Exercises

Chapter 1

1. Generate spikes for 10 s (or longer if you want better statistics) using a Poisson spike generator with a constant rate of 100 Hz, and record their times of occurrence. Compute the coefficient of variation of the interspike intervals, and the Fano factor for spike counts obtained over counting intervals ranging from 1 to 100 ms. Plot the interspike interval histogram.

2. Add a refractory period to the Poisson spike generator by allowing the firing rate to depend on time. Initially, set the firing rate to a constant value, $\mathsf{r}(t) = \mathsf{r}_0$. After every spike, set $\mathsf{r}(t)$ to 0, and then allow it to recover exponentially back to r_0 with a time constant τ_{ref} that controls the refractory recovery rate. In other words, have $\mathsf{r}(t)$ obey the equation

   $$\tau_{\text{ref}} \frac{d\mathsf{r}}{dt} = \mathsf{r}_0 - \mathsf{r}$$

   except immediately after a spike, when it is set to 0. Plot the coefficient of variation as a function of τ_{ref} over the range 1 ms $\leq \tau_{\text{ref}} \leq$ 20 ms, and plot interspike interval histograms for a few different values of τ_{ref} in this range. Compute the Fano factor for spike counts obtained over counting intervals ranging from 1 to 100 ms for the case $\tau_{\text{ref}} = 10$ ms.

3. Compute autocorrelation histograms of spike trains generated by a Poisson generator with a constant firing rate of 100 Hz, a constant firing rate of 100 Hz together with a refractory period modeled as in exercise 2 with $\tau_{\text{ref}} = 10$ ms, and a variable firing rate $\mathsf{r}(t) = 100(1 + \cos(2\pi t/25 \text{ ms}))$ Hz. Plot the histograms over a range from 0 to 100 ms.

4. Generate a Poisson spike train with a time-dependent firing rate $\mathsf{r}(t) = 100(1 + \cos(2\pi t/300 \text{ ms}))$ Hz. Approximate the firing rate from this spike train using a variable $\mathsf{r}_{\text{approx}}$ that satisfies

   $$\tau_{\text{approx}} \frac{d\mathsf{r}_{\text{approx}}}{dt} = -\mathsf{r}_{\text{approx}} ,$$

   except that $\mathsf{r}_{\text{approx}} \rightarrow \mathsf{r}_{\text{approx}} + 1/\tau_{\text{approx}}$ every time a spike occurs. Make plots of the true rate, the spike sequence generated, and the estimated rate. Experiment with a few different values of τ_{approx} in the range of 1 to 100 ms. Determine the best value of τ_{approx} by computing the average squared error of the estimate, $\int dt(\mathsf{r}(t) - \mathsf{r}_{\text{approx}}(t))^2$, for different values of τ_{approx}, and finding the value of τ_{approx} that minimizes this error.

5. For a constant rate Poisson process, every specific (up to a finite resolution) sequence of N spikes occurring over a given time interval

is equally likely. This seems paradoxical because we certainly do not expect to see all N spikes appearing within the first 1% of the time interval. Resolve this paradox.

6. Build an approximate white-noise stimulus by choosing random values at discrete times separated by a time-step interval Δt. Plot its autocorrelation function and power spectrum (use the MATLAB® function `spectrum` or `psd`). Discuss how well this stimulus matches an ideal white-noise stimulus given the value of Δt you used.

7. Consider a model with a firing rate determined in terms of a stimulus $s(t)$ by integrating the equation

$$\tau_r \frac{d\mathrm{r}_{\mathrm{est}}(t)}{dt} = [\mathrm{r}_0 + s]_+ - \mathrm{r}_{\mathrm{est}}(t)\,,$$

   where r_0 is a constant that determines the background firing rate and $\tau_r = 20$ ms. Drive the model with an approximate white-noise stimulus. Adjust the amplitude of the white-noise and the parameter r_0 so that rectification is not a big effect (i.e. $\mathrm{r}_0 + s > 0$ most of the time). From the responses of the model, compute the stimulus-response correlation function, Q_{rs}. Next, generate spikes from this model using a Poisson generator with a rate $\mathrm{r}_{\mathrm{est}}(t)$, and compute the spike-triggered average stimulus from the spike trains produced by the white-noise stimulus. By comparing the stimulus-response correlation function with the spike-triggered average, verify that equation 1.22 is satisfied. Examine what happens if you set $\mathrm{r}_0 = 0$, so that the white-noise stimulus becomes half-wave rectified.

8. MATLAB® file `c1p8.mat` contains data collected and provided by Rob de Ruyter van Steveninck from a fly H1 neuron responding to an approximate white-noise visual motion stimulus. Data were collected for 20 minutes at a sampling rate of 500 Hz. In the file, `rho` is a vector that gives the sequence of spiking events or nonevents at the sampled times (every 2 ms). When an element of `rho` is one, this indicates the presence of a spike at the corresponding time, whereas a zero value indicates no spike. The variable `stim` gives the sequence of stimulus values at the sampled times. Calculate and plot the spike-triggered average from these data over the range from 0 to 300 ms (150 time steps). (Based on a problem from Sebastian Seung.)

9. Using the data of problem 8, calculate and plot stimulus averages triggered on events consisting of a pair of spikes (which need not necessarily be adjacent) separated by a given interval (as in figure 1.10). Plot these two-spike-triggered average stimuli for various separation intervals ranging from 2 to 100 ms. (Hint: in MATLAB® , use convolution for pattern matching: e.g. `find(conv(rho,[1 0 1])==2)` will contain the indices of all the events with two spikes separated by 4 ms.) Plot, as a function of the separation between the two spikes, the magnitude of the difference between the two-spike-triggered average and the sum of two single-spike-triggered averages (obtained

in exercise 8) separated by the same time interval. At what temporal separation does this difference become negligibly small. (Based on a problem from Sebastian Seung.)

10. Using the data of problem 8, find the spike-triggered average stimulus for events that contain exactly two adjacent spikes separated by various different intervals ranging from 2 to 100 ms (e.g. for 4 ms, the event `[1 0 1]` but not the event `[1 1 1]`). This is distinct from exercise 9 in which we only required two spikes separated by a given interval, but did not restrict what happened between the two spikes. Compare results of the exclusive case considered here with those of the inclusive two-spike-triggered average computed in exercise 9. In what ways and why are they different? (Based on a problem from Sebastian Seung.)

Exercises

Chapter 2

1. Use the rate given by equation 2.1 with $r_0 = 50$ Hz and

$$D(\tau) = -\cos\left(\frac{2\pi(\tau - 20\text{ ms})}{140\text{ ms}}\right)\exp\left(-\frac{\tau}{60\text{ ms}}\right)\text{ Hz/ms}$$

to predict the response of a neuron of the electrosensory lateral-line lobe to a stimulus. The above equation is an approximation for the linear kernel obtained from the spike-triggered average shown in figure 1.9. Use an approximate Gaussian white noise stimulus constructed by choosing a stimulus value every 10 ms ($\Delta t = 10$ ms) from a Gaussian distribution with zero mean and variance $\sigma_s^2/\Delta t$, with $\sigma_s^2 = 10$. Compute the firing rate over a 10 s period. From the results, compute the firing rate-stimulus correlation function $Q_{rs}(\tau)$. Using equation 2.6, compare $Q_{rs}(-\tau)/\sigma_s^2$ with the kernel $D(\tau)$ given above.

2. MATLAB® file `c1p8.mat` contains the data described in exercise 8 of chapter 1. Use the spike-triggered average (calculated in that exercise) to construct a linear kernel and use it in equation 2.1 to provide a model of the response of the H1 neuron. Choose r_0 so that the average firing rate predicted by the model in response to the stimulus used for the data matches the actual average firing rate. Use a Poisson generator with the computed rate to generate a synthetic spike train from this linear estimate of the firing rate in response to the stimulus `stim`. Plot examples of the actual and synthetic spike trains. How are they similar and how do they differ? Plot the autocorrelation function of the actual and the synthetic spike trains over the range 0 to 100 ms. Why is there a dip at a lag of 2 ms in the autocorrelation of the actual spike train? Is there a dip for the synthetic train too? Plot the interspike interval histogram for both spike trains. Why is there a dip below 6 ms in the histogram for the actual spike train? What are the coefficients of variation for the two spike trains and why might they differ? (Based on a problem from Sebastian Seung).

3. MATLAB® file `c2p3.mat` contains the responses of a cat LGN cell to two-dimensional visual images (these data are described in Kara, P, Reinagel, P, & Reid, RC (2000) Low response variability in simultaneously recorded retinal, thalamic, and cortical neurons. *Neuron* **30**:803-817 and were kindly provided by Clay Reid). In the file, `counts` is a vector containing the number of spikes in each 15.6 ms bin, and `stim` contains the 32767, 16 × 16 images that were presented at the corresponding times. Specifically, `stim(x, y, t)` is the stimulus presented at the coordinate `(x,y)` at time-step `t`. Note that `stim` is an `int8` array that must to be converted into `double` using the command `stim=double(stim)` in order to be manipulated within MATLAB®. Calculate the spike-triggered average images for each of

the 12 time steps before each spike and show them all (using the `imagesc` command). Note that in this example, the time bins can contain more than one spike, so the spike-triggered average must be computed by weighting each stimulus by the number of spikes in the corresponding time bin, rather than weighting it by either 1 or 0 depending on whether a spike is present or not. In the averaged images, you should see a central receptive field that reverses sign over time. By summing up the images across one spatial dimension, produce a figure like that of figure 2.25C. (Based on a problem from Sebastian Seung.)

4. For a Gaussian random variable x with zero mean and standard deviation σ, prove that

$$\langle xF(\alpha x)\rangle = \alpha\sigma^2\langle F'(\alpha x)\rangle\,,$$

where α is a constant, F is any function, F' is its derivative,

$$\langle xF(\alpha x)\rangle = \int dx\,\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2}{2\sigma^2}\right)xF(\alpha x)\,,$$

and similarly for $\langle F'(\alpha x)\rangle$. This is the basis of the identity 2.64, which can be derived by extending this basic result first to multivariate functions and then to functionals.

5. Using the inverses of equations 2.15 and 2.17

$$\epsilon = \epsilon_0\left(\exp(X/\lambda) - 1\right) \quad \text{and} \quad a = -\frac{180^\circ(\epsilon_0+\epsilon)Y}{\lambda\epsilon\pi}\,,$$

map from cortical coordinates back to visual coordinates and determine what various patterns of activity across the primary visual cortex would "look like". Ermentrout and Cowan (Ermentrout, GB, & Cowan, J (1979) A mathematical theory of visual hallucination patterns. *Biological Cybernetics* **34**:137–150) used these results as a basis of a mathematical theory of visual hallucinations. The figure generated by the MATLAB® program `c2p5.m` shows an illustrative example. This program simulates a plane sine wave of activity across the primary visual cortex with a specified spatial frequency and direction, and then maps it back into retinal coordinates to see what visual pattern would be perceived due to this activity. Consider various other patterns of activity and show the visual hallucinations they would generate.

6. Perform the integrals in equations 2.31 and 2.32 for the case $\sigma_x = \sigma_y = \sigma$ to obtain the results

$$\begin{aligned} L_s &= \frac{A}{2}\exp\left(-\frac{\sigma^2(k^2+K^2)}{2}\right)\Big(\cos(\phi-\Phi)\exp\left(\sigma^2 kK\cos(\Theta)\right) \\ &\quad + \cos(\phi+\Phi)\exp\left(-\sigma^2 kK\cos(\Theta)\right)\Big)\,. \end{aligned}$$

and

$$L_t(t) = \frac{\alpha^6|\omega|\sqrt{\omega^2+4\alpha^2}}{(\omega^2+\alpha^2)^4}\cos(\omega t-\delta)\,.$$

with

$$\delta = 8\arctan\left(\frac{\omega}{\alpha}\right) + \arctan\left(\frac{2\alpha}{\omega}\right) - \pi\,.$$

From these results, verify the selectivity curves in figures 2.15 and 2.16. In addition, plot δ as a function of ω.

7. Numerically compute the spatial part of the linear response of a simple cell with a separable space-time receptive field to a sinusoidal grating, as given by equation 2.31. Use a stimulus oriented with $\Theta = 0$. For the spatial receptive field kernel, use equation 2.27 with $\sigma_x = \sigma_y = 1^\circ$, $\phi = 0$, and $1/k = 0.5^\circ$. Plot L_s as a function of K taking $\Phi = 0$ and $A = 50$. This determines the spatial frequency selectivity of the cell. What is its preferred spatial frequency? Plot L_s as a function of Φ taking $1/K = 0.5^\circ$ and $A = 50$. This determines the spatial phase selectivity of the cell. What is its preferred spatial phase?

8. Consider a complex cell with the spatial part of its response given by $L_1^2 + L_2^2$, where L_1 and L_2 are linear responses determined by equation 2.31 with kernels given by equation 2.27 with $\sigma_x = \sigma_y = 1^\circ$, and $1/k = 0.5^\circ$; and with $\phi = 0$ for L_1 and $\phi = -\pi/2$ for L_2. Use a stimulus oriented with $\Theta = 0$. Compute and plot $L_1^2 + L_2^2$ as a function of K taking $\Phi = 0$ and $A = 5$. This determines the spatial frequency selectivity of the cell. Compute and plot $L_1^2 + L_2^2$ as a function of Φ taking $1/K = 0.5^\circ$ and $A = 5$. This determines the spatial phase selectivity of the cell. Does the spatial phase selectivity match what you expect for a complex cell?

9. Consider the linear temporal response for a simple or complex cell given by equation 2.32 with a temporal kernel given by equation 2.29 with $1/\alpha = 15$ ms. Compute and plot $L_\mathrm{t}(t)$ for $\omega = 6\pi/s$. This determines the temporal response of the simple cell. Do not plot the negative part of $L_\mathrm{t}(t)$ because the cell cannot fire at a negative rate. Compute and plot $L_\mathrm{t}^2(t)$ for $\omega = 6\pi/s$. This determines the temporal response of a complex cell. What are the differences between the temporal responses of the simple and complex cells?

10. Compute the response of a model simple cell with a separable space-time receptive field to a moving grating

$$s(x, y, t) = \cos(Kx - \omega t)\,.$$

For D_s, use equation 2.27 with $\sigma_x = \sigma_y = 1^\circ$, $\phi = 0$, and $1/k = 0.5^\circ$. For D_t, use equation 2.29 with $1/\alpha = 15$ ms. Compute the linear estimate of the response given by equation 2.24 and assume that the actual response is proportional to a rectified version of this linear response estimate. Plot the response as a function of time for $1/K = 1/k = 0.5^\circ$ and $\omega = 8\pi/\mathrm{s}$. Plot the response amplitude as a function of ω for $1/K = 1/k = 0.5^\circ$ and as a function of K for $\omega = 8\pi/\mathrm{s}$.

11. Compute the response of a model complex cell to the moving grating

$$s(x, y, t) = \cos(Kx - \omega t)\,.$$

The complex cell should be modeled by squaring the unrectified linear response estimate of a simple cells with a spatial receptive field given by equation 2.27 with $\sigma_x = \sigma_y = 1^\circ$, $\phi = 0$, and $1/k = 0.5^\circ$, and adding this to the square of the unrectified linear response of a second simple cell with identical properties except that its spatial phase preference is $\phi = -\pi/2$ instead of $\phi = 0$. Both linear responses are computed from equation 2.24. For both of these, use equation 2.29 with $1/\alpha = 15$ ms for the temporal receptive field. Plot the complex cell response as a function of time for $1/K = 1/k = 0.5^\circ$ and $\omega = 8\pi/\text{s}$. Plot the response amplitude as a function of ω for $1/K = 1/k = 0.5^\circ$ and as a function of K for $\omega = 8\pi/\text{s}$.

12. Construct a model simple cell with the nonseparable space-time receptive field described in the caption of figure 2.21B. Compute its response to the moving grating

$$s(x, y, t) = \cos(Kx - \omega t)\,.$$

Plot the amplitude of the response as a function of the velocity of the grating, ω/K, using $\omega = 8\pi/\text{s}$ and varying K to obtain a range of both positive and negative velocity values (use negative K values for this). Show that the response is directionally selective.

13. Construct a model complex cell that is disparity tuned but insensitive to the absolute position of a grating. The complex cell is constructed by summing the squares of the unrectified linear responses of two simple cells, but disparity effects are now included. For this exercise, we ignore temporal factors and only consider the spatial dependence of the response. Each simple cell response is composed of two terms that correspond to inputs coming from the left and right eyes. Because of disparity, the spatial phases of the image of a grating in the two eyes, Φ_L and Φ_R, may be different. We write the spatial part of the linear response estimate for a grating with the preferred spatial frequency ($k = K$) and orientation ($\Theta = \theta = 0$) as

$$L_1 = \frac{A}{2}\left(\cos(\Phi_L) + \cos(\Phi_R)\right)\,,$$

assuming that $\phi = 0$ (this equation is a generalization of equation 2.34). Let the complex cell response be proportional to $L_1^2 + L_2^2$, where L_2 is similar to L_1 but with the cosine functions replaced by sine functions. Show that the response of this neuron is tuned to the disparity, $\Phi_L - \Phi_R$, but is independent of the absolute spatial phase of the grating, $\Phi_L + \Phi_R$. Plot the response tuning curve as a function of disparity. (See DeAngelis, GC, Ohzawa, I, & Freeman, RD (1991) Depth is encoded in the visual cortex by a specialized receptive field structure. *Nature* **352**:156–159.)

14. Determine the selectivity of the LGN receptive field of equation 2.45 to spatial frequency by computing its integrals when multiplied by the stimulus

$$s = \cos(Kx)$$

for a range of K values. Use $\sigma_c = 0.3°$, $\sigma_s = 1.5°$, $B = 5$ and plot the resulting spatial frequency tuning curve.

15. Construct the Hubel-Wiesel model of a simple-cell spatial receptive field, as depicted in figure 2.27A. Use difference-of-Gaussian functions (equation 2.45) to model the LGN receptive fields. Plot the spatial receptive field of the simple cell constructed by summing the spatial receptive fields of the LGN cells that provide its input. Compare the result of summing appropriately placed LGN center-surround receptive fields (figure 2.27A) with the results of an appropriately adjusted Gabor filter model of the simple cell that uses the spatial kernel of equation 2.27.

16. Construct the Hubel-Wiesel model of a complex cell, as depicted in figure 2.27B. Use Gabor functions (equation 2.27) to model the simple cell responses, which should be rectified before being summed. Plot the spatial receptive field of the complex cell constructed by summing the different simple cell responses. Compare the responses of a complex cell constructed by linearly summing the outputs of simple cells (figure 2.27B) with different spatial phase preferences with the complex cell model obtained by squaring and summing two unrectified simple cell responses with spatial phases 90° apart as in exercise 8.

Exercises

Chapter 3

1. Suppose that the probabilities that a neuron responds with a firing rate between r and $r + \Delta r$ to two stimuli labeled plus and minus are $p[r|\pm]\Delta r$ where

   $$p[r|\pm] = \frac{1}{\sqrt{2\pi}\sigma_r} \exp\left(-\frac{1}{2}\left(\frac{r - \langle r \rangle_\pm}{\sigma_r}\right)^2\right) .$$

   Assume that the two mean rate parameters $\langle r \rangle_+$ and $\langle r \rangle_-$ and the single variance σ_r^2 are chosen so that these distributions produce negative rates rarely enough that we can integrate over r values over the entire range $-\infty < r < \infty$. Suppose that you base discrimination of the plus and minus stimuli on whether the evoked firing rate is greater or less than a threshold z. Show that the size and power, $\alpha(z)$ and $\beta(z)$ of this test are given by

   $$\alpha(z) = \frac{1}{2}\mathrm{erfc}\left(\frac{z - \langle r \rangle_-}{\sqrt{2}\sigma_r}\right) \quad \text{and} \quad \beta(z) = \frac{1}{2}\mathrm{erfc}\left(\frac{z - \langle r \rangle_+}{\sqrt{2}\sigma_r}\right) .$$

   Show that the probability of a correct answer in the associated two-alternative forced choice task involving discriminating between plus-then-minus and minus-then-plus presentations of the two stimuli is given by equation 3.10. Also, derive the result of equation 3.17. Plot ROC curves for different values of the discriminability

   $$d' = \frac{\langle r \rangle_+ - \langle r \rangle_-}{\sigma_r} .$$

   By simulation, determine the fraction of correct discriminations that can be made in the two-alternative forced choice task. Show that the fractions of correct answer for different values of d' are equal to the areas under the corresponding ROC curves.

2. Simulate the random-dot discrimination experiment. Denote the stimulus by plus or minus, corresponding to the two directions of motion. On each trial, choose the stimulus randomly with equal probability for the two cases. When the minus stimulus is chosen, generate the responses of the neuron as 20 Hz plus a random Gaussian term with a standard deviation of 10 Hz (set any rates that come out negative to zero). When the plus stimulus is chosen, generate the responses as $20 + 10d$ Hz plus a random Gaussian term with a standard deviation of 10 Hz, where d is the discriminability (again, set any rates that come out negative to zero). First, choose a threshold $z = 20 + 5d$, which is half-way between the means of the two response distributions. Whenever $r \geq z$ guess "plus", otherwise guess "minus". Over a large number of trials (1000, for example) determine how often you get the right answer for different d values. Plot the

percent correct as a function of d over the range $0 \leq d \leq 10$. Next, by allowing z to vary over a range, plot ROC curves for several values of d (starting with $d = 2$). To do this, determine how frequently the guess is "plus" when the stimulus is, in fact, plus (this is β), and how often the guess is "plus" when the real stimulus is minus (this is α). Then, plot β versus α for z over the range $0 \leq z \leq 140$.

3. Simulate the responses of four interneurons in the cercal system of the cricket and check the accuracy of a vector decoding scheme. For a true wind direction θ, the average firing rates of the four interneurons should be generated as

$$\langle r_i \rangle = [50\ \mathrm{Hz} \cos(\theta - \theta_i)]_+ \ ,$$

where $[\,]_+$ indicates half-wave rectification, and $\theta_i = \pi/4, 3\pi/4, 5\pi/4, 7\pi/4$ for $i = 1, 2, 3, 4$. The actual rates, r_i, are then obtained by adding to these mean rates a random number chosen from a Gaussian distribution with zero mean and a standard deviation of 5 Hz (set any rates the come out negative to zero). From these rates, construct the x and y components of the population vector

$$x = \sum_{i=1}^{4} r_i \cos(\theta_i) \quad \text{and} \quad y = \sum_{i=1}^{4} r_i \sin(\theta_i)$$

and, from the direction of this vector, compute an estimate θ_{est} of the wind direction. Average the squared difference $(\theta - \theta_{\mathrm{est}})^2$ over 1000 trials. The square root of of this quantity is the error. Plot the error as a function of θ over the range $-90^\circ \leq \theta \leq 90^\circ$.

4. Show that if an infinite number of unit vectors $\vec{c}_a$ is chosen uniformly from a probability distribution that is independent of direction, $\sum(\vec{v} \cdot \vec{c}_a)\vec{c}_a \propto \vec{v}$ for any vector $\vec{v}$. How does the sum approach this limit for a finite number of terms?

5. Show that the Bayesian estimator that minimizes the expected average value of the the loss function $L(s, s_{\mathrm{bayes}}) = (s - s_{\mathrm{bayes}})^2$ is the mean of the distribution $p[s|\mathbf{r}]$, given by equation 3.27. Also show that the estimate that arises from minimizing the expected loss function $L(s, s_{\mathrm{bayes}}) = |s - s_{\mathrm{bayes}}|$ is the median of $p[s|\mathbf{r}]$.

6. Show that the equations for the Fisher information in equation 3.42 can also be written as in equation 3.43,

$$I_{\mathrm{F}}(s) = \left\langle \left(\frac{\partial \ln p[\mathbf{r}|s]}{\partial s} \right)^2 \right\rangle = \int d\mathbf{r}\, p[\mathbf{r}|s] \left(\frac{\partial \ln p[\mathbf{r}|s]}{\partial s} \right)^2$$

or as

$$I_{\mathrm{F}}(s) = \int d\mathbf{r}\, \frac{1}{p[\mathbf{r}|s]} \left(\frac{\partial p[\mathbf{r}|s]}{\partial s} \right)^2 .$$

Use the fact that $\int d\mathbf{r}\, p[\mathbf{r}|s] = 1$.

7. The discriminability for the variable Z defined in equation 3.19 is the difference between the average Z values for the two stimuli $s + \Delta s$ and s divided by the standard deviation of Z. The average of the difference in Z values is

   $$\langle \Delta Z \rangle = \int d\mathbf{r} \, \frac{\partial \ln p[\mathbf{r}|s]}{\partial s} \left(p[\mathbf{r}|s + \Delta s] - p[\mathbf{r}|s] \right) .$$

   Show that for small Δs, $\langle \Delta Z \rangle = I_{\mathrm{F}}(s)\Delta s$. Also prove that the average value of Z,

   $$\langle Z \rangle = \int d\mathbf{r} \, p[\mathbf{r}|s] \frac{\partial \ln p[\mathbf{r}|s]}{\partial s} ,$$

   is zero, and that the variance of Z is $I_{\mathrm{F}}(s)$. Computing the ratio, we find from these results that $d' = \Delta s \sqrt{I_{\mathrm{F}}(s)}$ which matches the discriminability 3.49 of the ML estimator.

8. Extend equation 3.46 to the case of neurons encoding a D-dimensional vector stimulus $\vec{s}$ with tuning curves given by

   $$f_a(\vec{s}) = r_{\max} \exp\left(-\frac{|\vec{s} - \vec{s}_a|^2}{2\sigma_r^2} \right)$$

   and perform the sum by approximating it as an integral over uniformly and densely distributed values of $\vec{s}_a$ to derive the result in equation 3.48.

9. Derive equation 3.54 by minimizing the expression 3.53. Use the methods of appendix A of chapter 2.

10. MATLAB® program `c3p10.m` performs acausal decoding using signal processing techniques to construct an approximate solution of equation 3.54, while suppressing unwanted effects of noise (for an illustration of these effects, see part (e) of this exercise). The program is called as `[est,K,ind]=ch3ex10(stim,spk,nfft)`, where `stim` and `spk` are the stimulus and response respectively, `nfft` is the length of a discrete Fourier transform (suitable values are `nfft=` $2^{10} = 1024$ or $2^{11} = 2048$), `K` is the acausal kernel, `est` is the resulting stimulus estimate, and `ind` is a vector of indices that specifies the range over which the estimate and stimulus should be compared. Specifically, `est` provides an estimate of `stim(ind)`.

    a) Compute an estimated firing rate $\mathrm{r}_{\mathrm{est}}$ from equation 2.1 with $\mathrm{r}_0 = 50$ Hz and

    $$D(\tau) = -\cos\left(\frac{2\pi(\tau - 20\mathrm{ms})}{140\mathrm{ms}} \right) \exp\left(-\frac{\tau}{60\mathrm{ms}} \right) \text{ Hz/ms}$$

    in response to an approximate white noise stimulus (roughly 500 seconds long) calculated by choosing at each time step (with $\Delta t = 10$ ms) a stimulus value from a Gaussian distribution with mean 0 and variance 2. Generate a kernel and estimate for this stimulus using `c3p10.m` with $\mathrm{r}_{\mathrm{est}}$ playing the role of the argument `spk`. Verify that `est` and `stim(ind)` are closely related, and describe the relationship between the coding kernel D and the decoding kernel `K`.

b) Repeat (a) with a rectification non-linearity, so that $[\mathrm{r}_{\mathrm{est}}]_+$ is used in place of $\mathrm{r}_{\mathrm{est}}$. Measure the effect of the nonlinearity by comparing the average (over time) of the squared difference between `est` and `stim(ind)`, divided by the variance of `stim`, for the rectified and nonrectified cases. What is the effect of rectification on the optimal decoding kernel `K`, and why? Assess the accuracy with which different frequency components in `stim` are captured in `est` by considering the power spectrum of the average squared difference between `est` and `stim(ind)`.

c) Generate a spike sequence `spk` from the rectified firing rate $[\mathrm{r}_{\mathrm{est}}]_+$ using a Poisson generator. The sequence `spk` should consist of a one or a zero at each time step, depending on whether or not a spike occurred. Recompute the acausal kernel as in part (a), but using `spk` as the response rather than $\mathrm{r}_{\mathrm{est}}$. How accurate is the resulting decoding, and what is the effect of using spikes rather than rates on the decoding kernel `K`?

d) What happens to decoding accuracy as the value of Δt, which defines the approximation to a white noise stimulus, increases and why? In the general case, the approximate white noise should be generated by choosing a stimulus value at each time step from a Gaussian distribution with mean 0 and variance 20 ms/Δt.

e) Attempt to repeat the decoding in (a) using the cross-correlation function `xcorr` and the fast Fourier transform `fft` to solve equation 3.54. Why is the answer so noisy?

Exercises

Chapter 4

1. Show that the firing-rate distribution that maximizes the entropy when the firing rate is constrained to lie in the range $0 \leq r \leq r_{\max}$ is given by equation 4.22, and that its entropy for a fixed resolution Δr is given by equation 4.23. Use a Lagrange multiplier (see the Mathematical Appendix) to constrain the integral of $p[r]$ to one.

2. Show that the firing-rate distribution that maximizes the entropy when the mean of the firing rate is held fixed is an exponential, and compute its entropy for a fixed resolution Δr. Assume that the firing rate can fall anywhere in the range from 0 to ∞. Use Lagrange multipliers (see the Mathematical Appendix) to constrain the integral of $p[r]$ to 1 and the integral of $p[r]r$ to the fixed average firing rate.

3. Show that the distribution that maximizes the entropy when the mean and variance of the firing rate are held fixed is a Gaussian, and compute its entropy for a fixed resolution Δr. To simplify the mathematics, allow the firing rate to take any value between $-\infty$ and $+\infty$. Use Lagrange multipliers (see the Mathematical Appendix) to constrain the integral of $p[r]$ to 1, the integral of $p[r]r$ to the fixed average firing rate $\langle r \rangle$, and the integral of $p[r](r - \langle r \rangle)^2$ to the fixed variance.

4. Using Fourier transforms, solve equation 4.37, using equation 4.36, to obtain the result of equation 4.42.

5. Suppose the filter $L_{\rm s}(\vec{a})$ has a correlation function that satisfies equation 4.37. Consider a new filter constructed in terms of this old one by writing
$$L'_{\rm s}(\vec{a}) = \int d\vec{c}\, U(\vec{a},\vec{c}) L_{\rm s}(\vec{c}) \, . \tag{1}$$
Show that if $U(\vec{a},\vec{c})$ satisfies the condition of an orthogonal transformation,
$$\int d\vec{c}\, U(\vec{a},\vec{c}) U(\vec{b},\vec{c}) = \delta(\vec{a} - \vec{b}) \, , \tag{2}$$
the correlation function for this new filter also satisfies equation 4.37.

6. Consider a stimulus $s_{\rm r} = s_{\rm s} + \eta$ that is given by the sum of a true stimulus $s_{\rm s}$ and a noise term η. Values of the true stimulus $s_{\rm s}$ are drawn from a Gaussian distribution with mean 0 and variance $Q_{\rm ss}$. Values of the noise term η are also obtained from a Gaussian distribution, with mean 0 and variance $Q_{\eta\eta}$. The two terms η and $s_{\rm s}$ are independent of each other. Using the formula for the continuous entropy of a Gaussian random variable calculated in problem 3, calculate the mutual information between $s_{\rm r}$ and $s_{\rm s}$.

7. Consider a multivariate signal $\mathbf{s}_{\mathrm{s}}$ drawn from a Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $\mathbf{Q}_{\mathrm{ss}}$. Compute the continuous entropy of $\mathbf{s}$ in terms of the eigenvalues of $\mathbf{Q}_{\mathrm{ss}}$, up to the usual resolution term for a continuous entropy.

8. Suppose that a stimulus at one point on the retina, and at a given time, $s_{\mathrm{r}} = s_{\mathrm{s}} + \eta$, is the sum of a true stimulus s_{s} and a noise term η, as in exercise 6. Model the retinal processing at this particular location as producing a signal at the thalamus

$$s_{\mathrm{l}} = D_{\mathrm{s}} s_{\mathrm{r}} + \eta_{\mathrm{l}} ,$$

where D_{s} is a parameter called the transfer constant, and η_{l} represents an additional, independent source of noise that can be modeled as being drawn from a Gaussian distribution with mean 0 and variance $Q_{\eta_{\mathrm{l}}\eta_{\mathrm{l}}}$. Calculate the mutual information I_{l} between s_{l} and s_{s} as a function of D_{s}. The power of the signal produced by the retina is defined as $P_{\mathrm{r}} = \langle (D_{\mathrm{s}} s_{\mathrm{r}})^2 \rangle$. By maximizing

$$I_{\mathrm{l}} - kP_{\mathrm{r}}$$

as a function of D_{s}, find the transfer constant that maximizes the mutual information for a given value of k (with $k > 0$), a parameter that controls the trade-off between information and power. What happens when Q_{ss}, describing the visual signal, gets much smaller than $Q_{\eta\eta}$? (Based on a problem from Dawei Dong.)

9. Consider two independent inputs s and s' drawn from Gaussian distributions with means 0 and with different variances Q_{ss} and $Q_{\mathrm{s's'}}$. These generate two thalamic signals, as in exercise 8.

$$s_{\mathrm{l}} = D_{\mathrm{s}} s + \eta \quad and \quad s'_{\mathrm{l}} = D_{\mathrm{s'}} s' + \eta' ,$$

defined by two separate transfer constants, D_{s} and $D_{\mathrm{s'}}$, and two independent noise terms with variances $Q_{\eta\eta}$ and $Q_{\eta'\eta'}$. Find the transfer constants that maximize the total mutual information $I_{\mathrm{l}} + I'_{\mathrm{l}}$ for a fixed total power $P_{\mathrm{r}} + P'_{\mathrm{r}}$, where the non-primes and primes denote the information and power for s_{l} and s'_{l}, respectively.

Exercises

Chapter 5

1. The Nernst equation (equation 5.4) was derived in chapter 5 under the assumption that the membrane potential was negative and the ion being considered was positively charged. Rederive this result for a negatively charged ion and for the case when E is positive to verify that it applies in all these cases.

2. Verify that equation 5.47 is a solution of equation 5.46 when V_∞ is independent of time. Then, solve equation 5.46 for the case when V_∞ is an arbitrary function of time. In this solution, $V(t)$ is expressed in terms of integrals involving $V_\infty(t)$.

3. Build a model integrate-and-fire neuron from equation 5.8. Use $V_{\text{rest}} = -70$ mV, $R_{\text{m}} = 10$ MΩ, and $\tau_{\text{m}} = 10$ ms. Initially set $V = V_{\text{rest}}$. When the membrane potential reaches $V_{\text{th}} = -54$ mV, make the neuron fire a spike and reset the potential to $V_{\text{reset}} = -80$ mV. Show sample voltage traces (with spikes) for a 300-ms-long current pulse (choose a reasonable current I_{e}) centered in a 500-ms-long simulation. Determine the firing rate of the model for various magnitudes of constant I_{e} and compare the results with equation 5.11.

4. Include an extra current in the integrate-and-fire model to introduce spike-rate adaptation, as described in equations 5.13 and 5.14, and in the caption to figure 5.6.

5. Add an excitatory synaptic conductance to the integrate-and-fire neuron of exercise 3 by adding the extra synaptic conductance term in equation 5.43 with $E_{\text{s}} = 0$. Set the external current to zero, $I_{\text{e}} = 0$, in this example, and assume that the probability of release on receipt of a presynaptic spike is 1. Use $r_{\text{m}}\overline{g}_{\text{s}} = 0.5$ and describe P_{s} using the alpha function of equation 5.35 with $\tau_{\text{s}} = 10$ ms and $P_{\text{max}} = 0.5$. To incorporate multiple presynaptic spikes, P_{s} should be described by a pair of differential equations,

$$\tau_{\text{s}} \frac{dP_{\text{s}}}{dt} = eP_{\text{max}}z - P_{\text{s}}$$

   with $e = \exp(1)$, and

$$\tau_{\text{s}} \frac{dz}{dt} = -z\,,$$

   with the additional rule that z is set to 1 whenever a presynaptic spike arrives. Plot V(t) in one graph and the synaptic current in another. Trigger synaptic events at times 50, 150, 190, 300, 320, 400, and 410 ms. Explain what you see.

6. The equations in exercise 5 generate an α function response to a single input spike, but they do not prevent P_s from growing greater than 1 when the model synapse is driven by multiple spikes at a sufficiently

high frequency. In other words, this model synapse does not saturate. To introduce saturation, modify the equations of exercise 5 to

$$\tau_s \frac{dP_s}{dt} = eP_{max} z(1 - P_s) - P_s$$

with $e = \exp(1)$, and

$$\tau_s \frac{dz}{dt} = -z\,,$$

with the additional rule that z is set to 1 whenever a presynaptic spike arrives. Compare $P_s(t)$ computed using these equations with $P_s(t)$ computed using the equations of exercise 5 for constant rate, regular (periodic) presynaptic spike trains with frequencies ranging from 1 to 100 Hz. In both cases, use $\tau_s = 10$ ms and $P_{max} = 0.5$.

7. Construct a model of two coupled integrate-and-fire neurons similar to that of figure 5.20. Both model neurons obey equation 5.43 with $E_L = -70$ mV, $V_{th} = -54$ mV, $V_{reset} = -80$ mV, $\tau_m = 20$ ms, $r_m \overline{g}_s = 0.15$, and $R_m I_e = 18$ mV. Both synapses should be described as in exercise 5 with $P_{max} = 0.5$ and $\tau_s = 10$ ms. Consider cases where both synapses are excitatory, with $E_s = 0$ mV, and both are inhibitory, with $E_s = -80$ mV. Show how the pattern of firing for the two neurons depends on the type (excitatory or inhibitory) of the reciprocal synaptic connections. For these simulations, set the initial membrane voltages of the two neurons to slightly different values, randomly, and run the simulation until an equilibrium situation has been reached, which may take a few seconds of simulated run time. Start from a few different random initial conditions to study whether the results are consistent. Investigate what happens if you change the strengths and time constants of the reciprocal synapses.

8. Build a Hodgkin-Huxley model neuron by numerically integrating the equations for V, m, h, and n given in chapter 5 (see, in particular equations 5.6, 5.17–5.19, 5.22, 5.24, and 5.25). Take $c_m = 10$ nF/mm^2, and as initial values take: $V = -65$ mV, $m = 0.0529$, $h = 0.5961$, and $n = 0.3177$. Use an integration time step of 0.1 ms. Use an external current with $I_e/A = 200$ nA/mm^2 and plot V, m, h, and n as functions of time for a suitable interval. Also, plot the firing rate of the model as a function of I_e/A over the range from 0 to 500 nA/mm^2. Show that the firing rate jumps discontinuously from zero to a finite value when the current passes through the minimum value required to produce sustained firing. Finally, apply a pulse of negative current with $I_e/A = -50$ nA/mm^2 for 5 ms followed by $I_e/A = 0$ and show what happens.

9. Construct and simulate the K$^+$ channel model of figure 5.12. Plot the mean squared deviation between the current produced by N such model channels and the Hodgkin-Huxley current as a function of N over the range $1 \leq N \leq 100$, matching the amplitude of the Hodgkin-Huxley model so that the mean currents are the same.

10. Compute analytically the value of the release probability P_{rel} just before the time of each presynaptic spike for a regular (periodic rather than Poisson), constant-frequency presynaptic spike train as a function of the presynaptic firing rate. Do this for both the depression and facilitation models described by equation 5.37.

Exercises

Chapter 6

1. Build a Connor-Stevens model neuron by numerically integrating the equations for V, m, h, n, a, and b given in chapter 6 (see, in particular, equations 6.1, 6.4, and appendix A). Use $c_{\mathrm{m}} = 10$ nF/mm^2, and as initial values take: $V = -68$ mV, $m = 0.0101$, $h = 0.9659$, $n = 0.1559$, $a = 0.5404$, and $b = 0.2887$. Use an integration time step of 0.1 ms. Use an external current with $I_{\mathrm{e}}/A = 200$ nA/mm^2 and plot V, m, h, n, a, and b as functions of time over a suitable interval. Plot the firing rate of the model as a function of I_{e}/A over the range from 0 to 500 nA/mm^2. How does this differ from what you got for the Hodgkin-Huxley model in exercise 8 of chapter 5. Finally, apply a pulse of negative current with $I_{\mathrm{e}}/A = -500$ nA/mm^2 for 5 ms followed by $I_{\mathrm{e}}/A = 200$ nA/mm^{2}and show what happens.

2. Construct a Morris-Lecar model neuron (Morris, C & Lecar, H (1981) Voltage oscillations in the barnacle giant muscle fiber. *Biophysical Journal* **35**:193–213). Instead of simulating the fast sodium spikes of an action potential, this model describes slower calcium spikes. The model has just two active currents, an instantaneous voltage-dependent Ca^{2+} current and a persistent K$^+$ current, described by a single dynamical gating variable N:

   $$i_{\mathrm{m}} = \overline{g}_L(V - E_{\mathrm{L}}) + \overline{g}_{\mathrm{Ca}} M_\infty(V)(V - E_{\mathrm{Ca}}) + \overline{g}_K N(V - E_{\mathrm{K}})$$

   with $\overline{g}_L = 0.005$ mS/mm^2, $\overline{g}_{\mathrm{Ca}} = 0.01$ mS/mm^2 and $\overline{g}_{\mathrm{K}} = 0.02$ mS/mm^2, $E_{\mathrm{L}} = -50$ mV, $E_{\mathrm{Ca}} = 100$ mV and $E_{\mathrm{K}} = -70$ mV. Use $c_{\mathrm{m}} = 10$ nF/mm^2. The function $M_\infty(V)$ is given by

   $$M_\infty(V) = \frac{1}{1 + \exp[-.133(V+1)]}$$

   and the gating variable N is given by

   $$\tau_N(V)\frac{dN}{dt} = N_\infty(V) - N$$

   with

   $$\tau_N(V) = \frac{3}{\cosh[.0345(V-10)]}$$

   and

   $$N_\infty(V) = \frac{1}{1 + \exp[-.138(V-10)]} .$$

   Here, V is understood to be in mV units, and τ_N is expressed in ms units. Determine the firing rate as a function of injected current and plot the membrane potential and N as a functions of time. Also, show a phase-plane trajectory, which is a plot of that path taken by these variables in the two-dimensional space described by the points (V,

N), while the model is firing. In the phase plane, plot the nullclines for the V and N equations. These are lines in the V-N plane along which either $dV/dt = 0$ or $dN/dt = 0$. (Phase-plane descriptions and nullclines are described in chapter 7.)

3. The FitzHugh-Nagumo equations (see FitzHugh, R (1961) Impulses and physiological states in models of nerve membrane. *Biophysical Journal* **1**:445–466) are given by

$$\frac{dv}{dt} = v(1 - v^2) - u + I_e \quad \text{and} \quad \frac{du}{dt} = \epsilon(v - 0.5u)$$

   Draw the nullclines for these equations for $I_e = 0$ and $I_e = -1$. These are the lines in the v-u plane where the right side of one or the other of these two equations is zero. In which case or cases do you think the model will produce oscillations? Next simulate the model to see what happens when these equations are integrated over time. Determine what happens for $I_e = 0$ with $\epsilon = 0.3$, 0.1, and 1 and for $I_e = -1$ with $\epsilon = 0.3$. (Phase-plane descriptions and nullclines are described in chapter 7.)

4. Show that solution of equation 6.19 satisfies the cable equation along an infinite cable in response to the injected current $i_e = I_e \tau_m \delta(x)\delta(t)/(2\pi a)$.

5. Verify that the solution for an isolated junction given by equations 6.21 and 6.22 satisfies the correct boundary conditions at the junction point: $v_1(0) = v_2(0) = v_3(0)$ and

$$\sum_{i=1}^{3} a_i^2 \left.\frac{\partial v_i}{\partial x}\right|_{x=0} = 0\,.$$

6. Generalize the solution for an isolated junction of equation 6.21 to the time-dependent case when the injected current on segment 2 is $i_e = I_e \tau_m \delta(x_2 - y)\delta(t)/(2\pi a)$.

7. Show that the expression for $v(x)$ given in figure 6.10, with R_1 and R_2 given by equations 6.23 and 6.24, satisfies the cable equation and the boundary conditions, $v(0) = v_{\text{soma}}$ and $\partial v/\partial x = 0$ when $x = L$.

8. Show that the expression for $v(x)$ given in figure 6.12, with R_3 and R_4 given by equations 6.26 and 6.27, satisfies the cable equation and the boundary conditions, $v(0) = 0$ and $\partial v/\partial x = 0$ when $x = L$.

9. Construct a non-branching axonal cable with conductances in each compartment described by the Connor-Stevens model (as in exercise 1). Solve for the membrane potential using the methods of appendix B of chapter 6. Initiate action potential propagation at one end of the cable by injecting current into the terminal compartment of the cable. Plot the action potential propagation velocity as a function of the axon radius. Inject current into the middle of the cable to generate two, opposite-moving action potentials. Generate action

potentials from each end of the cable and show that they annihilate each other when they collide.

10. Determine the numerical solution for a multi-compartment cable with a single branching node (where a single cable splits into two branches) analogous to the solution for a non-branching cable (equations 6.53–6.56) given in appendix B of chapter 6.

Exercises

Chapter 7

1. a) Consider network activities $v(\theta)$ that are steady-state solutions of equation 7.36, satisfying

$$v(\theta) = \left[h(\theta) + \int_{-\pi/2}^{\pi/2} \frac{d\theta'}{\pi} \left(-\lambda_0 + \lambda_1 \cos(2(\theta - \theta'))\right) v(\theta')\right]_+ , \quad (1)$$

in response to input $h(\theta) = Ac(1 - \epsilon + \epsilon \cos(2\theta))$ as in equation 7.37. Assuming that $v(\theta)$ is symmetric about $\theta = 0$, show that $v(\theta)$ takes either the form

$$v(\theta) = \alpha \left[\cos(2\theta) - \cos(2\theta_C)\right]_+ \quad (2)$$

or the form

$$v(\theta) = \alpha \cos(2\theta) + v_0 \, . \quad (3)$$

In the case of equation 2, which applies when $\theta_C < \pi/2$ and for which θ_C defines the width of the orientation tuning curve, by calculating the integral

$$\int_{-\pi/2}^{\pi/2} \frac{d\theta'}{\pi}(-\lambda_0 + \lambda_1 \cos(2(\theta - \theta'))v(\theta') \, ,$$

show that α and θ_C must satisfy the consistency conditions

$$\begin{aligned} \alpha &= \frac{Ac\epsilon}{1 - \lambda_1 \left(\theta_C - \sin(4\theta_C)/4\right)/\pi} \\ \cos(2\theta_C) &= \frac{\lambda_0}{\pi}(\sin(2\theta_C) - 2\theta_C \cos(2\theta_C)) - \\ &\quad \frac{(1-\epsilon)}{\epsilon}\left(1 - \frac{\lambda_1}{\pi}\left(\theta_C - \frac{\sin(4\theta_C)}{4}\right)\right) . \end{aligned} \quad (4)$$

b) In the case of equation 3, calculate α and v_0.

c) For values $\lambda_0 = 7.3$, $\lambda_1 = 11$, $c = 1$, and $A = 40$ Hz, use the MATLAB® function `fzero` to find the value of θ_C that satisfies the consistency condition in equation 4 as a function of ϵ for $0 < \epsilon \leq 1$. For $\epsilon = 0.1$ and $c = 0.1, 0.2, 0.4$, and 0.8, solve for α, and thereby reproduce figure 7.10B. Repeat the plots for $\lambda_1 = 0$. At what value of ϵ does θ_C fall below $\pi/2$. This corresponds to a model in which feedforward orientation tuning is sharpened only by inhibition. [Corrected by Sune Jespersen]

d) Numerically integrate equation 7.36 for the sets of parameters in (c) to confirm your results. Use 100 neurons with preferred values evenly spaced between $-\pi/2$ and $\pi/2$.

e) Plot $\theta_C - \sin(4\theta_C)/4$ for $0 \leq \theta_C \leq \pi/2$. What is its maximum value? As $\epsilon \to 0$ (so that $(1-\epsilon)/\epsilon \to \infty$), calculate (from equation 4) a condition on λ_1 that ensures there will always be a solution with $\theta_C < \pi/2$. This defines a marginal phase in which the recurrent connections create a tuned output even from untuned input, and it constitutes what is called a continuous attractor.

2. A Hopfield associative memory network has activities for individual units, v_a for $a = 1, 2, \ldots, N$ (or collectively $\mathbf{v}$), that take values of either $+1$ or -1, and are updated at every discrete time step of the network dynamics by the rule

$$v_a(t+1) = \text{sgn}\left(\sum_{a'=1}^{N} M_{aa'} v_{a'}(t)\right), \tag{5}$$

where

$$\text{sgn}(z) = \begin{cases} +1 & \text{if} \quad z \geq 0 \\ -1 & \text{if} \quad z < 0 \,. \end{cases}$$

Here $\mathbf{M}$ is a matrix constructed from P "memory" vectors $\mathbf{v}^m$ ($m = 1, 2, \ldots P$), also composed of elements that are either $+1$ or -1, through the sum of outer products

$$M_{aa'} = (1 - \delta_{aa'}) \sum_{m=1}^{P} v_a^m v_{a'}^m \,. \tag{6}$$

Note that the diagonal elements of $\mathbf{M}$ are set to zero by this equation. Consider a 100-element network ($N = 100$). Construct P memory states by randomly assigning $+1$ and -1 values with equal probabilities to the N elements of each $\mathbf{v}^m$. Using these memory vectors, set the matrix of synaptic weights according to equation 6. Then, study the behavior of the network by iterating equation 5. To measure how close the state of the network at time t, $\mathbf{v}(t)$, is to a particular memory state, define the overlap function $q(t) = \mathbf{v}(t) \cdot \mathbf{v}^m / N$. This is equal to 1 if $\mathbf{v}(t) = \mathbf{v}^m$, is near zero if $\mathbf{v}(t)$ is unrelated to $\mathbf{v}^m$, and is equal to -1 if $\mathbf{v}(t) = -\mathbf{v}^m$. Set the initial state $\mathbf{v}(0)$ so that it has a positive overlap, $q(0)$, with memory state $\mathbf{v}^1$. Plot $q(t)$ as the network evolves from this state according to equation 5. Final values of $q(t)$ near one indicate successful recovery of the memory. Do the same starting from $\mathbf{v}(0)$ close to the inverse of the memory state $-\mathbf{v}^1$. What accounts for this behavior? Determine the range of $q(0)$ values (about $\mathbf{v}^1$) that assures successful memory recovery for different values of P. Start with $P = 1$ and increase it until memory recovery fails even for $q(0) = 1$. At what P value does this occur?

3. Repeat exercise 2 with the matrix $\mathbf{M}$ replaced by

$$M_{aa'} = (1 - \delta_{aa'}) \sum_{m,m'=1}^{P} v_a^m C_{mm'} v_{a'}^{m'} \,,$$

where $C_{mm'}$ is the m, m' element of the inverse of the matrix

$$\sum_{a=1}^{N} v_a^m v_a^{m'} \,.$$

Compare the performance and capacity of the associative memory constructed using this matrix with that of the associative memory in exercise 2.

4. Build and study a simple model of oscillations arising from the interaction of excitatory and inhibitory populations of neurons. The firing rate of the excitatory neurons is r_E, and that of the inhibitory neurons is r_I and these are characterized by equations 7.50 and 7.51. Set $M_\mathrm{EE} = 1.25$, $M_\mathrm{IE} = 1$, $M_\mathrm{II} = 0$, $M_\mathrm{EI} = -1$, $\gamma_\mathrm{E} = -10$ Hz, $\gamma_\mathrm{I} = 10$ Hz, $\tau_\mathrm{E} = 10$ ms, and vary the value of τ_I. The negative value of γ_E means that this parameter serves as a source of background activity (activity in the absence of excitatory input) rather than as a threshold. Show what happens for $\tau_\mathrm{I} = 30$ ms and for $\tau_\mathrm{I} = 50$ ms. Find the value of τ_I for which there is a transition between fixed-point and oscillatory behavior, thereby verifying the results obtained analytically in chapter 7 on the basis of equation 7.53.

5. MATLAB® files `c7p5.m` and `c7p5sub.m` perform a numerical integration of a two-unit, nonlinear, symmetric recurrent network with a threshold linear activation function $F(I) = \beta[I]_+$ and plot the results. Here, the dynamics come from

$$\frac{d\mathbf{v}}{dt} = -\mathbf{v} + \mathbf{F}(\mathbf{M}\cdot\mathbf{v} + \mathbf{h})$$

with $\mathbf{v} = (v_1, v_2)$ and $h_1 = h_2 = 1$. The weight matrix in this example is $\mathbf{M} = [0\ {-1}\ ;\ {-1}\ 0]$, which tends to make v_1 and v_2 compete. Execute `c7p5.m` to see the consequences of regimes of high ($\beta = 2$) and low ($\beta = 0.5$) activation (which is equivalent to large and small recurrent weights). For these two values of β, plot the nullclines (the locations in the v_1-v_2 phase plane where $dv_1/dt = 0$ and $dv_2/dt = 0$). You should find one fixed point for $\beta = 0.5$ and three for $\beta = 2$. Linearize the network about the fixed point with $v_1 = v_2$ and derive a condition on β for this fixed point to be stable. (Based on a problem from Dawei Dong.)

6. Plot the results of exercise 5 for the inputs $\mathbf{h} = (0.75, 1.25)$ and $\mathbf{h} = (0.5, 1.5)$. By plotting nullclines for these values of $\mathbf{h}$, explain the resulting behavior. (Based on a problem from Dawei Dong.)

7. Use the expression

$$f_u(s-\xi, g-\gamma) = A\exp\left(-\frac{(s-\xi)^2}{2\sigma_s^2}\right) N\left(\frac{g-\gamma}{\sigma_g}\right),$$

where A, ξ, σ_s, γ, and σ_g are parameters and N is the (sigmoidal) cumulative normal function

$$N(z) = \int_{-\infty}^{z} dx\, \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) = 1 - \frac{1}{2}\mathrm{erfc}\left(\frac{z}{\sqrt{2}}\right).$$

Plot $f_u(s - \xi, g - \gamma)$ and find values of the parameters that make it roughly match the gain-modulated response of figure 7.6B. Using $w(\xi, \gamma) = \exp(-(\xi + \gamma)^2/2\sigma_w^2)$, evaluate the integral in equation 7.15 in terms of a single cumulative normal function to show that the resulting tuning curves are functions of $s + g$, and assess how the tuning width depends on σ_s, σ_g and σ_w.

Exercises

Chapter 8

1. Simulate the course of Hebbian learning for the case of figure 8.3. Find the ranges of initial weight values, (w_1, w_2), that lead to saturation at $(1, 1)$. Can you predict the result analytically? If the off-diagonal term in the correlation matrix is -2 instead of -0.4 and there are no saturation boundaries, what happens to the sum of the weights? Could this be used as a way of normalizing the weights?

2. Show that the averaged form of the single-trial Oja rule in equation 8.16 is given by

   $$\tau_w \frac{d\mathbf{w}}{dt} = \mathbf{Q} \cdot \mathbf{w} - \alpha(\mathbf{w} \cdot \mathbf{Q} \cdot \mathbf{w})\mathbf{w} \, .$$

   Prove that if it converges, the averaged learning rule produces a set of weights proportional to an eigenvector of the correlation matrix $\mathbf{Q}$, normalized so that $|\mathbf{w}|^2 = 1/\alpha$.

3. Simulate the ocular dominance model of figure 8.7 using a subtractively normalized version of equation 8.31 (i.e. equation 8.14) with saturation limits at 0 and 1, and cortical interactions generated as in figure 8.8 from

   $$\mathbf{K}_{aa'} = \exp\left(-\frac{(a-a')^2}{2\sigma^2}\right) - \frac{1}{9}\exp\left(-\frac{(a-a')^2}{18\sigma^2}\right) ,$$

   where $\sigma = 0.066$ mm. Use 512 cortical cells with locations a spread evenly over a nominal 10 mm of cortex, and periodic boundary conditions (this means that you can use Fourier transforms to calculate the effect of the cortical interactions). Also use the discrete form of equation 8.31

   $$\mathbf{W} \rightarrow \mathbf{W} + \epsilon \mathbf{K} \cdot \mathbf{W} \cdot \mathbf{Q}$$

   with a learning rate of $\epsilon = 0.01$. Plot $\mathbf{w}_-$ as it evolves from near $\mathbf{0}$ to the final form of ocular dominance. Calculate the magnitude of the discrete Fourier transform of $\mathbf{w}_-$. Repeat this around 100 times, work out the average of the magnitudes of the Fourier transforms, and compare this to the Fourier transform of $\mathbf{K}$.

4. Construct two-dimensional input data sets similar to those shown in figure 8.4 and use them to train a two-input, one output linear network using correlation- and covariance-based Hebbian learning rules with multiplicative normalization. Compare the final outcome for the weights with the principal components of the data when the mean of the input distribution is zero and when it is nonzero.

5. Repeat exercise 4 for a data set with zero mean, but with subtractive normalization and saturation. Start with initial values for the weights that are chosen randomly over the full range from 0 to their saturation

limit. When does this algorithm produce a weight vector aligned with the principal component axis of the input data set, and when does it fail to do so. Why does the weight vector sometimes fail to align with the principal component axis?

6. Consider minimizing the function $E(w) = (w-2)^2$ using the gradient descent rule for w,
$$w \rightarrow w - \epsilon \frac{dE}{dw} \, .$$
Plot $E(w)$ together with the trajectories of w starting from $w = 5$ for $\epsilon = 0.01, 0.1, 1, 2, 3$. Why does learning diverge as ϵ gets large?

7. Consider $E(\mathbf{w}) \propto \left\langle (h(s) - \mathbf{w} \cdot \mathbf{f}(s))^2 \right\rangle$, as in equation 8.52, in the case that matrix $\langle \mathbf{f}(s)\mathbf{f}(s) \rangle$ is invertible. An extended delta rule can be written as
$$\mathbf{w} \rightarrow \mathbf{w} + \langle (h(s) - \mathbf{w} \cdot \mathbf{f}(s))\mathbf{H} \cdot \mathbf{f}(s) \rangle \, ,$$
where $\mathbf{H}$ is a matrix that generalizes the learning rate ϵ of the standard delta rule. For what matrix $\mathbf{H}$ does this rule go from any initial value $\mathbf{w}$ to the optimal weights in one single step. This amounts to a form of the Newton-Raphson method.

8. Train the feedforward network of figure 8.13 to produce the output $v = \cos(0.6s)$ when the input tuning curves are given as in the caption to figure 8.14. Train the network by using the stochastic delta learning rule (equation 8.61) with s values chosen randomly in the range between -10 and 10.

9. Construct a perceptron (equation 8.46) that classifies 10 binary inputs according to whether their sum $\sum u_a$ is positive or negative. Use a random set of binary inputs during training and compare the performance (both the learning rate and the final accuracy) of the Hebbian (equation 8.47), delta, and perceptron learning rules. Repeat this training protocol, but this time attempt to make the output of the perceptron classify according to the parity of the inputs, which is the sign of their product $\prod u_a$. Why is this example so much harder than the first case?

Exercises

Chapter 9

1. Implement acquisition and extinction as in figure 9.1 using the Rescorla-Wagner (delta) rule (equation 9.2).

2. Add a second stimulus and demonstrate that the delta rule can describe blocking, but that it fails to exhibit secondary conditioning.

3. Consider the case of partial reinforcement (studied in figure 9.1) in which reward $r = 1$ is provided randomly with probability p on any given trial. Assume that there is a single stimulus with $u = 1$, so that $\epsilon\delta u$, with $\delta = r - v = r - wu$, is equal to $\epsilon(r - w)$. By considering the expected value $\langle w + \epsilon(r - w)\rangle$ and the expected square value $\langle (w + \epsilon(r - w))^2\rangle$ of the new weights, calculate the self-consistent equilibrium values of the mean and variance of the weight w. What happens to your expression for the variance if $\epsilon = 2$ or $\epsilon > 2$? To what features of the learning rule do these effects correspond?

4. The original application of temporal difference learning to conditioning (Sutton & Barto, 1990) considered the use of stimulus traces (as a preliminary to the linear filter of equation 9.5). That is, the prediction of sum future reward at time t is $v(t) = \mathbf{w} \cdot \mathbf{u}(t)$ where $u_i(t)$, with prediction weight is w_i, marks the presence (when $u_i(t) = 1$) or absence (when $u_i(t) = 0$) of stimulus i at time t. Also, the temporal difference learning rule of equation 9.10 is replaced by

   $$w_i \rightarrow w_i + \epsilon\delta(t)\bar{u}_i(t)\,,$$

   where
   $$\bar{u}_i(t) = \lambda\bar{u}_i(t-1) + (1-\lambda)u_i(t)$$

   is the stimulus trace for stimulus i, and $\delta(t)$ is as in equation 9.10. Here λ is the trace parameter which governs the length of the memory of the past occurrence of stimuli (see equation 9.30). Construct a trace learning model for a case similar to that of figure 9.2, but taking $r(t)$ to be the hat-function $r(t) = 1/5, 200 \leq t \leq 210$ and $r(t) = 0$ otherwise. Note that to match figure 9.2, you must use $\Delta t = 5$ for each time step rather than $\Delta t = 1$. Show the signals as in figure 9.2B for $\lambda = 0.5, 0.9, 0.99$, using $\epsilon = 0.2$. Could this model account for the data on the activity of the dopamine cells? Would it show secondary conditioning?

5. Use the prediction model of equation 9.5 and the standard temporal difference learning rule of equation 9.10 to reproduce figure 9.2. Take $r(t)$ to be the hat-function $r(t) = 1/5, 200 \leq t \leq 210$ and $r(t) = 0$ otherwise. In this figure, the increments of time are in steps of $\Delta t = 5$, and $\epsilon = 0.4$. Consider what happens if the time between the stimulus and the reward is stochastic, drawn from a uniform distribution between 50 and 150. Show the average prediction error signal $\delta(t)$

time-locked to the stimulus and the reward. How does this differ from those in figure 9.2.

6. Implement a stochastic three-armed bandit using the indirect actor and the action choice softmax rule 9.12. Let arm a produce a reward of p_a, with $p_1 = 1/4, p_2 = 1/2, p_3 = 3/4$, and use a learning rate of $\epsilon = 0.01, 0.1, 0.5$ and $\beta = 1, 10, 100$. Consider what happens if after every 250 trials, the arms swap their reward probabilities at random. Averaging over a long run, explore to see which values of ϵ and β lead to the greatest cumulative reward. Can you account for this behavior?

7. Repeat exercise 6 using the direct actor (with learning rule 9.22). For $\bar{r}$, use a low-pass filtered version of the actual reward, which is obtained by using the update rule

   $$\bar{r} \rightarrow \lambda \bar{r} + (1 - \lambda) r$$

   with $\lambda = 0.95$. Study the effect of the different values of ϵ and β in controlling the average rate of rewards when the arms swap their reward probabilities at random every 250 trials.

8. Implement actor critic learning (equations 9.24 and 9.25) in the maze of figure 9.7, with learning rate $\epsilon = 0.5$ for both actor and critic, and $\beta = 1$ for the critic. Starting from zero weights for both the actor and critic, plot learning curves as in figures 9.8 and 9.9. Start instead from a policy in which the agent is biased to go left at both B and C, with initial probability 0.99. How does this affect learning at A?

9. Implement actor critic learning for the maze, as in exercise 8, except using vectorial state representations as in equations 9.26, 9.27, and 9.28. If $\mathbf{u}(\mathrm{A}) = (1, 0, 0), \mathbf{u}(\mathrm{B}) = (0, 1, 0)$ and $\mathbf{u}(\mathrm{C}) = (0, 0, 1)$, then the result should be exactly as in exercise 8. What happens to the speed of leaning if $\mathbf{u}(\mathrm{A}) = (1, a, a)$ (while retaining $\mathbf{u}(\mathrm{B}) = (0, 1, 0)$ and $\mathbf{u}(\mathrm{C}) = (0, 0, 1)$) for $a = +0.5$ and $a = -0.5$, and why?

Exercises

Chapter 10

1. Data file `c10p1.mat` contains 100 data points drawn from the same two-dimensional distribution as those in figure 10.1. Fit a mixture of two circular Gaussian distributions to these data using EM, as in equation 10.4. Do not allow the variance of either of the Gaussians to become smaller than a minimal value of 0.0001.

2. Explore what happens to the fit of the mixture of Gaussians model from exercise 1 as the number of data points from each Gaussian is reduced and the number of potential Gaussians is increased. If you set the minimal variance given in exercise 1 to 0, a Gaussian distribution can settle around a single sample point and then have its variance shrink to 0. Why does this pathological behavior occur?

3. Modify your code from exercise 1 to calculate function $\mathcal{F}$ of equation 10.14 during each E and M step of EM. Check that $\mathcal{F}$ changes monotonically. Explicitly calculate the true log likelihood of the data from equation 10.7 at the end of each M phase. Is it equal to $\mathcal{F}$?

4. Modify the code in exercise 1 to fit a K-means model rather than a mixture of Gaussians. Can you see any practical differences in the solutions that arise?

5. Consider the factor analysis model of figure 10.3 (discussed in more generality later in the chapter). Using the joint probability over v and $\mathbf{u}$ given in equation 10.15, derive an expression for $\mathcal{F}$, and thus the learning rules of equation 10.5.

6. Using the EM version of factor analysis (see the appendix of chapter 10), reproduce figure 10.4. MATLAB® file `c10p6.m` shows how to generate data `u1` for figure 10.4A and B, and `u2` for C and D. First perform factor analysis on these data and reproduce figures 10.4A and C. Next, use the `eig` function to perform principal components analysis on `u1` and `u2`, and thereby produce the rest of the figures. For some initial conditions, the cloud of points in the figures might slope downwards instead of upwards. Why? Calculate the expression for $\mathcal{F}$ derived in exercise 5 as factor analysis progresses and show that it changes monotonically.

7. Apply a rotation matrix to the data set `u2` from exercise 6 (an example rotation matrix is given as `rot` in MATLAB® file `c10p6.m`). Perform factor analysis and principal components analysis on the rotated data. How do the results compare with those for the unrotated data (remember to rotate your results back, if necessary, so that appropriate comparisons can be made)?

8. Construct a data set $\mathbf{u}$ from a set of independent, heavy-tailed, "sources" $\mathbf{v}$ through the relation $\mathbf{u} = \mathbf{G} \cdot \mathbf{v}$. Both $\mathbf{u}$ and $\mathbf{v}$ should

be four-dimensional vectors. Choose the components of $\mathbf{v}$ independently and randomly from a double exponential distribution, for which the probability of getting the value v is proportional to $\exp(-|v|)$ (note that a one-sided exponentially distributed random variable can be generated using either `exprnd(·)` or `-log(rand(·))`). Choose a random matrix $\mathbf{G}$ and generate the corresponding $\mathbf{u}$ values as $\mathbf{u} = \mathbf{G} \cdot \mathbf{v}$. Use 2000 randomly chosen $\mathbf{v}$'s and their corresponding $\mathbf{u}'s$. Then, use independent components analysis, as in equation 10.40, to learn generative sources from the inputs $\mathbf{u}$. How well do the values of the extracted sources match those of the original sources?

9. Compare the actual $\mathbf{G}$ you used to generate the data in problem 8 with the $\mathbf{G}$ that is recovered by independent components analysis. Plot the six two-dimensional projections of the input data (u_1 versus u_2; u_1 versus u_3; *etc*) together with the projections of the mixing axes coming from $\mathbf{G}$ (it is good to use more data points for this, say 10000). The mixing axes are lines parallel to vectors with components G_{1i}, G_{2i}, G_{3i}, and G_{4i}, for $i = 1, 2, 3, 4$. What relationship exists between these mixing axes and the envelope of the data points, and why? Plot $\mathbf{u}$ generated in the same way when the components of $\mathbf{v}$ are chosen independently from identical Gaussian distributions, together with the mixing axes coming from $\mathbf{G}$. What differences do you see?

10. Implement wake-sleep learning for the Helmholtz machine with binary units when the input data is derived from a square "retina" of size `ndim`×`ndim`. The `ndim` columns of the input array are independently turned "on" with probability `pbar`. Each unit in a column that is "on" takes the value 1 with probability 1−`pout` and 0 with probability `pout`, and each unit in a column that is "off" takes the value 1 with probability `pout` and 0 with probability 1−`pout`. MATLAB® program `c10p10.m` is an example. In what way does the activity of the $\mathbf{v}$ units in the model capture the way that each input $\mathbf{u}$ was actually generated? What happens if there are not enough hidden units to represent each column separately?

11. Implement wake-sleep learning for the binary Helmholtz machine as in problem 10, except now make a correlational structure between the columns – so that for half the input patterns, only columns 1 . . .`ndim`/2 are eligible to be turned on (with probability `pbar`), and for the other half, only the other columns `ndim`/2 + 1 . . .`ndim` are eligible. Program `c10p11.m` shows one way to generate such inputs. Train a Helmholtz machine with two representational layers ($\mathbf{v}$ and z), the top layer (z) having just one unit, the middle layer ($\mathbf{v}$) with `ndim`+1 units. Does this build a generative model that captures the hierarchical way in which each input pattern is generated?