Contents

Introduction

Computational neuroscience is an approach to understanding the information content of neural signals by modeling the nervous system at many different structural scales, including the biophysical, the circuit, and the systems levels. Theoretical analysis and computational modeling are important tools for characterizing what nervous systems do, determining how they function, and understanding why they operate in particular ways.

Concept 1 (*Descriptive Models*)**.** Summarizing large amounts of experimental data compactly yet accurately, thereby characterizing what neurons and neural circuits do.

Concept 2 (*Mechanistic Models*)**.** Addressing the question of how nervous systems operate on the basis of known anatomy, physiology, and circuitry.

Concept 3 (*Interpretive Models*)**.** Using computational and information-theoretic principles to explore the behavioral and cognitive significance of various aspects of nervous system function, addressing the question of why nervous systems operate as they do.

Chapter 1

Neural Encoding I: Firing Rates and Spike Statistics

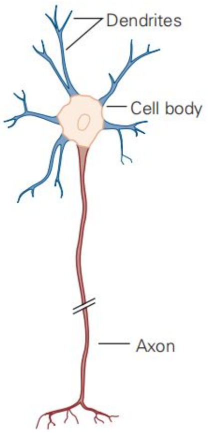
1.1 Introduction

1.1.1 The Explanation of Some Terms

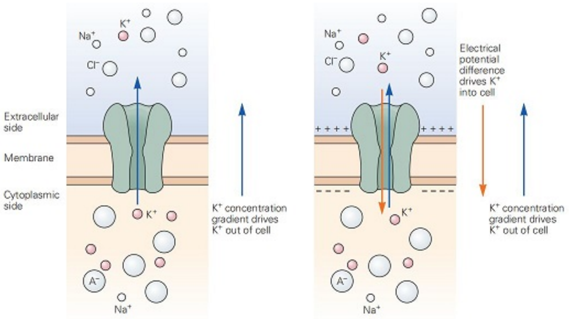
Remark 1.1. *Neurons* are highly specialized for generating electrical signals in response to chemical and other inputs, and transmitting them to other cells.

Remark 1.2. *Dendrites* receives information inputs from other neurons.

Remark 1.3. *Axon* carries the neuronal output to other cells.



Remark 1.4. *Ion channels* control the flow of ions across the cell membrane by opening and closing in response to voltage changes and to both internal and external signals.


Na$^+$
K$^+$
Cl$^-$
Extracellular side
Membrane
Cytoplasmic side
K$^+$
A$^-$
Na$^+$
K$^+$ concentration gradient drives K$^+$ out of cell
Na$^+$
K$^+$
Cl$^-$
Electrical potential difference drives K$^+$ into cell
+ + + +
+ + + +
K$^+$
A$^-$
Na$^+$
K$^+$ concentration gradient drives K$^+$ out of cell

Concept 4 (Membrane Potential)**.** The potential difference between two solutions separated by membranes, generally refers to the electrical phenomenon accompanying the life activities of cells, which exists on both side of cells.

Remark 1.5. Under resting conditions,the potential inside the cell membrane(mainly K$^+$) is negative, outside the cell membrane(mainly Na$^+$) is positive, and the cell is said to be *polarized*.

Definition 1.1 (Action Potential)**.** *Action potential* is the characteristic electrical pulses or, more simply, spikes that can travel down nerve fibers.

Concept 5 (Hyperpolarization)**.** Current in the form of positively charged ions flowing out of the cell (or negatively charged ions flowing into the cell) through open channels makes the membrane potential more negative, a process called *hyperpolarization*.

Concept 6 (Depolarization)**.** Current flowing into the cell changes the membrane potential to less negative or even positive values. This is called *depolarization*.

Remark 1.6. If a neuron is depolarized sufficiently to raise the membrane potential above a threshold level, a positive feedback process is initiated, and the neuron generates an *action potential*.

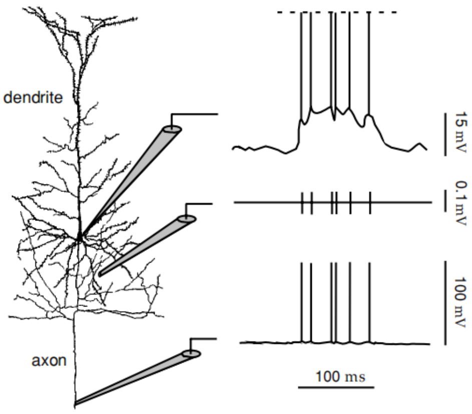
Concept 7 (Absolute Refractory Period)**.** For a few milliseconds just after an action potential has been fired, it may be virtually impossible to initiate another spike.

Concept 8 (Relative Refractory Period)**.** After the absolute refractory period, the excitability of cells gradually recovers. After stimulation, excitement can occur, but the stimulation must be greater than the original threshold intensity.

Remark 1.7. *Absolute refractory period* and *relative refractory period* are two basic phenomena in the process of neural response.

1.1.2 Recording Neuronal Responses

Example 1.2. Intracellular and extracellular methods for recording neuronal responses electrically



(i) The top trace represents a recording from an intracellular electrode connected to the soma of the neuron.

(ii) The middle trace is a simulated extracellular recording.

(iii) The bottom trace represents a recording from an intracellular electrode connected to the axon some distance away from the soma.

1.1.3 From Stimulus to Response

Remark 1.8. Neurons typically respond by producing complex spike sequences that reflect both the intrinsic dynamics of the neuron and the temporal characteristics of the stimulus.

Definition 1.3. Neural encoding refers to the map from stimulus to response.

Example 1.4. We can catalog how neurons respond to a wide variety of stimuli, and then construct models that attempt to predict responses to other stimuli.

Definition 1.5. Neural decoding refers to the reverse map, from response to stimulus.

Remark 1.9. The complexity and trial-to-trial variability of action potential sequences make it unlikely that we can describe and predict the timing of each spike deterministically. Instead, we seek a model that can account for the probabilities that different spike sequences are evoked by a specific stimulus.

1.2 1.2

Remark 1.10. Many physical laws are cumbersome when written in coordinate form but become more compact and attractive looking when written in tensorial form. For example, the incompressible Navier-Stokes equations in cylindrical coordinates are

$$\begin{aligned}\rho\left(\frac{Dv_r}{Dt}-\frac{v_\theta^2}{r}\right)&=\rho f_r-\frac{\partial p}{\partial r}+\mu\left(\Delta v_r-\frac{v_r}{r^2}-\frac{2}{r^2}\frac{\partial v_\theta}{\partial\theta}\right),\\ \rho\left(\frac{Dv_\theta}{Dt}+\frac{v_r v_\theta}{r}\right)&=\rho f_\theta-\frac{1}{r}\frac{\partial p}{\partial\theta}+\mu\left(\Delta v_\theta+\frac{2}{r^2}\frac{\partial v_r}{\partial\theta}-\frac{v_\theta}{r^2}\right),\\ \rho\frac{Dv_z}{Dt}&=\rho f_z-\frac{\partial p}{\partial z}+\mu\Delta v_z,\end{aligned}$$

where

$$\Delta=\frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right)+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}+\frac{\partial^2}{\partial z^2},$$

and

$$\frac{D}{Dt}=\frac{\partial}{\partial t}+v_r\frac{\partial}{\partial r}+\frac{v_\theta}{r}\frac{\partial}{\partial\theta}+v_z\frac{\partial}{\partial z}.$$

1.3 What Makes a Neuron Fire?

Remark 1.11. Response tuning curves characterize the average response of a neuron to a given stimulus. We now consider the complementary procedure of averaging the stimuli that produce a given response.

Remark 1.12. To average stimuli in this way, we need to specify what fixed response we will use to “trigger” the average. The most obvious choice is the firing of an action potential. Thus, we ask, “What, on average, did the stimulus do before an action potential was fired?” The resulting quantity, called the spike-triggered average stimulus, provides a useful way of characterizing neuronal selectivity.

1.3.1 Describing the Stimulus

Notation 1. Weber measured how different the intensity of two stimuli had to be for them to be reliably discriminated, the “just noticeable” difference Δs.

Principle 1.6 (Weber’s law)**.** Δs is proportional to the magnitude of the stimulus s, so that $\Delta s/s$ is constant for a given stimulus.

Remark 1.13. Fechner suggested that noticeable differences set the scale for perceived stimulus intensities.

Principle 1.7 (Fechner’s law)**.** Integrating Weber’s law, the perceived intensity of a stimulus of absolute intensity s varies as $\log s$.

Example 1.8. To deal with such wide-ranging stimuli, sensory neurons often respond most strongly to rapid changes in stimulus properties and are relatively insensitive to steady-state levels. Steady-state responses are highly compressed functions of stimulus intensity, typically with logarithmic or weak power-law dependences. This compression has an interesting psychophysical correlate.

Remark 1.14. Sensory systems make numerous adaptations, using a variety of mechanisms, to adjust to the average level of stimulus intensity. When a stimulus generates such adaptation, the relationship between stimulus and response is often studied in a potentially simpler regime by describing responses to fluctuations about a mean stimulus level.

Assumption 1.9. We frequently impose this condition that $s(t)$ satisfies

$$\frac{1}{T}\int_0^T s(t)dt = 0,$$

that is, $s(t)$'s time average over the duration of a trial is 0.

Remark 1.15. Our analysis of neural encoding involves two different types of averages: averages over repeated trials that employ the same stimulus, which we denote by angle brackets, and averages over different stimuli. We could introduce a second notation for averages over stimuli, but this can be avoided when using time-dependent stimuli.

Definition 1.10 (Stimulus and time averages)**.** Instead of presenting a number of different stimuli and averaging over them, we can string together all of the stimuli we wish to consider into a single time-dependent stimulus sequence and average over time. Thus, stimulus averages are replaced by *time averages.*

Proposition 1.11 (Periodic stimulus)**.** If integrals involving the stimulus are time-translationally invariant, then for any function h and time interval τ

$$\int_0^T h(s(t+\tau))dt = \int_\tau^{T+\tau} h(s(t))dt = \int_0^T h(s(t))dt. \quad (1.1)$$

To assure the last equality, we define the stimulus outside the time limits of the trial by the relation $s(T+\tau) = s(\tau)$ for any τ, thereby making the stimulus periodic.

1.3.2 The Spike-Triggered Average

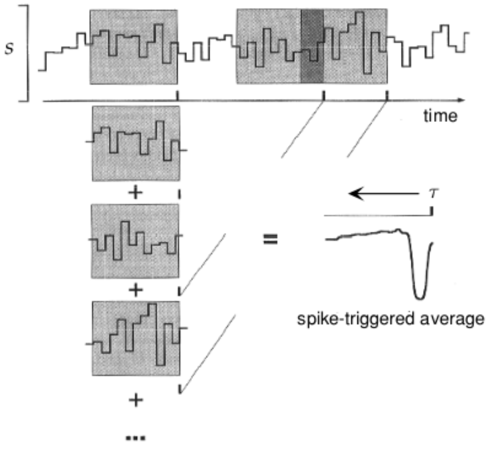
Definition 1.12. *The spike-triggered average* stimulus, $C(\tau)$, is the average value of the stimulus s a time interval τ before a spike is fired. The *spike-triggered average* $C(\tau)$ is a number of the form

$$C(\tau) = \left\langle \frac{1}{n}\sum_{i=1}^{n} s(t_i - \tau) \right\rangle \approx \frac{1}{\langle n\rangle}\left\langle \sum_{i=1}^{n} s(t_i - \tau) \right\rangle, \quad (1.2)$$

where n is the total number of spikes on each trial and t_i is the time when the spike occurs for $i = 1, 2, \cdots, n$. If n is large, it is well approximated by $\langle n\rangle$.

Remark 1.16. We make use of this approximation in Equation 1.2 because it allows us to relate the spike-triggered average to other quantities commonly used to characterize the relationship between stimulus and response (see Proposition 1.14).

Example 1.13. The following figure provides a schematic description of the computation of the spike-triggered average. Each time a spike appears, the stimulus in a time window preceding the spike is recorded. Although the range of τ values in Equation 1.2 is unlimited, the response is typically affected only by the stimulus in a window a few hundred milliseconds wide immediately preceding a spike. More precisely, we expect $C(\tau)$ to approach 0 for positive τ values larger than the correlation time between the stimulus and the response. If the stimulus has no temporal correlations with itself, we also expect $C(\tau)$ to be 0 for $\tau < 0$, because the response of a neuron cannot depend on future stimuli. In practice, the stimulus is recorded only over a finite time period, as indicated by the shaded areas in the figure below. The recorded stimuli for all spikes are then summed and the procedure is repeated over multiple trials.



Proposition 1.14. The spike-triggered average stimulus can be expressed as an integral of the stimulus times the neural response function of Equation **??**. If we replace the sum over spikes with an integral, as in Equation **??**, and use the approximate expression for $C(\tau)$ in Equation 1.2, we find

$$C(\tau) = \frac{1}{\langle n\rangle}\int_0^T \langle\rho(t)\rangle s(t-\tau)dt = \frac{1}{\langle n\rangle}\int_0^T \mathrm{r}(t)s(t-\tau)dt. \quad (1.3)$$

The second equality is due to the equivalence of $\langle\rho(t)\rangle$ and $\mathrm{r}(t)$ within integrals. Equation 1.3 allows us to relate the spike-triggered average to the correlation function of the firing rate and the stimulus.

Definition 1.15. The correlation function of the continuous functions f and g on $[0,T]$ is

$$R(\tau) = \frac{1}{T}\int_0^T f(t)g(t+\tau)dt. \quad (1.4)$$

Proposition 1.16. The correlation function of the firing rate and the stimulus (also called *firing-rate stimulus correlation function*) is

$$Q_{\mathrm{rs}} = \frac{1}{T}\int_0^T \mathrm{r}(t)s(t+\tau)dt. \quad (1.5)$$

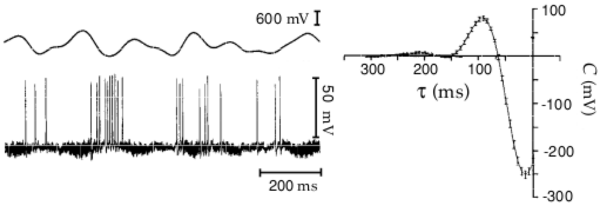
Remark 1.17. Correlation functions are a useful way of determining how two quantities that vary over time are related to one another. The two quantities being related are evaluated at different times, one at time t and the other at time $t+\tau$. The correlation function is then obtained by averaging their product over all t values, and it is a function of τ.

Proposition 1.17. By comparing Equations 1.3 and 1.5, we find that

$$C(\tau) = \frac{1}{\langle r\rangle}Q_{\mathrm{rs}}(-\tau), \quad (1.6)$$

where $\langle r\rangle = \langle n\rangle/T$ is the average firing rate over the set of trials. Because the argument of the correlation function in Equation 1.6 is $-\tau$, the spike-triggered average stimulus is often called *the reverse correlation function.*

Example 1.18. The following figure shows the spike-triggered average stimulus for a neuron in the electrosensory lateral-line lobe of the weakly electric fish *Eigenmannia.* Fluctuating electrical potentials, such as that shown in the upper left trace of the figure below, elicit responses from electrosensory lateral-line lobe neurons, as seen in the lower left trace. The spike-triggered average stimulus, plotted at the right, indicates that, on average, the electric potential made a positive upswing followed by a large negative deviation prior to a spike being fired by this neuron.



Remark 1.18. The spike-triggered average stimulus is widely used to study and characterize neural responses. Because $C(\tau)$ is the average value of the stimulus at a time τ before a spike, larger values of τ represent times farther in the past relative to the time of the triggering spike. For this reason, we spike-triggered averages with the time axis going backward compared to the normal convention. This allows the average spike-triggering stimulus to be read off from the plots in the usual left-to-right order.

Remark 1.19. The results obtained by spike-triggered averaging depend on the particular set of stimuli used during an experiment. How should this set be chosen? In chapter 2, we show that there are certain advantages to using a stimulus that is uncorrelated from one time to the next, a white-noise stimulus. A heuristic argument supporting the use of such stimuli is that in asking what makes a neuron fire, we may want to sample its responses to stimulus fluctuations at all frequencies with equal weight (i.e., equal power), and this is one of the properties of white-noise stimuli. In practice, white-noise stimuli can be generated with equal power only up to a finite frequency cutoff, but neurons respond to stimulus fluctuations only within a limited frequency range anyway. The figure in Example 1.18 is based on such an approximate white-noise stimulus.The power in a signal as a function of its frequency is called the power spectrum or power spectral density. White noise has a flat power spectrum.

1.3.3 White-Noise Stimuli

Definition 1.19. The defining characteristic of *white-noise stimulus* is that its value at any one time is uncorrelated with its value at any other time.

Proposition 1.20. The stimulus-stimulus correlation function (also called the *stimulus autocorrelation*) for white-noise stimulus $s(t)$ can be expressed by

$$Q_{ss}(\tau) = \sigma_s^2\delta(\tau) \tag{1.7}$$

with some constant σ_s.

Proof. By Definition 1.15, we have

$$Q_{ss}(\tau) = \frac{1}{T}\int_0^T s(t)s(t+\tau)dt. \tag{1.8}$$

Just as a correlation function provides information about the temporal relationship between two quantities, so an autocorrelation function tells us about how a quantity at one time is related to itself evaluated at another time. For white noise, the stimulus autocorrelation function is 0 in the range $-T/2 < \tau < T/2$ except when $\tau = 0$, thus, over this range we have Equation 1.7. □

Definition 1.21. The *power spectrum* for a stimulus $s(t)$ is the Fourier transform of the autocorrelation function of $s(t)$.

$$\widetilde{Q}_{ss}(\omega) = \frac{1}{T}\int_{-T/2}^{T/2} Q_{ss}(\tau)\exp(i\omega\tau)d\tau. \tag{1.9}$$

Remark 1.20. Because we have defined the stimulus as periodic outside the range of the trial T, we have used a finite-time Fourier transform and ω should be restricted to values that are integer multiples of $2\pi/T$.

Lemma 1.22. The power spectrum for a white-noise stimulus $s(t)$ is

$$\widetilde{Q}_{ss}(\omega) = \frac{\sigma_s^2}{T}, \tag{1.10}$$

which is the defining characteristic of white noise; its power spectrum is independent of frequency.

Proof. Using the fact that $Q_{ss}(\tau) = \sigma_s^2\delta(\tau)$ for white noise, we have

$$\widetilde{Q}_{ss}(\omega) = \frac{\sigma_s^2}{T}\int_{-T/2}^{T/2}\delta(t)\exp(i\omega\tau)d\tau = \frac{\sigma_s^2}{T}.$$

□

Proposition 1.23. Equation 1.7 is equivalent to the statement that white noise has equal power at all frequencies.

Solution. This conclusion is directly derived from Lemma 1.22.

Proposition 1.24. The *power spectrum* for a stimulus $s(t)$ satisfies

$$\widetilde{Q}_{ss}(\omega) = |\widetilde{s}(\omega)|^2. \tag{1.11}$$

Proof. Using the definition of the stimulus autocorrelation function, we can also write

$$\begin{aligned}
\widetilde{Q}_{ss}(\omega) &= \frac{1}{T}\int_0^T s(t)\frac{1}{T}\int_{-T/2}^{T/2} s(t+\tau)e^{i\omega\tau}d\tau dt \\
&= \frac{1}{T}\int_0^T s(t)e^{-i\omega t}\frac{1}{T}\int_{-T/2+t}^{T/2+t} s(t+\tau)e^{i\omega(t+\tau)}d(t+\tau)dt \\
&= \frac{1}{T}\int_0^T s(t)e^{-i\omega t}\frac{1}{T}\int_{-T/2}^{T/2} s(\tau)e^{i\omega(\tau)}d(\tau)dt \\
&= \frac{1}{T}\int_0^T s(t)e^{-i\omega t}dt\frac{1}{T}\int_{-T/2}^{T/2} s(\tau)e^{i\omega(\tau)}d(\tau),
\end{aligned}$$

where the seond step and third step follow from the variable substitution and the periodicity of the stimulus. The first integral on the right side of the forth equality is the complex conjugate of the Fourier transform of the stimulus,

$$\widetilde{s}(\omega) = \frac{1}{T}\int_0^T s(t)\exp(i\omega\tau)d\tau. \tag{1.12}$$

The second integral, because of the periodicity of the integrand (when ω is an integer multiple of $2\pi/T$) is equal to $\widetilde{s}(\omega)$. Therefore,

$$\widetilde{Q}_{ss}(\omega) = |\widetilde{s}(\omega)|^2, \tag{1.13}$$

which provides another definition of the stimulus power spectrum. It is the absolute square of the Fourier transform of the stimulus. □

Remark 1.21. No physical system can generate noise that is white to arbitrarily high frequencies. Approximations of white noise that are missing high-frequency components can be used, provided the missing frequencies are well above the sensitivity of the neuron under investigation.

Notation 2. To approximate white noise, we consider times that are integer multiples of a basic unit of duration Δt, that is, times $t = m\Delta t$ for $m = 1, 2, \cdots, M$ where $M\Delta t = T$. The function $s(t)$ is then constructed as a discrete sequence of stimulus values.

Proposition 1.25. In terms of the discrete-time values $s(t) = s_m$ for $(m-1)\Delta t \le t < m\Delta t$, the condition that the stimulus is uncorrelated is

$$\frac{1}{M}\sum_{m=1}^{M} s_m s_{m+p} = \begin{cases} \sigma_s^2/\Delta t & \text{if} \quad p = 0 \\ 0 & \text{otherwise.} \end{cases} \tag{1.14}$$

Remark 1.22. The factor of $1/\Delta t$ on the right side of Equation 1.14 reproduces the δ function of Equation 1.7 in the limit $\Delta t \to 0$. For approximate white noise, the autocorrelation function is 0 except for a region around $\tau = 0$ with width of order Δt. Similarly, the binning of time into discrete intervals of size Δt means that the noise generated has a flat power spectrum only up to frequencies of order $1/(2\Delta t)$.

Remark 1.23. An approximation to white noise can be generated by choosing each s_m independently from a probability distribution with mean 0 and variance $\sigma_s^2/\Delta t$. Any reasonable probability function satisfying these two conditions can be used to generate the stimulus values within each time bin. The factor of $1/\Delta t$ in the variance indicates that the variability must be increased as the time bins get smaller.

Example 1.26. A special class of white-noise stimuli, Gaussian white noise, results when the probability distribution used to generate the s_m values is a Gaussian function.

Remark 1.24. Although Equations 1.9 and 1.13 are both sound, they do not provide a statistically efficient method of estimating the power spectrum of discrete approximations to white-noise sequences generated by the methods described in this chapter.

Definition 1.27. The apparently natural procedure of taking a white-noise sequence $s(m\Delta t)$ for $m = 1, 2, \cdots, T/\Delta t$, and computing the square amplitude of its Fourier transform at frequency ω,

$$\frac{\Delta t}{T}\left|\sum_{m=1}^{T/\Delta t} s(t)\exp(-i\omega m\Delta t)\right|^2,$$

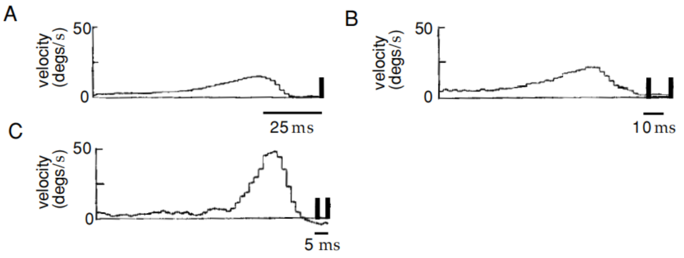
is a biased and extremely noisy way of estimating $\widetilde{Q}_{ss}(\omega)$. This estimator is called the *periodogram.*

Remark 1.25. The statistical problems with the periodogram, and some of the many suggested solutions, are discussed in almost any textbook on spectral analysis.

1.3.4 Multiple-Spike-Triggered Averages and Spike-Triggered Correlations

Remark 1.26. In addition to triggering on single spikes, stimulus averages can be computed by triggering on various combinations of spikes.

Example 1.28. The following pictures shows some examples of two-spike triggers. These results come from a study of the H1 movement-sensitive visual neuron of the blowfly. The H1 neuron detects the motion of visual images during flight in order to generate and guide stabilizing motor corrections. It responds to motion of the visual scene. In the experiments, the fly is held fixed while a visual image with a time-varying velocity $s(t)$ is presented. Figure A, showing the spike-triggered average stimulus, indicates that this neuron responds to positive angular velocities after a latency of about 15 ms. Figure B is the average stimulus prior to the appearance of two spikes separated by 10 ± 1 ms. In this case, the two-spike average is similar to the sum of two single-spike-triggered average stimuli displaced from one another by 10 ms. Thus, for 10 ms separations, two spikes occurring together tell us no more as a two-spike unit than they would individually. This result changes when shorter separations are considered. Figure C shows the average stimulus triggered on two spikes separated by 5 ± 1 ms. The average stimulus triggered on a pair of spikes separated by 5 ms is not the same as the sum of the average stimuli for each spike separately.



Remark 1.27. Spike-triggered averages of other stimulus-dependent quantities can provide additional insight into neural encoding, for example, spike-triggered average autocorrelation functions. Obviously, spike-triggered averages of higher-order stimulus combinations can be considered as well.

Assumption 1.29. From now on, assume that

$$\text{force on } S \text{ per unit area} = -p(\mathbf{x}, t)\mathbf{n} + \mathbf{n} \cdot \boldsymbol{\sigma}(\mathbf{x}, t), \quad (1.15)$$

where $\boldsymbol{\sigma}$ is the *(deviatoric) stress tensor* and $\mathbf{n}$ is the unit outward normal of S.

1.4 Spike-Train Statistics

Remark 1.28. A complete description of the stochastic relationship between a stimulus and a response would require us to know the probabilities corresponding to every sequence of spikes that can be evoked by the stimulus.

Lemma 1.30. The probability that z takes a value between z and $z + \Delta z$, for small Δ(strictly speaking, as $\Delta z \to 0$), is equal to $p[z]\Delta z$, where $p[z]$ is called a probability density.

Notation 3. Throughout this book, we use the notation $P[\ ]$ to denote probabilities and $p[\ ]$ to denote probability densities.

Theorem 1.31. The probability of a spike sequence appearing is proportional to the probability density of spike times, $p[t_1, t_2, ..., t_n]$. In other words, the probability $P[t_1, t_2, ..., t_n]$ that a sequence of n spikes occurs with spike i falling between times t_i and $t_i + \Delta t$ for i =1,2,...,n is given in terms of this density by the relation

$$P[t_1, t_2, ..., t_n] = p[t_1, t_2, ..., t_n](\Delta t)^n. \tag{1.16}$$

Proof.

$$P[t_1, t_2, ..., t_n] = \int ... \int p[s_1, s_2, ..., s_n]dS$$

$$= \int_{t_n-\Delta t/2}^{t_n+\Delta t/2} \int_{t_{n-1}-\Delta t/2}^{t_{n-1}+\Delta t/2} ... \int_{t_1-\Delta t/2}^{t_1+\Delta t/2} p[s_1, s_2, ..., s_n]ds_1...ds_{n-1}ds_n$$

According to the integral mean value theorem ($\Delta t \to 0$)
$\Rightarrow P[t_1, t_2, ..., t_n] = p[t_1, t_2, ..., t_n](\Delta t)^n$.

□

Definition 1.32 (*point process*)**.** A stochastic process that generates a sequence of events, such as action potentials ,is called a point process.

Remark 1.29. In general, the probability of an event occurring at any given time could depend on the entire history of preceding events.

Definition 1.33 (*renewal process*)**.** If this dependence extends only to the immediately preceding event, so that the intervals between successive events are independent, the point process is called a renewal process.

Definition 1.34. The Poisson process provides an extremely useful approximation of stochastic neuronal firing. To make the presentation easier to follow, we separate two cases, the homogeneous Poisson process, for which the firing rate is constant over time, and the inhomogeneous Poisson process, which involves a time-dependent firing rate.

1.4.1 The Homogeneous Poisson Process

Notation 4. We denote the firing rate for a homogeneous Poisson process by r(t) =r, because it is independent of time.

Definition 1.35 (*probality of n spikes occuring*)**.** The probality that an arbitrary sequence of exactly n spikes occurs within a trial of duration T is $P_T[n]$.

Theorem 1.36. For a homogeneous Poisson process, the Poisson distribution is

$$P_T[n] = \frac{(rn)^n}{n} exp(-rT). \tag{1.17}$$

Proof. To compute $P_T[n]$, we divide the time T into M bins of size $\Delta t = T/M$. We assume that Δt is small enough so that we never get two spikes within any one bin because, at the end of the calculation,we take the limit $\Delta t \to 0$.
$P_T[n]$ is the product of three factors:
(a) The probability of generating n spikes within a specified set of the M bins,$\frac{M!}{(M-n)!n!}$;
(b) The probability of not generating spikes in the remaining $M - n$ bins,$(r\Delta t)^n$;
(c) A combinatorial factor equal to the number of ways of putting n spikes into M bins,$(1 - r\Delta t)^{M-n}$;
To sum up,

$$P_T[n] = \lim_{\Delta t \to 0} \frac{M!}{(M-n)!n!}(r\Delta t)^n(1 - r\Delta t)^{M-n}. \tag{1.18}$$

As $\Delta t \to 0, M$ grows without bound because $M\Delta t = T$. Because n is fixed, we can write $M - n \approx M = T/\Delta t$. Using this approximatin and defining $\epsilon = -r\Delta t$, we find that

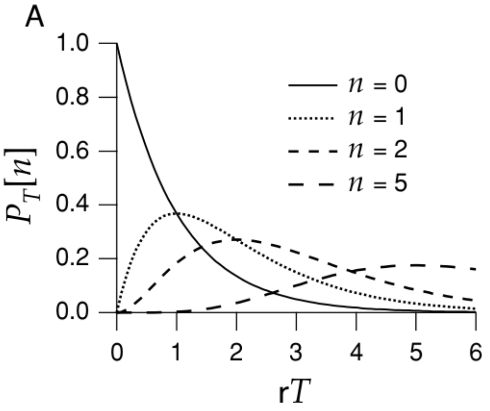
$$\lim_{\Delta t \to 0}(1-r\Delta t)^{M-n} = \lim_{\epsilon \to 0}(((1+\epsilon)^{\frac{1}{\epsilon}})^{-rT} = \exp(-rT) \tag{1.19}$$

For large M, $\frac{M!}{(M-n)!} \approx M^n = (T/\Delta t)^n$, so

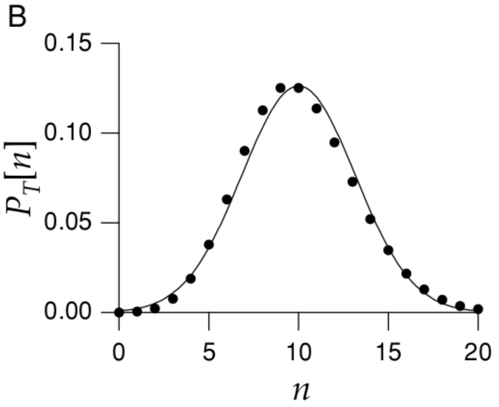
$$P_T[n] = \frac{(rn)^n}{n} exp(-rT). \tag{1.20}$$

□

Example 1.37. The probabilities $P_T[n]$, for a few n values, are plotted as a function of rT in the following firgue. Note that as n increase, the probability reaches its maximum at larger T values and that large n values are more likely than small ones for large T.



Example 1.38. The following figure shows the probabilities of various numbers of spikes occurring when the average number of spikes is 10. For large rT, which corresponds to a large expected number of spikes, the Poisson distribution approaches a Gaussian distribution with mean and variance equal to rT. This figure shows that this approximation is already quite good for $rT = 10$.



Theorem 1.39. The probability $P[t_1, t_2, ..., t_n]$ can be expressed in terms of another probability function $P_T[n]$, which is the probality that an arbitrary sequence of exactly n spikes occurs within a trial of duration T. Assuming that the spike times are ordered $0 \leq t_1 \leq t_2 \leq ... \leq t_n \leq T$, so that, the relationship is

$$P[t_1, t_2, ..., t_n] = n! P_T[n] \left(\frac{\Delta t}{T}\right)^n. \tag{1.21}$$

Proof. The probability of docking is $n!(\frac{\Delta t}{T})^n$ in a specific time order $(t_1, t_2, ..., t_n)$. so,

$$P[t_1, t_2, ..., t_n] = P_T[n](n(\frac{\Delta t}{T})(n-1)(\frac{\Delta t}{T})...1(\frac{\Delta t}{T})) \tag{1.22}$$

$$= n! P_T[n] \left(\frac{\Delta t}{T}\right)^n \tag{1.23}$$

□

Corollary 1.40. We can compute the variance of spike counts produced by a Poisson process from the probabilities in Equation 1.17. The spike count is

$$\sigma_n^2 = \langle n^2 \rangle - \langle n \rangle^2 = rT. \tag{1.24}$$

Proof. The average number of spikes generated by a Poisson process with constasnt rate r over a time T is

$$\langle n \rangle = \sum_{n=0}^{\infty} n P_T[n] = \sum_{n=0}^{\infty} \frac{n(rT)^n}{n!} \exp(-rT). \tag{1.25}$$

and the variance in the spike count is

$$\sigma_n^2(T) = \sum_{n=0}^{\infty} n^2 P_T[n] - \langle n \rangle^2 = \sum_{n=0}^{\infty} \frac{n^2 (rT)^n}{n!} \exp(-rT) - \langle n \rangle^2. \tag{1.26}$$

To compute the quantities,we need to calculate the two sums appearing in these Equations.A good way to do this is to compute the moment-generating function

$$g(\alpha) = \sum_{n=0}^{\infty} \frac{(rT)^n \exp(\alpha n)}{n!} \exp(-rT). \tag{1.27}$$

The kth derivative of g with respect to α,evaluated at the point $\alpha = 0$, is

$$\frac{dg}{d\alpha^k}|_{\alpha=0} = \sum_{n=0}^{\infty} \frac{n^k (rT)^n}{n!} \exp(-rT), \tag{1.28}$$

so once we have computed g,we need to calculate only its first and second derivative to determine the sums we need. Rearranging the terms a bit, and recalling that $\exp(z) = \sum z^n/n!$, we find

$$g(\alpha) = \exp(-rT) \sum_{n=0}^{\infty} \frac{(rT\exp(\alpha))^n}{n!} = \exp(-rT)\exp(rTe^{\alpha}). \tag{1.29}$$

The derivatives are then

$$\frac{dg}{d\alpha} = rTe^{\alpha} \exp(-rT)\exp(rTe^{\alpha}) \tag{1.30}$$

and

$$\frac{d^g}{d\alpha^2} = (rTe^{\alpha})^2 \exp(-rT)\exp(rTe^{\alpha}) + rTe^{\alpha}\exp(-rT)\exp(rTe^{\alpha}). \tag{1.31}$$

Evaluating these at $\alpha = 0$and putting the results into Equation 1.25 and 1.26 gives the result $\langle n \rangle = rT$ and

$$\sigma_n^2(T) = (rT)^2 + rT - (rT)^2 = rT.$$

□

Definition 1.41 (*Fano factor*)**.** The ratio of the variance and mean of the spike count, $\sigma_n^2/\langle n \rangle$, is called the Fano factor.

Example 1.42. The Fano factor takes the value 1 for a homogeneous Poisson process, independent of the time interval T.

Lemma 1.43. The probability of an interspike interval-falling between τ and $\tau + \Delta t$ is

$$P[\tau \leq t_{i+1} - t_i < \tau + \Delta t] = r\Delta t \ \exp(-r\tau). \tag{1.32}$$

Proof. Suppose that a spike occurs at a time t_i for some value of i. The probability of a homogeneous Poisson process generating the next spike somewhere in the interval

$$t_i + \tau \leq t_{i+1} \leq t_i + \tau + \Delta t,$$

for small Δt, is the probabilities that no spike is fired for a time τ, times the probability, $r\Delta t$, of generating a spike within the following small interval Δt. From Equation 1.17, with $n = 0$, the probability of not firing a spike for period τ is $\exp(-r\tau)$. So the probability of an interspike interval falling between τ and $\tau + \Delta t$ is

$$P[\tau \leq t_{i+1} - t_i < \tau + \Delta t] = r\Delta t \ \exp(-r\tau).$$

□

Theorem 1.44. From the interspike interval distribution of a homogeneous Poisson spike train, we can compute the mean interspike interval,

$$\langle \tau \rangle = \int_0^{\infty} \tau r \ \exp(-r\tau) d\tau = \frac{1}{r} \tag{1.33}$$

and the variance of the interspike intervals,

$$\sigma_\tau^2 = \int_0^{\infty} \tau^2 r \ \exp(-r\tau) d\tau - \langle \tau \rangle^2 = \frac{1}{r^2}. \tag{1.34}$$

Definition 1.45. The ratio of the standard deviation and the mean of interspike interval distribution.

$$C_V = \frac{\sigma_\tau}{\langle \tau \rangle}, \tag{1.35}$$

is the *the coefficient of variation*

Remark 1.30. The coefficient of variation takes the value 1 for a homogeneous Poisson process. This is a necessary, though not sufficient, condition to identify a Poisson spike train. Recall that the Fano factor for a Poisson process is also 1. For any renewal process, the Fano factor evaluated over long time intervals approaches the value C_V^2.

1.4.2 The Spike-Train Autocorrelation Funciton

Definition 1.46. The spike-train autocorrelation function,

$$Q_{\rho\rho}(\tau) = \frac{1}{T}\int_0^T \langle(\rho(t) - \langle r\rangle)(\rho(t+\tau) - \langle r\rangle)\rangle dt, \tag{1.36}$$

is the autocorrelation of the neural response function of Equation **??** with its average over time and trials substracted out.

Theorem 1.47. The autocorrelation function for a Poisson spike train generated at a constant rate $\langle r\rangle = r$ is

$$Q_{\rho\rho}(\tau) = r\delta(\tau) \tag{1.37}$$

Proof. The spike-train auto correlation function is constructed from data in the form of a histogram by dividing time into bins. The value of the histogram for a bin labeled with a positive or negative integer m is computed by determining the number of the times that any two spikes in the train are separated by a time interval lying between $(m-1/2)\Delta t$ and $(m+1/2)\Delta$ with Δt the bin size. This includes all pairings, even between a spike and itself. We call this number N_m. If the intervals between the n^2 spike pairs in the train were uniformly distributed over the range from 0 to T, there would be $n^2\Delta t/T$ intervals in each bin. This uniform term is removed from the autocorrelation histogram by subtracting $n^2\Delta t/T$ from N_m for all m. The spike-train autocorrelation histogram is then defined by dividing the resulting numbers by T, so the value of the histogram in bin m is $H_m = N_m/T - n^2\Delta/T^2$. For small bin sizes, the $m = 0$ term in the histogram counts the average number of spikes, that is $N_m = \langle n\rangle$ and in the limit $\Delta t \to 0$, $H_0 = \langle n\rangle/T$ is the average firing rate $\langle r\rangle$. Because other bins have H_m of order Δt, large $m = 0$ term is often removed from histogram plots. The spike-train autocorrlation function is defined as $H_m/\Delta t$ in the limit $\Delta t \to 0$, and it has the units of a firing rate squared. In this limit, the $m = 0$ bin becomes a δfuncitn, $H_0/\Delta t \to \langle r\rangle\delta(\tau)$.
As we can seen, the distribution of interspikde intervals for adjacent spikes in a homogeneous Poisson spike train is exponential(Equation 1.32). By contrast, the intervals between any two spikes(not necessarily adjacent) in such a train are uniformly distributed. As a result, the subtraction procedure outlined above gives $H_m = 0$ for all bins except for the $m = 0$ bin that contains the contribution of the zero intervals between spikes and themselves. The autocorrlation function for a Poisson spike train generated at a constant rate $\langle r\rangle = r$ is

$$Q_{\rho\rho}(\tau) = r\delta(\tau).$$

□

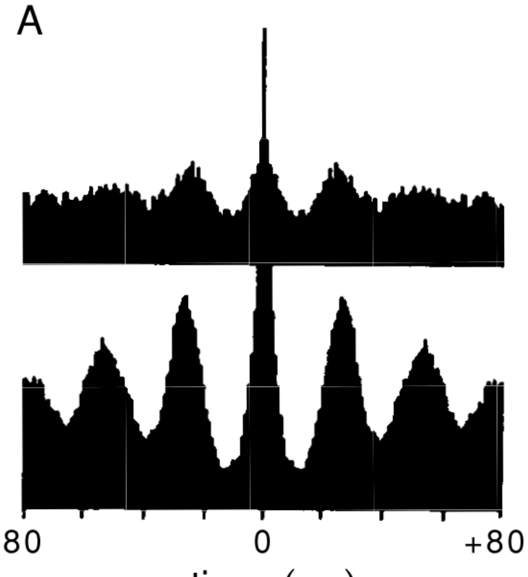
Definition 1.48. The spike-train correlation function ,

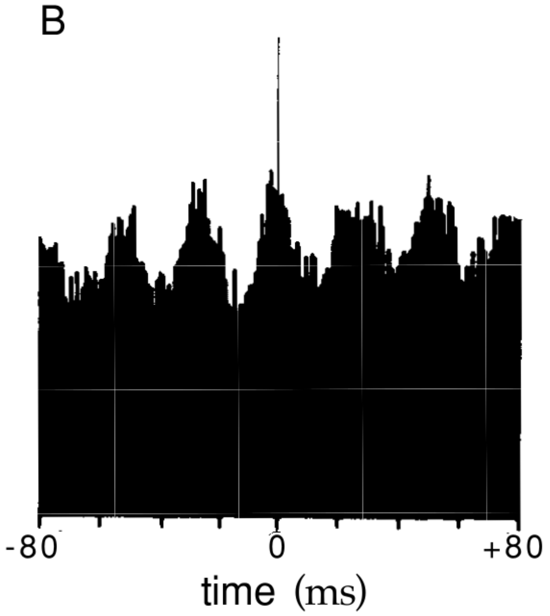
$$Q_{\rho_1\rho_2}(\tau) = \frac{1}{T}\int_0^T \langle(\rho_1(t) - \langle r_1\rangle)(\rho_2(t+\tau) - \langle r_2\rangle)\rangle dt, \tag{1.38}$$

is the correlation of different neural response function $\rho_1(t)$ and $\rho_2(t)$ with their average over time and trials which are r_1 and r_2 substracted out.

Remark 1.31. The spike-train autocorrelation function is an even function of τ, $Q_{\rho\rho}(\tau) = Q_{\rho\rho}(-\tau)$, but the cross-correlation function is not necessarily even.

Example 1.49. Asymmetric shifts in this peak away from 0 result from fixed delays between the firing of the twoneurons, and they indicate nonsynchronous but phase-locked firing. Periodic structure in either an autocorrelation or a cross-correlation function or histogram indicates that the firing probability oscillates. Such periodic structure is seen in the histograms of the following firgue, showing 40 Hz oscillations in neurons of catprimary visual cortex that are roughly synchronized between the two cerebral hemispheres.





1.4.3 The Inhomogeneous Poisson Process

Theorem 1.50. The probability density of the inhomogeneous Poisson Process for n spike times is

$$p[t_1,t_2,...,t_n]=\exp\left(-\int_0^T r(t)dt\right)\prod_{i=1}^{n} r(t_i), \tag{1.39}$$

The spike times are ordered $0\leq t_1\leq t_2\leq ...\leq t_n\leq T$.

Proof. The probability density for a particular spike sequence with spike times t_i for $i=1,2,...,n$ is obtained from the corresponding probability distribution by multiplying the probability that the spikes occur when they do by the probability that no other spikes occur.We begin by computing the probability that no spikes are generated during the time interval from t_i to t_{i+1} between two adjacent spikes. We determine this by dividing the interval into M bins of size Δt and setting $M\Delta t=t_{i+1}-t_i$. We will ultimately take the limit $\Delta t\to 0$. The firing rate during bin m within this interval is $r(t_i+m\Delta t)$. Because the probability of firing a spike in this bin is $r(t_i{+}m\Delta t)\Delta t$, the probabilities of not firing a spike is $1-r(t_i+m\Delta t)\Delta t$. To have no spikes during the entire interval, we must string together M such bins, and the probability of this occurring is the product of the individual probabilities,

$$P[\text{no spikes}]=\prod_{m=1}^{M}(1-r(t_i+m\Delta t)\Delta t). \tag{1.40}$$

We evaluate this expression by taking its logarithm,

$$\ln P[\text{no spikes}]=\sum_{m=1}^{M}\ln(1-r(t_i+m\Delta t)\Delta t), \tag{1.41}$$

using the fact that the logarithm of a product is the sum of the logarithms of the multiplied terms. Using the approximation $\ln(1-r(t_i+m\Delta t)\Delta t)\approx -r(t_i+m\Delta t)\Delta t$, valid for small Δt, we can simplify this to

$$\ln P[\text{no spikes}]=-\sum_{m=1}^{M} r(t_i+m\Delta t)\Delta t. \tag{1.42}$$

In the limit $\Delta t\to 0$, the approximation becomes exact and this sum becomes the integral of $r(t)$ from t_i to t_{i+1},

$$\ln P[\text{no spikes}]=-\int_{t_i}^{t_{i+1}} r(t)dt. \tag{1.43}$$

Exponentiating this Equation gives the result we need,

$$P[\text{no spikes}]=\exp\left(-\int_{t_i}^{t_{i+1}} r(t)dt\right). \tag{1.44}$$

The probability density $p[t_1,t_2,...,t_n]$is the product of the densities for the individual spikes and the probabilities of not generating spikes during the interspikde intervals, between time 0 and the first spike, and between the time of the last spike and the end of the trial period:

$$p[t_1,t_2,...t_n]=\exp\left(-\int_0^{t_1} r(t)dt\right)\exp\left(-\int_{t_n}^{T} r(t)dt\right)\times \\ r(t_n)\prod_{i=1}^{n-1} r(t_i)\exp\left(-\int_{t_i}^{t_{i+1}} r(t)dt\right). \tag{1.45}$$

The exponentials in this expression all combine because the product of exponentials is the exponential of the sum, so the different integrals in this sum add up to form a single integral:

$$\exp\left(-\int_0^{t_1} r(t)dt)\right)\exp\left(-\int_{t_n}^{T} r(t)dt\right)\prod_{i=1}^{n-1}\exp\left(-\int_{t_i}^{t_{i+1}} r(t)dt\right) \\ =\exp\left(-\left(\int_0^{t_1} r(t)dt+\sum_{i=1}^{n-1}\int_{t_i}^{t_{i+1}} r(t)dt+\int_{t_n}^{T} r(t)dt\right)\right) \\ =\exp\left(-\int_0^{T} r(t)dt\right). \tag{1.46}$$

Substituting this into Equation 1.45 gives the result in Equation 1.39 □

Remark 1.32. The eqution 1.21 is a special case of Equation 1.39.

1.4.4 The Poisson Spike Generator

Rule 1.51 (*Estimated firing rate*)**.** Spike sequences can be simulated by using some estimate of the firing rate, $r_{\text{est}}(t)$, predicted from knowledge of the stimulus, to drive a Poisson process.

Algorithm 1.52. The program progresses through time in small steps of size Δt and generates, at each time step, a random number x_{rand} chosen uniformly in the range between 0 and 1. If $r_{\text{est}}(t)\Delta t>x_{\text{rand}}$ at that time step, a spike is fired; otherwise it is not.

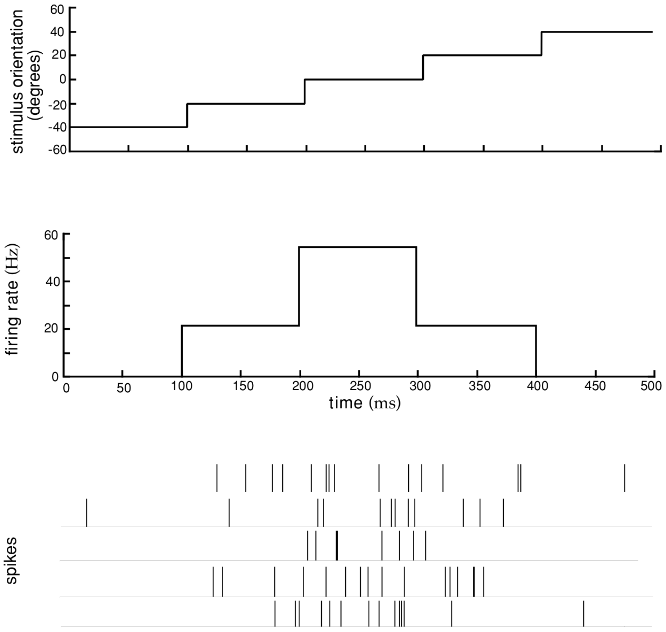
Algorithm 1.53. For a constant firing rate, it is faster to compute spike times t_i for $i=1,2,...,n$ iteratively by generating interspike intervals from an exponential probability density(Equation 1.32). Thus we can generate spike times iteratively from the formula $t_{i+1}=t_i-\ln(x_{\text{rand}}/r)$.

Remark 1.33. If x_{rand} is uniformly distributed over the range between 0 and 1, the negative of its logarithm is exponentially distributed.

Algorithm 1.54 (*Spike thinning*)**.** The thinning technique requires a bound $r_{\max}$ on the estimated firing rate such that $r_{\text{est}}(t) \leq r_{\max}$ at all times. We first generate a spike sequence corresponding to the constant rate r_{max} by iterating the rule $t_{i+1} = t_i - \ln(x_{\text{rand}})/r_{\max}$. The spike are then thinned by generating another x_{rand} for each i and removing the spike at time t_i from the train if $r_{\text{est}(t_i)}/r_{\max} < x_{\text{rand}}$. If $r_{\text{est}}(t_i)/r_{\max} \geq x_{\text{rand}}$, spike i is retained. Thinning corrects for the difference between the estimated timedependent rate and the maximum rate.

Example 1.55. The following figures shows an example of a model of an orientation-selective V1 neuron constructed by Spike thinning. In this model, the estimated firing rate is determined from the response tuning curve

$$r_{est}(t) = f(s(t)) = r_{max} \exp\left(-\frac{1}{2}\left(\frac{s(t) - s_{max}}{\sigma_f}\right)^2\right). \tag{1.47}$$



This figure Model of an orientation-selective neuron. The orientation angle (top panel) was increased from an initial value of $-40°$ by $20°$ every 100 ms. The firing rate (middle panel) was used to generate spikes (bottom panel) using a Poisson spike generator. The bottom panel shows spike sequences generated on five different trials.

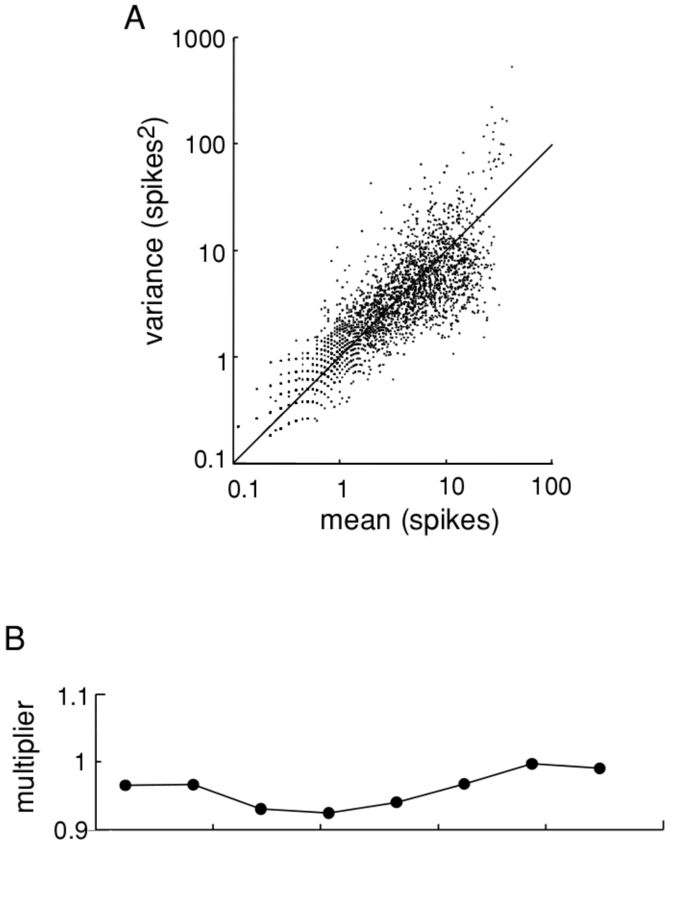
1.4.5 Comparison with Data

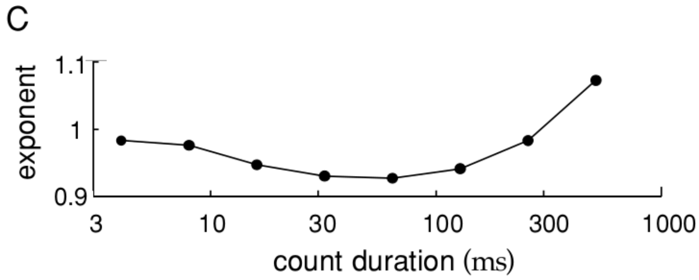
Remark 1.34. The Poisson process is simple and useful, but does it match data on neural response variability? To address this question, we examine Fano factors, interspike interval distributions, and coefficients of variation.

Proposition 1.56. The Fano factor describes the relationship between the mean spike count over a given interval and the spike-count variance. Mean spike counts $\langle n \rangle$ and variances σ_n^2 from a wide variety of neuronal recordings have been fitted to the Equation $\sigma_n^2 = A\langle n \rangle^B$, and the *multiplier* A and exponent B have been determined. The values of both A and B typically lie between 1.0 and 1.5.

Remark 1.35. Because the Poisson model predicts $A = B = 1$, this indicates that the data show a higher degree of variability than the Poisson model would predict. However, many of these experiments involve anesthetized animals, and it is known that response variability is higher in anesthetized than in alert animals.

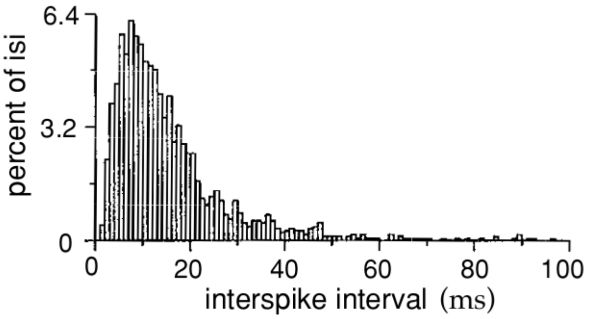
Example 1.57 (*comparison of the Fano factor*)**.** The following figures shows data for spike-count means and variances extracted from recordings of MT neurons in alert macaque monkeys using a number of different stimuli. The MT (medial temporal) area is a visual region of the primate cortex where many neurons are sensitive to image motion. The individual means and variances are scattered in figure A, but they cluster around the diagonal which is the Poisson prediction. Similarly, the results show A and B values close to 1, the Poisson values (figure B). Of course, many neural responses cannot be described by Poisson statistics, but it is reassuring to see a case where the Poisson model seems a reasonable approximation. As mentioned previously, when spike trains are not described very accurately by a Poisson model, refractory effects are often the primary reason.





Algorithm 1.58. Interspike interval distributions are extracted from data as interspike histograms by counting the number of intervals falling in discrete time bins.

Example 1.59 (*the Poisson model of interspike interval*)**.** The following figure presents an example from the responses of a nonbursting cell in area MT of a monkey in response to images consisting of randomly moving dots with a variable amount of coherence imposed on their motion (see chapter 3 for a more detailed description).



For interspike intervals longer than about 10 ms, the shape of this histogram is exponential, in agreement with Equation 1.32. However, for shorter intervals there is a discrepancy. While the homogeneous Poisson distribution of Equation 1.32 rises for short interspike intervals, the experimental results show a rapid decrease. This is the result of refractoriness making short interspike intervals less likely than the Poisson model would predict.

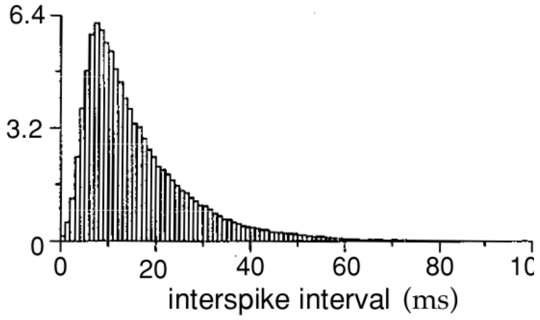
Remark 1.36.

Proposition 1.60. The data of the Poisson model of interspike interval with a stochastic refractory period can be fitted more accurately by a gamma distribution,

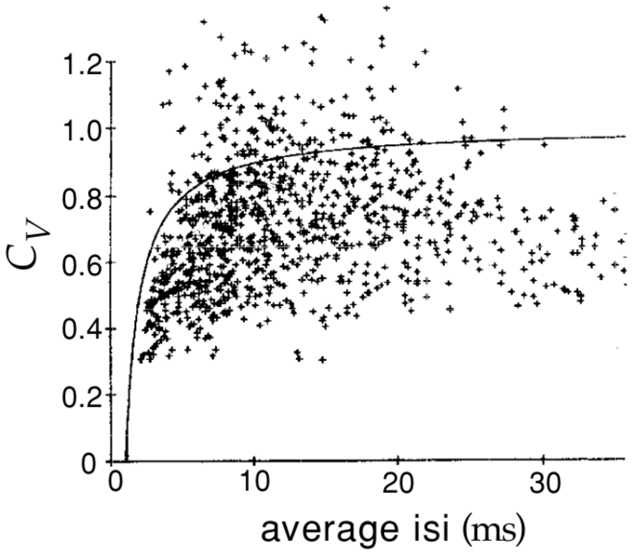
$$p[\tau] = \frac{r(r\tau)^k \exp(-r\tau)}{k!} \tag{1.48}$$

with $k > 0$, than by the exponential distribution of the Poisson model, which has $k = 0$.

Example 1.61 (*the Poisson model of interspike interval with a stochastic refractory period*)**.** The following figure shows a theoretical histogram obtained by adding a refractory period of variable duration to the Poisson model. Spiking was prohibited during the refractory period, and then was described once again by a homogeneous Poisson process. The refractory period was randomly chosen from a Gaussian distribution with a mean of 5 ms and a standard deviation of 2 ms (only random draws that generated positive refractory periods were included). The resulting interspike interval distribution of figure 1.4.5 agrees quite well with the data.

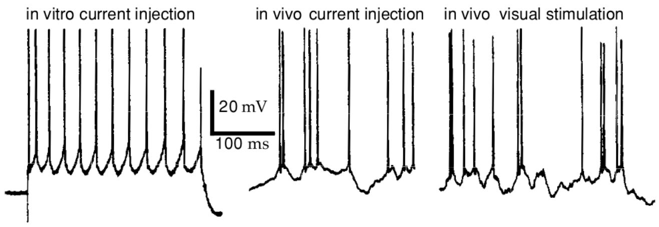


Example 1.62 (*comparion of the coefficients of variation*)**.** C_V values extracted from the spike trains of neurons recorded in monkeys from area MT and primary visual cortex(V1) are shown in this figure. The data have been divided into groups based on the mean interspike interval, and the coefficient of variation is plotted as a function of the mean interval, equivalent to $1/\langle r \rangle$. Except for short mean interspike intervals, the values are near 1, although they tend to cluster slightly lower than 1, the Poisson value. The small C_V values for short interspike intervals are due to the refractory period. The solid curve is the prediction of a Poisson model with refractoriness.



Remark 1.37. However, there are cases in which the accuracy in the timing and numbers of spikes fired by a neuron is considerably higher than would be implied by Poisson statistics. Furthermore, even when it successfully describes data, the Poisson model does not provide a mechanistic explanation of neuronal response variability.

Example 1.63. The following figure compares the response of V1 cells to constant current injection in vivo and in vitro. The in vitro response is a regular and reproducible spike train(left panel). The same current injection paradigm applied in vivo produces a highly irregular pattern of firing(center panel) similar to the response to a moving bar stimulus(right panel).



Although some of the basic statistical properties of firing variability may be captured by the Poisson model of spike generation, the spike generating mechanism itself in real neurons is clearly not responsible for the variability. We explore ideas about possible sources of spike-train variability in chapter 5.

Remark 1.38. Some neurons fire action potentials in clusters or bursts of spikes that can not be described by a Poisson process with a fixed rate. Bursting can be included in a Poisson model by allowing the firing rate to fluctuate in order to describe the high rate of firing during a burst. Sometimes the distribution of bursts themselves can be described by a Poisson process (such a doubly stochastic process is called a Cox process).

1.5 The Neural code

Example 1.64. Assuming that the neural response and its relation to the stimulus are completely characterized by the probability distribution of spike times as a function of the stimulus. If spike generation can be described as an inhomogeneous Poisson process, this probability distribution can be computed from the time-dependent firing rate r(t), using equation 1.37. In this case, r(t) contains all the information about the stimulus that can be extracted from the spike train, and the neural code could reasonably be called a rate code.

Remark 1.39. The central issue in neural coding is whether individual action potentials and individual neurons encode independently of each other, or whether correlations between different spikes and different neurons carry significant amounts of information.

1.5.1 Independent-Spike,Independent-Neuron,and Correlation Codes

Remark 1.40. All information in this section refers to stimulating information.

Definition 1.65 (*Independent-Spike Code*)**.** A code based solely on the time-dependent firing rate. This refers to the fact that the generation of each spike is independent of all the other spikes in the train.

Definition 1.66 (*Correlation Codes*)**.** Individual spikes do not encode independently of each other, correlations between spike times may carry additional correlation code information.

Remark 1.41. It has been found that some information is carried by correlations between two or more spikes, but this information is rarely larger than 10% of the information carried by spikes considered independently.Information could be carried by more complex relationships between spikes.Independent-spike codes are much simpler to analyze than correlation codes, and most work on neural coding assumes spike independence.

Rule 1.67. Information is typically encoded by neuronal populations.

Remark 1.42. We still consider whether individual neurons act independently, or whether correlations between different neurons carry additional information.

Remark 1.43. The analysis of population coding is easiest if the response of each neuron is considered statistically independent.It means that they can be combined without taking correlations into account.

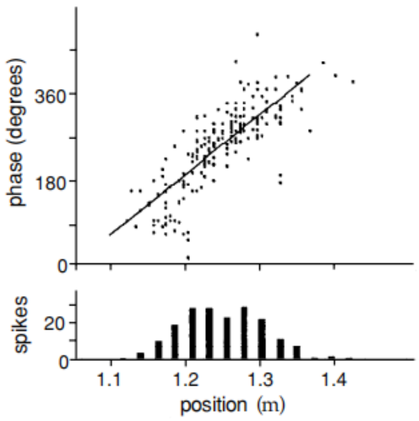
Definition 1.68 (*Independent-Neuron Code*)**.** *Independent-neuron code* means the response of each neuron in a neural population is considered statistically independent.

Remark 1.44. The assumption of independent-neuron coding is a useful simplification that is not in gross contradiction with experimental data, but it is less well established and more likely to be challenged in the future than the independent-spike hypothesis.

Remark 1.45. To test the validity of independent-neuron, we should know whether correlations between the spiking of different neurons provide additional information about a stimulus that cannot be obtained by considering all of their firing patterns individually.

Principle 1.69. Synchronous firing of two or more neurons and rhythmic oscillations of population activity are mechanism for conveying information in a population correlation code.

Example 1.70. Place-cell coding of spatial location in the rat hippocampus, which at least some additional information appears to be carried by correlations between the firing patterns of neurons in a population. The firing rates of many hippocampal neurons, recorded when a rat is moving around a familiar environment, depend on the location of the animal and are restricted to spatially localized areas called the place fields of the cells. When a rat explores an environment, hippocampal neurons fire collectively in a rhythmic pattern with a frequency in the theta range, 7-12 Hz. The spiking time of an individual place cell relative to the phase of the population theta rhythm gives additional information about the location of the rat not provided by place cells considered individually.



Each dot in the upper figure shows the phase of the theta rhythm plotted against the position of the animal at the time when a spike was fired. The linear relation shows that information about position is contained in the relative phase of firing. The lower plot is a conventional place field tuning curve of spike count versus position.

1.5.2 Temporal Codes

Rule 1.71. Precise spike timing is a significant element in neural encoding. When precise spike timing or high-frequency firing-rate fluctuations are found to carry information, the neural code is often identified as a temporal code.

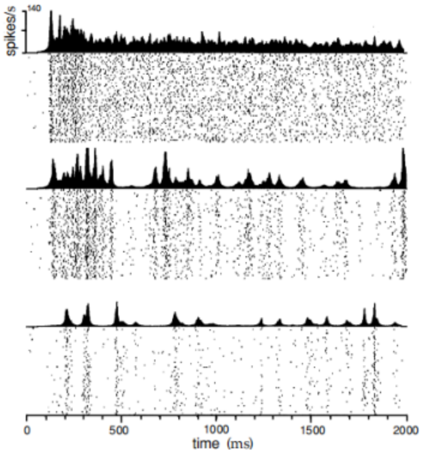
Remark 1.46. The temporal structure of a spike train or firing rate evoked by a stimulus is determined both by the dynamics of the stimulus and by the nature of the neural encoding process. Stimuli that change rapidly tend to generate precisely timed spikes and rapidly changing firing rates.

Remark 1.47. Temporal coding refers to temporal precision in the response that not only arise from the dynamics of the stimulus but also relates to properties of the stimulus.

Rule 1.72. If the independent-spike hypothesis is valid, the temporal character of the neural code is determined by the behavior of $\mathrm{r}(t)$.

Definition 1.73. If $\mathrm{r}(t)$ varies slowly with time, the code is typically called a *rate code*, and if it varies rapidly, the code is called *temporal code*.

Example 1.74. Different firing-rate behaviors for a neuron in area MT of a monkey recorded over multiple trials with three different stimuli(consisting of moving random dots). The activity in the top panel would typically be regarded as reflecting rate coding, and the activity in the bottom panel as reflecting temporal coding.



Remark 1.48. It is not obvious what criterion should be used to characterize the changes in $\mathrm{r}(t)$as slow or rapid. The identification of rate and temporal coding in this way is ambiguous.

Example 1.75. Using the spikes to distinguish slow from rapid, so that a temporal code is identified when peaks in the firing rate occur with roughly the same frequency as the spikes themselves. In this case, each peak corresponds to the firing of only one, or at most a few action potentials.

Remark 1.49. When many neurons are involved, any single neuron may fire only a few spikes before its firing rate changes, but the population may produce a large number of spikes over the same time period. Thus, it is not targeted at populations.

Example 1.76. Using the stimulus to establish what makes a temporal code. In this case, a temporal code is defined as one in which information is carried by details of spike timing on a scale shorter than the fastest time characterizing variations of the stimulus. This requires that frequencies higher than those present in the stimulus.

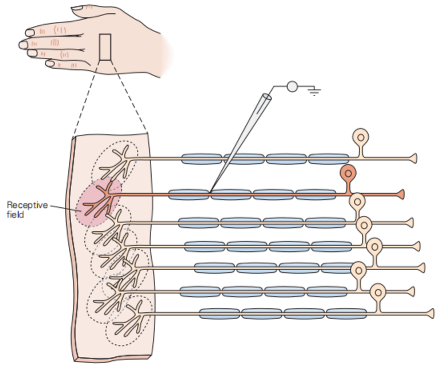
Rule 1.77. A temporal code has been reported when using spikes to define the nature of the code, and it would be called rate codes if the stimulus were used instead.

Chapter 2

Neural Encoding II: Reverse Correlation and Visual Receptive Fields

Definition 2.1. The skin area, location in the body, retinal area, or tonal domain in which stimuli can activate a sensory neuron is called its *receptive field*.

Remark 2.1. The following figure is the receptive field of a touch-sensitive neuron, which denotes the region of skin where gentle tactile stimuli evoke action potentials in that neuron. Sometimes receptive fields would change over time.



Definition 2.2. *Reverse-correlation* is a technique for studying how sensory neurons add up signals from different locations in their receptive fields, and also how they sum up stimuli that they receive at different times, to generate a response.

Remark 2.2. The goal of the reverse-correlation technique is to find a function $r = f(s)$ that maps from the stimulus s to the neuronal response r, where the stimulus is a function dependent on spatial location and time $s = s(x, y, z, t)$.

Remark 2.3. The reason that this technique is called "reverse" is that we align the time origin with the neuron's response and then reverse the timeline to find what stimulus ($t < 0$) triggered the neuron's response at the current moment ($t = 0$).

Assumption 2.3. As discussed in chapter 1, sensory systems tend to adapt to the absolute intensity of a stimulus. We therefore assume throughout this chapter that the stimulus parameter $s(t)$ has been defined with its mean value subtracted out, that is,

$$\frac{1}{T}\int_0^T s(t)dt = 0. \tag{2.1}$$

2.1 Estimating Firing Rates

Remark 2.4. The response tuning curve discussed in Chapter 1 is a simple model in which firing rates were estimated as instantaneous functions of the stimulus. Nevertheless, the activity of a neuron at time t typically depends on the behavior of the stimulus over a period of time starting a few hundred milliseconds prior to t and ending perhaps tens of milliseconds before t. Reverse-correlation methods can be used to construct a more accurate model that includes the effects of the stimulus over such an extended period of time.

Remark 2.5. The **basic problem** is to construct an estimate $r_{\text{est}}(t)$ of the firing rate $r(t)$ evoked by a stimulus $s(t)$.

2.1.1 The Linear Rate Estimate

Definition 2.4. The *linear rate estimate* at any given time t is the weighted sum of the values taken by the stimulus at earlier times. With the continuous change in time, this sum actually takes the form of an integral, that is,

$$r_{\text{est}}(t) = r_0 + \int_0^\infty D(\tau)s(t-\tau)d\tau, \tag{2.2}$$

where r_0 accounts for any background firing that may occur when $s = 0$, $D(\tau)$ is a weighting factor that determines how strongly, and with what sign, the value of the stimulus at time $t - \tau$ affects the firing rate at time t.

Remark 2.6. The integral in Equation 2.2 is a linear filter.

Definition 2.5. The *error* of an estimate $r_{\text{est}}(t)$ to an actual neural response $r(t)$ is defined as

$$E = \frac{1}{T}\int_0^T (r_{\text{est}}(t) - r(t))^2 dt. \tag{2.3}$$

Definition 2.6. The kernel D that minimizes the linear rate estimate error E defined in Equation 2.3 is called *optimal linear kernel* or simply called *optimal kernel*.

Proposition 2.7. The optimal kernel D satisfies

$$\int_0^\infty Q_{ss}(\tau-\tau')D(\tau')d\tau' = Q_{rs}(-\tau), \tag{2.4}$$

where $Q_{ss}(\tau)=\int s(t)s(t+\tau)/T$ is the stimulus autocorrelation function, and $Q_{rs}(\tau)=\int r(t)s(t+\tau)/T$ is the firing rate-stimulus correlation function, both of which were defined in chapter 1.

Proof. Using Equation 2.2 for the estimated firing rate, the expression in Equation 2.3 to be minimized is

$$E=\frac{1}{T}\int_0^T\left(r_0+\int_0^\infty D(\tau)s(t-\tau)d\tau - r(t)\right)^2. \tag{2.5}$$

The minimum is obtained by setting the derivative of E with respect to functional derivative the function D to 0. E that depends on a function D is a functional. Finding the extrema of functionals is the subject of a branch of mathematics called the calculus of variations. A simple way to define a functional derivative is to introduce a small time interval Δt and evaluate all functions at integer multiples of Δt. We define $r_i=r(i\Delta t)$, $D_k=D(k\Delta t)$ and $s_{i-k}=s((i-k)\Delta t)$. If Δt is small enough, the integrals in Equation 2.5 can be approximated by sums,

$$E=\frac{\Delta t}{T}\sum_{i=0}^{T/\Delta t}\left(r_0+\Delta t\sum_{k=0}^{\infty}D_k s_{i-k}-r_i\right)^2. \tag{2.6}$$

E is minimized by setting its derivative with respect to D_j for all values of j to 0,

$$\frac{\partial E}{\partial D_j}=0=\frac{2\Delta t}{T}\sum_{i=0}^{T/\Delta t}\left(r_0+\Delta t\sum_{k=0}^{\infty}D_k s_{i-k}-r_i\right)s_{i-j}\Delta t. \tag{2.7}$$

Rearranging and simplifying this expression gives the condition,

$$\Delta t\sum_{k=0}^{\infty}D_k\left(\frac{\Delta t}{T}\sum_{i=0}^{T/\Delta t}s_{i-k}s_{i-j}\right)=\frac{\Delta t}{T}\sum_{i=0}^{T/\Delta t}(r_i-r_0)s_{i-j}. \tag{2.8}$$

If we take the limit $\Delta t\to 0$ and make the replacements $i\Delta t\to t$, $j\Delta t\to\tau$, and $k\Delta t\to\tau'$, the sums in Equation 2.8 turn back into integrals, the indexed variables become functions, and we find

$$\begin{aligned}&\int_0^\infty D(\tau')\left(\frac{1}{T}\int_0^T s(t-\tau')s(t-\tau)dt\right)d\tau'\\ &=\frac{1}{T}\int_0^T(r(t)-r_0)s(t-\tau)dt.\end{aligned} \tag{2.9}$$

And,

$$\begin{aligned}&\frac{1}{T}\int_0^T s(t-\tau')s(t-\tau)dt\\ &=\frac{1}{T}\int_0^T s(t-\tau+\tau-\tau')s(t-\tau)d(t-\tau)\\ &=\frac{1}{T}\int_{-\tau}^{T-\tau}s(t+\tau-\tau')s(t)dt\\ &=\frac{1}{T}\int_0^T s(t+\tau-\tau')s(t)dt=Q_{ss}(\tau-\tau'),\end{aligned}$$

where the third step follows from the translation invariance of $s(t)$. Also,

$$\begin{aligned}&\frac{1}{T}\int_0^T(r(t)-r_0)s(t-\tau)dt\\ &=\frac{1}{T}\int_0^T r(t)s(t-\tau)dt+r_0\frac{1}{T}\int_0^T s(t-\tau)dt\\ &=\frac{1}{T}\int_0^T r(t)s(t-\tau)dt=Q_{ss}(-\tau),\end{aligned}$$

where the second step follows from Assumption 2.3. Thus, Equation 2.9 can be re-expressed in the form of Equation 2.4. □

Remark 2.7. The method we are describing is a kind of reverse-correlation technique because the firing rate-stimulus correlation function is evaluated at $-\tau$ in equation2.4.

Definition 2.8. The *white-noise kernel* is the optimal kernel with a white-noise stimulus that satisfies $Q_{ss}(\tau)=\sigma_s^2\delta(\tau)$.

Proposition 2.9. The *white-noise kernel* satisfies

$$D(\tau)=\frac{\langle r\rangle C(\tau)}{\sigma_s^2}, \tag{2.10}$$

where $C(\tau)$ is the spike-triggered average stimulus and $\langle r\rangle$ is the average firing rate of the neuron.

Proof. The left side of Equation 2.4 is

$$\sigma_s^2\int_0^\infty\delta(\tau-\tau')D(\tau')d\tau'=\sigma_s^2 D(\tau). \tag{2.11}$$

Thus, we have

$$D(\tau)=\frac{Q_{rs}(-\tau)}{\sigma_s^2}=\frac{\langle r\rangle C(\tau)}{\sigma_s^2}, \tag{2.12}$$

where the second step follows from the relation $Q_{rs}(-\tau)=\langle r\rangle C(\tau)$ from chapter1. □

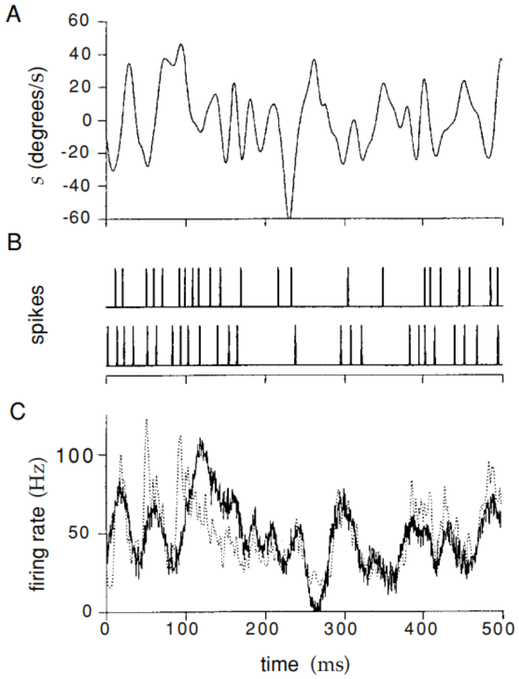
Proposition 2.10. The general solution of Equation 2.4 for an arbitrary stimulus is

$$D(\tau)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\frac{\tilde{Q}_{rs}(-\omega)}{\tilde{Q}_{ss}(\omega)}\exp(-i\omega\tau)\,d\omega, \tag{2.13}$$

where $\tilde{Q}_{rs}(\omega)$ and $\tilde{Q}_{ss}(\omega)$ are the Fourier transforms of $Q_{rs}(\omega)$ and $Q_{ss}(\omega)$, respectively.

Proof. The result could be obtained by the method of Fourier transforms. □

Example 2.11. The H1 neuron of the fly visual system responds to moving images. The following figure shows a prediction of the firing rate of this neuron obtained from a linear filter. The velocity of the moving image is plotted in A, and two typical responses are shown in B. The linear rate estimate with optimal kernel (the solid line) and the firing rate computed from the data by binning and counting spikes (the dashed line) are compared in figure C.



Definition 2.12. Neuronal selectivity is often characterized by describing stimuli that evoke maximal responses, subject to a constraint. This stimulus is called the *most effective stimulus.*

Remark 2.8. A constraint is essential because the linear estimate in Equation 2.2 is unbounded.

Definition 2.13. The *fixed energy constraint* is

$$\int_0^T \left(s(t')\right)^2 dt' = constant, \tag{2.14}$$

where the integral $\int_0^T \left(s(t')\right)^2 dt'$ is called *stimulus energy.*

Proposition 2.14. With the optimal kernel $D(\tau)$ and the fixed energy constraint 2.14, the most effective stimulus $s(t)$ is proportional to the optimal kernel $D(\tau)$ with

$$D(\tau) = -2\lambda s(t-\tau), \tag{2.15}$$

where $\lambda < 0$.

Proof. We impose this constraint by the method of Lagrange multipliers, which means that we must find the unconstrained maximum value with respect to s of

$$\begin{aligned} r_{\text{est}}(t) + \lambda \int_0^T s^2(t')dt' = r_0 + \int_0^\infty D(\tau)s(t-\tau)d\tau \\ + \lambda \int_0^T \left(s(t')\right)^2 dt', \end{aligned} \tag{2.16}$$

where λ is the Lagrange multiplier. Setting the derivative of this expression with respect to the function s to 0 (similar with the derivative of E in the solution to the proposition 2.7) gives 2.15. □

Remark 2.9. The value of λ (which is less than 0) in Equation 2.15 is determined by requiring that condition 2.14 is satisfied, but the precise value is not important for our purposes. The essential result is the proportionality between the optimal stimulus and $D(\tau)$.

Remark 2.10. The most effective stimulus analysis provides an intuitive interpretation of the linear rate estimate 2.2. At fixed stimulus energy, the integral in 2.2 measures the overlap between the actual stimulus and the most effective stimulus. In other words, it indicates how well the actual stimulus matches the most effective stimulus. Mismatches between these two reduce the value of the integral and result in lower predictions for the firing rate.

Remark 2.11. As the Example 2.11 shows, the linear rate estimate is a good agreement in regions where the measured rate varies slowly, but the estimate fails to capture high-frequency fluctuations of the firing rate, presumably because of nonlinear effects not captured by the linear kernel. Some such effects can be described by a static nonlinear function or including higher-order terms in a Volterra or Wiener expansion, as discussed below.

2.1.2 Volterra and Wiener Expansion

Definition 2.15. The *Volterra expansion* is the functional equivalent of the Taylor series expansion used to generate power series approximations of functions. For the case we are considering, it takes the form

$$\begin{aligned} r_{\text{est}}(t) =& r_0 + \int D(\tau)s(t-\tau)d\tau \\ &+ \iint D_2(\tau_1,\tau_2)s(t-\tau_1)s(t-\tau_2)d\tau_1 d\tau_2 \\ &+ \iiint D_3(\tau_1,\tau_2,\tau_3)s(t-\tau_1)s(t-\tau_2)s(t-\tau 3)d\tau_1 d\tau_2 d\tau_3 \\ &+ \cdots. \end{aligned} \tag{2.17}$$

Definition 2.16. The series rearranged by Wiener from Equation 2.17 to make the terms easier to compute has the same first two terms of the Volterra expansion, and it is called *Wiener expansion.* And the linear kernel D is called the *first Wiener kernel.*

2.1.3 Static Nonlinearities

Remark 2.12. The linear prediction has two obvious problems:

(i) there is nothing to prevent the predicted firing rate from becoming negative,

(ii) the predicted rate does not saturate, but instead increases without bound as the magnitude of the stimulus increases.

One way to deal with these and some of the other deficiencies of a linear prediction is to write the firing rate as a background rate plus a nonlinear function of the linearly filtered stimulus.

Definition 2.17. The *estimate with static nonlinearity* is

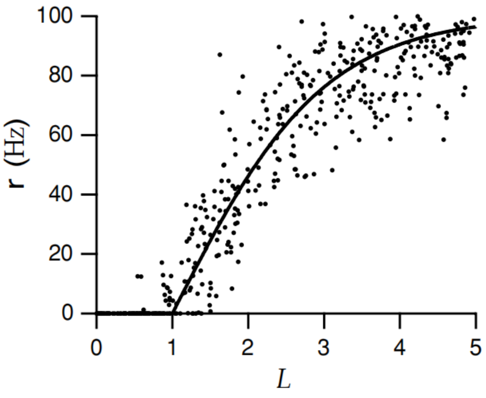
$$r_{\text{est}}(t) = r_0 + F(L(t)), \tag{2.18}$$

where F is an arbitrary function and

$$L(t) = \int_0^\infty D(\tau)s(t-\tau)d\tau. \tag{2.19}$$

F is called a *static nonlinearity* to stress that it is a function of the linear filter value evaluated instantaneously at the time of the rate estimation.

Example 2.18. F can be extracted from data by means of the graphical procedure illustrated in the following figure. First, a linear estimate of the firing rate is computed using the optimal kernel defined by Equation 2.4. Next, a plot is made of the pairs of points $(L(t), r(t))$ at various times and for various stimuli, where $r(t)$ the actual rate extracted from the data. There will be a certain amount of scatter in this plot due to the inaccuracy of the estimation. F can be extracted by fitting a function to the points on the scatter plot.



Remark 2.13. The function F typically contains constants used to set the firing rate to realistic values. These give us the freedom to normalize $D(\tau)$ in some convenient way, correcting for the arbitrary normalization by adjusting the parameters within F.

Example 2.19. The *threshold function*

$$F(L) = G[L - L_0]_+, \tag{2.20}$$

is a static nonlinearity used to introduce firing thresholds into estimates of neural responses. Here L_0 is the threshold value that L must attain before firing begins.

Remark 2.14. Above the threshold, the firing rate is a linear function of L, with G acting as the constant of proportionality. Half-wave rectification is a special case of this with $L_0 = 0$. That this function does not saturate is not a problem if large stimulus values are avoided.

Example 2.20. The *sigmoidal function*

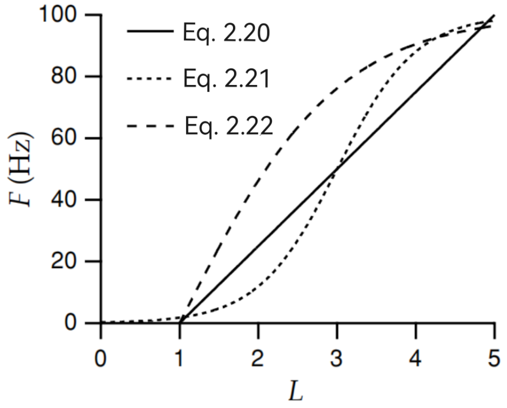
$$F(L) = \frac{r_{\max}}{1 + \exp\big(g_1(L_{1/2} - L)\big)}, \tag{2.21}$$

is a static nonlinearity used to introduce saturation into estimates of neural responses. Here $r_{\max}$ is the maximum possible firing rate, $L_{1/2}$ is the value of L for which F achieves half of this maximal value, and g_1 determines how rapidly the firing rate increases as a function of L.

Example 2.21.

$$F(L) = r_{\max}[\tanh(g_2(L - L_0))]_+ \tag{2.22}$$

is a static nonlinearity that combines a hard threshold with saturation uses a rectified hyperbolic tangent function. Here $r_{\max}$ and g_2 play similar roles as in Equation 2.21, and L_0 is the threshold.



Remark 2.15. Although the static nonlinearity can be any function, the estimate of Equation 2.18 is still restrictive because it allows for no dependence on weighted autocorrelations of the stimulus or other higher-order terms in the Volterra series.

Remark 2.16. Once the static nonlinearity is introduced, the linear kernel derived from Equation 2.4 is no longer optimal because it was chosen to minimize the squared error of the linear estimate $r_{\text{est}}(t) = r_0 + L(t)$, not the estimate with the static nonlinearity $r_{\text{est}}(t) = r_0 + F(L(t))$.

Definition 2.22. The *self-consistency condition* is that when the nonlinear estimate $r_{\text{est}}(t) = r_0 + F(L(t))$ is substituted into Equation 2.12, the relationship between the linear kernel and the firing rate-stimulus correlation function should still hold. In other words, we require that

$$D(\tau) = \frac{1}{\sigma_s^2 T}\int_0^T r_{\text{est}}(t)s(\tau - t)dt = \frac{1}{\sigma_s^2 T}\int_0^T F(L(t))s(\tau - t)dt, \tag{2.23}$$

where the second step follows from Assumption 2.3.

Theorem 2.23 (Bussgang Theorem)**.** An estimate based on the optimal kernel for linear estimation can still be self-consistent (although not necessarily optimal) when nonlinearities are present, if the stimulus used to extract the optimal kernel is Gaussian white noise.

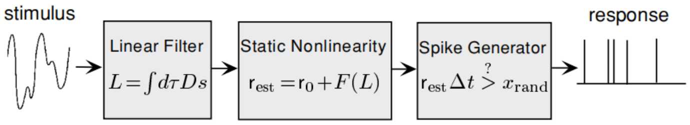
Proof. If stimulus used to extract D is Gaussian white noise, we have

$$\frac{1}{\sigma_s^2 T}\int_0^T F(L(t))s(\tau - t)dt = \frac{D(\tau)}{T}\int_0^T \frac{dF(L(t))}{dL}dt. \quad (2.24)$$

For the right side of this equation to be $D(\tau)$, the remaining expression must be equal to 1 by appropriate scaling of F. The critical identity 2.24 is based on integration by parts for a Gaussian weighted integral. □

Remark 2.17. The Bussgang Theorem suggests that Equation 2.12 will provide a reasonable kernel, even in the presence of a static nonlinearity, if the white noise stimulus used is Gaussian.

Example 2.24. A model of the spike trains evoked by a stimulus can be constructed by using the firing-rate estimate of Equation 2.18 to drive a Poisson spike generator (see chapter 1). The following figure shows the structure of such a model with a linear filter, a static nonlinearity, and a stochastic spike generator.


stimulus
Linear Filter
$L = \int d\tau D s$
Static Nonlinearity
$r_{\text{est}} = r_0 + F(L)$
Spike Generator
$r_{\text{est}}\Delta t \overset{?}{>} x_{\text{rand}}$
response

Remark 2.18. In some cases, the linear term fails to predict even when static nonlinearities are included and in practice including more terms in the Volterra series is quite difficult to go beyond the first few terms. We can replace the parameter s in Equation 2.19 with an appropriately chosen function of s to improve the accuracy, that is,

$$L(t) = \int_0^\infty D(\tau)f(s(t-\tau))d\tau. \quad (2.25)$$

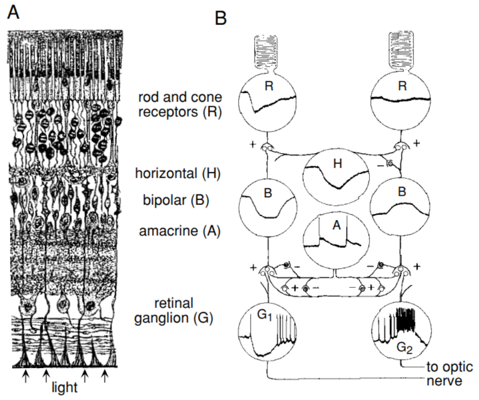
A reasonable choice for this function is the response tuning curve. For time-dependent stimuli, we can think of Equation 2.25 as a dynamic extension of the response tuning curve.

2.2 The Early Visual System

Principle 2.25 (Retinal Signal Conversion)**.** The conversion of a light stimulus into an electrical signal and ultimately an action potential sequence occurs in the retina. The retina is roughly composed of 3 layers of cells, *photoreceptor cells*, *bipolar cells* and *ganglion cells.* First, photoreceptor cells convert light signals into electrical signals. And then, bipolar cells are responsible for sorting and processing these electrical signals. Finally, ganglion cells will convert electrical signals into action potential sequences. In the intact eye, counterintuitively, light enters through the side opposite from the photoreceptors because Vertebrate retinal cell layers are arranged in reverse order of signaling.

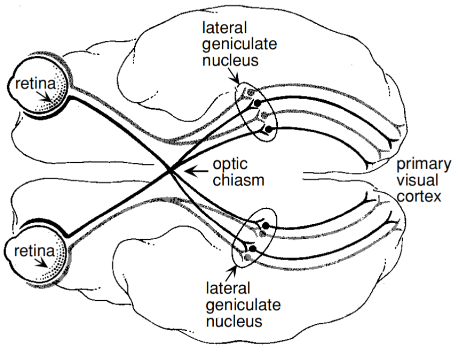
Remark 2.19. Changing membrane potentials is adequate for signaling within the retina, where distances are small. However, it is inadequate for the task of conveying information from the retina to the brain. Thus, the ganglion cells are needed.

Example 2.26. The following figure A is an anatomical diagram showing the five principal cell types of the retina and figure B is a rough circuit diagram and intracellular recordings made in neurons of the retina of a mud puppy (an amphibian). The rod cells, especially the one on the left side of figure B, are hyperpolarized by the light flash. This electrical signal is passed along to bipolar and horizontal cells through synaptic connections. Note that in one of the bipolar cells, the signal has been inverted, leading to depolarization. Pluses and minuses represent excitatory and inhibitory synapses, respectively. The two retinal ganglion cells shown in the figure have different responses and transmit different sequences of action potentials. G_2 fires while the light is on, and G_1 fires when it turns off. These are called *ON* and *OFF responses*, respectively



Definition 2.27. The output neurons of the retina are the retinal ganglion cells, whose axons form the *optic nerve.*

Principle 2.28 (Visual Pathway)**.** As the following figure shows, the optic nerve carry information from each visual hemifield up to the *optic chiasm*, where some retinal ganglion cell axons cross the midline at the optic chiasm, and then to the LGN. Cells in this nucleus send their axons along the optic radiation to the primary visual cortex.



Definition 2.29. The restricted regions of the visual field where light stimuli could active Neurons in the retina, LGN, and primary visual cortex is called *receptive fields of* the corresponding *visual neuron.*

Assumption 2.30. Patterns of illumination outside the receptive field of a given neuron cannot generate a response directly, although they can significantly affect responses to stimuli within the receptive field. We do not consider such effects, although they are of considerable experimental and theoretical interest.

Remark 2.20. Within the receptive fields, there are regions where illumination higher than the background light intensity enhances firing, and other regions where lower illumination enhances firing. The spatial arrangement of these regions determines the selectivity of the neuron to different inputs. The term *receptive field* is often generalized to refer not only to the overall region where light affects neuronal firing, but also to the spatial and temporal structure within this region.

Definition 2.31. Visually responsive neurons in the retina, LGN, and primary visual cortex are divided into two classes, depending on whether or not the contributions from different locations within the visual field sum linearly. *Simple cells* in primary visual cortex appear to satisfy this assumption. *Complex cells* in primary visual cortex do not show linear summation across the spatial receptive field, and nonlinearities must be included in descriptions of their responses.

Assumption 2.32. To streamline the discussion in this chapter, we consider only gray-scale images, although the methods presented can be extended to include color. We also restrict the discussion to two-dimensional visual images, ignoring how visual responses depend on viewing distance and encode depth.

Remark 2.21. In discussing the response properties of retinal, LGN, and V1 neurons, we do not follow the path of the visual signal, nor the historical order of experimentation, but instead begin with primary visual cortex and then move back to the LGN and retina. And the emphasis of this chapter is on properties of individual neurons.

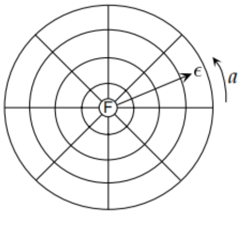
2.2.1 The Retinotopic Map

Definition 2.33. The *retinotopic map* is a map from the visual world to the cortical surface that make sure neighboring points in a visual image evoke activity in neighboring regions of visual cortex.

Remark 2.22. A striking feature of most visual areas in the brain, including primary visual cortex, is that the visual world is mapped onto the cortical surface in this topographic manner. The retinotopic map refers to the transformation from the coordinates of the visual world to the corresponding locations on the cortical surface.

Definition 2.34. The image point that focuses onto the fovea or center of the retina is called the *fixation point.*

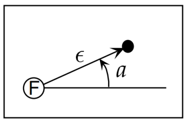
Definition 2.35. Locations on a sphere can be represented using the same longitude and latitude angles used for the surface of the earth, which called *spherical coordinate system.*



The north pole is located at the fixation point, the latitude coordinate is called the *eccentricity* ϵ, and the longitude coordinate, measured from the horizontal meridian, is called the *azimuth* a.

Principle 2.36. In primary visual cortex, the visual world is split in half, with the region $-90^\circ \leq a \leq 90^\circ$ for ϵ from 0° to about 70° (for both eyes) represented on the left side of the brain, and the reflection of this region about the vertical meridian represented on the right side of the brain.

Definition 2.37. Polar coordinate system used to parameterize image location is shown in the following figure.



The rectangle represents a *tangent screen*, the filled circle is the location of a particular image point on the screen, the origin of the polar coordinate system is the fixation point F, the *eccentricity* ϵ is proportional to the radial distance from the fixation point to the image point, and a is the angle between the radial line from F to the image point and the horizontal axis.

Remark 2.23. In most experiments, images are displayed on a tangent screen that does not coincide exactly with the sphere discussed in the previous paragraph. However, if the tangent screen is not too large, the difference is negligible, and the eccentricity and azimuth angles approximately coincide with polar coordinates on the screen.

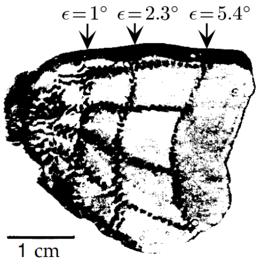
Assumption 2.38. The eccentricity ϵ and the x and y coordinates of the Cartesian system that are based on measuring distances on the screen are converted to degrees by

$$\frac{l}{r} \times \frac{180^\circ}{\pi}, \tag{2.26}$$

where l is the distance on the screen, r is the distance from the eye to the screen.

Remark 2.24. Assumption 2.38 makes sense because it is the angular, not the absolute size and location of an image that is typically relevant for studies of the visual system. And Equation 2.26 is similar to the arc length-radian relationship.

Example 2.39. The following figure shows An autoradiograph of the posterior region of the primary visual cortex from the left side of a macaque monkey brain. The pattern is a radioactive trace of the activity evoked by an image like that in Definition 2.35 figure. The vertical lines correspond to circles at eccentricities of 1°, 2.3°, and 5.4°, and the horizontal lines (from top to bottom) represent radial lines in the visual image at a values of −90°, −45°, 0°, 45°, and 90°. Only the part of cortex corresponding to the central region of the visual field on one side is shown.



Remark 2.25. To construct the retinotopic map, we assume that eccentricity is mapped onto the horizontal coordinate X of the cortical sheet, and a is mapped onto its Y coordinate.

Definition 2.40. The *cortical magnification factor* determines the distance across a flattened sheet of cortex separating the activity evoked by two nearby image points.

Assumption 2.41. We assume the cortical magnification factor is isotropic, denoted by $M(\epsilon)$.

Example 2.42. Suppose that there are two image points in question (ϵ, a) and $(\epsilon + \Delta\epsilon, a)$, the angular distance between these two points is $\Delta\epsilon$, and the distance separating the activity evoked by these two image points on the cortex is ΔX. By the definition of $M(\epsilon)$, these two quantities satisfy $\Delta X = M(\epsilon)\Delta\epsilon$ or, taking the limit as ΔX and $\Delta\epsilon$ go to 0,

$$\frac{dX}{d\epsilon} = M(\epsilon). \tag{2.27}$$

Suppose that there are the other two image points in question (ϵ, a) and $(\epsilon, a+\Delta a)$, the angular distance between these two points is

$$\Delta a \times \frac{\epsilon\pi}{180^\circ},$$

where ϵ corrects for the increase of arc length as a function of eccentricity, and $\frac{\pi}{180^\circ}$converts from degrees to radians. The separation on the cortex ΔY corresponding to these points satisfies $\Delta Y = \Delta a \frac{\epsilon\pi}{180^\circ} M(\epsilon)$. Taking the limit $\Delta a \to 0$,

$$\frac{dY}{da} = -\frac{\epsilon\pi}{180^\circ} M(\epsilon). \tag{2.28}$$

The minus sign in this relationship appears because the visual field is inverted on the cortex.

Example 2.43. The cortical magnification factor for the macaque monkey, obtained from results such as the figure in Example 2.39, is approximately

$$M(\epsilon) = \frac{\lambda}{\epsilon_0 + \epsilon}, \tag{2.29}$$

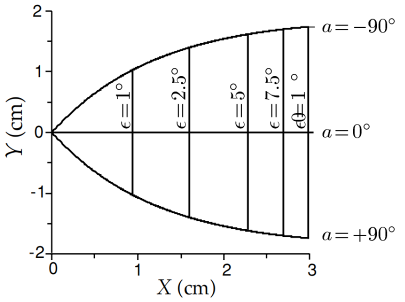
with $\lambda \approx 12$ mm and $\epsilon_0 \approx 1^\circ$. Integrating Equation 2.27 and defining $X = 0$ to be the point representing $\epsilon = 0$, we find

$$X = \lambda \ln(1 + \frac{\epsilon}{\epsilon_0}). \tag{2.30}$$

Similarly,

$$Y = -\frac{\lambda\epsilon a\pi}{(\epsilon_0 + \epsilon)180^\circ}. \tag{2.31}$$

The following figure shows that these coordinates agree fairly well with the map in Example 2.39.



Example 2.44. For $\epsilon \gg 1^\circ$, equations 2.30 and 2.31 reduce to

$$X \approx \lambda \ln(\frac{\epsilon}{\epsilon_0}), Y \approx -\frac{\lambda\pi a}{180^\circ}.$$

These two formulas can be combined by defining the complex numbers $Z = X + iY$ and $z = \frac{\epsilon}{\epsilon_0}\exp(-i\pi a/180^\circ)$, and writing

$$Z = \lambda \ln(z).$$

For this reason, the cortical map is sometimes called a *complex logarithmic map.*

For an image scaled radially by a factor γ, eccentricities change according to $\epsilon \to \gamma\epsilon$ while a is unaffected. Scaling of the eccentricity produces a shift

$$X \to X + \lambda \ln(\gamma)$$

over the range of values where the simple logarithmic form of the map is valid. The logarithmic transformation thus causes images that are scaled radially outward on the retina to be represented at locations on the cortex translated in the X direction.

Remark 2.26. For smaller ϵ, the map we have derived is only approximate even in the complete form given by equations 2.30 and 2.31. This is because the cortical magnification factor is not really isotropic, as we have assumed in this derivation, and a complete description requires accounting for the curvature of the cortical surface.

2.2.2 Visual Stimuli

Remark 2.27. Pixel locations are parameterized by Cartesian coordinates x and y. However, pixel-by-pixel light intensities are not a useful way of parameterizing a visual image for the purposes of characterizing neuronal responses. This is because visually responsive neurons, like many sensory neurons, adapt to the overall level of screen illumination.

Definition 2.45. we describe the *visual stimulus* by a function $s(x, y, t)$ that is proportional to the difference between the luminance at the point (x, y) at time t and the average or background level of luminance.

Remark 2.28. Definition 2.45 could avoid dealing with adaptation effects.

Definition 2.46. The *contrast* is the resulting quantity that $s(x, y, t)$ divided by the background luminance level, making it dimensionless.

Definition 2.47. A commonly used stimulus, the *counterphase sinusoidal grating*, is described by

$$s(x, y, t) = A\cos(Kx\cos\Theta + Ky\sin\Theta - \Phi)\cos(\omega t), \quad (2.32)$$

where K and ω are the *spatial* and *temporal frequencies* of the grating (these are angular frequencies), Θ is its *orientation*, Φ is its *spatial phase*, and A is its *contrast amplitude*.

Example 2.48. The following figure shows a similar grating (a spatial square wave is drawn rather than a sinusoid) and illustrates the significance of the parameters in Definition 2.47. The lighter stripes are regions where $s > 0$, and $s < 0$ within the darker stripes. This stimulus oscillates in both space and time:

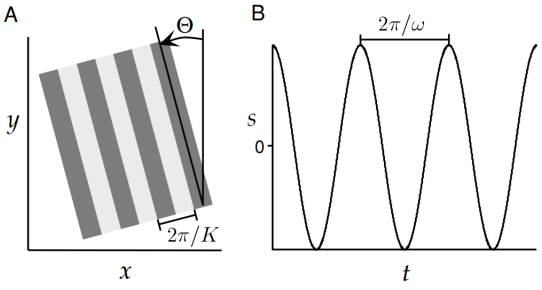
(i) At any fixed time, it oscillates in the direction perpendicular to the orientation angle Θ as a function of position, with wavelength $\frac{2\pi}{K}$ (figure A). A stimulus with $\Theta = 0$ varies in the x direction.

(ii) At any fixed position, it oscillates in time with period $\frac{2\pi}{\omega}$ (figure B).

Changing Φ by an amount $\Delta\Phi$ shifts the grating in the direction perpendicular to its orientation by a fraction $\frac{\Delta\Phi}{2\pi}$ of its wavelength, that is, $\frac{\Delta\Phi}{K}$, derived from

$$\begin{aligned} &Kx\cos\Theta + Ky\sin\Theta - (\Phi + \Delta\Phi)\\ &= Kx\cos\Theta + Ky\sin\Theta - \Delta\Phi(\sin\Theta^2 + \cos\Theta^2) - \Phi\\ &= K(x - \frac{\Delta\Phi}{K}\cos\Theta)\cos\Theta + K(y - \frac{\Delta\Phi}{K}\sin\Theta)\sin\Theta - \Phi. \end{aligned}$$

The contrast amplitude A controls the maximum degree of difference between light and dark areas.



Exercise 2.49. Prove that units of parameters in Definition 2.47 are as follows:

parameter	unit
K	radians per degree
$\frac{K}{2\pi}$	cycles per degree
Φ	radians
ω	radians/s(second)
$\frac{\omega}{2\pi}$	HZ

Remark 2.29. Experiments that consider reverse correlation and spike-triggered averages use various types of random and white-noise stimuli in addition to bars and gratings.

Definition 2.50. A *white-noise image* is one visual stimulus that is uncorrelated in both space and time so that

$$\frac{1}{T}\int_0^T s(x, y, t)s(x', y', t + \tau)dt = \sigma_s^2\delta(\tau)\delta(x - x')\delta(y - y'). \quad (2.33)$$

Remark 2.30. In practice a discrete approximation of such a stimulus must be used by dividing the image space into pixels and time into small bins. In addition, more structured random sets of images (randomly oriented bars, for example) are sometimes used to enhance the responses obtained during stimulation.

2.2.3 The Nyquist Frequency

Remark 2.31. Many factors limit the maximal spatial frequency that can be resolved by the visual system, one interesting effect arises from the size and spacing of individual photoreceptors on the retina. The region of the retina with the highest resolution is the fovea at the center of the visual field. Within the macaque or human fovea, cone photoreceptors are densely packed in a regular array.

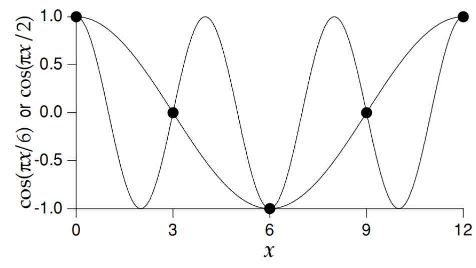
Definition 2.51. Along any direction in the visual field, a regular array of tightly packed photoreceptors of size Δx samples points at locations $m\Delta x$ for $m = 1, 2, \ldots$. The (angular) frequency that defines the *resolution* of such an array is called the *Nyquist frequency* and is given by

$$K_{\text{nyq}} = \frac{\pi}{\Delta x}. \quad (2.34)$$

Example 2.52. Consider sampling two cosine gratings with spatial frequencies of K and $2K_{\text{nyq}} - K$, with $K < K_{\text{nyq}}$. These are described by $s = \cos(Kx)$ and $s = \cos((2K_{\text{nyq}} - K)x)$. At the sampled points $m\Delta x$, these functions are identical because

$$\cos((2K_{\text{nyq}} - K)m\Delta x) = \cos(2\pi m - Km\Delta x) = \cos(Km\Delta x),$$

which follows from the periodicity and evenness of the cosine function (see the following figure).



As a result, these two gratings cannot be distinguished by examining them only at the sampled points.

Remark 2.32 (The importance of Nyquist frequency)**.** As discussed in Example 2.52, any two spatial frequencies $K < K_{\text{nyq}}$ and $2K_{\text{nyq}} - K$ can be confused with one another in this way, a phenomenon known as aliasing. Conversely, if an image is constructed solely of frequencies less than K_{nyq}, it can be reconstructed perfectly from the finite set of samples provided by the array. (Note that, images with smaller K will be easier to distinguish because of their bigger wavelengths.)

Example 2.53. There are 120 cones per degree at the fovea of the macaque retina, which makes $\Delta x = 1/120$ and

$$\frac{K_{\text{nyq}}}{2\pi} = \frac{1}{2\Delta x} = 60 \text{ cycles per degree.}$$

In this result, we have divided the right side of Equation 2.34, which gives K_{nyq} in units of radians per degree, by 2π to convert the answer to cycles per degree.

2.3 Reverse-Correlation Methods: Simple Cells

Definition 2.54. Given the light intensity of a visual image $s(s, y, t)$ and a spike sequence $\{t_i\}_{i=1}^n$, the *spike-triggered average stimulus* is a function of three variables

$$C(x, y, \tau) = \frac{1}{\langle n \rangle} \left\langle \sum_{i=1}^{n} s(s, y, t_i - \tau) \right\rangle, \qquad (2.35)$$

where the brackets denote trial averaging, and we have used the approximation $1/n \approx 1/\langle n \rangle$.

Definition 2.55. The *correlation function* between the firing rate at time t and the stimulus at time $t + \tau$, for trials of duration T is defined as

$$Q_{rs}(x, y, \tau) = \frac{1}{T} \int_0^T r(t) s(x, y, t + \tau) dt. \qquad (2.36)$$

Proposition 2.56. The spike-triggered average is related to the reverse-correlation function by

$$C(x, y, \tau) = \frac{Q_{rs}(x, y, -\tau)}{\langle r \rangle}. \qquad (2.37)$$

where $\langle r \rangle = \langle n \rangle / T$ is as usual, the average firing rate over the entire trial.

Proof. The proof is similar with the one in Chapter 1. □

Remark 2.33. To estimate the firing rate of a neuron in response to a particular image, we add a function of the output of a linear filter of the stimulus to the background firing rate r_0, as in Equation 2.18, $r_{\text{est}}(t) = r_0 + F(L(t))$. Because visual stimuli depend on spatial location, we must decide how contributions from different image locations are to be combined to determine $L(t)$. Note that, firing rates are not a function of x and y.

Definition 2.57. Suppose that the contributions from linear response estimate different spatial points sum linearly, the *linear response estimate* $L(t)$ is obtained by integrating over all x and y values:

$$L(t) = \int_0^\infty \iint D(s, y, \tau) s(x, y, t - \tau) dx dy d\tau, \qquad (2.38)$$

where the kernel $D(x, y, \tau)$ determines how strongly, and with what sign, the visual stimulus at the point (x, y) and at time $t - \tau$ affects the firing rate of the neuron at time t.

Proposition 2.58. The optimal kernel is given in terms of the firing rate-stimulus correlation function, or the spike-triggered average, for a white-noise stimulus with variance parameter σ_s^2 by

$$D(x, y, \tau) = \frac{Q_{rs}(x, y, -\tau)}{\sigma_s^2} = \frac{\langle r \rangle C(x, y, \tau)}{\sigma_s^2}. \qquad (2.39)$$

Proof. The proof is Similar with the one in Chapter 1. □

Definition 2.59. The kernel $D(x, y, \tau)$ defines the *space-time receptive field* of a neuron.

Remark 2.34. Because $D(x, y, \tau)$ is a function of three variables, it can be difficult to measure and visualize.

Definition 2.60. If the spatial structure of one neuron's receptive field does not change over time except by an overall multiplicative factor, the kernel can be written as a product of two functions, one that describes the *spatial receptive field* and the other, the *temporal receptive field*,

$$D(x, y, \tau) = D_s(x, y) D_t(\tau). \qquad (2.40)$$

Such neurons are said to have *separable space-time receptive field.*

Remark 2.35. When $D(x, y, \tau)$ cannot be written as the product of two terms, the neuron is said to have a *nonseparable space-time receptive field.*

Assumption 2.61 (Normalization)**.** Given the freedom in Equation 2.18 to set the scale of D (by suitably adjusting the function F), we typically normalize D_s so that its integral is 1, and use a similar rule for the components from which D_t is constructed, that is,

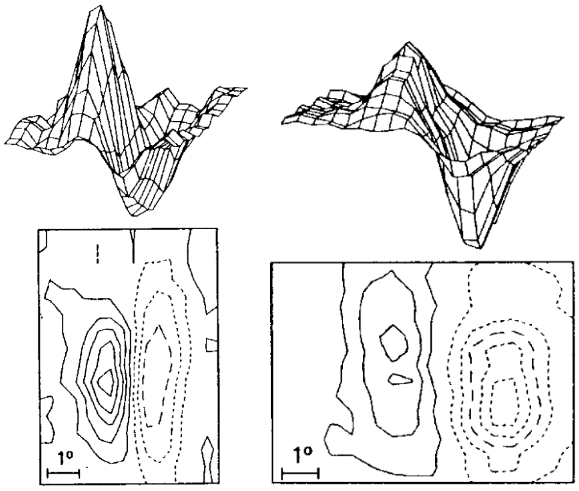
$$\iint D_s(x, y) dx dy = 1 \text{ and } \int_0^\infty \mathrm{D_t}(\tau) \mathrm{d}\tau = 1.$$

Remark 2.36. We begin our analysis by studying first the spatial and then the temporal components of a separable space-time receptive field, and then proceed to the nonseparable case. For simplicity, we ignore the possibility that cells can have slightly different receptive fields for the two eyes, which underlies the disparity tuning considered in chapter 1.

2.3.1 Spatial Receptive Fields

Definition 2.62. Regions within the receptive field where D_s is positive, is called *ON regions*, and where it is negative, is called *OFF regions.*

Example 2.63. The following figures show the spatial structure of the receptive fields of two neurons in cat primary visual cortex determined by averaging stimuli between 50 ms and 100 ms prior to an action potential, that is, $\int_{50}^{100} C(x, y, \tau) d\tau / 50$.



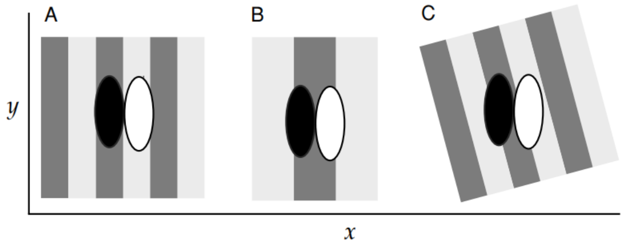
The upper plots are three-dimensional representations, with the horizontal dimensions acting as the x-y plane and the vertical dimension indicating the magnitude and sign of $D_s(x, y)$. The lower contour plots represent the x-y plane. Regions with solid contour curves are ON areas where $D_s(x, y) > 0$, and regions with dashed contours are OFF areas where $D_s(x, y) < 0$.

Proposition 2.64. The response of a neuron is enhanced if ON regions are illuminated ($s > 0$) or if OFF regions are darkened ($s < 0$) relative to the background level of illumination. Conversely, they are suppressed by darkening ON regions or illuminating OFF regions.

Proof. The integral of the linear kernel times the stimulus can be visualized by noting how the OFF and ON regions overlap the image. □

Remark 2.37. From Proposition 2.64, the neurons of figures in Example 2.63 respond most vigorously to light-dark edges positioned along the border between the ON and OFF regions, and oriented parallel to this border and to the elongated direction of the receptive fields, shown below.

Example 2.65. Grating stimuli superimposed on spatial receptive fields similar to those shown in Example 2.63. The receptive field is shown as two oval regions, one dark to represent an OFF area where $D_s < 0$ and one white to denote an ON region where $D_s > 0$.



Remark 2.38. The above examples show receptive fields with two major subregions. Simple cells are found with from one to five subregions. Along with the ON-OFF patterns we have seen, another typical arrangement is a three-lobed receptive field with OFF-ON-OFF or ON-OFF-ON subregions.

Definition 2.66. A *Gabor function* is a product of a Gaussian function and a sinusoidal function.

Example 2.67. If the coordinates x and y are chosen such that the borders between the ON and OFF regions are parallel to the y axis and the origin of the coordinates is placed at the center of the receptive field, the Gabor function

$$D_s(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \cos(kx - \phi) \quad (2.41)$$

can be used to approximate the observed receptive field structures. The parameters in this function determine the properties of the spatial receptive field:

(i) σ_x and σ_y, the receptive field size, determine its extent in the x and y directions, respectively.

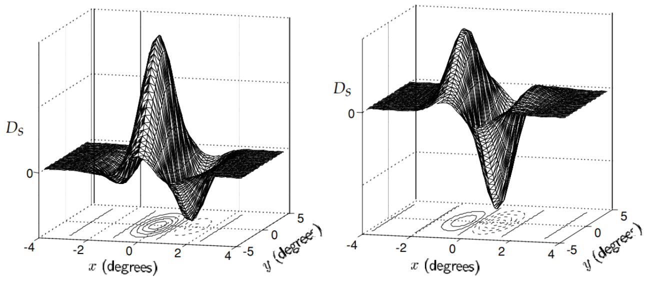
(ii) k, the preferred spatial frequency, determines the spacing of light and dark bars that produce the maximum response (the preferred spatial wavelength is $2\pi/k$).

(iii) ϕ, the preferred spatial phase, determines where the ON-OFF boundaries fall within the receptive field.

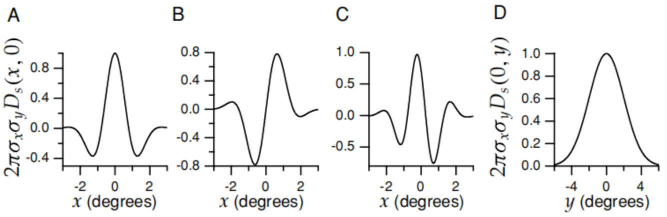
For this spatial receptive field, the sinusoidal grating described by Equation 2.32 that produces the maximum response for a fixed value of A has $K = k$, $\Phi = \phi$, and $\Theta = 0$.

Remark 2.39. The Gabor functions mentioned below refer to Equation 2.41.

Example 2.68. The Gabor functions chosen specifically to match the data in Example 2.63. These two figures are plotted with $\sigma_x = 1°$, $\sigma_y = 2°$, $1/k = 0.56°$ and $\phi = 1 - \pi/2$ (left), $\phi = 1 - \pi$ (right).



Example 2.69. x- and y- plots of a variety of Gabor functions (Equation 2.41) with different parameter values. For convenience, these plots are the dimensionless function $2\pi\sigma_x\sigma_y D_s$.



Responding parameters are as follows:

(i) A with $\sigma_x = 1°$, $1/k = 0.5°$, $\phi = 0$ and $y = 0$ is symmetric about $x = 0$.

(ii) B with $\sigma_x = 1°$, $1/k = 0.5°$, $\phi = \pi/2$ and $y = 0$ is antisymmetric about $x = 0$ and corresponds to using a sine instead of a cosine function in Equation 2.41.

(iii) C with $\sigma_x = 1°$, $1/k = 0.33°$, $\phi = \pi/4$ and $y = 0$ has no particular symmetry properties with respect to $x = 0$.

(iv) D with $\sigma_y = 2°$ and $x = 0$ is simply a Gaussian.

Remark 2.40. As seen in Example 2.69, Gabor functions can have various types of symmetry, and variable numbers of significant oscillations (or subregions) within the Gaussian envelope.

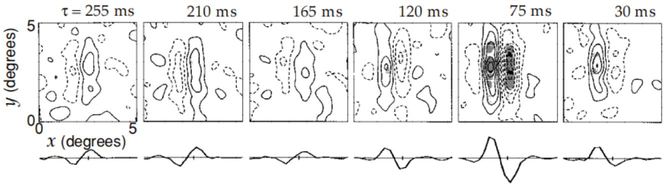
Remark 2.41. The response characterized by Equation 2.41 is maximal if light-dark edges are parallel to the y axis, so the preferred orientation orientation for a stimulus is 0. An *arbitrary preferred orientation* θ can be generated by rotating the coordinates, making the substitutions $x \to a\cos(\theta) + y\sin(\theta)$ and $y \to y\cos(\theta) - x\sin(\theta)$ in Equation 2.41. This produces a spatial receptive field that is maximally responsive to a grating with $\Theta = \theta$. Actually, this rotation is from the rotation matrix

$$M(\theta) = \begin{pmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{pmatrix}.$$

Remark 2.42. Similarly, a receptive field centered at the point (x_0, y_0) rather than at the origin can be constructed by making the substitutions $x \to x - x_0$ and $y \to y - y_0$, where the (x_0, y_0) is called the *receptive field center.*

2.3.2 Temporal Receptive Fields

Example 2.70. The following figure reveals the temporal development of the space-time receptive field of a neuron in the cat primary visual cortex through a series of snapshots of its spatial receptive field. Each panel is a plot of $D(x, y, \tau)$ for a different value of τ. The curves below the contour diagrams are one-dimensional plots of the receptive field as a function of x alone. Around $\tau = 210$ ms, a two-lobed OFF-ON receptive field is evident. As τ decreases, this structure first fades away and then reverses, so that the receptive field at $\tau = 75$ ms has the opposite sign from what appeared at $\tau = 210$ ms.



The stimulus preferred by this cell is thus an appropriately aligned dark-light boundary that reverses to a light-dark boundary over time.

Remark 2.43. In the above figure, for $\tau > 300$ ms, there is little correlation between the visual stimulus and the upcoming spike. Due to latency effects, the spatial structure of the receptive field is less significant for $\tau < 75$ ms.

Proposition 2.71. The neuron in Example 2.70 has approximately a separable space-time receptive field.

Proof. Although the magnitudes and signs of the different spatial regions vary over time, their locations and shapes remain fairly constant. This indicates that the neuron has, to a good approximation, a separable space-time receptive field. □

Definition 2.72. The phenomenon that a neuron's receptive field reverses with time is called the *Reversal effect.*

Remark 2.44. Reversal effects are a common feature of space-time receptive fields.

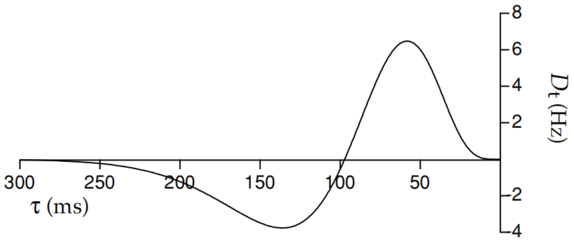
Proposition 2.73. The reversal can be described by a function

$$D_t(\tau) = \begin{cases} \alpha \exp(-\alpha\tau) \left(\frac{(\alpha\tau)^5}{5!} - \frac{(\alpha\tau)^7}{7!} \right) & \tau \geq 0, \\ 0 & \tau < 0. \end{cases} \tag{2.42}$$

Here, α is a constant that sets the scale for the temporal development of the function.

Proof. The function $D_t(\tau)$ of Equation 2.42 rises from 0, becomes positive, then negative, and ultimately returns to 0 as τ increases. □

Example 2.74. Temporal structure of a receptive field. (Plot $D_t(\tau)$ in Equation 2.42 with $\alpha = 1/(15\text{ ms})$.)



Remark 2.45. Single-phase responses are also seen for V1 neurons, and these can be described by eliminating the second term in Equation 2.42. Three-phase responses, which are sometimes seen, must be described by a more complicated function.

2.3.3 Response of a Simple Cell to a Counterphase Grating

Remark 2.46. The response of a simple cell to a counterphase grating stimulus (Equation 2.32) can be estimated by computing the function $L(t)$.

Proposition 2.75. For the separable receptive field, the linear estimate of the response can be written as the product of two terms,

$$L(t) = L_s L_t(t), \tag{2.43}$$

where

$$L_s = \iint D_s(x, y) A \cos\left(Kx\cos(\Theta) + Ky\sin(\Theta) - \Phi\right) dxdy \tag{2.44}$$

and

$$L_t(t) = \int_0^\infty D_t(\tau)\cos(\omega(t - \tau))d\tau. \tag{2.45}$$

Assumption 2.76. We assume that $D_s(x, y)$ and $D_t(\tau)$ used in this section are from Equation 2.41 and 2.42, respectively.

Exercise 2.77. Compute these integrals in equations 2.44 and 2.45 for $D_s(x,y)$ in Equation 2.41 with $\sigma_x = \sigma_y = \sigma$ and $D_t(\tau)$ in Equation 2.42.

Proposition 2.78. If the spatial phase of the stimulus and the preferred spatial phase of the receptive field are 0 ($\Phi = \phi = 0$),

$$L_s = A \exp\left(-\frac{\sigma^2(k^2+K^2)}{2}\cosh(\sigma^2 kK\cos(\Theta))\right), \quad (2.46)$$

which determines the orientation and spatial frequency tuning for an optimal spatial phase.

Lemma 2.79. For a grating with the preferred orientation $\Theta = 0$ and a spatial frequency that is not too small, the full expression for L_s can be simplified by noting that $\exp(-\sigma^2 kK) \approx 0$ for the values of $k\sigma$ noemally encountered.

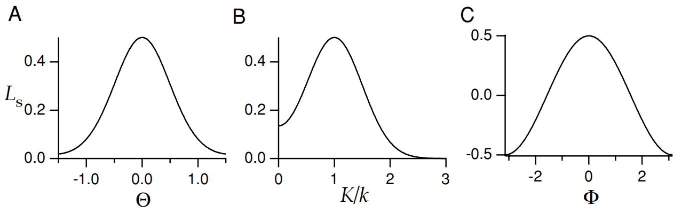
Example 2.80. If $K = k$ and $k\sigma = 2$, $\exp(-\sigma^2 kK) = 0.02$.

Proposition 2.81. Using the approximation in Lemma 2.79,

$$L_s = \frac{A}{2}\exp\left(-\frac{\sigma^2(k-K)^2}{2}\right)\cos(\phi - \Phi), \quad (2.47)$$

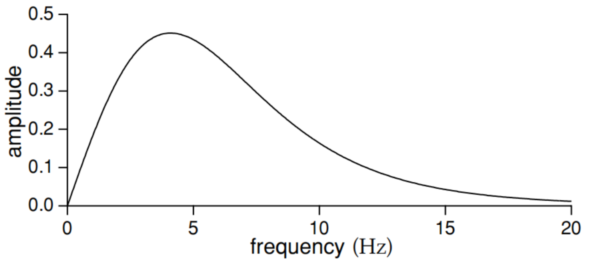
which reveals a Gaussian dependence on spatial frequency and a cosine dependence on spatial phase.

Example 2.82. Selectivity of a Gabor filter (Equation 2.44 with $D_s(x,y)$ from Equation 2.41) with $\theta = \phi = 0$, $\sigma_x = \sigma_y = \sigma$, and $k\sigma = 2$ acting on a cosine grating with $A = 1$.



(A) L_s as a function of stimulus orientation Θ for a grating with the preferred spatial frequency and phase, $K = k$ and $\Phi = 0$. (B) L_s as a function of the ratio of the stimulus spatial frequency to its preferred value, K/k, for a grating oriented in the preferred direction $\Theta = 0$ and with the preferred phase $\Phi = 0$. (C) L_s as a function of stimulus spatial phase Φ for a grating with the preferred spatial frequency and orientation, $K = k$ and $\Theta = 0$.

Example 2.83. The temporal frequency dependence of the amplitude of the linear response estimate is plotted as a function of the temporal frequency of the stimulus ($\omega/2\pi$ rather than the angular frequency ω).



The peak value around 4 Hz and roll-off above 10 Hz are typical for V1 neurons and for cortical neurons in other primary sensory areas as well.

2.3.4 Space-Time Receptive Fields

Remark 2.47. To display the function $D(x,y,\tau)$ in a space-time plot rather than as a sequence of spatial plots , we suppress the y dependence and plot an x-τ projection of the space-time kernel.

Example 2.84 (A separable space-time receptive field)**.** Figure A shows a space-time plot of the receptive field of a simple cell in the cat primary visual cortex. This receptive field is approximately separable, and it has side-by-side OFF and ON regions that reverse as a function of τ.

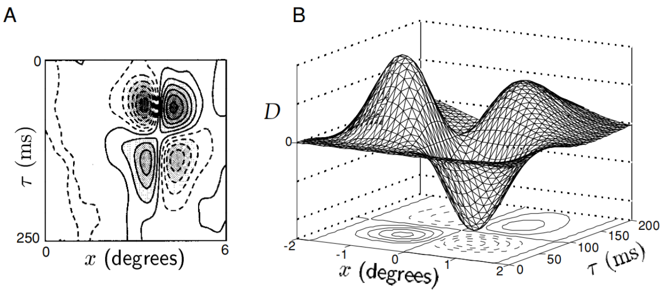
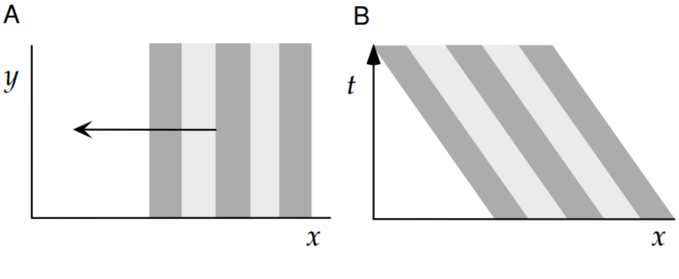


Figure B shows an x-τ plot of a separable space-time kernel, similar to the one in A, generated by multiplying a Gabor function (evaluated at $y = 0$) with $\sigma_x = 1°$, $1/k = 0.56°$ and $\phi = \pi/2$ by the temporal kernel of Equation 2.42 with $1/\alpha = 15$ ms.

Remark 2.48. We can also plot the visual stimulus in a space-time diagram, suppressing the y coordinate by assuming that the image does not vary as a function of y.

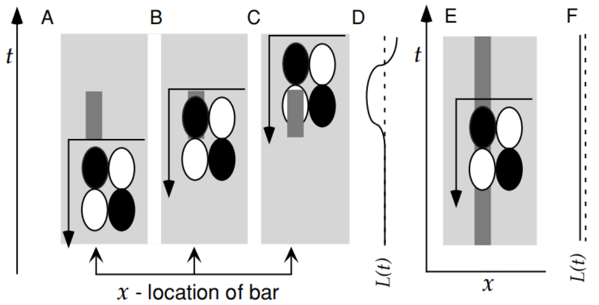
Example 2.85 (Space and space-time diagrams of a moving grating)**.** Figure A shows a grating of vertically oriented stripes moving to the left on an x-y plot. In the x-t plot of figure B, this image appears as a series of sloped dark and light bands. These represent the projection of the image in A onto the x axis evolving as a function of time. The leftward slope of the bands corresponds to the leftward movement of the image.



Principle 2.86. Most neurons in primary visual cortex do not respond strongly to static images, but respond vigorously to flashed and moving bars and gratings.

Remark 2.49. The receptive field structure of Example 2.84 reveals why this is the case in Principle 2.86.

Example 2.87 (Why a flashed bar is a effective stimulus)**.** The following figures show responses to dark bars estimated from a separable space-time receptive field.



The linear estimate of the response at any time is determined by positioning the receptive field diagram so that its horizontal axis matches the time of response estimation and noting how the OFF and ON regions overlap with the image.

(A-C) The image is a dark bar that is flashed on for a short interval of time. There is no response (A) until the dark image overlaps the OFF region (B) when $L(t) > 0$. The response is later suppressed when the dark bar overlaps the ON region (C) and $L(t) < 0$.

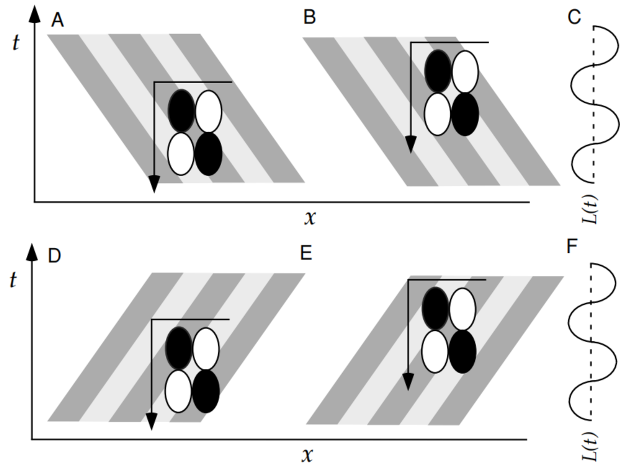
(D) A plot of $L(t)$ versus time corresponding to the responses generated in A-C. Time runs vertically in this plot, and $L(t)$ is plotted horizontally with the dashed line indicating the zero axis and positive values plotted to the left.

(E) The image is a static dark bar. The bar overlaps both an OFF (small τ) and an ON (large τ) region, generating opposing positive and negative contributions to $L(t)$.

(F) The weak response corresponding to E, plotted as in D.

The flashed dark bar of figures A-C is a more effective stimulus than the static bar of figure E.

Example 2.88 (Why a moving grating is a particularly effective stimulus)**.** The following figure shows responses to moving gratings estimated from a separable space-time receptive field. The receptive field is the same as in Example 2.87.



(A-C) The stimulus is a grating moving to the left.

- At the time corresponding to A, OFF regions overlap with dark bands and ON regions with light bands, generating a strong response.
- At the time of the estimate in B, the alignment is reversed, and $L(t)$ is negative.
- (C) is a plot of $L(t)$ versus time corresponding to the responses generated in A-B. Time runs vertically in this plot and $L(t)$ is plotted horizontally, with the dashed line indicating the zero axis and positive values plotted to the left.

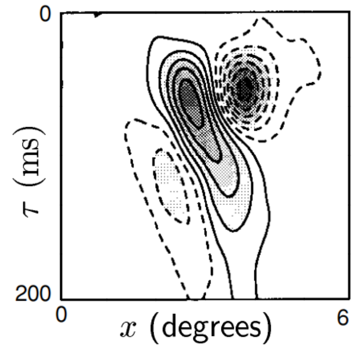
(D-F) The stimulus is a grating moving to the right. The responses are identical to those in A-C.

Remark 2.50. Separable space-time receptive fields can produce responses that are maximal for certain speeds of grating motion, but they are not sensitive to the direction of motion.

2.3.5 Nonseparable Receptive Fields

Remark 2.51. Many neurons in primary visual cortex are selective for the direction of motion of an image. Accounting for direction selectivity requires nonseparable space-time receptive fields.

Example 2.89. An example of a nonseparable receptive field is shown as below. This neuron has a three-lobed OFF-ON-OFF spatial receptive field, and these subregions shift to the left as time moves forward (and τ decreases). This means that the optimal stimulus for this neuron has light and dark areas that move toward the left.



Remark 2.52. One way to describe a nonseparable receptive field structure is to use a separable function constructed from a product of a Gabor function for D_s and Equation 2.42 for D_t, but to write these as functions of a mixture or rotation of the x and τ variables.

Lemma 2.90. The rotation matrix

$$M'(\psi) = \begin{pmatrix} \cos(\psi) & -\sin(\psi) \\ \sin(\psi) & \cos(\psi) \end{pmatrix} \tag{2.48}$$

can rotate a vector in two-dimensional space by ψ counter-clockwise.

Proposition 2.91. The rotation of the space-time receptive field is achieved by mixing the space and time coordinates, using the transformation

$$D(x, y, \tau) = D_s(x', y) D_t(\tau') \tag{2.49}$$

with

$$x' = x\cos(\psi) - c\tau\sin(\psi) \tag{2.50}$$

and

$$\tau' = \tau\sin(\psi) + \frac{x}{c}\sin(\psi), \tag{2.51}$$

where factor c converts between the units of time (ms) and space (degrees), and ψ is the space-time rotation angle.

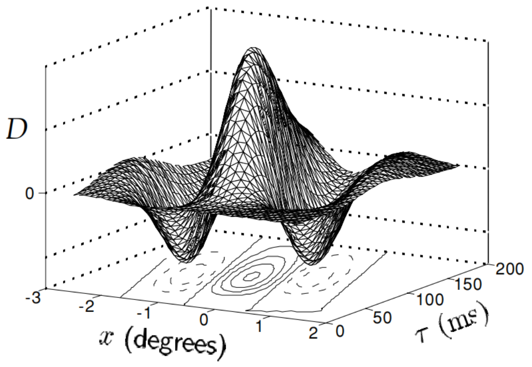
Proof. Note that the origin of x-τ plot in Example 2.89 is on the upper left corner, thus this rotation should be counter-clockwise. Equation 2.50 and 2.51 follow from Lemma 2.90 directly. □

Example 2.92. Mathematical description of the space-time receptive field in Example 2.89 constructed from equations 2.49 - 2.51. The parameters are selected as follows:

(i) The Gabor function used (evaluated at $y = 0$) had $\sigma_x = 1°$, $1/k = 0.5°$, and $\phi = 0$.

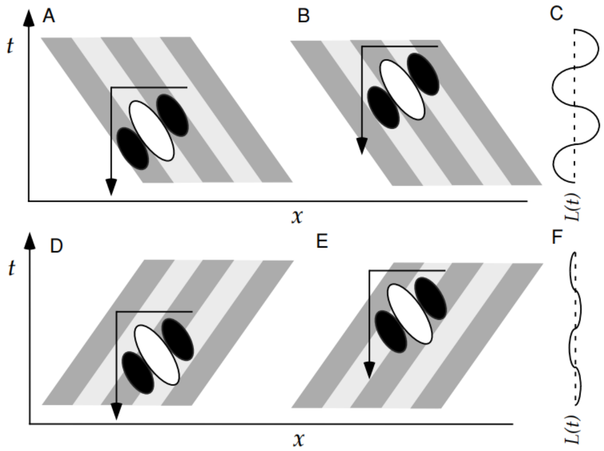
(ii) D_t is given by the expression in Equation 2.42 with $\alpha = 20$ ms, except that the second term, with the seventh power function, was omitted because the receptive field does not reverse sign in this example.

(iii) The x-τ rotation angle used was $\psi = \pi/9$, and the conversion factor was $c = 0.02°/\text{ms}$.



Remark 2.53. The rotation operation is not the only way to generate nonseparable space-time receptive fields. They are often constructed by adding together two or more separable space-time receptive fields with different spatial and temporal characteristics.

Example 2.93 (Direction Sensitivity)**.** The following figures show responses to moving gratings estimated from a nonseparable spacetime receptive field.



(A-C) The stimulus is a grating moving to the left.

- At the time corresponding to A, OFF regions overlap with dark bands and the ON region overlaps a light band, generating a strong response.
- At the time of the estimate in B, the alignment is reversed, and $L(t)$ is negative.
- (C) is a plot of $L(t)$ versus time corresponding to the responses generated in A-B.

(D-F) The stimulus is a grating moving to the right. Because of the tilt of the space-time receptive field, the alignment with the right-moving grating is never optimal and the response is weak (F).

Remark 2.54. Although x-corrdination of the receptive field in the Example 2.93 changes over time, the variation range of x is only the range displayed by the three oval regions. x does not always move in one direction over time, but move periodically within this range over time.

Remark 2.55. As a result, a neuron with a nonseparable space-time receptive field can be *selective for the direction of motion* of a grating and *for its velocity*, responding most vigorously to an optimally spaced grating moving at a velocity given by $c\tan(\psi)$. That is, the moving velocity of the space-time receptive field is the *preferred velocity.*

2.3.6 Static Nonlinearities: Simple Cells

Remark 2.56. Once the linear response estimate $L(t)$ has been computed, the firing rate of a visually responsive neuron can be approximated by using Equation 2.18, $r_{\text{est}}(t) = r_0 + F(L(t))$, where F is an appropriately chosen static nonlinearity.

Example 2.94. The simplest choice for F consistent with the positive nature of firing rates is rectification, $F = G[L]_+$, with G set to fit the magnitude of the measured firing rates.

Remark 2.57. However, the choice in Example 2.94 makes the firing rate a linear function of the contrast amplitude, which does not match the data on the contrast dependence of visual responses.

Definition 2.95. The *contrast saturation* means neural responses saturate as the contrast of the image increases, and are more accurately described contrast saturation by

$$r \propto A^n/(A_{1/2}^n + A^n)$$

where n is near 2, and $A_{1/2}$ is a parameter equal to the contrast amplitude that produces a half-maximal response.

Proposition 2.96. A static nonlinearity defined by

$$F(L) = \frac{G[L]_+^2}{A_{1/2} + G[L]_+^2} \tag{2.52}$$

reproduces the observed contrast dependence.

2.4 Static Nonlinearities: Complex Cells

Remark 2.58. The spatial receptive fields of complex cells cannot be divided into separate ON and OFF regions that sum linearly to generate the response. Areas where light and dark images excite the neuron overlap, making it difficult to measure and interpret spike-triggered average stimuli.

Principle 2.97. Like simple cells, complex cells are *selective to the spatial frequency and orientation* of a grating. However, unlike simple cells, complex cells respond to bars of light or dark no matter where they are placed within the overall receptive field. Likewise, the responses of complex cells to grating stimuli show *little dependence on spatial phase.*

Definition 2.98. The phenomenon that a complex cell is selective for a particular type of image independent of its exact spatial position within the receptive field is called the *spatial-phase invariance.*

Remark 2.59. The spatial-phase invariance of complex cells may represent an early stage in the visual processing that ultimately leads to position-invariant object recognition.

Remark 2.60. Complex cells also have temporal response characteristics that distinguish them from simple cells.

Principle 2.99. Complex cell responses to moving gratings are approximately constant, not oscillatory as in examples 2.88 and 2.93. The firing rate of a complex cell responding to a counterphase grating oscil lating with frequency ω has both a constant component and an oscillatory component with a frequency of 2ω, a phenomenon known as *frequency doubling.*

Remark 2.61. To give a first approximation to complex-cell responses, the key observation comes from Equation 2.47, which shows how the linear response estimate of a simple cell depends on spatial phase for an optimally oriented grating with K not too small.

Proposition 2.100. Consider two such responses, labeled L_1 and L_2, with preferred spatial phases ϕ and $\phi - \pi/2$. Including both the spatial and the temporal response factors, we find, for preferred spatial phase ϕ

$$L_1 = AB(\omega, K)\cos(\phi - \Phi)\cos(\omega t - \delta), \tag{2.53}$$

where $B(\omega, K)$ is a temporal and spatial frequency-dependent amplitude factor. For preferred spatial phase $\phi - \pi/2$,

$$L_2 = AB(\omega, K)\sin(\phi - \Phi)\cos(\omega t - \delta). \tag{2.54}$$

Thus we have,

$$L_1^2 + L_2^2 = A^2 B^2(\omega, K)\cos^2(\omega t - \delta). \tag{2.55}$$

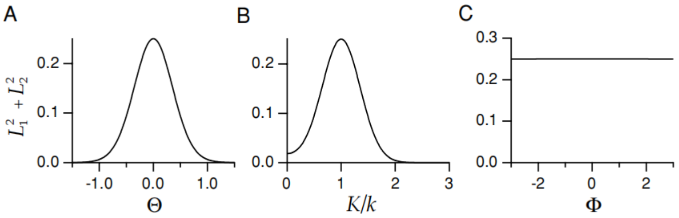
Proof. Equation 2.53 follows from Equation 2.47, Equation 2.54 follows from $\cos(\phi - \pi/2 - \Phi) = \sin(\phi - \Phi)$ and Equation 2.55 follows from $\cos^2(\phi - \Phi) + \sin^2(\phi - \Phi) = 1$. □

Definition 2.101. We can describe the *spatial-phase-invariant response* of a complex cell by writing

$$r(t) = r_0 + G(L_1^2 + L_2^2), \tag{2.56}$$

for some constant G.

Example 2.102. Selectivity of the complex cell model (Equation 2.56) in response to a sinusoidal grating is shown in the following figures. The width and preferred spatial frequency of the Gabor functions underlying the estimated firing rate satisfy $k\sigma = 2$.



A The complex cell response estimate, $L_1^2 + L_2^2$, as a function of stimulus orientation Θ for a grating with the preferred spatial frequency $K = k$.

B $L_1^2 + L_2^2$ as a function of the ratio of the stimulus spatial frequency to its preferred value, K/k, for a grating oriented in the preferred direction $\Theta = 0$.

C $L_1^2 + L_2^2$ as a function of stimulus spatial phase Φ for a grating with the preferred spatial frequency and orientation, $K = k$ and $\Theta = 0$.

Remark 2.62. The response of the model complex cell is tuned to orientation and spatial frequency, but the spatial phase dependence, illustrated for a simple cell in Example 2.102 figure C, is absent. In computing the curve for Example 2.102 figure C, we used the exact expressions for L_1 and L_2 from the integrals in equations 2.44 and 2.45, not the approximation used in Equation 2.47 to simplify the previous discussion. Although it is not visible in the figure, there is a weak dependence on Φ when the exact expressions are used.

Proposition 2.103. The complex cell response given by equations 2.55 and 2.56 reproduces the frequency-doubling effect seen in complex cell responses.

Proof. This follows from the identity

$$\cos^2(\omega t - \delta) = \frac{1}{2}\cos(2(\omega t - \delta)) + \frac{1}{2}, \tag{2.57}$$

where the last term on the right side of this equation generates the constant component of the complex cell response to a counterphase grating. □

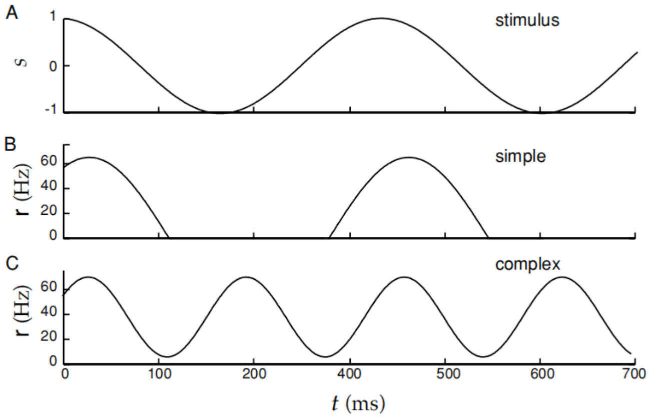
Example 2.104. Temporal responses of model simple and complex cells to a counterphase grating is shown in the following figures.

(A) The stimulus $s(x, y, t)$ at a given point (x, y) plotted as a function of time.

(B) The rectified linear response estimate of a model simple cell to this grating with a temporal kernel given by Equation 2.42 with $\alpha = 1/(15 \text{ ms})$.

(C) The "frequency-doubled" response of a model complex cell with the same temporal kernel but with the estimated rate given by a squaring operation rather than rectification. The background firing rate is $r_0 = 5$ Hz.

Note the temporal phase shift of both B and C relative to A.



Definition 2.105. The description of a complex cell response that we have presented is called an *energy model* because of its resemblance to the equation for the energy of a simple harmonic oscillator. The pair of linear filters used, with preferred spatial phases separated by $\pi/2$, is called a *quadrature pair.*

Proposition 2.106. We can write the complex cell response as the sum of the squares of four rectified simple cell responses,

$$r(t) = r_0 + G([L_1]_+^2 + [L_2]_+^2 + [L_3]_+^2 + [L_4]_+^2), \tag{2.58}$$

where the different $[L]_+$ terms represent the responses of simple cells with preferred spatial phases ϕ, $\phi + \pi/2$, $\phi + \pi$, and $\phi + 3\pi/2$.

Proof. Because of rectification, the terms L_1^2 and L_2^2 cannot be constructed by squaring the outputs of single simple cells. However, they can each be constructed by summing the squares of rectified outputs from two simple cells with preferred spatial phases separated by π. □

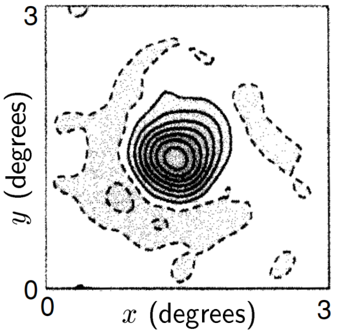
Remark 2.63. While such a construction is possible, it should not be interpreted too literally because complex cells receive input from many sources, including the LGN and other complex cells. Rather, this model should be viewed as purely descriptive. Simple mechanistic models of complex cells are described at the end of this chapter.

2.5 Receptive Fields in the Retina and LGN

Definition 2.107. A receptive field with a center-surround structure consisting either of a circular central ON region surrounded by an annular OFF region is called *ON-center*, or the opposite arrangement of a central OFF region surrounded by an ON region is called *OFF-center.*

Remark 2.64. Retinal ganglion cells display a wide variety of response characteristics, including nonlinear and direction-selective responses. However, a class of retinal ganglion cells (X cells in the cat or P cells in the monkey retina and LGN) can be described by a linear model built using reverse-correlation methods. The receptive fields of this class of retinal ganglion cells and an analogous type of LGN relay neurons are similar. The receptive fields of these neurons are ON-center or OFF-center.

Example 2.108. The following figure shows the center-surround spatial structure of the receptive field of a cat LGN X cell. This has a central ON region (solid contours) and a surrounding OFF region (dashed contours).

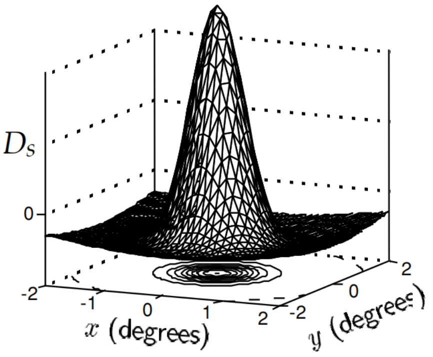


Definition 2.109. A *difference-of-Gaussians model* capturing the spatial structure of retinal ganglion and LGN receptive fields is expressed as

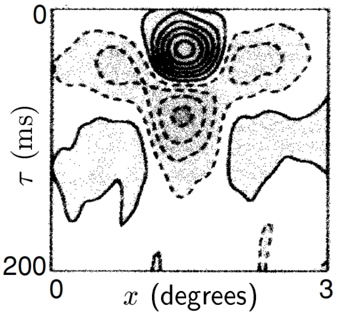
$$D_s(x, y) = \pm\left(\frac{1}{2\pi\sigma_{\text{cen}}^2} e^{-\frac{x^2+y^2}{2\sigma_{\text{cen}}^2}} - \frac{B}{2\pi\sigma_{\text{sur}}^2} e^{-\frac{x^2+y^2}{2\sigma_{\text{sur}}^2}}\right), \tag{2.59}$$

where the first Gaussian function describes the center, the second describes the surround, σ_{cen} determines the size of the central region, σ_{sur}, which is greater than σ_{cen}, determines the size of the surround, B controls the balance between center and surround contributions, the ± sign allows both ON-center (+) and OFF-center (−) cases to be represented.

Example 2.110. A fit of the receptive field shown in Example 2.108 using a difference-of-Gaussians function (Equation 2.59) with $\sigma_{\text{cen}} = 0.3°$, $\sigma_{\text{csur}} = 1.5°$, and $B = 5$.



Example 2.111. The following figure shows the space-time receptive field of a cat LGN X cell. Note that the center and surround regions both reverse sign as a function of τ and that the temporal evolution is slower for the surround than for the center.



Because of the difference between the time course of the center and of the surround regions, the space-time receptive field is not separable, although the center and surround components are individually separable.

Definition 2.112. A model capturing basic features of LGN neuron space-time receptive fields is expressed as

$$D(x, y, \tau) = \pm \left(\frac{D_t^{\text{cen}}(\tau)}{2\pi\sigma_{\text{cen}}^2} e^{-\frac{x^2+y^2}{2\sigma_{\text{cen}}^2}} - \frac{B D_t^{\text{sur}}(\tau)}{2\pi\sigma_{\text{sur}}^2} e^{-\frac{x^2+y^2}{2\sigma_{\text{sur}}^2}} \right), \tag{2.60}$$

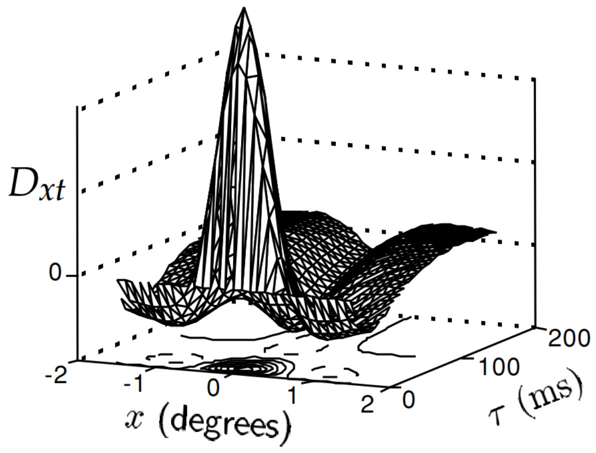
where $D_t^{\text{cen}}(\tau)$ and $D_t^{\text{sur}}(\tau)$ can both be described by the same functions, using two sets of parameters,

$$D_t^{\text{cen,sur}}(\tau) = \alpha_{\text{cen,sur}}^2 \tau e^{-\alpha_{\text{cen,sur}}\tau} - \beta_{\text{cen,sur}}^2 \tau e^{-\beta_{\text{cen,sur}}\tau}, \tag{2.61}$$

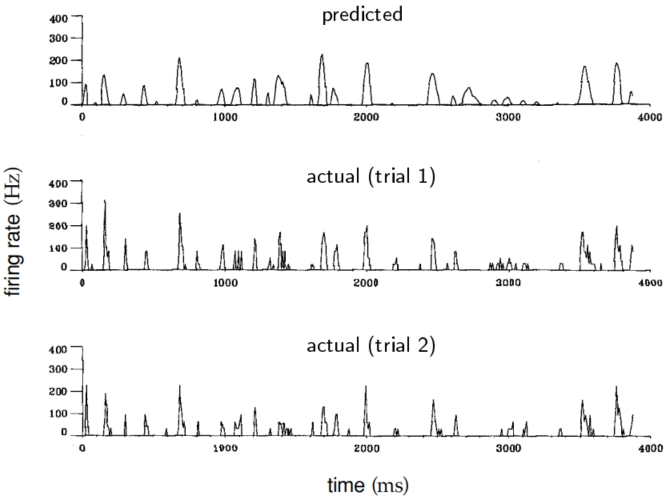
where α_{cen} and α_{sur} control the latency of the response in the center and surround regions, respectively, and β_{cen} and β_{sur} affect the time of the reversal.

Remark 2.65. This function in Equation 2.61 has characteristics similar to the function in Equation 2.42, but the latency effect is less pronounced.

Example 2.113. A fit of the space-time receptive field in Example 2.111 using Equation 2.60 with the same parameters for the Gaussian functions as in Example 2.110, and temporal factors given by Equation 2.61 with $1/\alpha_{\text{cen}} = 16$ ms, $1/\alpha_{\text{sur}} = 32$ ms, and $1/\beta_{\text{cen}} = 1/\beta_{\text{sur}} = 64$ ms.



Example 2.114 (A direct test of a reverse-correlation model of an LGN neuron)**.** Comparison of predicted and measured firing rates for a cat LGN neuron responding to a video movie is shown in the following figures.



(i) The top panel is the rate predicted by integrating the product of the video image intensity and the kernel needed to describe this neuron was first extracted by using a white-noise stimulus. The resulting linear prediction was rectified, that is, $F(L) = G[L]_+$.

(ii) The middle and lower panels are measured firing rates extracted from two different sets of trials.

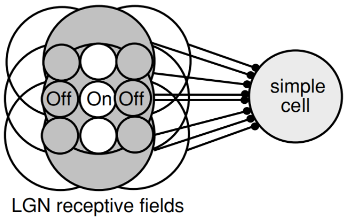
Remark 2.66. In Example 2.114, the correlation coefficient between the predicted and actual firing rates was 0.5, which was very close to the correlation coefficient between firing rates extracted from different groups of trials. This means that the error of the prediction was no worse than the variability of the neural response itself.

2.6 V1 Receptive Fields Construction

Remark 2.67. The models of visual receptive fields we have been discussing are purely descriptive, but they provide an important framework for studying how the circuits of the retina, LGN, and primary visual cortex generate neural responses.

Definition 2.115. The *Hubel-Wiesel Model* propose the oriented receptive fields of cortical neurons could be generated by summing the input from appropriately selected LGN neurons.

Example 2.116 (The Hubel-Wiesel model of a simple cell)**.** The Hubel-Wiesel model of orientation selectivity is shown below.



The spatial arrangement of the receptive fields of nine LGN neurons are shown, with a column of three ON-center fields flanked on either side by columns of three OFF-center fields. White areas denote ON fields and gray areas, OFF fields. In the model, the converging LGN inputs are summed by the simple cell. *This arrangement produces a receptive field oriented in the vertical direction.* Note that, two center types are represented as

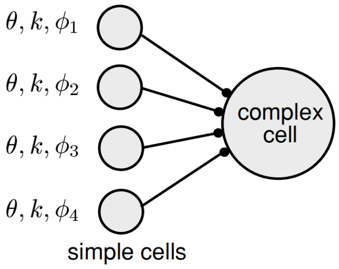
(i) ON-center field: a concentric circle with an white inner circle and an gray outer circle;

(ii) OFF-center field: a concentric circle with an gray inner circle and an white outer circle.

Remark 2.68. This model accounts for the selectivity of a simple cell purely on the basis of feedforward input from the LGN.

Remark 2.69. In a previous section, we showed how the properties of complex cell responses could be accounted for by using a squaring static nonlinearity. While this provides a good description of complex cells, there is little indication that complex cells actually square their inputs. Models of complex cells can be constructed without introducing a squaring nonlinearity.

Example 2.117 (The Hubel-Wiesel model of a complex cell)**.** Inputs from a number of simple cells with similar orientation and spatial frequency preferences (θ and k), but different spatial phase preferences (ϕ_1, ϕ_2, ϕ_3, and ϕ_4), converge on a complex cell and are summed. *This produces a complex cell output that is selective for orientation and spatial frequency, but not for spatial phase (phase-invariant response).*



The figure shows four simple cells converging on a complex cell, but additional simple cells can be included to give a more complete coverage of spatial phase.

Remark 2.70. While the model generates complex cell responses, there are indications that complex cells in primary visual cortex are not driven exclusively by simple cell input. An alternative model is considered in chapter 7.

2.7 Questions

This section states the questions that we can't solve or the concepts that we can't understand.

2.7.1 the Bandwidth

This sunsection belongs to Chapter 2 section 2.3.

Definition 2.118. The *bandwidth* is defined as

$$b = \log_2(K_+/K_-),$$

where $K_+ > k$ and $K_- < k$ are the spatial frequencies of gratings that produce one-half the response amplitude of a grating with $K = k$.

Chapter 3

Neural Decoding

3.1 3.1

Principle 3.1 (Conservation of angular momentum)**.** The rate of change of angular momentum of a system is equal to the net torque acting on the system, i.e.,

$$\frac{\mathrm{d}\mathbf{L}}{\mathrm{d}t} = \boldsymbol{\tau}, \tag{3.1}$$

where $\boldsymbol{\tau}$ is the torque of all external forces on the system about any chosen axis, and $\mathrm{d}\mathbf{L}/\mathrm{d}t$ is the rate of change of angular momentum of the system about the same axis.

3.2 3.2

Remark 3.1. Many physical laws are cumbersome when written in coordinate form but become more compact and attractive looking when written in tensorial form. For example, the incompressible Navier-Stokes equations in cylindrical coordinates are

$$\begin{aligned}
\rho\left(\frac{Dv_r}{Dt} - \frac{v_\theta^2}{r}\right) &= \rho f_r - \frac{\partial p}{\partial r} + \mu\left(\Delta v_r - \frac{v_r}{r^2} - \frac{2}{r^2}\frac{\partial v_\theta}{\partial \theta}\right), \\
\rho\left(\frac{Dv_\theta}{Dt} + \frac{v_r v_\theta}{r}\right) &= \rho f_\theta - \frac{1}{r}\frac{\partial p}{\partial \theta} + \mu\left(\Delta v_\theta + \frac{2}{r^2}\frac{\partial v_r}{\partial \theta} - \frac{v_\theta}{r^2}\right), \\
\rho\frac{Dv_z}{Dt} &= \rho f_z - \frac{\partial p}{\partial z} + \mu\Delta v_z,
\end{aligned}$$

where

$$\Delta = \frac{1}{r}\frac{\partial}{\partial r}\left(r\frac{\partial}{\partial r}\right) + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial z^2},$$

and

$$\frac{D}{Dt} = \frac{\partial}{\partial t} + v_r\frac{\partial}{\partial r} + \frac{v_\theta}{r}\frac{\partial}{\partial \theta} + v_z\frac{\partial}{\partial z}.$$

3.3 3.3

Proposition 3.2. The mass of fluid in a region W at time t is

$$m(W,t) = \int_W \rho(\mathbf{x},t)\mathrm{d}V, \tag{3.2}$$

where $\mathrm{d}V$ is the area element in the plane or the volume element in space.

3.4 3.4

Assumption 3.3. From now on, assume that

$$\text{force on } S \text{ per unit area} = -p(\mathbf{x},t)\mathbf{n} + \mathbf{n}\cdot\boldsymbol{\sigma}(\mathbf{x},t), \tag{3.3}$$

where $\boldsymbol{\sigma}$ is the *(deviatoric) stress tensor* and $\mathbf{n}$ is the unit outward normal of S.