

Probabilistic Theories of the Visual Cortex

A.L. Yuille and D. Kersten

Abstract

MODELING NEURONS BY PROBABILITIES

The purpose of this section is to introduce models of neurons, some simple examples (e.g., ON/OFF cells), probabilistic modeling in general, MRF's, and how these models can be tested with single- and multi-electrode recordings.

What are neurons and how should we model them? Neurons consist of a cell body, or soma, which receives inputs from a tree of dendrites and which outputs action potentials along an axon. In a group of neurons, the input to the dendrites are the action potentials from other neurons mediated by the synapses. The biophysics of neurons is well understood but the resulting models are often very complex. Researchers attempt to find simplified models which capture the underlying structure and predict the responses well (neurons can be tested extensively "in vitro" where they are studied in isolation).

In this section we introduce the classic integrate-and-fire model of neurons. We start by considering a single neuron and then discuss the activities of groups of neurons. We treat these from a probabilistic perspective because it uses the same mathematical language as the computational models which we will discuss in later sections. But we will also discuss how alternative non-probabilistic models can be derived from this probabilistic framework.

A. Probability Models of Single Neurons

Purpose is to describe probability model for a single neuron. Say how this often relates to linear filtering plus thresholding.

In the classic integrate-and-fire model a neuron i receives input I_j at a dendrite j . These inputs are weighted by the synaptic strengths T_{ij} and sent along the dendrites to the soma. At the soma, these weighted inputs are summed linearly to yield $\sum_j T_{ij}I_j$. If this sum is above a threshold value T_i then the neuron fires an action potential. In the probabilistic framework, the probability of firing is modeled stochastically. In this chapter we will use a simple model where the neuron fires $x_i = 1$, or does not fire $x_i = 0$, with probability:

$$P(x_i|\vec{I}) = \frac{\exp\{x_i(\sum_j T_{ij}I_j - T_i)\}}{1 + \exp\{\sum_j T_{ij}I_j - T_i\}}, \quad (1)$$

The probability of firing ($x_i = 1$) is given by a sigmoid function:

$$\sigma(\sum_j T_{ij}I_j - T_i) = \frac{1}{1 + \exp\{\sum_j T_{ij}I_j - T_i\}}. \quad (2)$$

The sigmoid function is approximately linear for small inputs, saturates at value 1 for large positive inputs, and suppresses large negative inputs to 0. Hence there is a linear regime where the probability of firing is $\sum_j T_{ij}I_j - T_i$.

We can obtain a deterministic model of a neuron by calculating the expected response of the neuron (variants of this trick will be used in this chapter to relate probabilistic and deterministic models of neurons and computations). The mean firing rate is given by:

$$\sum_{ox_i} x_i P(x_i|\vec{I}) = P(x_i = 1|\vec{I}) = \sigma(\sum_j T_{ij}I_j - T_i). \quad (3)$$

This is equivalent to linear filtering the input to obtain $\sum_j T_{ij} I_j$, subtracting T_i , and then applying a sigmoid non-linearity. Hence for certain values of the input, the mean output will appear as the linear filter of the input. We will discuss this further in the next subsection.

A more advanced model of neurons treats the probability of firing by a Poisson distribution. Let λ be a measure of the activity at the soma, for example $\lambda = \exp\{\sum_j T_{ij} I_j - T_i\}$, and Δ be a time constant, then the probability of the neuron firing n times within period Δ is given by:

$$P(n|\lambda) = \frac{(\lambda\Delta)^n}{n!} \exp\{-\lambda\Delta\}. \quad (4)$$

Our simple model, see equation (1), is very similar to the Poisson model in the case when $\lambda\Delta$ is very small (e.g., by making the time period Δ small). In this situation, the Poisson model will have almost zero probability of firing more than once (i.e. $P(n|\lambda) \approx 0$ for $n \geq 2$) and the probability of firing or not-firing can be approximated by $\lambda\Delta/(1 + \lambda\Delta)$ and $1/(1 + \lambda\Delta)$, in agreement with our simple model.

Caveats: Real neurons are more complicated. The detailed biophysics of neurons has been studied (how electrical current flow along dendrites and spikes along the axons). Part of this is modeled by the Hodgkin-Huxley equation [4]. There are also studies about how computations which might be performed on the dendritic trees of the neurons [7],[1]. But many of these models are too complex to be useful. Instead some researchers try to derive analytic models which, while more complex than the linear-sum and fire model, are simpler than the detailed biophysical models. In particular, it has been shown that in a population of thin dendrites emanating from a main trunk or soma, the overall input-output behavior of a dendritic subtree can be described by a 2-layer model, where the first level consists of a set of independent dendritic subunits with a stereotyped input-output function, and the second layer, corresponding to the soma sums all the dendritic outputs and applies a response function [6]. The 2-layer model is $f(\vec{x}) = S(\sum_i D_i(x_i) + I_0)$. f is the firing rate, \vec{x} is the input, S is the axo-somatic nonlinearity, D_i is the dendrite i 's current input-output function, x_i is input configuration in dendrite i . Linear regression was used to determine the parameters $D_i(x_i)$ for the model. More recent work from the same research group (Mel's group at USC) used regression method to fit more complex 2-layer models (multi-layer perceptrons) to both detailed biophysical models of neurons and data obtained from in vitro experiments (where a neuron is extracted from the brain, is stimulated electrically at different positions of the neuron, and the out at the axon is measured).

B. Experimental Evidence: Linear and Non-Linear Filters

Purpose is to describe how linear models model the receptive fields of neurons in the retina and LGN. I.e. the models do relate to real data. We might want to describe the computational rationale for these models better – e.g., show that they deal with contrast, also at least mention Atick's information theory analysis.

How well do classic models of neurons describe the performance of neurons in the early visual system? Recall that the mean firing rate is a non-linear function (a sigmoid function) of terms which are linear in the input $\sum_j T_{ij} I_j - T$. Many of methods for analyzing neurons also assume that the neuron can also be expressed in this general form (sometimes referred to as a Generalized Linear Model). For convenience, we assume that the non-linear function is known (in practice it is estimated) and we work in the linear regime. see equation (3).

An advantage of studying the early visual system is that we can directly observe the properties of neurons by showing simple stimuli. The classic receptive field of a sensory neuron is the region of space in which the presence of a stimulus will alter the firing of that neuron. The receptive field of a photoreceptor in the eye is specified by the set of directions from which light will stimulate the cell. The receptive field of a ganglion cell in the retina of the eye is composed of input from all of the photoreceptors which synapse with it, and a group of ganglion cells in turn forms the receptive field for a cell in the brain. In this way, simple small receptive fields in the retina and lateral geniculate nucleus (LGN) can combine to build larger and more complex receptive fields for cells in the visual cortex and extrastriate cortex.

The receptive field properties of cells in the retina, LGN, and early visual cortex have been classified. Below we give simple mathematical models for receptive fields of ON-OFF cells (in retinal ganglion and lateral geniculate nucleus) and for the oriented simple and complex cells found in the early visual cortex (V1). We stress that these mathematical models are first order approximations and there is considerable evidence suggesting that the receptive field properties of cell are more complicated. In particular, there are the so-called non-classical receptive fields, where the cells' responses are affected by inputs outside their classical receptive fields. These responses may be mediated by lateral (sideways) connections and/or to feedback from higher level cortical areas (both will be discussed later in the chapter). Other complications arise because of selection bias (the tendency to record only from those neurons which appear sensitive to the stimuli) and the restricted class of stimuli used to measure receptive fields (e.g., does the response of cells to oriented bars reflect their response to more realistic natural images). For a thorough review of these issues see [3].

More specifically, retinal ganglion cells and cells in the LGN are modeled as center-surround cells, which are either on-center cells – where light in the central and surround regions excite and inhibit the cell – or off-center cells where the surround is excitatory and the center is inhibitory. The purpose of this is believed to be adaptive gain control which reduces the enormous range of light which reaches the retina. Formally, we can express models of this type as:

$$X(\vec{p}) = T * I(\vec{p}) = \int T(\vec{p} - \vec{q}) I(\vec{q}) d\vec{q}, \quad \text{where } T(\cdot) \text{ is a linear filter,} \\ \text{and } * \text{ denotes convolution.} \quad (5)$$

In this equation, $I(\vec{p})$ represents the input image where $\vec{p} = (x, y)$ specifies the spatial position. For on-center and off-center cells, the function $T(\cdot)$ is radially symmetric and takes the form of a Mexican hat or inverted Mexican hat [5]. This is specified by the second order derivative of a Gaussian:

$$-\left\{ \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} \right\} G(\vec{p} : \vec{0}, \sigma^2 \mathbf{I}) \quad (6)$$

where $G(\vec{p} : \vec{0}, \sigma^2 \mathbf{I})$ is a two-dimensional Gaussian with zero mean and symmetric covariance $\sigma^2 \mathbf{I}$, where \mathbf{I} is the identity matrix.

Center and surround cells have several purposes. They serve to deal with the large dynamic range of natural images by filtering out the constant components (e.g., the output of the cell does not change if we add a spatially constant term to the input I). The forms of their receptive fields can be predicted from coding theory – [2] – where the goal is to encode the information at the retina in order to transmit it efficiently to the visual cortex.

To make a connection to the models in the previous section, we restrict the position variables \vec{p} to take a finite set of discrete values i (like the lattice of a digital camera). In this case, the integrals in equation (5) will reduce to summations of form $x_i = \sum T_{ij} I_j$, which is similar to the models in equation (3) if we remove the non-linearity and the threshold. We will discuss other classes of linear models, and some quadratic non-linear extensions, in a later section on low-level vision.

The handout by Talebi and Baker describes in detail how receptive fields of neurons are estimated (the essential idea is to use to fit models to the data by regression – regression relates to learning and will be discussed in a later lecture). But their paper raises a very important issue. Their findings (for neurons in the cortex of anaesthetized cats) show that estimates of the receptive fields of neurons can depend on what types of stimuli are shown. They estimate receptive fields using three different types of input: (i) white noise (WN), (ii) oriented bars (B), and (iii) natural images (NI). They show that the receptive fields learnt from the WN stimuli are very bad at predicting the response of the neurons to B or to NI. Overall, natural images (NI) are best at predicting the response to all types of stimuli. This is worrying because many researchers have used white noise as a stimulus to estimate receptive fields. In fact, the other handout (Gillisch and Meister) argue that the retina is more complex than most scientists believe, partly because

many properties of retinal neurons have been estimated using white noise, and the receptive fields are sometimes more complicated if other stimuli are used. (But the fact that some scientists can reconstruct the input image from electrode recordings in either the retina or the LGN – but not the visual cortex – suggests that the current models of the retina cannot be too bad – see Yang Dan's work). Why can different stimuli give different results for receptive fields? This may happen because the stimulus set may not contain stimuli which match the receptive field properties and so do not stimulate the neurons very strongly (e.g., some neurons only fire a few spikes when they are shown white noise, but fire a lot when shown bars at a specific orientation).

REFERENCES

- [1] S. D. Antic, W.-L. Zhou, A. R. Moore, S. M. Short, and K. D. Ikonomu. The decade of the dendritic NMDA spike. *Journal of neuroscience research*, 88(14):2991–3001, Nov. 2010.
- [2] J. Atick and A. Redlich. What does the retina know about natural scenes? *Neural Computation*, 4(2):196–210, 1992.
- [3] M. Carandini. Do We Know What the Early Visual System Does? *Journal of Neuroscience*, 25(46):10577–10597, Nov. 2005.
- [4] A. Hodgkin and A. Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *J. Physiol.(London)*, 117:500–544.
- [5] D. Marr. Vision. *W.H. Freeman*, 1982.
- [6] P. Poirazi, T. Brannon, and B. W. Mel. Pyramidal neuron as two-layer neural network. *Neuron*, 37(6):989–999, 2003.
- [7] N. Spruston. Pyramidal neurons: dendritic structure and synaptic integration. *Nature Reviews Neuroscience*, 9(3):206–221, Mar. 2008.