Institut für Theoretische Physik
Lehrstuhl Univ.-Prof. Dr. J. Leo van Hemmen
der Technischen Universität München

Orientation Maps in Primary Visual Cortex: A Hebbian Model of Intracortical and Geniculocortical Plasticity

Armin P. Bartsch

Vollständiger Abdruck der von der Fakultät für Physik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. Erich Sackmann
Prüfer der Dissertation:

1. Univ.-Prof. Dr. J. Leo van Hemmen
2. Univ.-Prof. Dr. Paul Tavan
   Ludwig-Maximilians-Universität München

Die Dissertation wurde am 26.04.2000 bei der Technischen Universität München eingereicht und durch die Fakultät für Physik am 02.04.2001 angenommen.

Contents

Introduction

The visual system is one of the most extensively studied parts of the mammalian brain. Nevertheless, it is still largely unknown how visual information is processed in the cortex so that animals and humans can navigate with apparent ease in a complex environment. Understanding the underlying cortical circuitry would probably open new ways to treat brain diseases and would as well solve a difficult problem in the design of autonomous robots, namely vision.

The cortical area specialized for vision is called the visual cortex. The primary visual cortex is the subregion where visual information is first relayed into the cortex. In a pioneering series of experiments, Hubel and Wiesel have shown that neurons in the primary visual cortex respond to stimulation within a certain region on the retina, the so-called receptive field. Most cells respond preferentially to bar-like stimuli of a specific orientation and are activated predominantly through one of the two eyes (Hubel and Wiesel 1962).

The retinal position of the receptive field, the preferred orientation, and the ocular dominance of recorded neurons are graded functions of the cortical position of the recording site. Therefore, the relations between cortical location and the various neuronal response properties are referred to as cortical maps, viz., the retinotopic, the orientation preference, and the ocular dominance map. Current experimental data indicate that the final layout of these maps is shaped partially by neuronal activity and visual experience during a certain critical period early in life (Crair et al. 1998, Issa et al. 1999).

Several theoretical studies (von der Malsburg 1973, Linsker 1986a, Linsker 1986b, Miller 1994, Wimbauer et al. 1997a, Wimbauer et al. 1997b, Miller et al. 1999) have proposed that cortical orientation preference may arise from a Hebbian learning mechanism (Hebb 1949) driven by correlated activity of the afferent neurons. In the presence of a circularly symmetric lateral wiring within the cortex, these models predict the emergence of a nicely ordered orientation map that is formed by the resulting arrangement of afferent connections.

The afferent cells are located in the lateral geniculate nucleus (LGN), where visual information is relayed from the eyes to the cortex. As opposed to most cortical neurons, geniculate cells are monocular, i.e., they are activated exclusively by one eye. Hence, the primary visual cortex is the first stage along the visual pathway to combine input from both eyes, and so there are, precisely speaking, two separate orientation maps, one for either eye.

As such, the above models of correlation-based map formation have been challenged by reverse-lid-suturing experiments (Gödecke and Bonhoef-

fer 1996, Sengpiel et al. 1998) in which kittens are raised so that both eyes do never receive common visual input. If the development of geniculocortical afferents were driven by activity correlations in the LGN and these correlations were mainly determined by the correlations of the visual input, then the two maps would form independently and could be expected to be different. Optical recordings, however, show them to be nearly identical.

Recently, it has been demonstrated that the emergence of ocularly matched orientation maps can be explained within the framework of correlation-based development, if an appropriate amount of geniculate inter-eye activity correlations is assumed (Erwin and Miller 1998). Although strong inter-eye correlations have been found to be present in the ferret's LGN before eye opening (Weliky and Katz 1999), it is still unclear whether this model can actually account for the outcome of the reverse-lid-suturing experiments.

An alternative approach is based on a geometric argument (Wolf et al. 1996). As the authors have pointed out, the cortical region in which the experiments have been carried out is shaped as a narrow band. The process of pattern formation in this area may therefore be subject to strong confinement, i.e., the layout of the developing orientation map could be predetermined by the boundary conditions.

In this work we propose a model of layer 4 of the primary visual cortex consisting of laterally interconnected excitatory and inhibitory *spiking* neurons. It combines the idea of correlation-based learning of geniculocortical afferents with Hebbian development of short-range intracortical connections.

Using large-scale computer simulations we demonstrate that in this kind of network the spontaneous activity of cortical cells alone can drive the plasticity of lateral connections to produce a connectivity pattern whose overall layout closely resembles that of a measured orientation map. The formation of such an *intracortical* orientation map is independent of feedforward input from the LGN. It could therefore occur at an early stage of visual development, when geniculate afferents have not yet innervated cortical layer 4.

Furthermore, we show that a correlation-based development of the geniculate afferents can interact with the pattern of intracortical connections so that the emerging *feedforward* orientation map and the intracortical map will be in accord with each other. Thus, the layout of the feedforward map is more or less predetermined once the intracortical connectivity is fully established and remains fixed. This provides a very natural explanation for the remarkable stability of orientation maps that has been found in various experiments (Kim and Bonhoeffer 1994, Gödecke and Bonhoeffer 1996, Weliky and Katz 1997, Sengpiel et al. 1998, Sengpiel et al. 1999).

An analytic investigation of the network's learning dynamics helps to understand the mechanisms leading to these interesting results. In our model, the efficacy of each neuronal contact is modified in dependence on the relative timing of the spikes emitted by the cells it is connecting, and so the plasticity is driven mainly by spike-spike correlations in the network's activity. Hence, we first present a mathematical analysis of the spike statistics in recurrent networks of spiking cells. We derive approximative expressions for the mean activity and the two-spike correlation function and compare the results with a

series of numerical simulations.

Then we use the obtained approximations to study the learning process in a simplified version of our cortical network model. The simplified network comprises linear excitatory but no inhibitory neurons. Starting from spike-based learning rules, we derive two systems of non-linear differential equations to describe the temporal evolution of the feedforward and of the lateral connectivity. By means of a linear stability analysis we show how a feedforward orientation map can emerge and how the development of the map can be guided by an anisotropic intracortical connectivity pattern. As for the plasticity of lateral connections it turns out that the simplified model cannot account for the formation of intracortical orientation maps – probably because it lacks inhibitory neurons to limit the overall network activity.

The thesis is organized as follows. Chapter 1 provides a short survey of the visual pathways from the eye to the cortex and of their development during ontogeny. In Chapter 2 we then introduce our model of the primary visual cortex and examine the synaptic plasticity in this model using numerical simulations. Chapter 3 is devoted to a mathematical investigation of the spike statistics in recurrently interconnected networks of spiking neurons. The results are required in Chapter 4, where we present an analytical study of the synaptic dynamics using a simplified version of our model.

Chapter 1

Central Visual Pathways

The developmental model we shall discuss throughout the present work is concerned with processes of self-organization in the mammalian primary visual cortex. The following chapter is intended to introduce the most important biological concepts required for an understanding of this model. To this end, we will examine the flow of visual information along the visual pathway from the eye to the cortex and survey the ontogenetic development of the related anatomical structures. Of course, we will consider the most important stages only; for a detailed exposition the reader is referred to standard textbooks of (visual) neuroscience like those of Buser and Imbert (1992), Kandel et al. (1997), Nicholls et al. (1992), Thompson (1993), and Zeki (1994).

An overall view of the visual pathways from the retina to the primary visual cortex is given in Fig. 1.1. Light entering the eyes through the lenses is transformed into electrical signals by the retinae. In particular, light originating from the left half of the outside visual field, i.e., from the left visual hemifield reaches the right part of either eye's retina, whereas light originating from the right hemifield is mapped onto the left hemiretinae. The resulting electrical signals leave the eyes along the optic nerves. At the optic chiasm the nerve fibres are partially crossed so that the fibres from both right hemiretinae are merged into the right optic tract while the fibres from the left hemiretinae are bundled into the left optic tract. Each optic tract thus contains a complete representation of one visual hemifield.

The lateral geniculate nuclei (LGN) are the first relay stations of visual information outside the eyes. Either LGN is innervated by one optic tract and, accordingly, receives information from one visual hemifield. Therefore, the left visual hemifield is represented in the right LGN whereas the right hemifield is represented in the left LGN. The output fibres of the LGN, called the optic radiation, terminate in layer 4 of the primary visual cortex. Each LGN projects to the primary visual cortex of its own side so that again either part of the visual cortex receives a representation of the contralateral part of the visual field.

Before the involved structures and their function can be described in more detail we must investigate the elements they consist of, the individual nerve cells. We shall do this in the next section. After that we we will have a closer look at the retina, the lateral geniculate nucleus, and the primary visual cortex

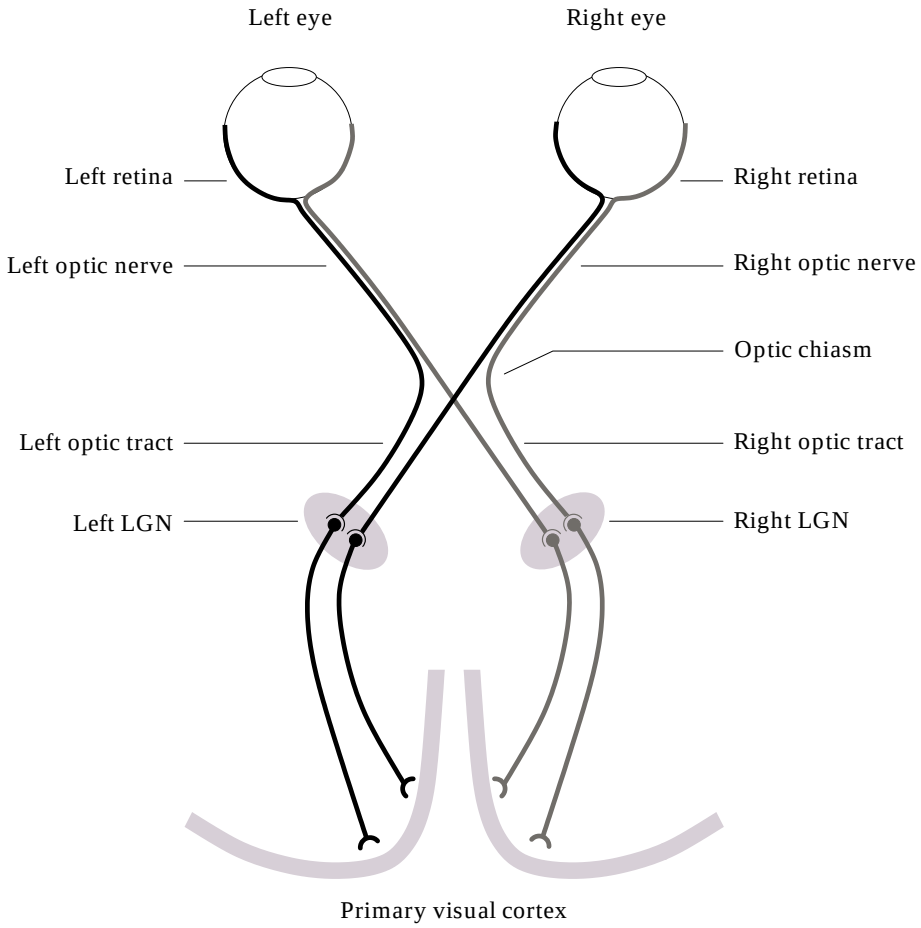


Figure 1.1: The visual pathways from the retina to the primary visual cortex are shown schematically. Light originating from the left half of the visual field reaches the right part of either retina, and vice versa. The electrical signals produced by the retinae leave the eyes along the optic nerves. At the optic chiasm the nerve fibres are partially crossed so that the fibres from both right hemiretinae and those from both left hemiretinae are merged into the right and the left optic tract, respectively. The optic tracts innervate the lateral geniculate nuclei (LGN), which are the first relay stations outside the eyes. The output fibres of the LGN, called the optic radiation, terminate in the primary visual cortex. As a consequence of this connectivity, the left LGN and the left part of the visual cortex receive a complete representation of the right visual hemifield, and vice versa. (After Nicholls et al. 1992.)

in Sects. 1.2 through 1.4. Section 1.5 will then give a survey of the developmental stages that will be relevant to our model of cortical self-organization. These include the emergence and the maturation of the cortical layers and the establishment of the geniculocortical connectivity, i.e., the projections from the LGN to the primary visual cortex. Finally, we shall summarize the main findings of this chapter in Sect. 1.6.

1.1 Neurons

The individual nerve cells, or *neurons*, are the building blocks of the central nervous system and are commonly thought to be the basic units of biological information processing. In spite of their conspicuous appearance, nerve cells are like normal cells in many respects. The major differences to other cells arise from their special function of transforming and transmitting electrical signals.

1.1.1 Structure and Electrical Properties

A typical cortical neuron is shown in panel (a) of Fig. 1.2, a more schematic view is given in panel (b). The cell body, called *soma*, contains the nucleus and the same organelles as other cells. Several fibres extend from the soma in different directions. Most of these fibres are to receive signals from other neurons. They are called *dendrites* and form a tree-like structure, the *dendritic tree*. Exactly one of the fibres is the *axon*, along which the neuron's output signals are conducted to their targets. Once the axon has left the cell body, it may split and send out numerous *axonal collaterals* so as to reach different targets at different locations of the nervous system.

Nerve cells transform and transmit electrical signals. The interior of a neuron that is not activated remains at a *resting membrane potential* of about -65 mV with respect to the outside. This potential results from differences in the concentrations of several types of ions between the intracellular and the extracellular medium. The most important types of ions include Na^+, K^+, Cl^-, and organic anions. The unequal distribution of ions across the cell membrane is established and maintained by an active mechanism, the metabolically driven Na^+-K^+ pump, which transports Na^+ from the inside of the cell to the outside and K^+ from the outside to the inside. The resting potential provides the baseline for the electrical signals traveling along the cell membrane. When the membrane potential becomes more negative then the membrane is said to be *hyperpolarized*, whereas a shift of the potential towards zero or positive values is called *depolarization*.

A pulse of current that is injected into one of the dendrites – e.g., as a result of an incoming signal from another neuron – will cause a transient change of the membrane potential to spread over the neuron and will thus induce a transient depolarization or hyperpolarization of the cell body. It is commonly assumed that the dendrites behave like passive cables so that the propagation of electrical signals along the dendrites can be described using linear cable theory. As a consequence, the perturbation of the membrane potential at the

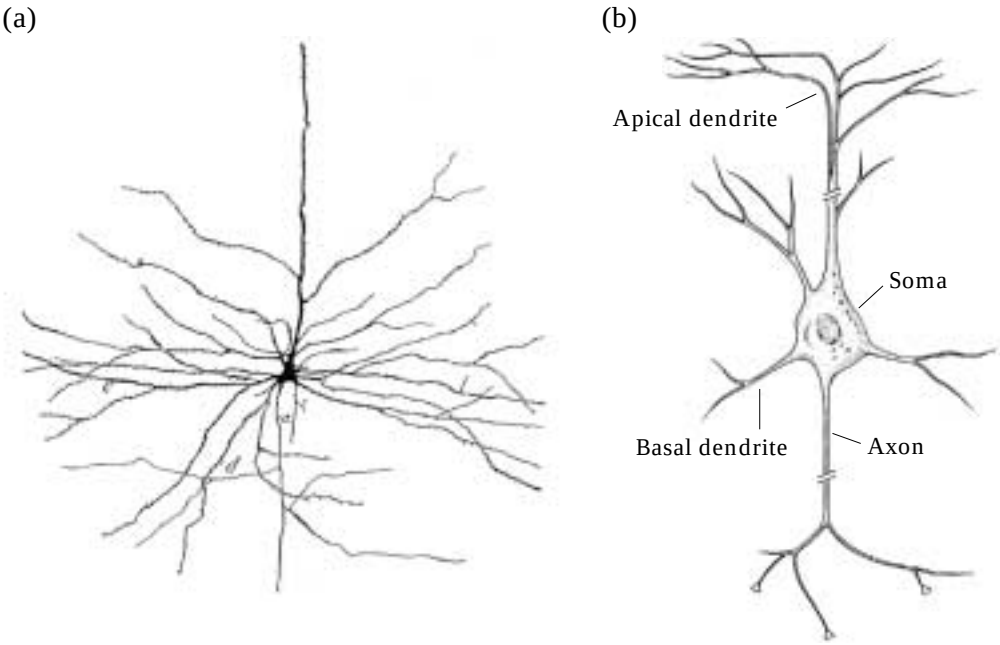


Figure 1.2: Panel (a) visualizes a typical cortical neuron in a drawing by y Cajal (1911). A more schematic presentation (adapted from Kandel et al. 1997) is given in panel (b). The cell body, or soma, contains the nucleus and the same organelles as other cells. Several fibres extend from the soma, most of which are so-called dendrites. They are to receive input from other nerve cells. Exactly one of the fibres, however, is the axon, along which the neuron's output signals are conducted to their targets. The axon may split and send out numerous collaterals so as to reach different targets at different locations in the nervous system.

soma in response to multiple dendritic current injections is simply the sum of the responses to the individual injections. In this sense, the dendritic tree can be regarded as an integrator where all the incoming signals are added up linearly.

In contrast, the propagation of electrical signals along the axon is an active and non-linear process. When the membrane potential at the soma is depolarized to a value of about -55 mV or beyond, then an *action potential*, or *spike*, is generated at the initial segment of the axon. An action potential is a large but very short depolarizing pulse in the membrane voltage. It lasts about 1 ms and reaches a maximal membrane voltage of some $+30$ mV. Action potentials are all-or-nothing events: their time course is stereotyped so that there is no information carried in their shape or amplitude. They emerge as the result of an exactly timed interplay between voltage-gated Na^+ and K^+ channels that are embedded in the axonal cell membrane. Once an action potential has been initiated at the soma it travels down the axon without changing its shape or amplitude.

1.1.2 Synapses

How are the output signals, i.e., the action potentials of a nerve cell transmitted to other neurons? In order to answer this question let us have a look at the interface, called *synapse*, between two nerve cells. Two fundamentally different types of synapses, electrical and chemical synapses have been found. At an electrical synapse the intracellular media of the involved neurons are coupled via gap-junction channels that enable a direct current flow. At a chemical synapse, on the other hand, the incoming electrical pulses are first transformed into chemical signals, which are then reconverted into electrical signals in the target cell. As chemical synapses represent the predominant type of synapses in the mammalian brain, the subsequent explanations will be restricted to this type.

A chemical synapse usually connects the axon of one cell with a dendrite or the soma of another cell. It is composed of two major components as sketched in panel (a) of Fig. 1.3. The *presynaptic* part is formed by the axon terminal of the incoming neuron. The *postsynaptic* component is formed by a specialized region of the target cell's membrane. The pre- and postsynaptic membranes are separated by the *synaptic cleft*, a tiny space about 20 nm wide. The axon terminal of the presynaptic neuron holds a large number of *vesicles* containing the synaptic transmitter substance. The postsynaptic membrane is enriched with specialized ion channels that are gated by chemical receptor molecules.

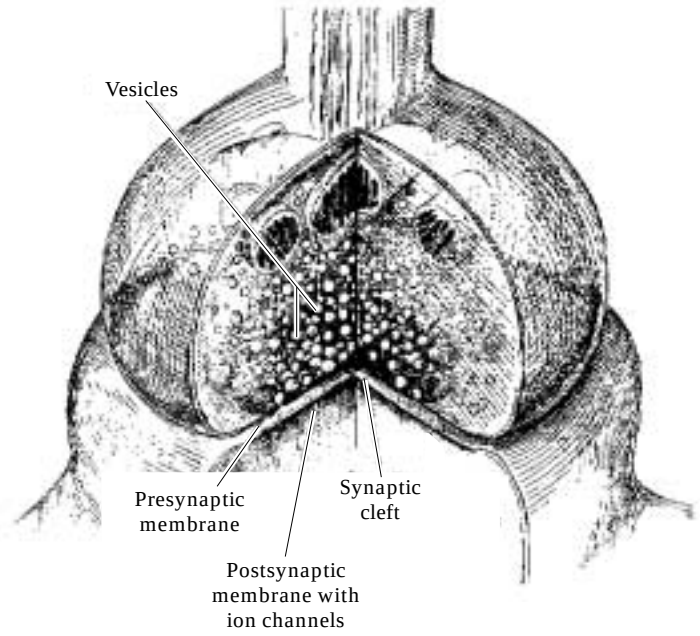
When a spike arrives at the presynaptic axon terminal, part of the vesicles fuse with the membrane and release their content of synaptic transmitter into the cleft; see panel (b) of Fig. 1.3. The transmitter molecules diffuse across the cleft to the postsynaptic membrane where they attach to the receptor molecules. In response, the connected ion channels are opened for a short period of time. The resulting ionic current depolarizes or hyperpolarizes the postsynaptic membrane and thus induces an electrical signal in the postsynaptic cell.

According to its effect onto the postsynaptic neuron, a synapse is categorized as *excitatory* or *inhibitory*. The activation of a depolarizing synapse induces a transient change of the postsynaptic membrane potential towards positive values. For the duration of this depolarization, the postsynaptic neuron is left closer to the firing threshold and is therefore more easily excited to fire an action potential. Correspondingly, a depolarizing synapse is called excitatory and the induced transient depolarization is called an *excitatory postsynaptic potential* or, in short, *EPSP*.

Conversely, a hyperpolarizing synapse is called inhibitory because its activation gives rise to a transient postsynaptic hyperpolarization, the *inhibitory postsynaptic potential (IPSP)*, which may inhibit the emission of an action potential by the postsynaptic cell. In the mammalian brain, the most important types of neurotransmitter used at inhibitory synapses are *gamma-aminobutyric acid (GABA)* and *glycine*, while *glutamate* is the most common type of excitatory neurotransmitter.

Current experimental data strongly suggest that in the cortex all outgoing synapses of one individual cell are of the same kind – excitatory or inhibitory

(a)



(b)

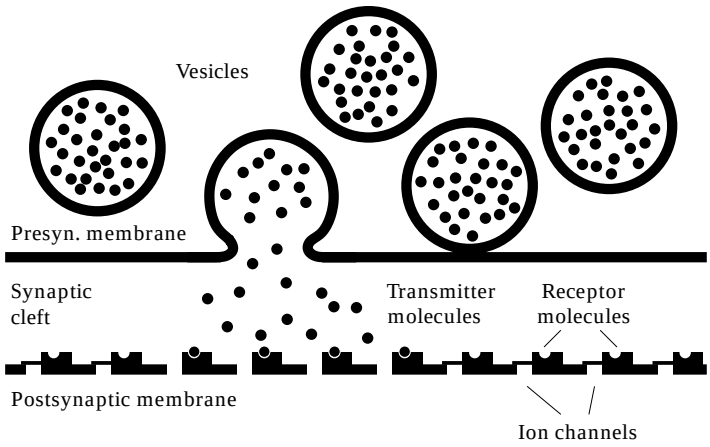


Figure 1.3: Panel (a) is a schematic of the components of a chemical synapse. The presynaptic component is formed by the axon terminal of the incoming neuron. The postsynaptic part is formed by a specialized region of the target cell's membrane. Both components are separated by the synaptic cleft. The presynaptic terminal holds vesicles containing the synaptic transmitter substance. The postsynaptic membrane is enriched with specialized ion channels. Panel (b) shows the processes occurring at an active chemical synapse. When a spike arrives at the axon terminal, part of the vesicles release their content of synaptic transmitter into the cleft. The transmitter molecules diffuse to the postsynaptic membrane where they attach to receptor molecules. In response, the connected ion channels are opened for a short period of time. The resulting ionic current depolarizes or hyperpolarizes the postsynaptic cell. (Adapted from Thompson 1993.)

(Douglas et al. 1989, Douglas and Martin 1990, Douglas and Martin 1992). This means that a given nerve cell can be classified either as an *excitatory neuron* with all its outgoing synapses being excitatory or as an *inhibitory neuron* with all its axonal synapses being inhibitory.

1.2 The Retina

The lens of the eye focuses an inverted image of the outside visual scene onto the *retina* in the same way as the lens of a camera projects an inverted image onto the photographic film. The retina is to convert the image into a sequence of electrical pulses suitable for further processing along the visual pathway.

The first step is taken by the *photoreceptor cells*. By means of photosensitive pigments, they transform variations in the intensity of the incoming light into graded changes of their membrane voltage. Two types of photoreceptors are found in the vertebrate eye, *rods* and *cones*. Rods are extremely sensitive to dim stimuli but are relatively slow. The pigment they contain is called rhodopsin. The cones, on the other hand, are much less sensitive but provide a better temporal resolution. Moreover, the cones also mediate colour vision: Different kinds of cones use different variations on the pigment rhodopsin and are thus sensitive to different wavelengths of light. The retina of humans and old-world monkeys contains rods and three types of cones, which are sensitive to red, green, and blue light. Animals lacking appropriate kinds of receptor cells are colour-blind like, e.g., cats.

The electrical signal produced by the photoreceptors is next relayed via *horizontal*, *bipolar*, and *amacrine cells* to the *retinal ganglion cells*. Similar to the photoreceptors, the horizontal and bipolar cells do not fire spikes but rather respond to stimulation with graded changes of the membrane potential. The ganglion cells, in contrast, emit trains of action potentials to code the result of retinal signal processing. Their axons merge into the optic nerve and leave the eye for the lateral geniculate nuclei.

Since the retinal ganglion cells are the output neurons of the retina, it is quite natural to ask how ganglion cells respond to visual stimulation of the retina. Investigations of this question have revealed that the activity of a given ganglion cell can be influenced by light directed onto a certain area of the retina, whereas its activity remains unaffected when other regions illuminated. This specific area was termed the ganglion cell's *receptive field* (Hartline 1940). As it turned out, the receptive fields of ganglion cells normally are circular and subdivide into a center and a surround having antagonistic effects onto the cell's activity (Kuffler 1953). Two opposite types of such receptive fields can be distinguished, namely *on-center* and *off-center* receptive fields.

An on-center receptive field consists of an excitatory region in its center with a surrounding inhibitory area, or, in short, it has an on-center with an off-surround. As shown in the left column of Fig. 1.4, the corresponding ganglion cell is excited by a spot of light directed into the center of the receptive field, but is inhibited when the outer region is illuminated. Because center and

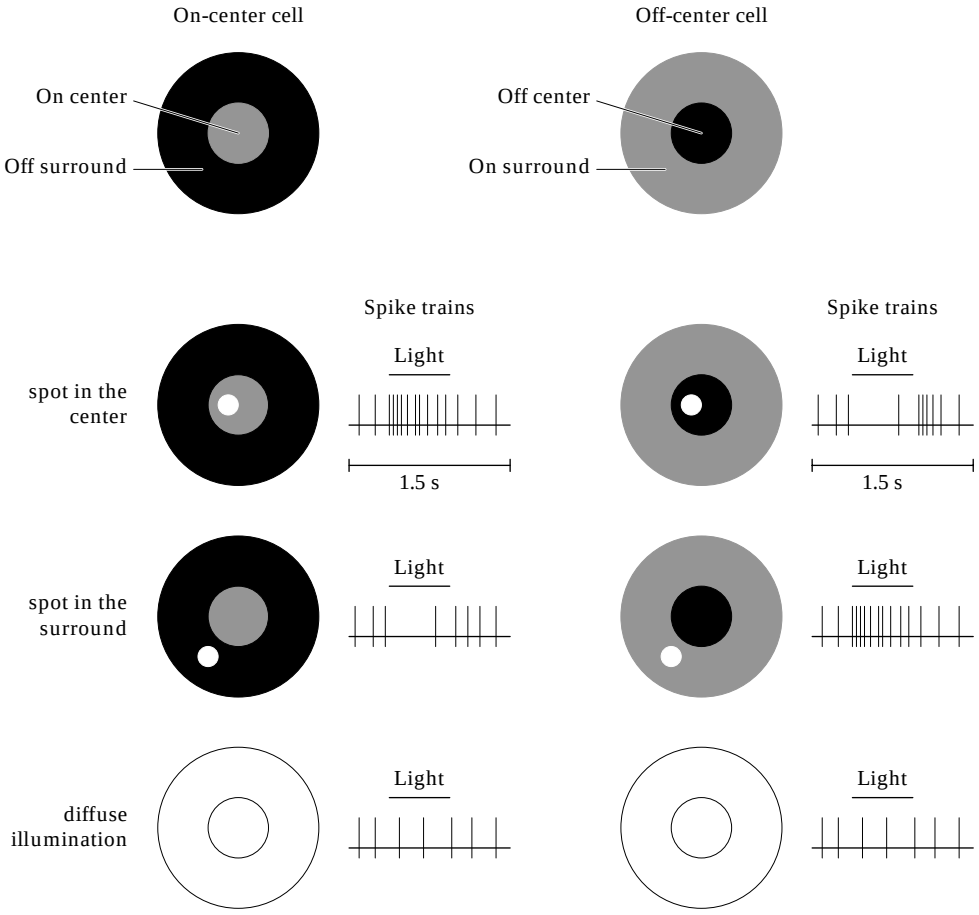


Figure 1.4: Retinal ganglion cells have circular receptive fields that are composed of a center area with an antagonistic surround. An on-center ganglion cell (left column) is excited by a spot of light directed into the center of the receptive field, but is inhibited when the surrounding region is illuminated. Because center and surround have antagonistic effects, the cell does not respond to a diffuse illumination of the receptive field as a whole. Off-center cells (right column) have the opposite response properties. (After Kuffler 1953.)

surround have antagonistic effects, the cell does not respond to a diffuse illumination of the receptive field as a whole. An off-center receptive field, conversely, consists of an off-center with an on surround. The respective ganglion cell is inhibited by a spot of light in the receptive field's center and excited by a stimulation of the surround; see the right column of Fig. 1.4.

Apart from the distinction between on- and off-receptive fields, ganglion cells in the cat retina can be classified into X , Y, and W cells according to anatomical and physiological criteria. W cells are specialized for detecting moving stimuli. In either one of the classes X and Y there are both on-center and off-center cells, but X cells have smaller receptive fields. Further differences between X and Y concern the size of the cell bodies and the properties regarding spatial stimulus summation. In the monkey retina, ganglion cells

can be categorized into two groups, parvocellular (P) and magnocellular (M), which are in some respects similar to the classes X and Y in the cat. P cells have small receptive fields and code for stimulus colour. The receptive fields of M ganglion cells are larger and rather insensitive to colour. Due to the larger diameter of their axons, M cells transmit action potentials more rapidly.

1.3 The Lateral Geniculate Nucleus

The retinal ganglion cells project along the optic nerve and the optic tract to three subcortical regions. In the pretectal area of the midbrain and in the superior colliculus these inputs are used to control pupillary reflexes and eye movements, respectively. Most retinal axons, however, terminate in the *lateral geniculate nucleus (LGN)*, where they are relayed to the visual cortex. On their way, the output fibres of the two eyes are partially crossed at the optic chiasm. As a result, either LGN receives a representation of one half of the visual field. The right LGN processes the left visual hemifield, whereas the left LGN processes the right hemifield.

The LGN of the cat consists of three layers of cells, named A, A1, and C. Inputs to A1 are provided by ganglion cells of the eye on the same side of the animal (ipsilateral), i.e., by axons that do not cross at the optic chiasm. Laminae A and C, in contrast, are supplied with inputs from the eye on the other side of the animal (contralateral). In the same way as the retinal ganglion cells, geniculate neurons fall into two classes, X and Y. Either class of geniculate cells receives input from the corresponding class of ganglion cells.

The monkey's LGN is divided into six layers, numbered from 1 to 6. Layers 2, 3, and 5 are innervated by ganglion cells of the ipsilateral eye; layers 1, 4, and 6 receive input from the contralateral retina. The neurons in laminae 1 and 2 have relatively large cell bodies as compared to those in the other layers. Accordingly, layers 1 and 2 are called magnocellular, layers 3 to 6 are termed parvocellular. The magnocellular layers are innervated by M type retinal ganglion cells, while the parvocellular laminae are innervated by P ganglion cells.

What do the receptive fields of geniculate neurons look like? Hubel and Wiesel (1961) were the first to answer this question. They found that geniculate cells have response properties closely resembling those of retinal ganglion cells. Their receptive fields are circular and consist of two concentrically arranged antagonistic regions. As in the retina, there are both on-center and off-center cells. Interestingly, the geniculate receptive fields are ordered *retinotopically*, which means that the receptive fields of neighbouring cells in the LGN correspond to neighbouring areas on the retina.

It is important to note that the details of the geniculate connectivity and its function are still unknown. Actually, the projections from the retina to the LGN are by far outnumbered by connections arriving from other regions like the reticular formation or the cortex. Not only do such feedback inputs control the flow of visual information from the eye to the cortex, they are also likely to be relevant for the generation of activity patterns needed during early development (Weliky and Katz 1999).

1.4 The Primary Visual Cortex

Axons originating in the lateral geniculate nucleus form the so-called optic radiation and innervate the visual cortex, which extends over the occipital lobe of either cortical hemisphere. The posterior part of the visual cortex, where most of the geniculate axons terminate, is the *primary visual cortex*, also called *visual area 1 (V1)*, *area 17*, or *striate cortex*. V1 is immediately surrounded by and projects to area V2 (area 18). In addition to these projections, however, area V2 also receives direct afferents from the LGN. Further visual areas include V3, V4 and V5. Areas V3 and V4 are thought to process information regarding the form of visual objects; V5 seems to be important for the detection of motion.

The neuronal network model we shall study in the subsequent chapters is concerned with the interface between the LGN and the primary visual cortex. Therefore, the following survey will concentrate on the organization of area V1.

1.4.1 Anatomical Structure

In the primary visual cortex, as in other cortical regions, two classes of neurons are predominant: the *pyramidal cells* and the *stellate cells*. Pyramidal neurons have spiny dendrites and relatively large cell bodies shaped as pyramids. Their axonal arbors often span long distances across the cortex. Stellate cells are smaller and project to local targets. Their dendrites are either spiny or smooth. The pyramidals and the spiny stellates are excitatory neurons, whereas the smooth stellate cells are inhibitory.

Perpendicular to the cortical surface, six layers of cells can be distinguished according to cell density, fiber density, and the principal cell type. Proceeding from the surface to the underlying white matter, the layers are numbered from 1 to 6. Layer 4 is further divided into sublayers 4A, 4B, and 4C. The majority of axons sent from the LGN to the cortex (*geniculocortical axons*) end on stellate cells in layer 4, but additional terminals can be found in layer 6. Layer 4 neurons then project to the upper layers 2 and 3, from where pyramidal cells project to layer 5 pyramidals. These, in turn, innervate pyramidal cells in layer 6, which project back to layer 4.

Apart from setting up this kind of internal circuitry, neurons in the different laminae also provide output signals for subsequent processing. Cells in laminae 2, 3 and 4, e.g., send axons to other cortical regions, especially to higher visual areas. Axons of cells in layer 5 primarily go to the superior colliculus, and layer 6 neurons project to the claustrum and supply feedback to the lateral geniculate nucleus.

1.4.2 Receptive Fields

The receptive fields of most neurons in the primary visual cortex are strikingly different from those of geniculate cells. As we have seen, cells in the various layers of the LGN are monocular, i.e. they are driven exclusively by one eye's ganglion cells. The majority of cortical cells, in contrast, receive input from

both eyes and thus respond to visual stimulation of either eye. Normally, the strength of a cell's response to a given stimulus depends on whether the stimulus is presented to the left or to the right eye, meaning that the cell "prefers" one eye to the other (Hubel and Wiesel 1962). This phenomenon is called *ocular dominance*.

What type of stimulus is most efficient for activating neurons in the primary visual cortex? Hubel and Wiesel (1959) found that cells in the cat's striate cortex respond best to bright or dark bars of a specific orientation within their receptive fields. The required orientation and the location of the receptive field vary from cell to cell. According to the detailed response properties, Hubel and Wiesel classified the neurons as *simple cells* and *complex cells*.

The receptive fields of simple cells, like those of geniculate cells, consist of antagonistic on- and off-subdivisions. Unlike geniculate receptive fields, however, they are not rotationally symmetric. Instead, they are composed of elongated subregions lying side by side in various configurations as shown in panel (a) of Fig. 1.5. Each cell is activated best if the on-regions of its receptive field are illuminated while the off-regions are left dark. The optimal stimulus therefore consists of a bar or edge of light presented with a specific orientation at a specific position within the receptive field so that it coincides with the subregions; see Fig. 1.5, panels (b) and (c).

A model of how such a simple-cell receptive field may be constructed from the center-surround receptive fields of geniculate cells has been proposed by Hubel and Wiesel (1962) and is illustrated in Fig. 1.6. A group of on-center geniculate cells whose receptive fields are aligned on the retina make excitatory synapses onto one cortical neuron. When a bar of light is projected on the retina in such a way that all the receptive-field centers are illuminated, then the cortical cell will be strongly excited and increase its firing rate accordingly. A bar of light projected at a different orientation or location, in contrast, will produce only a weak, if any, excitation of the cortical cell.

Complex cells usually have larger receptive fields than simple cells. They also respond best to stimuli consisting of specifically oriented edges or bars of light. However, as opposed to simple cells, the exact position of the edge or bar within the receptive field is less important. Hubel and Wiesel (1962) suggested that this may be the result of a wiring analogous to the one they proposed for the construction of simple-cell receptive fields. Consider a neuron receiving strong excitatory input from an array of simple cells whose receptive fields have similar substructure but are slightly shifted with respect to one another. Whenever one of these simple cells is activated, it drives the target neuron. The target neuron therefore behaves like a complex cell – it responds to any stimulus of the correct orientation, irrespective of the exact position within the receptive field.

It should be emphasized that the above models probably oversimplify the true wiring schemes, which are unknown and may vary from species to species. In the cat, most simple cells in cortical layer 4, the region receiving the dominant geniculate input, have orientation selective receptive fields as described above. This is different in other species like monkey (Hubel and Wiesel 1968, Blasdel and Fitzpatrick 1984, Hawken and Parker 1984), ferret (Chapman

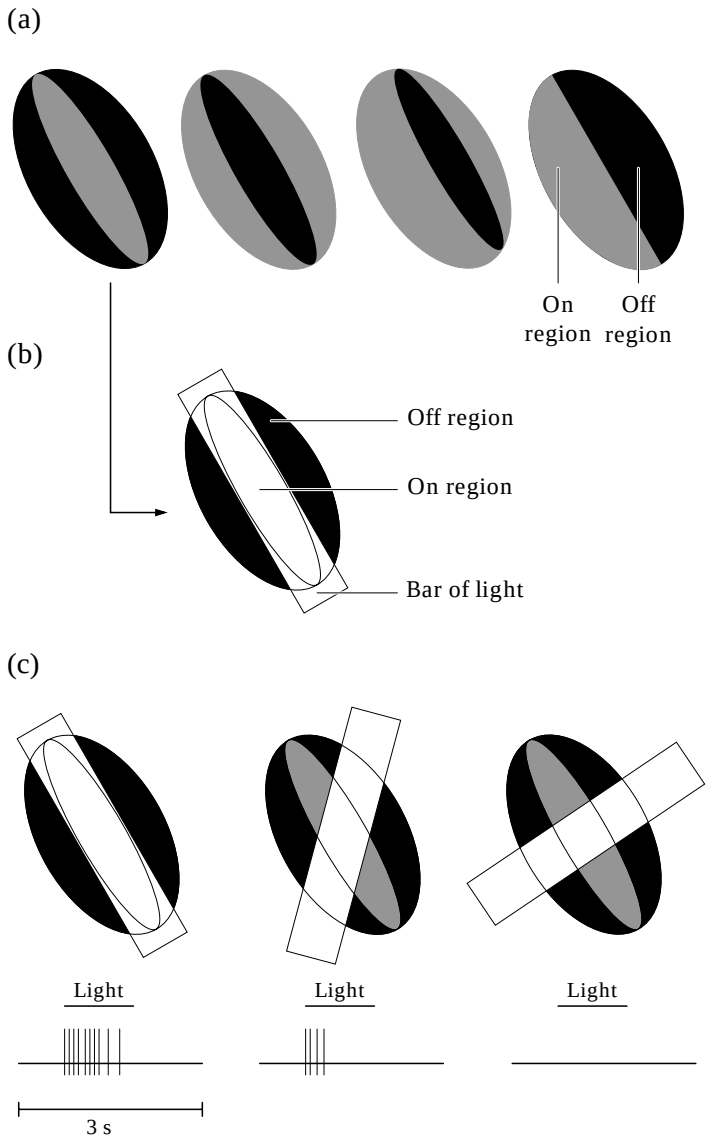


Figure 1.5: Panel (a): The receptive fields of simple cells consist of elongated subregions lying side by side in various configurations. Panel (b): A simple cell responds best to a bar of light whose edges coincide with the boundaries of the on-region. Panel (c): As a consequence of the elongated shape of the receptive field's subdivisions, simple cell responses are orientation selective. (Panel (a) after Hubel and Wiesel 1962; panel (c) after Hubel and Wiesel 1959.)

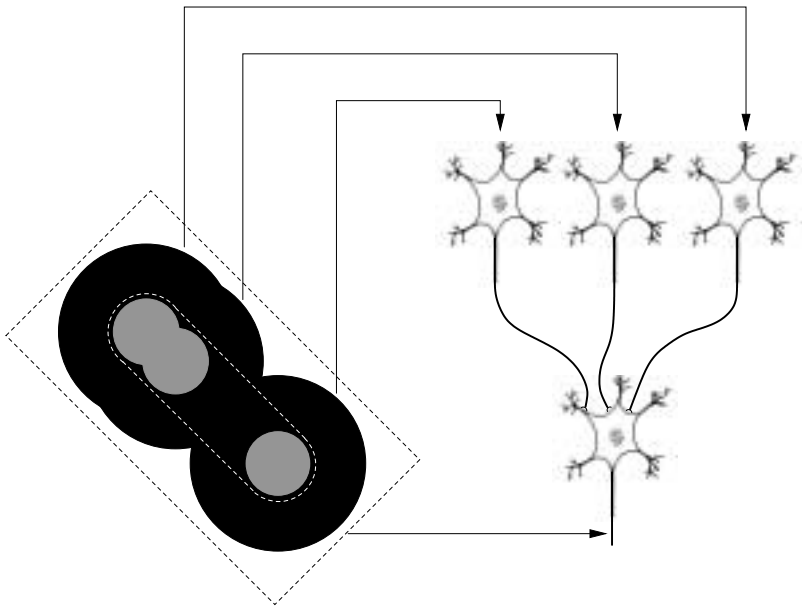


Figure 1.6: The figure illustrates how a simple cell's receptive field may be built up from center-surround receptive fields of geniculate cells. A group of on-center geniculate cells whose receptive fields are aligned on the retina contact one cortical neuron. When a bar of light is projected onto the retina in such a way that all the receptive-field centers are illuminated, then the cortical cell will be strongly excited. In contrast, a bar of light projected at a different orientation or location will produce only a weak, if any, excitation of the cortical cell. (Adapted from Hubel and Wiesel 1962.)

and Stryker 1993), or tree shrew (Humphrey and Norton 1980). Simple cells in cortical layer 4C of old world monkeys, for example, are usually monocular and their receptive fields have a center-surround substructure like those of LGN cells. Orientation selective receptive fields first occur in the monkey's cortical layer 4B, where inputs from layer 4C converge.

Today, it is still a matter of debate whether orientation selectivity in visual area 1 is primarily set up by converging feedforward projections as suggested by the model of Hubel and Wiesel or by recurrent intracortical connections (Ben-Yishai et al. 1995, Somers et al. 1995, Hansel and Sompolinsky 1996, Ben-Yishai et al. 1997, Adorján et al. 1999). Currently, there is experimental support for both types of models. A recent review on this topic has been given by Ferster and Miller (2000).

1.4.3 Columnar Organization and Cortical Maps

When a recording electrode is driven through the cortex perpendicularly to the surface, the majority of cells it encounters are found to have similar ocular dominance, preferred orientation, and retinal position of the receptive field. On the other hand, tangential penetrations reveal that these properties change gradually as the recording site is moved across the cortical surface. Hence,

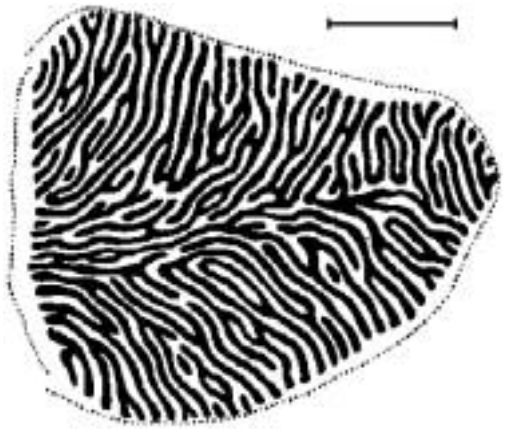


Figure 1.7: The ocular dominance map can be visualized, e.g., by injecting a radiolabeled amino acid into one of the eyes, where it is taken up by the retinal neurons and transported through the lateral geniculate nucleus to the visual cortex. Cortical cells receiving input from the injected eye can then be detected using an X-ray sensitive photographic film. The figure shows a drawing of the ocular dominance map obtained in layer 4C of a monkey's primary visual cortex. The black and white stripes indicate the cortical regions where cells were predominantly driven by one or the other eye. Scale bar, 5 mm. (Adapted from LeVay et al. 1975)

the cortex seems to be organized into *columns* of cells with similar response properties (Hubel and Wiesel 1962). A cluster of cells dominated by the same eye is referred to as an *ocular dominance column*; a stack of neurons preferring similar orientations is called an *orientation column*.

The above-mentioned observations gave rise to the concept of cortical maps. The continuous relation between a neuron's cortical position and the location of its receptive field on the retina has been named the *retinotopic map*. Analogously the relations of cortical position versus ocular dominance and preferred orientation have been termed *ocular dominance map* and *orientation map*, respectively.

These cortical maps can be measured in different ways. One method to visualize the ocular dominance map is to inject a radiolabeled amino acid into one of the eyes, where it is taken up by the retinal neurons and transported to the lateral geniculate nucleus. As it turns out, the radiolabel is transferred across synaptic contacts (!) to the geniculate cells and further transported to the primary visual cortex. In cortical sections, the cells receiving input from the injected eye can then be detected using an X-ray sensitive photographic film. Figure 1.7 shows the typically striped pattern of ocular dominance columns obtained in layer 4C of a monkey's primary visual cortex: Cells in the white regions received input from one eye, neurons in the black areas were innervated by the other eye.

More recent techniques, known as *optical imaging*, allow the maps to be visualized in vivo using a CCD camera. To this end, the cortex can be pre-

pared with a voltage sensitive dye so that the pattern of electrical activity is converted into an optical signal for the camera (Grinvald et al. 1984, Orbach et al. 1985). Alternatively, activity-dependent intrinsic signals can be recorded, such as variations of the tissue's "colour" resulting from changes of the relative concentrations of oxy- and deoxy-hemoglobin (Grinvald et al. 1986).

In order to measure an animal's cortical orientation map by means of optical imaging, the animal is visually stimulated using several gratings of different orientations while its cortical activity is recorded with the camera. By comparing the observed patterns of activity, the preferred stimulus orientation can be determined for each point of the recorded area. The resulting map is usually visualized as an image of the cortical surface in which each pixel is assigned a color that codes for the preferred orientation found at the corresponding cortical point.

Figure 1.8 (a) shows a colour-coded orientation map that Bonhoeffer and Grinvald (1993) have recorded from a cat's area 18. Panel (b) of the same figure displays an orientation map obtained from the striate cortex (V1) of a monkey (Blasdel 1992b). The colours coding for the different orientations are indicated below each panel. In both maps orientation columns appear as curved patches of uniform colour. A typical feature of such orientation maps is the existence of point-like singularities where the preferred orientation changes by 90 degrees. Around each of these singularities, every orientation is represented once. They have been termed "orientation centers" or "pinwheel centers" (Bonhoeffer and Grinvald 1993).

From the above explanations it is clear that on the cortical surface several different maps are superimposed onto one another. So far we have mentioned the retinotopic, the ocular dominance, and the orientation map, but there is also a map of preferred direction of stimulus movement (Swindale et al. 1987, Weliky et al. 1996, Shmuel and Grinvald 1996). The question of how the various maps are woven into one another is a topic of ongoing research (Swindale et al. 1987, Bartfeld and Grinvald 1992, Blasdel 1992a, Blasdel 1992b, Obermayer and Blasdel 1993, Shmuel and Grinvald 1996, Crair et al. 1997, Hübener et al. 1997). Current experimental data indicate that pinwheel centers of the orientation map tend to lie in the centers of ocular dominance bands and that a cell's preferred direction is approximately perpendicular to its preferred orientation.

1.5 Development

In the preceding sections we have discussed the main stages of information processing and transmission along the visual pathway from the retina to the primary visual cortex. In the next chapter we will present a model of activity-driven synaptic refinement in layer 4 of cortical area V1 which may explain the emergence of orientation maps and their marked stability observed in various experiments. The related processes are thought to occur early in an animal's life, when the ontogenetic development of the brain is not yet complete. Let us therefore survey a few relevant milestones in the development of the mam-

(a) (b)

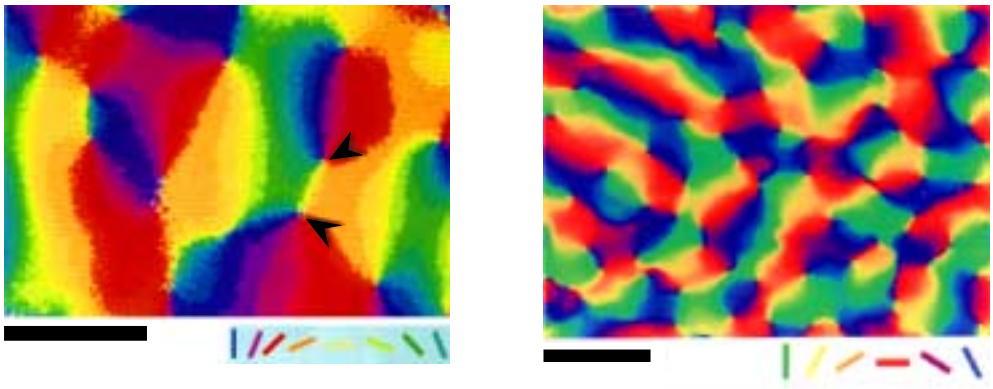


Figure 1.8: Optical imaging of cortical activity allows to determine the preferred stimulus orientation for each point of the recorded cortical area. The resulting orientation map is usually visualized as an image of the cortical surface with colours coding for the preferred orientations. Panel (a) shows an orientation map that Bonhoeffer and Grinvald (1993) have recorded from a cat's area 18. Panel (b) displays an orientation map obtained from the striate cortex (V1) of a monkey (Blasdel 1992b). The colour codes that have been applied are indicated at the bottom of the panels; the black scale bars are 1 mm. A typical feature of such orientation maps is the existence of point-like singularities around which each orientation is represented once; see the arrows in panel (a). These singularities have been named "orientation centers" or "pinwheel centers" (Bonhoeffer and Grinvald 1993).

malian brain. In particular, we will focus on the cat's primary visual cortex.

The mammalian embryo consists of three principal layers, namely an innermost layer called endoderm, an intermediary layer called mesoderm, and an outermost layer termed ectoderm. By a process named neural induction a certain region of the ectoderm becomes committed to the formation of the nervous system. This region extends along the dorsal midline of the embryo and is called neural plate. The detailed mechanisms of its commitment are still unknown, but they involve some kind of interaction between the ectoderm and the mesoderm. The edges of the neural plate then begin to move upward until they fuse at the midline so that the neural plate is folded and finally takes the shape of a tube termed neural tube.

At its head end, the neural tube forms three swellings that give rise to the three main parts of the brain: forebrain, midbrain, and hindbrain. The forebrain further subdivides into the telencephalon (endbrain) and the diencephalon (betweenbrain). Similarly, the hindbrain gives rise to the metencephalon (afterbrain) and the myelencephalon. During later development, the cerebral hemispheres, including the cerebral cortex, emerge from the telencephalon; the retinae and the thalamus with LGN emerge from the diencephalon.

At the beginning of cortical development the corresponding region of the

neural tube is only a few cells thick and consists of two layers. The inner layer is called *ventricular zone*, the outer layer is the *marginal zone*. Cells in the ventricular zone proliferate so that the total number of cells increases continuously. Having divided several times, some cells loose their ability to proliferate and move outward forming the *intermediate zone* between the ventricular and the marginal zone. The upper part of the intermediate zone is the so-called *subplate*.

Above the subplate a fourth layer, the *cortical plate*, emerges. Within the cortical plate the future cortical laminae are built up. This occurs in an inside-out sequence, i.e., cells destined for lamina 6 are generated before lamina 5 cells, and so on. All cells are born in the ventricular zone and migrate through the intermediate zone to their target layer. Thus, the deeper cortical layers are formed before the upper layers and cells destined for the upper layers have to pass through the deeper layers. In the cat, for example, layer 2 cells reach their target position about one month after layer 6 cells have finished their migration.

At what time is the developing visual cortex innervated by axons from the LGN? Shatz and Luskin (1986) have found that the cat's optic radiation reaches the primary cortex relatively early, when the cells of future layer 4 have just been born in the ventricular zone and are still on the way to their target locations. Interestingly, however, the geniculate axons do not enter the cortical plate immediately, but instead "wait" in the subplate for a prolonged period of time. As a result, layer 4 cells first receive geniculate input about one week *after* they have finished their migration.

It is still an open question how the extending nerve fibers find their targets so as to set up the intricate circuitry of the adult brain. According to current knowledge, several different mechanisms are important (for a review see Goodman and Shatz 1993). Some fibers seem to grow along gradients in the concentration of certain molecules. Others may be guided by existing fibres or by so-called guide-post cells which act as landmarks indicating the correct pathway. Sperry (1943) was the first to demonstrate that in certain neural systems an axon's target can be predetermined very accurately. To this end he cut the optic nerve of a frog and rotated the eye by 180 degrees. Unlike in mammals, a frog's optic nerve regenerates and grows back into the optic tectum. After the frog had regained vision, it struck down to catch a fly that was up, and vice versa – its visual world was rotated by 180 degrees. Thus, the optic nerve had reconnected to the optic tectum in the same way as before the rotation of the eye, which shows that, by some mechanism, each axon "knows" its specific target position.

On the other hand, it is well-known that the fine tuning of the mammalian cortex is crucially dependent on sensory input and neuronal activity. This has been demonstrated in several experiments. For example, Wiesel and Hubel (1963b) have shown that when a kitten's eye is sutured within the first two weeks after birth and reopened one or two months later, then the animal is left practically blind in that eye. While the responses of geniculate cells to stimulation of the deprived eye appear relatively normal, only a few cortical cells can be driven by the deprived eye and their receptive fields are abnormal (Wiesel

and Hubel 1963a, Wiesel and Hubel 1965). These effects are reversible if the deprived eye is reopened within the first six weeks postnatal and the previously open eye is closed (Blakemore and van Sluyters 1974). The visual system of *adult* cats, in contrast, does not seem to be affected by even prolonged periods of monocular occlusion. It has thus been concluded that there is a *critical period* early in life during which the cortical development is influenced by visual experience (Crair et al. 1998, Issa et al. 1999). In the cat this critical period is thought to extend over some three months after birth.

The results obtained from the above experiments might lead to the view that the disuse of the deprived eye is mainly responsible for the disruption of normal visual responses. One could therefore expect that binocular deprivation results in most cortical cells being unresponsive to both eyes and the animal being blind. Interestingly, however, it has been found (Wiesel and Hubel 1965) that after binocular deprivation the majority of cortical cells respond largely normal. Therefore, some kind of *competition* between inputs representing the two eyes rather than disuse of the closed eye appears to be the relevant factor in producing the striking changes during monocular deprivation.

Anatomical studies in layer 4 of the cat's primary visual cortex have shown that at the time of birth, afferents from the lateral geniculate nucleus representing the two eyes are intermingled and do not form the typical striped pattern of ocular dominance described in Sect. 1.4.3. A morphological segregation of the geniculate axons into ocular dominance columns becomes first visible about three weeks after birth (LeVay et al. 1978). This segregation does not appear to require visual experience, but it seems to depend critically on spontaneous neuronal activity: The formation of ocular dominance columns is not prevented by rearing the animal in the dark or with both eyes closed (Stryker and Harris 1986, Crair et al. 1998), but the geniculocortical afferents do not segregate if the spontaneous activity of retinal ganglion cells is inhibited by injecting tetrodotoxin into both eyes (Stryker and Harris 1986).

Other experiments indicate that the development of orientation selectivity of cortical cells is also activity-dependent. Kittens that are reared in environments where they can see only contours of a single orientation remain practically blind for contours of the perpendicular orientation; microelectrode recordings reveal that the distribution of preferred orientations in the primary visual cortex of these *stripe-reared* animals is biased towards the experienced orientation (Blakemore and Cooper 1970, Hirsch and Spinelli 1970, Blasdel et al. 1977). Recently, the influence of experience on orientation maps in cat visual cortex has been analyzed by means of optical imaging techniques (Sengpiel et al. 1999). In stripe-reared animals the total fraction of cortical area representing the experienced orientation turned out to be increased as compared to the area occupied by columns preferring other orientations. Nevertheless, all orientations were still represented and the overall layout of the maps (e.g. pinwheel density) appeared largely unaltered. These results suggest that that the formation of orientation maps is guided by environmental as well as by intrinsic factors.

Figure 6 summarizes a few important stages in the development of the cat's

visual system. Even from this simple scheme it is evident that the circuitry of the adult brain arises from a complex interplay of a large number of simultaneous processes. Many investigations like those described above have shown that the involved processes are partially dependent on neuronal activity and on sensory experience. This has been studied extensively in the visual system, but similar results have been obtained also in other sensory pathways as, e.g., in the somatic sensory and the auditory system.

1.6 Summary

In the preceding sections we have presented a survey of the visual pathway from the retina to the primary visual cortex; an overall view is given in Fig. 1.1. As prerequisite the basic units of biological information processing – neurons and synapses – have been introduced in Sect. 1.1. As we have seen, a neuron or nerve cell is specialized for transmitting and transforming electrical signals. Most neurons use short electrical pulses, so-called spikes, to code their output. The connection between two neurons is called synapse. Some synapses provide a direct electrical contact, but the majority of connections are chemical synapses, where an incoming spike is first transformed into a chemical signal on the presynaptic side and then reconverted into an electrical signal at the postsynaptic nerve cell.

The retina is the input stage of the visual system. An inverted image of the outside visual scene is focused onto the retina by the lens of the eye. The retina contains photoreceptor cells so as to convert this image into electrical signals. After some processing, these signals arrive at the retinal ganglion cells and, from there, leave the eye along the optic nerve. How does a ganglion cell respond to visual stimulation of the retina? The output of a given ganglion cell can be influenced only by stimulation of a certain retinal area. This area is called the ganglion cell's receptive field. The receptive fields of ganglion cells are usually circular and can be divided into two concentric subregions with antagonistic effects on the cells' output signal (Fig. 1.4) They are therefore called center-surround receptive fields.

At the optic chiasm, the retinal output fibres are partially crossed and rebundled in such a way that the left optic tract carries a representation of the right half of the visual field, whereas the right optic tract represents the left visual hemifield. Each optic tract then projects to one of the lateral geniculate nuclei (LGN), and so each LGN receives information from one half of the visual field. The receptive fields of neurons in the LGN are very similar to those of retinal ganglion cells: They are circular, have a center-surround substructure, and are monocular, i.e., are driven exclusively by one eye. Therefore, the LGN is often regarded as a simple relay station although it is well known today that most inputs to the LGN do not arrive from the retina but from other regions of the brain and include massive feedback connections from the cortex. Notably, the cells in the LGN are arranged retinotopically with neighbouring cells having neighbouring receptive fields on the retina.

Output fibres of geniculate neurons innervate the primary visual cortex,

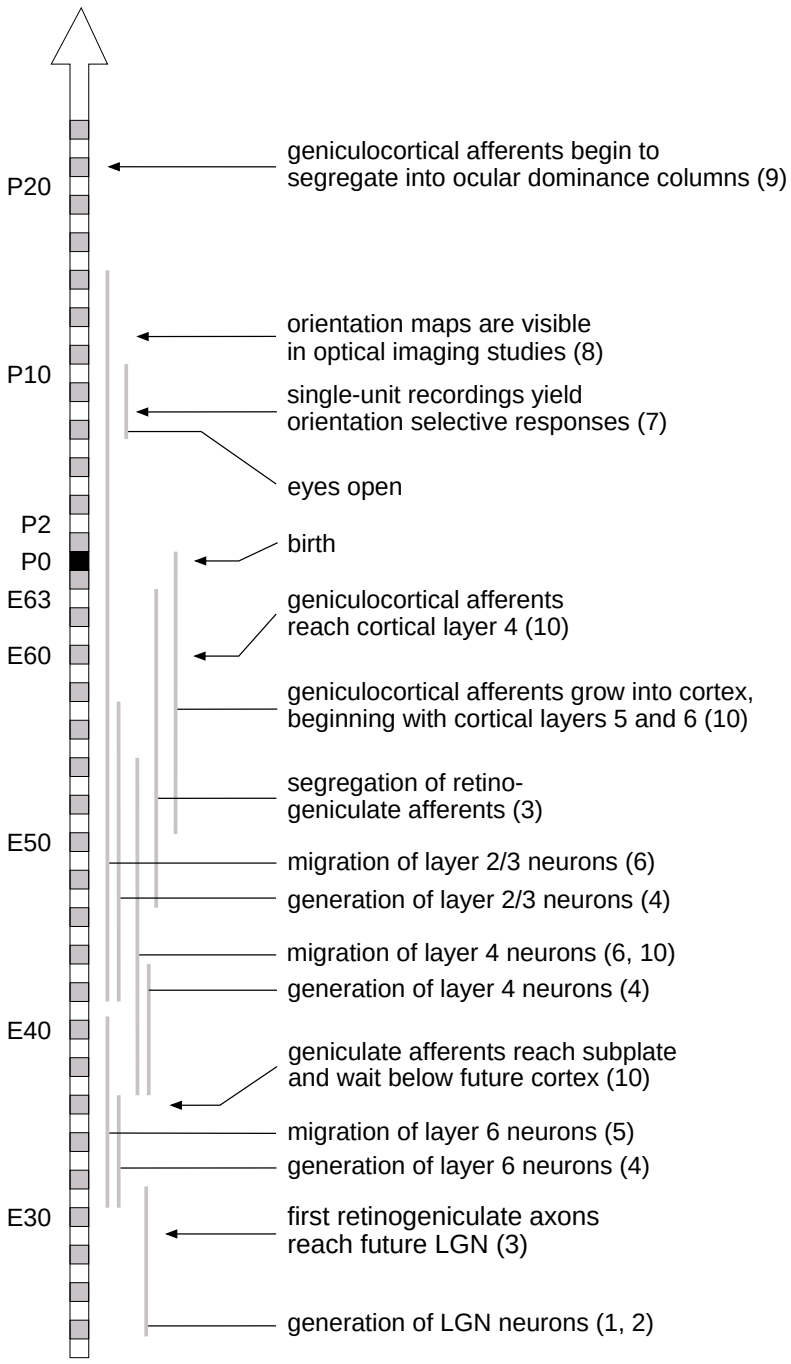


Figure 1.9: Important stages in the development of the cat visual system are depicted. A few corresponding references are denoted in parentheses. (1) Hickey and Hitchcock 1984; (2) Shatz 1981; (3) Shatz 1983; (4) Luskin and Shatz 1985b; (5) Luskin and Shatz 1985a; (6) Shatz and Luskin 1986; (7) Hubel and Wiesel 1963; (8) Crair et al. 1998; (9) LeVay et al. 1978; (10) Gosh and Shatz 1992.

also referred to as striate cortex, visual area 1 (V1), or area 17. The visual cortex consists of six layers of cells, numbered from 1 to 6. The majority of geniculocortical projections end in layer 4, from where the incoming signals are fed into a complicated intracortical circuitry whose function is largely unknown. Most neurons in area V1 are orientation selective, meaning that they respond best when a specifically oriented edge of light is projected onto the retina. According to the properties of their receptive fields, cortical neurons are divided into so-called simple and complex cells. Simple cell receptive fields consist of elongated antagonistic subregions lying closely side by side as shown in Fig. 1.5. The response properties of complex cells may result from a nonlinear combination of several simple cells' outputs. Although most cortical cells respond to a stimulation of either eye, one eye is usually preferred. This phenomenon is called ocular dominance.

As in the LGN, neurons in the primary visual cortex are arranged in a retinotopic fashion so that the receptive fields neighbouring cells cover neighbouring regions on the retina. Moreover, neighbouring neurons are normally dominated by the same eye and prefer similar stimulus orientations. When a recording electrode is moved tangentially to the cortical surface, then the encountered cells' receptive field position, ocular dominance, and preferred orientation are found to change continuously. The relation between a neuron's cortical location and the retinal position of its receptive field is referred to as the retinotopic map. Analogously, the relations of cortical location versus ocular dominance and preferred orientation are called ocular dominance map and orientation map, respectively.

During ontogeny the visual cortex develops from a certain region of the so-called neural tube. All cortical cells are born in the ventricular zone at the inner surface of the tube and then migrate outward to the cortical layer they are destined for. The layers are formed in a reverse order, i.e. the deeper layers are generated before the upper layers. Already at a very early stage, nerve fibers from the developing LGN arrive immediately below area V1. Interestingly, however, they undergo a prolonged waiting period before entering their target region.

How do all the neuronal fibers find their correct targets so as to set up the complicated wiring of the brain? As we have discussed, current experimental data indicate that the nervous system uses several different mechanisms for shaping its circuitry. Some of them appear to depend crucially on neuronal activity. Extensive studies, especially in the mammalian visual system, have revealed a so-called critical period early in an animal's life during which the development of the cortex is influenced significantly by neuronal activity and sensory experience. Subsequently, we will present a theoretical model of activity-driven self-organization in the primary visual cortex which provides a new explanation for some experimental results difficult to understand in the framework of earlier models.

Chapter 2

Modeling Self-Organization in the Primary Visual Cortex

Measurements of neuronal activity in primary visual cortex of cat have shown that cortical cells respond well to stimulation within a certain receptive field on the retina. Many cells respond preferentially to bar-like stimuli of a specific orientation and are activated predominantly through one of the two eyes. The location of the receptive field, the preferred orientation, and the ocular dominance of recorded cells change gradually as the recording site is moved tangentially to the cortical surface (Hubel and Wiesel 1962). The global organization of these cortical response properties has been mapped by anatomical, electrophysiological, and optical imaging methods for cat (Tusa et al. 1978, Bonhoeffer and Grinvald 1991, Bonhoeffer and Grinvald 1993), monkey (LeVay et al. 1975, Hubel et al. 1977, Blasdel and Salama 1986, Blasdel 1992a, Blasdel 1992b), ferret (Law et al. 1988, Chapman and Stryker 1993, Chapman et al. 1996, Weliky and Katz 1997), and tree shrew (Humphrey and Norton 1980, Humphrey et al. 1980, Bosking et al. 1997).

It is, however, still a matter of debate how cortical orientation selectivity is set up. Hubel and Wiesel (1962) originally proposed that geniculocortical connections are arranged so that the receptive field centers of thalamic cells projecting onto a single cortical simple cell cover an elongated region in the visual field. While there are many experimental studies claiming that the response properties of simple and complex cells are mainly determined by feedforward projections as suggested by this model (Ferster 1987, Ferster 1988, Reid and Alonso 1995, Ferster et al. 1996, Chung and Ferster 1998), others find that intracortical links provide the main contribution (Sillito 1979, Sillito et al. 1980, Crook and Eysel 1992, Nelson et al. 1994). At present it seems most likely that both feedforward and recurrent intracortical processes of both excitatory and inhibitory nature participate in the formation of orientation selectivity (Vidyasagar et al. 1996).

Theoretical studies (von der Malsburg 1973, Linsker 1986b, Linsker 1986a, Kammen and Yuille 1988, Stetter et al. 1993, Miller 1994, Wimbauer et al. 1997b, Wimbauer et al. 1997a) have proposed a Hebbian development of geniculocortical synapses that is driven by correlated feedforward input

from thalamic neurons (for a recent review see Miller et al. 1999). In these correlation-based approaches the intracortical connectivity is assumed to be rotationally symmetric and fixed. Under relatively general conditions they predict the emergence of an orientation map for cortical simple cells which is formed by the resulting arrangement of feedforward connections.

As such, the above explanation of cortical map formation has been challenged by work of Gödecke and Bonhoeffer (1996) and of Sengpiel et al. (1998). In their experiments, cats were raised so that both eyes never received visual input at the same time, which was achieved by reverse lid suture. If geniculocortical refinement were driven by activity correlations in the lateral geniculate nucleus (LGN) and these correlations were mainly determined by the correlations of the visual input, then the left-eye orientation map would form independently of the right-eye map and the two maps could be expected to be different. Optical imaging of area 18, however, showed them to be nearly identical. The authors concluded that the map's layout was fixed by some internal mechanism either a priori or during the period when the first eye was open. They proposed long-range horizontal projections within primary visual cortex as a potential substrate of this mechanism.

Recently, however, Erwin and Miller (1998) have demonstrated that the emergence of ocularly matched orientation maps can be well explained within the framework of correlation-based development, if an appropriate amount of thalamic inter-eye activity correlations is assumed. Experimental findings of Weliky and Katz (1999) indicate that strong inter-eye correlations are indeed present in the ferret's LGN before eye opening. Nevertheless, it is still unclear, whether this model can actually account for the outcome of the reverse-lid-suture experiments.

Wolf et al. (1996) have pointed out that cortical area 18 of the cat is shaped as a narrow band on the cortical surface, so that pattern formation within this region is subject to strong confinement. In computer simulations they have shown that different feedforward orientation maps developing under this constraint are always very similar – in accordance with reverse-suturing experiments. The authors argued that experimental results should be qualitatively different in a larger area as, e.g., area 17, because boundary conditions are less important. "Unfortunately this idea is difficult to test, as in cats the main part of area 17 lies buried in the medial bank and is therefore inaccessible to optical imaging", as Bonhoeffer and Gödecke (1996) have explained.

In this chapter we propose a model of layer 4 of primary visual cortex consisting of laterally interconnected *spiking* neurons of both excitatory and inhibitory type. It combines the idea of correlation-based learning of geniculocortical afferents with Hebbian development of *short-range* intracortical synapses. Inhibitory interneurons and plastic inhibitory synapses have been included into the model so as to control overall network activity.

Large-scale computer simulations show that in this kind of network it is possible to obtain an *intracortical* orientation map from a Hebbian learning process driven by cortical activity alone. The process does not depend on the presence of feedforward input and could therefore occur at early stages of visual development, when thalamic axons have not yet entered cortical layer

4. The resulting map structure resembles that of typical orientation maps obtained from optical imaging experiments. This might indicate that intracortical circuitry does contribute significantly to the orientation selective response properties of cells in the primary visual cortex.

Experiments by Ferster (Ferster 1987, Ferster 1988, Ferster et al. 1996) do indicate, however, that feedforward input from the LGN is relevant as well. Consistently with these data, we demonstrate that correlation-based development of geniculocortical projections can interact with emerging intracortical connectivity so as to give a matched feedforward *and* intracortical orientation tuning for each cortical cell. As a consequence, the developing pattern of feedforward projections is more or less predetermined, once the intracortical connectivity is fixed. This provides a very natural explanation of the remarkable stability of orientation maps that has been found experimentally (Kim and Bonhoeffer 1994, Weliky and Katz 1997, Gödecke and Bonhoeffer 1996, Sengpiel et al. 1998, Sengpiel et al. 1999). In contrast to the proposition of Wolf et al. (1996), our model predicts an agreement of orientation maps in reverse-suturing experiments not only for small but also for larger visual areas such as area 17 of the cat.

The chapter is organized as follows. In Sect. 2.1, we introduce the model of spiking neurons that we have used in our simulations. We then turn to a description of the full network, explain the learning rules that govern synaptic development, and present our data-analyzing procedure in Sects. 2.2 through 2.4. The numerical simulations in Sect. 2.5 demonstrate that intracortical orientation maps can develop from spontaneous cortical activity alone, without feedforward input. In Sects. 2.6 and 2.7, we show how this intracortical development can be combined with a plasticity of geniculocortical connectivity and show that such a combined development can explain the outcome of the reverse-suturing experiments. Section 2.8 provides the values of all simulation parameters that are needed to reproduce our results. At the end, we summarize and give a short discussion. Part of this work has been presented in preliminary form in (Bartsch and van Hemmen 1999).

2.1 Spiking Neurons

To keep our network as close to biology as possible, we decided to build it up from spiking neurons. A neuron model that is able to reproduce many biological features and yet allows to handle large network simulations is given by the usual integrate-and-fire neuron. For the present investigation we have chosen the stochastic spike response model (Gerstner and van Hemmen 1992, Gerstner and van Hemmen 1994), which is a more flexible, extended version of the integrate-and-fire neuron.

The state of a spike response-neuron is described by its membrane potential h as a function of time. Every incoming spike evokes a transient change of this membrane potential called postsynaptic potential and modeled as a response kernel $\varepsilon(t)$. After spike emission the neuron enters a refractory period that is described by a second response function $\eta(t)$, the refractory poten-

tial. Spike generation itself is governed by a generalized Poisson process. This means that the probability $P^{\mathrm{f}}(t, \mathrm{d}t)$ that the neuron emits a spike during the infinitesimal time interval $[t, t + \mathrm{d}t]$ can be written

$$P^{\mathrm{f}}\left(t, \mathrm{d}t\right) = \lambda\left(t\right) \mathrm{d}t \, ,$$

and the probability of more than one spike being emitted during that period is $o(\mathrm{d}t)$. The spike emission rate $\lambda(t)$ is given by some function q of the membrane potential h,

$$\lambda\left(t\right) = q\left[h\left(t\right)\right] \, .$$

Each time the neuron generates an action potential, a negative contribution $\eta(t)$ is added to the membrane potential to account for the reduced excitability during the refractory period. At a postsynaptic neuron i each action potential arriving from a presynaptic neuron j induces a postsynaptic potential that is given by $\varepsilon(t)$ multiplied by a synaptic weight J_{ij}.

Thus the total membrane potential h_i of neuron i is a sum of a synaptic contribution h_i^{syn} evoked by spikes of other neurons and a refractory contribution h_i^{refr} as a result of its own spiking,

$$h_i\left(t\right) = h_i^{\mathrm{syn}}\left(t\right) + h_i^{\mathrm{refr}}\left(t\right) \, ,$$

where

$$\begin{aligned} h_i^{\mathrm{refr}}\left(t\right) &= -\sum_{t_i^{\mathrm{f}} \leq t} \eta\left(t - t_i^{\mathrm{f}}\right) \, , \\ h_i^{\mathrm{syn}}\left(t\right) &= \sum_j \sum_{t_j^{\mathrm{f}} \leq t} J_{ij} \varepsilon\left(t - t_j^{\mathrm{f}}\right) \, , \end{aligned}$$

and t_i^{f} denotes the firing times of neuron i.

2.2 Network Setup and Simulation Algorithm

We have designed a network of spike response neurons that is intended to model (a) excitatory and inhibitory cells in a small patch of layer 4 in primary visual cortex as well as (b) excitatory cells in a corresponding patch of LGN, providing input to cortical layer 4. The full network is separated into three equally sized square grids with periodic boundary conditions; see Fig. 2.1.

Cortical excitatory and inhibitory cells make up two of these grids. We have separated them into distinct layers so as to clarify the connectivity structure within the network; we do not presume that such a separation is present anatomically. As we assume that every neuron has a limited arborization range of axon collaterals and dendrites within layer 4, each of the excitatory neurons receives lateral connections from a region of neighbouring excitatory cells as well as from a region of inhibitory cells. This region has been chosen to be a circle centered on the cell under consideration and extending over 11 cells in diameter. In the present version of our model, inhibitory neurons also

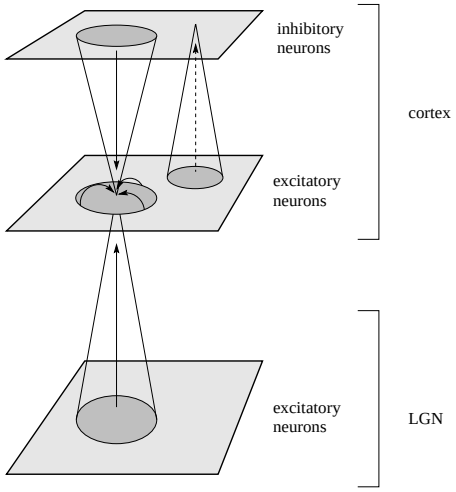


Figure 2.1: The network model we use to study intracortical and feedforward learning dynamics can be decomposed into three equally sized square grids of spike response neurons. The two upper layers are designed to represent a small patch of layer 4 in primary visual cortex. Excitatory and inhibitory neurons have been separated into distinct layers only for a clear illustration of the connectivity structure. We do not presume that such a separation is present anatomically. Each of the excitatory cortical neurons receives lateral input from neighbouring excitatory and inhibitory cortical cells. Inhibitory neurons are driven by the activity of excitatory neurons. Feedforward input into the cortex is provided by LGN neurons in the bottom layer, representing geniculate relay cells. They produce correlated spike activity of given statistics. All synaptic connections except for those from excitatory to inhibitory cells (dashed arrow) are subject to activity-driven learning dynamics. Periodic boundary conditions have been applied throughout each layer. In that part of our simulations where we consider cortical learning dynamics alone (Sect. 2.5), the LGN grid is not present so that we are left with the two upper layers only.

receive lateral excitatory input from within a neighbourhood of 11 neurons in diameter, but do not make synapses with other inhibitory cells. They produce an inhibitory background of spikes that is used to normalize overall network activity. The number 11 is not realistic but suffices for simulation purposes.

A third layer in our setup represents geniculate cells, providing input to the cortex. Since our current investigations are restricted to the case of monocular input, we need only consider neurons located within one eye-specific lamina in the LGN, say lamina A. Every excitatory cortical neuron obtains input from geniculate neurons lying within a circular arborization area centered at the retinotopic position of the cortical cell. Again the diameter of the circle extends over 11 neurons. To save computation time, which is a really limiting factor of our simulations, we only model LGN cells with ON-center receptive fields and do not account for OFF-center cells explicitly. It will become clear in the following why this is possible. All synaptic connections in the network, except for those from excitatory to inhibitory cells (dashed arrow), are subject

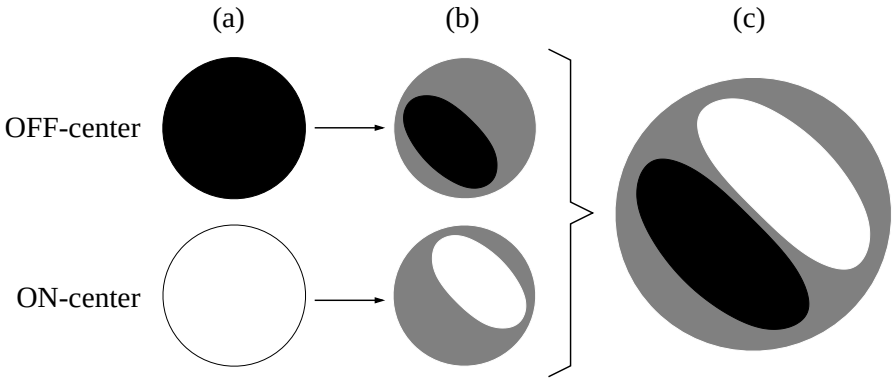


Figure 2.2: Emergence of an orientation selective simple-cell receptive field through competition between inputs from LGN neurons with ON- and OFF-center receptive fields (Linsker 1986b, Miller 1994, Wimbauer et al. 1997b, Wimbauer et al. 1997a, Erwin and Miller 1998). Initially the cell receives input of both ON- and OFF-center type from a circularly symmetric region in the visual field, indicated by a white and a black circle, respectively (a). In the course of a symmetry breaking process both input types become restricted to non-isotropic subfields (b). This symmetry breaking can be thought of as occurring separately for ON- and OFF-center inputs. An assumed anticorrelation between the activities of the two geniculate cell types ensures that the subfields are in anti-phase with each other. Together they form the receptive field shown on the right (c), where a black or a white shading indicates whether dominant input is from OFF- or ON-type cells.

to activity-driven learning dynamics.

In many existing models of correlation-based geniculocortical self-organization (Linsker 1986b, Miller 1994, Wimbauer et al. 1997b, Wimbauer et al. 1997a, Erwin and Miller 1998), orientation selectivity of simple cells develops through a competition between inputs from LGN neurons with ON-center and OFF-center receptive fields. Correlated random activity of geniculate cells is the driving force for this competition. The typical simple-cell receptive field structure emerging from such a model can be understood as the result of a symmetry-breaking process taking place separately for ON- and OFF-center inputs. This is illustrated in Fig. 2.2 from left to right. There are three stages. (a) Initially a simple cell receives input of both ON- and OFF-center type from a circularly symmetric region in the visual field, indicated by a white and a black circle, respectively. (b) At the end of the symmetry breaking process both input types are restricted to non-isotropic receptive fields, which we will call *subfields*. An assumed anticorrelation between the activities of ON-center and OFF-center geniculate neurons ensures that the subfields are in anti-phase with each other. Together they form the receptive field shown on the right (c), where a black or a white shading indicates whether dominant input is from OFF- or ON-type cells.

In our model, like in previous ones, the emergence of orientation selectivity in the geniculocortical projections is driven by correlated random activity

of LGN cells. Since, however, ON-OFF competition can be viewed as two coupled but separable symmetry-breaking processes, we take into account only one geniculate cell type, with ON-center receptive fields, say. It should be emphasized, though, that we do *not* presume that OFF-center LGN cells do not contribute to orientation selectivity. Neglecting OFF-center inputs in our model is just a computational shortcut. Because of the aforementioned anticorrelations between ON-center and OFF-center activities, the complete receptive field structure shown in Fig. 2.2 (c) is fully determined by either one of the two substructures displayed in Fig. 2.2 (b). It is therefore sufficient to study the development of geniculocortical synapses from either ON- or OFF-center LGN cells. The lower part of Fig. 2.2 (b) gives a schematic view of a simple cell's typical feedforward input connectivity emerging in our reduced model.

In short, the simulation algorithm that has been applied is as follows. At the beginning of each time step, the firing probability of every neuron is determined from its membrane potential h. To be specific, we let

$$\begin{aligned} &\Pr\{\text{Spike during } \Delta t \mid h\} \approx q(h)\,\Delta t \\ &:= \quad \{1+\exp[-(h-\theta)/T]\}^{-1} \\ &= \quad (1/2)\{1+\tanh[(h-\theta)/2T]\}\ , \end{aligned}$$

with $\Delta t = 1$ ms denoting the size of the time step, θ being the neural threshold and T a noise parameter.

Next, it is decided randomly from the firing probability whether a neuron emits a spike or remains silent during the current time step. Subsequently, the membrane potential of each cell is updated using the spike trains of all the neurons projecting onto it. Both the postsynaptic potential $\varepsilon(t)$ and the refractory potential $\eta(t)$ have been chosen to be exponentials (see Fig. 2.3),

$$\begin{aligned} \varepsilon(t) &= \exp(-t/\tau_\varepsilon)\ , \\ \eta(t) &= \eta_0 \exp(-t/\tau_\eta)\ , \end{aligned}$$

for $t > 0$ and vanishing for $t \leq 0$, with $\tau_\varepsilon = 6$ ms, $\tau_\eta = 10$ ms, and $\eta_0 = 10$. At the end of each time step, all synaptic weights are modified as will be explained in the next section.

The membrane potentials of the LGN neurons are prescribed externally, in order to simulate correlated thalamic activity. Their values are drawn from a two-dimensional Gaussian random field. A new realization of this random field is generated every 10 time steps. In this simple way we obtain spatially and temporally correlated spike trains from LGN neurons.

Depending on the simulation, these steps are repeated for $2\,500\,000$ or $5\,000\,000$ iterations (see simulation parameter values in Sect. 2.8), which corresponds to roughly 40 or 80 minutes in real time.

2.3 Learning Mechanisms

There is increasing evidence that synaptic plasticity in the developing brain is to a large extent dependent on neural activity (Wiesel and Hubel 1963b,

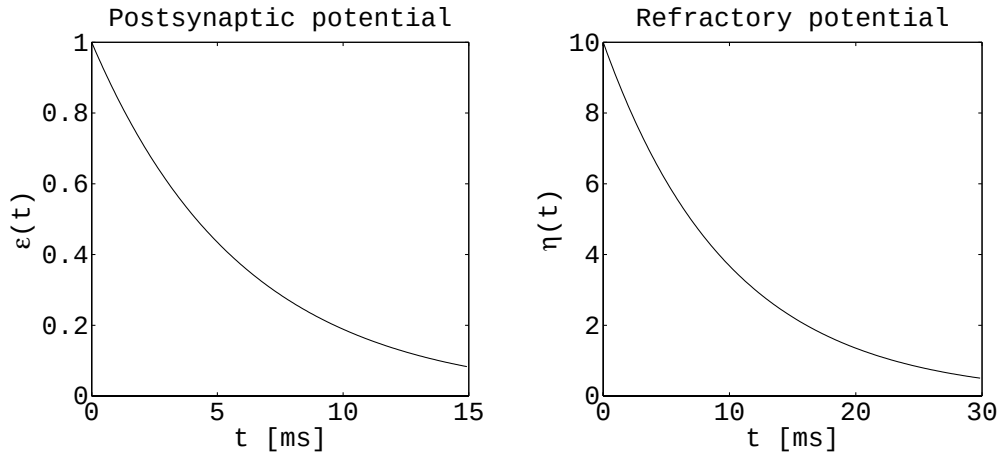


Figure 2.3: In the spike response model (cf. Sect. 2.1) every incoming spike induces a transient change of the postsynaptic neuron's membrane potential. This postsynaptic potential is described by a response function $\varepsilon(t)$, which we have chosen to be an exponential as displayed in the left panel. To model refractoriness, a second response kernel $\eta(t)$, shown in the right panel, is subtracted from the membrane potential after spike emission.

Crair et al. 1998, Issa et al. 1999). Originally, Hebb (1949) postulated that the efficacy of a connection between two cells is increased, if the postsynaptic cell is repeatedly activated by the presynaptic one. In the meantime, many studies have shown that by appropriately coactivating a pair of neurons, the strength of their connection can indeed be modified (Brown and Chattarji 1994, Fregnac et al. 1994).

In the model presented in this chapter, we have incorporated plasticity of both excitatory and inhibitory connections. Modification of *excitatory* synapses is governed by the following rules.

(1) A Hebbian mechanism increases synaptic efficacy, whenever a presynaptic action potential is immediately followed by a postsynaptic one (see the learning window in Fig. 2.4).

(2) Each time a neuron emits a spike, the weights of all its incoming synapses are reduced by a certain amount. Such a process prevents a neuron from enhancing its inputs ad infinitum, since growing input increases the neuron's firing rate, which in turn diminishes the weight of incoming synapses. As a consequence, different input synapses of the same neuron have to compete for synaptic weight, because an increased efficacy of one group of synapses leads to a down-regulation of the remaining ones due to the cell's increased firing rate. Experimental support for this mechanism has been provided by work of Turrigiano et al. (1998), who have found the total input strength of rat cortical pyramidal cells to be increased or decreased as a function of activity.

Taken together, the rules (1) and (2) effect a synapse to be up-regulated,

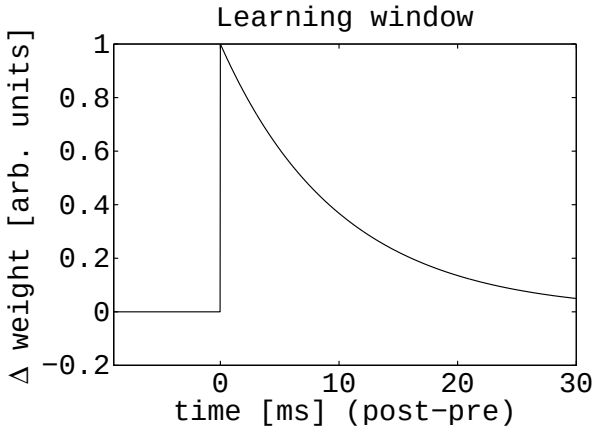


Figure 2.4: Learning window $W(t)$ for excitatory synapses. In our model, an excitatory synapse is strengthened, whenever a presynaptic spike is immediately followed by a postsynaptic one. The graph displays the change of synaptic weight (in arbitrary units) that is applied in dependence of the time difference between post- and presynaptic spike.

when a postsynaptic spike immediately follows a presynaptic one, and to be down-regulated otherwise. This is at least qualitatively in agreement with experimental findings (Markram et al. 1997, Zhang et al. 1998).

(3) The efficacy of every synapse is slowly reduced at a rate proportional to its current weight. Recently, Engert and Bonhoeffer (1999) gave some indirect evidence for such an activity-independent decay in rat hippocampal slice cultures.

(4) Each synaptic weight is enhanced at a constant rate, independent of neuronal activity. Together with rule (3) this means that without activation a synapse will slowly approach some non-zero efficacy. This process can be considered as an activity-independent formation of synapses that is driven by some sort of nerve growth factor.

(5) Finally, we limit each synaptic weight to a finite range between 0 and some upper bound. Note, however, that in contrast to some previous models (Miller 1994, Wimbauer et al. 1997b, Wimbauer et al. 1997a, Erwin and Miller 1998) these bounds are not sticky, i.e., all synapses remain plastic, whether they are saturated or not.

With $J^{\mathrm{e}}_{ij}(t)$ denoting the excitatory weight from neuron j to neuron i, the change $\Delta J^{\mathrm{e}}_{ij}(t)$ that results from contributions (1) to (4) to this weight during the current time step can be summarized in the following formula,

$$\Delta J^{\mathrm{e}}_{ij}(t) \;=\; A^{\mathrm{e}}_{ij}\Bigg[\underbrace{a_i(t)\sum_{t^{\mathrm{f}}_j<t} W\left(t-t^{\mathrm{f}}_j\right)}_{(1)} + \underbrace{a_i(t)\,\sigma^{\mathrm{e}}}_{(2)} + \underbrace{\xi^{\mathrm{e}}}_{(4)}\Bigg]$$

$$- \underbrace{\vartheta^{\mathrm{e}}\, J_{ij}^{\mathrm{e}}(t)}_{(3)} \; . \tag{2.1}$$

In this expression, $W(t)$ denotes the learning window (see Fig. 2.4)

$$W\,(t) = \exp\left(-t/\tau_W\right) \; ,$$

where $\tau_W = 11$ ms, and $a_i(t)$ is the activity of neuron i at the current time step, with $a_i(t) = 1$ if an action potential is emitted and $a_i(t) = 0$ otherwise. As in Sect. 2.1, the t_i^{f} indicate the firing times of cell i. Below each term, the corresponding number of the above list of items is given.

In real cortex, the number of synapses connecting two neurons as a function of their distance is probably a random quantity (Braitenberg and Schüz 1991). A similar argument holds for the connections from thalamic relay cells to cortical neurons and their respective retinotopic coordinates. However, in our simulations we neglect the statistical nature of this process. Instead, we have introduced an 'arbor function' A_{ij}^{e}, expressing the expected number of synapses between the neurons j and i. Within this framework, J_{ij}^{e} is taken to be the effective weight connecting cell j to cell i, i.e., the efficacy of a single synapse multiplied by the number of synapses from j to i.

In addition to the above learning rules, an unspecific spread of weight changes onto neighbouring but non-activated synapses can be included into the model, as is suggested by measurements of Engert and Bonhoeffer (1997). Although the results presented in this chapter have been obtained without this kind of plasticity, we have performed a corresponding set of numerical simulations where such an unspecific modification of excitatory synapses has been taken into account. Except for a smoothening of the emerging connectivity patterns, we did not find a qualitative change of our results.

To control overall activation, i.e., to avoid epileptiform bursts, our network takes advantage of the subsequent learning rules for inhibitory connections,

(1) Whenever an excitatory neuron of the cortical layer generates an action potential, its incoming inhibitory synapses are enhanced by a certain amount. This implements a negative feedback loop with a higher firing rate leading to a cell receiving more inhibition, thereby limiting its activity.

(2) The efficacy of every synapse is slowly reduced at a rate proportional to its current weight.

(3) The upper bound 0 prevents inhibitory weights from becoming excitatory. In contrast to excitation, there is no lower bound for inhibition.

Using the same notation as above and neglecting the upper bound 0, the change of the inhibitory efficacy $J_{ij}^{\mathrm{i}}(t)$ that results during one time step can be written

$$\Delta J_{ij}^{\mathrm{i}}(t) = -A_{ij}^{\mathrm{i}} \underbrace{a_i(t)\,\sigma^{\mathrm{i}}}_{(1)} - \underbrace{\vartheta^{\mathrm{i}}\, J_{ij}^{\mathrm{i}}(t)}_{(2)} \; . \tag{2.2}$$

Again, we have used an arbor function A^{i}_{ij} to express the expected number of synapses that the inhibitory neuron j makes onto cell i. It should be noted that, within the framework of the spike response model, inhibitory synapses are represented by a negative coupling constant J_{ij}. To strengthen an inhibitory synapse thus means to make its weight J_{ij} more negative, which is the reason for the negative sign in front of the first term.

In order to allow inhibition to respond sufficiently fast to changes of excitation during synaptic development, parameters have been chosen so that the typical relaxation time of inhibitory weights is significantly shorter than that of excitatory weights. More specifically, we let $\vartheta^{\mathrm{e}} = 2.5 \times 10^{-6}$ for excitatory intracortical synapses, $\vartheta^{\mathrm{e}} = 1.25 \times 10^{-6}$ for geniculocortical synapses, and $\vartheta^{\mathrm{i}} = 10^{-4}$ for inhibitory synapses, corresponding to relaxation time constants of 400 s, 800 s, and 10 s, respectively.

Comparing these time constants with the typical time course of pattern formation in the visual system of higher mammals during the critical period, it is obvious that nature takes more time for development than we do in our simulations. Unfortunately, numerical simulations with *spiking* neurons are computationally very expensive, meaning that at the moment it is impossible to run simulations of suitably sized networks representing periods of a week or more in real time. We thus have to reduce the relevant time constants and speed up learning. According to our experience, however, the process of pattern formation does not become instable, but rather more stable as learning is slowed down. This is reasonable because reducing the speed of learning means averaging over a larger number of pre- and postsynaptic spikes, which reduces the effects of noise. We therefore think that the results presented throughout the subsequent sections could be reproduced with biological learning time constants by simply rescaling time.

2.4 Data Analysis

In the following sections we will investigate the connectivity structures emerging from the learning mechanisms in the network described so far. To this end, we will focus on the incoming excitatory synapses of all excitatory neurons. As it turns out, these coupling structures very often show an elongated shape of a certain orientation. Here the method is described that we have applied to determine these orientations.

In our model, the excitatory input efficacies of one cortical cell, say at location $(0, 0)$, are given as a two-dimensional synaptic array $\tilde{J}_{xy}$ of 11×11 positive real values. Here the coordinates (x, y) can range through either the geniculate or the cortical layer while the reference point $(0, 0)$ is in V1; cf. Fig. 2.1. In order to obtain the orientation of a connectivity pattern, we calculate the overlap of the corresponding synaptic array with a set of Gaussian 'bars' S_{xy},

$$S_{xy}(\phi, p) := \exp\left\{ \frac{-\left[x\cos\phi + y\sin\phi - p\right]^2}{2 \cdot 0.5^2} \right\}$$

$$\times \quad \exp\left\{\frac{-\left[-x\sin\phi + y\cos\phi\right]^2}{2\cdot 4^2}\right\} - S_0(\phi,p)\ ,$$

where $x, y \in \{-5,-4,\ldots,5\}$. The parameter ϕ determines the orientation of the bar and p controls its position within a frame of 11×11 pixels. $S_0(\phi,p)$ is chosen so that $\sum_{xy} S_{xy}(\phi,p) = 0$. For each of the four orientations $\phi \in \{0°, 45°, 90°, 135°\}$, p is varied to determine the maximal overlap

$$R(\phi) := \max_p \left[\sum_{x,y} \tilde{J}_{xy}\, S_{xy}(\phi,p)\right]\ .$$

These values are then taken as the lengths of four vectors $\mathbf{x}(\phi)$ pointing in the directions given by $2\,\phi$, i.e.,

$$\mathbf{x}(\phi) := [R(\phi), 2\phi]$$

in polar coordinates. We sum the four vectors so as to obtain a polar vector

$$\mathbf{y} = (R_y, \psi_y) := \mathbf{x}(0°) + \mathbf{x}(45°) + \mathbf{x}(90°) + \mathbf{x}(135°)\ ,$$

and take the resulting polar angle ψ_y divided by 2 as the required orientation $\phi^{\rm or}$ of the synaptic weight pattern,

$$\phi^{\rm or} := \psi_y/2\ .$$

The described procedure is a formal method to analyze the neuronal connectivity. It does not directly yield the cells' preferred orientation as it would be measured in an optical imaging experiment. To achieve this, we would have to apply test stimuli to the geniculate layer of our network and record the activity in the cortical layer. The obtained neural response properties would then be determined by the feedforward connectivity, the lateral connectivity, and the nonlinear gain function. In this work, however, we focus on the developmental interplay of feedforward and lateral projections and thus have to investigate their structures separately.

2.5 Intracortical Self-Organization

Anatomical studies of the cat's developing visual system (Shatz and Luskin 1986, Gosh and Shatz 1992) have revealed that geniculate axons have reached the cortical subplate by embryonic day (E) 36 but do *not* enter the future cortex during the following week days. Although most of the cells destined for cortical layer 4 have finished their migration by E55, a geniculate projection to layer 4 could only be detected by E60. Thus, for layer 4 of area 17 in the cat, there is likely to be a waiting period of about 1 week between the completion of neuronal migration and thalamic innervation. This suggests the possibility of an intracortical synaptic refinement going on for about one week *without* feedforward input from LGN.

As long as the primary visual cortex does not receive external input, spike-spike correlations of spontaneous activity are determined by lateral interactions. If the lateral connections are shaped by an activity-driven, Hebbian learning rule, as we assume in our model, their development is in turn determined by the cortical spike-spike correlations. It is therefore an interesting question, what the connectivity patterns emerging under these conditions will look like.

In order to investigate this, we used the network described in Sect. 2.2 but removed the geniculate layer. It turned out that under certain conditions depending on network activity the development of excitatory lateral projections can undergo a symmetry breaking process As a result, every neuron will receive most of its lateral input from neighbouring excitatory cells lying within an elongated region of a certain orientation. Without this symmetry breaking process the emerging patterns of connectivity will be rotationally invariant.

Both cases are illustrated schematically for one excitatory cell in Fig. 2.5. The panels (a) to (c) display different developmental stages of the cell's intracortical input as they can arise in our model. As we have explained in Sect. 2.2, the cell receives lateral excitatory projections from neurons within a circular arborization area. In the different panels of the figure, the outline of this area is shaded according to the synaptic efficacy of the corresponding inputs. Projections from dark regions are weak whereas those from light regions are relatively strong. At the beginning of our simulations, all synaptic weights are zero (a). As the development proceeds, the neuron continually receives more and more input from the region within its circular arborization radius (b). At some point the abovementioned symmetry breaking process may lead to the emergence of an oriented connectivity pattern (bottom of panel c). Otherwise the input region will remain rotationally symmetric (top of panel c).

Figure 2.6 presents the outcome of two different runs of a network comprising 16×16 excitatory and an equal number of inhibitory neurons. Panels (a) and (b) visualize the final excitatory lateral input connectivities by means of two arrays of 16×16 grey level plots. In these plots, each of the small squares consists of 11×11 pixels, representing one cell's incoming synaptic weights from 11×11 neighbouring neurons. As in Fig. 2.5, dark and white shaded pixels indicate low and high synaptic efficacies, respectively. Lateral input is restricted to a circular region of 11 cells in diameter, as explained in Sect. 2.2, i.e, synaptic weights from neurons outside this circle are zero and the corresponding pixels are black.

For the first network run (a), we have chosen a low value for the parameter ξ^{e} in Eq. (2.1). The emerging coupling patterns are rotationally symmetric. For the second run (b), ξ^{e} has been increased, resulting in higher synaptic efficacies on average. In this case, rotational symmetry is broken during the learning process and the emerging connectivity patterns obtain an elongated shape.

To each of these elongated patterns an orientation can be assigned as described in Sect. 2.4. Figure 2.6 (c) visualizes the resulting array of orientations. For every neuron, a small rectangle is plotted with a color coding for the orientation of the corresponding input connectivity pattern. The continuous color code that has been applied is indicated for a few orientations below the plot.

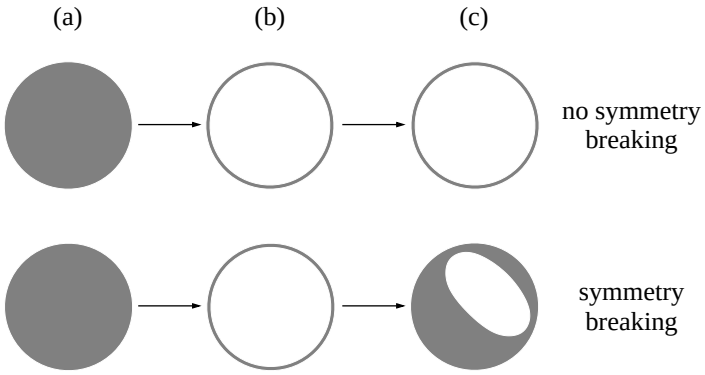


Figure 2.5: The panels (a) to (c) display different developmental stages of a cortical cell's intracortical input that can arise in our model. The cell receives lateral excitatory projections from neurons within a circular arborization area. In the different panels of the figure, the outline of this area is shaded according to the synaptic efficacy of the corresponding inputs. Projections from dark regions are weak whereas those from white regions are relatively strong. At the beginning of our simulations, all synaptic weights are zero (a). As the development proceeds, the neuron continually receives more and more input from within its circular arborization radius (b). Under certain conditions depending on network activity an oriented connectivity pattern will emerge due to a symmetry breaking process (bottom of panel c). Otherwise the input region will remain rotationally symmetric (top of panel c).

In Fig. 2.7 we display the lateral projections that have emerged in a simulated network whose grid size was 32×32. Panel (a) is again a set of grey level plots showing each cells synaptic input weights. Panel (b) is a color map of the orientations in the intracortical connectivity. The color code is the same as in Fig. 2.6.

The global structure of the map resembles that of typical orientation maps obtained from optical imaging experiments in primary visual cortex (Bonhoeffer and Grinvald 1991, Blasdel 1992b, Bonhoeffer and Grinvald 1993, Chapman et al. 1996, Bosking et al. 1997). This is even more obvious in panel (c), where a smoothened version of panel (b) is presented. It can be seen clearly that orientation normally changes continuously across the surface of the simulated cortical patch. The only exceptions are point-like singularities where orientation changes by 90 degrees. Such singularities are well-known from experimental maps of preferred orientation and have been termed "orientation centers" or "pinwheel centers" (Bonhoeffer and Grinvald 1993). Around each of these pinwheel centers, every orientation is represented once. Two kinds of pinwheels can be distinguished according to whether orientation changes clockwise or counterclockwise around the center. Both kinds appear in approximately equal numbers per unit area of simulated cortical surface. This is again in agreement with experimental findings.

Our simulations demonstrate that a Hebbian learning mechanism driven by spontaneous activity can generate oriented patterns of intracortical con-

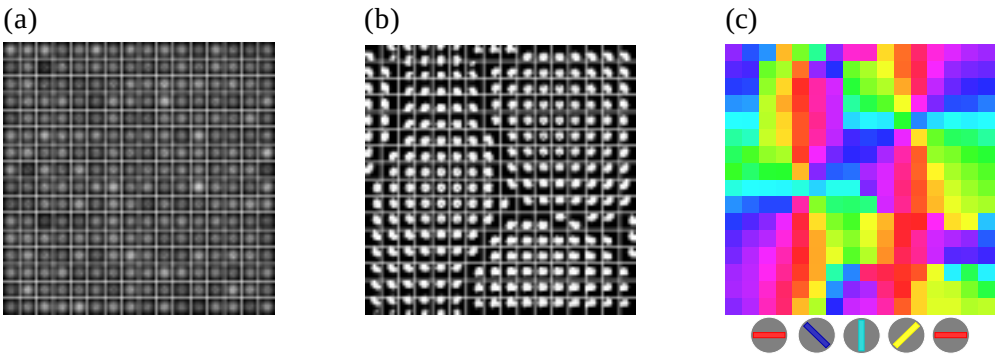


Figure 2.6: The figure shows the outcome of two different runs of a network *without* feedforward input. It comprises 16×16 excitatory and an equal number of inhibitory neurons. Panels (a) and (b) visualize the final excitatory lateral connectivities by means of two arrays of 16×16 grey level plots. In these plots, each of the small squares consists of 11×11 pixels, representing one cell's incoming synaptic weights from 11×11 neighbouring neurons. Dark and bright shaded pixels indicate low and high synaptic efficacies, respectively. For run (a), we have chosen a low value of the parameter ξ^e in Eq. (2.1). The emerging coupling patterns are rotationally symmetric. For run (b), ξ^e has been increased, resulting in higher synaptic efficacies on average. In this case, rotational symmetry is broken during the learning process and the emerging connectivity patterns obtain an elongated shape. As described in Sect. 2.4, an orientation can be assigned to each of these elongated patterns. In panel (c), we visualize the resulting array of orientations. For every neuron, a small rectangle is plotted with a color coding for the orientation of the corresponding input connectivity pattern. The continuous color code that has been applied is indicated for a few orientations below the plot.

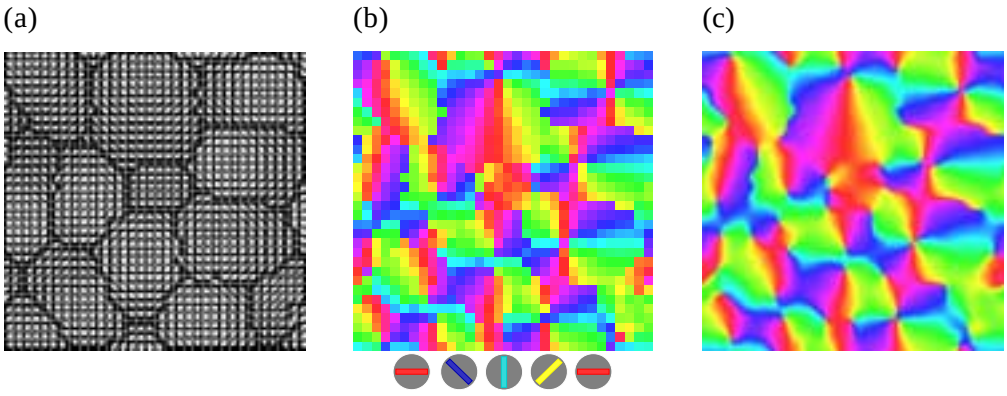


Figure 2.7: The intracortical projections that have emerged from Hebbian learning in a network without geniculocortical input and with a grid size of 32×32 cells are presented as grey level plots in panel (a) and as a colored orientation map in panel (b). Panel (c) is a smoothened version of (b). The color code that has been used is shown below panel (b).

nectivity. The overall organization of this connectivity can be considered as an intracortical orientation map that is similar to measured maps of preferred orientation in striate cortex. This suggests that lateral projections may play a crucial role in the shaping of orientation-selective response properties of cells in the primary visual cortex.

2.6 Combined Feedforward and Intracortical Plasticity

We are now going to investigate the effect that the existence of an intracortical orientation map can have on the development of geniculocortical feedforward projections. To this end, we activate the geniculate layer (that had been removed for the analysis in Sect. 2.5) and let the geniculocortical connections evolve according to the Hebbian learning dynamics described in Sect. 2.3. During that process, the pattern of excitatory intracortical projections is kept fixed, whereas inhibitory synapses remain plastic. This inhibitory plasticity allows for a normalization of overall network activity on short time scales and thus helps to avoid epileptiform discharges by shifting the balance of cortical excitation and inhibition towards an increased inhibition (Varela et al. 1999, Douglas et al. 1995).

As a first step we assume the lateral excitatory connections to have rotational symmetry, which corresponds to the scenario studied in earlier work (Linsker 1986a, Miller 1994, Wimbauer et al. 1997b, Wimbauer et al. 1997a, Erwin and Miller 1998). These analyses were based on graded-response neurons and did not distinguish between excitatory and inhibitory cortical cells, whereas the present network consists of spiking neurons with inhibition mediated via inhibitory interneurons. It is therefore important to check whether previous results can be reproduced in our setup, which is closer to biology.

Figure 2.8 (a) visualizes each excitatory cell's intracortical synaptic efficacies in the form of grey-level plots (see Fig. 2.6 for details). Grid size is 32×32 cells. As mentioned above, the coupling patterns have been chosen to be rotationally invariant, viz. shaped as a two-dimensional Gaussian. The result of a simulation of geniculocortical development in this setup is presented in Fig. 2.8 (b). It shows an array of 32×32 grey level plots, each corresponding to the feedforward input connectivity of one excitatory cortical neuron. All these small plots consist of 11×11 pixels representing the efficacies the input synapses that an excitatory cortical neuron receives from LGN. A dark shaded pixel means that the corresponding synapse is weak, whereas a white shaded pixel indicates a strong synapse. All pixels outside a circle of diameter 11 are black, because the arborization area of the geniculocortical projection is a circle of that diameter, i.e., the synaptic efficacy from any LGN cell outside this area is zero.

Obviously, most of the excitatory cortical neurons receive their feedforward input from elongated patches on the LGN grid. As we have set forth in Sect. 2.2, these patches can be considered as the ON-center part of an orientation selective simple-cell receptive field. To extract the orientation of the connectivity pattern we apply the same method that we have used for the in-

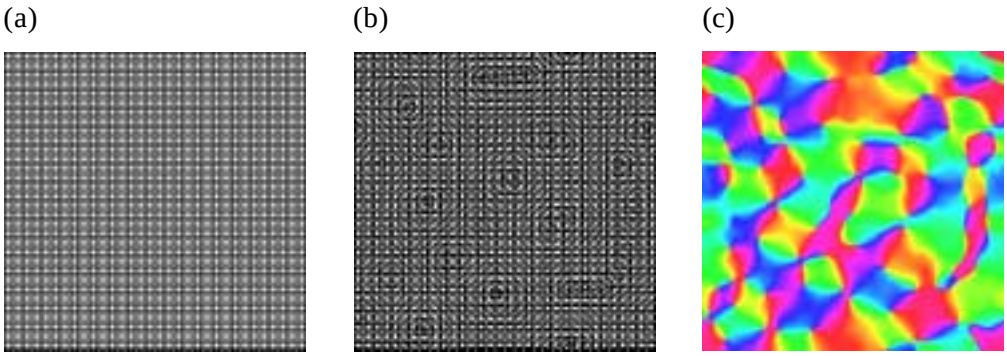


Figure 2.8: An activity-driven development of geniculocortical afferents in the presence of an isotropic intracortical connectivity leads to the emergence of a feedforward orientation map; grid size is 32×32. This confirms the results of earlier studies carried out in models that used graded-response neurons and / or did not distinguish between excitatory and inhibitory cells (Linsker 1986a, Miller 1994, Wimbauer et al. 1997b, Wimbauer et al. 1997a, Erwin and Miller 1998, Choe and Miikkulainen 1998). For each excitatory cortical cell, panel (a) displays the lateral input strengths as a grey level plot, and similarly panel (b) visualizes the efficacies of the feedforward input synapses from the LGN. Extracting orientations as described in Sect. 2.4, this feedforward connectivity pattern can be transformed into a color coded orientation map, which is shown in a smoothened version in panel (c). The color code that has been used is given in Figs. 2.6 (c) and 2.7 (b).

tracortical projections and which has been explained in Sect. 2.4.

Using again the color code shown in Figs. 2.6 (c) and 2.7 (b) we obtain the feedforward orientation map presented in Fig. 2.8 (c). We find that the map is in a good qualitative agreement with experimental data (Bonhoeffer and Grinvald 1991, Bonhoeffer and Grinvald 1993). This demonstrates that the results of previous models of correlation-based geniculocortical development can be nicely reproduced in our more detailed approach.

Now we turn to the case of an anisotropic intracortical connectivity. We assume this connectivity to form an orientation map which has emerged during a preceding learning process as described in Sect. 2.5. To be specific, we use the map of Fig. 2.7. The efficacies of the intracortical excitatory synapses are fixed at their respective values while the spontaneous activity of the thalamic relay cells drives the development of geniculocortical feedforward projections. The grey-level coded results are presented in Fig. 2.9.

Panel (a) reproduces the intracortical map. Panel (b) displays the final feedforward orientation map. Its overall organization is very similar to that of the intracortical map. That means that, in our model, a pre-existing intracortical connectivity can guide the synaptic refinement of geniculocortical afferents. This may provide a natural explanation for the enormous stability of orientation maps that has been found in experiments disturbing normal geniculocortical development (Kim and Bonhoeffer 1994, Weliky and Katz 1997, Gödecke

(a) (b)

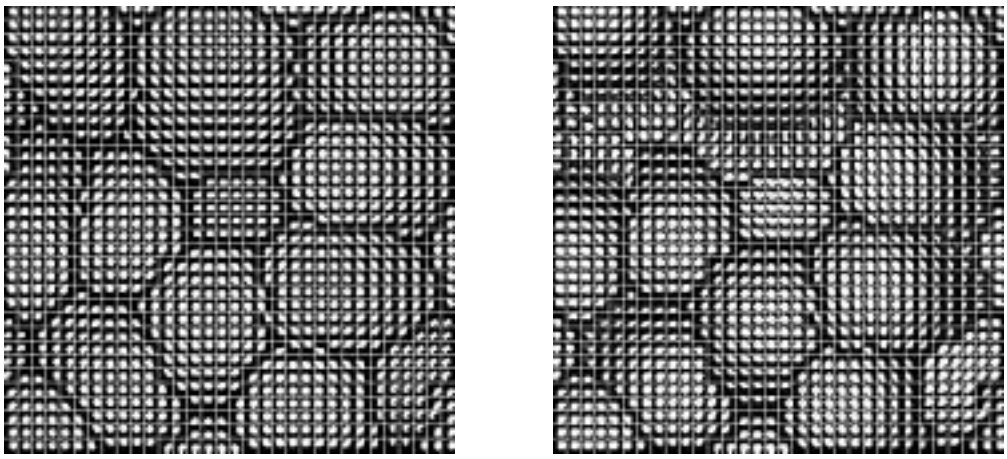


Figure 2.9: Panel (a) reproduces the intracortical map of Fig. 2.7. When feedforward connectivity develops in the presence of such an *intracortical* map, the resulting *feedforward* orientation map – shown in panel (b) – is essentially in agreement with the intracortical pattern, as can be seen from the similarity between the two images.

and Bonhoeffer 1996, Sengpiel et al. 1998).

2.7 Simulating Reverse Lid-Suture

Models explaining cortical orientation maps to emerge from correlation-based development of feedforward projections have been challenged by the outcome of so-called reverse-lid-suturing experiments (Gödecke and Bonhoeffer 1996, Sengpiel et al. 1998). In these experiments, kittens were raised so that both eyes never received visual input at the same time. This was achieved using the following protocol. Immediately after birth, one eye-lid was sutured and the animal received monocular input through the other eye. After a period of a few weeks, the cortical orientation map in area 18 was recorded via optical imaging through the open eye. Then the open eye was closed and the initially closed eye was opened. After one or two weeks the second eye's orientation map in the same cortical area was measured. The two maps turned out to be nearly identical.

Assuming the activity correlations in the LGN to be chiefly determined by input from the retinae, the models predict the left-eye map and the right-eye map to develop independently of each other. Their global layouts can then be expected to be different, in striking contrast to the experimental finding.

In the meantime there are two different theoretical approaches to explain the experimental result. The first is based on a geometric argument (Wolf et al. 1996). As the authors pointed out, the cat's visual area 18 is shaped as a narrow band on the cortical surface. The formation of an orientation map in this area is therefore subject to strong confinement, i.e., the layout of a developing

map is predetermined by boundary conditions. Numerical simulations have confirmed that different orientation maps emerging under this condition are indeed very similar.

The second explanation relies on activity correlations in the LGN. Erwin and Miller (1998) have shown that with an appropriate amount of inter-eye correlations in the activity of geniculate relay cells, a Hebbian development of geniculocortical afferents can nicely produce ocularly matched cortical orientation maps. They assume that these correlations are present in LGN activity even without visual input. Recently it has been shown that strong inter-eye correlations can in fact be found in the ferret LGN before eye opening (Weliky and Katz 1999).

It is also well-known, however, that monocular deprivation over a period of a few days during the critical period causes most cortical neurons to loose their responsiveness to stimulation of the deprived eye (Blakemore and van Sluyters 1974). As a consequence, optical imaging experiments find the orientation map of the deprived eye to be largely eliminated (Kim and Bonhoeffer 1994, Gödecke and Bonhoeffer 1996). This may indicate that LGN activity is strongly affected by monocular occlusion. According to the model of Erwin and Miller, ocularly matched orientation selectivity can be maintained in a reverse-suturing experiment if one of the two following conditions is fulfilled. Either the remainder of the deprived eye's map is strong enough so as to act as a seed for the restoration of the original map, or there is a sufficient amount of geniculate inter-eye correlations during the period of reverse occlusion so that the deprived map can be reinstated as a copy of the non-deprived map. It is still an open question whether one of these conditions does indeed hold true.

In consideration of our results presented in Sects. 2.5 and 2.6 we propose a third mechanism that may reconcile models of correlation-based geniculocortical development with experimental data. As we have shown, a Hebbian learning mechanism driven by spontaneous cortical activity may lead to the formation of an intracortical orientation map. This map is set up by lateral projections of cortical layer 4 neurons and can emerge very early in visual development – even before thalamic afferents reach cortical layer 4. The intracortical connectivity can then *guide* the refinement of thalamocortical inputs, so that the developing feedforward orientation map is in agreement with the intracortical map.

Let us now assume that the critical period for the development of short-range interactions in cat layer 4 ends after about three weeks postnatal. Within the framework of our model this would mean that at this time the layout of the intracortical orientation map is fixed, thereby fixing also the structure of both eyes' feedforward maps. The outcome of the reverse-suturing experiments can then be explained as follows.

In the course of – or possibly before – the first period of monocular deprivation, cortical layer 4 cells finish their refinement of short-range lateral connections. During the same period of time, a feedforward orientation map for the open eye emerges. At the end, the feedforward and the intracortical map are in accord with each other. Together they form the orientation map that can be recorded by means of optical imaging. After the open eye has been closed and

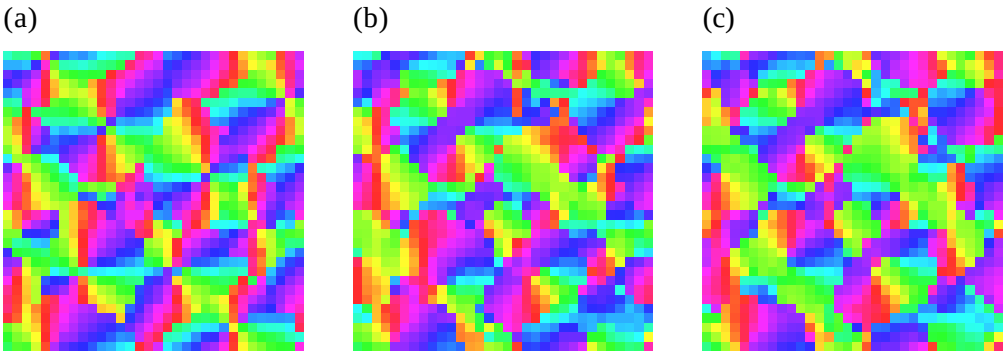


Figure 2.10: The reverse lid-suturing protocol is mimicked by a sequence of numerical simulations. Panel (a): Before the onset of geniculocortical development an intracortical orientation map is assumed to emerge from a Hebbian development of lateral connectivity in layer 4 as explained in Sect. 2.5. Panel (b): The geniculocortical afferents of the initially open eye are learned and produce a feedforward orientation map. Panel (c): After the first eye has been closed, the second eye's afferents develop independently from scratch. This is simulated by re-running the feedforward learning process but initialized with a different seed value of the random number generator. Comparing panels (b) and (c) reveals them to be very similar. This is in accordance with the results of reverse lid-suture experiments, in which both eyes' orientation maps have been found to be virtually identical (Gödecke and Bonhoeffer 1996, Sengpiel et al. 1998).

the closed eye has been opened, cortical neurons reorganize their geniculate input synapses so that an orientation map for the newly opened eye is formed. This process is again guided by the intracortical connectivity, which does not change any more. As a consequence, the second feedforward map develops in accord with the same intracortical map as did the first one. Optical imaging at the end of the experiment therefore yields a second orientation map that is very similar to the previous one.

This mechanism can be demonstrated in the following sequence of numerical simulations. First, the Hebbian development of short-range lateral connections in cortical layer 4 is simulated according to the explanations of Sect. 2.5. This process is driven by spontaneous cortical activity and for simplicity we assume that it is finished before the onset of thalamocortical input. The emerging intracortical map is displayed in Fig. 2.10 (a).

In the presence of this intracortical map, two independent network runs as described Sect. 2.6 are then performed to mimick the development of the two eyes' geniculocortical afferents during the reverse-suturing protocol. Starting the runs with different seed values for the random number generator ensures that their LGN activities are uncorrelated. Despite the lack of inter-eye correlations in thalamic activity the resulting maps are strikingly similar, as can be seen in panels (b) and (c) of Fig. 2.10. This similarity arises because both maps develop in such a way that they are in agreement with the intracortical

connectivity, which is the same during both runs.

Taken together, our results show that a patterned intracortical connectivity arising early in visual development can guide the development of feedforward projections so that the emerging cortical orientation maps are matched in the two eyes. In our simulations, there were no inter-eye correlations in thalamic activity. This provides evidence that in the present scenario inter-eye correlations are not necessary for ocularly matched orientation maps to develop.

2.8 Simulation Parameter Values

In order to allow the reader to reproduce our results, this section lists all the relevant simulation parameter values. A description of the simulation algorithm has already been given in Sect. 2.2.

2.8.1 Intracortical Plasticity

We have assumed the refinement of short-ranged connections within cortical layer 4 to occur before the onset of geniculocortical input. For the corresponding simulation runs the geniculate layer has thus been removed from our network (Fig. 2.1). The remaining cortical part consists of excitatory neurons and inhibitory interneurons arranged on two equally sized square grids.

The model neurons (c.f. Sect. 2.1) are determined by specifying the postsynaptic potential $\varepsilon(t)$, the refractory potential $\eta(t)$ and the firing probability for given value of the membrane potential h. We let (see Fig. 2.3)

$$\varepsilon(t) = \exp(-t/\tau_\varepsilon)\ ,\ \tau_\varepsilon = 6\text{ms}\ , \tag{2.3}$$

$$\eta(t) = \eta_0 \exp(-t/\tau_\eta)\ ,\ \tau_\eta = 10\text{ms}\ ,\ \eta_0 = 10\ , \tag{2.4}$$

$$\Pr\{\text{Spike in } \Delta t \mid h\} = \{1 + \exp[-(h-\theta)/T]\}^{-1}\ , \tag{2.5}$$

and

$$\begin{aligned} &\theta = 3\ ,\ T = 0.5 \qquad \text{for excitatory cells,} \\ &\theta = 3\ ,\ T = 0.5 \qquad \text{for inhibitory neurons.} \end{aligned}$$

As mentioned in Sect. 2.2, the size of one simulated time step is $\Delta t = 1$ ms.

Within a circular arborization region of 11 neurons in diameter, the excitatory neurons connect to one another as well as to the inhibitory neurons, and the latter project back to the excitatory cells. Thus there is no connection between two cells if the distance of their grid positions is larger than $11/2$.

Within this radius the connections from excitatory to inhibitory neurons are kept at fixed values

$$J_{ij} = 0.3\ \exp\left[-\mathrm{d}(i,j)^2\ /\ \left(2\cdot 3^2\right)\right]\ ,$$

with $\mathrm{d}(i,j)$ denoting the distance between the grid positions of inhibitory neuron i and excitatory neuron j.

#	Figure	Grid size	time steps	σ^{e}	ξ^{e}
1	6 (a)	16×16	$2.5 \cdot 10^6$	-0.57	$8.0 \cdot 10^{-4}$
2	6 (b), (c)	16×16	$2.5 \cdot 10^6$	-0.57	$9.5 \cdot 10^{-4}$
3	7, 9 (a)	32×32	$2.5 \cdot 10^6$	-0.57	$9.5 \cdot 10^{-4}$
4	10 (a)	32×32	$2.5 \cdot 10^6$	-0.57	$9.5 \cdot 10^{-4}$

Table 2.1: Each row specifies the quantities σ^{e}, ξ^{e}, the total number of simulated time steps, and the grid size for one network run simulating Hebbian development of intracortical connections. The values are listed together with the number of the figure that presents the corresponding results. All the other simulation parameters are held constant as given in Sect. A.1. Parameter sets $\#3$ and $\#4$ are equal except for the seed value of the random number generator.

In contrast, the synapses between excitatory cells and those from inhibitory to excitatory neurons are plastic. They develop according to the learning rules described in Sect. 2.3. We use an exponential learning window

$$W(t) = \exp(-t/\tau_W) \ , \ \tau_W = 11\mathrm{ms} \ ,$$

and Gaussian shaped arbor functions

$$\begin{aligned} A^{\mathrm{e}}_{ij} &= 0.025 \ \exp\left[-\mathrm{d}(i,j)^2 \ / \ \left(2 \cdot 3^2\right)\right] \ , \\ A^{\mathrm{i}}_{ij} &= 0.05 \ \exp\left[-\mathrm{d}(i,j)^2 \ / \ \left(2 \cdot 3^2\right)\right] \ , \end{aligned}$$

where $\mathrm{d}(i,j)$ denotes again the distance between the grid positions of the respective neurons. An upper limit $J^{\max} = 0.8$ is applied for the weight of excitatory synapses. Furthermore, we let

$$\vartheta^{\mathrm{e}} = 2.5 \times 10^{-6} \ , \ \ \vartheta^{\mathrm{i}} = 10^{-4} \ , \ \ \sigma^{\mathrm{i}} = 1,$$

as introduced in Eqs. (2.1) and (2.2). The values of the remaining parameters σ^{e}, ξ^{e} and the grid size are varied from run to run. They are summarized in Table 2.1.

2.8.2 Geniculocortical Plasticity

For our simulations of geniculocortical development we use the full network presented in Fig. 2.1. The grid size is 32×32. We fix the excitatory intracortical connectivity, while we let excitatory feedforward and inhibitory intracortical synapses develop according to Eqs. (2.1) and (2.2), respectively.

The postsynaptic potential $\varepsilon(t)$ and the refractory potential $\eta(t)$ are set as given in Eqs. (2.3) and (2.4). For the firing probability we use again the parameterization (2.5) but with different values of θ and T,

$$\begin{aligned} \theta = 13 \ , \ T = 0.25 \qquad & \text{for excitatory cells,} \\ \theta = 3 \ , \ T = 0.25 \qquad & \text{for inhibitory neurons.} \end{aligned}$$

At the very early developmental stage that we are concerned with, real cortical cells are still in a process of maturation, i.e., their response properties are very likely to change with time. Thus, there is no need to assume the response properties of our model neurons, i.e., θ and T to be constant.

The spatiotemporal activity correlations in the geniculate layer that are required to drive thalamocortical pattern formation, are generated as follows. Every 10 simulated time steps, the membrane potentials h_i of all the neurons i on the thalamic layer are drawn as a new realization of a Gaussian random field and are then kept fixed over the subsequent 10 iterations. Each geniculate cell produces a random spike train according to its membrane potential and the firing probability (2.5) with $\theta = 7$, $T = 1$.

In order to describe the Gaussian random field of membrane potentials, the expectation values $\langle h_i \rangle$ as well as the correlation matrix $\langle h_i\, h_j \rangle$ must be given. Throughout this chapter, we have set $\langle h_i \rangle = 0$ and

$$\begin{aligned} \langle h_i h_j \rangle &= 16.3\, \exp\left[-\mathrm{d}\,(i,j)^2 \,/\, \left(2 \cdot 1^2\right)\right] \\ &- 1.82\, \exp\left[-\mathrm{d}\,(i,j)^2 \,/\, \left(2 \cdot 3^2\right)\right] \ , \end{aligned}$$

which is a Mexican-hat-like function of the distance $\mathrm{d}\,(i,j)$ between the respective geniculate neurons i and j.

The parameters governing the dynamics of inhibitory weights are the same as specified in the preceding subsection. For the development of feedforward synapses, we use an arbor function

$$A^{\mathrm{e}}_{ij} = 0.0125\, \exp\left[-\mathrm{d}\,(i,j)^2 \,/\, \left(2 \cdot 3^2\right)\right] \ ,$$

and apply an upper weight limit $J^{\max} = 1$. Moreover we let

$$\vartheta^{\mathrm{e}} = 1.25 \times 10^{-6} \ .$$

and

$$W\,(t) = \exp\left(-t/\tau_W\right) \ , \ \tau_W = 11\mathrm{ms} \ .$$

Table 2.2 lists the values of those parameters that are varied for the different simulation runs. During each run, the excitatory lateral weights are kept constant. Their values are either taken from a previous simulation of intracortical development or determined by the Gaussian

$$J_{ij} = J_0^{\mathrm{ic}}\, \exp\left[-\mathrm{d}\,(i,j)^2 \,/\, \left(2 \cdot 3^2\right)\right] \ , \tag{2.6}$$

where $\mathrm{d}\,(i,j)$ is the distance between the two cortical cells i and j. In the former case, Table 2.2 references the corresponding intracortical simulation by its respective number in Table 2.1. In the latter case, the value of J_0^{ic} is given.

2.9 Summary and Discussion

In the preceding sections we have introduced a neuronal model of combined lateral *and* geniculocortical plasticity in layer 4 of the primary visual cortex.

Figure	time steps	σ^{e}	ξ^{e}	J_0^{ic}	# in T. 2.1
8 (b), (c)	$5 \cdot 10^6$	0.85	$8 \cdot 10^{-4}$	0.7	–
9 (b)	$5 \cdot 10^6$	0.4	$8 \cdot 10^{-4}$	–	3
10 (b)	$5 \cdot 10^6$	0.4	$8 \cdot 10^{-4}$	–	4
10 (c)	$5 \cdot 10^6$	0.4	$8 \cdot 10^{-4}$	–	4

Table 2.2: For each simulation of geniculocortical development, the network parameters σ^{e} and ξ^{e}, the total number of simulated time steps, and the corresponding figures are listed. In addition, the lateral excitatory connectivity is specified in the following way. If the pattern of lateral weights is isotropic, then the amplitude J_0^{ic}, corresponding to Eq. (2.6), is given. If the connectivity pattern is the result of a previous simulation of intracortical development, then we indicate that simulation by specifying the respective row of Table 2.1. The last two parameter sets are identical except for the seed value of the random number generator.

The network consists of stochastically spiking neurons to model geniculate relay cells as well as excitatory cortical cells interacting with inhibitory interneurons. The development of both lateral and feedforward connectivity is governed by activity-driven Hebbian learning dynamics. Plasticity of inhibitory synapses on short time scales has been incorporated so as to stabilize cortical activity and prevent epileptiform discharges (Varela et al. 1999, Douglas et al. 1995).

We have found that in this model spontaneous cortical activity in the absence of geniculate input can drive the plasticity of lateral connections to form an intracortical orientation map. The layout of such an intracortical map resembles that of typical orientation maps obtained from optical imaging experiments in primary visual cortex. It exhibits linear zones, where orientation changes smoothly across cortical surface, as well as so called pinwheel centers – point-like singularities at which orientation changes discontinuously by 90 degrees (Bonhoeffer and Grinvald 1991, Blasdel 1992b).

Furthermore, a Hebbian development of geniculocortical afferents in the presence of intracortical interaction and correlated activity in the LGN leads to the emergence of orientation selective receptive fields of cortical cells. As in previous correlation-based models of pattern formation (Linsker 1986b, Linsker 1986a, Miller 1994, Wimbauer et al. 1997b, Wimbauer et al. 1997a, Erwin and Miller 1998, Choe and Miikkulainen 1998), an isotropic pattern of intracortical connectivity is sufficient to obtain nicely ordered feedforward orientation maps very similar to measured ones.

We have then investigated the effect that a non-isotropic intracortical interaction in layer 4 can have on feedforward plasticity. It turned out that a previously formed intracortical orientation map can guide the development of geniculocortical afferents. As a consequence, the emerging feedforward map will be in accord with the existing intracortical map. Comparing the lateral arborization radius of layer 4 cells in the model (≤ 5.5 cells) with the typical distance of pinwheel centers in the final orientation maps (see e.g. Fig. 2.10) shows

that lateral interactions are normally confined to approximately one hypercolumn. The intracortical connectivity under consideration is thus short-ranged and must not be confused with the patchy patterns of horizontal projections that have been found anatomically to exist in layers 2/3 and 5 (Callaway and Katz 1990, Katz and Callaway 1992).

Since in our setup the same intracortical connectivity guides the development of afferents from both the left and the right eye, orientation maps will be matched in the two eyes. We have shown that this mechanism may provide a new explanation of the outcome of reverse lid-suturing experiments (Gödecke and Bonhoeffer 1996, Sengpiel et al. 1998). In contrast to earlier propositions (Wolf et al. 1996, Erwin and Miller 1998) it is neither dependent on the geometry of the respective cortical area nor on inter-eye correlations in thalamic activity.

Further support for our model comes from recent observations of Sengpiel et al. (1999). The authors raised kittens in an environment where they could see contours of only one orientation. The orientation maps that were obtained from these animals via optical imaging exhibited an only moderate shift towards this orientation. This is in full agreement with a scenario in which the development of geniculocortical afferents is guided by an existing pattern of intracortical connectivity and visual experience has an only minor influence.

Current anatomical data (Shatz and Luskin 1986, Gosh and Shatz 1992) show that in the cat geniculate afferents reach layer 4 of the primary visual cortex only about one week *after* the cells in this layer have finished their migration. This raises the possibility that these cells start an activity-driven development of lateral interactions at least one week before a Hebbian modification of feedforward afferents can begin. Although we have focused on the case of intracortical learning occurring strictly before geniculocortical refinement, the above considerations remain valid in a scenario in which both types of connections develop jointly for a certain period of time. The crucial requirement is that the plasticity of short-ranged lateral synapses in layer 4 ceases relatively early so that (i) the emerging intracortical map is independent of visual experience and (ii) the feedforward afferents can adapt to this map.

According to our knowledge, there is only one other spiking network model dedicated to the problem of combined development of feedforward and lateral projections (Choe and Miikkulainen 1998). In contrast to the aim of the present work, however, the authors did not analyze the possibility of an intracortical map formation and its implications for the refinement of geniculocortical afferents. Rather they concentrated on the formation of feedforward maps and the role that fast-adapting intracortical synapses can take in the segmentation of visual scenes afterwards.

Since our network is made up of spiking neurons and strictly distinguishes between excitatory cells and inhibitory interneurons, it is closer to biology than many previous models of the visual cortex. We have verified that in this relatively detailed approach, earlier results on the formation of orientation maps can be reproduced. This was not obvious a priori because stochastically spiking networks produce noise which may prevent ordered receptive fields to develop within biologically plausible time scales.

Quite to the contrary and most importantly, we have demonstrated that the new synaptic order, viz., the intracortical orientation map, is due to inherent cortical noise. An external source of noise would be necessary to obtain similar results in a network consisting of the usual graded-response neurons. The similarity of such intracortical maps with optically recorded maps of preferred orientation and the finding that they can guide the refinement of feedforward afferents emphasizes the role of horizontal connections in shaping neuronal response properties. Not only are lateral interactions likely to modulate a cortical cell's tuning curve "on the run", they can even predetermine its input connectivity during development.

Chapter 3

Correlations of Activity in Networks of Spiking Neurons

Throughout the preceding chapters we have introduced a neuronal model that allowed us to study the plasticity of both lateral and geniculocortical connections in layer 4 of the primary visual cortex. The network consists of stochastically spiking neurons so as to model geniculate relay cells as well as cortical pyramidal cells and inhibitory interneurons. The Hebbian development of lateral and feedforward connections has been analyzed by means of numerical simulations.

In the sequel, we want to gain some analytical insight into the mechanisms governing intracortical and feedforward development. To this end, we will have to investigate the dynamics of synaptic weights as induced by the learning rules described in Sect. 2.3. According to these rules, the modification of a synapse is determined by the spike trains of the two neurons it is connecting. Since we are dealing with a network of stochastically spiking neurons, some knowledge concerning the network's spike statistics and especially the spike-spike correlations is a prerequisite for our analysis. We will therefore use the present chapter to derive analytical expressions for the two-spike correlation functions in a recurrent network of stochastically spiking neurons. The reader who is willing to take the results for granted and is mainly interested in an analysis of the learning dynamics may safely skip this very technical part of our work and proceed directly to the next chapter.

The subsequent discussion is organized as follows. Section 3.1 first recapitulates the spike response model, which is the neuron model our calculations will be based on. In Sect. 3.2 we then introduce the concept of a network *ensemble* as an entirety of independent random realizations of one given neural network. Within this framework the spike statistics we are interested in can be formulated in terms of ensemble averages of the network activity. Section 3.3 presents two methods for obtaining approximations of the input-output statistics of a rather simple network, namely a single neuron without feedback connections. The second of these approaches is then generalized in Sect. 3.4 so as to find analytical estimates of the average firing rate and the two-spike correlation functions in recurrently interconnected networks of arbitrary architec-

ture. In Sect. 3.5 we compare our theoretical predictions to the corresponding quantities as obtained from numerical simulations. At the end we give a short summary of this chapter's main results.

3.1 Neuron Model

The model of spiking neurons that we have implemented in our computational model of primary visual cortex is the spike response model (Gerstner and van Hemmen 1994), which we have already introduced in Sect. 2.1. To render this chapter self-contained, we give a short summary in the following.

The state of a spike response-neuron is described by a time-dependent internal variable $h(t)$ which can be considered as the membrane potential. The effects of spike reception and spike generation are modeled as transient changes of h, described by two response functions $\varepsilon(t)$ and $\eta(t)$. Spike generation is governed by a generalized Poisson process that depends only on the cell's membrane potential. That means, the probability that neuron number i emits a spike during the infinitesimal time interval $[t, t + \mathrm{d}t]$, given the value of its membrane potential $h_i(t)$, can be written

$$\Pr\{\text{Neuron } i \text{ fires during } [t, t+\mathrm{d}t) \mid h_i(t) = h\} := q_i(h)\,\mathrm{d}t\,, \tag{3.1}$$

and the probability of more than one spike being emitted during that period is $o(\mathrm{d}t)$. We will refer to $q_i(h)$ as the neuron's activation function.

Whenever an action potential has been generated at neuron j, it is transmitted along the axon to other neurons. At the same time, a negative contribution $-\eta(t)$ is added to the membrane voltage h_j to account for the reduced excitability of cell j during its refractory period. At a postsynaptic neuron i, every action potential that arrives from a presynaptic neuron j induces a postsynaptic potential that is given by the response kernel $\varepsilon(t)$ multiplied by the synaptic weight J_{ij}. Thus, the total membrane potential h_i of the postsynaptic neuron is

$$h_i(t) = \sum_j \sum_{t_j^{\mathrm{f}} \le t} J_{ij}\,\varepsilon(t - t_j^{\mathrm{f}}) - \sum_{t_i^{\mathrm{f}} \le t} \eta(t - t_i^{\mathrm{f}})\,, \tag{3.2}$$

where the t_i^{f} are the firing times of neuron i. Obviously, the membrane potential h_i at time t is determined by the sequence of spikes emitted at cell i and its presynaptic neighbours j up to time t.

For each neuron i, let us now use a sum of Dirac-δ pulses to represent its spike train $a_i(t)$,

$$a_i(t) := \sum_{t_i^{\mathrm{f}}} \delta(t - t_i^{\mathrm{f}})\,. \tag{3.3}$$

We can then rewrite Eq. (3.2) in the form

$$\begin{aligned} h_i(t) &= \sum_j J_{ij} \int_{-\infty}^{t} a_j(t')\,\varepsilon(t - t')\,\mathrm{d}t' + \int_{-\infty}^{t} a_i(t')\,\eta(t - t')\,\mathrm{d}t' \\ &=: [\hat{\kappa}\,a]_i(t)\,, \end{aligned} \tag{3.4}$$

thereby defining the linear coupling operator $\hat{\kappa}$. In the sequel, we will always assume, that $\varepsilon(t \leq 0) = 0$ and $\eta(t \leq 0) = 0$, so that the upper integration limits in (3.4) can formally be extended to $+\infty$.

3.2 Network Ensembles

In a network of stochastic spike response-neurons, the cells' membrane potentials as well as their spike trains are stochastic processes in time. That means that different 'runs' of the same network, started from the same initial conditions, will produce different spike trains and different tracks of the membrane potentials.

Here we are interested in the statistics of spikes in such a system, as e.g., the probability (density) of a spike being emitted at neuron i and time t, given that a spike has occurred at cell j and time $t' \leq t$. Deriving the probability of such an event means to consider infinitely many runs of the same network, and to determine the fraction of those runs in which this event occurs. Instead of an entirety of infinitely many runs of the same network, we can equivalently consider an ensemble of infinitely many instances of the same network running all at the same time. Henceforth we will term this an 'ensemble of networks' or, in short, an 'ensemble'.

3.2.1 Notation

To simplify notation, we introduce the following abbreviations for specifying events of spike emission.

$S_i(t)$: A spike occurs in spike train number i at time t,

$S_i(t, \Delta t)$: At least one spike occurs in spike train number i within time interval $[t, t + \Delta t)$,

$O_i(t, t')$: No spike is emitted in spike train number i within time interval $[t, t')$,

$S_i^n(t, t', \{t_1, \ldots, t_n\})$: Exactly n spikes are emitted in spike train number i within the time interval $[t, t')$, viz., at times $t_1, \ldots, t_n$.

The probability of a given event will be denoted using the symbol $\Pr(\cdot)$, whereas probability densities will be written as $\mathrm{pr}(\cdot)$,e.g.,

$$\mathrm{pr}\left[S_i(t)\right] := \lim_{\Delta t \to 0} \frac{\Pr\left[S_i(t, \Delta t)\right]}{\Delta t} \,.$$

Throughout this work, probability densities are to be understood in the sense of distributions, i.e., they may contain Dirac-delta contributions. We will assume that the probability of more than one spike occurring within an infinitesimal time interval $\mathrm{d}t$ is $o(\mathrm{d}t)$. The probability of exactly one spike occurring within that interval is then equal to $\Pr\left[S_i(t, \mathrm{d}t)\right]$. This can be written as an integral

$$\Pr\left[S_i(t, \mathrm{d}t)\right] = \int_t^{t+\mathrm{d}t} \mathrm{pr}\left[S_i(t')\right] \mathrm{d}t' \,,$$

which we will often write in a sloppy but shorter notation as

$$\Pr\left[S_i(t,\mathrm{d}t)\right] = \mathrm{pr}\left[S_i(t)\right]\mathrm{d}t\ .$$

In the subsequent sections, we will very often perform averages over ensembles of networks. We let $\langle x\rangle$ denote the ensemble average – or expectation value – of the quantity x. Since the state of a network of spike response-neurons at a given time τ is fully determined by the sequence of spikes that have occurred in the network up to time τ, the quantity x will in general be a function of all the firing times $t_i^{\mathrm{f}} \leq \tau$ of all neurons i. If we were given the probabilities of all the spike trains that the network can produce, then we could directly calculate the ensemble average $\langle x\rangle$.

As an example consider a quantity $x := x_i(\{t_1,\ldots,t_n\})$ that depends on the spike times of neuron i only. According to the above declarations, the expression $\mathrm{pr}[S_i^n(0,\tau,\{t_1,\ldots,t_n\})]$ denotes the probability density that up to time τ this cell emits exactly n spikes, viz., at the times $t_1\ldots t_n$. From that we have

$$\begin{aligned}\langle x_i\rangle &= \sum_{n=0}^{\infty}\int_0^\tau \mathrm{d}t_1\int_{t_1}^\tau \mathrm{d}t_2\cdots\int_{t_{n-1}}^\tau \mathrm{d}t_n\, x_i\left(\{t_1,\ldots,t_n\}\right)\,\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right]\\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\int_0^\tau \mathrm{d}t_1\cdots\int_0^\tau \mathrm{d}t_n\, x_i\left(\{t_1,\ldots,t_n\}\right)\,\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right]\ . \qquad (3.5)\end{aligned}$$

We are now going to analyze two special cases of (3.5) because they will be of particular importance in this chapter.

3.2.2 The Mean Activity

First, we choose x_i to be the spike train of cell i as defined in (3.3). As we will show, the ensemble average of this quantity evaluated at a given time t is equal to the probability density of spike emission at that time, which is often referred to as the neuron's mean activity or firing rate. From (3.5) we have with $\tau > t$,

$$\begin{aligned}\langle a_i(t)\rangle &= \sum_{n=0}^{\infty}\frac{1}{n!}\int_0^\tau \mathrm{d}t_1\cdots\int_0^\tau \mathrm{d}t_n\, a_i(t)\,\mathrm{pr}\left[S_i(0,\tau,\{t_1,\ldots,t_n\})\right]\\ &= \sum_{n=0}^{\infty}\frac{1}{n!}\int_0^\tau \mathrm{d}t_1\cdots\int_0^\tau \mathrm{d}t_n\,\left[\sum_{k=1}^{n}\delta(t-t_k)\right]\,\mathrm{pr}\left[S_i(0,\tau,\{t_1,\ldots,t_n\})\right]\ .\end{aligned}$$

For $n=0$ the sum over k on the right hand side vanishes, because $n=0$ means that the neuron does not spike at all up to time τ, so its spike train is $a_i(t)\equiv 0$. Therefore we find

$$\begin{aligned}\langle a_i(t)\rangle &= \sum_{n=1}^{\infty}\frac{1}{n!}\int_0^\tau \mathrm{d}t_1\cdots\int_0^\tau \mathrm{d}t_n\, a_i(t)\,\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] \qquad (3.6)\\ &= \sum_{n=1}^{\infty}\frac{1}{n!}\sum_{k=1}^{n}\int_0^\tau \mathrm{d}t_1\cdots\int_0^\tau \mathrm{d}t_n\,\delta(t-t_k)\,\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right]\ .\end{aligned}$$

Since

$$\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_k,\ldots,t_n\})\right] = \mathrm{pr}\left[S_i^n(0,\tau,\{t_k,t_1,\ldots,t_{k-1},t_{k+1},\ldots,t_n\})\right] \ ,$$

the variables of integration in (3.6) can be rearranged in such a way that

$$\begin{aligned}
\langle a_i(t)\rangle &= \sum_{n=1}^{\infty}\frac{1}{n!}\sum_{k=1}^{n}\int_0^\tau \mathrm{d}t_1\cdots\int_0^\tau \mathrm{d}t_n\,\delta(t-t_1)\,\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] \\
&= \sum_{n=1}^{\infty}\frac{n}{n!}\int_0^\tau \mathrm{d}t_1\cdots\int_0^\tau \mathrm{d}t_n\,\delta(t-t_1)\,\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] \\
&= \mathrm{pr}\left[S_i^1(0,\tau,\{t\})\right] + \\
&+ \sum_{n=2}^{\infty}\frac{1}{(n-1)!}\int_0^\tau \mathrm{d}t_2\cdots\int_0^\tau \mathrm{d}t_n\,\mathrm{pr}\left[S_i^n(0,\tau,\{t,t_2,\ldots,t_n\})\right] \\
&= \mathrm{pr}\left[S_i^1(0,\tau,\{t\})\right] + \\
&+ \sum_{n=1}^{\infty}\frac{1}{n!}\int_0^\tau \mathrm{d}t_1\cdots\int_0^\tau \mathrm{d}t_n\,\mathrm{pr}\left[S_i^{n+1}(0,\tau,\{t,t_1,\ldots,t_n\})\right] \\
&= \sum_{n=0}^{\infty}\frac{1}{n!}\int_0^\tau \mathrm{d}t_1\cdots\int_0^\tau \mathrm{d}t_n\,\mathrm{pr}\left[S_i^{n+1}(0,\tau,\{t,t_1,\ldots,t_n\})\right] \ .
\end{aligned}$$

Inspecting the last and the last but one equality we find that the right hand side represents a sum over the probabilities (densities) of all possible spike trains in which there is a spike at time t. This yields the probability density for a spike occurring at time t, no matter whether and when there are any additional spikes. We thus find

$$\langle a_i(t)\rangle = \mathrm{pr}\left[S_i(t)\right] \ . \tag{3.7}$$

At first sight, it may be surprising that an ensemble average over sequences of infinitely thin δ-spikes yields the probability density $\mathrm{pr}\left[S_i(t)\right]$, which may be smooth and finite. A demonstrative way of thinking about the above averaging procedure is to consider every spike as being smeared out by a Gaussian $g(t,\sigma)$ of finite width σ with $\int_{-\infty}^{+\infty} g(t,\sigma)\mathrm{d}t = 1$. The cell's spike train then reads

$$a_i(t,\sigma) := \sum_{t^\mathrm{f}} g(t-t^\mathrm{f},\sigma) \ .$$

Now perform the ensemble average $\langle a_0(t,\sigma)\rangle$, which yields a smooth function of t, and take the limit $\sigma \to 0$ afterwards. In terms of mathematics this approach is nothing but a replacement of the δ-distribution by its representation as a sequence of functions, but it provides a more intuitive understanding of the meaning of averages over spike trains.

3.2.3 Joint Probability Density of Two Spike Events

As a second application of Eq. (3.5) we will derive that $\langle a_i(t)\,a_i(t')\rangle$ is equal to the joint probability density $\mathrm{pr}\left[S_i(t)\wedge S_i(t')\right]$ of a spike occurring at both

times the intervals t *and* t. To be prepared, let us first see how to write this probability in terms of $\mathrm{pr}[S_i(0,\tau,\{t_1,\ldots,t_n\})]$.

If $t \neq t'$, we obtain $\mathrm{pr}[S_i(t) \wedge S_i(t')]$ by summing the probability densities of all spike trains with spikes occurring in the respective time intervals, i.e.,

$$\begin{aligned} &\mathrm{pr}\left[S_i(t) \wedge S_i(t')\right] \qquad (3.8)\\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \,\mathrm{pr}\left[S_i^{n+2}(0,\tau,\{t,t',t_1,\ldots,t_n\})\right] \text{ , for } t \neq t', \end{aligned}$$

with $\tau > t, t'$. If on the other hand $t = t'$, then there is a spike in $[t', t' + \mathrm{d}t')$ whenever there is a spike in $[t, t+\mathrm{d}t)$. For this case we thus find the probability $\mathrm{pr}[S_i(t) \wedge S_i(t')]\,\mathrm{d}t\,\mathrm{d}t'$ of a occurring in both intervals $[t', t'+\mathrm{d}t')$ and $[t', t'+\mathrm{d}t')$ to be equal to the probability of a spike being emitted in $[t, t + \mathrm{d}t)$, which is $\mathrm{pr}[S_i(t)]\,\mathrm{d}t$. Together with (3.8) this implies

$$\begin{aligned} &\mathrm{pr}\left[S_i(t) \wedge S_i(t')\right] = \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \,\mathrm{pr}\left[S_i^{n+2}(0,\tau,\{t,t',t_1,\ldots,t_n\})\right] \\ &+ \mathrm{pr}\left[S_i(t)\right]\,\delta(t-t') \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \,\mathrm{pr}\left[S_i^{n+2}(0,\tau,\{t,t',t_1,\ldots,t_n\})\right] \\ &+ \delta(t-t') \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \,\mathrm{pr}\left[S_i^{n+1}(0,\tau,\{t,t_1,\ldots,t_n\})\right] . \qquad (3.9) \end{aligned}$$

Now we start evaluating $\langle a_i(t)\, a_i(t')\rangle$. With $\tau > t, t'$ we have

$$\begin{aligned} \left\langle a_i(t)\, a_i(t')\right\rangle &= \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \, a_i(t)\, a_i(t') \,\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] \\ &= \sum_{n=1}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \,\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] \sum_{k,j=1}^{n} \delta(t-t_k)\,\delta(t'-t_j) \\ &= \sum_{n=2}^{\infty} \frac{1}{n!} \sum_{\substack{k,j=1\\k\neq j}}^{n} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \,\delta(t-t_k)\,\delta(t'-t_j) \,\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] \\ &+ \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{\substack{k=1\\(k=j)}}^{n} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \,\delta(t-t_k)\,\delta(t'-t_k) \,\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] . \end{aligned}$$

As in (3.6), we rearrange the variables of integration and obtain

$$\begin{aligned} &\left\langle a_i(t)\, a_i(t')\right\rangle = \\ &= \sum_{n=2}^{\infty} \frac{n(n-1)}{n!} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \,\delta(t-t_1)\,\delta(t'-t_2) \,\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] \\ &+ \sum_{n=1}^{\infty} \frac{n}{n!} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \,\delta(t-t_1)\,\delta(t'-t_1) \,\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] \end{aligned}$$

$$= \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \, \mathrm{pr}\left[S_i^{n+2}(0,\tau,\{t,t',t_1,\ldots,t_n\})\right]$$
$$+ \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \, \delta(t'-t) \, \mathrm{pr}\left[S_i^{n+1}(0,\tau,\{t,t_1,\ldots,t_n\})\right] \ .$$

Comparing the last equality with Eq. (3.9), we find

$$\left\langle a_i(t)\, a_i(t')\right\rangle = \mathrm{pr}\left[S_i(t) \wedge S_i(t')\right] \ . \tag{3.10}$$

The above calculations can easily be extended to show that quite generally

$$\left\langle a_i(t)\, a_j(t')\right\rangle = \mathrm{pr}\left[S_i(t) \wedge S_j(t')\right] \ , \tag{3.11}$$

with arbitrary $i, j = 1, \ldots N$.

3.2.4 Conditional Probability Density

At the end of this section, let us work out an expression for the conditional probability density of exactly n spikes occurring at times $t_1, \ldots, t_n$ at input i, given that there is a spike in some interval $[t, t+\Delta t)$ at the same input. From the definition of a conditional probability, we have

$$\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\}) \,|\, S_i(t,\Delta t)\right] = \frac{\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\}) \ \wedge \ S_i(t,\Delta t)\right]}{\mathrm{Pr}\left[S_i(t,\Delta t)\right]} \ . \tag{3.12}$$

For a further evaluation of the right hand side, we check whether at least one of the $t_1, \ldots, t_n$ is within the interval $[t, t+\Delta t)$. If this is the case, then the event $S_i(0,\tau,\{t_1,\ldots,t_n\})$ implies the event $S_i(t,\Delta t)$ and the joint probability density of both events is equal to that of $S_i(0,\tau,\{t_1,\ldots,t_n\})$ alone. Otherwise the events $S_i(0,\tau,\{t_1,\ldots,t_n\})$ and $S_i(t,\Delta t)$ are disjunct and their joint probability density is zero. Therefore it is

$$\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\}) \wedge S_i(t,\Delta t)\right] = \begin{cases} 0 \ , \text{ if } t_1,\ldots,t_n \notin [t,t+\Delta t) \ , \\ \mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] \text{ otherwise.} \end{cases}$$

Because of our assumption that the probability of more than two spikes occurring within in an interval $[t, t+\Delta t)$ is $o(\Delta t)$, we can take for granted that there is at most one spike in that interval, if $\Delta t \to 0$. With $\Delta t \to 0$ we thus write

$$\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\}) \wedge S_i(t,\Delta t)\right]$$
$$= \int_t^{t+\Delta t} \mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] \sum_{k=1}^{n} \delta\left(s - t_k\right) \mathrm{d}s \ ,$$

which implies

$$\lim_{\Delta t \to 0} \frac{\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\}) \wedge S_i(t,\Delta t)\right]}{\Delta t}$$
$$= \mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] \sum_{k=1}^{n} \delta\left(t - t_k\right) \ .$$

Inserting this into Eq. (3.12) yields

$$\lim_{\Delta t \to 0} \; \mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\}) \mid S_i(t,\Delta t)\right] = \frac{\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] \sum_{k=1}^{n} \delta\left(t-t_k\right)}{\mathrm{pr}\left[S_i(t)\right]} \; .$$

Carrying out the limit $\Delta t \to 0$ on the left hand side leaves us with

$$\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\}) \mid S_i(t)\right] = \frac{\mathrm{pr}\left[S_i^n(0,\tau,\{t_1,\ldots,t_n\})\right] \sum_{k=1}^{n} \delta\left(t-t_k\right)}{\mathrm{pr}\left[S_i(t)\right]} \; . \tag{3.13}$$

If $n=0$ this is to be read $\mathrm{pr}\left[S_i^0(0,\tau,\{\}) \mid S_i(t)\right] = 0$.

3.3 Feedforward Networks

In this section we will derive approximations for the two-spike correlation between input and output of a single neuron. It is assumed that the input spike trains do not depend causally on the output spike train, i.e., there is no feedback from output to input.

3.3.1 General Considerations

Let us consider a single neuron receiving N input spike trains from a given stochastic process and generating an output spike train according to the spike response model described in section 3.1. We assign the index $i=0$ to this neuron, while the input spike trains are labeled by $i = 1 \ldots N$. The input-output correlation function is defined as

$$C_{0i}(t_1,t_2) := \frac{\langle a_0(t_1)\, a_i(t_2)\rangle}{\langle a_0(t_1)\rangle \, \langle a_i(t_2)\rangle} \; , \tag{3.14}$$

where $\langle a_i(t)\rangle$ is given from the statistics of the i^{th} input process. We thus need to find suitable expressions for $\langle a_0(t)\rangle$ and $\langle a_0(t_1)\, a_i(t_2)\rangle$. Subsequently, some general relations are derived, which will be useful to approximate these quantities.

From Eq. (3.7) we know that $\langle a_0(t)\rangle$ is equal to the firing probability density $\mathrm{pr}[S_0(t)]$. It is easy to see that this is in turn equal to the expectation value $\langle q_0[h_0(t)]\rangle$, where $q_0(h)$ is the neural activation function. In order to show that, we expand

$$\mathrm{pr}\left[S_0(t)\right]\mathrm{d}t = \Pr\left[S_0(t,\mathrm{d}t)\right] = \int_{-\infty}^{+\infty} \mathrm{d}h \, \mathrm{pr}\left[S_0(t,\mathrm{d}t) \wedge h_0(t) = h\right] \; , \tag{3.15}$$

where $\mathrm{pr}[S_0(t,\mathrm{d}t) \wedge h_0(t)=h]$ is the joint probability density that the cell emits a spike during $[t, t+\mathrm{d}t)$ *and* the value of its membrane potential h_0 at time t is h. The first equality is nothing but the definition of the probability density $\mathrm{pr}[S_0(t)]$. The second equality holds true because the membrane potential $h_0(t)$ must take some value between $-\infty$ and $+\infty$. The integral thus sums the

probabilities of a complete set of disjoint events in which a spike occurs during $[t, t + \mathrm{d}t)$. This yields the probability of the event $S_0(t, \mathrm{d}t)$.

The joint probability density in the integral on the right hand side of (3.15) can be written as the product of a conditional probability and an unconditional probability density, resulting in

$$\mathrm{pr}\left[S_0(t)\right] \mathrm{d}t = \int_{-\infty}^{+\infty} \mathrm{d}h \, \Pr\left[S_0(t, \mathrm{d}t) \mid h_0(t) = h\right] \, \mathrm{pr}\left[h_0(t) = h\right] \ .$$

In (3.1) we have defined for our neuron model $\Pr\left[S_0(t, \mathrm{d}t) \mid h_0(t) = h\right]$ to be equal to $q_0(h)\,\mathrm{d}t$, so that we obtain

$$\mathrm{pr}\left[S_0(t)\right] \mathrm{d}t = \int_{-\infty}^{+\infty} q_0(h) \, \mathrm{pr}\left[h_0(t) = h\right] \, \mathrm{d}h \, \mathrm{d}t = \langle q_0[h_0(t)] \rangle \, \mathrm{d}t \, .$$

This proves the statement

$$\langle a_0(t) \rangle = \langle q_0[h_0(t)] \rangle \ . \tag{3.16}$$

In an analogous manner, we can treat the second moment $\langle a_0(t_1)\, a_i(t_2) \rangle$, where $i = 1 \ldots N$. From Eq. (3.11) we know that $\langle a_0(t_1)\, a_i(t_2) \rangle$ is equal to $\mathrm{pr}[S_0(t_1) \wedge S_i(t_2)]$. We start by expanding

$$\begin{aligned} \mathrm{pr}[S_0(t_1) \wedge S_i(t_2)] \mathrm{d}t_1 \, \mathrm{d}t_2 &= \Pr\left[S_0(t_1, \mathrm{d}t_1) \wedge S_i(t_2, \mathrm{d}t_2)\right] \\ &= \int \mathrm{d}h \, \mathrm{pr}\left[S_0(t_1, \mathrm{d}t_1) \wedge S_i(t_2, \mathrm{d}t_2) \wedge h_0(t_1) = h\right] \\ &= \int \mathrm{d}h \, \Pr\left[S_0(t_1, \mathrm{d}t_1) \mid h_0(t_1) = h \wedge S_i(t_2, \mathrm{d}t_2)\right] \, \mathrm{pr}\left[S_i(t_2, \mathrm{d}t_2) \wedge h_0(t_1) = h\right] \ . \end{aligned} \tag{3.17}$$

Since according to the spike response model, the neuron's firing probability density does only depend on the value of its membrane potential, it is

$$\begin{aligned} \Pr\left[S_0(t_1, \mathrm{d}t_1) \mid h_0(t) = h \wedge S_i(t_2, \mathrm{d}t_2)\right] &= \Pr\left[S_0(t_1, \mathrm{d}t_1) \mid h_0(t_1) = h\right] \\ &= q_0(h) \, \mathrm{d}t_1 \ . \end{aligned}$$

Inserting into Eq. (3.17) yields

$$\begin{aligned} \mathrm{pr}[S_0(t_1) \wedge S_i(t_2)] \mathrm{d}t_1 \, \mathrm{d}t_2 &= \int q_0(h) \, \mathrm{pr}\left[h_0(t_1) = h \wedge S_i(t_2)\right] \, \mathrm{d}h \, \mathrm{d}t_1 \, \mathrm{d}t_2 \\ &= \int q_0(h) \, \mathrm{pr}\left[h_0(t_1) = h \mid S_i(t_2)\right] \, \mathrm{pr}\left[S_i(t_2)\right] \, \mathrm{d}h \, \mathrm{d}t_1 \, \mathrm{d}t_2 \\ &= \langle a_i(t_2) \rangle \, \mathrm{d}t_2 \int q_0(h) \, \mathrm{pr}\left[h_0(t_1) = h \mid S_i(t_2)\right] \, \mathrm{d}h \, \mathrm{d}t_1 \ , \end{aligned}$$

and that means

$$\langle a_0(t_1)\, a_i(t_2) \rangle = \langle a_i(t_2) \rangle \int q_0(h) \, \mathrm{pr}\left[h_0(t_1) = h \mid S_i(t_2)\right] \, \mathrm{d}h \ . \tag{3.18}$$

The integral on the right hand side of this expression is the conditional expectation of $q_0[h_0(t_1)]$ given that there is a spike at input i and time t_2. An

alternative representation of (3.18) can be found by further expanding the conditional probability according to

$$\begin{aligned}
\langle a_i(t_2)\rangle \ \mathrm{pr}\Big[h_0(t_1) = h \mid S_i(t_2)\Big] \ &= \ \mathrm{pr}\Big[h_0(t_1) = h \wedge S_i(t_2)\Big] \\
&= \ \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t'_1 \cdots \int_0^\tau \mathrm{d}t'_n \, \mathrm{pr}\Big[h_0(t_1) = h \ \wedge S_i(t_2) \wedge S_i^n(0,\tau,\{t'_1,\ldots,t'_n\})\Big] \\
&= \ \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t'_1 \cdots \int_0^\tau \mathrm{d}t'_n \, \mathrm{pr}\Big[h_0(t_1) = h \mid S_i^n(0,\tau,\{t'_1,\ldots,t'_n\}) \wedge S_i(t_2)\Big] \\
&\times \ \mathrm{pr}\Big[S_i^n(0,\tau,\{t'_1,\ldots,t'_n\}) \wedge S_i(t_2)\Big] \ .
\end{aligned}$$

We notice that the joint probability density in the last equality is equal to $\mathrm{pr}[S_i(t_2)]$ times the conditional probability density of $S_i^n(\ldots)$, given $S_i(t_2)$. The joint probability density can therefore be replaced using Eq. (3.13), multiplied by $\mathrm{pr}[S_i(t_2)]$. This results in

$$\begin{aligned}
\langle a_i(t_2)\rangle \ \mathrm{pr}\Big[h_0(t_1) = h \mid S_i(t_2)\Big] \ &= \ \mathrm{pr}\Big[h_0(t_1) = h \wedge S_i(t_2)\Big] \ = \\
&= \ \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t'_1 \cdots \int_0^\tau \mathrm{d}t'_n \, \mathrm{pr}\Big[h_0(t_1) = h \mid S_i^n(0,\tau,\{t'_1,\ldots,t'_n\}) \wedge S_i(t_2)\Big] \\
&\times \ \mathrm{pr}\Big[S_i^n(0,\tau,\{t'_1,\ldots,t'_n\})\Big] \sum_{k=1}^{n} \delta\left(t_2 - t'_k\right) \ .
\end{aligned}$$

Because of the delta functions, the integrand vanishes but when any $t'_1,\ldots,t'_n$ is equal to t_2. That means that the integral is contributed only by events $S_i^n(\ldots)$ which imply the event $S_i(t_2)$. Thus, we can drop $S_i(t_2)$ from the conditional probability density and merge the product of the two densities so as to obtain

$$\begin{aligned}
&\langle a_i(t_2)\rangle \, \mathrm{pr}\Big[h_0(t_1) = h \mid S_i(t_2)\Big] = \qquad (3.19)\\
&= \ \mathrm{pr}\Big[h_0(t_1) = h \ \wedge \ S_i(t_2)\Big] \ = \ \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t'_1 \cdots \int_0^\tau \mathrm{d}t'_n \\
&\times \ \mathrm{pr}\Big[h_0(t_1) = h \ \wedge S_i^n(0,\tau,\{t'_1,\ldots,t'_n\})\Big] \sum_{k=1}^{n} \delta\left(t_2 - t'_k\right) \\
&= \ \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t'_1 \cdots \int_0^\tau \mathrm{d}t'_n \, a_i(t_2) \, \mathrm{pr}\Big[h_0(t_1) = h \ \wedge S_i^n(0,\tau,\{t'_1,\ldots,t'_n\})\Big] \ .
\end{aligned}$$

The last equality holds true, because the sum over delta functions can be identified with the i^{th} input spike train evaluated at time t_2. Inserting this result into the right hand side of Eq. (3.18) we finally get

$$\begin{aligned}
\langle a_0(t_1)\, a_i(t_2)\rangle \ &= \ \langle a_i(t_2)\rangle \int q_0(h) \, \mathrm{pr}\left[h_0(t_1) = h \mid S_i(t_2)\right] \, \mathrm{d}h \\
&= \ \langle a_i(t_2)\, q_0\left[h_0(t_1)\right]\rangle \ . \qquad (3.20)
\end{aligned}$$

Usually, it will be difficult to find an exact expression for $\langle a_0(t_1)\, a_i(t_2)\rangle$, depending on the stochastic processes generating the input spike trains. In the remainder of this section we will introduce two methods that can be used to obtain approximative results.

3.3.2 Non-Refractory Neuron with a Large Number of Statistically Independent Poisson Inputs

Let us assume that the firing rate of the output neuron is low, such that the expected time interval between two individual output spikes is large as compared to the decay time of the neuron's refractory potential. In this case we may neglect the effect of the refractory potential and thus obtain from Eq. (3.2)

$$h_0(t) = \sum_i \sum_{t_i^{\mathrm{f}} \le t} J_{0i}\, \varepsilon(t - t_i^{\mathrm{f}})\ .$$

If the input spike trains are statistically independent and the number of input synapses is sufficiently large, then we know by the central limit theorem that, under quite general conditions concerning the distribution of weights J_{0i} and the postsynaptic potential $\varepsilon(t)$, the distribution of $h_0(t)$ is approximately Gaussian. Here we will presume, that the input spikes are generated from statistically independent generalized Poisson processes. This allows for a direct calculation of mean and variance of $h_0(t)$, and the resulting expressions are simple. A detailed proposition of the inhomogeneous – or generalized – Poisson process in the context of neural spike trains has been given by Kempter et al. (1998).

For a given input synapse i, we consider the probability density that up to time τ exactly n spikes arrive at times $t_1, \ldots, t_n$. In terms of the previously introduced spike events S_i, this writes

$$\begin{aligned}\mathrm{pr}[S_i^n(0,\tau,\{t_1,\ldots,t_n\})]\,\mathrm{d}t_1\cdots\mathrm{d}t_n = \Pr[&O_i(0,t_1)\wedge S_i(t_1,\mathrm{d}t_1)\\ &\wedge O_i(t_1+\mathrm{d}t_1,t_2)\wedge S_i(t_2,\mathrm{d}t_2)\wedge\ldots\wedge S_i(t_n,\mathrm{d}t_n)\wedge O_i(t_n+\mathrm{d}t_n,\tau)]\ .\end{aligned}$$

By definition of the Poisson process, disjunct intervals of the input spike train are statistically independent. The right hand side of the last equation can thus be factorized, so that

$$\begin{aligned}&\Pr[O_i(0,t_1)\wedge S_i(t_1,\mathrm{d}t_1)\wedge\ldots\wedge S_i(t_n,\mathrm{d}t_n)\wedge O_i(t_n+\mathrm{d}t_n,\tau)]\\ &\quad= \Pr[O_i(0,t_1)\wedge O_i(t_1+\mathrm{d}t_1,t_2)\wedge\cdots\wedge O_i(t_n+\mathrm{d}t_n,\tau)]\\ &\quad\times\ \Pr[S_i(t_1,\mathrm{d}t_1)]\cdots\Pr[S_i(t_n,\mathrm{d}t_n)]\ ,\end{aligned}$$

and for $\mathrm{d}t_1,\ldots,\mathrm{d}t_n \to 0$ we obtain

$$\mathrm{pr}[S_i^n(0,\tau,\{t_1,\ldots,t_n\})] = \Pr[O_i(0,\tau)]\,\mathrm{pr}[S_i(t_1)]\cdots\mathrm{pr}[S_i(t_n)]\ . \tag{3.21}$$

After this preparation we are in the situation to derive an expression for the conditional probability density $\mathrm{pr}\,[h_0(t) = h \mid S_i(t')]$ required in (3.18). For simplicity, let us consider for a moment only one single input spike train (index

$i = 1$) arriving at the output neuron. Let the corresponding synaptic weight be J_{01}. Then, with $\tau > t, t'$, it is

$$\begin{aligned}\mathrm{pr}\left[h_0(t) = h \mid S_1(t')\right] = \sum_{n=0}^{\infty} \int_0^\tau \mathrm{d}t_1 \int_{t_1}^\tau \mathrm{d}t_2 \cdots \int_{t_{n-1}}^\tau \mathrm{d}t_n \\ \times \quad \mathrm{pr}\left[S_1(0,\tau,\{t_1,\ldots,t_n\}) \mid S_1(t')\right] \delta\Big[h - J_{01} \sum_{m=1}^{n} \varepsilon(t-t_m)\Big] \; .\end{aligned}$$

Using (3.13) to replace the conditional probability, we have

$$\begin{aligned}&\mathrm{pr}\left[h_0(t) = h \mid S_1(t')\right] = \\ &= \sum_{n=1}^{\infty} \frac{1}{n!} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \, \delta\Big[h - J_{01} \sum_{m=1}^{n} \varepsilon(t-t_m)\Big] \\ &\times \quad \frac{\mathrm{pr}\left[S_1^n(0,\tau,\{t_1,\ldots,t_n\})\right]}{\mathrm{pr}\left[S_1(t')\right]} \sum_{k=1}^{n} \delta\left(t' - t_k\right) \\ &= \sum_{n=1}^{\infty} \frac{1}{n!} \sum_{k=1}^{n} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \, \delta\Big[h - J_{01} \sum_{m=1}^{n} \varepsilon(t-t_m)\Big] \\ &\times \quad \frac{\mathrm{pr}\left[S_1^n(0,\tau,\{t_1,\ldots,t_n\})\right]}{\mathrm{pr}\left[S_1(t')\right]} \, \delta\left(t' - t_k\right) \\ &= \sum_{n=1}^{\infty} \frac{n}{n!} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \, \delta\Big[h - J_{01} \sum_{m=1}^{n} \varepsilon(t-t_m)\Big] \\ &\times \quad \frac{\mathrm{pr}\left[S_1^n(0,\tau,\{t_1,\ldots,t_n\})\right]}{\mathrm{pr}\left[S_1(t')\right]} \, \delta\left(t' - t_n\right) \\ &= \sum_{n=1}^{\infty} \frac{1}{(n-1)!} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_{n-1} \, \delta\Big[h - J_{01} \sum_{m=1}^{n-1} \varepsilon(t-t_m) - J_{01}\varepsilon(t-t')\Big] \\ &\times \quad \frac{\mathrm{pr}\left[S_1^n(0,\tau,\{t_1,\ldots,t_{n-1},t'\})\right]}{\mathrm{pr}\left[S_1(t')\right]} \; .\end{aligned}$$

Because of Eq. (3.21), the rightmost factor in the last equality can be replaced by $\mathrm{pr}[S_1^{n-1}(0,\tau,\{t_1,\ldots,t_{n-1}\})]$. Furthermore, we substitute $(n-1) \to n$ and obtain

$$\begin{aligned}\mathrm{pr}\left[h_0(t) = h \mid S_1(t')\right] = \sum_{n=0}^{\infty} \int_0^\tau \mathrm{d}t_1 \cdots \int_0^\tau \mathrm{d}t_n \\ \times \quad \delta\Big[h - J_{01}\varepsilon(t-t') - J_{01} \sum_{m=1}^{n} \varepsilon(t-t_m)\Big] \, \mathrm{pr}\left[S_1^n(0,\tau,\{t_1,\ldots,t_n)\right] \\ = \quad \mathrm{pr}\left[h_0(t) = h - J_{01}\varepsilon(t-t')\right] \; .\end{aligned}$$

Analogously, it is easily proved that for the general case of N input spike trains with corresponding synaptic weights J_{0i} we have

$$\mathrm{pr}\left[h_0(t) = h \mid S_i(t')\right] = \mathrm{pr}\left[h_0(t) = h - J_{0i}\,\varepsilon(t-t')\right] \; ,$$

since the individual spike trains are presumed to be statistically independent. Inserting this result into equation (3.18) yields

$$\begin{aligned}\langle a_0(t_1)\, a_i(t_2)\rangle &= \langle a_i(t_2)\rangle \int \mathrm{d}h\, q_0(h)\, \mathrm{pr}\left[h_0(t_1) = h - J_{0i}\,\varepsilon(t_1 - t_2)\right] \\ &= \langle a_i(t_2)\rangle \int \mathrm{d}h\, q_0(h + J_{0i}\,\varepsilon(t_1 - t_2))\, \mathrm{pr}\left[h_0(t_1) = h\right] \\ &= \langle a_i(t_2)\rangle \left\langle q_0\left[h_0(t_1) + J_{0i}\,\varepsilon(t_1 - t_2)\right]\right\rangle \ .\end{aligned}$$

From that, we can finally calculate the desired input-output correlation function as defined in (3.14),

$$C_{0i}(t_1, t_2) = \frac{\left\langle q_0\left[h_0(t_1) + J_{0i}\,\varepsilon(t_1 - t_2)\right]\right\rangle}{\left\langle q_0\left[h_0(t_1)\right]\right\rangle} \ ,$$

where we have used Eq. (3.16) to replace $\langle a_0(t_1)\rangle$. Thanks to our assumption that the distribution of the neuron's membrane potential h_0 is approximately Gaussian, the above averages can be carried out at least numerically, once the first two moments of h_0 are given. According to (3.20) these are

$$\langle h_0(t)\rangle = \sum_i J_{0i} \int_0^t \mathrm{d}t' \varepsilon(t - t')\left\langle a_i(t')\right\rangle \ ,$$

and

$$\left\langle h_0(t)^2\right\rangle = \sum_{i,j} J_{0i} J_{0j} \int_0^t \int_0^t \mathrm{d}t' \mathrm{d}t''\, \varepsilon(t - t')\, \varepsilon(t - t'')\left\langle a_i(t')\, a_j(t'')\right\rangle \ ,$$

where all the quantities on the right hand sides are known from the model. The second moment can be further reduced to

$$\left\langle h_0(t)^2\right\rangle = \langle h_0(t)\rangle^2 + \sum_i J_{0i}^2 \int_{-\infty}^t \mathrm{d}t' \varepsilon(t - t')^2 \left\langle a_i(t')\right\rangle \ ,$$

because, by the statistical independence of the input spike trains, it is

$$\left\langle a_i(t') a_j(t'')\right\rangle = \left\langle a_i(t')\right\rangle \left\langle a_j(t'')\right\rangle + \delta_{ij}\delta(t' - t'')\left\langle a_i(t')\right\rangle \ .$$

Example

Following Gerstner and van Hemmen (1994) we let

$$q_0(h) := \tau_0^{-1} \exp\left[\beta\,(h - \theta)\right] \ .$$

For this choice, one finds

$$\begin{aligned}\left\langle q_0\left[h_0(t_1) + J_{0i}\,\varepsilon(t_1 - t_2)\right]\right\rangle &= \left\langle q_0\left[h_0(t_1)\right] \exp\left[\beta\, J_{0i}\,\varepsilon(t_1 - t_2)\right]\right\rangle \\ &= \left\langle q_0\left[h_0(t_1)\right]\right\rangle \exp\left[\beta\, J_{0i}\,\varepsilon(t_1 - t_2)\right] \ ,\end{aligned}$$

and therefore

$$C_{0i}(t_1, t_2) = \exp\left[\beta\, J_{0i}\,\varepsilon(t_1 - t_2)\right] \ .$$

Note, that we did not make use of the Gaussian approximation to derive this formula. As a consequence, it remains valid, even if the number of input spike trains is extremely low.

In some applications, however, the knowledge of $C_{0i}(t_1, t_2)$ may not be sufficient, but rather $\langle a_0(t_1)\, a_i(t_2)\rangle$ might be needed. Since $\langle a_0(t_1)\, a_i(t_2)\rangle$ is the product of $C_{0i}(t_1, t_2)$ with $\langle q_0\,[h_0(t_1)]\rangle$ and $\langle a_i(t_2)\rangle$, a good approximation of $\langle q_0\,[h_0(t_1)]\rangle$ is then required. Assuming that the distribution of h_0 is a Gaussian,

$$\mathrm{pr}\,[h_0(t) = h] \approx \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(h-\mu)^2}{2\sigma^2}\right] ,$$

with

$$\mu = \mu(t) := \langle h_0(t)\rangle \ , \ \ \sigma^2 = \sigma(t)^2 := \left\langle h_0(t)^2\right\rangle - \langle h_0(t)\rangle^2 \ ,$$

yields

$$\begin{aligned}
\langle q_0\,[h_0(t)]\rangle &\approx \frac{1}{\tau_0\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \mathrm{d}h\, \exp\left[\beta\,(h-\theta)\right] \exp\left[-\frac{(h-\mu)^2}{2\sigma^2}\right] \\
&= \frac{\exp\left[\beta\,(\mu-\theta)\right]}{\tau_0\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \mathrm{d}h\, \exp\left[\beta\,(h-\mu)\right] \exp\left[-\frac{(h-\mu)^2}{2\sigma^2}\right] \\
&= \frac{\exp\left[\beta\,(\mu-\theta)\right]}{\tau_0\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \mathrm{d}h\, \exp\left[\beta h\right] \exp\left[-\frac{h^2}{2\sigma^2}\right] \\
&= \frac{\exp\left[\beta\,(\mu-\theta)\right]}{\tau_0\sqrt{2\pi\sigma^2}} \int_{-\infty}^{+\infty} \mathrm{d}h\, \exp\left[-\frac{h^2}{2\sigma^2} + \beta h\right] .
\end{aligned}$$

Fortunately, the integral on the right hand side of the last equality can be carried out analytically (Gradshteyn and Ryzhik 1994), so that we obtain

$$\langle a_0(t)\rangle = \langle q_0\,[h_0(t)]\rangle \approx \tau_0^{-1} \exp\left\{\beta\,[\mu(t)-\theta] + \frac{\beta^2\sigma(t)^2}{2}\right\} .$$

Finally, we have

$$\begin{aligned}
\langle a_0(t_1)\, a_i(t_2)\rangle &= \langle a_i(t_2)\rangle\ \langle a_0(t_1)\rangle\ C_{0i}(t_1, t_2) \\
&\approx \frac{\langle a_i(t_2)\rangle}{\tau_0}\ \exp\left\{\beta\,[\mu(t_1) + J_{0i}\,\varepsilon(t_1 - t_2) - \theta] + \frac{\beta^2\sigma(t_1)^2}{2}\right\} ,
\end{aligned}$$

where all quantities are known from the model parameters.

3.3.3 Few Input Synapses

If the number N of input spike trains is very small or the inputs are strongly correlated, then a Gaussian approximation of the distribution of h_0 may not be good enough to achieve an acceptable precision when evaluating Eq. (3.18). We are now going to present an alternative method, that may help to arrive nevertheless at satisfactory results. The basic idea is to expand the nonlinear neural activation function in a Taylor series. As we will show, the moments

of the output spike train can then be expressed as a series of moments of the input processes. This approach has been studied in some detail by Kuznetsov et al. (1965).

The nth order power series expansion of $q_0(h)$ in h about some given point χ reads

$$q_0(h) =: q_0^{(0)} + q_0^{(1)}\,(h-\chi) + \cdots + q_0^{(n)}\,(h-\chi)^n + r_0^{(n)}\,(h-\chi)\ , \tag{3.22}$$

with appropriate constants $q_0^{(0)}, \ldots, q_0^{(n)}$ and the remainder $r_0^{(n)}(h-\chi)$. For the moment we do not specify χ. The problem of making a good choice for this value will be discussed below. We combine the above ansatz with (3.16) and obtain

$$\begin{aligned} \langle a_0(t)\rangle &= \langle q_0\,[h_0(t)]\rangle \\ &= q_0^{(0)} + q_0^{(1)}\,\langle [h_0(t)-\chi]\rangle + \cdots \\ &+ q_0^{(n)}\,\langle [h_0(t)-\chi]^n\rangle + \left\langle r_0^{(n)}\,[h_0(t)-\chi]\right\rangle\ . \end{aligned} \tag{3.23}$$

On the right hand side of this equation, each term $\langle [h_0(t)-\chi]^m\rangle$ can be computed, if the moments $\langle h_0(t)^m\rangle$ for $m = 1 \ldots n$ are known. But the latter can be derived directly from the input statistics, using the relation

$$\begin{aligned} h_0(t)^m &= \sum_{j_1,\ldots,j_m=1}^{N} J_{0j_1}\cdots J_{0j_m} \\ &\times \int_0^t \mathrm{d}t_1 \cdots \int_0^t \mathrm{d}t_m\, \varepsilon(t-t_1)\cdots\varepsilon(t-t_m)\, a_{j_1}(t_1)\cdots a_{j_m}(t_m)\ , \end{aligned} \tag{3.24}$$

so that

$$\begin{aligned} \langle h_0(t_1)^m\rangle &= \sum_{j_1,\ldots,j_m=1}^{N} J_{0j_1}\cdots J_{0j_m} \int_0^{t_1}\mathrm{d}t_1' \cdots \int_0^{t_1}\mathrm{d}t_m' \\ &\times \varepsilon(t_1-t_1')\cdots\varepsilon(t_1-t_m')\,\left\langle a_{j_1}(t_1')\cdots a_{j_m}(t_m')\right\rangle\ . \end{aligned}$$

Obviously, we are still left with the problem of calculating the value of the rightmost term in Eq. (3.23), the ensemble average of the remainder $r_0^{(n)}$. In general, this will turn out to be as difficult as the original problem of determining $\langle q_0[h_0(t)]\rangle$. If, however, the activation function $q_0(h)$ is analytic within a certain region around χ, then we know that within this region $r_0^{(n)}(h-\chi)$ goes to zero as n approaches infinity. If, furthermore, the probability distribution of the membrane potential $h_0(t)$ – which is determined by the input processes – is sufficiently concentrated near χ, then we may hope that the average $\left\langle r_0^{(n)}(h-\chi)\right\rangle$ decreases to zero with increasing n as well. In that case, the mean activity $\langle a_0(t)\rangle$ can be derived from (3.23) to any degree of precision by choosing n large enough and neglecting the remainder $r_0^{(n)}$.

The above considerations emphasize, that an appropriate choice of the point χ may be crucial for the quality of the results achieved. Since the chances

for the method to work fine are best if the distribution of $h_0(t)$ is concentrated near χ, it would be natural to let χ be equal to the expectation value of $h_0(t)$,

$$\chi = \chi(t) := \langle h_0(t) \rangle = \sum_{k=1}^{N} J_{0k} \int_0^t \mathrm{d}t' \varepsilon(t-t') \left\langle a_k(t') \right\rangle \ .$$

Notice that, in general, with χ being dependent on t, the series coefficients $q_0^{(0)} \ldots q_0^{(n)}$ and the remainder $r_0^{(n)}$ will also be time dependent.

In order to find an expression for the second moment of the input-output statistics, $\langle a_0(t_1) a_i(t_2) \rangle$ with $i = 1 \ldots N$, we now insert the expansion (3.22) into Eq. (3.20). That yields

$$\begin{aligned} \langle a_0(t_1)\, a_i(t_2) \rangle &= \langle a_i(t_2)\, q_0\,[h_0(t_1)] \rangle \\ &= q_0^{(0)} \langle a_i(t_2) \rangle + q_0^{(1)} \langle a_i(t_2)\,[h_0(t_1) - \chi] \rangle + \cdots \\ &+ q_0^{(n)} \langle a_i(t_2)\,[h_0(t_1) - \chi]^n \rangle + \left\langle a_i(t_2) r_0^{(n)}\,[h_0(t_1) - \chi] \right\rangle \ . \end{aligned} \tag{3.25}$$

Let us again postpone the specification of χ until the end of the calculations. The quantities $\langle a_i(t_2)[h_0(t_1) - \chi]^m \rangle$ needed on the right hand side of (3.25) can be determined from the input processes, because

$$\begin{aligned} \langle a_i(t_2) h_0(t_1)^m \rangle &= \sum_{j_1,\ldots,j_m=1}^{N} J_{0j_1} \cdots J_{0j_m} \int_{-\infty}^{t_1} \cdots \int_{-\infty}^{t_1} \mathrm{d}t_1' \cdots \mathrm{d}t_m' \\ &\times \ \varepsilon(t_1 - t_1') \cdots \varepsilon(t_1 - t_m') \left\langle a_i(t_2)\, a_{j_1}(t_1') \cdots a_{j_m}(t_m') \right\rangle \ , \end{aligned}$$

where Eq. (3.24) has been applied to replace $h_0(t_1)^m$. What remains unknown in (3.24) is the term $\left\langle a_i(t_2) r_0^{(n)}[h_0(t_1) - \chi] \right\rangle$. In the same way as we have derived Eq. (3.20) from (3.18), it is easily shown that

$$\left\langle a_i(t_2)\, r_0^{(n)}\,[h_0(t_1) - \chi] \right\rangle = \langle a_i(t_2) \rangle \int r_0^{(n)}(h - \chi)\, \mathrm{pr}\,[h_0(t_1) = h \mid S_i(t_2)]\ \mathrm{d}h \ . \tag{3.26}$$

If the conditional probability density $\mathrm{pr}\,[h_0(t_1) = h \mid S_i(t_2)]$ is sufficiently concentrated within a region about χ where $q_0(h)$ is analytic, then, hopefully, the conditional expectation on the right hand side of (3.26) approaches zero as n goes to infinity. Under that condition, the last term in (3.25) can be dropped if n is chosen large enough.

Presuming that the probability density under consideration is concentrated near its first moment, it would be natural to choose χ as the conditional expectation value of $h_0(t_1)$ given $S_i(t_2)$,

$$\chi = \chi(t_1, t_2) := \int h\, \mathrm{pr}\,[h_0(t_1) = h \mid S_i(t_2)]\ \mathrm{d}h \ .$$

Using Eq. (3.19), this can equivalently be written

$$\chi(t_1, t_2) = \frac{\langle a_i(t_2) h_0(t_1) \rangle}{\langle a_i(t_2) \rangle} = \sum_{k=1}^{N} J_{0k} \int_0^{t_1} \mathrm{d}t' \varepsilon(t_1 - t') \frac{\langle a_i(t_2)\, a_k(t') \rangle}{\langle a_i(t_2) \rangle} \ ,$$

and is thus easily derived from the input statistics.

3.4 Recurrent Networks

In Sect. 3.3 we have analyzed a very simple neural network consisting of a single neuron receiving at its dendritic synapses input from statistically independent stochastic processes that were not causally connected to the output. Throughout that whole section the neuron's refractory potential, a sort of self-input, was neglected. In what follows we want to address the general case of a fully connected network of spike response-neurons, in which each neuron can receive input from all the other neurons as well as from itself.

We will concentrate our calculations to the derivation of the mean activity and the two-spike correlation function in a recurrent neural network, because these are the quantities that will be needed for an analytical investigation of the Hebbian learning dynamics in our model of primary visual cortex. Nevertheless, the methods that will be presented can easily be generalized to obtain expressions for higher-order correlation functions.

3.4.1 Notation

To render expressions as simple as possible, we will extensively use the following conventions of notation.

- **Multiplication of functions.** For two functions $f, g : \mathbb{N}_0^n \times \mathbb{R}^n \to \mathbb{R}$, with $n \in \mathbb{N}$, we define the product $(fg) : \mathbb{N}_0^n \times \mathbb{R}^n \to \mathbb{R}$ as

$$\begin{aligned} [fg]_{i_1 \ldots i_n}(t_1, \ldots, t_n) &:= f_{i_1 \ldots i_n}(t_1, \ldots, t_n)\, g_{i_1 \ldots i_n}(t_1, \ldots, t_n) \\ &= [gf]_{i_1 \ldots i_n}(t_1, \ldots, t_n) , \end{aligned}$$

  where $i_1, \ldots, i_n \in \mathbb{N}_0$ and $t_1, \ldots, t_n \in \mathbb{R}$. If the function f has no zeroes, then we can also write

$$\left[\frac{1}{f} g\right]_{i_1 \ldots i_n}(t_1, \ldots, t_n) := \frac{1}{f_{i_1 \ldots i_n}(t_1, \ldots, t_n)}\, g_{i_1 \ldots i_n}(t_1, \ldots, t_n) .$$

  If, furthermore, $\hat{A}$ is an operator acting upon a function h so that the result is a function $\hat{A}h : \mathbb{N}_0^n \times \mathbb{R}^n \to \mathbb{R}$, then we define the operator $(f\hat{A})$ by

$$\begin{aligned} \Big[(f\hat{A})h\Big]_{i_1 \ldots i_n}(t_1, \ldots, t_n) &:= \Big[f(\hat{A}h)\Big]_{i_1 \ldots i_n}(t_1, \ldots, t_n) \\ &= f_{i_1 \ldots i_n}(t_1, \ldots, t_n) \Big[\hat{A}h\Big]_{i_1 \ldots i_n}(t_1, \ldots, t_n) . \end{aligned}$$

- **Tensor product of functions.** With two functions $f, g : \mathbb{N}_0 \times \mathbb{R} \to \mathbb{R}$, we denote

$$[f \otimes g]_{ij}(t_1, t_2) := f_i(t_1)\, g_j(t_2) .$$

- **Tensor product of operators.** With two operators $\hat{A}$ and $\hat{B}$ acting upon functions f and g so that $(\hat{A}f), (\hat{B}g) : \mathbb{N}_0 \times \mathbb{R} \to \mathbb{R}$, we define

$$[(\hat{A} \otimes \hat{B})(f \otimes g)]_{ij}(t_1, t_2) := [\hat{A}f]_i(t_1)\, [\hat{B}g]_j(t_2) .$$

- **Time ordering operator.** Given a function $f_{ij}(t_1, t_2)$, let us define the time ordering operator $\hat{\tau}$ as

$$[\hat{\tau} f]_{ij}(t_1, t_2) := \begin{cases} f_{ij}(t_1, t_2) , & \text{if } t_1 \leq t_2 , \\ f_{ji}(t_2, t_1) , & \text{if } t_1 > t_2 . \end{cases}$$

- **Ensemble average.** As in the previous sections, the average of a quantity x over an ensemble of networks will be written $\langle x \rangle$. When applying the average to a function, we will often draw the arguments of the function outside the brackets, as in

$$\langle f_{i_1 \ldots i_n}(t_1 \ldots t_n) \rangle =: \langle f \rangle_{i_1 \ldots i_n}(t_1 \ldots t_n) .$$

3.4.2 General Considerations

We consider a network consisting of N mutually connected spike response-neurons. The spike train of neuron number $i \in \{1, \ldots, N\}$ is denoted by $a_i(t)$ as introduced in Sect. 3.1, and the two-spike correlation function of two spike trains i and j is defined as

$$C_{ij}(t_1, t_2) := \frac{\langle a \otimes a \rangle_{ij}(t_1, t_2)}{\langle a \rangle_i (t_1) \langle a \rangle_j (t_2)} , \quad i, j = 1, \ldots, N . \tag{3.27}$$

Given the the matrix J_{ij} of synaptic weights in the network, i.e., given any fixed network architecture, it is our aim to find an approximation for this correlation function. We begin by deriving general expressions for the mean and the second moment of the neural spike activity, namely,

$$\langle a \rangle = \langle \hat{q} h \rangle , \tag{3.28}$$

$$\langle a \otimes a \rangle = \hat{\tau} \langle a \otimes \hat{q} h \rangle + \hat{\delta} \langle a \rangle , \tag{3.29}$$

where a nonlinear operator $\hat{q}$ and a linear operator $\hat{\delta}$ have been defined as

$$[\hat{q} f]_i (t) := q_i [f_i(t)] ,$$

$$\left[\hat{\delta} f\right]_{ij}(t_1, t_2) := \delta_{ij} \delta(t_1 - t_2) f_i(t_1) ,$$

with a test function $f : \mathbb{N}_0 \times \mathbb{R} \to \mathbb{R}$.

Equation (3.28) is easily proved. Recalling the line of arguments that has led to Eq. (3.16) for the case of a feedforward network, we can quickly convince ourselves that for a recurrent neural network it is

$$\langle a_i(t) \rangle = \langle q_i[h_i(t)] \rangle , \quad i = 1 \ldots N ,$$

which is equivalent to (3.28).

In contrast, the derivation (3.29) is a bit more involved, so let us proceed step by step. In analogy to (3.17), we first find

$$\langle a_i(t_1)\, a_j(t_2) \rangle \, \mathrm{d}t_1 \mathrm{d}t_2 = \mathrm{pr}\left[S_i(t_1) \wedge S_j(t_2)\right] \mathrm{d}t_1 \, \mathrm{d}t_2 = \tag{3.30}$$

$$= \int \mathrm{d}h \, \Pr\left[S_j(t_2, \mathrm{d}t_2) \mid h_j(t_2) = h \ \wedge S_i(t_1, \mathrm{d}t_1)\right] \mathrm{pr}\left[S_i(t_1, \mathrm{d}t_1) \wedge h_j(t_2) = h\right] .$$

Next, we will replace the conditional probability on the right hand side. For the moment, we presume that either $i \neq j$ or $t_1 \neq t_2$, i.e., the event $S_i(t_1, \mathrm{d}t_1)$ does not imply $S_j(t_2, \mathrm{d}t_2)$. Since within the framework of the spike response model, a neuron's firing probability density is coupled to earlier spikes in the network only via its membrane potential, we have

$$\begin{aligned} &\Pr\left[S_j(t_2, \mathrm{d}t_2) \mid h_j(t_2) = h \ \wedge S_i(t_1, \mathrm{d}t_1)\right] \\ &= \ \Pr\left[S_j(t_2, \mathrm{d}t_2) \mid h_j(t_2) = h\right] = q_j(h)\mathrm{d}t_2 \ , \ \text{ if } t_1 \leq t_2 \text{ and } (i, t_1) \neq (j, t_2) \ , \end{aligned} \tag{3.31}$$

where $q_j(h)$ is the activation function of neuron j as introduced in Sect. 3.1.

Notice, however, that (3.31) does not hold true for $t_1 > t_2$, because an action potential at time t_1 does carry information not only about the membrane potential but also about the actual presence of action potentials at an earlier time t_2: A spike at time t_1 can be generated in response to the postsynaptic potential elicited by a presynaptic spike at t_2. Hence, the probability that a spike has occurred at time t_2, given the membrane voltage at t_2 *and* the presence of an action potential at $t_1 > t_2$, is different from the probability to find a spike at t_2, given only the corresponding membrane voltage.

Now assume $i = j$ and $t_1 = t_2$. In that case we have

$$\Pr[S_j(t_2, \mathrm{d}t_2) \wedge S_i(t_1, \mathrm{d}t_1) \wedge h_j(t_2) = h] = \Pr[S_i(t_1, \mathrm{d}t_1) \wedge h_j(t_2) = h] \ ,$$

so that

$$\Pr\left[S_j(t_2, \mathrm{d}t_2) \mid h_j(t_2) = h \ \wedge S_i(t_1, \mathrm{d}t_1)\right] = 1 \ , \ \text{ if } (i, t_1) = (j, t_2) \ . \tag{3.32}$$

Putting together Eqs. (3.31) and (3.32), we arrive at

$$\mathrm{pr}\left[S_j(t_2) \mid h_j(t_2) = h \ \wedge S_i(t_1)\right] = q_j(h) + \delta_{ij}\delta(t_1 - t_2), \ \text{ if } t_1 \leq t_2 \ . \tag{3.33}$$

This can be inserted into (3.30) and thus yields

$$\begin{aligned} \langle a_i(t_1)\, a_j(t_2)\rangle &= \int \mathrm{d}h \ \left[q_j(h) + \delta_{ij}\delta(t_1 - t_2)\right] \mathrm{pr}\left[S_i(t_1) \wedge h_j(t_2) = h\right] \\ &= \ \delta_{ij}\delta(t_1 - t_2)\mathrm{pr}\left[S_i(t_1)\right] + \int \mathrm{d}h\, q_j(h)\mathrm{pr}\left[S_i(t_1) \wedge h_j(t_2) = h\right] \\ &= \ \delta_{ij}\delta(t_1 - t_2)\left\langle a_i(t_1)\right\rangle + \left\langle a_i(t_1)\right\rangle \int \mathrm{d}h\, q_j(h)\, \mathrm{pr}\left[h_j(t_2) = h \mid S_i(t_1)\right] \\ &= \ \delta_{ij}\delta(t_1 - t_2)\left\langle a_i(t_1)\right\rangle + \left\langle a_i(t_1)\, q_j\left[h_j(t_2)\right]\right\rangle \ , \ \text{ if } t_1 \leq t_2 \ , \end{aligned} \tag{3.34}$$

where the last equality results from applying the same transformations that have been used to derive Eq. (3.20) from (3.18). In order to put (3.34) into a more compact form, we define the (nonlinear) operator $\hat{q}$ by

$$\left[\hat{q}h\right]_i(t) := q_i[h_i(t)] \ ,$$

and the operator $\hat{\delta}$ as

$$\left[\hat{\delta}f\right]_{ij}(t_1, t_2) := \delta_{ij}\, \delta(t_1 - t_2)\, f_i(t_1) \ ,$$

so that we can write

$$\langle a \otimes a \rangle_{ij}(t_1, t_2) = \langle a \otimes \hat{q}h \rangle_{ij}(t_1, t_2) + \left[\hat{\delta}\langle a \rangle\right]_{ij}(t_1, t_2) , \quad \text{for } t_1 \leq t_2 .$$

Instead of continuously keeping track of the constraint $t_1 \leq t_2$, let us now simply apply the time ordering operator $\hat{\tau}$ to both sides of this equation. As the time ordering operator consistently rearranges all arguments of the involved functions in such a way that the time arguments are in the required order, the constraint can be dropped and we have

$$[\hat{\tau}\langle a \otimes a \rangle]_{ij}(t_1, t_2) = [\hat{\tau}\langle a \otimes \hat{q}h \rangle]_{ij}(t_1, t_2) + \left[\hat{\tau}\hat{\delta}\langle a \rangle\right]_{ij}(t_1, t_2) ,$$

for arbitrarily ordered times t_1 and t_2. Because of the relations

$$\langle a \otimes a \rangle_{ij}(t_1, t_2) = \langle a_i(t_1)\, a_j(t_2) \rangle = \langle a_j(t_2)\, a_i(t_1) \rangle = \langle a \otimes a \rangle_{ji}(t_2, t_1)$$

and

$$\left[\hat{\delta}\langle a \rangle\right]_{ij}(t_1, t_2) = \left[\hat{\delta}\langle a \rangle\right]_{ji}(t_2, t_1) ,$$

the time ordering operator is lost upon the functions $\langle a \otimes a \rangle$ and $\hat{\delta}\langle a \rangle$. Hence it is

$$\langle a \otimes a \rangle_{ij}(t_1, t_2) = [\hat{\tau}\langle a \otimes \hat{q}h \rangle]_{ij}(t_1, t_2) + \left[\hat{\delta}\langle a \rangle\right]_{ij}(t_1, t_2) ,$$

or in short

$$\langle a \otimes a \rangle = \hat{\tau}\langle a \otimes \hat{q}h \rangle + \hat{\delta}\langle a \rangle ,$$

as claimed in (3.29).

3.4.3 Expanding the Activation Function

As motivated in Sect. 3.3, we will now try to gain access to Eqs. (3.28) and (3.29) by expanding the neural activation function $q_i(h)$ into a power series in the membrane potential h,

$$q_j(h) = q^{(0)} + q^{(1)}(h - \chi) + r(h - \chi) .$$

We will apply this expansion for the purpose of obtaining approximations of the ensemble averages $\langle \hat{q}h \rangle$ and $\langle a \otimes \hat{q}h \rangle$. Normally, the precision of the resulting approximations will depend on the choice of χ. One may therefore consider using two different expansions for approximating $\langle \hat{q}h \rangle_i(t)$ and $\langle a \otimes \hat{q}h \rangle_{ij}(t, t')$. In the former χ might be chosen as a function of i and t, whereas in the latter χ might depend on i, j, t, and t'. Here we do not need such a general ansatz so that we simply write

$$q_i[h_i(t)] = q_i^{(0)}(t) + q_i^{(1)}(t)\,[h_i(t) - \chi_i(t)] + r_i[h_i(t) - \chi_i(t), t] . \tag{3.35}$$

For convenience, we define a new (non-linear) operator $\hat{r}$ acting on a function $f : \mathbb{N} \times \mathbb{R} \to \mathbb{R}$,

$$[\hat{r}f]_i(t) := r_i[f_i(t), t] .$$

This allows to rewrite (3.35) in the form

$$[\hat{q}h]_i(t) = q_i^{(0)}(t) + q_i^{(1)}(t)\,[h_i(t) - \chi_i(t)] + [\hat{r}(h-\chi)]_i(t)\,,$$

where the arguments can be dropped so as to obtain

$$\hat{q}h = q^{(0)} + q^{(1)}\,(\mathbb{1}\otimes h - \chi) + \hat{r}(h-\chi)\,. \tag{3.36}$$

Inserting this result into Eq. (3.28) we obtain

$$\langle a\rangle = q^{(0)} + q^{(1)}\,(\langle h\rangle - \chi) + \langle \hat{r}(h-\chi)\rangle\ . \tag{3.37}$$

In a similar way we proceed for Eq. (3.29) and get

$$\langle a\otimes a\rangle = \hat{\tau}\left\langle a\otimes\left[q^{(0)} + q^{(1)}\,(h-\chi) + \hat{r}(h-\chi)\right]\right\rangle + \hat{\delta}\,\langle a\rangle\ . \tag{3.38}$$

Next, we solve (3.37) for the function $q^{(0)}$ and insert the result into (3.38). After a few simple transformations this leads us to

$$\begin{aligned}\langle a\otimes a\rangle &= \langle a\rangle\otimes\langle a\rangle + \hat{\delta}\,\langle a\rangle + \hat{\tau}\,\left[\left\langle a\otimes(q^{(1)}h)\right\rangle - \langle a\rangle\otimes\left\langle q^{(1)}h\right\rangle\right]\\ &+ \hat{\tau}\,\langle a\otimes\hat{r}(h-\chi)\rangle - \hat{\tau}\,[\langle a\rangle\otimes\langle\hat{r}(h-\chi)\rangle]\ .\end{aligned} \tag{3.39}$$

According to the definition of our neuron model (Sect. 3.1), the membrane potential h of any neuron at any time t is a function of the network's spike activity a up to time t. To be specific, h can be written as a linear coupling operator $\hat{\kappa}$ acting upon a,

$$h = \hat{\kappa}a\,,$$

as introduced in Eq. (3.4). Using this relation in (3.39), we find

$$\begin{aligned}\langle a\otimes a\rangle &= \langle a\rangle\otimes\langle a\rangle + \hat{\tau}\left[(\hat{\mathbb{E}}\otimes q^{(1)}\hat{\kappa})\,(\langle a\otimes a\rangle - \langle a\rangle\otimes\langle a\rangle)\right] + \hat{\delta}\,\langle a\rangle\\ &+ \hat{\tau}\,\langle a\otimes\hat{r}(\hat{\kappa}a-\chi)\rangle - \hat{\tau}\langle a\rangle\otimes\langle\hat{r}(\hat{\kappa}a-\chi)\rangle\,,\end{aligned}$$

where $\hat{\mathbb{E}}$ denotes the identity operator, i.e., $\hat{\mathbb{E}}\,a = a$. Analogously, we can replace h in Eq. (3.37) and get

$$\langle a\rangle = q^{(0)} + q^{(1)}\,(\langle\hat{\kappa}a\rangle - \chi) + \langle\hat{r}(\hat{\kappa}a-\chi)\rangle\ .$$

After defining the functions

$$K := \langle a\otimes a\rangle - \langle a\rangle\otimes\langle a\rangle \tag{3.40}$$

and

$$R := \hat{r}\,(\hat{\kappa}a - \chi)\ , \tag{3.41}$$

we finally obtain the equations

$$\langle a\rangle = q^{(0)} + q^{(1)}\,(\langle\hat{\kappa}a\rangle - \chi) + \langle R\rangle\ , \tag{3.42}$$

$$\left[(\hat{\mathbb{E}}\otimes\hat{\mathbb{E}}) - \hat{\tau}(\hat{\mathbb{E}}\otimes q^{(1)}\hat{\kappa})\right]K = \hat{\delta}\,\langle a\rangle + \hat{\tau}\,\langle a\otimes R\rangle - \hat{\tau}\langle a\rangle\otimes\langle R\rangle\,, \tag{3.43}$$

from which we shall try to determine the mean activity $\langle a \rangle$ and the second moment of the spike statistics K.

It should be emphasized that up to now we did not introduce any kind of approximation, i.e., within the framework of the spike response model, equations (3.42) and (3.43) are exact. The problem is, however, to determine $\langle R \rangle$ and $\langle a \otimes R \rangle$. In the following subsections, we will present two simple methods to handle these quantities and obtain approximations of $\langle a \rangle$ and K. Especially the second method yields quite respectable results as we will demonstrate in Sect. 3.5 by means of numerical simulations.

3.4.4 Linear Neurons

Let us now assume that within our network's operating regime, the neural activation function $q_j(h)$ is given by a linear function. Then we have

$$q_j(h) = \varrho_j^{(0)} + \varrho_j^{(1)} h \,, \tag{3.44}$$

with two real constants $\varrho_j^{(0)}$ and $\varrho_j^{(1)}$.

The activation function $q_j(h)$ is defined to be the probability density that neuron j emits a spike, given that its membrane potential h_j is equal to h. Since any probability density must be greater than or equal to zero, we have to require $q_j(h) \geq 0$, which means that $\varrho_j^{(1)} h_j(t) \geq -\varrho_j^{(0)}$ must hold for the membrane potential h_j of any neuron at any time. If the network under consideration is free of inhibition and the neurons are non-refractory, then the membrane potentials are always positive or zero, so that the above requirement is automatically fulfilled for $\varrho_j^{(0)}, \varrho_j^{(1)} \geq 0$. If on the other hand, the condition $\varrho_j^{(1)} h_j(t) \geq -\varrho_j^{(0)}$ is not fulfilled in all but in most network runs, then the linear ansatz may nevertheless yield good approximative results.

A comparison of (3.44) with (3.35) suggests to choose $\chi_i(t) \equiv 0$ and to identify

$$q_i^{(0)}(t) \equiv \varrho_i^{(0)} \,, \;\; q_i^{(1)}(t) \equiv \varrho_i^{(1)} \,, \;\; r_i(h,t) \equiv 0 \,,$$

so that Eqs. (3.42) and (3.43) read

$$\langle a \rangle = \varrho^{(0)} + \varrho^{(1)} \hat{\kappa} \langle a \rangle \,,$$

and

$$\left[(\hat{\mathbb{E}} \otimes \hat{\mathbb{E}}) - \hat{\tau} (\hat{\mathbb{E}} \otimes \varrho^{(1)} \hat{\kappa}) \right] K = \hat{\delta} \langle a \rangle \,. \tag{3.45}$$

Assuming that the operators $[\hat{\mathbb{E}} - \varrho^{(1)} \hat{\kappa}]$ and $[(\hat{\mathbb{E}} \otimes \hat{\mathbb{E}}) - \hat{\tau} (\hat{\mathbb{E}} \otimes \varrho^{(1)} \hat{\kappa})]$ can be inverted, we arrive at

$$\langle a \rangle = \left[\hat{\mathbb{E}} - \varrho^{(1)} \hat{\kappa} \right]^{-1} \varrho^{(0)} \,, \tag{3.46}$$

$$K = \left[(\hat{\mathbb{E}} \otimes \hat{\mathbb{E}}) - \hat{\tau} (\hat{\mathbb{E}} \otimes \varrho^{(1)} \hat{\kappa}) \right]^{-1} \hat{\delta} \langle a \rangle \,. \tag{3.47}$$

If $\|\varrho^{(1)}\|$ is sufficiently small, then $\|\varrho^{(1)}\hat{\kappa}\| < 1$ and $\|\hat{\tau}(\hat{\mathbb{E}} \otimes \varrho^{(1)}\hat{\kappa})\| < 1$. In that case the inverse operators in (3.46) and (3.47) can be expressed by their von-Neumann series,

$$\left[\hat{\mathbb{E}} - \varrho^{(1)}\hat{\kappa}\right]^{-1} = \hat{\mathbb{E}} + \varrho^{(1)}\hat{\kappa} + \left[\varrho^{(1)}\hat{\kappa}\right]^2 + \cdots , \tag{3.48}$$

$$\left[(\hat{\mathbb{E}} \otimes \hat{\mathbb{E}}) - \hat{\tau}(\hat{\mathbb{E}} \otimes \varrho^{(1)}\hat{\kappa})\right]^{-1} = \hat{\mathbb{E}} \otimes \hat{\mathbb{E}} + \hat{\tau}(\hat{\mathbb{E}} \otimes \varrho^{(1)}\hat{\kappa}) + \left[\hat{\tau}(\hat{\mathbb{E}} \otimes \varrho^{(1)}\hat{\kappa})\right]^2 + \cdots , \tag{3.49}$$

and we can obtain approximations for the left hand sides by summing over a finite number of terms on the right hand sides.

Hawkes (1971a, b) has presented a set of equations which is equivalent to (3.45) and (3.46) under the condition that the network ensemble is stationary, so that $\langle a \rangle_i$ is independent of time and $K_{ij}(t_1, t_2)$ does only depend on the time difference $(t_1 - t_2)$. The author introduces a matrix $\gamma = (\gamma_{ij})$ to represent the coupling operator $\hat{\kappa}$ as follows,

$$\varrho_j^{(1)} \left[\hat{\kappa} g\right]_j (t) =: \int_{-\infty}^{t} \mathrm{d}t' \sum_k \gamma_{jk}(t - t') g_k(t') ,$$

and furthermore defines the Fourier transforms

$$f_{jk}(\omega) := \int_{-\infty}^{\infty} \mathrm{d}t\, e^{-i\omega t} K_{jk}(t) , \quad G_{jk}(\omega) := \int_{-\infty}^{\infty} \mathrm{d}t\, e^{-i\omega t} \gamma_{jk}(t) .$$

From that he obtains

$$\langle a \rangle = \left[\mathbb{E} - G(0)\right]^{-1} \varrho^{(0)} ,$$

where $\mathbb{E}$ denotes the identity matrix and $G(\omega)$ is to be read as the matrix whose elements are given by $G_{jk}(\omega)$. Then he defines the diagonal matrix

$$D := \mathrm{diag}\left(\langle a \rangle_1 \ldots, \langle a \rangle_N\right) ,$$

and shows that a solution K of (3.45) has a Fourier transform $f(\omega) = [f_{ij}(\omega)]$ given by

$$f(\omega) = \left[\mathbb{E} - G(\omega)\right]^{-1} D \left[\mathbb{E} - G^{\mathrm{T}}(-\omega)\right]^{-1} .$$

3.4.5 Non-Linear Approximation

How can we proceed from Eqs. (3.42) and (3.43), if the linear ansatz shown in the previous subsection does not yield satisfactory results? In that case we have to find useful approximations for the terms $\langle R \rangle$ and $\langle a \otimes R' \rangle$. From (3.41) we notice that both of these are ensemble averages of some nonlinear operator acting upon the vector of neural spike trains a. Unfortunately, we do not know the probability density of a so as to calculate these averages. Otherwise we would not need Eqs. (3.42) and (3.43), because we could directly derive $\langle a \rangle$ and $\langle a \otimes a \rangle$ from the given density.

A widely-used heuristic technique for estimating complicated averages is to replace the average $\langle f(x)\rangle$ of some function $f(x)$ by the value of the function taken at the average $\langle x\rangle$ of its argument, i.e.,

$$\langle f(x)\rangle \approx f(\langle x\rangle)\,. \tag{3.50}$$

In (3.42) we can take advantage of this method by substituting

$$\langle R\rangle = \langle \hat{r}\,(\hat{\kappa}a - \chi)\rangle \approx \hat{r}\,(\hat{\kappa}\,\langle a\rangle - \chi)\ , \tag{3.51}$$

so that using (3.36) we arrive at

$$\langle a\rangle \approx \hat{q}\,(\hat{\kappa}\,\langle a\rangle)\ . \tag{3.52}$$

This is a system of N non-linear integral equations for the N functions $\langle a\rangle_i$. It can be solved numerically.

The most simple case arises if the network ensemble is in a stationary and homogeneous state. Stationarity means that the averaged activities $\langle a\rangle_i$ as well as the mean membrane potentials $\langle h\rangle_i$ are constants in time. Homogeneity implies these constants to be the same for all neurons. Let us therefore denote $\langle a\rangle_i(t) \equiv a_*$ and $\langle h\rangle_i(t) \equiv h_*$. Since neural activity and membrane potential are related via Eq. (3.4), we have $\langle h\rangle = \hat{\kappa}\,\langle a\rangle$. The ensemble being in a stationary and homogeneous state thus requires $\hat{\kappa}$ to be of an appropriate form so that for constant $\langle a\rangle \equiv a_*$ this relation writes

$$\langle h\rangle_i(t) = [\hat{\kappa}\,\langle a\rangle]_i(t) = \kappa\, a_* = h_* \quad \forall\, i \in \{1,\ldots,N\}\,,$$

with some real constant κ. Then the system (3.52) reduces to one single non-linear equation

$$a_* \approx \hat{q}\,(\kappa a_*)\ ,$$

which is to be solved for the constant $a_* \in \mathbb{R}$. Any root a_* of this equation can be analyzed for stability by linearizing (3.52), and it represents a stationary and homogeneous network state if it turns out to be stable.

Now we apply the approximation (3.50) to Eq. (3.43). Substituting

$$\langle R\rangle \approx \hat{r}\,(\hat{\kappa}\,\langle a\rangle - \chi)$$

and

$$\langle a \otimes R\rangle \approx \langle a\rangle \otimes \hat{r}\,(\hat{\kappa}\,\langle a\rangle - \chi) \tag{3.53}$$

yields

$$\Big[(\hat{\mathbb{E}}\otimes\hat{\mathbb{E}}) - \hat{\tau}(\hat{\mathbb{E}}\otimes q^{(1)}\hat{\kappa})\Big]\,K \approx \hat{\delta}\,\langle a\rangle\ , \tag{3.54}$$

and we are practically back to Eq. (3.45), which has been derived for a network of linear neurons. We are still free, however, to choose the points $\chi_i(t)$ about which we apply the expansion (3.35) of the neural activation functions $q_i(h)$. Since this expansion is intended to yield a good approximation for the value $q_i[h_i(t)]$ at any time t, a sensible choice for the point of expansion would be the expectation $\langle h_i(t)\rangle$ of h_i at that time. We thus let

$$\chi_i(t) := \langle h\rangle_i(t) = [\hat{\kappa}\,\langle a\rangle]_i(t)\,, \tag{3.55}$$

so that

$$q_i^{(1)}(t) := \left.\frac{\mathrm{d}q_i(h)}{\mathrm{d}h}\right|_{h=[\hat{\kappa}\langle a\rangle]_i(t)} , \tag{3.56}$$

where $\langle a\rangle$ is given by the solution of (3.52). Taken together, Eqs. (3.52), (3.54), and (3.56) provide a first non-linear approximation for the moments $\langle a\rangle$ and K of the spike activity a in a recurrent neural network.

There is, however, a better way to treat the quantity $\langle a \otimes R\rangle$. It arises from the observation that this average can equivalently be written in the form of a conditional expectation,

$$\begin{aligned}\langle a \otimes R\rangle_{ij}(t_1,t_2) = \\ = \langle a_i(t_1)\rangle \int \mathrm{d}h\, r_j\left[h - \chi_j(t_2), t_2\right]\ \mathrm{pr}\left[h_j(t_2) = h \mid S_i(t_1)\right] .\end{aligned} \tag{3.57}$$

That means that $\langle a \otimes R\rangle_{ij}(t_1,t_2)$ is equal to the probability of a spike occurring at neuron i and time t_1 times the conditional expectation of r_j given the presence of a spike at neuron i and time t_1. This relation is easily verified in analogy to the derivation of (3.20) from (3.18).

As is suggested by (3.50), we replace in (3.57) the conditional expectation of the function r_j by r_j applied to the conditional expectation of its argument, i.e., we set

$$\begin{aligned}\int r_j\left[h - \chi_j(t_2), t_2\right]\ \mathrm{pr}\left[h_j(t_2) = h \mid S_i(t_1)\right]\ \mathrm{d}h \approx \\ \approx r_j\left[\int h\ \mathrm{pr}\left[h_j(t_2) = h \mid S_i(t_1)\right]\ \mathrm{d}h - \chi_j(t_2), t_2\right] .\end{aligned} \tag{3.58}$$

This yields

$$\langle a \otimes R\rangle_{ij}(t_1,t_2) \approx \langle a_i(t_1)\rangle\ r_j\left[\frac{\langle a_i(t_1)h_j(t_2)\rangle}{\langle a_i(t_1)\rangle} - \chi_j(t_1,t_2), t_2\right] , \tag{3.59}$$

where we have used

$$\langle a_i(t_1)\rangle \int \mathrm{d}h\, h\ \mathrm{pr}\left[h_j(t_2) = h \mid S_i(t_1)\right] = \langle a_i(t_1)h_j(t_2)\rangle ,$$

which can be proved using (3.19). Let us now introduce the constant function

$$\mathbb{1}_i(t) \equiv 1 , \tag{3.60}$$

and an extended version of the operator $\hat{r}$,

$$[\hat{r}_* f]_{ij}(t_1,t_2) := r_j\left[f_{ij}(t_1,t_2), t_2\right] , \tag{3.61}$$

acting upon a test function $f : \mathbb{N}^2 \times \mathbb{R}^2 \to \mathbb{R}$. As a result, Eq. (3.59) can be written in the form

$$\langle a \otimes R\rangle \approx \langle a \otimes \mathbb{1}\rangle\, \hat{r}_* \left[\frac{1}{\langle a \otimes \mathbb{1}\rangle}\langle a \otimes h\rangle - \mathbb{1} \otimes \chi\right] ,$$

and by choosing χ according to (3.55) we obtain

$$\begin{aligned}
\langle a \otimes R \rangle &\approx \langle a \otimes \mathbb{1} \rangle \hat{r}_* \left[\frac{1}{\langle a \otimes \mathbb{1} \rangle} \langle a \otimes h \rangle - \mathbb{1} \otimes \langle h \rangle \right] \\
&= \langle a \otimes \mathbb{1} \rangle \hat{r}_* \left[\frac{1}{\langle a \otimes \mathbb{1} \rangle} \langle a \otimes \hat{\kappa} a \rangle - \mathbb{1} \otimes \hat{\kappa} \langle a \rangle \right] \\
&= \langle a \otimes \mathbb{1} \rangle \hat{r}_* \left[\frac{1}{\langle a \otimes \mathbb{1} \rangle} \langle a \otimes \hat{\kappa} a \rangle - \mathbb{1} \otimes \hat{\kappa} \langle a \rangle \right] \\
&= \langle a \otimes \mathbb{1} \rangle \hat{r}_* \left[\frac{1}{\langle a \otimes \mathbb{1} \rangle} \left(\langle a \otimes \hat{\kappa} a \rangle - \langle a \rangle \otimes \hat{\kappa} \langle a \rangle \right) \right] \\
&= \langle a \otimes \mathbb{1} \rangle \hat{r}_* \left[\frac{1}{\langle a \otimes \mathbb{1} \rangle} (\hat{\mathbb{E}} \otimes \hat{\kappa}) \left(\langle a \otimes a \rangle - \langle a \rangle \otimes \langle a \rangle \right) \right] \\
&= \langle a \otimes \mathbb{1} \rangle \hat{r}_* \left[\frac{1}{\langle a \otimes \mathbb{1} \rangle} (\hat{\mathbb{E}} \otimes \hat{\kappa}) K \right] .
\end{aligned}$$

For the same choice of χ we obtain from Eq. (3.51)

$$\langle R \rangle \approx \hat{r} \left(\hat{\kappa} \langle a \rangle - \hat{\kappa} \langle a \rangle \right) = 0 ,$$

so that (3.43) yields

$$\left[(\hat{\mathbb{E}} \otimes \hat{\mathbb{E}}) - \hat{\tau} (\hat{\mathbb{E}} \otimes q^{(1)} \hat{\kappa}) \right] K \approx \hat{\delta} \langle a \rangle + \hat{\tau} \left[\langle a \otimes \mathbb{1} \rangle \hat{r}_* \left(\frac{1}{\langle a \rangle \otimes \mathbb{1}} (\hat{\mathbb{E}} \otimes \hat{\kappa}) K \right) \right] , \tag{3.62}$$

where $q^{(1)}$ is given by (3.56).

We have thus found a second relation that allows us to derive approximative expressions for K. In difference to the first one, Eq. (3.54), it introduces an additional term so as to correct for the non-linearity of the neural activation function. On the other hand, it is this additional term which makes (3.62) difficult to handle. However, if the operator $[(\hat{\mathbb{E}} \otimes \hat{\mathbb{E}}) - \hat{\tau} (\hat{\mathbb{E}} \otimes q^{(1)} \hat{\kappa})]$ is invertible, then we can proceed writing

$$K \approx \left[(\hat{\mathbb{E}} \otimes \hat{\mathbb{E}}) - \hat{\tau} (\hat{\mathbb{E}} \otimes q^{(1)} \hat{\kappa}) \right]^{-1} \left\{ \hat{\delta} \langle a \rangle + \hat{\tau} \left[\langle a \otimes \mathbb{1} \rangle \hat{r}_* \left(\frac{1}{\langle a \rangle \otimes \mathbb{1}} (\hat{\mathbb{E}} \otimes \hat{\kappa}) K \right) \right] \right\}, \tag{3.63}$$

and we may hope to obtain useful estimates of K by iterating this last equation, starting from

$$K^{(0)} := \left[(\hat{\mathbb{E}} \otimes \hat{\mathbb{E}}) - \hat{\tau} \hat{\mathbb{E}} \otimes (q^{(1)} \hat{\kappa}) \right]^{-1} \hat{\delta} \langle a \rangle . \tag{3.64}$$

3.4.6 External Input

The methods we have presented so far in this section are well suited to analyze the spike statistics in closed recurrent neural networks that do not receive any kind of structured external input. Most models of neural systems, however, are designed to perform some sort of data processing and thus do receive input from external sources. Our previously developed techniques should therefore be extended accordingly.

For this purpose we allow an external voltage h^{ext} to be added to each neuron's membrane potential. Hence, the total membrane potential of cell number i is the sum of the synaptic contributions that it receives from within the network plus its refractory potential plus the external voltage h_i^{ext}, which is to account for all kinds of inputs that the neuron may receive from outside the network. Instead of Eq. (3.4) we thus have

$$\begin{aligned} h_i(t) &= \sum_j J_{ij} \int_{-\infty}^{t} a_j(t')\, \varepsilon(t-t')\, \mathrm{d}t' + \int_{-\infty}^{t} a_i(t')\, \eta(t-t')\, \mathrm{d}t' + h_i^{\mathrm{ext}}(t) \\ &=: [\hat{\kappa}\, a]_i(t) + h_i^{\mathrm{ext}}(t)\,, \end{aligned} \tag{3.65}$$

Most of the calculations that are necessary to extend our previous results are fully analogous to those we have shown in the preceding subsections so that we will present only a few major steps in the following. We start from the relations (3.28), (3.30), and

$$\left\langle a \otimes h^{\mathrm{ext}} \right\rangle = \left\langle (\hat{q} h) \otimes h^{\mathrm{ext}} \right\rangle\,. \tag{3.66}$$

In order to prove this last equation we write

$$\begin{aligned} \left\langle a_i(t)\, h_j^{\mathrm{ext}}(t') \right\rangle &= \int \mathrm{d}h \sum_{n=0}^{\infty} \frac{1}{n!} \int_0^T \mathrm{d}t_1 \cdots \int_0^T \mathrm{d}t_n \\ &\times \; a_i(t)\, h\, \mathrm{pr}\left[h_j^{\mathrm{ext}}(t') = h \wedge\; S_i^n(0, T, \{t_1, \ldots, t_n\}) \right] \end{aligned}$$

and then adapt (3.19) so as to obtain

$$\begin{aligned} \left\langle a_i(t)\, h_j^{\mathrm{ext}}(t') \right\rangle &= \int \mathrm{d}h\, h\, \mathrm{pr}\left[S_i(t) \wedge\; h_j^{\mathrm{ext}}(t') = h \right] \\ &= \int \mathrm{d}h \int \mathrm{d}h'\, h\, \mathrm{pr}\left[S_i(t) \wedge\; h_j^{\mathrm{ext}}(t') = h \wedge h_i(t) = h' \right] \\ &= \int \mathrm{d}h \int \mathrm{d}h'\, h\, \mathrm{pr}\left[S_i(t) \mid h_j^{\mathrm{ext}}(t') = h \wedge h_i(t) = h' \right] \\ &\times \; \mathrm{pr}\left[h_j^{\mathrm{ext}}(t') = h \wedge h_i(t) = h' \right]\,. \end{aligned}$$

As the probability density of neuron i emitting a spike at time t is given by its membrane potential $h_i(t)$ and is not related to the value of $h_j^{\mathrm{ext}}(t')$, we have

$$\mathrm{pr}\left[S_i(t) \mid h_j^{\mathrm{ext}}(t') = h \wedge h_i(t) = h' \right] = \mathrm{pr}\left[S_i(t) \mid h_i(t) = h' \right] = q_i(h')\,,$$

From that we find

$$\left\langle a_i(t)\, h_j^{\mathrm{ext}}(t') \right\rangle = \int \mathrm{d}h \int \mathrm{d}h'\, h\, q_i(h')\, \mathrm{pr}\left[h_i(t) = h' \wedge h_j^{\mathrm{ext}}(t') = h \right]\,,$$

which yields (3.66).

Expanding the neural activation function $q_i(h)$ in (3.28), (3.30), and (3.66) according to (3.35) we obtain

$$\langle a \rangle = q^{(0)} + q^{(1)} \left(\langle h \rangle - \chi \right) + \left\langle \hat{r}(h - \chi) \right\rangle\,, \tag{3.67}$$

$$\begin{aligned}
\langle a \otimes a \rangle &= \langle a \rangle \otimes \langle a \rangle + \hat{\delta}\,\langle a \rangle + \hat{\tau}\,\left[\left\langle a \otimes (q^{(1)} h)\right\rangle - \langle a \rangle \otimes \left\langle q^{(1)} h \right\rangle\right] \\
&+ \hat{\tau}\,\langle a \otimes \hat{r}(h-\chi)\rangle - \hat{\tau}\,[\langle a \rangle \otimes \langle \hat{r}(h-\chi)\rangle]\;, \qquad (3.68) \\
\left\langle a \otimes h^{\mathrm{ext}} \right\rangle &= \langle a \rangle \otimes \left\langle h^{\mathrm{ext}} \right\rangle + \left\langle (q^{(1)} h) \otimes h^{\mathrm{ext}} \right\rangle - q^{(1)}\,\langle h \rangle \otimes \left\langle h^{\mathrm{ext}} \right\rangle \\
&+ \left\langle [\hat{r}\,(h-\chi)] \otimes h^{\mathrm{ext}} \right\rangle - \langle \hat{r}\,(h-\chi)\rangle \otimes \left\langle h^{\mathrm{ext}} \right\rangle\;, \qquad (3.69)
\end{aligned}$$

where (3.67) has been used to eliminate $q^{(0)}$ from Eqs. (3.68) and (3.69). Next, by virtue of (3.65), we can express h in terms of $\hat{\kappa} a$ and h^{ext}. After defining the functions

$$\begin{aligned}
K &:= \langle a \otimes a \rangle - \langle a \rangle \otimes \langle a \rangle\;, \\
K^{\mathrm{ext}} &:= \left\langle a \otimes (q^{(1)} h^{\mathrm{ext}}) \right\rangle - \langle a \rangle \otimes \left\langle q^{(1)} h^{\mathrm{ext}} \right\rangle\;,
\end{aligned}$$

and

$$R' := \hat{r}\left(\hat{\kappa} a + h^{\mathrm{ext}} - \chi\right)\;,$$

we thus get

$$\begin{aligned}
\langle a \rangle &= q^{(0)} + q^{(1)}\left(\langle \hat{\kappa} a \rangle + \left\langle h^{\mathrm{ext}} \right\rangle - \chi\right) + \left\langle R' \right\rangle\;, \qquad (3.70) \\
\left[\hat{\mathbb{E}} \otimes \hat{\mathbb{E}} - \hat{\tau}(\hat{\mathbb{E}} \otimes q^{(1)} \hat{\kappa})\right] K &= \hat{\delta}\,\langle a \rangle + \hat{\tau} K^{\mathrm{ext}} \\
&+ \hat{\tau}\left\langle a \otimes R' \right\rangle - \hat{\tau} \langle a \rangle \otimes \left\langle R' \right\rangle\;, \qquad (3.71) \\
\left[\hat{\mathbb{E}} \otimes \hat{\mathbb{E}} - (q^{(1)} \hat{\kappa}) \otimes \hat{\mathbb{E}}\right] K^{\mathrm{ext}} &= (q^{(1)} \otimes q^{(1)})\left(\left\langle h^{\mathrm{ext}} \otimes h^{\mathrm{ext}} \right\rangle - \left\langle h^{\mathrm{ext}} \right\rangle \otimes \left\langle h^{\mathrm{ext}} \right\rangle\right) \\
&+ \left\langle R' \otimes (q^{(1)} h^{\mathrm{ext}}) \right\rangle - \left\langle R' \right\rangle \otimes \left\langle q^{(1)} h^{\mathrm{ext}} \right\rangle\;. \qquad (3.72)
\end{aligned}$$

These expressions generalize Eqs. (3.42) and (3.43) so as to account for an external membrane potential h^{ext}. They will now be evaluated following our explanations in the preceding subsections. We shall presume that the statistics of h^{ext} and, especially, its two first moments $\left\langle h^{\mathrm{ext}} \right\rangle$ and $\left\langle h^{\mathrm{ext}} \otimes h^{\mathrm{ext}} \right\rangle$ are known, because they describe the properties of the external input that is provided to the network.

Linear Neurons

If the neural activation function is given by a linear function

$$q_i(h) = \varrho_i^{(0)} + \varrho_i^{(1)} h\;,$$

then we can equate $q^{(0)}$ and $q^{(1)}$ with $\varrho^{(0)}$ and $\varrho^{(1)}$, respectively. This implies $\chi \equiv 0$ and $\hat{r} h \equiv 0$ so that Eqs. (3.70) through (3.72) reduce to

$$\begin{aligned}
\left[\hat{\mathbb{E}} - \varrho^{(1)} \hat{\kappa}\right] \langle a \rangle &= \varrho^{(0)} + \varrho^{(1)} \left\langle h^{\mathrm{ext}} \right\rangle\;, \\
\left[\hat{\mathbb{E}} \otimes \hat{\mathbb{E}} - \hat{\tau}(\hat{\mathbb{E}} \otimes \varrho^{(1)} \hat{\kappa})\right] K &= \hat{\delta}\,\langle a \rangle + \hat{\tau} K^{\mathrm{ext}}\;, \\
\left[\hat{\mathbb{E}} \otimes \hat{\mathbb{E}} - (\varrho^{(1)} \hat{\kappa}) \otimes \hat{\mathbb{E}}\right] K^{\mathrm{ext}} &= (\varrho^{(1)} \otimes \varrho^{(1)})\left(\left\langle h^{\mathrm{ext}} \otimes h^{\mathrm{ext}} \right\rangle - \left\langle h^{\mathrm{ext}} \right\rangle \otimes \left\langle h^{\mathrm{ext}} \right\rangle\right)\;.
\end{aligned}$$

Provided the operators on the left hand sides can be inverted we therefore find

$$\begin{aligned}
\langle a \rangle &= \left[\hat{\mathbb{E}} - \varrho^{(1)}\hat{\kappa}\right]^{-1} \left(\varrho^{(0)} + \varrho^{(1)} \left\langle h^{\mathrm{ext}} \right\rangle\right) , && (3.73)\\
K &= \left[\hat{\mathbb{E}} \otimes \hat{\mathbb{E}} - \hat{\tau}(\hat{\mathbb{E}} \otimes \varrho^{(1)}\hat{\kappa})\right]^{-1} \left(\hat{\delta} \langle a \rangle + \hat{\tau} K^{\mathrm{ext}}\right) , && (3.74)\\
K^{\mathrm{ext}} &= \left[\hat{\mathbb{E}} \otimes \hat{\mathbb{E}} - (\varrho^{(1)}\hat{\kappa}) \otimes \hat{\mathbb{E}}\right]^{-1} \left[(\varrho^{(1)} \otimes \varrho^{(1)})\right.\\
&\times \left.\left(\left\langle h^{\mathrm{ext}} \otimes h^{\mathrm{ext}} \right\rangle - \left\langle h^{\mathrm{ext}} \right\rangle \otimes \left\langle h^{\mathrm{ext}} \right\rangle\right)\right] . && (3.75)
\end{aligned}$$

Non-Linear Approximation

In order to obtain estimates of those terms in (3.70), (3.71), and (3.72) that involve the nonlinear function R', let us again utilize the heuristic method of approximating the average of a non-linear function by taking the value of the function at the average of its argument. This yields

$$\begin{aligned}
\left\langle R' \right\rangle &\approx \hat{r}\left(\hat{\kappa}\langle a \rangle - \left\langle h^{\mathrm{ext}} \right\rangle - \chi\right) ,\\
\left\langle R' \otimes h^{\mathrm{ext}} \right\rangle &\approx \hat{r}\left(\hat{\kappa}\langle a \rangle - \left\langle h^{\mathrm{ext}} \right\rangle - \chi\right) \otimes h^{\mathrm{ext}} ,
\end{aligned}$$

so that (3.70) and (3.72) read

$$\begin{aligned}
\langle a \rangle &\approx \hat{q}\left(\hat{\kappa}\langle a \rangle + \left\langle h^{\mathrm{ext}} \right\rangle\right) , && (3.76)\\
K^{\mathrm{ext}} &\approx \left[\hat{\mathbb{E}} \otimes \hat{\mathbb{E}} - (\varrho^{(1)}\hat{\kappa}) \otimes \hat{\mathbb{E}}\right]^{-1} \left[(q^{(1)} \otimes q^{(1)})\right.\\
&\times \left.\left(\left\langle h^{\mathrm{ext}} \otimes h^{\mathrm{ext}} \right\rangle - \left\langle h^{\mathrm{ext}} \right\rangle \otimes \left\langle h^{\mathrm{ext}} \right\rangle\right)\right] . && (3.77)
\end{aligned}$$

Equation (3.76) is the analogon to (3.52). It is to be solved numerically for the functions $\langle a \rangle_i(t)$, from which we determine the mean membrane potential $\langle h \rangle = \hat{\kappa}\langle a \rangle + \left\langle h^{\mathrm{ext}} \right\rangle$. Then we let

$$\chi_i(t) := \langle h \rangle_i(t) = \left[\hat{\kappa}\langle a \rangle\right]_i(t) + \left\langle h^{\mathrm{ext}} \right\rangle_i(t)$$

and

$$q_i^{(1)}(t) := \left.\frac{\mathrm{d}q_i(h)}{\mathrm{d}h}\right|_{h=\chi_i(t)=[\hat{\kappa}\langle a \rangle]_i(t)+\langle h^{\mathrm{ext}} \rangle_i(t)} .$$

As to Eq. (3.71) we notice that $\langle a \otimes R' \rangle$ can be written in the form of a conditional expectation; see (3.57). Following (3.58) and (3.59) we replace the conditional expectation of r_j by the value of r_j taken at the conditional expectation of its argument. Substituting $\chi = \langle h \rangle$ and $h = \hat{\kappa}a + h^{\mathrm{ext}}$ we thus find

$$\left\langle a \otimes R' \right\rangle \approx \langle a \otimes \mathbb{1} \rangle \, \hat{r}_* \left\{ \frac{1}{\left\langle a \otimes q^{(1)} \right\rangle} \left[(\hat{\mathbb{E}} \otimes q^{(1)}\hat{\kappa})K + K^{\mathrm{ext}}\right] \right\} ,$$

with $\mathbb{1}$ and $\hat{r}_*$ as defined in (3.60) and (3.61). This result together with

$$\left\langle R' \right\rangle \approx \hat{r}\left(\hat{\kappa}\langle a \rangle + \left\langle h^{\mathrm{ext}} \right\rangle - \hat{\kappa}\langle a \rangle - \left\langle h^{\mathrm{ext}} \right\rangle\right) = 0$$

inserted into Eq. (3.71) yields

$$K \approx \left[\hat{\mathbb{E}} \otimes \hat{\mathbb{E}} - \hat{\tau}(\hat{\mathbb{E}} \otimes q^{(1)}\hat{\kappa})\right]^{-1} \left\{ \hat{\delta}\,\langle a \rangle + \hat{\tau} K^{\mathrm{ext}} + \hat{\tau}\left[\langle a \otimes \mathbb{1} \rangle\, \hat{r}_* \left(\frac{1}{\langle a \otimes q^{(1)} \rangle} \left((\hat{\mathbb{E}} \otimes q^{(1)}\hat{\kappa}) K + K^{\mathrm{ext}} \right) \right) \right] \right\}, \quad (3.78)$$

presuming that the inverse operator on the right hand side does exist. Equation (3.78) is the generalized version of (3.63). We propose it to be iterated starting from

$$K^{(0)} \approx \left[\hat{\mathbb{E}} \otimes \hat{\mathbb{E}} - \hat{\tau}(\hat{\mathbb{E}} \otimes q^{(1)}\hat{\kappa})\right]^{-1} \left[\hat{\delta}\,\langle a \rangle + \hat{\tau} K^{\mathrm{ext}}\right] .$$

3.5 Numerical Simulations

Throughout the previous section we have derived analytical approximations of the two-spike correlations in recurrent networks of spiking neurons. Subsequently, we will compare these estimates with results obtained from numerical simulations.

3.5.1 Numerical Simulation of an Ensemble of Networks

The network we will use for the numerical simulations is a very simple one. It consists of two mutually connected spike response-neurons as shown in Fig. 3.1; refer to Sect. 3.1 for a complete definition of the spike response model. Each neuron produces a random spike train according to the time course of its membrane potential. Every spike emitted by cell 1 evokes a transient change of its own membrane potential due to refractoriness and, in addition, induces a transient change of the membrane potential at cell 2 via the synaptic coupling J_{21}. Analogously, every spike of neuron 2 induces a refractory contribution to its own membrane potential as well as a synaptic contribution at neuron 1 due to the coupling J_{12}. Formally, the refractory contributions can be considered as resulting from a sort of self-coupling, which we have denoted by two connections labeled η in Fig. 3.1.

The simulation algorithm is as follows. For a given set of network parameters, an ensemble of $4\,000\,000$ networks is constructed. We start at time $t = 0$ from the condition that no spike has occurred at $t < 0$, i.e., initially all the membrane potentials and all the integrators are set to zero.

At the beginning of each simulated time step, every cell's firing probability is determined from its membrane potential. We use a sigmoidal activation function

$$\Pr\{S_i(t, \Delta t) \mid h_i(t) = h\} := \{1 + \exp[-(h - \theta)/T]\}^{-1} \approx q_i[h]\,\Delta t\,,$$

where $\Delta t = 1$ ms is the size of the time step, $\theta = 0.5$ denotes the neural threshold, and $T = 0.5$ is a noise parameter. From the obtained firing probabilities it

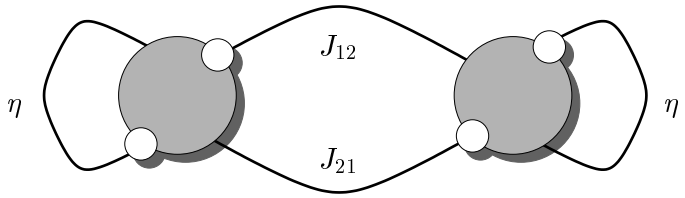


Figure 3.1: The network we use to compare our analytical results with numerical simulations consists of two mutually connected spike response-neurons. Both neurons produce random spike trains according to the time course of their membrane potentials. Every spike emitted by cell 1 evokes a transient change of the membrane potential at cell 1 due to refractoriness and in addition induces a transient change of the membrane potential at cell 2 via the synaptic coupling J_{21}. Analogously, every spike of neuron 2 induces a refractory contribution to the membrane potential of cell 2 as well as a synaptic contribution at neuron 1 due to the coupling J_{12}. Formally, the refractory contributions can be considered as resulting from a sort of self-coupling, which we have denoted by two connections labeled η.

is then decided randomly for each neuron whether a spike is emitted during the current time step.

Next, the membrane potential of each cell is updated according to the incoming spikes, which originate from its own activity and that of the second neuron within the same network. The transient response of a cell's membrane potential to a spike occurring at the other neuron is given by the corresponding synaptic weight J_{ij} times the postsynaptic potential $\varepsilon(t)$. We have chosen $\varepsilon(t)$ to be a decaying exponential,

$$\varepsilon(t) = \begin{cases} \exp(-t/\tau_\varepsilon)\,, & \text{if } t > 0\,, \\ 0 \ \text{ otherwise,} \end{cases}$$

with $\tau_\varepsilon = 6$ ms. Similarly, the refractory potential $\eta(t)$ models the response of a neuron to its own spikes. It is given by

$$\eta(t) = \begin{cases} \eta_0 \, \exp(-t/\tau_\eta)\,, & \text{if } t > 0\,, \\ 0 \ \text{ otherwise,} \end{cases}$$

with $\tau_\eta = 10$ ms. The functions $\varepsilon(t)$ and $\eta(t)$ are plotted in Fig. 3.2.

Initially, we repeat the above procedure for 200 time steps, after which we assume the the ensemble to have relaxed into a stationary state. Then start measuring the average firing rates and the correlation functions.

To determine the average firing rates $\langle a_1(t_1)\rangle$ and $\langle a_2(t_1)\rangle$ at a given time t_1 we count the number of networks in which a spike occurs at during the corresponding time step at cell 1 or cell 2, respectively. Dividing these numbers by Δt and by the total number of networks in the simulated ensemble yields estimates of the required average firing rates. We do this for 600 different

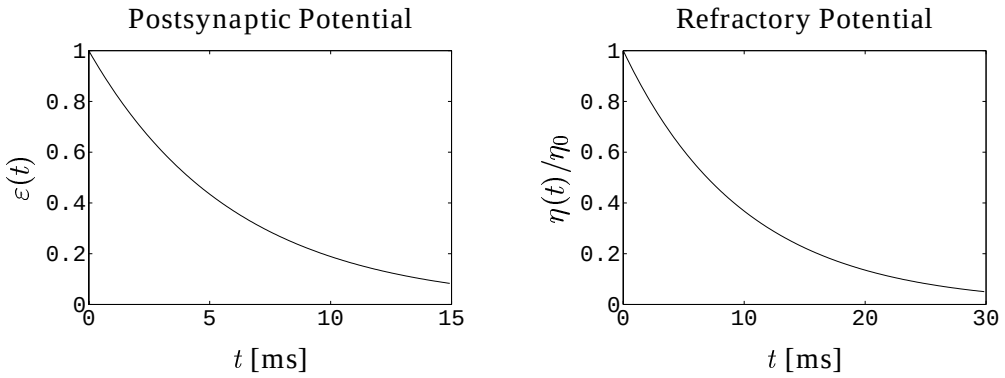


Figure 3.2: Both the postsynaptic potential $\varepsilon(t)$ and the refractory potential $\eta(t)$ have been chosen to be decaying exponentials with corresponding time constants $\tau_\varepsilon = 6$ ms and $\tau_\eta = 10$ ms.

values of t_1, viz.,

$$t_1 \in \{200\,\mathrm{ms},\ 202\,\mathrm{ms},\ 204\,\mathrm{ms}, \ldots, 1390\,\mathrm{ms}\}\ . \tag{3.79}$$

In an analogous manner approximations of the second moments $\langle a_i(t_1)\, a_j(t_2)\rangle$ of the networks' spike statistics are derived. With $i, j = 1, 2$ and for given times t_1 and t_2 we count the number of networks that emit a spike at time t_2 at neuron j *and* have emitted a spike at time t_1 at neuron i. We divide this number by Δt^2 and by the total number of networks so as to obtain an estimate of the probability density $\mathrm{pr}\,[S_i(t_1) \wedge S_j(t_2)]$, which is equal to $\langle a_i(t_1)\, a_j(t_2)\rangle$ by Eq. (3.11). Following this procedure we determine $\langle a_i(t_1)\, a_j(t_2)\rangle$ for the 600 different values of t_1 given in (3.79) and for t_2 varying from $t_1 + 1$ ms to $t_1 + 70$ ms in steps of 1 ms. Using definition (3.27) together with the measured values of the mean firing rates yields 600 traces of the correlation functions $C_{ij}(t_1, t_2)$ for consecutive values of t_1 with $(t_2 - t_1)$ ranging from 1 to 70 ms.

3.5.2 Numerical versus Analytical Results

We have performed numerical simulations of network ensembles as described in the previous subsection for the five different sets of parameters given in Table 3.1. In all these sets we have set $J_{12} = J_{21}$, so that the two cells in each network are functionally equivalent. As the initial conditions of our simulations are symmetric with respect to cell 1 and cell 2, the spike statistics in the ensemble must therefore be identical for both neurons. In the sequel we will thus restrict our analyses to only one mean firing rate, $\langle a_1(t_1)\rangle$ say, and to the two correlation functions C_{11} and C_{12}, because $C_{11} = C_{22}$ and $C_{12} = C_{21}$. Since by definition of the correlation functions it is $C_{11}(t_1, t_2) = C_{11}(t_2, t_1)$ and $C_{12}(t_1, t_2) = C_{21}(t_2, t_1)$, we need only consider $C_{11}(t_1, t_2)$ and $C_{12}(t_1, t_2)$ with $t_2 \geq t_1$.

If our assumption is correct that the simulated ensemble has reached a stationary state at the time when the measurements are started, then the derived firing rate $\langle a_1(t_1)\rangle$ must turn out to be independent of t_1 and, similarly, the

#	η_0	J_{12}	J_{21}
1	0.5	0.2	0.2
2	1	0.5	0.5
3	2	0.5	0.5
4	2	1	1
5	5	1	1

Table 3.1: Numerical simulations of ensembles of neural networks have been performed for five different sets of parameters. The neural threshold $\theta = 3$, the noise parameter $T = 0.5$, and the decay time constants $\tau_\eta = 10$ ms, $\tau_\varepsilon = 6$ ms of refractory and postsynaptic potentials were held constant, whereas the synaptic weights J_{12}, J_{21} and the amplitude η_0 of the refractory potential have been varied.

correlation functions $C_{ij}(t_1, t_2)$ should depend on the difference $t_2 - t_1$ only. We test these conditions by plotting $\langle a_1(t_1)\rangle$ against t_1 and $C_{ij}(t_1, t_2)$ against t_1 and $(t_2 - t_1)$. Figure 3.3 shows the corresponding graphs as obtained from a simulation using parameter set #1.

In order to reduce the scatter in these plots we do not display every single measured value but, instead, average our data over bins of 20 consecutive measurements along the t_1-axis. The graph in panel (a) displaying $\langle a_1(t_1)\rangle$ as a function of t_1 thus consists of 30 data points, each of which represents the average over 20 consecutively measured values. To be explicit, the ordinate a_n of the n^{th} data point has been calculated according to the formula

$$a_n = \frac{1}{20}\sum_{k=0}^{19} \langle a_1(t_1 = 200\,\mathrm{ms} + k \times 2\,\mathrm{ms} + n \times 40\,\mathrm{ms})\rangle \ , \quad n \in 0, 1, \ldots 29 \ .$$

In analogy, each point in the three-dimensional plots of C_{11} and C_{12} shown in panels (b) and (c), respectively, has been obtained from averaging over 20 consecutive measurements along the t_1-axis.

The figure reveals that for $t_1 > 200$ ms both the mean firing rate and the correlation functions are – within numerical accuracy – independent of t_1. This finding supports our presumption that the ensemble of networks has reached an almost stationary state after the initial 200 simulated time steps. We can exploit this fact by averaging our data along the t_1-axis so as to obtain more accurate estimates of the mean firing rate and the correlation functions. The firing rate as derived from this approach is 2.44 Hz; the correlation functions are presented in Fig. 3.4.

Let us now compare these results to the corresponding quantities as obtained from our theory presented in Sect. 3.4. We begin by calculating the average firing rates by numerically solving Eq. (3.52) under the assumption of stationarity. This yields

$$\langle a_1(t_1)\rangle = \langle a_2(t_1)\rangle \approx 2.43 \text{ Hz} \ ,$$

which is close to the value derived from the simulations. From this approximation of the firing rates we can now compute a first estimate of the correlation functions using Eqs. (3.54) and (3.56) together with the definitions (3.27) and (3.40). Since we do not have an exact expression for the required inverse of

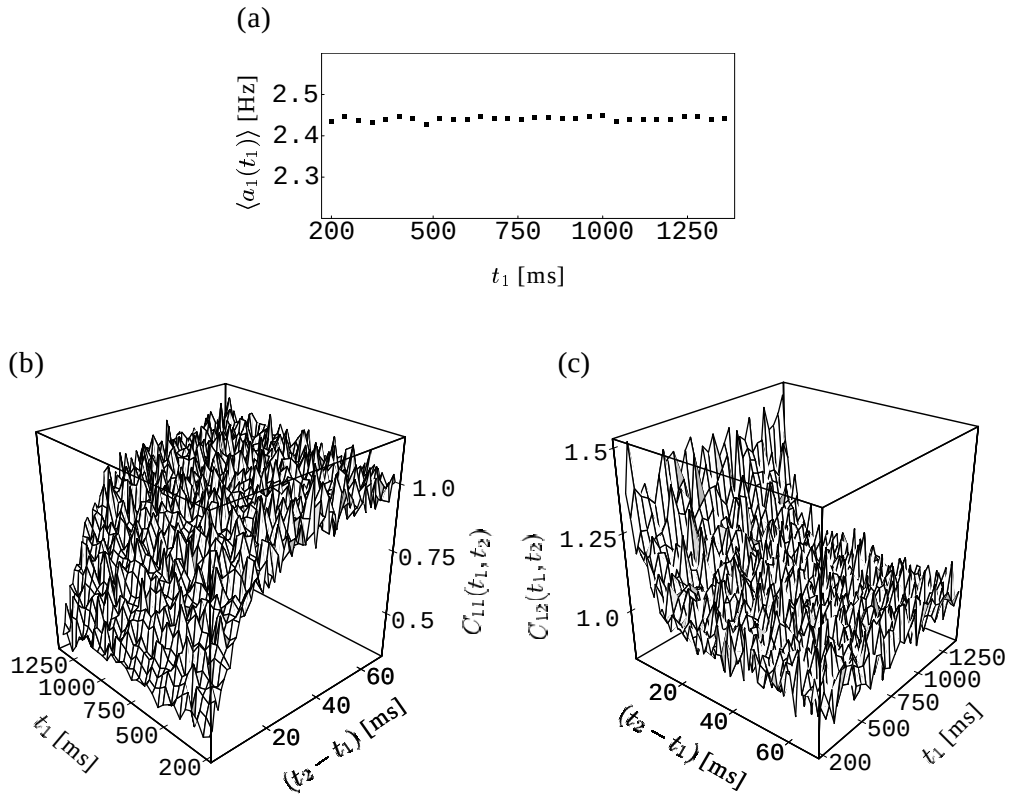


Figure 3.3: The average firing rate and the correlation functions are displayed as they have been obtained from a numerical simulation of an ensemble of networks using the network topology presented in Fig. 3.1 and the parameter set #1 given in Table 3.1. In panel (a) the average firing rate $\langle a_1(t_1)\rangle$ is shown as a function of t_1. Panels (b) and (c) visualize the correlation functions $C_{11}(t_1,t_2)$ and $C_{12}(t_1,t_2)$ in dependence of t_1 and (t_2-t_1). The plots reveal that for $t_1 > 200$ ms both the mean firing rate and the correlation functions are – within the numerical accuracy – independent of t_1. This indicates that the simulated ensemble has reached an almost stationary state after the initial 200 time steps.

the operator on the right hand side, we use the von-Neumann series up to the second order,

$$\left[(\hat{\mathbb{E}}\otimes\hat{\mathbb{E}}) - \hat{\tau}(\hat{\mathbb{E}}\otimes\varrho^{(1)}\hat{\kappa})\right]^{-1} \approx \hat{\mathbb{E}}\otimes\hat{\mathbb{E}} + \hat{\tau}(\hat{\mathbb{E}}\otimes\varrho^{(1)}\hat{\kappa}) + \left[\hat{\tau}(\hat{\mathbb{E}}\otimes\varrho^{(1)}\hat{\kappa})\right]^2 , \quad (3.80)$$

taking for granted that it provides us with a helpful approximation. The functions $C_{11}(t_1,t_2)$ and $C_{12}(t_1,t_2)$ calculated in this way depend only on the time difference (t_2-t_1). They are plotted as solid lines in Fig. 3.5. For comparison the numerically-derived curves from Fig. 3.4 have also been included into the graphs.

As can be seen from the two panels, the analytical estimates are systematically to low in those regions where the values of the correlation functions are significantly greater or lower than 1. This can be explained in the following way. Equation (3.54) is essentially identical to Eq. (3.47), which has been

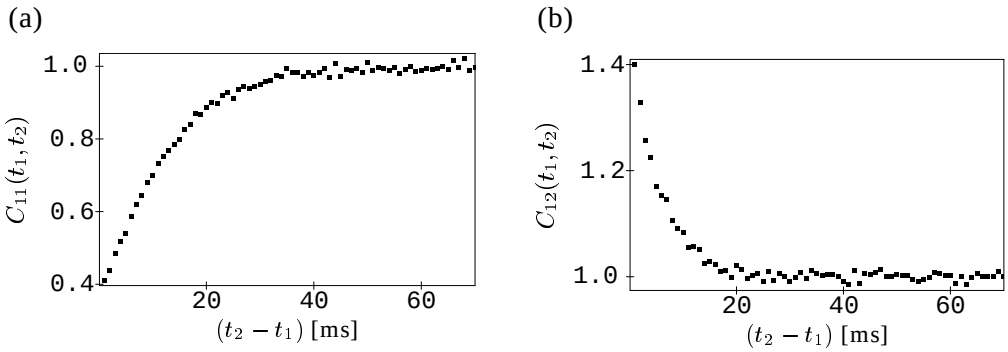


Figure 3.4: Since according to Fig. 3.3 the mean firing rate and the two-spike correlation functions are stationary, we can take their averages over time, i.e., we collapse the corresponding graphs along their t_1-axes. In this way we obtain more accurate estimates of the respective quantities. The resulting correlations $C_{11}(t_1, t_2)$ and $C_{12}(t_1, t_2)$ are plotted in panels (a) and (b), respectively; the value of the mean firing rate is found to be $\langle a_1(t_1)\rangle \approx 2.44$ Hz.

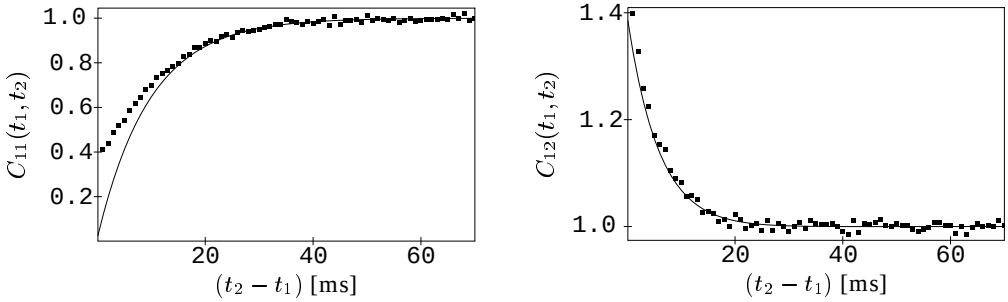


Figure 3.5: The solid curves display the correlation functions C_{11} and C_{12} as derived analytically for parameter set #1 (Table 3.1) using Eqs. (3.54) and (3.56). For comparison the results of the numerical simulation have also been included; cf. Fig. 3.4. The discrepancies can be explained by the fact that Eq. (3.54) does not account for the non-linearity of the neuronal activation function (see text and Fig. 3.6). As demonstrated in Fig. 3.7, a more elaborate analytical method can approximately correct for this error.

derived by linearizing the neural activation functions $q_i(h)$, and does therefore not account for the nonlinearity of $q_i(h)$. For this reason, the correlation function obtained from (3.54) is the correlation function arising in a network in which each neuron's activation function has been linearized about the expectation value of the corresponding membrane potential.

For the presently considered ensemble of networks with parameter set #1 the resulting linearization is identical for cell 1 and cell 2. It is displayed in Fig. 3.6 together with the true non-linear activation function. Obviously, in a wide range around the mean membrane potential, the value of the linearized activation function is lower than that of the real one. Only at a very large value of the membrane potential will the straight line cross the sigmoidal curve of the true activation function. The firing probability as derived from the lin-

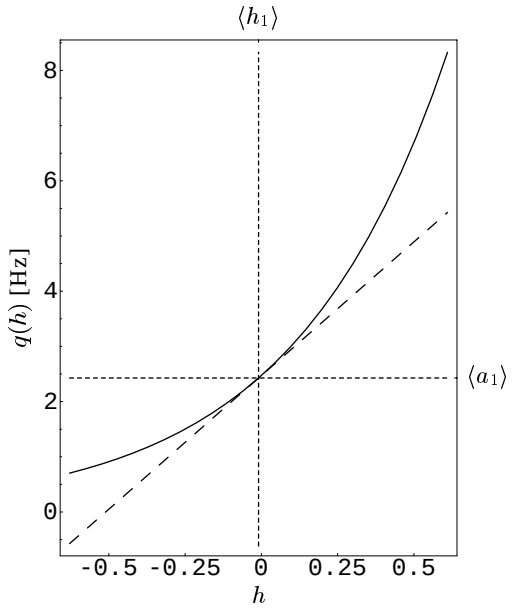


Figure 3.6: The neural activation function used in the network ensemble under consideration (parameter set $\#1$, Table 3.1) is plotted as a solid curve, whereas its linearization is shown as a dashed line. The dotted lines indicate the values of the mean membrane potential $\langle h_1 \rangle = \langle h_2 \rangle$ and the average firing rate $\langle a_1 \rangle = \langle a_2 \rangle$. Since the value of the linearized activation function is systematically lower than that of the real one, the linear estimates of the correlation functions shown in Fig. 3.5 are to low at small inter-spike time intervals.

earization therefore underestimates the true value, and this effect is the more pronounced the more the firing probability deviates from its average. As a consequence, Eq. (3.54) yields approximations of the correlations that are especially low in those regions where they differ significantly from a value of 1.

These deviations can in part be corrected by deriving the correlation functions from Eq. (3.63) instead of (3.54), because it accounts for the non-linearity of the activation function, although it is only an approximation either. Since the required quantity K is involved on both sides of (3.63) in a way that prevents us from solving for K, we will try to iterate this equation. We start from $K^{(0)}$ given by (3.64), which is just the solution of the linearized problem discussed so far. After one step of iteration we obtain a function $K^{(1)}$ that is equal to $K^{(0)}$ plus some correction for the non-linearity of the neural activation function.

From $K^{(1)}$ we can derive corrected estimates of the correlation functions C_{11} and C_{12}. These are plotted in Fig. 3.7 as solid lines together with the results of the numerical simulation, which have been taken from Fig. 3.4. The figure reveals that our non-linear analytical approximation yields very accurate estimates of the true correlation functions arising in the network ensemble for parameter set $\#1$.

In analogy to the above explanations let us now compare the analytical method to the numerical simulations for the four remaining parameter sets given in Table 3.1.

As a first step, we check the numerical results for stationarity of the network ensemble by plotting the mean firing rate and the correlation functions against the elapsed time t_1 as we did in Fig. 3.3. We do not display the resulting graphs because they are qualitatively similar to those of Fig. 3.3 and thus indicate that for all parameter sets the simulated ensemble has reached a stationary state after 200 time steps. In order to obtain good numerical estimates

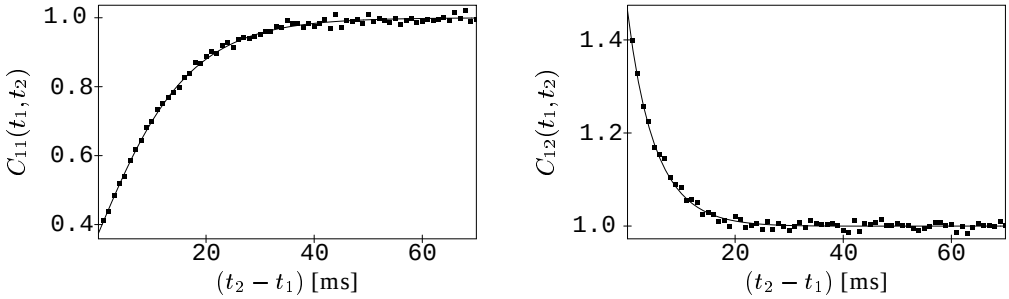


Figure 3.7: The solid lines display the correlation functions C_{11} and C_{12} as derived analytically for parameter set $\#1$ (Table 3.1) using one iteration of Eq. (3.63). Obviously, the outcome is in very good agreement with the results of the numerical simulation, which are given by the dotted curves; cf. Fig. 3.4.

#	$\langle a_1 \rangle$ from simulation	$\langle a_1 \rangle$ from theory
1	2.44 Hz	2.43 Hz
2	2.44 Hz	2.39 Hz
3	2.40 Hz	2.29 Hz
4	2.46 Hz	2.32 Hz
5	2.40 Hz	2.06 Hz

Table 3.2: For each parameter set the mean firing rate $\langle a_1 \rangle$ as obtained from a numerical simulation of the network ensemble is presented together with the corresponding theoretically-derived value. The numbers in the first column correspond to those of Table 3.1. Due to the symmetry of the parameter sets and the initial conditions it is always $\langle a_1 \rangle = \langle a_2 \rangle$, so that only $\langle a_1 \rangle$ is listed.

of the mean firing rate and the correlations, we may therefore average over the individual measurements taken at different times during one simulation.

For each parameter set, Table 3.2 presents the value of the mean firing rate as derived from the simulation as well as the corresponding value found by numerically solving Eq. (3.52). It is clearly visible that the difference between these two values is smallest for parameter set $\#1$ and largest for $\#5$. The reason is that from parameter set to parameter set in Table 3.1 we have increased either the synaptic weights J_{12} and J_{21} or the amplitude η_0 of the refractory potential or both. The larger these quantities are the more does a cell's membrane potential vary in response to a single spike in the network, which gives rise to a larger variance of the membrane potentials in an ensemble of networks. As a consequence of this increasing variance, the approximation of setting $\langle q(h) \rangle \approx q(\langle h \rangle)$ whence we have derived Eq. (3.52) becomes less accurate. The average firing rate as obtained from this approximation thus deviates more and more from its numerically-derived 'true' value.

Next, we compare the auto-correlations C_{11} and the cross-correlations C_{12} for parameter sets $\#2, \ldots, \#5$ as determined from the numerical simulations (black squares) with those obtained from the linearized theory (solid curves), i.e., from Eq. (3.54). Each row of Fig. 3.8 corresponds to one set of simulation parameters, the left and right panels showing C_{11} and C_{12}, respectively. As we have already found for parameter set $\#1$, the linear predictions of C_{11} and C_{12} deviate substantially from the true correlations. This is especially obvious

for the auto-correlations, which are predicted to be negative at low values of $(t_2 - t_1)$. According to Eqs. (3.11) and (3.27), however, $C_{11}(t_1, t_2)$ is a probability density divided by a positive quantity and must therefore be positive or zero.

Consistently with our results for parameter set #1, we find the discrepancies to be largely reduced, when the linear approximations are replaced by the non-linear estimates obtained from Eq. (3.63). This is shown in Fig. 3.9. For most of the parameter sets the theoretical curves are in a remarkably good agreement with the numerical data. Nevertheless, the bottom row of the figure also reveals that our analytical estimates quickly deteriorate, when the amplitude of the refractory potential η_0 is increased to much. This is due to the following two reasons.

First, the analytical estimates are based on the approximation (3.58), where the conditional expectation of a non-linear function r is replaced by the value of r applied to the conditional expectation of its argument. This approximation deteriorates for increasing variance of the function's argument, i.e., of the neurons' membrane potentials. An increased amplitude of the refractory potential thus affects the accuracy of the results because it induces a larger variance of the membrane potentials.

Second, causing the norm of the coupling operator $\hat{\kappa}$ to increase, a growing value of η_0 will slow down or even prevent the convergence of the von-Neumann series used in Eq. (3.80). Consequently, the second order expansion on the right hand side of (3.80) may not yield a good approximation of the inverse operator on the left hand side resulting in erroneous predictions of the correlation functions. This is, however, not a fundamental problem of the analytical methods, as it can be solved by calculating an exact or at least a more accurate expression for the required inverse operator. In the next subsection we demonstrate how to circumvent this difficulty in a very simple way that is useful in situations where the focus of interest lies on the cross-correlations rather than on the auto-correlations.

3.5.3 On the Role of the Refractory Potential

According to Table 3.2, the average firing rates in the network ensemble is about 2.5 Hz for the investigated sets of parameter values. This means that the duration of the inter-spike intervals at each cell is typically about 400 ms. This is much longer than the decay time constant of the refractory potential $\tau_\eta = 10$ ms. Intuitively one might therefore argue that neglecting the refractory potential will not have to much an effect on the mean firing rate and the cross-correlation function in the network ensemble. So why not simply neglect the refractory potential?

On the other hand, it is immediately clear that the time course of auto-correlation function must strongly depend on the properties of the refractory potential as it measures the probability of two subsequent spike events occurring at the same neuron. For short inter-spike time intervals this probability is obviously dominated by the neuron's refractoriness.

In order to investigate the influence that the refractory potential exerts on the mean firing rate and the correlation functions we analyze the results of

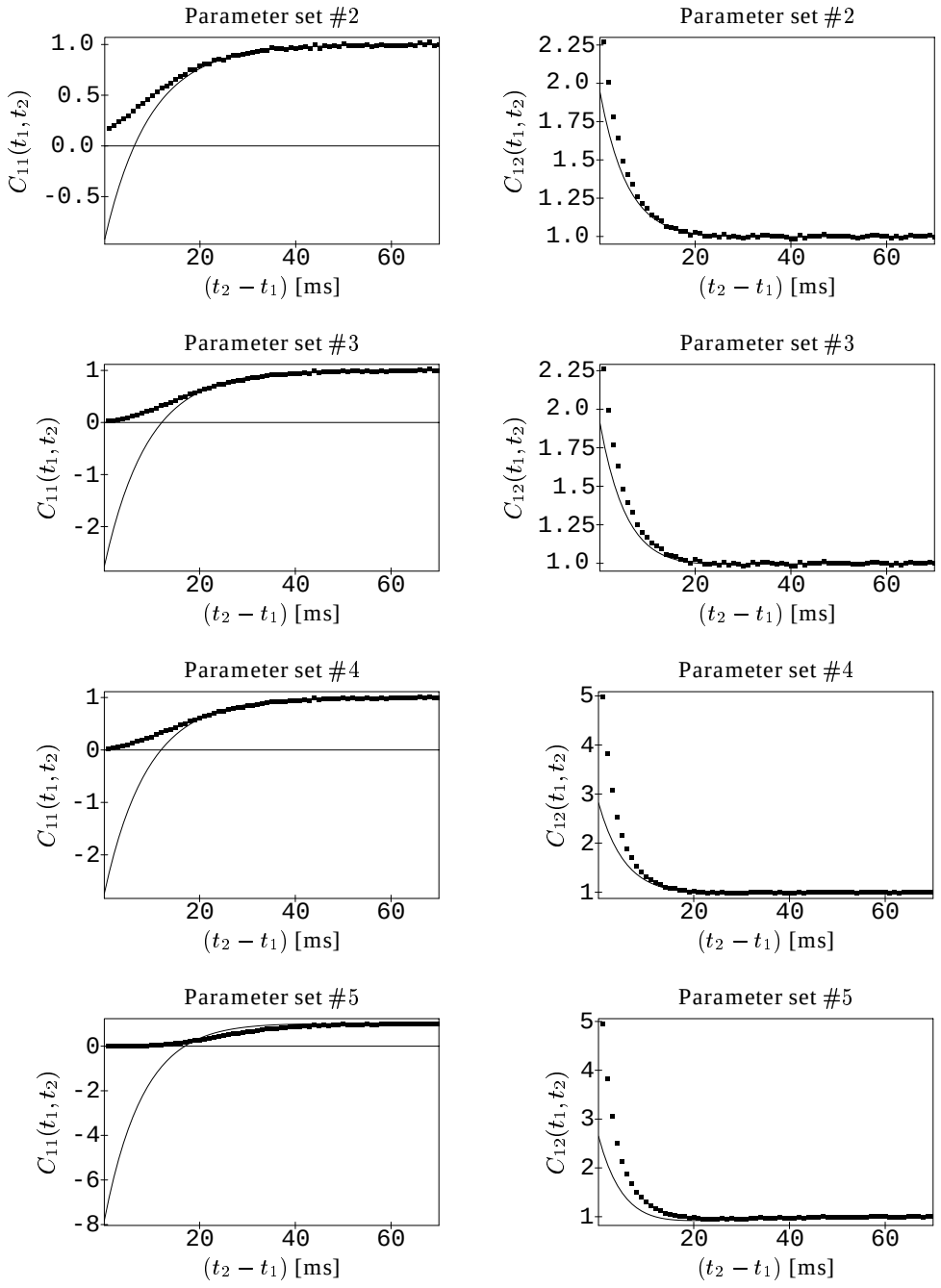


Figure 3.8: The correlation functions C_{11} and C_{12} as derived from the numerical simulations (black squares) and from the linearized theory (solid curves) are plotted against $(t_2 - t_1)$; cf. Fig. 3.5. Each row corresponds to one set of simulation parameters as indicated above the individual panels. The linear predictions deviate substantially from the true correlations, especially at low values of $(t_2 - t_1)$. To a great extent, these deviations are corrected by our non-linear approximation (3.63), as can be seen in Fig. 3.9.

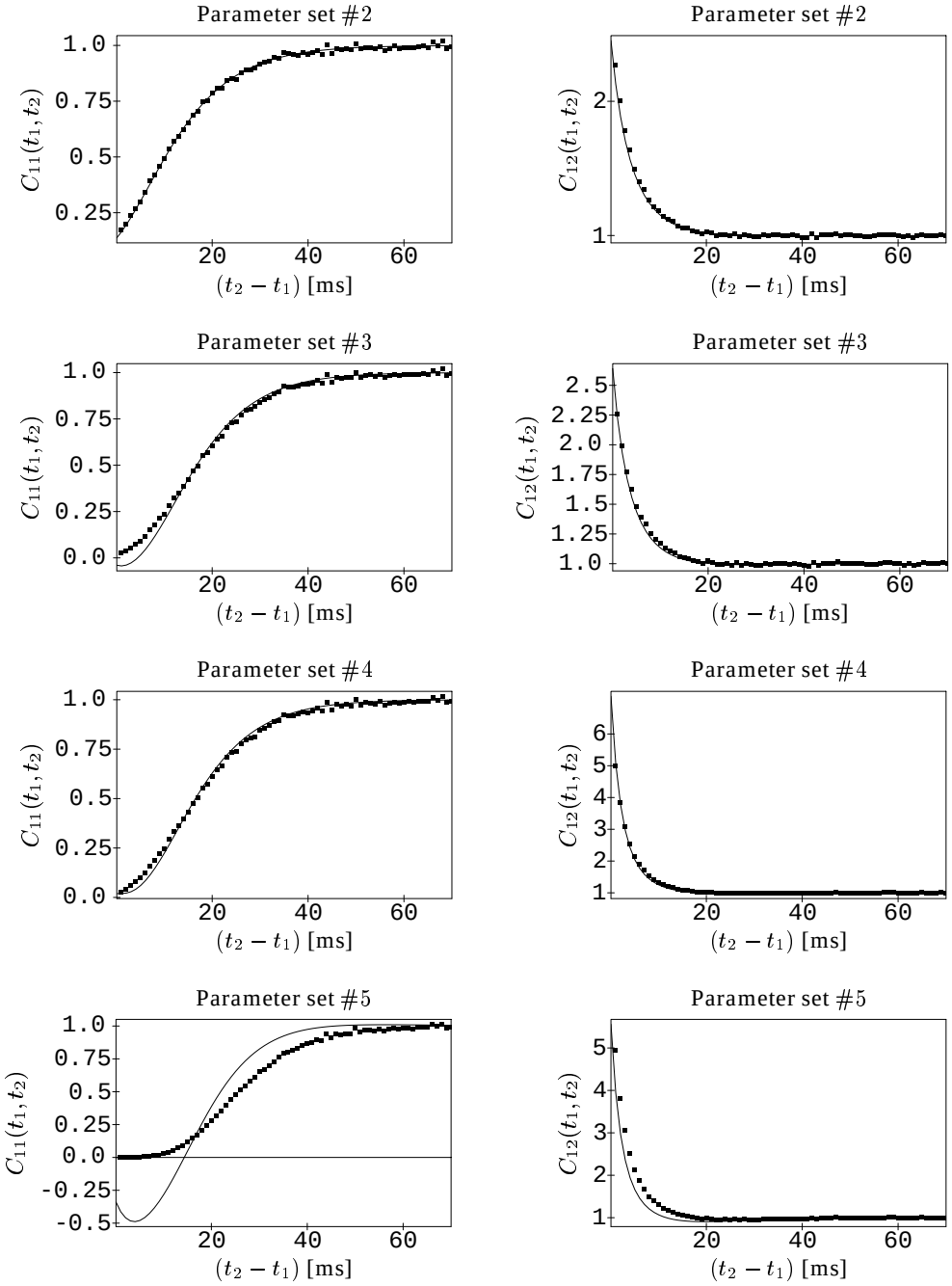


Figure 3.9: As Fig. 3.8 but with theoretical curves obtained from Eq. (3.63), which corrects for the non-linearity of the neural activation function. Except for parameter set #5 (see Table 3.1) the analytical predictions are in very good agreement with the numerical data.

#	η_0	J_{12}	J_{21}
1	5	1	1
2	10	1	1
3	0	1	1

Table 3.3: In order to investigate the influence of the refractory potential on the spike statistics in an ensemble of networks we perform three different numerical simulations in which we vary only the amplitude of the refractory potential. All the parameters not listed are held constant at their respective values given in Table 3.1.

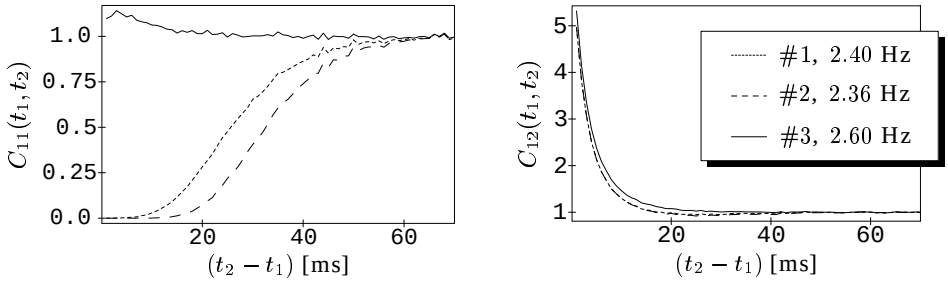


Figure 3.10: The auto- and cross-correlation functions obtained from numerical simulations using the parameter sets of Table 3.3 are compared in this figure. The legend is given as an inset of the right panel. Each parameter set is referred to by its respective number in Table 3.3. In addition the legend denotes for each parameter set the mean firing rate as determined from the simulations. It can be seen clearly that (i) the mean firing rate and the cross-correlations are only weakly affected by changing the amplitude η_0 of the refractory potential, whereas (ii) the auto-correlations are strongly dependent on η_0.

three different numerical simulations using the parameter values presented in Table 3.3. The first set of parameters is identical to set $\#5$ in Table 3.1. We have already seen in the previous subsection that due to the large contribution of the refractory field it is difficult to obtain good analytical predictions of the auto-correlations for this set of parameters. In the second parameter set the amplitude of the refractory potential is even larger. In Fig. 3.10 the results obtained from these two simulations are compared with those derived from a third simulation in which the refractory potential has been set to zero. In the figure legend, which is given as an inset of the right panel, each parameter set is referred to by its respective number in Table 3.3. In addition the legend denotes the mean firing rate as determined from the simulations.

Essentially the figure confirms what we have already suspected. Changing the amplitude η_0 of the refractory potential over a wide range of values has only a weak effect on the firing rate and the cross-correlations C_{12}, whereas it strongly affects the auto-correlation function $C_{11}(t_1, t_2)$ especially at low values of the inter-spike time $(t_2 - t_1)$. We thus conclude that in an investigation of neural dynamics where the main interest is on the cross-correlations rather than on the auto-correlations there may be no harm in dropping the refractory potential, provided its decay time is much shorter than the typical inter-spike interval.

This finding will prove to be very helpful in the following chapter, where we will present a mathematical analysis of the learning dynamics going on in the network model of primary visual cortex introduced in Chapter 2. As it will turn out, the dynamics of a synapse in this model can be expressed in terms of the spike statistics of its pre- and postsynaptic cells, namely their mean firing rates and the correlations of their activity. The auto-correlation function of a given cell's activity thus determines the plasticity of those synapses that this cell makes onto itself. In our network there is only one such autapse for each neuron as compared to a large number of incoming connections from other cells so that its contribution to the neuron's membrane potential will normally be negligible. The dynamics that we are interested in is therefore mainly determined by the crosscorrelation functions. In regard of the typical firing rates in the model being a few spikes per second only, this means that we may safely drop the refractory potential from our analysis so as to simplify the required calculations.

3.6 Summary

In the present chapter we have developed a mathematical method to derive analytical approximations of the two-spike correlation functions in ensembles of networks of stochastically spiking neurons. The underlying neuron model is the stochastic spike response model suggested by Gerstner and van Hemmen (1994) and summarized in Sect. 3.1. It is an extended version of the usual integrate-and-fire model that can be configured by means of two response kernels, viz., the postsynaptic potential and the refractory potential to resemble the characteristics of real spiking neurons (Kistler et al. 1997). Spikes are generated from an inhomogeneous Poisson process with a probability density given by some function of the neuron's membrane potential, which is called the activation function.

The concept of an ensemble of networks has been introduced in Sect. 3.2. We consider a network ensemble as an entirety of infinitely many different realizations or independent 'runs' of the same stochastically spiking neural network. Within this framework the probability of an event can be taken as the fraction of those realizations in which this event occurs. We have derived that the probability density of a neuron emitting a spike is equal to the ensemble average of its activity and, similarly, that the probability density of two spike events occurring at two given cells is given by the ensemble average of the product of the respective activities.

In Sect. 3.3 we then turned to the problem of deriving expressions for the input-output correlations of a single non-refractory neuron. In a first approach we have assumed that the neuron receives are large number of statistically independent Poissonian input spike trains so that the probability distribution of the membrane potential is approximately Gaussian. This presumption has been released in our second approach. From expanding the activation function into a power series we have obtained an approximative formula converting the the moments of the input spike statistics into the input-output two-spike corre-

lations. If necessary, the method can be generalized to higher-order correlation functions (cf. Kuznetsov et al. 1965).

In the main part of this chapter, Sect. 3.4, we have generalized our previously introduced methods so as to derive analytical expressions for the average firing rate and the two-spike correlation functions in networks of arbitrary architecture. First, we have treated the case of a linear network, i.e., a network consisting of neurons whose activation function is linear in the membrane potential. We were able to give exact equations for the required quantities in such a linear network. Then we have presented an extended method that allows to obtain analytical approximations of the firing rate and the correlation functions in networks of non-linear neurons. Finally, we have further generalized our results so as to account for external input that the network may be provided with.

We have then compared our theoretical predictions with results obtained from numerical simulations in Sect. 3.5. The simulations were carried out for an ensemble of networks where each network consisted of two mutually connected non-linear neurons. Corresponding theoretical results were derived using both the linear method and its non-linear extension. In order to apply the linear method each neuron's activation function was linearized about the estimated value of the average membrane potential.

It turned out that within the investigated parameter regimes the linear predictions of the two-spike correlations are systematically below the true values as obtained from the simulations. This is easily explained as an effect of the linearization. On the other hand, we have found that our non-linear theory yields remarkably good estimates of the true firing rate and the correlation functions except when the refractory potential is very large.

At last, we have shown in our simulations that the average firing rate and the cross-correlation functions are only weakly affected by modifications of the refractory potential, provided that the decay time of the refractory potential is much shorter than the typical inter-spike interval. Under this condition the refractory potential may therefore be negligible when the main interest of an investigation is on the cross-correlations rather than on the auto-correlations. This may prove to be especially helpful in theoretical analyses like the one we will present in the following chapter because dropping the refractory potential can significantly simplify the involved mathematical expressions.

Chapter 4

Analytical Investigation of the Learning Dynamics

In the previous chapter we have developed a technique that allows to derive analytical estimates of the spike-spike correlations in recurrent neuronal networks. Due to these considerations we are now in the situation to gain some analytical insight into the dynamics of Hebbian synaptic plasticity in such networks. We can thus obtain a mathematical understanding of the basic mechanisms that lead to the emergence of orientation maps in the model of primary visual cortex presented in Chapter 2.

Since the dynamics of the full model are rather intractable we will instead concentrate on a reduced version that will be described in Sect. 4.1. Within the framework of this reduced network, as in the full model, each neuronal spike train is a random process so that all synaptic weights subject to spike-based Hebbian plasticity are stochastic processes as well. In Sect. 4.2 a mathematical approach to the resulting synaptic dynamics will be presented. We will start from the learning rules set forth in Sect. 2.3 and derive a differential equation for the temporal evolution of a synapse's efficacy as a function of the pre- and postsynaptic spike statistics. By means of the methods provided in the preceding chapter the required statistics can be expressed in terms of the synaptic efficacies so as to arrive at a closed system of differential equations describing the synaptic dynamics.

In Sect. 4.3 we will apply this method and investigate (i) how a cortical orientation map can emerge from a Hebbian development of geniculocortical synapses and (ii) how the layout of the emerging map can be influenced by the structure of the intracortical connectivity. For the sake of simplicity the intracortical connectivity will be assumed to be fixed during the geniculocortical development. In Sect. 4.4 we will then analyze the Hebbian dynamics of intracortical synapses in the absence of feedforward input. Although the reduced network model turns out to be too simple as to account for the emergence of an intracortical orientation map, it will nevertheless be interesting to find out *why* it is too simple.

The chapter will be closed by a summary and a short discussion of the obtained results.

4.1 Reduced Network Model

The network model of layer 4 in primary visual cortex that we have proposed and investigated numerically in Chapter 2 consists of three distinct layers of stochastically spiking non-linear neurons. Its architecture is reproduced in the left panel of Fig. 4.1. The two upper layers are designed to represent cortical excitatory and inhibitory neurons. Excitatory cortical cells receive lateral input from neighbouring excitatory and inhibitory cortical cells. Inhibitory neurons are driven by the activity of excitatory neurons. Geniculate relay cells in the bottom layer provide correlated feedforward input to the cortex. The statistics of their spike trains are determined by the statistics of their membrane potentials, which are prescribed externally from a given Gaussian random field. All synaptic connections except for those from excitatory to inhibitory cells are subject to activity-driven learning. As compared to the development of excitatory synapses the plasticity of inhibitory connections proceeds on a short time scale and serves to normalize overall cortical activity.

Our large-scale numerical simulations have revealed that in this network it is possible to obtain an intracortical orientation map from a Hebbian development of excitatory intracortical synapses and a feedforward orientation map from Hebbian plasticity of geniculocortical connections. We have found that the intracortical map can guide the geniculocortical development so that the layout of the emerging geniculocortical map is in accord with the layout of the intracortical map.

The mathematical analysis that will be presented throughout the remaining sections of this chapter is intended to provide some insight into the underlying processes of pattern formation. The dynamics of the full model, however, turn out to be extremely involved so that we will study a simplified version instead. The architecture of this simplified network is shown in the right panel of Fig. 4.1. Being interested mainly in the development of excitatory intracortical and feedforward connectivity, we have dropped the layer of inhibitory cells. In addition to this reduction of the network's topology, we also simplify the representation of individual neurons by (i) replacing their non-linear activation function by a linear one and by (ii) neglecting the refractory potential.

Although the reduced model certainly oversimplifies the dynamics of the full network in many respects it will nevertheless be very useful for an analytical investigation. We will see that its synaptic dynamics are relatively easy to understand and yet capture important features of the full model. It will serve us to understand how a feedforward orientation map can emerge from Hebbian development of geniculocortical synapses and how the layout of this map can be predetermined by an existing intracortical connectivity pattern. On the other hand, we will also find an essential limitation of the simplified model: It does not provide a satisfactory explanation of the emergence of an intracortical orientation map in the absence of feedforward input that we have found in our simulations of the full model.

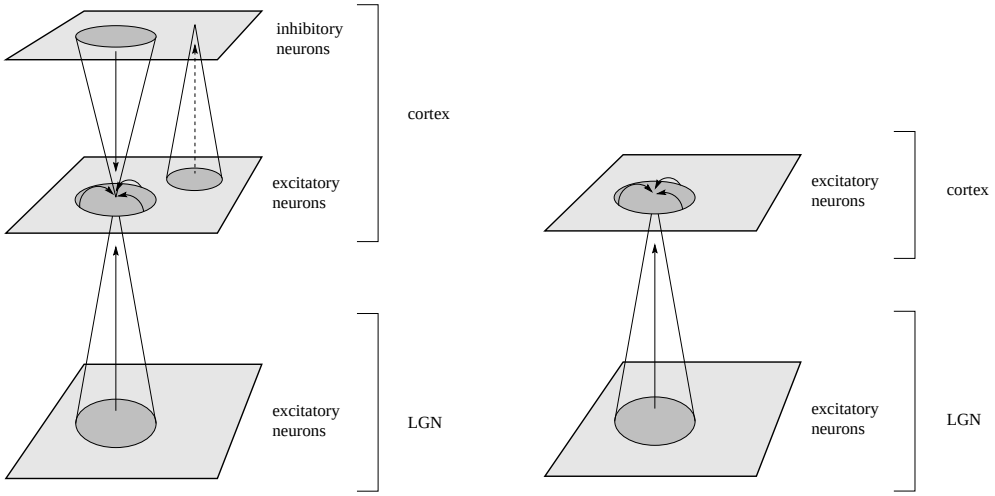


Figure 4.1: The left panel reproduces the architecture of our network model of layer 4 in primary visual cortex as introduced in Sect. 2.2. It can be separated into three layers of non-linear neurons. The two upper layers are designed to represent cortical excitatory and inhibitory neurons. Each of the excitatory cells receives lateral input from neighbouring excitatory and inhibitory cortical cells. Inhibitory neurons are driven by the activity of excitatory neurons. Feedforward input into the cortex is provided by LGN neurons in the bottom layer, producing correlated spike activity of given statistics. All synaptic connections except for those from excitatory to inhibitory cells (dashed arrow) are subject to activity-driven learning. The full dynamics of this model are very complicated and rather intractable analytically. For the subsequent study the network will therefore be simplified in three respects. First, the inhibitory layer will be removed so as to arrive at the architecture shown in the right panel. Second, the neurons' non-linear activation function will be replaced by a linear one and, third, their refractory potential will be neglected.

4.2 Dynamics of Spike Based Learning

The dynamics of synaptic plasticity in our model are governed by the Hebbian learning rules described in Sect. 2.3. According to these rules the efficacy of a synapse is changed in dependence on the spike trains emitted by the two neurons it is connecting. Within the framework of the stochastic spike response model, which we have adopted to represent the individual neurons (cf. Sects. 2.1 and 3.1), these spike trains are random processes and are thus different in each run of the same network. As a consequence, the resulting synaptic weights are random processes as well.

This consideration raises the question as to whether it is possible to obtain detailed temporal information concerning the pattern of synaptic connectivity emerging during one individual network run. In fact, this might be very difficult in situations where learning proceeds on the timescale of the fluctuations in neuronal activity as is the case in short term plasticity (Zucker 1989, Thomson and Deuchars 1994, Varela et al. 1997, Tsodyks et al. 1998, Varela et al.

1999). Hebbian plasticity, on the other hand, is usually assumed to occur on a much longer time scale so that a large number of spike events is needed to induce a significant change of a synapse's efficacy. This means that the current weight of each synapse is the result of a slow process averaging over the stochastic fluctuations in neuronal activity (Kempter et al. 1999, Kempter et al. 2000).

In our simulations of intracortical and feedforward map formation we have chosen the relaxation time constants for excitatory synapses to be 400s and 800s, respectively. In view of all the neuronal time constants being of the order of a few milliseconds and the neural firing rate being several spikes per second, the development of excitatory synapses can therefore be considered a slow process in the above sense. The change that is applied to the excitatory weight J_{ij}^{e} during one simulated time step of size Δt is given by Eq. (2.1). As explained in Sect. 2.3, J_{ij}^{e} is taken to be the *effective* weight connecting cell j to cell i. That means that J_{ij}^{e} is equal to the efficacy K_{ij}^{e} of a single synapse multiplied by the number n_{ij} of synapses from neuron j to neuron i. In continuous time, the corresponding dynamics of K_{ij}^{e} read

$$\frac{\mathrm{d}}{\mathrm{d}t}K_{ij}^{\mathrm{e}}(t) = \nu \left[a_i(t) \int_{-\infty}^{t} W\left(t-t'\right) a_j(t')\,\mathrm{d}t' + a_i(t)\sigma^{\mathrm{e}} + \zeta^{\mathrm{e}} \right] - \vartheta^{\mathrm{e}} K_{ij}^{\mathrm{e}}(t)\,, \quad (4.1)$$

where ν is a constant learning parameter, $\zeta^{\mathrm{e}} := \xi^{\mathrm{e}}/\Delta t$, and $a_i(t)$ denotes the spike train of neuron i as defined in (3.3). Equation (2.1) can be regained from (4.1) by integrating over the time interval $[t, t+\Delta t)$, multiplying both sides by the number n_{ij} of synapses from j to i, and substituting

$$\begin{aligned} A_{ij}^{\mathrm{e}} &:= \nu n_{ij}\,, \\ J_{ij}^{\mathrm{e}} &:= n_{ij} K_{ij}^{\mathrm{e}}\,. \end{aligned}$$

According to the above considerations there is no harm in averaging the right hand side of (4.1) over temporal fluctuations of neuronal activity as long as we are only interested in the long-term synaptic dynamics. We thus write

$$\frac{\mathrm{d}}{\mathrm{d}t}K_{ij}^{\mathrm{e}}(t) = \overline{\nu \left[a_i(t) \int_{-\infty}^{t} W\left(t-t'\right) a_j(t')\,\mathrm{d}t' + a_i(t)\sigma^{\mathrm{e}} + \zeta^{\mathrm{e}} \right]} - \vartheta^{\mathrm{e}} K_{ij}^{\mathrm{e}}(t)\,,$$

with $\overline{x(t)}$ denoting the temporal average of a quantity $x(t)$, i.e.,

$$\overline{x(t)} := \lim_{T\to\infty} \frac{1}{T} \int_0^T \mathrm{d}t'\, x(t+t')\,.$$

Since by assumption the synaptic efficacies do not change significantly on the time scale of neuronal activity variations, temporal averaging is to be carried out with all the synaptic weights held constant. This yields

$$\frac{\mathrm{d}}{\mathrm{d}t}K_{ij}^{\mathrm{e}}(t) = \nu \left[\int_0^{\infty} W\left(t'\right) \overline{a_i(t) a_j(t-t')}\,\mathrm{d}t' + \overline{a_i(t)}\sigma^{\mathrm{e}} + \zeta^{\mathrm{e}} \right] - \vartheta^{\mathrm{e}} K_{ij}^{\mathrm{e}}(t)\,. \quad (4.2)$$

Although Eqs. (4.1) and (4.2) both describe the long-term dynamics of synaptic plasticity, they are of course *not* equivalent on short timescales. The

right hand side of (4.1) includes Dirac delta-pulses in the form of the neural spike trains $a_i(t)$, so that the resulting weights $K_{ij}^{\mathrm{e}}(t)$ are discontinuous functions of time. On the right hand side of (4.2) in contrast the delta pulses have vanished due to the procedure of temporal averaging. Equation (4.2) thus yields weights $K_{ij}^{\mathrm{e}}(t)$ that develop continuously in time. Technically speaking, Eq. (4.2) results from (4.1) by separating the time scale of long-term synaptic dynamics from the time scale of short-term random fluctuations.

To proceed from (4.2) we now have to express the temporal averages on the right hand side in terms of the network's synaptic weights so as to obtain a closed set of dynamical equations. How can we do that? The key idea is to assume that the dynamics of neuronal activity in the network is ergodic, so that the temporal averages are identical in almost all network runs. Under this condition a temporal average can be replaced by the respective quantity averaged over time *and* over an ensemble of networks as introduced in the previous chapter, i.e., we may replace

$$\overline{x(t)} = \left\langle \overline{x(t)} \right\rangle \ ,$$

where angular brackets denote the average over an ensemble of networks. It must be emphasized, though, that this is in fact an *assumption* which is not guaranteed to hold true for a network of arbitrary architecture consisting of spike response-neurons with arbitrary response kernels $\varepsilon(t)$ and $\eta(t)$.

Numerical simulations (Gerstner 1995, Mar et al. 1999) do indicate, however, that in an asynchronously firing network of spike response-neurons with exponentially decaying response kernels the temporal averages $\overline{a_i(t)}$ and $\overline{a_i(t)\, a_j(t')}$ are normally identical to the respective ensemble averages. Since this is the setup that we are going to investigate, we will take for granted that in the following we can identify

$$\overline{a_i(t)} = \left\langle \overline{a_i(t)} \right\rangle = \overline{\langle a_i(t) \rangle} \ , \tag{4.3}$$

$$\overline{a_i(t) a_j(t-t')} = \left\langle \overline{a_i(t) a_j(t-t')} \right\rangle = \overline{\langle a_i(t) a_j(t-t') \rangle} \ . \tag{4.4}$$

Here the rightmost equalities arise from exchanging the integration over time with the summation over the distribution of realizations in the network ensemble.

Inserting (4.3) and (4.4) into (4.2) we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t} K_{ij}^{\mathrm{e}}(t) = \nu \left[\int_0^\infty W\left(t'\right) \overline{\langle a_i(t) a_j(t-t') \rangle}\, \mathrm{d}t' + \overline{\langle a_i(t) \rangle} \sigma^{\mathrm{e}} + \zeta^{\mathrm{e}} \right] - \vartheta^{\mathrm{e}} K_{ij}^{\mathrm{e}}(t) \ . \tag{4.5}$$

We will now use the methods developed in the previous chapter to derive analytical expressions for the required ensemble averages. Averaging these expressions over time, we turn Eq. (4.5) into a closed system of equations describing the dynamics of synaptic plasticity in the network under consideration.

4.3 Plasticity of Feedforward Synapses

How does correlated spike activity in the LGN drive the development of a feedforward orientation map? How is the layout of this map influenced by an intracortical orientation map? These questions will be addressed in the following. Earlier mathematical analyses of so-called correlation-based pattern formation (Kammen and Yuille 1988, Yuille et al. 1989, MacKay and Miller 1990, Stetter et al. 1993, Wimbauer et al. 1994, Wimbauer 1996, Wimbauer et al. 1998) were performed on the basis of graded-response cells and did therefore not account for the spiking nature of biological neurons. Moreover, they did not investigate how an emerging orientation map is affected by the presence of oriented patterns of intracortical connectivity.

In contrast, the investigation that we will give subsequently is based on spiking neurons. For the sake of simplicity we will utilize the reduced network model described in Sect. 4.1. As in our numerical simulations we will assume the intracortical connectivity to be fixed during the development of feedforward synapses. In this way we separate the geniculocortical learning dynamics from the mechanisms of intracortical plasticity so as to reduce the model's dynamical complexity. In the Discussion of Chapter 2 we have already argued that we need not presume this to be strictly fulfilled in biological systems.

4.3.1 The Learning Equation

The network that we are going to analyze consists of two square grids of linear stochastic spike response-neurons, representing geniculate relay cells and cortical pyramidal neurons, respectively; see Fig. 4.1. We define their linear activation functions by

$$q(h) := c_0 + c_1\, h \tag{4.6}$$

for cortical cells and

$$q_{\mathrm{g}}(h) := d_0 + d_1\, h \tag{4.7}$$

for geniculate cells. Since we neglect the refractory potential, the coupling operator $\hat{\kappa}$ introduced in Eq. (3.3) reads

$$\left[\hat{\kappa} a\right]_i (t) := \sum_j J_{ij} \int_{-\infty}^{\infty} a_j(t')\, \varepsilon(t - t')\, \mathrm{d}t' \,, \tag{4.8}$$

with J_{ij} denoting the total synaptic weight connecting cell j to cell i and $\varepsilon(t)$ being the postsynaptic potential

$$\varepsilon(t) := \exp(-t/\tau_\varepsilon)$$

for $t > 0$ and vanishing for $t \leq 0$.

Notice that in Eq. (4.8) the sum is to be taken over the cortical *and* the geniculate layer. In the following calculations it will be convenient to distinguish explicitly between these two layers. To this end we introduce two sets

$\mathcal{L}_{\mathrm{G}}$ and $\mathcal{L}_{\mathrm{C}}$ comprising the indices of geniculate and cortical cells, respectively, so that (4.8) reads

$$[\hat{\kappa}a]_i(t) := \sum_{j\in\mathcal{L}_{\mathrm{G}}} J_{ij}\int_{-\infty}^{\infty} a_j(t')\,\varepsilon(t-t')\,\mathrm{d}t' + \sum_{j\in\mathcal{L}_{\mathrm{C}}} J_{ij}\int_{-\infty}^{\infty} a_j(t')\,\varepsilon(t-t')\,\mathrm{d}t' \ . \quad (4.9)$$

As explained in Sect. 2.2, geniculate cells in our model do not receive input from other cells within the network but rather have their membrane potential prescribed externally. This external potential is intended to mimick the effect of retinogeniculate input as well as feedback from the cortex. In our simulations, the external potential is drawn every 10 ms as a realization of a Gaussian random field with zero mean and a Mexican-hat-like correlation function of the form

$$M_{ij} := \exp\left[-\mathrm{d}(i,j)^2/(2\cdot\sigma_{\mathrm{c}}^2)\right] - (1/9)\exp\left[-\mathrm{d}(i,j)^2/(18\cdot\sigma_{\mathrm{c}}^2)\right] \ , \quad (4.10)$$

where $\mathrm{d}(i,j)$ denotes the distance between the two geniculate cells $i,j\in\mathcal{L}_{\mathrm{G}}$ and σ_{c} is the correlation length.

The correlation of the membrane potentials of two geniculate neurons i and j taken at times t and t' is therefore determined by M_{ij} if the realization of the random field does not change during the time interval $[t,t')$. Otherwise the two values are uncorrelated, as they belong to statistically independent realizations of the Gaussian random field. With new realizations being drawn at $t=0$, $t=10$ ms, $t=20$ ms, etc., the resulting spatio-temporal correlation function can be written as

$$\left\langle h_i^{\mathrm{ext}}(t)h_j^{\mathrm{ext}}(t')\right\rangle - \left\langle h_i^{\mathrm{ext}}(t)\right\rangle\left\langle h_j^{\mathrm{ext}}(t')\right\rangle = H_0\,T\left(t,t'\right)\,M_{ij}\ ,$$

where $i,j\in\mathcal{L}_{\mathrm{G}}$ and

$$T\left(t,t'\right) := \begin{cases} 1 \ \text{ if } \ n\cdot 10\,\mathrm{ms} \le t,t' < (n+1)\cdot 10\,\mathrm{ms} \ \text{ for some } n\in\mathbb{N}_0\ , \\ 0 \ \text{ otherwise.}\end{cases}$$

Because in our model cortical cells do not receive external input, we have $h_i^{\mathrm{ext}}(t)\equiv 0$ for $i\in\mathcal{L}_{\mathrm{C}}$ and so the correlation of two cells' external membrane potential is zero if at least one of these cells is in the cortical layer. Putting things together we thus obtain

$$\begin{aligned}\left\langle h_i^{\mathrm{ext}}(t)h_j^{\mathrm{ext}}(t')\right\rangle - \left\langle h_i^{\mathrm{ext}}(t)\right\rangle\left\langle h_j^{\mathrm{ext}}(t')\right\rangle = \\ = \left\langle h_i^{\mathrm{ext}}(t)h_j^{\mathrm{ext}}(t')\right\rangle = \begin{cases} H_0\,T\left(t,t'\right)\,M_{ij}\ , & \text{if } i,j\in\mathcal{L}_{\mathrm{G}}\ , \\ 0 \ \text{ otherwise,}\end{cases}\end{aligned} \quad (4.11)$$

and furthermore

$$\left\langle h_i^{\mathrm{ext}}(t)\right\rangle \equiv 0\,, \ \text{ for any } i\in\mathcal{L}_{\mathrm{G}}\cup\mathcal{L}_{\mathrm{C}}. \quad (4.12)$$

In Sect. 3.4, Eqs. (3.73) through (3.75) we have found analytical expressions for the mean activity and the two-spike correlations in a recurrent network of

linear stochastic spike response-neurons in the presence of an external membrane potential. Subsequently we shall assume that the inverse operators required in these expressions do exist and can be approximated to a sufficient degree of precision by their respective von-Neumann series up to second order. Using the above definitions and neglecting all the terms beyond second order in the weights we thus have

$$\begin{aligned} \alpha_i &:= \\ \langle a_i(t)\rangle &= \begin{cases} d_0\ , \quad \text{if } i \in \mathcal{L}_{\rm G}\ , \\ c_0 + d_0 c_1 \tau_\varepsilon \sum_{j\in\mathcal{L}_{\rm G}} J_{ij} + c_0 c_1 \tau_\varepsilon \sum_{j\in\mathcal{L}_{\rm C}} J_{ij} \\ + d_0 c_1^2 \tau_\varepsilon^2 \sum_{j\in\mathcal{L}_{\rm C}} \sum_{k\in\mathcal{L}_{\rm G}} J_{ij}J_{jk} + c_0 c_1^2 \tau_\varepsilon^2 \sum_{j\in\mathcal{L}_{\rm C}} \sum_{k\in\mathcal{L}_{\rm C}} J_{ij}J_{jk}\ , \quad \text{if } i \in \mathcal{L}_{\rm C}\ , \end{cases} \end{aligned} \tag{4.13}$$

and, for $i \in \mathcal{L}_{\rm C}$, $j \in \mathcal{L}_{\rm G}$,

$$\begin{aligned} \langle a_i(t_1)a_j(t_2)\rangle &= \alpha_i\alpha_j + c_1 J_{ij}\alpha_j \varepsilon(t_2-t_1) \\ &+ \ [1-\theta(t_2-t_1)]\ c_1^2 \sum_{k\in\mathcal{L}_{\rm C}} J_{ik}J_{kj}\alpha_j \int_{-\infty}^{\infty} {\rm d}t' \varepsilon(t_1-t')\varepsilon(t'-t_2) \\ &+ \ c_1 d_1^2 \sum_{k\in\mathcal{L}_{\rm G}} J_{ik} \int_{-\infty}^{\infty} {\rm d}t' \varepsilon(t_1-t') \left\langle h_k^{\rm ext}(t') h_j^{\rm ext}(t_2)\right\rangle \\ &+ \ c_1^2 d_1^2 \sum_{\substack{k\in\mathcal{L}_{\rm G} \\ l\in\mathcal{L}_{\rm C}}} J_{il}J_{lk} \int_{-\infty}^{\infty} {\rm d}t' \int_{-\infty}^{\infty} {\rm d}t'' \varepsilon(t_1-t')\varepsilon(t'-t'') \left\langle h_k^{\rm ext}(t'') h_j^{\rm ext}(t_2)\right\rangle\ , \end{aligned} \tag{4.14}$$

where $\theta(t)$ is Heaviside's step function,

$$\theta(t) := \begin{cases} 1 & \text{if } t > 0\ , \\ 0 & \text{otherwise.} \end{cases} \tag{4.15}$$

In the derivation of (4.13) and (4.14) we have utilized the fact that $J_{ij} = 0$ for $i \in \mathcal{L}_{\rm G}$ because in our network there is neither a synaptic connection from the cortical to the geniculate layer nor within the geniculate layer. Averaging the above expressions over time yields $\overline{\langle a_i\rangle} = \alpha_i$ and

$$\begin{aligned} \overline{\langle a_i(t)a_j(t-t')\rangle} &= \alpha_i\alpha_j + c_1 J_{ij}\alpha_j \varepsilon(t') \\ &+ \ \left[1-\theta(-t')\right]\ c_1^2 \sum_{k\in\mathcal{L}_{\rm C}} J_{ik}J_{kj}\alpha_j \int_{-\infty}^{\infty} {\rm d}t'' \varepsilon(t'')\varepsilon(t'-t'') \\ &+ \ c_1 d_1^2 \sum_{k\in\mathcal{L}_{\rm G}} J_{ik} \int_{-\infty}^{\infty} {\rm d}t'' \varepsilon(t'') \overline{\left\langle h_k^{\rm ext}(t-t'') h_j^{\rm ext}(t-t')\right\rangle} \\ &+ \ c_1^2 d_1^2 \sum_{\substack{k\in\mathcal{L}_{\rm G} \\ l\in\mathcal{L}_{\rm C}}} J_{il}J_{lk} \int_{-\infty}^{\infty} {\rm d}t'' \int_{-\infty}^{\infty} {\rm d}t''' \varepsilon(t'')\varepsilon(t''') \overline{\left\langle h_k^{\rm ext}(t-t''-t''') h_j^{\rm ext}(t-t')\right\rangle}\ , \end{aligned} \tag{4.16}$$

with $i \in \mathcal{L}_{\rm C}$, $j \in \mathcal{L}_{\rm G}$. From Eq. (4.11) we derive the required temporal average of the external correlation function,

$$\overline{\left\langle h_i^{\rm ext}(t-t'') h_j^{\rm ext}(t-t')\right\rangle} \quad = \quad H_0 M_{ij} \overline{T(t-t'', t-t')}$$

$$
\begin{aligned}
&= H_0 M_{ij}\, \theta(\tau_C - \left|t' - t''\right|) \left(1 - \frac{1}{\tau_C}\left|t' - t''\right|\right) \\
&=: H_0 M_{ij}\, \Lambda(t' - t'')\,,
\end{aligned}
$$

where $i, j \in \mathcal{L}_{\mathrm{G}}$ and $\tau_C = 10$ ms.

Let us now insert these quantities into the learning equation (4.5). As in our numerical simulations in Chapter 2 we use an exponentially decaying time window,

$$W(t) := \exp(-t/\tau_W)$$

for $t > 0$ and vanishing for $t \leq 0$. In addition, we define

$$
\begin{aligned}
\tau_*^2 &:= \int_{-\infty}^{\infty} \mathrm{d}t' \int_{-\infty}^{\infty} \mathrm{d}t''\, W(t')\, \varepsilon(t'')\, \varepsilon(t' - t'') = \left(\frac{\tau_\varepsilon \tau_W}{\tau_\varepsilon + \tau_W}\right)^2 , \\
\tau_1^2 &:= \int_{-\infty}^{\infty} \mathrm{d}t' \int_{-\infty}^{\infty} \mathrm{d}t''\, W(t')\, \varepsilon(t'')\, \Lambda(t' - t'')\,, \\
\tau_2^3 &:= \int_{-\infty}^{\infty} \mathrm{d}t' \int_{-\infty}^{\infty} \mathrm{d}t'' \int_{-\infty}^{\infty} \mathrm{d}t'''\, W(t')\, \varepsilon(t'')\, \varepsilon(t''')\, \Lambda(t' - t'' - t''')\,,
\end{aligned}
$$

and thus get

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} K_{ij}^{\mathrm{e}} = -\vartheta^{\mathrm{e}} J_{ij}^{\mathrm{e}} + \nu \Bigg\{ & (d_0 c_0 \tau_W + \sigma^{\mathrm{e}} c_0 + \zeta^{\mathrm{e}}) + d_0 c_1 \tau_* J_{ij}^{\mathrm{e}} \\
& + \left(d_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} d_0 c_1 \tau_\varepsilon\right) \sum_{k \in \mathcal{L}_{\mathrm{G}}} J_{ik}^{\mathrm{e}} + \left(d_0 c_0 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} c_0 c_1 \tau_\varepsilon\right) \sum_{k \in \mathcal{L}_{\mathrm{C}}} J_{ik} \\
& + \left(d_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} d_0 c_1^2 \tau_\varepsilon^2\right) \sum_{\substack{k \in \mathcal{L}_{\mathrm{G}} \\ l \in \mathcal{L}_{\mathrm{C}}}} J_{il} J_{lk}^{\mathrm{e}} + \left(d_0 c_0 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} c_0 c_1^2 \tau_\varepsilon^2\right) \sum_{\substack{k \in \mathcal{L}_{\mathrm{C}} \\ l \in \mathcal{L}_{\mathrm{C}}}} J_{il} J_{lk} \\
& + d_0 c_1^2 \tau_*^2 \sum_{k \in \mathcal{L}_{\mathrm{C}}} J_{ik} J_{kj}^{\mathrm{e}} + c_1 d_1^2 \tau_1^2 H_0 \sum_{k \in \mathcal{L}_{\mathrm{G}}} M_{kj} J_{ik}^{\mathrm{e}} + c_1^2 d_1^2 \tau_2^3 H_0 \sum_{\substack{k \in \mathcal{L}_{\mathrm{G}} \\ l \in \mathcal{L}_{\mathrm{C}}}} J_{il} M_{kj} J_{lk}^{\mathrm{e}} \Bigg\}\,,
\end{aligned}
\tag{4.17}
$$

with $i \in \mathcal{L}_{\mathrm{C}}$ and $j \in \mathcal{L}_{\mathrm{G}}$. Here the upper index 'e' has been applied only to the geniculocortical weights so that they can be distinguished clearly from the intracortical efficacies, which are kept fixed during the course of the learning process.

Equation (4.17) yields the temporal change of the efficacy K_{ij}^{e} of a single synapse connecting neuron $j \in \mathcal{L}_{\mathrm{G}}$ to neuron $i \in \mathcal{L}_{\mathrm{C}}$. Since this temporal change is the same for all the synapses from j to i, all their efficacies will be equal at any given time provided they are equal at the beginning of the learning dynamics. The total geniculocortical weight J_{ij}^{e} between two cells j and i can therefore be written as the efficacy K_{ij}^{e} of a single synapse multiplied by the number n_{ij} of synapses connecting j to i, i.e.,

$$J_{ij}^{\mathrm{e}} = n_{ij} K_{ij}^{\mathrm{e}}\,. \tag{4.18}$$

Thus, Eq. (4.17) multiplied by n_{ij} becomes

$$
\frac{\mathrm{d}}{\mathrm{d}t} n_{ij} K_{ij}^{\mathrm{e}} = -\vartheta^{\mathrm{e}}\, n_{ij} K_{ij}^{\mathrm{e}} + \nu\, n_{ij} \Bigg\{ (d_0 c_0 \tau_W + \sigma^{\mathrm{e}} c_0 + \zeta^{\mathrm{e}}) + d_0 c_1 \tau_* n_{ij} K_{ij}^{\mathrm{e}} +
$$

$$
\begin{aligned}
&+ \left(d_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} d_0 c_1 \tau_\varepsilon\right) \sum_{k \in \mathcal{L}_{\mathrm{G}}} n_{ik} K^{\mathrm{e}}_{ik} + \left(d_0 c_0 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} c_0 c_1 \tau_\varepsilon\right) \sum_{k \in \mathcal{L}_{\mathrm{C}}} J_{ik} \\
&+ \left(d_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} d_0 c_1^2 \tau_\varepsilon^2\right) \sum_{\substack{k \in \mathcal{L}_{\mathrm{G}} \\ l \in \mathcal{L}_{\mathrm{C}}}} J_{il}\, n_{lk} K^{\mathrm{e}}_{lk} \\
&+ \left(d_0 c_0 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} c_0 c_1^2 \tau_\varepsilon^2\right) \sum_{\substack{k \in \mathcal{L}_{\mathrm{C}} \\ l \in \mathcal{L}_{\mathrm{C}}}} J_{il} J_{lk} + d_0 c_1^2 \tau_*^2 \sum_{k \in \mathcal{L}_{\mathrm{C}}} J_{ik}\, n_{kj} K^{\mathrm{e}}_{kj} \\
&+ c_1 d_1^2 \tau_1^2 H_0 \sum_{k \in \mathcal{L}_{\mathrm{G}}} M_{kj}\, n_{ik} K^{\mathrm{e}}_{ik} + c_1^2 d_1^2 \tau_2^3 H_0 \sum_{\substack{k \in \mathcal{L}_{\mathrm{G}} \\ l \in \mathcal{L}_{\mathrm{C}}}} J_{il} M_{kj}\, n_{lk} K^{\mathrm{e}}_{lk} \Bigg\}, \qquad (4.19)
\end{aligned}
$$

where n_{ij} is taken to be a constant in time.

4.3.2 Modeling Neuronal Arborization

Before we can go on and investigate the synaptic dynamics described by the learning equation (4.19) we have to fix the quantities n_{ij}. According to the above explanations n_{ij} specifies the number of synapses that a given geniculate neuron j in our network makes onto a given cortical neuron i. Since the n_{ij} are taken to be constant during the learning process, they are parameters of the model. To a certain degree, the choice of these parameters predetermines the patterns of synaptic connectivity that can emerge from the learning dynamics. They can be chosen in several plausible ways; here we will concentrate on three of them.

Flat Arborization

As a very simple approach let us assume that a cortical neuron i receives exactly one synaptic input from all those geniculate neurons j whose retinotopic positions are within a certain circular area around the cortical neuron's position and that there is no input from all the remaining geniculate cells. In our model, a cells retinotopic coordinate directly corresponds to its position within the respective network layer so that we have

$$
n_{ij} = \begin{cases} 1 & \text{if } \mathrm{d}(i,j) < r_{\max}, \\ 0 & \text{otherwise,} \end{cases}
$$

where $\mathrm{d}(i,j)$ denotes the distance between the grid positions of cells i and j – which is proportional to their retinotopic distance. Taking into account that $n_{ij}^2 = n_{ij}$ and defining the arbor function

$$
A_{ij} := \nu\, n_{ij} \qquad (4.20)
$$

the learning equation (4.19) can now be written

$$
\frac{\mathrm{d}}{\mathrm{d}t} J^{\mathrm{e}}_{ij} = -\vartheta^{\mathrm{e}}\, J^{\mathrm{e}}_{ij} + \nu d_0 c_1 \tau_* J^{\mathrm{e}}_{ij} + A_{ij} \Big\{ \left(d_0 c_0 \tau_W + \sigma^{\mathrm{e}} c_0 + \zeta^{\mathrm{e}}\right) + \qquad (4.21)
$$

$$+ \left(d_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} d_0 c_1 \tau_\varepsilon\right) \sum_{k \in \mathcal{L}_{\mathrm{G}}} J^{\mathrm{e}}_{ik} + \left(d_0 c_0 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} c_0 c_1 \tau_\varepsilon\right) \sum_{k \in \mathcal{L}_{\mathrm{C}}} J_{ik}$$

$$+ \left(d_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} d_0 c_1^2 \tau_\varepsilon^2\right) \sum_{\substack{k \in \mathcal{L}_{\mathrm{G}} \\ l \in \mathcal{L}_{\mathrm{C}}}} J_{il} J^{\mathrm{e}}_{lk} + \left(d_0 c_0 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} c_0 c_1^2 \tau_\varepsilon^2\right) \sum_{\substack{k \in \mathcal{L}_{\mathrm{C}} \\ l \in \mathcal{L}_{\mathrm{C}}}} J_{il} J_{lk}$$

$$+ d_0 c_1^2 \tau_*^2 \sum_{k \in \mathcal{L}_{\mathrm{C}}} J_{ik} J^{\mathrm{e}}_{kj} + c_1 d_1^2 \tau_1^2 H_0 \sum_{k \in \mathcal{L}_{\mathrm{G}}} M_{kj} J^{\mathrm{e}}_{ik} + c_1^2 d_1^2 \tau_2^3 H_0 \sum_{\substack{k \in \mathcal{L}_{\mathrm{G}} \\ l \in \mathcal{L}_{\mathrm{C}}}} J_{il} M_{kj} J^{\mathrm{e}}_{lk} \Bigg\} ,$$

where J^{e}_{ij} is the effective weight given in (4.18).

Dense Arborization

Probably a biologically more realistic view of neuronal organization is considering the number of synapses between a geniculate cell j and a cortical cell i a random quantity whose expectation value depends on the overlap of the geniculate neuron's axonal arbor with the cortical cell's dendritic arbor (Braitenberg and Schüz 1991). For simplicity we neglect the statistical nature of this process. Instead we let n_{ij} be the expected number of synapses from j to i if the retinotopic distance of i and j is below a certain maximal value, whereas we let $n_{ij} = 0$ beyond this maximal distance. This approach implies a dense arborization, where each cortical neuron receives input from all geniculate cells within the given maximal retinotopic radius. As both the axonal and the dendritic arbors extend over a finite region and are densest in their center, the number n_{ij} decreases with the retinotopic distance of the two neurons. In our model, the retinotopic distance of two cells is proportional to the distance $\mathrm{d}(i,j)$ of their respective grid positions. Therefore, n_{ij} is a decreasing function of $\mathrm{d}(i,j)$. For the numerical simulations presented in Chapter 2 we have chosen

$$n_{ij} = n_0 \exp\left[-\mathrm{d}(i,j)^2/(2 \cdot \sigma_{\mathrm{n}}^2)\right]$$

if $\mathrm{d}(i,j) < r_{\max}$ and $n_{ij} = 0$ otherwise, with $r_{\max} = 5.5$ and $\sigma_{\mathrm{n}} = 3$. Using the arbor function (4.20) the learning equation (4.19) can be put into the form

$$\frac{\mathrm{d}}{\mathrm{d}t} J^{\mathrm{e}}_{ij} = -\vartheta^{\mathrm{e}} J^{\mathrm{e}}_{ij} + A_{ij} \Bigg\{ \left(d_0 c_0 \tau_W + \sigma^{\mathrm{e}} c_0 + \zeta^{\mathrm{e}}\right) + d_0 c_1 \tau_* J^{\mathrm{e}}_{ij} + \tag{4.22}$$

$$+ \left(d_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} d_0 c_1 \tau_\varepsilon\right) \sum_{k \in \mathcal{L}_{\mathrm{G}}} J^{\mathrm{e}}_{ik} + \left(d_0 c_0 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} c_0 c_1 \tau_\varepsilon\right) \sum_{k \in \mathcal{L}_{\mathrm{C}}} J_{ik}$$

$$+ \left(d_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} d_0 c_1^2 \tau_\varepsilon^2\right) \sum_{\substack{k \in \mathcal{L}_{\mathrm{G}} \\ l \in \mathcal{L}_{\mathrm{C}}}} J_{il} J^{\mathrm{e}}_{lk} + \left(d_0 c_0 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} c_0 c_1^2 \tau_\varepsilon^2\right) \sum_{\substack{k \in \mathcal{L}_{\mathrm{C}} \\ l \in \mathcal{L}_{\mathrm{C}}}} J_{il} J_{lk}$$

$$+ d_0 c_1^2 \tau_*^2 \sum_{k \in \mathcal{L}_{\mathrm{C}}} J_{ik} J^{\mathrm{e}}_{kj} + c_1 d_1^2 \tau_1^2 H_0 \sum_{k \in \mathcal{L}_{\mathrm{G}}} M_{kj} J^{\mathrm{e}}_{ik} + c_1^2 d_1^2 \tau_2^3 H_0 \sum_{\substack{k \in \mathcal{L}_{\mathrm{G}} \\ l \in \mathcal{L}_{\mathrm{C}}}} J_{il} M_{kj} J^{\mathrm{e}}_{lk} \Bigg\} ,$$

where J^{e}_{ij} denotes the effective weight given in (4.18).

Sparse Arborization

As in the previous approach we consider n_{ij} a random quantity whose expectation value n^*_{ij} depends on the overlap between the axonal and dendritic arbors of the respective cells j and i. Here we presume this expected number of synapses to be small, $n^*_{ij} \ll 1$. This means that only a small fraction of all geniculate cells j within the arborization radius of a given cortical cell i do actually make a synaptic contact onto this cortical neuron and the number of synapses between j and i is either 0 or 1. Thus, it is $n^2_{ij} = n_{ij}$ as in the case of a flat arborization (see above).

Inspecting the right hand side of (4.19) we find that the sums involving the geniculocortical efficacies are either of the form $\sum_{l\in\mathcal{L}_\mathrm{C}} B_{il} n_{lj} K^\mathrm{e}_{lj}$ or of the form $\sum_{k\in\mathcal{L}_\mathrm{G}} n_{ik} K^\mathrm{e}_{ik} C_{kj}$. If the functions B_{il}, C_{kj}, K^e_{lk}, and n^*_{lk} vary sufficiently slow as l runs over neighbouring neurons on the cortical surface or as k runs over neighbouring neurons in the LGN, then the above sums are self-averaging and we can write

$$\begin{aligned}
\sum_{l\in\mathcal{L}_\mathrm{C}} B_{il} n_{lj} K^\mathrm{e}_{lj} &\approx \sum_{l\in\mathcal{L}_\mathrm{C}} B_{il} n^*_{lj} K^\mathrm{e}_{lj} \,, \\
\sum_{k\in\mathcal{L}_\mathrm{G}} n_{ik} K^\mathrm{e}_{ik} C_{kj} &\approx \sum_{k\in\mathcal{L}_\mathrm{G}} n^*_{ik} K^\mathrm{e}_{ik} C_{kj} \,.
\end{aligned}$$

Similarly, averaging over a sufficiently large number N of cortical neurons k in the vicinity of cell i yields

$$\frac{1}{N} \sum\nolimits'_k D_{kj} n_{kj} K^\mathrm{e}_{kj} \approx D_{ij} n^*_{ij} K^\mathrm{e}_{ij} \,,$$

if D_{kj}, K^e_{kj}, and n^*_{kj} are slow-varying functions so that they can be considered approximately constant across the corresponding cortical area.

Taking for granted that the above approximations can be applied we average both sides of (4.19) over a neighbourhood of i so that we arrive at a learning equation that is formally identical to (4.21) but with the definitions

$$\begin{aligned}
J^\mathrm{e}_{ij} &:= n^*_{ij} K^\mathrm{e}_{ij} \,, \\
A_{ij} &:= \nu\, n^*_{ij} \,.
\end{aligned}$$

According to the preceding considerations this provides a good description of the learning dynamics if the following conditions are fulfilled.

(i) The expected number of synapses n^*_{ij} connecting a geniculate neuron j to a cortical neuron i is much smaller than 1 and varies slowly as i runs over the cortical layer or j runs over the geniculate layer. This implies that the arborization A_{ij} extends over a large number of cells in both layers.

(ii) The efficacies J_{ij} of intracortical connections and the geniculate activity correlations M_{ij} vary slowly in i and j.

(iii) The emerging geniculocortical efficacies K^e_{ij} and, correspondingly, the effective weights J^e_{ij} are sufficiently smooth in i and j.

The requirements (i) and (ii) can be met by choosing an appropriate set of network parameters. Specifically, we may take n_{ij}^* to be a wide Gaussian function of the retinotopic distance $\mathrm{d}(i,j)$ between the neurons i and j. As a prerequisite, of course, the analyzed network model must be large enough, i.e., comprise a sufficient number of neurons and synapses. Condition (iii) can only be verified a posteriori when the emerging connectivity patterns have been determined. If they turn out to be smooth, then the use of Eq. (4.21) is justified. Otherwise, the obtained results are probably useless.

4.3.3 Linear Stability Analysis

The learning equations that are to be analyzed in the following, viz., Eqs. (4.21) and (4.22) can be summarized in the formula

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}J_{ij}^{\mathrm{e}} = -\vartheta^{\mathrm{e}}\,J_{ij}^{\mathrm{e}} + B_{ij}d_0c_1\tau_*J_{ij}^{\mathrm{e}} + A_{ij}\Bigg\{&(d_0c_0\tau_W + \sigma^{\mathrm{e}}c_0 + \zeta^{\mathrm{e}}) + \\ &+ \left(d_0^2c_1\tau_\varepsilon\tau_W + \sigma^{\mathrm{e}}d_0c_1\tau_\varepsilon\right)\sum_{k\in\mathcal{L}_{\mathrm{G}}}J_{ik}^{\mathrm{e}} + \left(d_0c_0c_1\tau_\varepsilon\tau_W + \sigma^{\mathrm{e}}c_0c_1\tau_\varepsilon\right)\sum_{k\in\mathcal{L}_{\mathrm{C}}}J_{ik} \\ &+ \left(d_0^2c_1^2\tau_\varepsilon^2\tau_W + \sigma^{\mathrm{e}}d_0c_1^2\tau_\varepsilon^2\right)\sum_{\substack{k\in\mathcal{L}_{\mathrm{G}}\\ l\in\mathcal{L}_{\mathrm{C}}}}J_{il}J_{lk}^{\mathrm{e}} + \left(d_0c_0c_1^2\tau_\varepsilon^2\tau_W + \sigma^{\mathrm{e}}c_0c_1^2\tau_\varepsilon^2\right)\sum_{\substack{k\in\mathcal{L}_{\mathrm{C}}\\ l\in\mathcal{L}_{\mathrm{C}}}}J_{il}J_{lk} \\ &+ d_0c_1^2\tau_*^2\sum_{k\in\mathcal{L}_{\mathrm{C}}}J_{ik}J_{kj}^{\mathrm{e}} + c_1d_1^2\tau_1^2H_0\sum_{k\in\mathcal{L}_{\mathrm{G}}}M_{kj}J_{ik}^{\mathrm{e}} + c_1^2d_1^2\tau_2^3H_0\sum_{\substack{k\in\mathcal{L}_{\mathrm{G}}\\ l\in\mathcal{L}_{\mathrm{C}}}}J_{il}M_{kj}J_{lk}^{\mathrm{e}}\Bigg\}\,,\end{aligned} \tag{4.23}$$

with $B_{ij} \equiv \nu$ or $B_{ij} = A_{ij}$, respectively. Obviously, this is a system of linear differential equations in the effective weights $J_{ik}^{\mathrm{e}}(t)$. Using vector notation, it can be written in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{J}^{\mathrm{e}} = \mathcal{Q}\,\mathbf{J}^{\mathrm{e}} + \mathbf{F}\,, \tag{4.24}$$

with a constant matrix $\mathcal{Q}$ and a constant vector $\mathbf{F}$. If $\mathcal{Q}$ has only non-zero eigenvalues, then there is a unique fixed point of the learning dynamics, which is obtained by setting $(\mathrm{d}/\mathrm{d}t)\mathbf{J}^{\mathrm{e}} = 0$ in (4.24) and solving for $\mathbf{J}^{\mathrm{e}}$,

$$\mathbf{J}^{\mathrm{e}} = -\mathcal{Q}^{-1}\,\mathbf{F}\,.$$

For reasons of symmetry it follows from (4.23) that the geniculocortical connectivity at this fixed point must be homogeneous and isotropic provided the arbor A_{ij}, the geniculate correlations M_{ij}, and the intracortical connectivity J_{ij} are homogeneous and isotropic. Quite in contrast, we have found in our numerical simulations that a nice feedforward orientation map can develop under just these conditions; see Fig. 2.8 in Sect. 2.6. The emerging distribution of geniculocortical synaptic weights is thus neither homogeneous nor isotropic. This indicates that the rotationally symmetric fixed point can become unstable, thereby enabling the growth of an anisotropic connectivity pattern.

In the sequel, we will therefore analyze the learning dynamics with respect to the stability of the fixed point. To this end we study the temporal development of a small deviation $\mathbf{L}$ of the geniculocortical weights from their fixed-point distribution. This yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{L} = \mathcal{Q}\,\mathbf{L}\,. \tag{4.25}$$

Since (4.25) is a linear differential equation, its solution $\mathbf{L}(t)$ is determined by the initial conditions and the eigenvalues and eigenvectors of the matrix $\mathcal{Q}$. If the real part of some eigenvalue λ is positive, then the respective eigenmode will grow exponentially in time. Conversely, any mode corresponding to an eigenvalue whose real part is negative will decay exponentially. If the real parts of all the eigenvalues are negative, then the fixed point of the learning dynamics is stable. Thus, we have to solve the eigenvalue problem

$$\lambda\mathbf{L} = \mathcal{Q}\,\mathbf{L}$$

or, explicitly, with $i \in \mathcal{L}_\mathrm{C}$ and $j \in \mathcal{L}_\mathrm{G}$ as before,

$$\begin{aligned}(\lambda + \vartheta^\mathrm{e})\,L_{ij} = {} & B_{ij} d_0 c_1 \tau_* L_{ij} + A_{ij} \Bigg\{ \left(d_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^\mathrm{e} d_0 c_1 \tau_\varepsilon\right) \sum_{k\in\mathcal{L}_\mathrm{G}} L_{ik} \\ & + \left(d_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^\mathrm{e} d_0 c_1^2 \tau_\varepsilon^2\right) \sum_{k\in\mathcal{L}_\mathrm{G}} \sum_{l\in\mathcal{L}_\mathrm{C}} J_{il} L_{lk} + d_0 c_1^2 \tau_*^2 \sum_{k\in\mathcal{L}_\mathrm{C}} J_{ik} L_{kj} \\ & + c_1 d_1^2 \tau_1^2 H_0 \sum_{k\in\mathcal{L}_\mathrm{G}} M_{kj} L_{ik} + c_1^2 d_1^2 \tau_2^3 H_0 \sum_{k\in\mathcal{L}_\mathrm{G}} \sum_{l\in\mathcal{L}_\mathrm{C}} J_{il} M_{kj} L_{lk} \Bigg\}\,. \end{aligned} \tag{4.26}$$

Henceforth, it will be convenient to use a slightly modified notation for specifying individual neurons and synapses. Up to now an integer index i has been assigned to each neuron so that the quantities belonging to this neuron could be identified. The membrane potential and the spike train of neuron number i, for example, have been denoted as h_i and a_i. Correspondingly, we have been using two indices i and j to denote the efficacy J_{ij} of a synapse connecting the presynaptic cell j to the postsynaptic cell i. As the network we are dealing with is arranged in the form of two retinotopically ordered grids of neurons, we can alternatively specify a neuron by giving the layer that it resides in together with a two-dimensional vector $\mathbf{x}$ determining its position within this layer. Accordingly, a synapse is uniquely identified by the layers and the coordinates $\mathbf{x}_2$ and $\mathbf{x}_1$ of its pre- and postsynaptic neurons. Therefore, we can denote individual synapses by means of the postsynaptic cell's position $\mathbf{x}_1$ and the presynaptic cell's relative position $\mathbf{y} := \mathbf{x}_1 - \mathbf{x}_2$. Since the arbor function A_{ij} depends on retinotopic distance only, it can be written as $A_\mathbf{y}$. Analogously, B_{ij} and M_{ij} take the form $B_\mathbf{y}$ and $M_\mathbf{y}$. Thus, Eq. (4.26) becomes

$$(\lambda + \vartheta^\mathrm{e})\,L_{\mathbf{x},\mathbf{y}} = B_\mathbf{y} d_0 c_1 \tau_* L_{\mathbf{x},\mathbf{y}} + A_\mathbf{y} \Bigg\{ \left(d_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^\mathrm{e} d_0 c_1 \tau_\varepsilon\right) \sum_{\mathbf{y}'} L_{\mathbf{x},\mathbf{y}'}$$

$$+ \left(d_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} d_0 c_1^2 \tau_\varepsilon^2\right) \sum_{\mathbf{y}',\mathbf{y}''} J^{\mathrm{ic}}_{\mathbf{x},\mathbf{y}'} L_{\mathbf{x}-\mathbf{y}',\mathbf{y}''} + d_0 c_1^2 \tau_*^2 \sum_{\mathbf{y}'} J^{\mathrm{ic}}_{\mathbf{x},\mathbf{y}'} L_{\mathbf{x}-\mathbf{y}',\mathbf{y}-\mathbf{y}'}$$

$$+ c_1 d_1^2 \tau_1^2 H_0 \sum_{\mathbf{y}'} M_{\mathbf{y}-\mathbf{y}'} L_{\mathbf{x},\mathbf{y}'} + c_1^2 d_1^2 \tau_2^3 H_0 \sum_{\mathbf{y}',\mathbf{y}''} M_{\mathbf{y}-\mathbf{y}'-\mathbf{y}''} J^{\mathrm{ic}}_{\mathbf{x},\mathbf{y}'} L_{\mathbf{x}-\mathbf{y}',\mathbf{y}''} \Bigg\} , \quad (4.27)$$

where $J^{\mathrm{ic}}_{\mathbf{x},\mathbf{y}}$ denotes the efficacy of an intracortical synapse, while $L_{\mathbf{x},\mathbf{y}}$ corresponds to a geniculocortical synapse.

For simplicity, our stability analysis will be restricted to the case of a homogeneous intracortical connectivity. This means that the efficacy $J^{\mathrm{ic}}_{\mathbf{x},\mathbf{y}}$ of an intracortical synapse depends only on the relative position $\mathbf{y}$ of the presynaptic cell with respect to the postsynaptic neuron. We can therefore drop the coordinate $\mathbf{x}$ of the postsynaptic coordinate from J^{ic} so that (4.27) reduces to

$$\begin{aligned}(\lambda + \vartheta^{\mathrm{e}})\, L_{\mathbf{x},\mathbf{y}} = B_{\mathbf{y}} d_0 c_1 \tau_* L_{\mathbf{x},\mathbf{y}} + A_{\mathbf{y}} \Bigg\{ & \left(d_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} d_0 c_1 \tau_\varepsilon\right) \sum_{\mathbf{y}'} L_{\mathbf{x},\mathbf{y}'} \\ & + \left(d_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} d_0 c_1^2 \tau_\varepsilon^2\right) \sum_{\mathbf{y}',\mathbf{y}''} J^{\mathrm{ic}}_{\mathbf{y}'} L_{\mathbf{x}-\mathbf{y}',\mathbf{y}''} + d_0 c_1^2 \tau_*^2 \sum_{\mathbf{y}'} J^{\mathrm{ic}}_{\mathbf{y}-\mathbf{y}'} L_{\mathbf{x}-\mathbf{y}+\mathbf{y}',\mathbf{y}'} \\ & + c_1 d_1^2 \tau_1^2 H_0 \sum_{\mathbf{y}'} M_{\mathbf{y}-\mathbf{y}'} L_{\mathbf{x},\mathbf{y}'} + c_1^2 d_1^2 \tau_2^3 H_0 \sum_{\mathbf{y}',\mathbf{y}''} M_{\mathbf{y}-\mathbf{y}'-\mathbf{y}''} J^{\mathrm{ic}}_{\mathbf{y}'} L_{\mathbf{x}-\mathbf{y}',\mathbf{y}''} \Bigg\} .\end{aligned}$$

Finally, we perform a Fourier transformation in $\mathbf{x}$ and define

$$\begin{aligned}\widetilde{L}_{\mathbf{k},\mathbf{y}} &:= \sum_{\mathbf{x}} L_{\mathbf{x},\mathbf{y}} \exp\left(-i\mathbf{k}\mathbf{x}\right) , \\ \widetilde{J}^{\mathrm{ic}}_{\mathbf{k}} &:= \sum_{\mathbf{y}} J^{\mathrm{ic}}_{\mathbf{y}} \exp\left(-i\mathbf{k}\mathbf{y}\right) , \\ \Upsilon_{\mathbf{k},\mathbf{y}} &:= \sum_{\mathbf{y}'} M_{\mathbf{y}-\mathbf{y}'} c_1 J^{\mathrm{ic}}_{\mathbf{y}'} \exp\left(-i\mathbf{k}\mathbf{y}'\right) ,\end{aligned}$$

for $\mathbf{k} \in (-\pi, \pi]^2$, which yields

$$\begin{aligned}\frac{1}{c_1}(\lambda + \vartheta^{\mathrm{e}})\, \widetilde{L}_{\mathbf{k},\mathbf{y}} = A_{\mathbf{y}} \sum_{\mathbf{y}'} \Bigg\{ & \left(d_0^2 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} d_0 \tau_\varepsilon\right) \widetilde{L}_{\mathbf{k},\mathbf{y}'} \\ & + \left(d_0^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} d_0 \tau_\varepsilon^2\right) c_1 \widetilde{J}^{\mathrm{ic}}_{\mathbf{k}} \widetilde{L}_{\mathbf{k},\mathbf{y}'} + d_0 \tau_*^2 \mathrm{e}^{-i\mathbf{k}(\mathbf{y}-\mathbf{y}')} c_1 J^{\mathrm{ic}}_{\mathbf{y}-\mathbf{y}'} \widetilde{L}_{\mathbf{k},\mathbf{y}'} \\ & + d_1^2 \tau_1^2 H_0 M_{\mathbf{y}-\mathbf{y}'} \widetilde{L}_{\mathbf{k},\mathbf{y}'} + d_1^2 \tau_2^3 H_0 \Upsilon_{\mathbf{k},\mathbf{y}-\mathbf{y}'} \widetilde{L}_{\mathbf{k},\mathbf{y}'} \Bigg\} + d_0 \tau_* B_{\mathbf{y}} \widetilde{L}_{\mathbf{k},\mathbf{y}} .\end{aligned} \quad (4.28)$$

4.3.4 Numerical Evaluation

What do the eigenfunctions $\widetilde{L}_{\mathbf{k},\mathbf{y}}$ look like? Since we do not have an analytic expression for them, let us approach Eq. (4.28) using numerical methods. For this purpose all the network parameters must be specified numerically. In view of the large number of such parameters it is clear that a complete scan of the parameter space is not feasible. We will therefore investigate only the

τ_ε	6 ms
τ_W	10 ms
$d_1^2 H_0$	900 s^{-2}
d_0	30 s^{-1}
σ_c	1.5

Table 4.1: For a numerical approach to the eigenvalue problem (4.28) the values of all the relevant model parameters must be specified numerically. Since a complete scan of the parameter space is not feasible, we will vary only the most important parameters and fix the rest of them at the values listed in this table.

most important scenarios and concentrate on variations of σ^e, ϑ^e, $c_1 J^\mathrm{ic}$, and of the arborization model. The remaining parameters will be kept fixed at the values listed in Table 4.1, which have been chosen according to the following considerations.

The time constants of the postsynaptic potential and the learning window, τ_ε and τ_W, have been chosen in agreement with the respective values used in the numerical simulations. The quantity $d_1^2 H_0$ determines the strength of the geniculate activity correlations. For the development of orientation selectivity these correlations must be sufficiently strong, i.e., $d_1^2 H_0$ must be sufficiently large. Before fixing the geniculate mean firing rate d_0 we should recall that our present computations are based on a linear neuron model. We must therefore take care that the linear gain function (4.7) does not yield negative firing rates or, at least, that the probability for a negative firing rate to occur is low because otherwise the obtained results are useless. Thus, the value of d_0 should not be smaller than the standard deviation of the firing rate, $d_1\sqrt{H_0}$; here we let $d_0 := d_1\sqrt{H_0}$. The width of the geniculate activity correlations, σ_c has been chosen to be a factor of 1.5 larger than in the numerical simulations. The reason for this difference is that in the following calculations the neuronal arborization radius will be rescaled by about the same factor so as to obtain larger receptive fields and hence a better resolution of the emerging connectivity patterns.

As to the intracortical connectivity J^ic two different scenarios will be investigated. First, we will determine the eigenfunctions $\widetilde{L}_{\mathbf{k},\mathbf{y}}$ with a rotationally symmetric connectivity J^ic so as to obtain insight into the mechanisms governing the development of a feedforward orientation map in the presence of an isotropic lateral connectivity; cf. Sect. 2.6, Fig. 2.8. Second, we will analyze (4.28) with a non-isotropic lateral connectivity. This will help us to understand how an existing intracortical orientation map can guide the development of geniculocortical connections; see Figs. 2.9 and 2.10.

With either type of intracortical connectivity we will investigate the three different arborization models that have been introduced in the preceding subsection: flat, dense and sparse arborization. Specifically, let us define

$$B_\mathbf{y} := \nu\,, \quad A_\mathbf{y} := \begin{cases} \nu & \text{if } |\mathbf{y}| < 7.5, \\ 0 & \text{otherwise} \end{cases}$$

for the flat,

$$B_\mathbf{y} \equiv A_\mathbf{y} := \begin{cases} \nu\, n_0 \exp\left[-|\mathbf{y}|^2/(2\cdot 4.5^2)\right] & \text{if } |\mathbf{y}| < 7.5, \\ 0 & \text{otherwise} \end{cases}$$

for the dense, and

$$B_{\mathbf{y}} := \nu \,, \quad A_{\mathbf{y}} := \begin{cases} 0.2\,\nu \, \exp\left[-|\mathbf{y}|^2 / (2 \cdot 4.5^2) \right] & \text{if } |\mathbf{y}| < 7.5, \\ 0 \quad \text{otherwise} \end{cases}$$

for the sparse arborization model.

Isotropic Intracortical Connectivity

In order to model an isotropic and homogeneous intracortical connectivity pattern similar to the one shown in panel (a) of Fig. 2.8 we choose

$$c_1 \, J^{\text{ic}}_{\mathbf{y}} := 1.5\,\text{s}^{-1} \, \exp\left[-|\mathbf{y}|^2 / (2 \cdot 4.5^2) \right]$$

for $|\mathbf{y}| < 7.5$ and vanishing for $|\mathbf{y}| \geq 7.5$. Hence, the eigenvalue problem (4.28) becomes almost circularly symmetric[1] so that we need to consider only one direction of the wave vector $\mathbf{k}$. In the sequel, we will therefore assume that the direction of $\mathbf{k}$ is parallel to the x_1-axis, i.e., $\mathbf{k} = (k_1, 0)$ with $k_1 \in [0, \pi]$. The eigenvectors for any other direction of $\mathbf{k}$ can be derived easily by means of an appropriate rotation.

Once the values of all the relevant parameters and the wave vector $\mathbf{k}$ are specified, the required eigenvalues and eigenvectors can be determined numerically. To this end, we will utilize the procedure `Eigensystem` of the mathematical toolkit `Mathematica4.0` (Wolfram 1999). The only parameter that has not yet been specified is σ^{e}. According to the learning rules set forth in Sect. 2.3 σ^{e} determines the amount by which each synaptic weight is reduced in response to a spike of the postsynaptic neuron. In order to clarify its effect onto the process of pattern formation let us first investigate the case $\sigma^{\text{e}} = 0$ and then see how decreasing σ^{e} changes the obtained results.

For $\sigma^{\text{e}} = 0$, Fig. 4.2 displays the largest eigenvalue λ of (4.28) as a function of k_1. Either panel is related to one arborization model, viz., flat (left panel) and dense arborization (right panel). The results for the sparse arborization model are not shown because they are qualitatively similar. The grey-level plots at the bottom of the figure visualize the corresponding eigenvector $\widetilde{L}_{\mathbf{k},\mathbf{y}}$ for a few specific values of k_1. Each one of them comprises two arrays of 15×15 pixels displaying the real part and the imaginary part of $\widetilde{L}_{\mathbf{k},\mathbf{y}}$ as functions of $\mathbf{y} = (y_1, y_2)$. The arrays' horizontal and vertical axes, from left to right and bottom to top, correspond to the vector components y_1 and y_2, respectively, and both range over the values $\{-7, \ldots, 7\}$. White pixels indicate positive values of $\widetilde{L}_{\mathbf{k},\mathbf{y}}$, whereas black pixels indicate negative values. Accordingly, pixels appearing in a medium grey represent values that are close to zero.

The results presented in Fig. 4.2 allow the following conclusions to be drawn. First, the eigenvalue λ depends on k_1 as well as on the model parameters ϑ^{e}, ν, c_1, and n_0, as can be seen from the expressions along the vertical axes of the graphs. With ν, c_1, and n_0 being positive constants, the value of

[1] except for slight deviations that are due to the discreteness of the neuronal grid

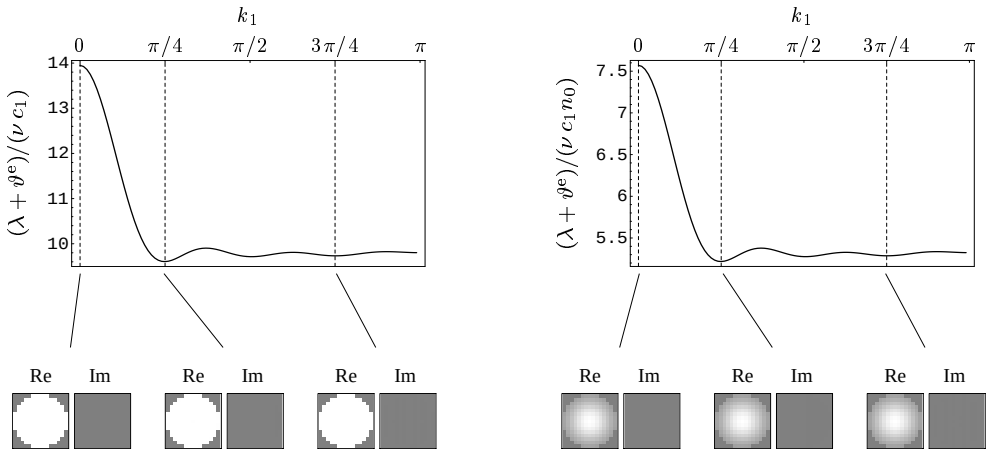


Figure 4.2: For $\sigma^e = 0$, the largest eigenvalue λ of the eigenvalue problem (4.28) is displayed as a function of the wave vector $\mathbf{k}$. Because of the rotational symmetry of Eq. (4.28) it is sufficient to consider only one direction of $\mathbf{k}$, e.g., $\mathbf{k} = (k_1, 0)$. Either panel is related to one of the previously introduced arborization models, viz., flat (left panel) and dense arborization (right panel). The results for the sparse arborization model are qualitatively similar. Note that the eigenvalue λ depends on k_1 as well as on the model parameters ϑ^e, ν, c_1, and n_0, as can be seen from the expressions along the vertical axes of the graphs.

The grey-level plots at the bottom visualize the respective eigenvector $\widetilde{L}_{\mathbf{k},\mathbf{y}}$ for a few specific values of k_1. Each one of them comprises two arrays of 15×15 pixels displaying the real part and the imaginary part of $\widetilde{L}_{\mathbf{k},\mathbf{y}}$ as functions of $\mathbf{y} = (y_1, y_2)$. The arrays' horizontal and vertical axes, from left to right and bottom to top, correspond to the vector components y_1 and y_2, respectively, and both range over the values $\{-7, \ldots, 7\}$. A bright pixel indicates a large value of $\widetilde{L}_{\mathbf{k},\mathbf{y}}$, whereas a dark pixel represents a low value.

With ν, c_1, and n_0 being positive constants, we find that λ is negative over the whole range of k_1, if the parameter ϑ^e is chosen large enough. Since the plotted eigenvalue λ is the largest eigenvalue of (4.28), the remaining eigenvalues are then negative as well and so the learning dynamics are stable. As the graphs show, λ is maximal for $k_1 = 0$. Therefore, the first eigenmode to become unstable with decreasing ϑ^e corresponds to $k_1 = 0$, i.e., to the wave vector $\mathbf{k} = 0$. The leftmost gey-level plot in either panel shows that this eigenmode is rotationally symmetric. Presuming that the final state of the learning process is dominated by the leading eigenmode it follows that the emerging pattern of geniculocortical connectivity is homogeneous ($\mathbf{k} = 0$) and isotropic.

λ for given k_1 is small, if ϑ^{e} is large, and vice versa. Therefore λ is negative for all values of k_1 if ϑ^{e} is chosen large enough. Since the plotted eigenvalue λ is the largest eigenvalue of (4.28), the remaining eigenvalues are then negative as well and so the learning dynamics are stable and converge towards the fixed point. On the other hand, if ϑ^{e} is chosen sufficiently small, then λ becomes positive at least within a limited range of k_1. In that case the synaptic dynamics are unstable.

Second, we find that λ is maximal for $k_1 = 0$. The first eigenvector to become unstable with decreasing ϑ^{e} thus corresponds to $k_1 = 0$, i.e., to the wave vector $\mathbf{k} = 0$. The leftmost gey-level plots in either panel of Fig. 4.2 show this eigenmode to be rotationally symmetric. Presuming that the final state of the learning process is dominated by the leading[2] eigenmode we conclude that the emerging connectivity pattern is homogeneous ($\mathbf{k} = 0$) and rotationally symmetric and hence does not establish an orientation map.

Third, the grey-level plots of the leading eigenvectors show that all components within the maximal arborization radius are positive – the corresponding pixels appear in white or in a light grey. Hence, we conclude that when the learning dynamics become unstable and the leading eigenmode begins to grow, then the geniculocortical synapses are either all strengthened simultaneously or they are all weakened. As a consequence, the total amount of geniculate input to the cortical network layer is either continuously increasing or continuously decreasing. This process is, of course, accompanied by an according change of the mean cortical activity.

The described behaviour of the synaptic dynamics changes completely when the parameter σ^{e} is decreased sufficiently, and it is easy to understand why. The value of σ^{e} determines the amount by which a synapse's efficacy is modified in response to a postsynaptic spike. It thus follows that all the dendritic synapses of a given neuron are modified by the same amount whenever this neuron emits a spike. Choosing a negative value of σ^{e} therefore establishes a negative feedback loop in the learning dynamics. When a neuron's total synaptic input is increasing then its increasing firing rate results in a growing contribution of σ^{e} to the synaptic weights, which in turn counteracts the growth of the input efficacies. Conversely, a declining synaptic input causes the cell's firing rate to decrease. This reduces the effect of σ^{e} and thereby opposes the decrease of synaptic input.

A negative value of σ^{e} can thus stabilize the sum of synaptic input converging onto a single neuron. As a result, different input synapses of the same neuron have to compete for their efficacy. The weight of one synapse can be increased only at the expense of another synapse because the sum of all synaptic weights must remain constant.

The above view is supported by an analysis of the eigenvalue problem (4.28) for negative σ^{e}. It turns out that the eigenvalue of the eigenmode displayed in Fig. 4.2 decreases with decreasing σ^{e}. A negative value of σ^{e} can

[2]Subsequently, the eigenvalue having the largest real part will be called the "leading" eigenvalue or the "principal" eigenvalue. Accordingly, the related eigenmode will be called the "leading" or "principal" eigenmode.

thus stabilize this eigenmode. On the other hand an inspection of Eq. (4.28) shows that the only terms involving the parameter σ^{e} are proportional to a spatial sum over the eigenvector, $\sum_{\mathbf{y}'} \widetilde{L}_{\mathbf{k},\mathbf{y}'}$. Consequently, a variation of σ^{e} does not affect eigenmodes with a sum equal to zero. We may therefore expect the eigenvalue of the rotationally symmetric eigenmode shown in Fig. 4.2 to decrease below the eigenvalues of the leading zero-sum eigenvectors when σ^{e} becomes small enough.

Figures 4.3 through 4.5 demonstrate that this is true. Each figure is related to one arborization model and presents the leading eigenmodes for $\sigma^{\mathrm{e}} = -0.5$ in dependence on the wave vector $\mathbf{k}$. Because of the rotational symmetry of Eq. (4.28) it is again sufficient to consider only one direction of $\mathbf{k}$, say, $\mathbf{k} = (k_1, 0)$. In the top panel of each figure the four largest eigenvalues are plotted as functions of k_1. Below this plot the four corresponding eigenvectors are displayed in a grey-level representation for three different values of k_1. The plots have been arranged in accordance with the sequence of the eigenvalues so that the leading eigenvector appears in the uppermost row.

As in the case $\sigma^{\mathrm{e}} = 0$ all eigenvalues are negative if the model parameter ϑ^{e} is chosen sufficiently large (see the expressions along the graphs' vertical axes). The first eigenvectors to become unstable when ϑ^{e} is reduced, however, are no longer rotationally symmetric and correspond to non-zero wave vectors with $|\mathbf{k}| \approx \pi/6$. Provided that the final state of the learning process is dominated by the leading eigenmodes, the resulting geniculocortical connectivity pattern is determined by a linear superposition of these eigenvectors.

One of the most simple cortical maps would thus emerge from a superposition of the modes $\widetilde{L}_{\mathbf{k},\mathbf{y}}$ and $\widetilde{L}_{-\mathbf{k},\mathbf{y}}$ with $\mathbf{k} = (\pi/6, 0)$. In order to obtain a spatial representation $J_{\mathbf{x},\mathbf{y}}$ of this map we perform a Fourier back-transformation from $\mathbf{k}$ to the cortical coordinate $\mathbf{x}$ and find

$$\begin{aligned} J_{\mathbf{x},\mathbf{y}} &\sim \widetilde{L}_{\mathbf{k},\mathbf{y}} \exp\left(i\mathbf{k}\mathbf{x} + i\phi\right) + \widetilde{L}_{-\mathbf{k},\mathbf{y}} \exp\left(-i\mathbf{k}\mathbf{x} - i\phi\right) \\ &= 2\,\mathrm{Re}[\widetilde{L}_{\mathbf{k},\mathbf{y}}]\,\cos\left(k_1 x_1 + \phi\right) - 2\,\mathrm{Im}[\widetilde{L}_{\mathbf{k},\mathbf{y}}]\,\sin\left(k_1 x_1 + \phi\right) \\ &= 2\,\mathrm{Re}[\widetilde{L}_{\mathbf{k},\mathbf{y}} \exp\left(i\mathbf{k}\mathbf{x} + i\phi\right)]\,, \end{aligned}$$

where ϕ denotes an arbitrary constant phase. The second and the third equalities of the above expression arise from the fact that $\widetilde{L}_{-\mathbf{k},\mathbf{y}}$ is the complex conjugate of $\widetilde{L}_{\mathbf{k},\mathbf{y}}$. Small sections of the connectivity pattern $J_{\mathbf{x},\mathbf{y}}$ with $x_1 = 0, \ldots, 7$ and $\phi = 0$ are shown at the bottom of the Figs. 4.3 through 4.5. Such a section consists of a series of 8 small grey-level plots, each of which is a representation of $J_{\mathbf{x},\mathbf{y}}$ as a function of $\mathbf{y}$ for fixed $\mathbf{x}$. These small grey-level plots can be regarded as the receptive fields of neighbouring cortical cells. Obviously, the emerging receptive fields are orientation selective with their preferred orientation perpendicular to the direction of the wave vector $\mathbf{k}$. This is in full agreement with the results of earlier analytical investigations that have been carried out on the basis of graded-response neurons (Wimbauer 1996, Wimbauer et al. 1998).

Of course, the actually developing orientation map will look much more complicated for several reasons. First, all eigenmodes corresponding to wave

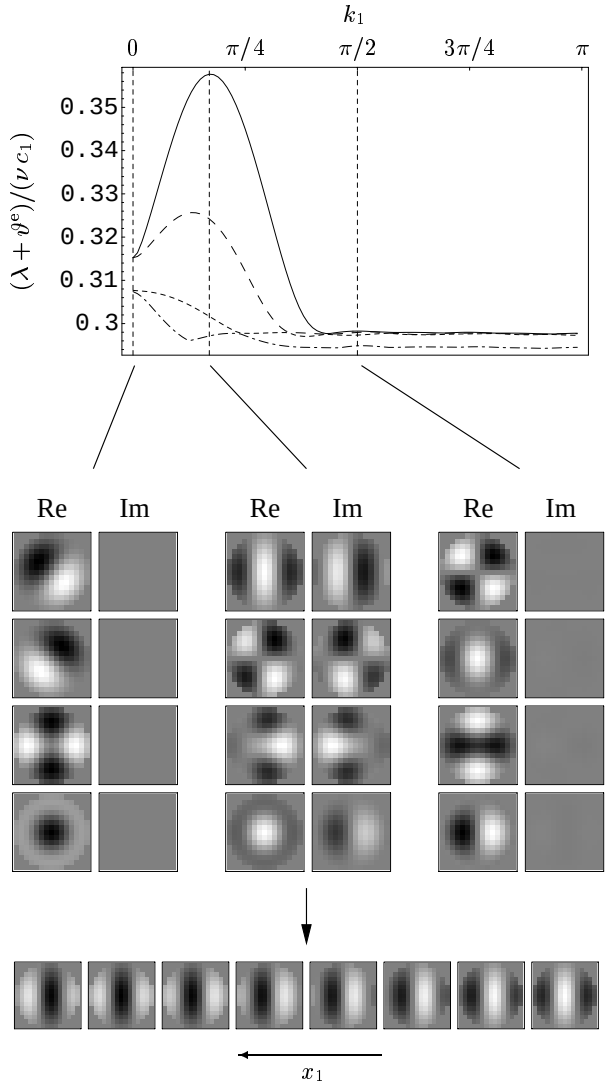


Figure 4.3: The leading eigenmodes are presented in dependence on the wave vector $\mathbf{k}$ for the flat arborization model and with $\sigma^{\mathrm{e}} = -0.5$. Because of the rotational symmetry of Eq. (4.28) it is again sufficient to consider only one direction of $\mathbf{k}$ as, e.g., $\mathbf{k} = (k_1, 0)$. In the top panel the four largest eigenvalues are plotted as functions of k_1. Below this plot the four corresponding eigenvectors are displayed in a grey-level representation for three specific values of k_1. Each grey-level plot consists of two arrays of 15×15 pixels that visualize separately the real part and the imaginary part of the respective eigenvector; cf. Fig. 4.2. The plots have been arranged in accordance with the sequence of the eigenvalues so that the leading eigenvector appears in the uppermost row. At $k_1 \approx \pi/6$ the largest eigenvalue reaches its maximum. The related eigenvector $\widetilde{L}_{\mathbf{k},\mathbf{y}}$ is shown at the top of the middle column of grey-level plots. A Fourier back-transformation of this mode from $\mathbf{k}$ to cortical position $\mathbf{x}$ yields the corresponding spatial eigenmode $L_{\mathbf{x},\mathbf{y}} \sim \widetilde{L}_{\mathbf{k},\mathbf{y}} \exp(i\mathbf{kx})$, whose real part is displayed in dependence on x_1 at the bottom of the figure. This series of plots can be regarded as a small section of a cortical orientation map with each of the small squares representing the receptive field of one cortical cell.

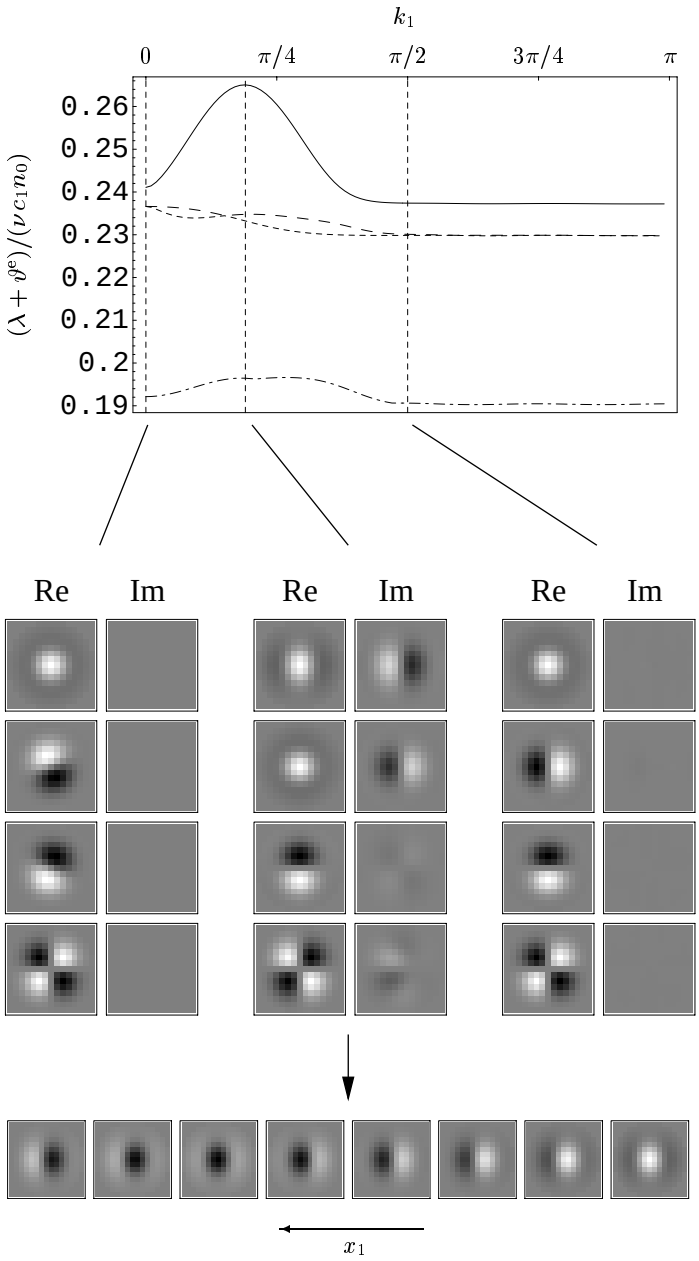


Figure 4.4: As Fig. 4.3 but for the dense arborization model.

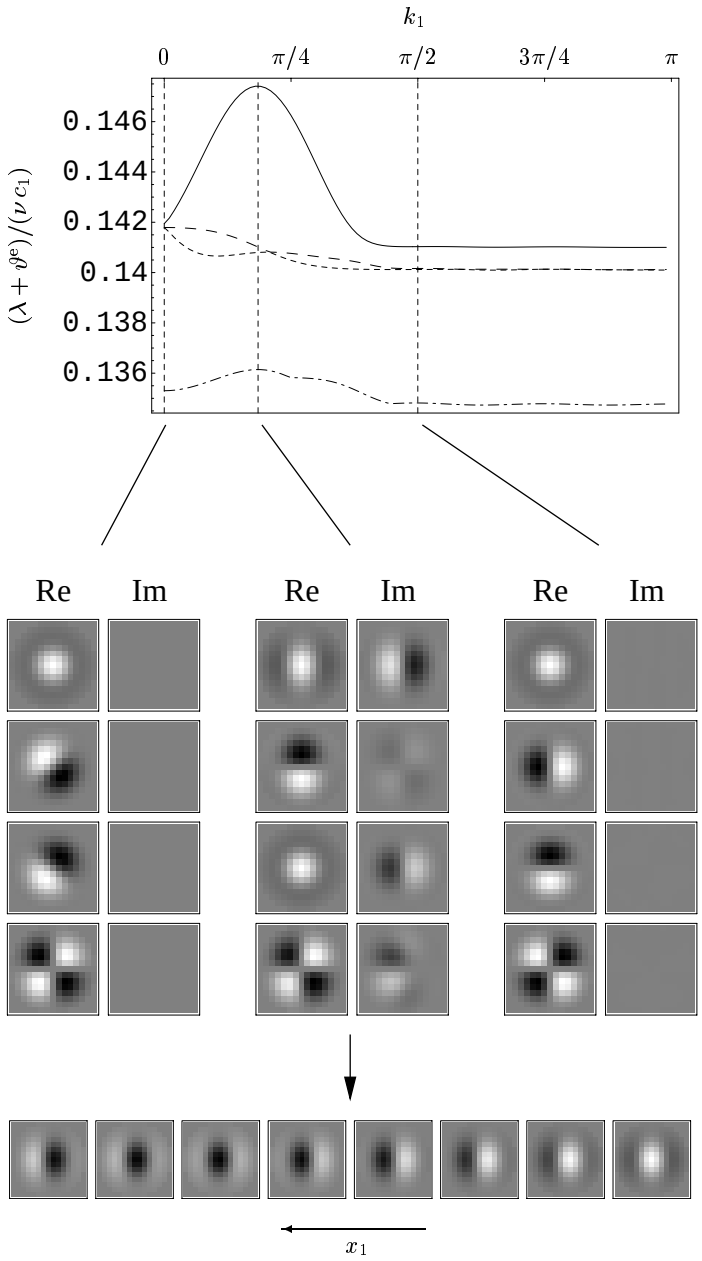


Figure 4.5: As Fig. 4.3 but for the sparse arborization model.

vectors of the same length are degenerate (cf. footnote 1 on p. 115). Therefore, the resulting connectivity pattern will be determined by a superposition of eigenmodes of many different orientations and phases. Second, variations of the length of the wave vector near the maximum of the leading eigenvalue induces only small variations of this eigenvalue. As a consequence, a whole range of wave vectors will contribute to the final map. Third, various non-linearities of the learning dynamics that have been neglected in the present analysis may have a significant influence on the emerging connectivity pattern. For example, a non-linearity that takes a very prominent role in our numerical simulations is established by the upper and lower bounds of the synaptic weights (see Sect. 2.3).

Non-Isotropic Intracortical Connectivity

Let us now investigate how the presence of an intracortical orientation map can guide the development of the geniculocortical connectivity as we have found in our numerical simulations. To this end we choose an intracortical connectivity

$$c_1 J_{\mathbf{y}}^{\mathrm{ic}} := 0.75\,\mathrm{s}^{-1}\ \exp\left[-|\mathbf{y}|^2/(2\cdot 4.5^2)\right]\ [1-\sin(\pi y_1/7)] \qquad (4.29)$$

for $|\mathbf{y}| < 7.5$ and vanishing for $|\mathbf{y}| \geq 7.5$. This yields an extremely simple intracortical map with all neurons exhibiting the same vertically oriented lateral receptive field. Applying the color code shown in Figs. 2.6 and 2.7 the colored representation of this map would appear in a uniform light blue.

Inserting the definition (4.29) into Eq. (4.28) we arrive at an eigenvalue problem that is no longer rotationally symmetric. As in the previous subsection it turns out that for $\sigma^{\mathrm{e}} = 0$ and decreasing ϑ^{e} the first eigenmode to become unstable corresponds to a homogeneous pattern of geniculocortical connectivity with all synaptic weights either growing or shrinking simultaneously. We have already explained what that means: The total geniculate input to each cortical neuron and to the cortical layer as a whole is unstable. We have also argued that this problem can be reconciled by choosing a sufficiently small (i.e., negative) value σ^{e}: The total input to each cortical cell is stabilized so that individual synapses have to compete for their efficacy. In the following, we let $\sigma^{\mathrm{e}} := -0.5$.

For each model of neuronal arborization, Fig. 4.6 displays the real part of the leading eigenvalue λ in dependence on the cortical wave vector $\mathbf{k}$. Since the solutions of (4.28) for a wave vector $-\mathbf{k}$ are the complex conjugate of the solutions for the wave vector $\mathbf{k}$, the plots have been restricted to one half-plane of $\mathbf{k}$. Obviously, the rotational symmetry is broken so that the maximum of the real part of λ along the k_1-axis is larger than its maximum along the k_2-axis. This is a consequence of the anisotropy of the intracortical connectivity. Accordingly, the first eigenmodes to become unstable with decreasing ϑ^{e} correspond to the wave vectors $\mathbf{k} = \pm(k_1, 0)$ with $k_1 \approx \pi/6$. When ϑ^{e} is further reduced, additional eigenmodes are destabilized, which are related to oblique directions and different lengths of $\mathbf{k}$. At last, eigenmodes corresponding to wave vectors directed parallel to the k_2-axis will become unstable.

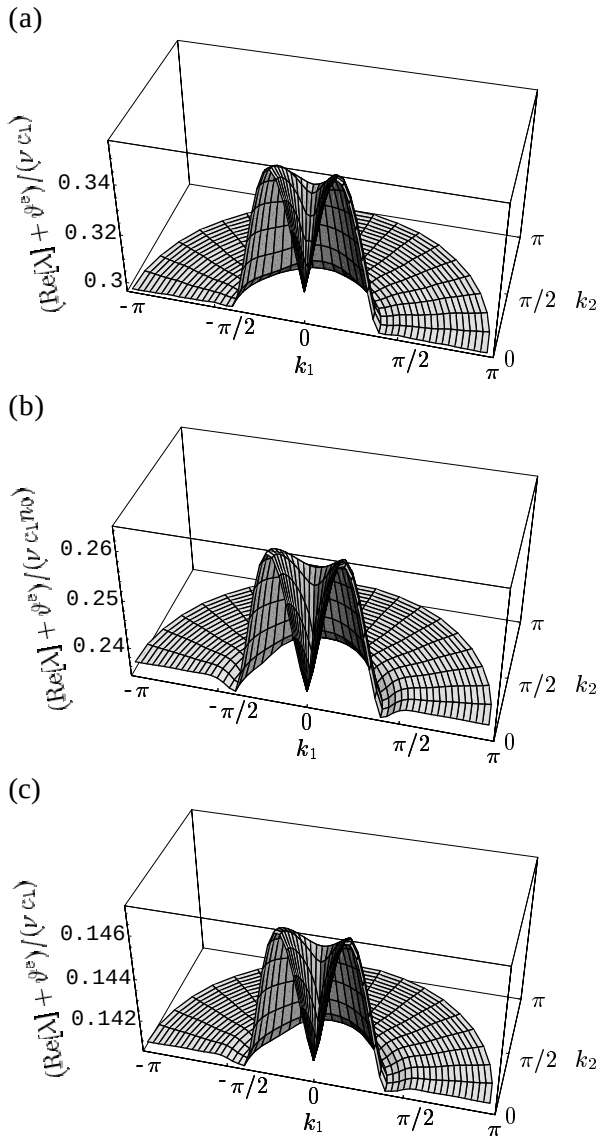


Figure 4.6: The real part of the leading eigenvalue λ of Eq. (4.28) is shown as a function of the wave vector $\mathbf{k}$ for $\sigma^{\mathrm{e}} = -0.5$ and the non-isotropic intracortical connectivity defined in (4.29). Each panel of the figure corresponds to one arborization model, viz., flat (a), dense (b), and sparse arborization (c). Because of the anisotropy of the intracortical connectivity the rotational symmetry is broken so that the maximal value of $\mathrm{Re}[\lambda]$ along the k_1-axis is larger than along the k_2-axis. In the following, we will therefore investigate more closely the eigenvalues and eigenvectors along these two directions of the wave vector.

Subsequently, we will investigate more closely the two most prominent directions of $\mathbf{k}$, namely the directions parallel to the k_1-axis and the k_2-axis. The real parts of the four leading eigenvalues with $\mathbf{k}$ varying along these directions are displayed in the Figs. 4.7, 4.8, and 4.9. In addition, grey-level plots of the four leading eigenvectors $\widetilde{L}_{\mathbf{k},\mathbf{y}}$ are shown for a few specific choices of $\mathbf{k}$. Each figure corresponds to one arborization model and consists of two panels presenting the results for $\mathbf{k} := (k_1, 0)$ (upper panel) and $\mathbf{k} := (0, k_2)$ (lower panel). As in the Figs. 4.3 through 4.5 the grey-level plots are ordered in such a way that their sequence is in accord with the sequence of the eigenvalues' real parts. Hence, the eigenvector corresponding to the leading eigenvalue is plotted at the top whereas the eigenmode corresponding to the fourth eigenvalue is plotted at the bottom.

In agreement with Fig. 4.6 we find that the real part of the leading eigenvalue λ reaches its maximum at $\mathbf{k} \approx \pm(\pi/6, 0)$ along the k_1-axis and at

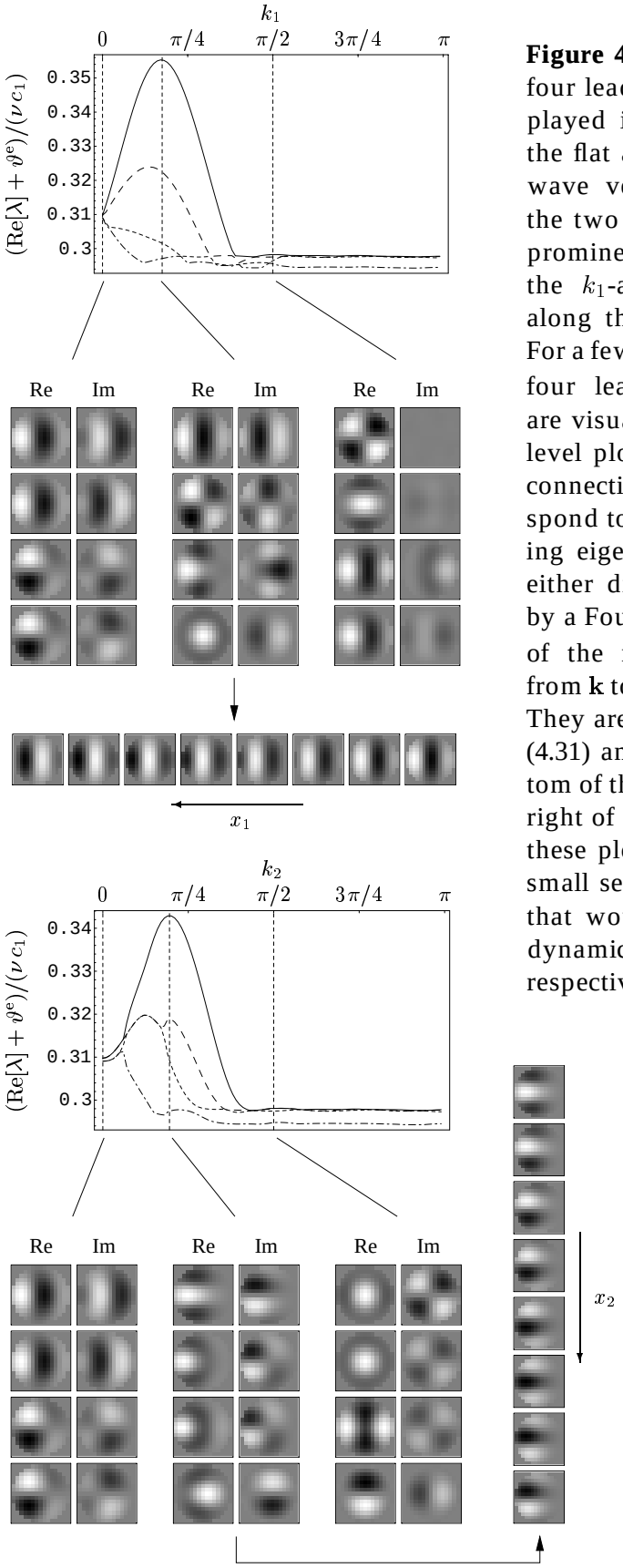


Figure 4.7: The real parts of the four leading eigenvalues are displayed in dependence on $\mathbf{k}$ for the flat arborization model. The wave vector $\mathbf{k}$ is varied along the two directions that are most prominent in Fig. 4.6, viz., along the k_1-axis (upper panel) and along the k_2-axis (lower panel). For a few specific choices of $\mathbf{k}$, the four leading eigenvectors $\widetilde{L}_{\mathbf{k},\mathbf{y}}$ are visualized by means of grey-level plots. The geniculocortical connectivity patterns that correspond to the maxima of the leading eigenvalue's real part along either direction can be obtained by a Fourier back-transformation of the related eigenmode $\widetilde{L}_{\mathbf{k},\mathbf{y}}$ from $\mathbf{k}$ to the spatial coordinate $\mathbf{x}$. They are given in Eqs. (4.30) and (4.31) and are plotted at the bottom of the upper panel and to the right of the lower panel. Each of these plots can be regarded as a small section of the cortical map that would arise if the learning dynamics were dominated by the respective eigenmode.

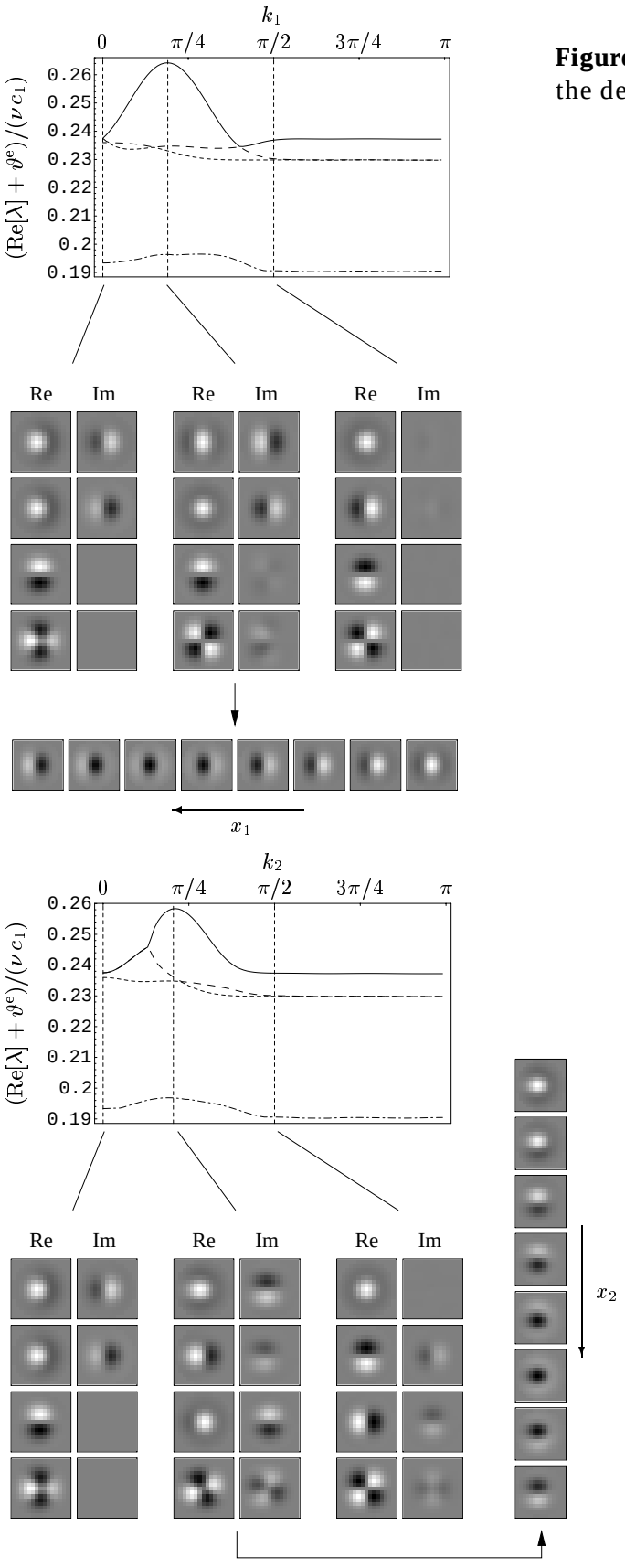


Figure 4.8: As Fig. 4.7 but for the dense arborization model.

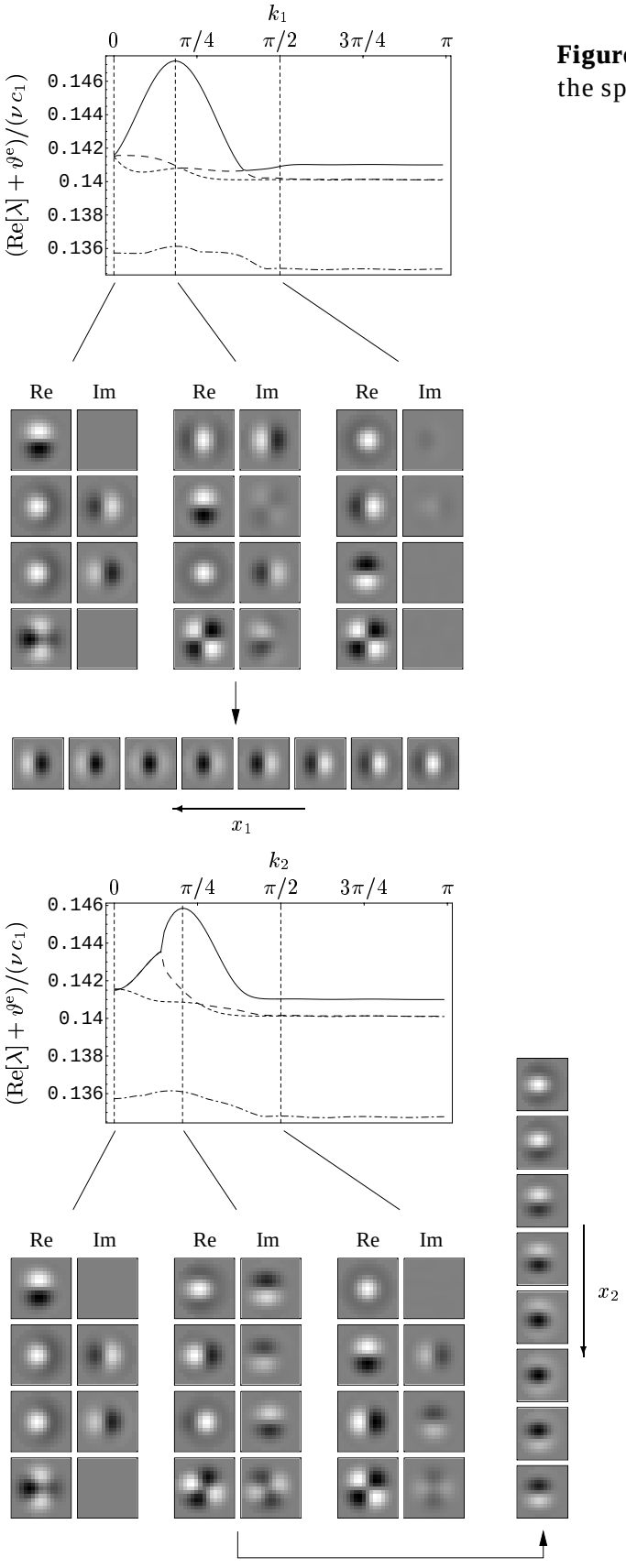


Figure 4.9: As Fig. 4.7 but for the sparse arborization model.

$\mathbf{k} \approx \pm(0, \pi/6)$ along the k_2-axis. Importantly, the maximal value along the k_2-axis is lower than that along the k_1-axis. Since we are interested in the typical layout of the orientation maps emerging as a result of the learning process, we perform a Fourier back-transformation of the respective eigenvectors $\widetilde{L}_{\mathbf{k},\mathbf{y}}$ so as to obtain their spatial representations over the cortex. With ϕ denoting an arbitrary constant phase we thus get the connectivity pattern

$$J^{(1)}_{\mathbf{x},\mathbf{y}} \sim \mathrm{Re}[\widetilde{L}_{\mathbf{k},\mathbf{y}}] \cos(k_1 x_1 + \phi) - \mathrm{Im}[\widetilde{L}_{\mathbf{k},\mathbf{y}}] \sin(k_1 x_1 + \phi) \ , \tag{4.30}$$

for $\mathbf{k} = (k_1, 0)$ and, analogously,

$$J^{(2)}_{\mathbf{x},\mathbf{y}} \sim \mathrm{Re}[\widetilde{L}_{\mathbf{k},\mathbf{y}}] \cos(k_2 x_2 + \phi) - \mathrm{Im}[\widetilde{L}_{\mathbf{k},\mathbf{y}}] \sin(k_2 x_2 + \phi) \ , \tag{4.31}$$

for $\mathbf{k} = (0, k_2)$. As to the flat arborization model, these connectivity patterns are displayed in Fig. 4.7. At the bottom of the upper panel $J^{(1)}_{\mathbf{x},\mathbf{y}}$ is shown by means of a series of grey-level plots. Each of them visualizes $J^{(1)}_{\mathbf{x},\mathbf{y}}$ in dependence on $\mathbf{y}$ for one fixed value of x_1. Between subsequent plots x_1 is increased in steps of 1. They can therefore be regarded as the receptive fields of neighbouring neurons that would emerge if the learning dynamics were dominated by the corresponding eigenmode $\widetilde{L}_{\mathbf{k},\mathbf{y}}$. In an analogous manner, the connectivity pattern $J^{(2)}_{\mathbf{x},\mathbf{y}}$ is shown to the right of the lower panel. Similar plots can be found in Figs 4.8 and 4.9 for the dense and the sparse arborization model.

All these plots show that the neuronal receptive fields arising from the connectivity pattern $J^{(1)}_{\mathbf{x},\mathbf{y}}$ are orientation selective and prefer vertically oriented stimuli, whereas the receptive fields corresponding to $J^{(2)}_{\mathbf{x},\mathbf{y}}$ prefer horizontally oriented stimuli. Provided that the final state of the learning dynamics is determined by a superposition of the leading unstable eigenmodes we arrive at the following view of pattern formation in the presence of the anisotropic intracortical connectivity given in Eq. (4.29).

On the one hand, the value of ϑ^{e} can be chosen in such a way that eigenmodes corresponding to a range of wave vectors parallel and oblique to the k_1 axis are unstable, while all eigenvectors corresponding to wave vectors parallel to the k_2-axis are stable. In this case the emerging orientation map will be dominated by receptive fields preferring vertical and oblique orientations because none of the eigenmodes corresponding to horizontally oriented receptive fields is unstable. However, such a scenario requires a relatively fine tuning of the parameter ϑ^{e} and may therefore seem unrealistic.

On the other hand, the value of ϑ^{e} may be a bit smaller so that eigenmodes corresponding to all directions of $\mathbf{k}$ are unstable. Then the developing receptive fields can be arbitrarily oriented. Nevertheless, the orientation map will be biased towards vertical orientations because the real parts of eigenvalues corresponding to wave vectors parallel to the k_1-axis are larger than the real parts of eigenvalues corresponding to the k_2-direction. For a given wavelength the growth rate of a vertically oriented connectivity pattern will therefore be larger than that of a horizontally oriented connectivity pattern. Moreover, the range of wavelengths corresponding to unstable eigenmodes is larger along

the k_1-axis than along the k_2-axis. As a consequence, a random superposition of unstable eigenvectors will comprise more eigenmodes corresponding to vertically oriented connectivity patterns than eigenmodes corresponding to horizontally oriented connectivity patterns.

We conclude that in our reduced network model the presence of a homogeneous but vertically oriented intracortical connectivity during the period of geniculocortical plasticity leads to the emergence of an orientation map that is biased towards vertical orientations. Hence, the developing geniculocortical map tends to be in accord with the intracortical "map", a finding that is in full agreement with the results of our numerical simulations presented in Sects. 2.6 and 2.7. The above analytical investigations thus demonstrate how an existing intracortical map can guide the development of the geniculocortical connectivity.

4.4 Plasticity of Intracortical Synapses

In the previous section we have seen how in our reduced model of primary visual cortex a feedforward orientation map can arise from Hebbian plasticity of geniculocortical synapses and how the layout of this map can be influenced by the structure of intracortical connectivity. The numerical simulations presented in Sect. 2.5 have shown that under appropriate conditions a Hebbian development of intracortical synapses – in the absence of feedforward input – leads to the emergence of a map-like structure of oriented connectivity patterns.

Here we will present an analytical investigation of intracortical plasticity on the basis of the reduced network model established in Sect. 4.1. As we want to study the development of lateral connections in the absence of input from the LGN we further simplify the reduced network by removing its geniculate layer. Hence, we are left with only one layer of linear stochastic spike response-neurons representing excitatory cortical cells.

Our analysis of the learning dynamics in this network will lead us to an eigenvalue problem quite similar to the one we have obtained for the feedforward dynamics. Despite this similarity, however, we will also find a fundamental difference between the dynamics of feedforward and intracortical synaptic plasticity, which is probably the reason for the fact that we could never observe a formation of an intracortical orientation map in numerical simulations using the reduced network model (results not shown). The following investigations will thus reveal the limitations of the reduced model and indicate how these limitations are overcome in the full network, in which the emergence of intracortical orientation maps is a robust phenomenon.

4.4.1 The Learning Equation

As in the previous section we use the stochastic spike response model to represent individual neurons and define their linear activation function as

$$q(h) := c_0 + c_1\, h\ .$$

Neglecting the refractory potential, we have the coupling operator

$$[\hat{\kappa}a]_i(t) := \sum_j J_{ij} \int_{-\infty}^{\infty} a_j(t')\,\varepsilon(t-t')\,\mathrm{d}t' ,$$

where J_{ij} denotes the total synaptic weight connecting the presynaptic cell j to the postsynaptic cell i and $\varepsilon(t)$ is the postsynaptic potential,

$$\varepsilon(t) := \exp(-t/\tau_\varepsilon)$$

for $t > 0$ and vanishing otherwise.

Since we are studying the intracortical plasticity in the absence of feedforward input, we need not account for an external membrane potential when calculating the mean activities and the spike-spike correlations so that we can use Eqs. (3.46) and (3.47) instead of the more complicated expressions (3.73) through (3.75). Applying the von-Neumann series up to second order to approximate the required inverse operators we have

$$\alpha_i := \langle a_i(t)\rangle = c_0 + c_1 \sum_j J_{ij}\,\tau_\varepsilon\,c_0 + c_1^2 \sum_{jk} J_{ij}J_{jk}\,\tau_\varepsilon^2\,c_0 , \tag{4.32}$$

and

$$\begin{aligned}
&\langle a_i(t_1)a_j(t_2)\rangle = \\
&= \alpha_i\alpha_j + \delta_{ij}\,\delta(t_1-t_2)\,\alpha_i + c_1\,J_{ji}\,\alpha_i\,\varepsilon(t_2-t_1) + c_1\,J_{ij}\,\alpha_j\,\varepsilon(t_1-t_2) \\
&+ \theta(t_2-t_1)\,c_1^2 \sum_k J_{jk}J_{ki}\,\alpha_i \int_{-\infty}^{\infty} \mathrm{d}t'\varepsilon(t_2-t')\,\varepsilon(t'-t_1) \\
&+ \theta(t_2-t_1)\,c_1^2 \sum_k J_{jk}J_{ik}\,\alpha_k \int_{-\infty}^{\infty} \mathrm{d}t'\varepsilon(t_2-t')\,\varepsilon(t_1-t') \\
&+ [1-\theta(t_2-t_1)]\,c_1^2 \sum_k J_{ik}J_{kj}\,\alpha_j \int_{-\infty}^{\infty} \mathrm{d}t'\varepsilon(t_1-t')\,\varepsilon(t'-t_2) \\
&+ [1-\theta(t_2-t_1)]\,c_1^2 \sum_k J_{ik}J_{jk}\,\alpha_k \int_{-\infty}^{\infty} \mathrm{d}t'\varepsilon(t_1-t')\,\varepsilon(t_2-t') ,
\end{aligned} \tag{4.33}$$

with $\theta(t)$ denoting Heaviside's step function as defined in (4.15).

Next we carry out the integrals on the right hand side of Eq. (4.33) and substitute α_i from Eq. (4.32). We consider only terms of up to quadratic order in the synaptic weights and obtain

$$\begin{aligned}
&\langle a_i(t_1)a_j(t_2)\rangle = \\
&= c_0^2 + c_0^2c_1\tau_\varepsilon \sum_k J_{ik} + c_0^2c_1\tau_\varepsilon \sum_k J_{jk} \\
&+ c_0^2c_1^2\tau_\varepsilon^2 \sum_{kl} J_{ik}J_{jl} + c_0^2c_1^2\tau_\varepsilon^2 \sum_{kl} J_{ik}J_{kl} + c_0^2c_1^2\tau_\varepsilon^2 \sum_{kl} J_{jk}J_{kl} \\
&+ \delta_{ij}\,\delta(t_1-t_2)\left(c_0 + c_0c_1\tau_\varepsilon \sum_k J_{ik} + c_0c_1^2\tau_\varepsilon^2 \sum_{kl} J_{ik}J_{kl}\right)
\end{aligned}$$

$$
\begin{aligned}
&+ \quad \theta(t_2-t_1) c_0 c_1 J_{ji} \exp\left(-\frac{t_2-t_1}{\tau_\varepsilon}\right) \\
&+ \quad [1-\theta(t_2-t_1)]\, c_0 c_1 J_{ij} \exp\left(-\frac{t_1-t_2}{\tau_\varepsilon}\right) \\
&+ \quad \theta(t_2-t_1) c_0 c_1^2 \tau_\varepsilon J_{ji} \sum_k J_{ik} \exp\left(-\frac{t_2-t_1}{\tau_\varepsilon}\right) \\
&+ \quad [1-\theta(t_2-t_1)]\, c_0 c_1^2 \tau_\varepsilon J_{ij} \sum_k J_{jk} \exp\left(-\frac{t_1-t_2}{\tau_\varepsilon}\right) \\
&+ \quad \theta(t_2-t_1)\, c_0 c_1^2 \sum_k J_{jk} J_{ki} \cdot (t_2-t_1) \exp\left(-\frac{t_2-t_1}{\tau_\varepsilon}\right) \\
&+ \quad [1-\theta(t_2-t_1)]\, c_0 c_1^2 \sum_k J_{ik} J_{kj} \cdot (t_1-t_2) \exp\left(-\frac{t_1-t_2}{\tau_\varepsilon}\right) \\
&+ \quad c_0 c_1^2 \frac{\tau_\varepsilon}{2} \sum_k J_{jk} J_{ik} \exp\left(-\frac{|t_1-t_2|}{\tau_\varepsilon}\right) \; . \qquad (4.34)
\end{aligned}
$$

Finally, we insert (4.32) and (4.34) into the learning equation (4.5). Again we use an exponentially decaying time window

$$W(t) := \exp(-t/\tau_W)$$

for $t > 0$ and vanishing for $t \leq 0$. This yields

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} K_{ij} = &-\vartheta^{\mathrm{e}} K_{ij} + \nu \Bigg[\left(c_0^2 \tau_W + \sigma^{\mathrm{e}} c_0 + \zeta^{\mathrm{e}}\right) + c_0 c_1 \tau_* J_{ij} \qquad (4.35) \\
&+ \quad c_0 c_1^2 \tau_\varepsilon \tau_* J_{ij} \sum_k J_{jk} + \left(c_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} c_0 c_1 \tau_\varepsilon\right) \sum_k J_{ik} + c_0^2 c_1 \tau_\varepsilon \tau_W \sum_k J_{jk} \\
&+ \quad \left(c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} c_0 c_1^2 \tau_\varepsilon^2\right) \sum_{kl} J_{ik} J_{kl} + c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{kl} J_{jk} J_{kl} \\
&+ \quad c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{kl} J_{ik} J_{jl} + c_0 c_1^2 \tau_*^2 \sum_k J_{ik} J_{kj} + c_0 c_1^2 \frac{\tau_\varepsilon}{2} \tau_* \sum_k J_{jk} J_{ik} \Bigg] \; ,
\end{aligned}
$$

with

$$\tau_* := \frac{\tau_W \, \tau_\varepsilon}{\tau_W + \tau_\varepsilon} \; .$$

Since we are dealing with a network consisting of only one layer of cortical pyramidal cells, there is no need to distinguish different types of synapses. In (4.35) we have therefore dropped the upper indices 'e' from J_{ij} and K_{ij}. As explained in the previous section, J_{ij} is to be regarded as the effective weight linking cell j to cell i and is equal to the efficacy K_{ij} of a single synapse multiplied by the total number n_{ij} of synapses from j to i. Within the framework of our model, n_{ij} is taken to be constant in time. After multiplying Eq. (4.35) by n_{ij} and substituting

$$J_{ij} = n_{ij} K_{ij} \qquad (4.36)$$

we thus obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}n_{ij}K_{ij} = -\vartheta^{\mathrm{e}}\, n_{ij}K_{ij} + \nu\, n_{ij}\Big[\big(c_0^2\tau_W + \sigma^{\mathrm{e}}c_0 + \zeta^{\mathrm{e}}\big) + c_0c_1\tau_* n_{ij}K_{ij} \tag{4.37}$$
$$+ \; c_0c_1^2\tau_\varepsilon\tau_* n_{ij}K_{ij}\sum_k n_{jk}K_{jk} + \big(c_0^2c_1\tau_\varepsilon\tau_W + \sigma^{\mathrm{e}}c_0c_1\tau_\varepsilon\big)\sum_k n_{ik}K_{ik}$$
$$+ \; c_0^2c_1\tau_\varepsilon\tau_W\sum_k n_{jk}K_{jk} + \big(c_0^2c_1^2\tau_\varepsilon^2\tau_W + \sigma^{\mathrm{e}}c_0c_1^2\tau_\varepsilon^2\big)\sum_{kl} n_{ik}K_{ik}n_{kl}K_{kl}$$
$$+ \; c_0^2c_1^2\tau_\varepsilon^2\tau_W\sum_{kl} n_{jk}K_{jk}n_{kl}K_{kl} + c_0^2c_1^2\tau_\varepsilon^2\tau_W\sum_{kl} n_{ik}K_{ik}n_{jl}K_{jl}$$
$$+ \; c_0c_1^2\tau_*^2\sum_k n_{ik}K_{ik}n_{kj}K_{kj} + c_0c_1^2\frac{\tau_\varepsilon}{2}\tau_*\sum_k n_{jk}K_{jk}n_{ik}K_{ik}\Big]\;.$$

4.4.2 Modeling Neuronal Arborization

As a prerequisite for the further analysis of (4.37) we must specify the n_{ij}. Representing the number of synapses between a given pair of neurons, each n_{ij} may be regarded as the result of the overlap between the axonal arbor of the presynaptic cell j with the dendritic arbor of the postsynaptic cell i. In this way, the entirety of the n_{ij} establishes a set of boundary conditions for the learning dynamics. In close analogy to our propositions of the preceding section we will consider three possible models of neuronal arborization: flat arborization, dense arborization, and sparse arborization.

Flat Arborization

In this approach it is assumed that a cortical neuron i receives exactly one synaptic input from all those neighbours j whose retinotopic positions are within a certain circular area around the cortical neuron's position. Conversely, no input is provided from any neuron outside this area. Therefore, we write

$$n_{ij} = \begin{cases} 1 & \text{if } \mathrm{d}(i,j) < r_{\max}, \\ 0 & \text{otherwise,} \end{cases}$$

where $\mathrm{d}(i,j)$ denotes the distance between the grid positions of cells i and j. Because of the relation $n_{ij}^2 = n_{ij}$ the learning equation (4.37) takes the form

$$\frac{\mathrm{d}}{\mathrm{d}t}J_{ij} = -\vartheta^{\mathrm{e}}J_{ij} + \nu\, c_0c_1\tau_* J_{ij} + \nu\, c_0c_1^2\tau_\varepsilon\tau_* J_{ij}\sum_k J_{jk} \tag{4.38}$$
$$+ A_{ij}\Big[\big(c_0^2\tau_W + \sigma^{\mathrm{e}}c_0 + \zeta^{\mathrm{e}}\big) + \big(c_0^2c_1\tau_\varepsilon\tau_W + \sigma^{\mathrm{e}}c_0c_1\tau_\varepsilon\big)\sum_k J_{ik} + c_0^2c_1\tau_\varepsilon\tau_W\sum_k J_{jk}$$
$$+ \big(c_0^2c_1^2\tau_\varepsilon^2\tau_W + \sigma^{\mathrm{e}}c_0c_1^2\tau_\varepsilon^2\big)\sum_{kl} J_{ik}J_{kl} + c_0^2c_1^2\tau_\varepsilon^2\tau_W\sum_{kl} J_{jk}J_{kl}$$
$$+ \; c_0^2c_1^2\tau_\varepsilon^2\tau_W\sum_{kl} J_{ik}J_{jl} + c_0c_1^2\tau_*^2\sum_k J_{ik}J_{kj} + c_0c_1^2\frac{\tau_\varepsilon}{2}\tau_*\sum_k J_{jk}J_{ik}\Big]\;,$$

with

$$A_{ij} := \nu\, n_{ij} \tag{4.39}$$

and the effective weight J_{ij} as defined in (4.36).

Dense Arborization

Alternatively, the number n_{ij} of synapses from cell j to cell i may be regarded as a random quantity whose expectation value depends on the overlap between the respective axonal and dendritic arbors (Braitenberg and Schüz 1991). We presume that neuronal arborization is dense so that (i) there is at least one synapse between each pair of cells whose retinotopic distance is below a certain maximal radius and (ii) there is no connection between cells beyond this radius. Neglecting the statistical fluctuations we let n_{ij} be the expected number of synapses from j to i. Specifically, we choose a Gaussian arborization

$$n_{ij} = n_0 \exp\left[-\mathrm{d}(i,j)^2/(2\cdot\sigma_\mathrm{n}^2)\right]$$

for $\mathrm{d}(i,j) < r_\mathrm{max}$ and $n_{ij} = 0$ otherwise. This ansatz corresponds to the arborization used the numerical simulations. Defining the effective weight J_{ij} and an arbor function A_{ij} as in Eqs. (4.36) and (4.39), respectively, we thus obtain

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}J_{ij} = -\vartheta^\mathrm{e}J_{ij} + A_{ij}\Bigg[& c_0c_1\tau_*J_{ij} + c_0c_1^2\tau_\varepsilon\tau_*J_{ij}\sum_k J_{jk} \\ & + \left(c_0^2\tau_W + \sigma^\mathrm{e}c_0 + \zeta^\mathrm{e}\right) + \left(c_0^2c_1\tau_\varepsilon\tau_W + \sigma^\mathrm{e}c_0c_1\tau_\varepsilon\right)\sum_k J_{ik} + c_0^2c_1\tau_\varepsilon\tau_W\sum_k J_{jk} \\ & + \left(c_0^2c_1^2\tau_\varepsilon^2\tau_W + \sigma^\mathrm{e}c_0c_1^2\tau_\varepsilon^2\right)\sum_{kl} J_{ik}J_{kl} + c_0^2c_1^2\tau_\varepsilon^2\tau_W\sum_{kl} J_{jk}J_{kl} \\ & + c_0^2c_1^2\tau_\varepsilon^2\tau_W\sum_{kl} J_{ik}J_{jl} + c_0c_1^2\tau_*^2\sum_k J_{ik}J_{kj} + c_0c_1^2\frac{\tau_\varepsilon}{2}\tau_*\sum_k J_{jk}J_{ik}\Bigg]\ .\end{aligned} \tag{4.40}$$

Sparse Arborization

Let us now consider a network in which the expectation value n_{ij}^* of the random number n_{ij} of synapses between two neurons is very low, $n_{ij}^* \ll 1$. In this case we may assume that the random value n_{ij} is either 0 or 1, so that $n_{ij}^2 = n_{ij}$ as in the case of a flat arborization. If n_{ij}^* and K_{ij} vary sufficiently slow in the neighbourhood of i and j, then the sums on the right hand side of (4.37) are self-averaging, i.e.,

$$\begin{aligned}\sum_k n_{ik}K_{ik} &\approx \sum_k n_{ik}^*K_{ik}\ , \\ \sum_{kl} n_{ik}K_{ik}n_{kl}K_{kl} &\approx \sum_{kl} n_{ik}^*K_{ik}n_{kl}^*K_{kl}\ , \\ \sum_k n_{ik}K_{ik}n_{kj}K_{kj} &\approx \sum_k n_{ik}^*K_{ik}n_{kj}^*K_{kj}\ ,\end{aligned}$$

$$\sum_k n_{ik} K_{ik} n_{jk} K_{jk} \quad \approx \quad \sum_k n^*_{ik} K_{ik} n^*_{jk} K_{jk} \; .$$

By averaging both sides of (4.37) over a neighbourhood of neuron i we arrive at

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} J_{ij} &= -\vartheta^{\mathrm{e}} J_{ij} + \nu \, c_0 c_1 \tau_* J_{ij} + \nu \, c_0 c_1^2 \tau_\varepsilon \tau_* J_{ij} \sum_k J_{jk} \\ &+ A_{ij} \Bigg[\left(c_0^2 \tau_W + \sigma^{\mathrm{e}} c_0 + \zeta^{\mathrm{e}} \right) + \left(c_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} c_0 c_1 \tau_\varepsilon \right) \sum_k J_{ik} + c_0^2 c_1 \tau_\varepsilon \tau_W \sum_k J_{jk} \\ &+ \left(c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} c_0 c_1^2 \tau_\varepsilon^2 \right) \sum_{kl} J_{ik} J_{kl} + c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{kl} J_{jk} J_{kl} \\ &+ c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{kl} J_{ik} J_{jl} + c_0 c_1^2 \tau_*^2 \sum_k J_{ik} J_{kj} + c_0 c_1^2 \frac{\tau_\varepsilon}{2} \tau_* \sum_k J_{jk} J_{ik} \Bigg] \; , \end{aligned} \tag{4.41}$$

which is formally identical to Eq. (4.38) but where

$$\begin{aligned} A_{ij} &:= \nu \, n^*_{ij} \; , \\ J_{ij} &:= n^*_{ij} K_{ij} \; . \end{aligned}$$

The requirement that n^*_{ij} is a slow-varying function of the positions of cells i and j can be met easily if the network under consideration is large enough so that the radius of neuronal arborization extends over sufficiently many neurons. For the present analysis we will choose n^*_{ij} as a wide Gaussian. In contrast, the additional requirement that K_{ij} varies slowly can only be verified a posteriori when the emerging connectivity patterns are known. If they turn out to be sufficiently smooth, then the use of Eq. (4.41) is justified. Otherwise, the obtained results are probably useless.

4.4.3 Fixed Points and Linear Stability Analysis

In the preceding parts of this section we have first derived the basic learning equation (4.37) governing the dynamics of intracortical synaptic plasticity in our reduced model of primary visual cortex. Depending on the assumed distribution of intracortical synapses – the neuronal arborization – we have then obtained different dynamical equations describing the development of the effective intracortical connectivity, viz., Eqs. (4.38), (4.40), and (4.41). We are now going to investigate the basic fixed-point properties of these equations so as to obtain some insight into the emerging patterns of intracortical connectivity. For the sake of a short notation we introduce the summarizing formula

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} J_{ij} &= -\vartheta^{\mathrm{e}} J_{ij} + B_{ij} \left[c_0 c_1 \tau_* J_{ij} + c_0 c_1^2 \tau_\varepsilon \tau_* J_{ij} \sum_k J_{jk} \right] \\ &+ A_{ij} \Bigg[\left(c_0^2 \tau_W + \sigma^{\mathrm{e}} c_0 + \zeta^{\mathrm{e}} \right) + \left(c_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} c_0 c_1 \tau_\varepsilon \right) \sum_k J_{ik} \end{aligned} \tag{4.42}$$

$$
\begin{aligned}
&+ \quad c_0^2 c_1 \tau_\varepsilon \tau_W \sum_k J_{jk} + \left(c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} c_0 c_1^2 \tau_\varepsilon^2\right) \sum_{kl} J_{ik} J_{kl} \\
&+ \quad c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{kl} J_{jk} J_{kl} + c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{kl} J_{ik} J_{jl} \\
&+ \quad c_0 c_1^2 \tau_*^2 \sum_k J_{ik} J_{kj} + c_0 c_1^2 \frac{\tau_\varepsilon}{2} \tau_* \sum_k J_{jk} J_{ik} \Bigg] \ ,
\end{aligned}
$$

where we set $B_{ij} \equiv \nu$ for the flat or the sparse arborization and $B_{ij} = A_{ij}$ for the dense arborization.

Suppose we were given a fixed point of the above learning dynamics, i.e. a configuration $J_{ij} = J_{ij}^{\mathrm{FP}}$ that yields $(\mathrm{d}/\mathrm{d}t)J_{ij} = 0$ when inserted on the right hand side of (4.42). In order to determine the stability of this fixed point we study the temporal evolution of a small perturbation $L_{ij}(t)$ that we add to J_{ij}^{FP}. If it turns out that any small perturbation L_{ij} approaches zero (or remains constant) as t goes to infinity, then J_{ij}^{FP} is a stable fixed point. On the other hand, J_{ij}^{FP} is unstable if there is a perturbation L_{ij} that grows with time. In the latter case we are especially interested in the shapes of these growing perturbations because they may be related to the network's final connectivity pattern at the end of the learning process.

Inserting $J_{ij} = J_{ij}^{\mathrm{FP}} + L_{ij}$ into Eq. (4.42) we obtain

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} L_{ij} &= -\vartheta^{\mathrm{e}} L_{ij} + B_{ij} \left[c_0 c_1 \tau_* L_{ij} + c_0 c_1^2 \tau_\varepsilon \tau_* \sum_k \left(J_{jk}^{\mathrm{FP}} L_{ij} + J_{ij}^{\mathrm{FP}} L_{jk} \right) \right] \qquad (4.43) \\
&+ A_{ij} \Bigg[\left(c_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} c_0 c_1 \tau_\varepsilon\right) \sum_k L_{ik} + c_0^2 c_1 \tau_\varepsilon \tau_W \sum_k L_{jk} \\
&+ \left(c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} c_0 c_1^2 \tau_\varepsilon^2\right) \sum_{kl} \left(J_{ik}^{\mathrm{FP}} L_{kl} + J_{kl}^{\mathrm{FP}} L_{ik} \right) \\
&+ c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{kl} \left(J_{jk}^{\mathrm{FP}} L_{kl} + J_{kl}^{\mathrm{FP}} L_{jk} \right) + c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{kl} \left(J_{ik}^{\mathrm{FP}} L_{jl} + J_{jl}^{\mathrm{FP}} L_{ik} \right) \\
&+ c_0 c_1^2 \tau_*^2 \sum_k \left(J_{ik}^{\mathrm{FP}} L_{kj} + J_{kj}^{\mathrm{FP}} L_{ik} \right) + c_0 c_1^2 \frac{\tau_\varepsilon}{2} \tau_* \sum_k \left(J_{jk}^{\mathrm{FP}} L_{ik} + J_{ik}^{\mathrm{FP}} L_{jk} \right) \Bigg] \ ,
\end{aligned}
$$

where it has been taken into account that L_{ij} is small so that terms of second order in L_{ij} are negligible. Since (4.43) is a linear differential equation, its solution $L_{ij}(t)$ is determined by the eigenfunctions of its right hand side and the corresponding eigenvalues. If the real parts of all the eigenvalues are negative or zero, then J_{ij}^{FP} is a stable fixed point. Conversely, if the real part of some eigenvalue is greater than zero, then the corresponding eigenmode will grow exponentially with time so that J_{ij}^{FP} is unstable. Therefore, we have to solve the eigenvalue problem

$$
\begin{aligned}
\lambda L_{ij} &= -\vartheta^{\mathrm{e}} L_{ij} + B_{ij} \left[c_0 c_1 \tau_* L_{ij} + c_0 c_1^2 \tau_\varepsilon \tau_* \sum_k \left(J_{jk}^{\mathrm{FP}} L_{ij} + J_{ij}^{\mathrm{FP}} L_{jk} \right) \right] \qquad (4.44) \\
&+ A_{ij} \Bigg[\left(c_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} c_0 c_1 \tau_\varepsilon\right) \sum_k L_{ik} + c_0^2 c_1 \tau_\varepsilon \tau_W \sum_k L_{jk}
\end{aligned}
$$

$$
\begin{aligned}
&+ \left(c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} c_0 c_1^2 \tau_\varepsilon^2\right) \sum_{kl} \left(J_{ik}^{\mathrm{FP}} L_{kl} + J_{kl}^{\mathrm{FP}} L_{ik}\right) \\
&+ c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{kl} \left(J_{jk}^{\mathrm{FP}} L_{kl} + J_{kl}^{\mathrm{FP}} L_{jk}\right) + c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{kl} \left(J_{ik}^{\mathrm{FP}} L_{jl} + J_{jl}^{\mathrm{FP}} L_{ik}\right) \\
&+ c_0 c_1^2 \tau_*^2 \sum_{k} \left(J_{ik}^{\mathrm{FP}} L_{kj} + J_{kj}^{\mathrm{FP}} L_{ik}\right) + c_0 c_1^2 \frac{\tau_\varepsilon}{2} \tau_* \sum_{k} \left(J_{jk}^{\mathrm{FP}} L_{ik} + J_{ik}^{\mathrm{FP}} L_{jk}\right) \Bigg] \ .
\end{aligned}
$$

Subsequently, it will be convenient to make use of a slightly changed notation. In the same way as explained in the previous section, let us replace every neuronal index i by a two-dimensional vector $\mathbf{x}$ specifying the position of the corresponding neuron within the network. This is possible because the network under consideration is arranged as a (retinotopically ordered) two-dimensional grid so that each cell can be identified uniquely by specifying its position on this grid. Correspondingly, a synapse can be identified by giving the two-dimensional coordinates of its pre- and postsynaptic cells, $\mathbf{x}_2$ and $\mathbf{x}_1$. Here we will use the following convention. The postsynaptic neuron will be specified by means of its absolute position $\mathbf{x_1}$; the presynaptic neuron will be given by means of its relative position $\mathbf{y} := \mathbf{x}_1 - \mathbf{x}_2$ with respect to the postsynaptic cell. Hence, $J_{\mathbf{x},\mathbf{y}}$ will denote the synaptic weight connecting the presynaptic cell located at $\mathbf{x} - \mathbf{y}$ to the postsynaptic neuron located at $\mathbf{x}$. With the arbor functions A_{ij} and B_{ij} depending only on the distance between the two neurons i and j they can be written as $A_{\mathbf{y}}$ and $B_{\mathbf{y}}$, where $\mathbf{y}$ is the relative position of cell j with respect to i as introduced above.

In the new notation the eigenvalue problem (4.44) reads

$$
\begin{aligned}
\lambda L_{\mathbf{x},\mathbf{y}} \ = \ & B_{\mathbf{y}} \left[c_0 c_1 \tau_* L_{\mathbf{x},\mathbf{y}} + c_0 c_1^2 \tau_\varepsilon \tau_* \sum_{\mathbf{y}'} \left(J_{\mathbf{x},\mathbf{y}}^{\mathrm{FP}} L_{\mathbf{x}-\mathbf{y},\mathbf{y}'} + J_{\mathbf{x}-\mathbf{y},\mathbf{y}'}^{\mathrm{FP}} L_{\mathbf{x},\mathbf{y}}\right) \right] \\
& + A_{\mathbf{y}} \Bigg[\left(c_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} c_0 c_1 \tau_\varepsilon\right) \sum_{\mathbf{y}'} L_{\mathbf{x},\mathbf{y}'} + c_0^2 c_1 \tau_\varepsilon \tau_W \sum_{\mathbf{y}'} L_{\mathbf{x}-\mathbf{y},\mathbf{y}'} \\
& + \left(c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} c_0 c_1^2 \tau_\varepsilon^2\right) \sum_{\mathbf{y}',\mathbf{y}''} \left(J_{\mathbf{x},\mathbf{y}'}^{\mathrm{FP}} L_{\mathbf{x}-\mathbf{y}',\mathbf{y}''} + J_{\mathbf{x}-\mathbf{y}',\mathbf{y}''}^{\mathrm{FP}} L_{\mathbf{x},\mathbf{y}'}\right) \\
& + c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{\mathbf{y}',\mathbf{y}''} \left(J_{\mathbf{x}-\mathbf{y},\mathbf{y}'}^{\mathrm{FP}} L_{\mathbf{x}-\mathbf{y}-\mathbf{y}',\mathbf{y}''} + J_{\mathbf{x}-\mathbf{y}-\mathbf{y}',\mathbf{y}''}^{\mathrm{FP}} L_{\mathbf{x}-\mathbf{y},\mathbf{y}'}\right) \\
& + c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{\mathbf{y}',\mathbf{y}''} \left(J_{\mathbf{x},\mathbf{y}'}^{\mathrm{FP}} L_{\mathbf{x}-\mathbf{y},\mathbf{y}''} + J_{\mathbf{x}-\mathbf{y},\mathbf{y}''}^{\mathrm{FP}} L_{\mathbf{x},\mathbf{y}'}\right) \\
& + c_0 c_1^2 \tau_*^2 \sum_{\mathbf{y}'} \left(J_{\mathbf{x},\mathbf{y}'}^{\mathrm{FP}} L_{\mathbf{x}-\mathbf{y}',\mathbf{y}-\mathbf{y}'} + J_{\mathbf{x}-\mathbf{y}',\mathbf{y}-\mathbf{y}'}^{\mathrm{FP}} L_{\mathbf{x},\mathbf{y}'}\right) \\
& + c_0 c_1^2 \frac{\tau_\varepsilon}{2} \tau_* \sum_{\mathbf{y}'} \left(J_{\mathbf{x},\mathbf{y}'}^{\mathrm{FP}} L_{\mathbf{x}-\mathbf{y},\mathbf{y}'-\mathbf{y}} + J_{\mathbf{x}-\mathbf{y},\mathbf{y}'-\mathbf{y}}^{\mathrm{FP}} L_{\mathbf{x},\mathbf{y}'}\right) \Bigg] - \vartheta^{\mathrm{e}} L_{\mathbf{x},\mathbf{y}} \ .
\end{aligned}
\tag{4.45}
$$

Before we can approach this problem we must determine the relevant fixed point(s) J^{FP} of the learning dynamics by setting $(\mathrm{d}/\mathrm{d}t)J_{ij} = 0$ on the left hand

side of Eq. (4.42), which yields the fixed-point equation

$$
\begin{aligned}
0 = &-\vartheta^{\mathrm{e}} J^{\mathrm{FP}}_{\mathbf{x},\mathbf{y}} + B_{\mathbf{y}} \left[c_0 c_1 \tau_* J^{\mathrm{FP}}_{\mathbf{x},\mathbf{y}} + c_0 c_1^2 \tau_\varepsilon \tau_* J^{\mathrm{FP}}_{\mathbf{x},\mathbf{y}} \sum_{\mathbf{y}'} J^{\mathrm{FP}}_{\mathbf{x}-\mathbf{y},\mathbf{y}'} \right] \\
&+ A_{\mathbf{y}} \Bigg[\left(c_0^2 \tau_W + \sigma^{\mathrm{e}} c_0 + \zeta^{\mathrm{e}} \right) + \left(c_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} c_0 c_1 \tau_\varepsilon \right) \sum_{\mathbf{y}'} J^{\mathrm{FP}}_{\mathbf{x},\mathbf{y}'} \\
&+ c_0^2 c_1 \tau_\varepsilon \tau_W \sum_{\mathbf{y}'} J^{\mathrm{FP}}_{\mathbf{x}-\mathbf{y},\mathbf{y}'} + \left(c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} c_0 c_1^2 \tau_\varepsilon^2 \right) \sum_{\mathbf{y}',\mathbf{y}''} J^{\mathrm{FP}}_{\mathbf{x},\mathbf{y}'} J^{\mathrm{FP}}_{\mathbf{x}-\mathbf{y}',\mathbf{y}''} \\
&+ c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{\mathbf{y}',\mathbf{y}''} J^{\mathrm{FP}}_{\mathbf{x}-\mathbf{y},\mathbf{y}'} J^{\mathrm{FP}}_{\mathbf{x}-\mathbf{y}-\mathbf{y}',\mathbf{y}''} + c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{\mathbf{y}',\mathbf{y}''} J^{\mathrm{FP}}_{\mathbf{x},\mathbf{y}'} J^{\mathrm{FP}}_{\mathbf{x}-\mathbf{y},\mathbf{y}''} \\
&+ c_0 c_1^2 \tau_*^2 \sum_{\mathbf{y}'} J^{\mathrm{FP}}_{\mathbf{x},\mathbf{y}'} J^{\mathrm{FP}}_{\mathbf{x}-\mathbf{y}',\mathbf{y}-\mathbf{y}'} + c_0 c_1^2 \frac{\tau_\varepsilon}{2} \tau_* \sum_{\mathbf{y}'} J^{\mathrm{FP}}_{\mathbf{x},\mathbf{y}'} J^{\mathrm{FP}}_{\mathbf{x}-\mathbf{y},\mathbf{y}'-\mathbf{y}} \Bigg] \; .
\end{aligned}
$$

Unfortunately, this equation is non-linear in the required weights $J^{\mathrm{FP}}_{\mathbf{x},\mathbf{y}}$ so that we will have to use an approximative method to find a solution. Moreover, there may be not only one but a whole bunch of solutions – each corresponding to a fixed point. In our numerical simulations we have found that for an intracortical orientation map to emerge the parameter ξ^{e} in Eq. (2.1) must be chosen sufficiently large. Otherwise, the developing pattern of intracortical connectivity is approximately homogeneous and isotropic; cf. Sect. 2.5. This indicates that within a certain parameter regime the learning dynamics converge towards a homogeneous and isotropic fixed point which becomes unstable outside this regime, thereby giving rise to the emergence of an orientation map. In the present analysis we will therefore concentrate on fixed points $J^{\mathrm{FP}}_{\mathbf{x},\mathbf{y}}$ that correspond to homogeneous intracortical connectivity patterns.

Since in such a homogeneous connectivity pattern the weights $J^{\mathrm{FP}}_{\mathbf{x},\mathbf{y}}$ are independent of the postsynaptic cell's position $\mathbf{x}$, we can write $J^{\mathrm{FP}}_{\mathbf{x},\mathbf{y}} = J^{\mathrm{FP}}_{\mathbf{y}}$ so that the fixed-point equation becomes

$$
\begin{aligned}
0 \;&=\; B_{\mathbf{y}} \Big[c_0 \tau_* c_1 J^{\mathrm{FP}}_{\mathbf{y}} + c_0 \tau_\varepsilon \tau_* c_1 J^{\mathrm{FP}}_{\mathbf{y}} \sum_{\mathbf{y}'} c_1 J^{\mathrm{FP}}_{\mathbf{y}'} \Big] \\
&+\; A_{\mathbf{y}} \Big[\left(c_0^2 \tau_W + \sigma^{\mathrm{e}} c_0 + \zeta^{\mathrm{e}} \right) + c_0 \tau_\varepsilon \left(2 c_0 \tau_W + \sigma^{\mathrm{e}} \right) \sum_{\mathbf{y}'} c_1 J^{\mathrm{FP}}_{\mathbf{y}'} \\
&+\; c_0 \tau_\varepsilon^2 \left(3 c_0 \tau_W + \sigma^{\mathrm{e}} \right) \Big(\sum_{\mathbf{y}'} c_1 J^{\mathrm{FP}}_{\mathbf{y}'} \Big)^2 \\
&+\; c_0 \tau_*^2 \sum_{\mathbf{y}'} c_1^2 J^{\mathrm{FP}}_{\mathbf{y}'} J^{\mathrm{FP}}_{\mathbf{y}-\mathbf{y}'} + c_0 \frac{\tau_\varepsilon}{2} \tau_* \sum_{\mathbf{y}'} c_1^2 J^{\mathrm{FP}}_{\mathbf{y}'} J^{\mathrm{FP}}_{\mathbf{y}'-\mathbf{y}} \Big] - \vartheta^{\mathrm{e}} J^{\mathrm{FP}}_{\mathbf{y}} \; .
\end{aligned}
\tag{4.46}
$$

Analogously, Eq. (4.45) reduces to

$$
\lambda L_{\mathbf{x},\mathbf{y}} \;=\; B_{\mathbf{y}} \Big[c_0 c_1 \tau_* L_{\mathbf{x},\mathbf{y}} + c_0 c_1^2 \tau_\varepsilon \tau_* \sum_{\mathbf{y}'} \left(J^{\mathrm{FP}}_{\mathbf{y}} L_{\mathbf{x}-\mathbf{y},\mathbf{y}'} + J^{\mathrm{FP}}_{\mathbf{y}'} L_{\mathbf{x},\mathbf{y}} \right) \Big]
$$

$$
\begin{aligned}
&+ A_{\mathbf{y}}\Big[\left(c_0^2 c_1 \tau_\varepsilon \tau_W + \sigma^{\mathrm{e}} c_0 c_1 \tau_\varepsilon\right) \sum_{\mathbf{y}'} L_{\mathbf{x},\mathbf{y}'} + c_0^2 c_1 \tau_\varepsilon \tau_W \sum_{\mathbf{y}'} L_{\mathbf{x}-\mathbf{y},\mathbf{y}'} \\
&+ \left(c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W + \sigma^{\mathrm{e}} c_0 c_1^2 \tau_\varepsilon^2\right) \sum_{\mathbf{y}',\mathbf{y}''} \left(J^{\mathrm{FP}}_{\mathbf{y}'} L_{\mathbf{x}-\mathbf{y}',\mathbf{y}''} + J^{\mathrm{FP}}_{\mathbf{y}''} L_{\mathbf{x},\mathbf{y}'}\right) \\
&+ c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{\mathbf{y}',\mathbf{y}''} \left(J^{\mathrm{FP}}_{\mathbf{y}'} L_{\mathbf{x}-\mathbf{y}-\mathbf{y}',\mathbf{y}''} + J^{\mathrm{FP}}_{\mathbf{y}''} L_{\mathbf{x}-\mathbf{y},\mathbf{y}'}\right) \\
&+ c_0^2 c_1^2 \tau_\varepsilon^2 \tau_W \sum_{\mathbf{y}',\mathbf{y}''} \left(J^{\mathrm{FP}}_{\mathbf{y}'} L_{\mathbf{x}-\mathbf{y},\mathbf{y}''} + J^{\mathrm{FP}}_{\mathbf{y}''} L_{\mathbf{x},\mathbf{y}'}\right) \\
&+ c_0 c_1^2 \tau_*^2 \sum_{\mathbf{y}'} \left(J^{\mathrm{FP}}_{\mathbf{y}'} L_{\mathbf{x}-\mathbf{y}',\mathbf{y}-\mathbf{y}'} + J^{\mathrm{FP}}_{\mathbf{y}-\mathbf{y}'} L_{\mathbf{x},\mathbf{y}'}\right) \\
&+ c_0 c_1^2 \frac{\tau_\varepsilon}{2} \tau_* \sum_{\mathbf{y}'} \left(J^{\mathrm{FP}}_{\mathbf{y}'} L_{\mathbf{x}-\mathbf{y},\mathbf{y}'-\mathbf{y}} + J^{\mathrm{FP}}_{\mathbf{y}'-\mathbf{y}} L_{\mathbf{x},\mathbf{y}'}\right)\Big] - \vartheta^{\mathrm{e}} L_{\mathbf{x},\mathbf{y}} \; .
\end{aligned}
$$

In this equation we perform a Fourier transformation of $L_{\mathbf{x},\mathbf{y}}$ in $\mathbf{x}$ and denote

$$
\begin{aligned}
\widetilde{L}_{\mathbf{k},\mathbf{y}} &:= \sum_{\mathbf{x}} L_{\mathbf{x},\mathbf{y}} \exp\left(-i\mathbf{k}\mathbf{x}\right) \; , \\
\widetilde{J}^{\mathrm{FP}}_{\mathbf{k}} &:= \sum_{\mathbf{y}} J^{\mathrm{FP}}_{\mathbf{y}} \exp\left(-i\mathbf{k}\mathbf{y}\right) \; ,
\end{aligned}
$$

for $\mathbf{k} \in (-\pi, \pi]^2$. Thus, we arrive at the eigenvalue problem

$$
\begin{aligned}
&\frac{1}{c_1}\left(\lambda + \vartheta^{\mathrm{e}}\right) \widetilde{L}_{\mathbf{k},\mathbf{y}} = \\
&= c_0 \left(\tau_* B_{\mathbf{y}} + \tau_\varepsilon \tau_* c_1 \widetilde{J}^{\mathrm{FP}}_0 B_{\mathbf{y}}\right) \widetilde{L}_{\mathbf{k},\mathbf{y}} + c_0 \tau_\varepsilon \tau_* c_1 J^{\mathrm{FP}}_{\mathbf{y}} B_{\mathbf{y}} \mathrm{e}^{-i\mathbf{k}\mathbf{y}} \sum_{\mathbf{y}'} \widetilde{L}_{\mathbf{k},\mathbf{y}'} \\
&+ c_0 A_{\mathbf{y}} \sum_{\mathbf{y}'} \Bigg[\left(c_0 \tau_\varepsilon \tau_W + 2 c_0 \tau_\varepsilon^2 \tau_W c_1 \widetilde{J}^{\mathrm{FP}}_0\right)\left(1 + \mathrm{e}^{-i\mathbf{k}\mathbf{y}}\right) + \sigma^{\mathrm{e}} \tau_\varepsilon^2 c_1 \widetilde{J}^{\mathrm{FP}}_0 \\
&+ \sigma^{\mathrm{e}} \tau_\varepsilon + c_0 \tau_\varepsilon^2 \tau_W c_1 \widetilde{J}^{\mathrm{FP}}_{\mathbf{k}} \left(1 + \mathrm{e}^{-i\mathbf{k}\mathbf{y}}\right) + \sigma^{\mathrm{e}} \tau_\varepsilon^2 c_1 \widetilde{J}^{\mathrm{FP}}_{\mathbf{k}} + \frac{\tau_\varepsilon}{2} \tau_* c_1 J^{\mathrm{FP}}_{\mathbf{y}'-\mathbf{y}} \\
&+ \tau_*^2 c_1 J^{\mathrm{FP}}_{\mathbf{y}-\mathbf{y}'} \left(1 + \mathrm{e}^{-i\mathbf{k}(\mathbf{y}-\mathbf{y}')}\right) + \frac{\tau_\varepsilon}{2} \tau_* \mathrm{e}^{-i\mathbf{k}\mathbf{y}} c_1 J^{\mathrm{FP}}_{\mathbf{y}'+\mathbf{y}}\Bigg] \widetilde{L}_{\mathbf{k},\mathbf{y}'} \; ,
\end{aligned}
\tag{4.47}
$$

which we are now going to analyze numerically.

4.4.4 Numerical evaluation

Since we do not have analytic solutions of the fixed-point equation (4.46) and the eigenvalue problem (4.47), we will try a numerical approach. To this end, we must fix the numerical values of all the relevant network parameters. The specific values that we will use in the subsequent investigations are given in Table 4.2; they are slightly different for the different arborization models.

The time constants of the postsynaptic potential and the learning window, τ_ε and τ_W, are the same as in the numerical simulations presented in Chapter 2;

arbor	flat	sparse	dense
τ_ε	6 ms	6 ms	6 ms
τ_W	10 ms	10 ms	10 ms
ζ^{e}	$1\ \mathrm{s}^{-1}$	$0.8\ \mathrm{s}^{-1}$	$1\ \mathrm{s}^{-1}$
σ^{e}	-0.6	-0.5	-0.6
c_0	$1\ \mathrm{s}^{-1}$	$1\ \mathrm{s}^{-1}$	$1\ \mathrm{s}^{-1}$
$\vartheta^{\mathrm{e}}/c_1$	$1.6 \cdot 10^{-4}$	$1.39 \cdot 10^{-4}$	0

Table 4.2: For a numerical approach to the fixed-point equation (4.46) and the eigenvalue problem (4.47) the values of all the relevant parameters must be fixed. This table lists the specific values that will be used subsequently.

cf. Sect.2.8. Similarly, the parameters $\zeta^{\mathrm{e}} \equiv \xi^{\mathrm{e}}/\Delta t$ and σ^{e} have been chosen in approximate agreement with the respective values used in these simulations. As in the dynamics of the feedforward connectivity it turns out that σ^{e} must be sufficiently small so as to stabilize the total synaptic input converging onto one cortical cell. A value of 1 Hz seems to be a plausible choice for the spontaneous firing rate c_0 of our model neurons in view of the typical spontaneous firing rate of cortical pyramidal cells being a few spikes per second (Buser and Imbert 1992). The quantity $\vartheta^{\mathrm{e}}/c_1$ has been chosen in such a way as to obtain the required distribution of eigenvalues from Eq. (4.47), as will be explained below. In addition, we define the arbor functions $A_{\mathbf{y}}$ and $B_{\mathbf{y}}$ as

$$B_{\mathbf{y}} := 0.025\ , \quad A_{\mathbf{y}} := \begin{cases} 0.025 & \text{if } |\mathbf{y}| < 7.5, \\ 0 & \text{otherwise} \end{cases}$$

for the flat,

$$B_{\mathbf{y}} := 0.025\ , \quad A_{\mathbf{y}} := \begin{cases} 0.2 \cdot 0.025\ \exp\left[-|\mathbf{y}|^2/(2 \cdot 4.5^2)\right] & \text{if } |\mathbf{y}| < 7.5, \\ 0 \quad \text{otherwise} \end{cases}$$

for the sparse, and

$$B_{\mathbf{y}} \equiv A_{\mathbf{y}} := \begin{cases} 0.025 \exp\left[-|\mathbf{y}|^2/(2 \cdot 4.5^2)\right] & \text{if } |\mathbf{y}| < 7.5, \\ 0 \quad \text{otherwise} \end{cases}$$

for the dense arborization model.

Let us now turn to the problem of deriving the fixed-point connectivity patterns $J^{\mathrm{FP}}_{\mathbf{y}}$ that arise from Eq. (4.46). First, we recall that our present calculations are all based on the learning equation (4.37), which, in turn, is based on the approximations (4.32) and (4.33). These approximations have been obtained from second-order von-Neumann series expansions of the inverse operators appearing on the right hand sides of Eqs. (3.46) and (3.47). For (4.32) to yield a useful estimate of the mean firing rate we must therefore require

$$\tau_\varepsilon c_1 \sum_{\mathbf{y}} J^{\mathrm{FP}}_{\mathbf{y}} < 1\ ; \tag{4.48}$$

otherwise the involved von-Neumann series does not converge. Note, however, that (4.48) is also a physical limit of the reduced network model because

the network's mean firing rate would diverge if the total lateral input converging onto each neuron reached the above limit, i.e., the network would enter a self-exciting state. This is a direct consequence of Eq. (3.46).

Second, we recognize that for arbitrary $\mathbf{y}$ we have

$$J_{\mathbf{y}}^{\mathrm{FP}} \ll \sum_{\mathbf{y}'} J_{\mathbf{y}'}^{\mathrm{FP}}$$

if the total synaptic weight $\sum_{\mathbf{y}'} J_{\mathbf{y}'}^{\mathrm{FP}}$ is distributed evenly over a large number of synapses. Thus, we may write

$$\tau_\varepsilon c_1 J_{\mathbf{y}}^{\mathrm{FP}} \ll \tau_\varepsilon \sum_{\mathbf{y}'} c_1 J_{\mathbf{y}'}^{\mathrm{FP}} < 1$$

and hence obtain

$$\tau_\varepsilon^2 \sum_{\mathbf{y}'} c_1^2 J_{\mathbf{y}'}^{\mathrm{FP}} J_{\mathbf{y}-\mathbf{y}'}^{\mathrm{FP}} \ll \tau_\varepsilon^2 \Big(\sum_{\mathbf{y}'} c_1 J_{\mathbf{y}'}^{\mathrm{FP}} \Big)^2 < \tau_\varepsilon \sum_{\mathbf{y}'} c_1 J_{\mathbf{y}'}^{\mathrm{FP}} \; .$$

Because of the above relations the fixed-point equation (4.46) can be simplified by neglecting the two rightmost sums over $\mathbf{y}'$ so that one is left with

$$\begin{aligned} 0 \;\approx\; & B_{\mathbf{y}} \Big[c_0 \tau_* c_1 J_{\mathbf{y}}^{\mathrm{FP}} + c_0 \tau_\varepsilon \tau_* c_1 J_{\mathbf{y}}^{\mathrm{FP}} \sum_{\mathbf{y}'} c_1 J_{\mathbf{y}'}^{\mathrm{FP}} \Big] \\ + \; & A_{\mathbf{y}} \Big[\big(c_0^2 \tau_W + \sigma^{\mathrm{e}} c_0 + \zeta^{\mathrm{e}} \big) + c_0 \tau_\varepsilon \left(2 c_0 \tau_W + \sigma^{\mathrm{e}} \right) \sum_{\mathbf{y}'} c_1 J_{\mathbf{y}'}^{\mathrm{FP}} \\ + \; & c_0 \tau_\varepsilon^2 \left(3 c_0 \tau_W + \sigma^{\mathrm{e}} \right) \Big(\sum_{\mathbf{y}'} c_1 J_{\mathbf{y}'}^{\mathrm{FP}} \Big)^2 \Big] - \vartheta^{\mathrm{e}} J_{\mathbf{y}}^{\mathrm{FP}} \; . \end{aligned} \tag{4.49}$$

Concerning the flat and the sparse arborization model, this can be turned into a quadratic equation for the total weight $\sum_{\mathbf{y}'} c_1 J_{\mathbf{y}'}^{\mathrm{FP}}$ by summing over $\mathbf{y}$ on both sides. The same approach is possible in the dense arborization model, if $\vartheta^{\mathrm{e}} = 0$ as in the parameter set given in Table 4.2. Without discussing the details, we simply state that in the parameter regime we are investigating the resulting equation has one positive and one negative root. Since $J_{\mathbf{y}}^{\mathrm{FP}}$ denotes the efficacies of *excitatory* synapses the subsequent stability analysis will be restricted to that fixed point which corresponds to the positive solution. Reinserting the obtained value of $\sum_{\mathbf{y}'} c_1 J_{\mathbf{y}'}^{\mathrm{FP}}$ into Eq. (4.49) and solving for $J_{\mathbf{y}}^{\mathrm{FP}}$ finally yields the required approximation of the synaptic weights at the fixed point of the learning dynamics.

Figures 4.10 and 4.11 show the leading eigenvalues and eigenvectors of the eigenvalue problem (4.47) for the flat and the sparse arborization model with $J_{\mathbf{y}}^{\mathrm{FP}}$ derived in the way just explained. They are arranged the same as Figs. 4.2 through 4.5. In the upper part of either figure the four largest eigenvalues are plotted as functions of the wave vector $\mathbf{k}$. In the lower parts the four corresponding eigenvectors are displayed in a grey-level representation for three specific values of $\mathbf{k}$. As $J_{\mathbf{y}}^{\mathrm{FP}}$ is rotationally symmetric, the eigenvalue problem is also rotationally symmetric so that it is sufficient to consider only one direction of the wave vector, e.g., $\mathbf{k} = (k_1, 0)$.

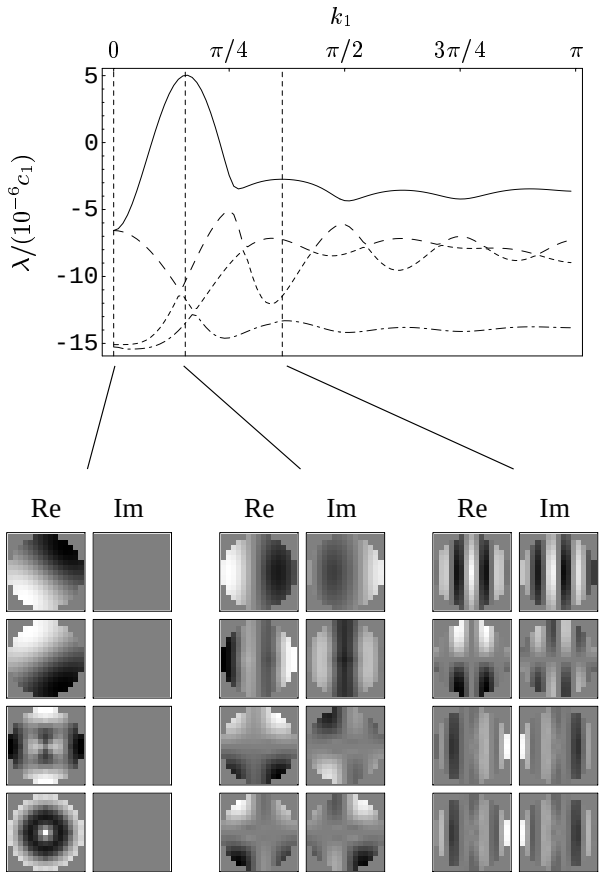


Figure 4.10: The leading eigenvalues and eigenmodes obtained from Eq. (4.47) are presented in dependence on the wave vector $\mathbf{k}$ for the flat arborization model and the parameter values given in Table 4.2. The corresponding fixed point of the intracortical learning dynamics, $J_{\mathbf{y}}^{\mathrm{FP}}$ has been approximated using Eq. (4.49) and is rotationally symmetric. As a consequence the underlying eigenvalue equation (4.47) is rotationally symmetric as well so that it is sufficient to consider only one direction of the wave vector as, e.g., $\mathbf{k} = (k_1, 0)$. The arrangement of the figure is analogous to that of Figs. 4.2 through 4.5. In the upper part the four largest eigenvalues are plotted as functions of k_1. In the lower part the four corresponding eigenvectors are displayed in a grey-level representation for three specific values of k_1. Each grey-level plot consists of two arrays of 15×15 pixels that visualize separately the real part and the imaginary part of the respective eigenvector. The plots have been arranged in accordance with the sequence of the eigenvalues so that the leading eigenvector appears in the uppermost row while the fourth eigenvector is shown at the bottom.

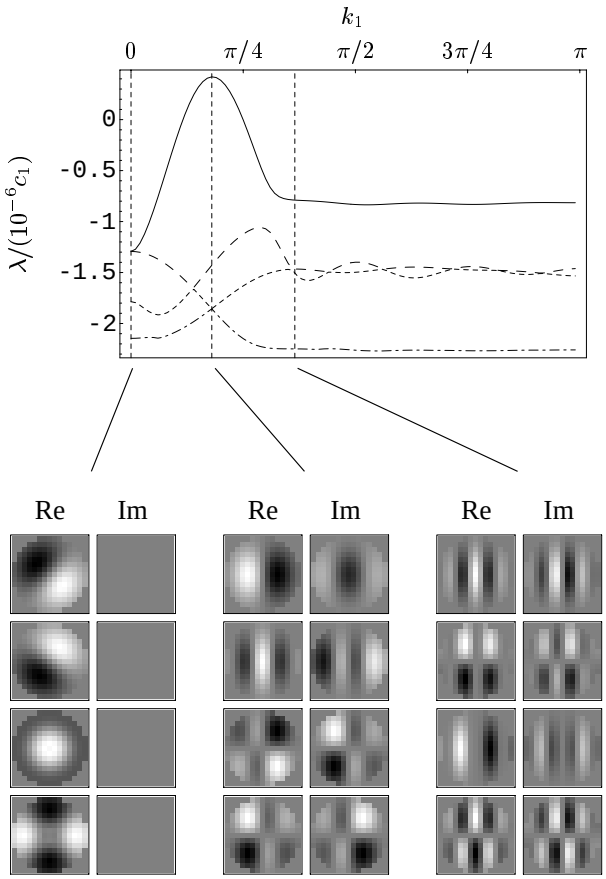


Figure 4.11: As Fig. 4.10 but for the sparse arborization model.

Obviously, the functional dependence of the principal eigenvector and its eigenvalue upon the wave vector $\mathbf{k}$ is qualitatively similar to what we have found in the Sect. 4.3.4 with regard to the geniculocortical development; cf. Figs. 4.3 and 4.5. Due to our particular choices of ϑ^{e} (see Table 4.2), the leading eigenvalue in both the flat and the sparse arborization model is negative except within a relatively narrow range of wavelengths. Under these conditions we may expect that the intracortical learning dynamics lead to the emergence of anisotropic connectivity patterns in the same way as an orientation map can develop from the plasticity of geniculocortical synapses. However, in the course of our investigations we have never found such an intracortical orientation map to develop in any numerical simulation that was based on the reduced network model.

This may be attributable to a fundamental difference between the dynamics of geniculocortical and intracortical plasticity as will be explained in the following. The development of feedforward synapses is mainly driven by correlations that are inherent to the activity of the geniculate relay cells and that are therefore prescribed externally. In contrast, the development of intracortical connections is driven by correlations in the activity of cortical cells and these correlations are induced by the intracortical connections themselves.

As a consequence, there is an important difference between the corresponding learning equations (4.23) and (4.42) and between the eigenvalue problems (4.28) and (4.47).

Obviously, in both eigenvalue equations the value of σ^{e} only affects eigenmodes $\widetilde{L}_{\mathbf{k},\mathbf{y}}$ with a non-vanishing sum $\sum_{\mathbf{y}} \widetilde{L}_{\mathbf{k},\mathbf{y}} \neq 0$. The real parts of their eigenvalues will normally decrease with decreasing σ^{e} as in the case of the eigenvectors shown in Fig. 4.2. Eigenvectors whose sum over $\mathbf{y}$ is zero, on the other hand, remain unaffected by changes of σ^{e}. It is thus plausible to assume that for a sufficiently negative value of σ^{e} the principal eigenvector will have a vanishing sum $\sum_{\mathbf{y}} \widetilde{L}_{\mathbf{k},\mathbf{y}} = 0$. This hypothesis is supported by analytical investigations of similar eigenvalue problems (MacKay and Miller 1990, Wimbauer 1996, Wimbauer et al. 1998) as well as by a numerical investigation of the leading eigenvectors presented in Figs. 4.3 through 4.5 and in Figs. 4.10 and 4.11. Within the numerical accuracy their sum is vanishing and their eigenvalues are independent of σ^{e}.

For these zero-sum eigenmodes the eigenvalue problems (4.28) and (4.47) reduce to

$$\begin{aligned}\frac{1}{c_1}\left(\lambda + \vartheta^{\mathrm{e}}\right)\widetilde{L}_{\mathbf{k},\mathbf{y}} = A_{\mathbf{y}} \sum_{\mathbf{y}'} \Big[& d_0 \tau_*^2 \mathrm{e}^{-i\mathbf{k}(\mathbf{y}-\mathbf{y}')} c_1 J^{\mathrm{ic}}_{\mathbf{y}-\mathbf{y}'} \\ + \quad & d_1^2 \tau_1^2 H_0 M_{\mathbf{y}-\mathbf{y}'} + d_1^2 \tau_2^3 H_0 \Upsilon_{\mathbf{k},\mathbf{y}-\mathbf{y}'} \Big] \widetilde{L}_{\mathbf{k},\mathbf{y}'} + d_0 \tau_* B_{\mathbf{y}} \widetilde{L}_{\mathbf{k},\mathbf{y}} \,,\end{aligned} \tag{4.50}$$

for the geniculocortical plasticity and

$$\begin{aligned}\frac{1}{c_1}\left(\lambda + \vartheta^{\mathrm{e}}\right)\widetilde{L}_{\mathbf{k},\mathbf{y}} = c_0 \tau_* A_{\mathbf{y}} \sum_{\mathbf{y}'} \Big[& \frac{\tau_\varepsilon}{2} c_1 J^{\mathrm{FP}}_{\mathbf{y}'-\mathbf{y}} + \frac{\tau_\varepsilon}{2} \mathrm{e}^{-i\mathbf{k}\mathbf{y}} c_1 J^{\mathrm{FP}}_{\mathbf{y}'+\mathbf{y}} \\ + \quad & \tau_* c_1 J^{\mathrm{FP}}_{\mathbf{y}-\mathbf{y}'} \left(1 + \mathrm{e}^{-i\mathbf{k}(\mathbf{y}-\mathbf{y}')}\right) \Big] \widetilde{L}_{\mathbf{k},\mathbf{y}'} + c_0 \tau_* B_{\mathbf{y}} \left(1 + \tau_\varepsilon c_1 \widetilde{J}_0^{\mathrm{FP}}\right) \widetilde{L}_{\mathbf{k},\mathbf{y}} \,,\end{aligned} \tag{4.51}$$

for the intracortical learning dynamics. A comparison of the above equations reveals that in (4.51) all the terms within the square brackets depend on the fixed point of the intracortical learning dynamics, while the corresponding terms in (4.50) do not depend on the fixed point of the geniculocortical plasticity. This is a result of the fact that the related terms in the learning equation (4.42) are quadratic in the dynamic variables, viz., the intracortical weights J_{ij}, whereas the respective terms in (4.23) are linear in the geniculocortical weights J^{e}_{ij}.

Thus, the process of pattern formation in the geniculocortical connectivity can be described, at least approximately, by a system of linear differential equations. The emerging structure is determined by the eigenmodes of a corresponding eigenvalue problem. In contrast, the dynamics of intracortical plasticity is described by a system of essentially non-linear differential equations – a linear approximation and its related eigenvalue problem can be valid only within a certain range in the vicinity of a fixed point. As an intracortical orientation map did not develop in our numerical simulations, this range seems to be too narrow in the reduced network model for the emerging intracortical connectivity to be determined by the eigenmodes of the linearized learning dynamics.

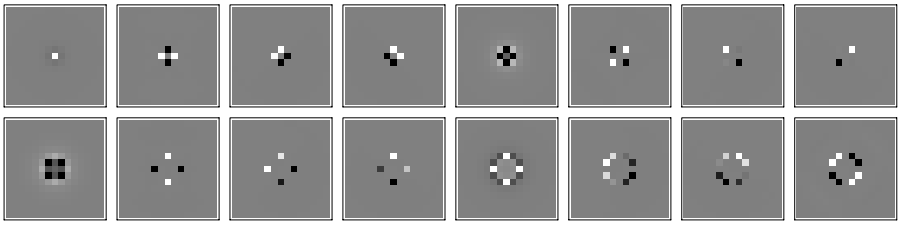


Figure 4.12: The first 16 eigenmodes of Eq. (4.47) are presented for the dense arborization model with $\mathbf{k} = 0$ and the parameter values given in Table 4.2 (sequence: top row from left to right, then bottom row from left to right). The corresponding fixed point of the intracortical learning dynamics, $J_{\mathbf{y}}^{\mathrm{FP}}$ has been approximated using Eq. (4.49).

There is also an important difference between Eqs. (4.50) and (4.51) concerning the size of the terms within the square brackets. The quantities $H_0\, M_{\mathbf{y}}$ and $H_0 \Upsilon_{\mathbf{k},\mathbf{y}}$ in (4.50) are determined by the geniculate activity correlations. The strength of these correlations and hence the size of $H_0\, M_{\mathbf{y}}$ and $H_0 \Upsilon_{\mathbf{k},\mathbf{y}}$ are prescribed by some external mechanisms and are not limited by the properties of our network model. In Eq. (4.51), on the other hand, all the terms within the square brackets are determined by the intracortical connectivity $J_{\mathbf{y}}^{\mathrm{FP}}$ and so their size is fundamentally limited by the upper bound (4.48).

As a matter of fact, the contribution of the respective terms to the eigenvalue problem (4.47) is so small that the eigenvectors resulting for the dense arborization model are almost identical to the eigenmodes that would arise if the contribution of these terms were exactly equal to zero. Figure 4.12 shows that these eigenmodes do not resemble the typical bilobed shape of orientation selective receptive fields. That means that even in the vicinity of the fixed point of the learning dynamics a growth of oriented intracortical connectivity patterns is prevented by the properties of the reduced network model.

Since the derivation of the eigenvalue problem (4.47) involved a second-order von-Neumann series expansion of the inverse operators in (3.46) and (3.47) the following approach may provide a possible way to circumvent this problem. Suppose the intracortical connectivity $J_{\mathbf{y}}^{\mathrm{FP}}$ at the fixed point is close to its limit given by (4.48). In that case a second-order approximation of the inverse operators will no longer be precise enough so that higher-order terms must be taken into account. If $\sum_{\mathbf{y}} J_{\mathbf{y}}^{\mathrm{FP}}$ is sufficiently close to its maximal value, then the number of required terms can become arbitrarily large. Since each of them will contribute to the square brackets in Eq. (4.51), their total contribution may also become arbitrarily large.

However, this approach raises new difficulties. First, the model parameters must be tuned very accurately so as to obtain a fixed point of the learning dynamics that is close to the limit (4.48). Second, it must be ensured that the network does not become self-exciting during the process of learning. With the network being already very close to a self-exciting state at the fixed point of the learning dynamics, this will also require an extremely exact adjustment of parameters.

We may therefore conclude that the reduced network model introduced in Sect. 4.1 is too limited as to account for the mechanisms which can, in a more elaborate model, robustly lead to the development of intracortical orientation maps as seen in Chapter 2.

4.5 Summary

In this chapter we have presented an analytical investigation of spike-based Hebbian plasticity in a simple model of the primary visual cortex. The purpose of this investigation was to understand the processes of pattern formation which we have found in our numerical simulations in Chapter 2. Because of the high complexity of the neuronal network used in these simulations we have introduced a simplified version in Sect. 4.1 so as to obtain an analytically tractable model. This reduced network comprises one geniculate and one cortical layer consisting of linear stochastic spike response-neurons that are intended to model geniculate relay cells and excitatory cortical cells, respectively.

According to the learning rules set forth in Sect. 2.3 each synaptic weight is modified in dependence on the pre- and postsynaptic spike trains, which are random processes. Starting from these learning rules we have derived a differential equation describing the temporal evolution of a synapse's efficacy as a function of the two-spike cross-correlations and the mean firing rates of the involved spike trains. Applying the methods developed in Chapter 3 the required spike statistics can be expressed in terms of the network's connectivity, i.e., in terms of the synaptic weights. In this way, it is possible to arrive at a closed system of differential equations for the temporal development of all synaptic weights.

Using this description of the learning dynamics we have investigated the plasticity of geniculocortical synapses in the presence of a fixed intracortical connectivity. First, the intracortical connections were assumed to be homogeneous and rotationally symmetric. A numerical evaluation of the resulting learning equation revealed that within a certain regime of network parameters the translational and rotational symmetries can be broken and an orientation selective connectivity pattern can emerge. As a prerequisite, however, the total synaptic input converging onto each neuron must be stabilized so that individual synapses have to compete for their efficacies. In agreement with earlier investigations (Kempter et al. 1999, Kempter et al. 2000) we have found that this can be achieved by inducing an appropriate reduction of the synaptic weights in response to every postsynaptic spike. To this end, the value of the parameter σ^{e} in Eq. (2.1) must be chosen sufficiently small (negative).

In order to mimick the presence of a previously established intracortical orientation map during the period of geniculocortical development, we then introduced a non-isotropic pattern of intracortical connectivity. The layout of the corresponding intracortical "orientation map" was very simple with all neurons' lateral receptive fields being identical and oriented vertically; its colored representation using the color code shown in Figs. 2.6 and 2.7 would

appear in a uniform light blue. A numerical investigation of the learning equation shows that the geniculocortical connectivity patterns emerging under these conditions can still form an orientation map. The resulting map, however, will be biased towards vertical orientations and thus tends to be in accord with the intracortical map. This finding is in agreement with the outcome of the numerical simulations presented in Sects. 2.6 and 2.7 where we have seen that a pre-existing intracortical orientation map can guide the development of the geniculocortical connectivity.

The above results are in accordance and extend the findings of earlier analytical investigations of cortical pattern formation (Kammen and Yuille 1988, Yuille et al. 1989, MacKay and Miller 1990, Stetter et al. 1993, Wimbauer 1996, Wimbauer et al. 1998). These investigations, based upon networks consisting of graded-response neurons, have shown how a Hebbian plasticity of geniculocortical synapses driven by correlated activity can induce the development of orientation selective receptive fields. In accordance with experimental data (Rauschecker and Singer 1979, Wiesel 1982) a competition of different synapses for synaptic weight turned out to be a prerequisite for pattern formation. Here we have demonstrated that in a biologically more realistic network of spiking neurons a similar process can result in an emergence of orientation selectivity and we have seen that synaptic competition plays a crucial role as well. As in the previously analyzed models the presence of an isotropic pattern of intracortical connectivity is sufficient to obtain nicely ordered orientation maps. However, we have also shown how the geniculocortical development can be guided by a non-isotropic intracortical connectivity.

Our numerical simulations presented in Sect. 2.5 have revealed that such a non-isotropic intracortical connectivity pattern, viz., an intracortical orientation map can develop from a Hebbian self-organization of lateral connections in the absence of feedforward input. Therefore, we tried to obtain some analytical insight into this process by investigating the properties of intracortical plasticity in the reduced model. We derived the corresponding learning equation and calculated the rotationally symmetric fixed point. Next, we linearized the learning dynamics about the fixed point so as to determine the connectivity patterns that would start growing when the fixed point is destabilized. Within an appropriate regime of network parameters these patterns turned out to be similar to those obtained previously for the geniculocortical development.

Hence, one may expect that the intracortical learning dynamics lead to the emergence of a map-like connectivity pattern in the same way as an orientation map can develop from the plasticity of geniculocortical synapses. However, we have never found an intracortical orientation map to emerge in a numerical simulation based on the reduced network model. We have discussed two possible sources of this discrepancy.

First, the differential equation describing the learning dynamics is essentially non-linear in the intracortical synaptic efficacies. Therefore, the results obtained from the linearized version of the learning equation are valid only within a certain range in the vicinity of the fixed point. This range may be too narrow for the linear solutions to determine the final state of the synaptic dynamics. Second, with the reduced network consisting merely of linear

excitatory neurons, the total synaptic input converging onto each cell at the fixed point must be limited by Eq. (4.48) so as to prevent the network from entering a self-exciting state (see also Douglas et al. 1995). Because of this limit the relevant terms in the learning equation are very small. We have thus concluded that the reduced network model introduced in Sect. 4.1 is too simple as to account for the mechanisms which can, in a more elaborate model with inhibitory neurons, robustly lead to the development of intracortical orientation maps, as we have seen in Chapter 2.

Summary

In the eye, an inverted image of the outside world is focused onto the retina, where it is converted into sequences of electrical spikes emitted by the retinal ganglion cells. Via the lateral geniculate nucleus (LGN), these electrical signals are relayed to layer 4 of the primary visual cortex. Detailed investigations have revealed how nerve cells in the primary visual cortex respond to visual stimulation (see Chapter 1). Normally, a given neuron can only be driven by stimulating a certain area on the retina, which is called its receptive field. The majority of cortical neurons respond to stimulation of either eye but are dominated by one eye. The best response is usually evoked by a specifically oriented linear edge of light projected onto the receptive field. As it turns out, there is a continuous relation between a cell's cortical location and each of its various response properties as, e.g., ocular dominance or preferred orientation (Hubel and Wiesel 1962). These relations are referred to as cortical maps; prominent examples are the ocular dominance map and the orientation map. Current experimental data suggest that the underlying neuronal circuitry is partially shaped by some kind of activity-driven learning mechanism during the so-called critical period early in an animal's life (Wiesel and Hubel 1963b, Blakemore and van Sluyters 1974, Stryker and Harris 1986, Crair et al. 1998, Issa et al. 1999).

In this work, we have investigated the development of orientation maps in a network model of layer 4 in the primary visual cortex. The network consists of stochastically spiking cells to mimick geniculate neurons as well as excitatory and inhibitory cortical neurons. The geniculate neurons provide feedforward input to the cortex. They are assumed to produce correlated spike activity of given statistics. The cortical cells are recurrently interconnected by lateral synapses. Both the geniculocortical feedforward connections and the recurrent intracortical connections are subject to activity-driven Hebbian learning (Chapter 2).

In a first set of numerical simulations we have considered a situation in which cortical cells produce random spontaneous activity but do not yet receive feedforward input from the LGN. This is an interesting scenario because anatomical investigations in the cat's developing brain have in fact demonstrated that geniculate afferents do not innervate layer 4 of the primary visual cortex until about one week after the target cells have finished their migration. The simulations in our model have revealed that spontaneous activity of cortical cells can drive the plasticity of lateral synapses to form an inhomogeneous and anisotropic connectivity pattern, even in the absence of geniculocortical

input. Since the overall organization of the emerging pattern typically resembles the layout of orientation maps recorded from real visual cortex (Bonhoeffer and Grinvald 1991, Blasdel 1992b), we have called it an *intracortical* orientation map.

Next, we have demonstrated that a Hebbian development of the geniculocortical afferents driven by correlated activity in the LGN leads to the formation of orientation selective receptive fields of cortical cells. As in previous models of correlation-based learning (von der Malsburg 1973, Linsker 1986b, Linsker 1986a, Miller 1994, Wimbauer et al. 1997b, Wimbauer et al. 1997a, Erwin and Miller 1998, Choe and Miikkulainen 1998), a uniform and circularly symmetric pattern of intracortical connectivity is sufficient for these receptive fields to arrange into a nicely ordered orientation map, which we refer to as a *feedforward* orientation map.

In view of these results, it was interesting to investigate how an intracortical orientation map would influence the development of the feedforward map. In a third series of numerical simulations we have therefore analyzed the learning dynamics of the feedforward connectivity in the presence of a previously formed intracortical map. Strikingly, the plasticity of the geniculocortical afferents can be guided by the lateral connections so that the emerging feedforward map is in accord with the intracortical map.

Since the same intracortical map guides the development of afferents from both eyes, the final orientation maps will be matched in the two eyes automatically. We have demonstrated that this mechanism provides a new explanation of the outcome of so-called reverse lid-suturing experiments (Gödecke and Bonhoeffer 1996, Sengpiel et al. 1998), by which previous theoretical approaches had been challenged. In these experiments, cats were raised so that both eyes never received common visual input. Nevertheless, the two eyes' orientation maps were found to be nearly identical. Our model is further supported by experiments in which kittens were raised in environments containing contours of only one orientation (Sengpiel et al. 1999). The orientation maps recorded from these animals exhibited only moderate shifts towards this orientation. This result is in full agreement with the assumption that the geniculocortical development is guided intracortically so that visual experience has only a minor influence regarding the distribution of orientation preferences.

In order to obtain a deeper understanding of the developmental processes observed in our network, we then turned to a mathematical analysis of the underlying learning dynamics. The neuronal contacts, i.e., the synapses in the network are modified according to a Hebbian learning mechanism. This means that a synapse is strengthened or weakened in dependence on the relative timing between the spikes emitted by the two cells it is connecting. Hence, the synaptic dynamics are primarily governed by the spike-spike correlations of the network's activity.

Chapter 3 was therefore devoted to an investigation of the spike statistics in recurrently interconnected networks of spiking neurons. We have introduced the concept of a network ensemble as an entirety of infinitely many realizations or independent 'runs' of the same stochastically spiking neural network.

Within this framework we have first derived expressions for the input-output correlations of a single neuron without feedback. Then we have generalized our approach so as to obtain analytical approximations for the mean activity and the two-spike correlation functions in arbitrarily interconnected networks of linear or non-linear neurons. A comparison with the results obtained from a series of numerical simulations has shown that for a suitable choice of network parameters our analytical expressions yield good approximations of the true spike statistics.

At that point we were prepared for a mathematical investigation of pattern formation in a recurrent network of spiking neurons (Chapter 4). The dynamics of the full model studied in Chapter 2, however, turn out to be extremely involved so that we have analyzed a simplified version. This reduced network comprises two layers of stochastically spiking linear neurons that are intended to model geniculate and cortical cells. Starting from the Hebbian learning rules used in the numerical simulations we have derived a differential equation describing the temporal evolution of a synapse's efficacy in dependence on the statistics of the pre- and postsynaptic spike trains.

Combining this equation with the methods developed in Chapter 3 we have studied the plasticity of geniculocortical synapses. First, the lateral connections were assumed to be circularly symmetric. A linear stability analysis of the learning dynamics has revealed that in an appropriate parameter regime a symmetry breaking process can lead to the formation of a feedforward orientation map. The underlying processes involve a competition of individual synapses and are quite similar to the mechanisms found previously (Kammen and Yuille 1988, MacKay and Miller 1990, Wimbauer et al. 1998) in models consisting of non-spiking neurons.

Then we have introduced a non-isotropic lateral connectivity so as to investigate its influence onto the developing feedforward map. The lateral projections established a very simple intracortical "orientation map" with all neurons' connectivity patterns being identical. In our investigation of the resulting learning equation we have seen that under these conditions orientation selective receptive fields can develop basically in the same way as with a circularly symmetric intracortical connectivity. The emerging feedforward orientation map, however, is biased towards the intracortical orientation because synaptic patterns corresponding to this orientation grow fastest. Thus, the feedforward map tends to be in accord with the intracortical map.

How does an intracortical orientation map arise? In order to address this question we have analyzed the dynamics of intracortical plasticity in the reduced network. At first glance, the results suggest that symmetry breaking may occur much like in the development of the geniculocortical connectivity. On the other hand, we have never found an intracortical orientation map to emerge in a numerical simulation based on the reduced model. As we have discussed, this discrepancy is probably a consequence of a linearization in the mathematical analysis. We have concluded that the reduced model is too simple as to account for the formation of intracortical orientation maps. One of its basic limitations results from its lack of inhibitory neurons, which are present in the full model – and also in real cortex.

The model of activity-driven cortical development that we have proposed and investigated provides a simple and natural explanation for experimental observations that were difficult to understand in the framework of earlier theoretical approaches. The key feature is an interaction of the geniculocortical development with an intracortical map. We have demonstrated that the required pattern of lateral connectivity can arise from activity-driven learning of intracortical synapses. Since our network consists of spiking neurons and strictly distinguishes between excitatory and inhibitory cells, it is significantly closer to biology than many previous models of the primary visual cortex.

Further investigations may extend the present model in several ways. In particular, it may be worthwhile to concentrate on the following questions. How does the layout of the emerging intracortical map depend on the model parameters and the details of the network's architecture? How well does it fit with the layout of measured orientation maps? Why are inhibitory neurons relevant for the development of an intracortical map? What is the role of the long-range lateral projections that have been found in cortical layers 2/3 and 5 (Callaway and Katz 1990)? How does the intracortical map influence the development of an ocular dominance map (Miller et al. 1989, Erwin and Miller 1998) or the development of a direction selectivity map (Wimbauer et al. 1997a, Wimbauer et al. 1997b)? Do the relations between the various emerging maps fit with experimental data (cf. Sect. 1.4.3)?

Hopefully, such theoretical studies will help to achieve a deeper understanding of cortical information processing and to arrive, some day, at a biologically realistic model of the visual cortex. For the moment, however, we do not have more than a faint idea about its functional organization and the underlying circuitry.

Bibliography

Adorján P., Levitt J. B., Lund J. S., Obermayer K., 1999, A model for the intracortical origin of orientation preference and tuning in macaque striate cortex. *Vis. Neurosci.*, 16:303–318.

Bartfeld E., Grinvald A., 1992, Relationships between orientation preference pinwheels, cytochrome oxidase blobs, and ocular dominance columns in primate striate cortex. *Proc. Natl. Acad. Sci. USA*, 89:11905–11909.

Bartsch A. P., van Hemmen J. L., 1999, Development of orientation maps from Hebbian learning of intracortical connections. In Elsner N., Eysel U., editors, *Göttingen Neurobiology Report, Volume II*, page 502, Stuttgart. Georg Thieme Verlag.

Ben-Yishai R., Bar-Or R. L., Sompolinsky H., 1995, Theory of orientation tuning in visual cortex. *Proc. Natl. Acad. Sci.*, 92:3844–3848.

Ben-Yishai R., Hansel D., Sompolinsky H., 1997, Travelling waves and the processing of weakly tuned inputs in a cortical network module. *J. Comput. Neurosci.*, 4:57–77.

Blakemore C., Cooper G. F., 1970, Development of the brain depends on the visual environment. *Nature*, 228:477–478.

Blakemore C., van Sluyters R. C., 1974, Reversal of the physiological effects of monocular deprivation in kittens: Further evidence for a sensitive period. *J. Physiol.*, 237:195–216.

Blasdel G. G., 1992a, Differential imaging of ocular dominance and orientation selectivity in monkey striate cortex. *J. Neurosci.*, 12:3115–3138.

Blasdel G. G., 1992b, Orientation selectivity, preference, and continuity in monkey striate cortex. *J. Neurosci.*, 12:3139–3161.

Blasdel G. G., Fitzpatrick D., 1984, Physiological organization of layer 4 in macaque striate cortex. *J. Neurosci.*, 4:880–895.

Blasdel G. G., Mitcell D. E., Muir D. W., Pettigrew J. D., 1977, A physiological and behavioural study in cats of the effect of early visual experience with contours of a single orientation. *J. Physiol. (Lond.)*, 265:615–636.

Blasdel G. G., Salama G., 1986, Voltage-sensitive dyes reveals a modular organization in the monkey striate cortex. *Nature*, 321:579–585.

Bonhoeffer T., Gödecke I., 1996, Organization of the visual cortex – reply. *Nature*, 382:306–307.

Bonhoeffer T., Grinvald A., 1991, Iso-orientation domains in cat visual cortex are arranged in pinwheel-like patterns. *Nature*, 353:429–431.

Bonhoeffer T., Grinvald A., 1993, The layout of iso-orientation domains in area 18 of cat visual cortex: Optical imaging reveals a pinwheel-like organization. *J. Neurosci.*, 13:4157–4180.

Bosking W. H., Zhang Y., Schofield B., Fitzpatrick D., 1997, Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. *J. Neurosci.*, 17:2112–2127.

Braitenberg V., Schüz A., 1991, *Anatomy of the cortex.* Springer, Berlin.

Brown T. H., Chattarji S., 1994, Hebbian synaptic plasticity: Evolution of the contemporary concept. In Domany E., van Hemmen J. L., Schulten K., editors, *Models of Neural Networks II*, pages 287–314. Springer, New York.

Buser P., Imbert M., 1992, *Vision.* MIT Press, Cambridge, Massachusetts.

y Cajal S. R., 1911, *Histologie du Système Nerveux de l'homme et de vertébrés.* A. Maloine, Paris.

Callaway E. M., Katz L. C., 1990, Emergence and refinement of clustered horizontal connections in cat striate cortex. *J. Neurosci.*, 10:1134–1153.

Chapman B., Stryker M. P., 1993, Development of orientation selectivity in ferret visual cortex and effects of deprivation. *J. Neurosci.*, 13:5251–5262.

Chapman B., Stryker M. P., Bonhoeffer T., 1996, Development of orientation preference maps in ferret primary visual cortex. *J. Neurosci.*, 16:6443–6453.

Choe Y., Miikkulainen R., 1998, Self-organization and segmentation in a laterally connected orientation map of spiking neurons. *Neurocomputing*, 21:51–60.

Chung S., Ferster D., 1998, Strength and orientation tuning of the thalamic input to simple cells revealed by electrically evoked cortical suppression. *Neuron*, 20:1177–89.

Crair M. C., Gillespie D. C., Stryker M. P., 1998, The role of visual experience in the development of columns in cat visual cortex. *Science*, 279:566–570.

Crair M. C., Ruthazer E. S., Gillespie D. C., Stryker M. P., 1997, Ocular dominance peaks at pinwheel center singularities of the orientation map in cat visual cortex. *J. Neurophysiol.*, 77:3381–3385.

Crook J. M., Eysel U. T., 1992, GABA-induced inactivation of functionally characterized sites in cat visual cortex (area 18): effects on orientation tuning. *J. Neurosci.*, 12:1816–1825.

Douglas R. J., Koch C., Mahowald M., Martin K. A. C., Suarez H. H., 1995, Recurrent excitation in neocortical circuits. *Science*, 269:981–985.

Douglas R. J., Martin K. A. C., 1990, Neocortex. In Shepherd G. M., editor, *The synaptic organization of the brain*, chapter 12, pages 389–438. Oxford University Press, Oxford, 3rd edition.

Douglas R. J., Martin K. A. C., 1992, Exploring cortical microcircuits: A combined anatomical, physiological, and computational approach. In McKenna T., Davis J., Zornetzer S. F., editors, *Single Neuron Computation*,

chapter 14, pages 381–412. Academic Press, Inc.

Douglas R. J., Martin K. A. C., Whitteridge D., 1989, A canonical microcircuit for neocortex. *Neural Comp.*, 1:480–488.

Engert F., Bonhoeffer T., 1997, Synapse specificity of long-term potentiation breaks down at short distances. *Nature*, 388:279–284.

Engert F., Bonhoeffer T., 1999, Dendritic spine changes associated with hippocampal long-term synaptic plasticity. *Nature*, 399:66–70.

Erwin E., Miller K. D., 1998, Correlation-based development of ocularly-matched orientation and ocular dominance maps: Determination of required input activities. *J. Neurosci.*, 18:9870–9895.

Ferster D., 1987, Origin of orientation-selective EPSPs in simple cells of cat visual cortex. *J. Neurosci.*, 7:1780–1791.

Ferster D., 1988, Spatially opponent excitation and inhibition in simple cells of the cat visual cortex. *J. Neurosci.*, 8:1172–1180.

Ferster D., Chung S., Wheat H., 1996, Orientation selectivity of thalamic input to simple cells of cat visual cortex. *Nature*, 380:249–252.

Ferster D., Miller K. D., 2000, Neural mechanisms of orientation selectivity in the visual cortex. *Annu. Rev. Neurosci.*, 23:441–471.

Fregnac Y., Burke J. P., Smith D., Friedlander M. J., 1994, Temporal covariance of pre- and postsynaptic activity regulates functional connectivity in the visual cortex. *J. Neurophysiol.*, 71:1403–1421.

Gerstner W., 1995, Time structure of the activity in neural network models. *Phys. Rev. E*, 51:738–758.

Gerstner W., van Hemmen J. L., 1992, Associative memory in a network of 'spiking' neurons. *Network*, 3:139–164.

Gerstner W., van Hemmen J. L., 1994, Coding and information processing in neural networks. In Domany E., van Hemmen J. L., Schulten K., editors, *Models of neural networks II*, pages 1–93, New York. Springer-Verlag.

Gödecke I., Bonhoeffer T., 1996, Development of identical orientation maps for two eyes without common visual experience. *Nature*, 379:251–254.

Goodman C. S., Shatz C. J., 1993, Developmental mechanisms that generate precise patterns of neuronal connectivity. *Neuron*, 10:1–20.

Gosh A., Shatz C. J., 1992, Pathfinding and target selection by developing geniculocortical axons. *J. Neurosci.*, 12:39–55.

Gradshteyn I. S., Ryzhik I. M., 1994, *Table of Integrals, Series, and Products*. Academic Press, San Diego, 5^{th} edition.

Grinvald A., Anglister L., Freeman J. A., Hildesheim R., Manker A., 1984, Real-time optical imaging of naturally evoked electrical activity in intact frog brain. *Nature*, 308:848–850.

Grinvald A., Lieke E., Frostig R. D., Gilbert C. D., Wiesel T. N., 1986, Functional architecture of cortex revealed by optical imaging of intrinsic signals. *Nature*, 324:361–364.

Hansel D., Sompolinsky H., 1996, Chaos and synchrony in a model of a hypercolumn in visual cortex. *J. Comput. Neurosci.*, 3:7–34.

Hartline H. K., 1940, The receptive fields of optic nerve fibers. *Am. J. Physiol.*, 130:690–699.

Hawken M. J., Parker A. J., 1984, Contrast sensitivity and orientation selectivity in lamina IV of the striate cortex of old world monkeys. *Exp. Brain Res.*, 54:367–372.

Hawkes A. G., 1971a, Point spectra of some mutually exciting point processes. *J. R. Statist. Soc. Ser. B*, 33:438–443.

Hawkes A. G., 1971b, Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, 58(1):83–90.

Hebb D. O., 1949, *The organization of behavior*. Wiley, New York.

Hickey T. L., Hitchcock P. F., 1984, Genesis of neurons in the dorsal lateral geniculate nucleus of the cat. *J. Comp. Neurol.*, 228:186–199.

Hirsch H. V., Spinelli D. N., 1970, Visual experience modifies distribution of horizontally and vertically oriented receptive fields in cats. *Science*, 168:869–871.

Hubel D. H., Wiesel T. N., 1959, Receptive fields of single neurones in the cat's striate cortex. *J. Physiol.*, 148:574–591.

Hubel D. H., Wiesel T. N., 1961, Integrative action in the cat's lateral geniculate body. *J. Physiol.*, 155:385–398.

Hubel D. H., Wiesel T. N., 1962, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J. Physiol.*, 160:106–154.

Hubel D. H., Wiesel T. N., 1963, Receptive fields of cells in striate cortex of very young, visually inexperienced kittens. *J. Neurophysiol.*, 26:994–1002.

Hubel D. H., Wiesel T. N., 1968, Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.*, 195:215–243.

Hubel D. H., Wiesel T. N., Stryker M. P., 1977, Orientation columns in macaque monkey visual cortex demonstrated by the 2-deoxyglucose autoradiographic technique. *Nature*, 269:328–330.

Hübener M., Shoham D., Grinvald A., Bonhoeffer T., 1997, Spatial relationships among three columnar systems in cat area 17. *J. Neurosci.*, 17:9270–9284.

Humphrey A. L., Norton T. T., 1980, Topographic organization of the orientation column system in the striate cortex of the tree shrew (tupaia glis). I. Microelectrode recording. *J. Comp. Neurol.*, 192:531–547.

Humphrey A. L., Skeen L. C., Norton T. T., 1980, Topographic organization of the orientation column system in the striate cortex of the tree shrew (tupaia glis). II. Deoxyglucose mapping. *J. Comp. Neurol.*, 192:549–566.

Issa N. P., Trachtenberg J. T., Chapman B., Zahs K. R., Stryker M. P., 1999, The critical period for ocular dominance plasticity in the ferret's visual cortex. *J. Neurosci.*, 19:6965–6978.

Kammen D. M., Yuille A. L., 1988, Spontaneous symmetry-breaking energy functions and the emergence of orientation selective cortical cells. *Biol. Cybern.*, 59:23–31.

Kandel E. R., Schwartz J. H., Jessell T. M., 1997, *Principles of Neural Science*. Prentice-Hall International Inc., 3rd edition.

Katz L. C., Callaway E. M., 1992, Development of local circuits in mammalian visual cortex. *Annu. Rev. Neurosci.*, 15:31–56.

Kempter R., Gerstner W., van Hemmen J. L., 1999, Hebbian learning and spiking neurons. *Phys. Rev. E*, 59:4498–4514.

Kempter R., Gerstner W., van Hemmen J. L., 2000, Intrinsic rate normalization in spike-based and rate-based hebbian learning. *Preprint*.

Kempter R., Gerstner W., van Hemmen J. L., Wagner H., 1998, Extracting oscillations: Neuronal coincidence detection with noisy periodic spike input. *Neural Comput.*, 10:1987–2017.

Kim D.-S., Bonhoeffer T., 1994, Reverse occlusion leads to a precise restoration of orientation preference maps in visual cortex. *Nature*, 370:370–372.

Kistler W., Gerstner W., van Hemmen J. L., 1997, Reduction of the Hodgkin-Huxley equations to a single-variable threshold model. *Neural Comput.*, 9:1015–1045.

Kuffler S. W., 1953, Discharge patterns and functional organization of the mammalian retina. *J. Neurophysiol.*, 16:37–68.

Kuznetsov P. I., Stratonovich R. L., Tikhonov V. I., 1965, The transmission of random functions through non-linear systems. In Kuznetsov P. I., Stratonovich R. L., Tikhonov V. I., editors, *Non-Linear Transformations of Stochastic Processes*, pages 29–58. Pergamon Press, Oxford.

Law M. I., Zahs K. R., Stryker M. P., 1988, Organization of primary visual cortex (area 17) in the ferret. *J. Comp. Neurol.*, 278:157–180.

LeVay S., Hubel D. H., Wiesel T. N., 1975, The pattern of ocular dominance columns in macaque visual cortex revealed by a reduced silver stain. *J. Comp. Neurol.*, 159:559–576.

LeVay S., Stryker M. P., Shatz C. J., 1978, Ocular dominance columns and their development in layer IV of the cat's visual cortex: A quantitative study. *J. Comp. Neurol.*, 179:223–244.

Linsker R., 1986a, From basic network principles to neural architecture: Emergence of orientation columns. *Proc. Natl. Acad. Sci. USA*, 83:8779–8783.

Linsker R., 1986b, From basic network principles to neural architecture: Emergence of orientation selective cells. *Proc. Natl. Acad. Sci. USA*, 83:8390–8394.

Linsker R., 1986c, From basic network principles to neural architecture: Emergence of spatial-opponent cells. *Proc. Natl. Acad. Sci. USA*, 83:7508–7512.

Luskin M. B., Shatz C. J., 1985a, Cogeneration of subplate and marginal zone cells in the cat's primary visual cortex. *J. Neurosci.*, 5:1062–1075.

Luskin M. B., Shatz C. J., 1985b, Neurogenesis of the cat's primary visual cortex. *J. Comp. Neurol.*, 242:611–631.

MacKay D. J. C., Miller K. D., 1990, Analysis of Linsker's application of Hebbain rules to linear networks. *Network*, 1:257–297.

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

Mar D. J., Chow C. C., Gerstner W., Adams R. W., Collins J. J., 1999, Noise shaping in populations of coupled model neurons. *Proc. Natl. Acad. Sci. USA*, 96:10450–10455.

Markram H., Lübke J., Frotscher M., Sakmann B., 1997, Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275:213–215.

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. *J. Neurosci.*, 14:409–441.

Miller K. D., Erwin E., Kayser A., 1999, Is the development of orientation selectivity instructed by activity? *J. Neurobiol.*, 41:44–57.

Miller K. D., Keller J. B., Stryker M. P., 1989, Ocular Dominance Column Development: Analysis and Simulation. *Science*, 245:605–615.

Nelson S., Toth L., Sheth B., Sur M., 1994, Orientation selectivity of cortical neurons during intracellular blockade of inhibition. *Science*, 265:774–777.

Nicholls J. G., Martin A. R., Wallace B. G., 1992, *From Neuron to Brain*. Sinauer Associates, Inc., Sunderland, Massachusetts, 2nd edition.

Obermayer K., Blasdel G. G., 1993, Geometry of orientation and ocular dominance columns in monkey striate cortex. *J. Neurosci.*, 13:4114–4129.

Orbach H. S., Cohen L. B., Grinvald A., 1985, Optical mapping of electrical activity in rat somatosensory and visual cortex. *J. Neurosci.*, 5:1886–1895.

Rauschecker J. P., Singer W., 1979, Changes in the circuitry of the kitten visual cortex are gated by postsynaptic activity. *Nature*, 280:58–60.

Reid R. C., Alonso J. M., 1995, Specificity of monosynaptic connections from thalamus to visual cortex. *Nature*, 378:281–284.

Sengpiel F., Gödecke I., Stawinski P., Hübener M., Lowel S., Bonhoeffer T., 1998, Intrinsic and environmental factors in the development of functional maps in cat visual cortex. *Neuropharmacology*, 37:607–621.

Sengpiel F., Stawinsky P., Bonhoeffer T., 1999, Influence of experience on orientation maps in cat visual cortex. *Nature Neurosci.*, 2:727–732.

Shatz C. J., 1981, Inside-out pattern of neurogenesis of the cat's lateral geniculate nucleus. *Soc. Neurosci. Abstr.*, 6:485.

Shatz C. J., 1983, The prenatal development of the cat's retinogeniculate pathway. *J. Neurosci.*, 3:482–499.

Shatz C. J., Luskin M. B., 1986, The relationship between the geniculocortical afferents and their cortical target cells during development of the cat's primary visual cortex. *J. Neurosci.*, 6:3655–3668.

Shmuel A., Grinvald A., 1996, Functional organization for direction of motion

and its relationship to orientation maps in cat area 18. *J. Neurosci.*, 16:6945–6964.

Sillito A. M., 1979, Inhibitory mechanisms influencing complex cell orientation selectivity and their modification at high resting discharge levels. *J. Physiol. (Lond.)*, 289:33–53.

Sillito A. M., Kern J. A., Milson J. A., Berardi N., 1980, A re-evaluation of the mechanisms underlying simple cell orientation selectivity. *Brain Res.*, 194:517–520.

Somers D. C., Nelson S. B., Sur M., 1995, An emergent model of orientation selectivity in cat visual cortical simple cells. *J. Neurosci.*, 15:5448–5465.

Sperry R. W., 1943, Effect of 180° rotation of the retinal field on visuomotor coordination. *J. Exp. Zool.*, 92:236–279.

Stetter M., Lang E. W., Müller A., 1993, Emergence of orientation selective simple cells simulated in deterministic and stochastic neural networks. *Biol. Cybern.*, 68:465–476.

Stryker M. P., Harris W. A., 1986, Binocular impulse blockade prevents the formation of ocular dominance columns in cat visual cortex. *J. Neurosci.*, 6:2117–2133.

Swindale N. V., Matsubara J. A., Cyander M. S., 1987, Surface organization of orientation and direction selectivity in cat area 18. *J. Neurosci.*, 7:1414–1427.

Thompson R. F., 1993, *The Brain. A Neuroscience Primer.* W. H. Freeman and Company, New York, 2^{nd} edition.

Thomson A. M., Deuchars J., 1994, Temporal and spatial properties of local circuits in neocortex. *Trends Neurosci.*, 17(3):119–126.

Tsodyks M., Pawelzik K., Markram H., 1998, Neural networks with dynamic synapses. *Neural Comput.*, 10:821–835.

Turrigiano G. G., Leslie K. R., Desai N. S., Rutherford L. C., Nelson S. B., 1998, Activity-dependent scaling of quantal amplitude in neocortical neurons. *Nature*, 391:892–896.

Tusa R. J., Palmer L. A., Rosenquist A. C., 1978, The retinotopic organization of area 17 (striate cortex) in the cat. *J. Comp. Neurol.*, 177:213–235.

Varela J. A., Sen K., Gibson J., Fost J., Abbott L. F., Nelson S. B., 1997, A Quantitative Description of Short-Term Plasticity at Excitatory Synapses in Layer 2/3 of Rat Primary Visual Cortex. *J. Neurosci.*, 17:7926–7940.

Varela J. A., Song S., Turrigiano G. G., Nelson S. B., 1999, Differential depression at excitatory and inhibitory synapses in visual cortex. *J. Neurosci.*, 19:4293–4304.

Vidyasagar T. R., Pei X., Volgushev M., 1996, Multiple mechanisms underlying the orientation selectivity of visual cortical neurones. *Trends Neurosci.*, 19:272–277.

Weliky M., Bosking W., Fitzpatrick D., 1996, A systematic map of direction preference in primary visual cortex. *Nature*, 379:725–728.

Weliky M., Katz L. C., 1997, Disruption of orientation tuning in visual cortex

by artificially correlated neuronal activity. *Nature*, 386:680–685.

Weliky M., Katz L. C., 1999, Correlational structure of spontaneous neuronal activity in the developing lateral geniculate nucleus in vivo. *Science*, 285:599–604.

Wiesel T. N., 1982, Postnatal development of the visual cortex and the influence of environment. *Nature*, 299:583–591.

Wiesel T. N., Hubel D. H., 1963a, Effects of visual deprivation on morphology and physiology of cells in the cat's lateral geniculate body. *J. Neurophysiol.*, 26:978–993.

Wiesel T. N., Hubel D. H., 1963b, Single-cell responses in striate cortex of kittens deprived of vision in one eye. *J. Neurophysiol.*, 26:1003–1017.

Wiesel T. N., Hubel D. H., 1965, Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *J. Neurophysiol*, 28:1029–1040.

Wimbauer S., 1996, *Raumzeitliche rezeptive Felder. Modellierung von Antworteigenschaften und Entwicklung einfacher Zellen im visuellen Cortex*, volume 62 of *Reihe Physik*. Harri-Deutsch Verlag, Frankfurt / Main. Dissertation 1996, TU München.

Wimbauer S., Gerstner W., van Hemmen J. L., 1994, Emergence of spatiotemporal receptive fields and its application to motion detection. *Biol. Cybern.*, 72:81–92.

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.

Wimbauer S., Wenisch O. G., van Hemmen J. L., Miller K. D., 1997a, Development of spatiotemporal receptive fields of simple cells: II. Simulation and analysis. *Biol. Cybern.*, 77:463–477.

Wimbauer S., Wenisch O. G., Miller K. D., van Hemmen J. L., 1997b, Development of spatiotemporal receptive fields of simple cells: I. Model formulation. *Biol. Cybern.*, 77:453–461.

Wolf F., Bauer H.-U., Pawelzik K., Geisel T., 1996, Organization of the visual cortex. *Nature*, 382:306–307.

Wolfram S., 1999, *The Mathematica Book*. Cambridge University Press, 4^{th} edition.

Yuille A. L., Kammen D. M., Cohen D. S., 1989, Quadrature and the development of orientation selective cortical cells by Hebb rules. *Biol. Cybern.*, 61:183–194.

Zeki S., 1994, *A Vision of the Brain*. Blackwell Scientific Publications, Oxford.

Zhang L. I., Tao H. W., Holt C. E., Harris W. A., Poo M.-M., 1998, A critical window for cooperation and competition among developing retinotectal synapses. *Nature*, 395:37–44.

Zucker R. S., 1989, Short-term synaptic plasticity. *Ann. Rev. Neurosci.*, 12:13–31.

Acknowledgements

It is a great pleasure to express my sincere gratitude to my supervisor Prof. Dr. J. Leo van Hemmen, who supported my work in a very open-minded way and was always willing to spend his time discussing my problems. His enthusiasm, advice and insight, his numerous ideas, critical questions and helpful comments have contributed substantially to the progress and success of this project.

I am also indebted to all my fellows at Prof. Dr. van Hemmen's institute. In particular, I thank Julian Eggert and Oliver Wenisch for a pleasant and cooperative atmosphere in the office and for many constructive discussions. I really appreciate their humor and their readiness to help whenever necessary. Oliver Wenisch and Christian Leibold have read large portions of the manuscript and have offered lots of valuable comments. Further thanks are due to Uli Hillenbrand, Werner Kistler, Richard Kempter, Moritz Franosch, Claudia Bergmann, Berthold Bäuml, Diana David, Stefan Wimbauer, Wulfram Gerstner, Marcus Böttcher, Alexander Raach, and Brigitte Raab. They were members of a very creative and friendly group that I was fortunate to join.

I am most grateful to my family, especially to my parents, and to my friends Sabine and Stefan for their great patience and permanent encouragement during the past years, but also for reminding me sometimes that my life is more than my current work. Owing to their help and support I am now in the situation to finish this thesis.

Finally, I gratefully acknowledge the Graduiertenkolleg "Sensorische Interaktion in biologischen und technischen Systemen" in Munich for support in the form of a scholarship, with particular thanks to Dr. Isolde von Bülow.