2}{2\sigma_1^2}} d\mathbf{x} &= \int_{-\infty}^{\infty} e^{-\frac{\mathbf{X}^2}{2\sigma_1^2}} d\mathbf{X} \\
&= \int_0^{2\pi}\int_0^{\infty} e^{-\frac{r^2}{2\sigma_1^2}} r dr d\theta \\
&= 2\pi\left[(-\sigma_1^2)e^{-\frac{r^2}{2\sigma_1^2}}\right]_0^{\infty} \\
&= 2\pi\sigma_1^2.
\end{aligned}$$

It follows

$$\widehat{C}(\mathbf{n}) = 2\pi\sigma_1^2 K_1 e^{-\frac{\sigma_1^2|\mathbf{n}|^2}{2}} - 2\pi\sigma_2^2 K_2 e^{-\frac{\sigma_2^2|\mathbf{n}|^2}{2}}.$$

□

Proof of Lemma 2.

Firstly we show that for any t,

$$\Delta w^{(t)}(\mathbf{x}) = \left(\frac{C_*}{4\pi^2}\right) \sum_{j=0}^{t} \left(\frac{C_*}{4\pi^2}\right)^j \binom{t}{j} \int_{|\mathbf{n}|=\sigma} \widehat{AW}^{(0)} * \widetilde{A}^{(j)_*}(\mathbf{n}) e^{i\mathbf{n}\mathbf{x}} d\mathbf{n}.$$

In the case of $t = 0$, it follows

$$\Delta w^{(0)}(\mathbf{x}) = \left(\frac{C_*}{4\pi^2}\right) \int_{|\mathbf{n}_0|=\sigma} \widehat{Aw}^{(0)}(\mathbf{n}_0) e^{i\mathbf{n}_0\cdot\mathbf{x}} d\mathbf{n}_0.$$

Suppose that it holds in the case of $t-1$ it, and we show the case of t.

$$
\begin{aligned}
\Delta w^{(t)}(\mathbf{x}) &= \left(\frac{C_*}{4\pi^2}\right)\int_{|\mathbf{n}_t|=\sigma}\widehat{Aw}^{(t)}(\mathbf{n}_t)e^{i\mathbf{n}_t\cdot\mathbf{x}}d\mathbf{n}_t \\
&= \left(\frac{C_*}{4\pi^2}\right)\int_{|\mathbf{n}_t|=\sigma}\left(\int_{-\infty}^{\infty}A(\mathbf{y})w^{(t)}(\mathbf{y})e^{-\mathbf{y}\cdot\mathbf{n}_t}d\mathbf{y}\right)e^{i\mathbf{n}_t\cdot\mathbf{x}}d\mathbf{n}_t \\
&= \left(\frac{C_*}{4\pi^2}\right)\int_{|\mathbf{n}_t|=\sigma}\left(\int_{-\infty}^{\infty}A(\mathbf{y})(w^{(t)}(\mathbf{y})+\Delta w^{(t-1)}(\mathbf{y}))e^{-\mathbf{y}\cdot\mathbf{n}_t}d\mathbf{y}\right)e^{i\mathbf{n}_t\cdot\mathbf{x}}d\mathbf{n}_t \\
&= \Delta w^{(t-1)}(\mathbf{x})+\left(\frac{C_*}{4\pi^2}\right)\int_{|\mathbf{n}_t|=\sigma}\left(\int_{-\infty}^{\infty}A(\mathbf{y})\Delta w^{(t-1)}(\mathbf{y})e^{-\mathbf{y}\cdot\mathbf{n}_t}d\mathbf{y}\right)e^{i\mathbf{n}_t\cdot\mathbf{x}}d\mathbf{n}_t \\
&= \Delta w^{(t-1)}(\mathbf{x}) \\
&+ \left(\frac{C_*}{4\pi^2}\right)\int_{|\mathbf{n}_t|=\sigma}\int_{-\infty}^{\infty}A(\mathbf{y})\left(\frac{C_*}{4\pi^2}\right)\sum_{j=0}^{t-1}\left(\frac{C_*}{4\pi^2}\right)^j\binom{t-1}{j} \\
&\quad \int_{|\mathbf{n}_{t-1}|=\sigma}\widehat{AW}^{(0)}*\widetilde{A}^{(j)*}(\mathbf{n}_{t-1})e^{i\mathbf{n}_{t-1}\mathbf{y}}d\mathbf{n}_{t-1}e^{-\mathbf{y}\cdot\mathbf{n}_t}d\mathbf{y}e^{i\mathbf{n}_t\cdot\mathbf{x}}d\mathbf{n}_t \\
&= \Delta w^{(t-1)}(\mathbf{x}) \\
&+ \left(\frac{C_*}{4\pi^2}\right)^2\int_{|\mathbf{n}_t|=\sigma}\int_{|\mathbf{n}_{t-1}|=\sigma}\sum_{j=0}^{t-1}\left(\frac{C_*}{4\pi^2}\right)^j\binom{t-1}{j}\widehat{AW}^{(0)}*\widetilde{A}^{(j)*}(\mathbf{n}_{t-1}) \\
&\quad \int_{-\infty}^{\infty}A(\mathbf{y})e^{i\mathbf{n}_{t-1}\mathbf{y}}e^{-\mathbf{y}\cdot\mathbf{n}_t}d\mathbf{y}d\mathbf{n}_{t-1}e^{i\mathbf{n}_t\cdot\mathbf{x}}d\mathbf{n}_t \\
&= \Delta w^{(t-1)}(\mathbf{x}) \\
&+ \left(\frac{C_*}{4\pi^2}\right)^2\int_{|\mathbf{n}_t|=\sigma}\sum_{j=0}^{t-1}\left(\frac{C_*}{4\pi^2}\right)^j\binom{t-1}{j}\widehat{AW}^{(0)}*\widetilde{A}^{(j+1)*}(\mathbf{n}_t)e^{i\mathbf{n}_t\cdot\mathbf{x}}d\mathbf{n}_t \\
&= \left(\frac{C_*}{4\pi^2}\right)\sum_{j=0}^{t-1}\left(\frac{C_*}{4\pi^2}\right)^j\binom{t-1}{j}\int_{|\mathbf{n}_t|=\sigma}\widehat{AW}^{(0)}*\widetilde{A}^{(j)*}(\mathbf{n}_t)e^{i\mathbf{n}_t\mathbf{x}}d\mathbf{n}_t \\
&+ \left(\frac{C_*}{4\pi^2}\right)^2\int_{|\mathbf{n}_t|=\sigma}\sum_{j=0}^{t-1}\left(\frac{C_*}{4\pi^2}\right)^j\binom{t-1}{j}\widehat{AW}^{(0)}*\widetilde{A}^{(j+1)*}(\mathbf{n}_t)e^{i\mathbf{n}_t\cdot\mathbf{x}}d\mathbf{n}_t \\
&= \left(\frac{C_*}{4\pi^2}\right)\sum_{j=0}^{t-1}\left(\frac{C_*}{4\pi^2}\right)^j\binom{t-1}{j}\int_{|\mathbf{n}_t|=\sigma}\widehat{AW}^{(0)}*\widetilde{A}^{(j)*}(\mathbf{n}_t)e^{i\mathbf{n}_t\mathbf{x}}d\mathbf{n}_t \\
&+ \left(\frac{C_*}{4\pi^2}\right)\int_{|\mathbf{n}_t|=\sigma}\sum_{j=1}^{t}\left(\frac{C_*}{4\pi^2}\right)^j\binom{t-1}{j-1}\widehat{AW}^{(0)}*\widetilde{A}^{(j)*}(\mathbf{n}_t)e^{i\mathbf{n}_t\cdot\mathbf{x}}d\mathbf{n}_t \\
&= \left(\frac{C_*}{4\pi^2}\right)\int_{|\mathbf{n}_0|=\sigma}\widehat{AW}^{(0)}e^{i\mathbf{n}_0\mathbf{x}}d\mathbf{n}_0 \\
&+ \left(\frac{C_*}{4\pi^2}\right)\sum_{j=1}^{t-1}\left(\frac{C_*}{4\pi^2}\right)^j\left(\binom{t-1}{j}+\binom{t-1}{j-1}\right)\int_{|\mathbf{n}_t|=\sigma}\widehat{AW}^{(0)}*\widetilde{A}^{(j)*}(\mathbf{n}_t)e^{i\mathbf{n}_t\mathbf{x}}d\mathbf{n}_t \\
&+ \left(\frac{C_*}{4\pi^2}\right)^{t+1}\int_{|\mathbf{n}_t|=\sigma}\widehat{AW}^{(0)}*\widetilde{A}^{(t)*}(\mathbf{n}_t)e^{i\mathbf{n}_t\mathbf{x}}d\mathbf{n}_t \\
&= \left(\frac{C_*}{4\pi^2}\right)\sum_{j=0}^{t}\left(\frac{C_*}{4\pi^2}\right)^j\binom{t}{j}\int_{|\mathbf{n}_t|=\sigma}\widehat{AW}^{(0)}*\widetilde{A}^{(j)*}(\mathbf{n}_t)e^{i\mathbf{n}_t\mathbf{x}}d\mathbf{n}_t.
\end{aligned}
$$

Then, it follows

$$\begin{aligned}\Delta w^{(t)}(\mathbf{x}) &= \left(\frac{C_*}{4\pi^2}\right)\sum_{j=0}^{t}\left(\frac{C_*}{4\pi^2}\right)^j\binom{t}{j}\int_{|\mathbf{n}|=\sigma}\widehat{AW}^{(0)} * \widetilde{A}^{(j)_*}(\mathbf{n})e^{i\mathbf{nx}}d\mathbf{n} \\ &= \left(\frac{C_*}{4\pi^2}\right)\int_{|\mathbf{n}|=\sigma}\widehat{AW}^{(0)} * \left(1+\left(\frac{C_*}{4\pi^2}\right)\widetilde{A}\right)^{(t)_*}(\mathbf{n})e^{i\mathbf{nx}}d\mathbf{n}.\end{aligned}$$

□

Proof of Lemma 3.

A is Gaussian.

$$\begin{aligned}\widetilde{A}(\mathbf{n}) &= \int_{-\infty}^{\infty} e^{-\frac{|\mathbf{x}|^2}{2\sigma_A^2}}e^{i\mathbf{nx}}d\mathbf{x} \\ &= \int_{-\infty}^{\infty} e^{-\frac{|\mathbf{x}|^2-2\sigma_A^2 i\mathbf{x}\cdot\mathbf{n}}{2\sigma_A^2}}d\mathbf{x} \\ &= \int_{-\infty}^{\infty} e^{-\frac{(|\mathbf{x}|^2-\sigma_A^2 i\mathbf{n})^2-(\sigma_A^2 i\mathbf{n})^2}{2\sigma_A^2}}d\mathbf{x} \\ &= e^{-\frac{\sigma_A^2|\mathbf{n}|^2}{2\sigma_A^2}}\int_{-\infty}^{\infty} e^{-\frac{(|\mathbf{x}|^2-\sigma_A^2 i\mathbf{n})^2}{2\sigma_A^2}}d\mathbf{x} \\ &= e^{-\frac{|\mathbf{n}|^2}{2}}\int_{-\infty}^{\infty} e^{-\frac{|\mathbf{x}|^2}{2\sigma_A^2}}d\mathbf{x} \\ &= 2\pi\sigma_A^2 e^{-\frac{|\mathbf{n}|^2}{2}}.\end{aligned}$$

□

A is Step.

$$
\begin{aligned}
\widetilde{A}(\mathbf{n}) &= \int_{|\mathbf{x}|=R} e^{i\mathbf{n}\mathbf{x}} d\mathbf{x} \\
&= \int_{|\mathbf{x}|=R} \cos(\mathbf{n}\mathbf{x}) + i\sin(\mathbf{n}\mathbf{x}) d\mathbf{x} \\
&= \int_{|\mathbf{x}|=R} \cos(\mathbf{n}\mathbf{x}) d\mathbf{x} \\
&= \int_{|\mathbf{x}|=R} \cos(|\mathbf{n}||\mathbf{x}|\cos\theta_{\mathbf{n},\mathbf{x}}) d\mathbf{x} \\
&= \int_0^R \int_0^{2\pi} \cos(|\mathbf{n}| r\cos\theta) r dr d\theta \\
&= \int_0^R r dr \int_{-\pi} \pi\cos(|\mathbf{n}| r\cos\theta) d\theta \\
&= 2\int_0^R r dr \int_0^{\pi} \pi\cos(|\mathbf{n}| r\cos\theta) d\theta \\
&\quad \left(\int_0^{\pi} \pi\cos(z\cos x)\cos nx dx = \pi\cos(n\pi/2) J_n(z). \right) \\
&= 2\int_0^R r dr \cdot \pi J_0(|\mathbf{n}| r) \\
&= 2\pi \int_0^R r J_0(|\mathbf{n}| r) dr \\
&\quad \left(\int x J_n(\alpha x) J_n(\beta x) dx = \frac{x(\alpha J_n(\beta x) J_n(\alpha x) - \beta J_n'(\alpha x) J_n'(\beta x))}{\beta^2 - \alpha^2}. \right) \\
&= 2\pi \left. \frac{r|\mathbf{n}| J_0'(|\mathbf{n}| r)}{-|\mathbf{n}|^2} \right|_0^R \\
&= 2\pi R \frac{J_0'(|\mathbf{n}| R)}{|\mathbf{n}|} \\
&\quad \left(J_{n-1}'(x) = J_n(x). \right) \\
&= 2\pi R \frac{J_1(|\mathbf{n}| R)}{|\mathbf{n}|}.
\end{aligned}
$$

□