Gatsby Computational Neuroscience Unit
Theoretical Neuroscience

Final examination, theoretical neuroscience
22 May 2025

Part II – long questions

There are four questions, one from each main section of the course. Please answer three out of the four, starting the answers for each question on a new page. Don't forget to write your name at the top of the answer to each question.

Good luck!

1 Biophysics

Consider the Izhikevich neuron, which is a quadratic integrate and fire (QIF) neuron coupled to a hyperpolarizing current,

$$\tau_v \frac{dv}{dt} = v^2 - u + w \tag{1a}$$

$$\tau_u \frac{du}{dt} = 2\beta v - u\,. \tag{1b}$$

When the voltage reaches threshold, taken to be at $v = v_\theta$, both the voltage and u are reset,

$$v \to 0 \tag{2a}$$

$$u \to u + \Delta u\,. \tag{2b}$$

The quantities v_θ, β and Δu are all constant and positive.

1. For this question, ignore the u-variable, by setting it permanently to 0 (which we can do by letting $\beta = \Delta u = 0$ and waiting for a long time). Show that as $w \to 0$ from above, the firing rate scales as $\sqrt{w}/\tau_v$. This corresponds to a type II neuron.

   (5 marks)

   Solution

   To compute the firing rate, we'll compute the time it takes to emit a spike (the time it takes to go from $v = 0$ to $v = v_\theta$), and then take the inverse. That time, denoted T, is found from

   $$\int_0^T dt = \int_0^{v_\theta} \frac{dv}{dv/dt}\,. \tag{3}$$

   Using Eq. (1a), and taking into account that for this question $u = 0$, this becomes

   $$T = \tau_v \int_0^{v_\theta} \frac{dv}{v^2 + w}\,. \tag{4}$$

   Making the change of variables $v = \sqrt{w}z$ gives us

   $$T = \frac{\tau_v}{\sqrt{w}} \int_0^{v_\theta/\sqrt{w}} \frac{dz}{z^2 + 1}\,. \tag{5}$$

   In the limit $w \to 0$, the upper limit of the z-integral goes to ∞, and the integral goes to a constant (which I think is $\pi/2$, but that's not important). Consequently, $T \propto \tau_v/\sqrt{w}$, which means the firing rate is proportional to $\sqrt{w}/\tau_v$.

2. For the rest of this question we'll include the u-variable. What kind of current does this variable correspond to?

   (5 marks)

Solution

Increasing u reduces the firing rate (it effectively decreases w), and in fact when $u > w$ the neuron doesn't fire at all. A natural candidate is a potassium current, since its reversal potential is around -80 mV and so tends to suppress firing. Given Eq. (1b), the current should be activated at high voltage.

3. Draw the nullclines assuming $w > 0$. Show, graphically if you want. that there are either zero, one or two equilibria, depending on the value of β. Quantitatively, how does the number of equilibria depend on β?

   (10 marks)

   Solution

   The nullclines are shown in Fig. 1. In the left panel β is large enough that there are two equilibria; in the right panel it's smaller, and there are no equilibria. In between there's a value of β that yields one equilibrium.

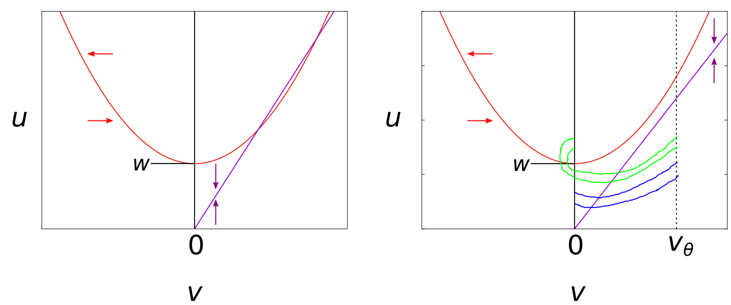


Figure 1: Nullclines for Eqns. (1a) and (1b). Red corresponds to v-nullclines (Eq. (1a)) and purple to u-nullcline (Eq. (1b)). Red arrows indicate dv/dt; purple arrows indicate du/dt. Left: two equilibria; the upper one is stable. Right: zero equilibria. The blue and green curves show, schematically, trajectories starting at $v = 0$. The blue trajectories go to the right; the green ones start off going to the left, but go to the right after they cross the v-nullcline.

To determine how the number of nullclines depends on β, we note that at equilibrium,

$$v^2 - 2\beta v + w = 0\,. \tag{6}$$

This has two solutions

$$v_{\pm} = \beta \pm \sqrt{\beta^2 - w}\,. \tag{7}$$

If $\beta^2 > w$ there are two real solutions, corresponding to two equilibria. If $\beta^2 = w$ there's one real solution, corresponding to one equilibrium. And if $\beta^2 < w$ there are no real solutions, and so no equilibria.

4. In the regime where there are two equilibria, show that one of them is unstable, and give conditions for the stability of the other one.

   (10 marks)

   Solution

To determine stability, we need the linearized dynamics. For that we let $v \to v_\pm + \delta v$ and $u \to u_\pm + \delta u$ where $v_\pm$ and $u_\pm$ are the steady state values of v. When δv and δu are small they obey the equations

$$\frac{d}{dt}\begin{pmatrix}\delta v\\ \delta u\end{pmatrix} = \begin{pmatrix}2v_\pm/\tau_v & -1/\tau_v\\ 2\beta/\tau_u & -1/\tau_u\end{pmatrix}\begin{pmatrix}\delta v\\ \delta u\end{pmatrix}. \tag{8}$$

This system of equations is stable (the eigenvalues have negative real part) if the trace is negative and the determinant is positive. Those are given respectively by

$$\begin{aligned}\text{trace} &= 2v_\pm/\tau_v - 1/\tau_u\\ &= 2\left(\beta \pm \sqrt{\beta^2 - w}\right)/\tau_v - 1/\tau_u\end{aligned} \tag{9a}$$

$$\begin{aligned}\text{determinant} &= 2(\beta - v_\pm)/(\tau_v\tau_u)\\ &= \mp(2/(\tau_v\tau_u))\sqrt{\beta^2 - w}\end{aligned} \tag{9b}$$

where the second equality in each equation follows from Eq. (7). For the upper equilibrium the determinant is negative, so it's unstable. For the lower equilibrium the determinant is positive, so we have to look at the trace. As is reasonably straightforward to show, the trace is negative, and thus the fixed point is stable, if $\beta > (\gamma_0^2 + w)/2\gamma_0$ where $\gamma_0 = \tau_v/2\tau_u$. We also need $\beta > w$; otherwise there won't be two fixed points.

The algebra to back this up is as follows. The trace is negative if

$$\beta - \sqrt{\beta^2 - w} < \gamma_0 . \tag{10}$$

Adding the square root to both sides and squaring gives us

$$(\beta - \gamma_0)^2 = \beta^2 - 2\beta\gamma_0 + \gamma_0^2 < \beta^2 - w . \tag{11}$$

Solving for β gives us

$$\beta > \frac{\gamma_0^2 + w}{2\gamma_0} . \tag{12}$$

We also need $\beta > w$; otherwise there won't be a stable equilibrium.

5. Consider a regime in which there are no equilibria, so the neuron spikes continuously. Let u_n be the value of u immediately after the n^{th} spike. Define the function $f(\cdot)$ via

$$u_{n+1} = f(u_n) . \tag{13}$$

Sketch $f(u)$ qualitatively. It will help to sketch some trajectories.

(10 marks)

Solution

The right hand panel of Fig. 1 contain sketches of four trajectories, all starting at $v = 0$. the blue ones move to the right; the green ones initially move to the left, and start moving to the right after they cross the v-nullcline. Thus, the starting point of these trajectories is u_n for some n. Comparing the two blue trajectories, both of which start below $u = w$, we see that

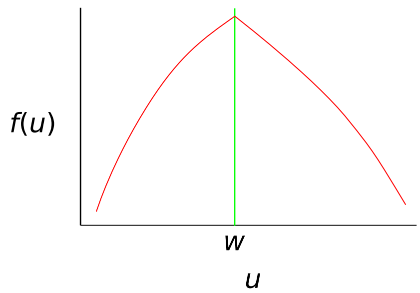


Figure 2: $f(u)$ (red curve). $u = w$ is marked with a vertical green line.

larger u_n leads to larger u_{n+1}. For the green trajectories, however, which start above $u = w$, it's the opposite: larger u_n leads to smaller u_{n+1}. Thus, $f(u)$ is an increasing function of u for $u < w$ and a decreasing function for $u > w$. I believe (although I admit to not checking very carefully) that the derivative is discontinuous at $u = w$. A plot showing the qualitative shape of $f(u)$ is shown in Fig. 2.

$$ \tag{14}$$

2 Networks

Consider a neuronal network in one dimension with Mexican-hat connectivity:

$$\frac{\partial s(x,t)}{\partial t} = -\frac{1}{\tau} s(x,t) + \phi \left[\int W_n^g(x-x') s(x',t) dx' + B(x) \right] \tag{1}$$

Where ϕ is a monotonic input-output function, $B(x)$ is an external input and $W_n^g(x-x')$ is the connection strength between units at x and x', given by a difference of Gaussians:

$$W_n^g(x-x') = W_n^g(\Delta x) = \alpha_E \; e^{-\frac{\Delta x^2}{2(\eta_E \sigma(n))^2}} - \alpha_I \; e^{-\frac{\Delta x^2}{2(\eta_I \sigma(n))^2}} \tag{2}$$

Where we defined $\Delta x = x - x'$, η_E, η_I are constants and $\sigma(n)$ is a linear function of n, $\sigma(n) = \sigma_0 + \sigma_1 n$.

1. (5 marks) Derive an equation for the fixed point solution of Eq. (1), s_n^0, when the input $B(x) = B^0$. Consider here the case of $\phi(x) = RELU(x)$ and sufficiently large B_0, such that a nontrivial solution exists.
2. (15 marks) Consider the evolution of a perturbation around that fixed point solution

$$s_n(x,t) = s_n^0 + \delta s_n(x,t) \tag{3}$$

   Using that the perturbation is small, find an explicit expression for $\delta s(x,t)$ in its Fourier basis and show that the critical spatial frequency that first emerges when the homogeneous fixed-point solution loses stability (i.e. spontaneous pattern formation) is given by

$$k^{*2} = \frac{1}{\sigma^2(n)} \frac{2}{(\eta_E^2 - \eta_I^2)} \log \left(\frac{\alpha_E}{\alpha_I} \; \frac{\eta_E^3}{\eta_I^3} \right) \tag{4}$$

   and therefore the spacing between the peaks of the oscillatory solution scales linearly with $\sigma(n)$. Plot how this solution will look like for $\sigma(n)$ linearly decreasing with n ($\sigma_1 < 0$).
3. Consider the network in Eq. (1) with a connectivity now given by the sum of a the global interaction term, $W_n^g(\Delta x)$ defined in Eq. (2) and a local interaction term, independent of n, $W^f(\Delta x)$

$$W_n(\Delta x) = W_n^g(\Delta x) + W^f(\Delta x) \qquad \text{with} \quad W^f(\Delta x) = \alpha_s \; e^{-\frac{(\Delta x - d)^2}{2\varepsilon_s^2}} + \alpha_s \; e^{-\frac{(\Delta x + d)^2}{2\varepsilon_s^2}} \tag{5}$$

(a) (5 marks) Write down $W_n(\Delta x)$ in Fourier space (i.e. find $\hat{W}_n(k)$). Can you write down an explicit analytical expression for the critical value of the frequency k^* at which the network loses linear stability in this case? why?

(b) (7.5 marks) Assume that $\alpha_E = \alpha_I$. From the Eq. of $\hat{W}_n(k)$, what is the condition for η_E and η_I such that $W_n^g(k)$ looks like the blue line in Fig. 3? And the relationship between ε and d such that $W^f(k)$ looks like the orange curve in Fig. 3?

(c) (7.5 marks) Plot your estimate of $W_n(k)$ from the curves in Fig. 3. The peaks of $\hat{W}^g(k)$ vary smoothly with $\sigma(n)$ as you showed in item (2). How does the critical value of the frequency k^* at which the network loses linear stability (i.e. the value of k that maximizes $\hat{W}_n(k)$) depend on $\sigma(n)$?

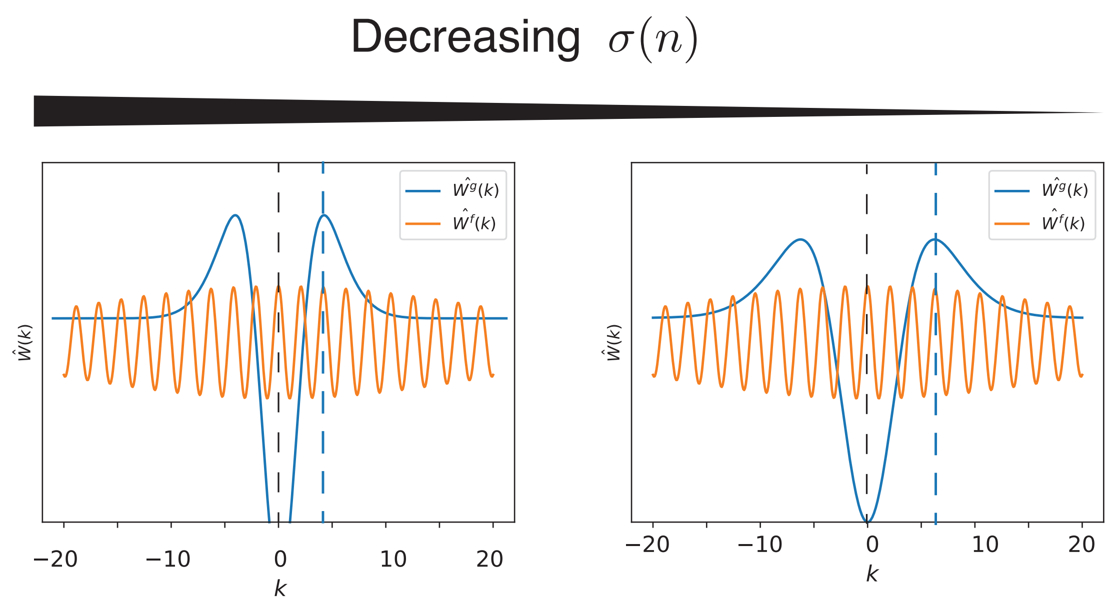


Figure 3: Left: W_n^g and W^f for small n. Right: W_n^g and W^f for larger n.

Reminder:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-d)^2}{2\sigma^2}} e^{\pm ikx} dx = e^{-\frac{k^2\sigma^2}{2}} e^{\pm ikd} \tag{6}$$

Solution

1. First one needs to compute the homogeneous steady-state solution to a constant input. If the input is constant we assume a homogeneous solution $s_n(x, y) = s_n^0$:

$$s_n^0 = \tau\phi(\int W_n^g(x-x')s_n^0 dx' + B_0) = \tau\phi(\bar{W}_n^g s_n^0 + B_0) \tag{7}$$

$$s_n^0 = \tau(\bar{W}_n^g s_n^0 + B_0) \tag{8}$$

$$s_n^0 = \frac{\tau B_0}{1-\tau\bar{W}_n^g} \tag{9}$$

$$\tag{10}$$

Where we used that $\int W_n^g(x-x')dx' = \bar{W}_n^g$.

2. We make a perturbation $s_n(x,t) = s_n^0 + \delta s_n(x,t)$ and derive under which conditions the system does not relax back to s_n^0.

$$\frac{\partial \delta s_n(x,t)}{\partial t} = -\frac{1}{\tau}(s_n^0 + \delta s_n(x,t)) + \phi\left(\int_{-\infty}^{\infty} W_n^g(x-x')(s_n^0 + \delta s_n(x',t))dx' + B_0\right) \tag{11a}$$

$$\frac{\partial \delta s_n(x,t)}{\partial t} = -\,-\frac{1}{\tau}\delta s_n(x,t) + \phi'(v_0)\int_{-\infty}^{\infty} W_n^g(x-x')\delta s(x',t)dx'\,. \tag{11b}$$

Because the system is linear we propose a plane wave solution $\delta s_n(x,t) = c(t)e^{-ikx}$

$$\dot{c}(t)e^{-ikx} = -\frac{1}{\tau}c(t)e^{-ikx} + \phi'(v_0)c(t)\int_{-\infty}^{\infty} W_n^g(x-x')e^{-ikx'}dx' \tag{12a}$$

$$\dot{c}(t) = -\frac{1}{\tau}c(t) + \phi'(v_0)c(t)\int_{-\infty}^{\infty} W_n^g(y)e^{iky}dy \tag{12b}$$

$$\dot{c}(t) = -c(t)(\frac{1}{\tau} - \phi'(v_0)\hat{W}_n^g(k))\,. \tag{12c}$$

Which means that the perturbation will decay exponentially if

$$1/\phi'(v_0) > \tau\hat{W}_n^g(k). \tag{13}$$

We first compute the Fourier transform of $W_n^g(x)$

$$W_n^g(x) = \alpha_E\ e^{-\frac{x^2}{2(\eta_E\sigma(n))^2}} - \alpha_I\ e^{-\frac{x^2}{2(\eta_I\sigma(n))^2}} \tag{14}$$

$$\hat{W}_n^g(k) = \alpha_E\ \sqrt{2\pi}\eta_E\sigma(n)\ e^{-\frac{(\eta_E\sigma(n)k)^2}{2}} - \alpha_I\ \sqrt{2\pi}\eta_I\sigma(n)\ e^{-\frac{(\eta_I\sigma(n)k)^2}{2}} \tag{15}$$

$$\tag{16}$$

need to find the value of k^* that maximizes $\hat{W}(k)$

$$\frac{d\hat{W}(k)}{dk}|_{k^*} = 0 \tag{17}$$

$$\frac{d\hat{W}(k)}{dk}|_{k^*} = -\sqrt{2\pi}\alpha_E\ \left(\eta_E\sigma(n)\right)^3\ k^*\ e^{-\frac{(\eta_E\sigma(n)k^*)^2}{2}} + \sqrt{2\pi}\alpha_I\ \left(\eta_I\sigma(n)\right)^3 k^*\ \ e^{-\frac{(\eta_I\sigma(n)k^*)^2}{2}} \tag{18}$$

From which we obtain

$$
\begin{aligned}
0 &= -\alpha_E \; \eta_E^3 \; e^{-\frac{(\eta_E \sigma(n) k^*)^2}{2}} + \alpha_I \; \eta_I^3 \; e^{-\frac{(\eta_I \sigma(n) k^*)^2}{2}} \\
e^{\sigma^2(n) \frac{k^{*2}}{2} (\eta_E^2 - \eta_I^2)} &= \frac{\alpha_E}{\alpha_I} \; \frac{\eta_E^3}{\eta_I^3} \\
k^{*2} &= \frac{1}{\sigma^2(n)} \frac{2}{(\eta_E^2 - \eta_I^2)} \log \left(\frac{\alpha_E}{\alpha_I} \; \frac{\eta_E^3}{\eta_I^3} \right)
\end{aligned}
$$

3. (a) We again need to find the maxima of the Fourier transom of the connectivity, $\hat{W}_n(k)$. This has now the form:

$$\hat{W}_n(k) = \hat{W_n}^g(k) + \hat{W}^f(k) \tag{20}$$

$$\hat{W_n}^g(k) = \sqrt{2\pi}\sigma(n) \left[\alpha_E \eta_E \exp\left(-\frac{\eta_E^2 \sigma(n)^2 k^2}{2} \right) - \alpha_I \eta_I \exp\left(-\frac{\eta_I^2 \sigma(n)^2 k^2}{2} \right) \right] \tag{21}$$

$$\hat{W}^f(k) = 2\sqrt{2\pi}\alpha_S \varepsilon_S \cos(kd) \exp\left(-\frac{\varepsilon_S^2 k^2}{2} \right)$$

This will be a transcendental equation, so we cannot obtain and explicit expression for k^*

(b) $\eta_E < \eta_I$ and $\varepsilon_s << d$

(c) Here what they should realize is that k^* is going to change discontinuously with n, through the peak selection mechanisms shown in Fig, 3C of Khona, Chandra and Fiete Nature 2025

3 Coding

The lobula giant motion detector (LGMD) is a neuron in the locust's lobula plate that responds to looming stimuli and seems to mediate an escape response. Hatsopolous et al (1995) and Gabbiani et al (1999) report that the firing rate of the neuron depends on both the angle subtended by a looming stimulus, and the angular velocity of its edges, according to:

$$f(t+\delta) = \dot{\theta}(t)e^{-\alpha\theta(t)} \tag{1}$$

where $\theta(t)$ is **half** the angle subtended by the object at time t relative to impact (the half will be useful to avoid having to carry many factors of 2); $\dot{\theta}$ is then the angular velocity of an edge; and $\alpha > 0$ and the delay δ are parameters that vary from animal to animal. Here, we will set δ to 0 for convenience.

1. Assume an object with a linear size of $2l$ approaches the locust's eye at a constant linear speed v. Let $t = 0$ be the time of impact. What angle does the object subtend at some time $t < 0$, expressed as a function of l and v? (3 marks)

2. Show that, if the LGMD neuron fires as in equation (1), its firing rate will peak at a time

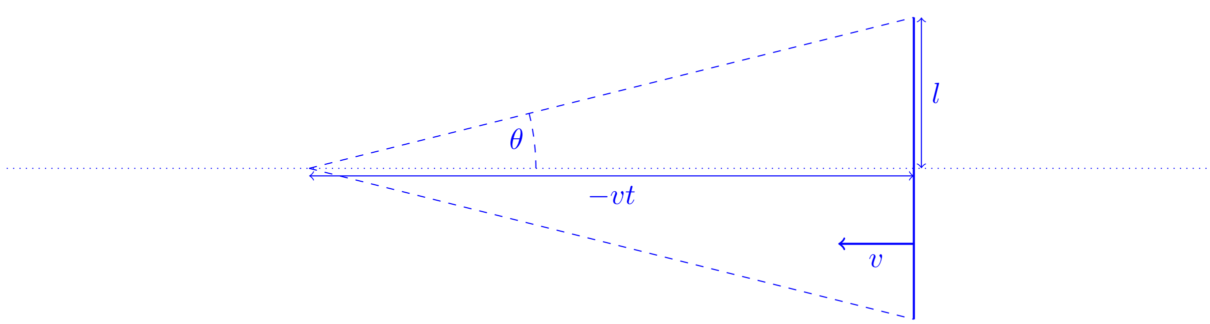
$$t_{\text{peak}} = -\alpha\frac{l}{v}$$

with a value

$$f_{\text{peak}} = F_{\max}(\alpha)\frac{v}{l}$$

where $F_{\max}$ is a function that depends on α alone. (12 marks)

Solution



1.

$$\theta = \tan^{-1}\left(\frac{l}{-vt}\right) = \frac{\pi}{2} - \tan^{-1}\left(\frac{-vt}{l}\right)$$

2.

$$\dot{\theta} = \frac{-1}{1+\left(\frac{vt}{l}\right)^2}\frac{-v}{l} = \frac{vl}{l^2+(vt)^2}$$

$$\ddot{\theta} = \frac{-vl}{(l^2+(vt)^2)^2}2v^2t$$

$$\dot{f} = (\ddot{\theta} - \alpha\dot{\theta}^2)e^{-\alpha\theta}$$

so at peak ($\dot{f} = 0$):

$$\ddot{\theta} = \alpha\dot{\theta}^2$$
$$\frac{-vl}{(l^2 + (vt)^2)^2} 2v^2 t = \alpha \left(\frac{vl}{l^2 + (vt)^2} \right)^2$$
$$-2v^3 lt = \alpha v^2 l^2$$
$$t = -\frac{\alpha}{2} \frac{l}{v}$$

May be a bug in the question (need to check original paper)!!

At that point

$$\theta = \tan^{-1}\left(\frac{l}{-vt} \right) = \tan^{-1} \frac{2}{\alpha}$$
$$\dot{\theta} = \frac{vl}{l^2 + (vt)^2} = \frac{vl}{l^2 + (\frac{\alpha}{2} l)^2} = \frac{v}{l} \frac{1}{1 + (\frac{\alpha}{2})^2}$$

so

$$f(t_{\text{peak}}) = \frac{v}{l} \frac{1}{1 + (\frac{\alpha}{2})^2} e^{-\alpha \tan^{-1} \frac{2}{\alpha}}$$

Based on this result, the authors above suggested that the occurance of the peak is an important cue to escape. This raises the question of how easy it might be for the downstream system to recognise the peak in the firing rate. One way to do so, would be to detect a negative rate of change in the LGMD firing rate. We will analyse the accuracy with which this might be possible, using a discretised simplification.

Suppose that LGMD fires according to an inhomogeneous Poisson process, and that the putative peak detector integrates spikes from the LGMD within two non-overlapping windows of the same size. Let the first have a mean spike rate of μ, and the second a mean rate of $\mu + \delta\mu$.

3. Calculate the Fisher Information matrix for the parameter vector $(\mu, \delta\mu)$. (5 marks)

4. Use this to derive the Cramér-Rao lower bound for the variance of an unbiased estimator of $\delta\mu$. (5 marks)

5. Assume an unbiased efficient estimator with a Gaussian distribution is available (not a good assumption). How large must $|\delta\mu|$ be for a simple peak detector that just looks at the sign of such an estimate to experience a false negative rate of about 2.5% [recall that if $\Phi(z)$ is the standard normal CDF, $\Phi(2) \approx 0.975$]. (5 marks)

6. Now, suppose the peak firing rate for a particular looming object is around 240 spikes s^{-1}, and that we can approximate equation (1) by a linear fall to 0 during the following 150 ms. Assume the peak detector integrates over 50 ms windows, and consider an ideal setting where the first window is centred over the peak. (Ignore the edge effects, and assume that the mean rate of 240 spikes s^{-1} is maintained throughout this window). How much later must the other 50 ms window be to achieve the 2.5% false negative rate? (5 marks)

7. How would this duration change if the looming object approached at the same speed, but was half the size (assume the time from peak to 0 is the same)? If the locust had time to escape the first object, would it also have time to escape this one? (5 marks)

4 Learning

Assume that the vector of firing rates of a neuron, $\mathbf{y}$, is related linearly to its input, $\mathbf{x}$,

$$\mathbf{y} = \mathbf{w} \cdot \mathbf{x} \tag{1}$$

where $\mathbf{y}$ is a vector in $\mathcal{R}^m$, $\mathbf{x}$ is a vector in $\mathcal{R}^n$, $\mathbf{w}$ is an $m \times n$ matrix of weights, and "·" denotes the usual dot product. The weights are learned via a generalization of Oja's rule,

$$\frac{d\mathbf{w}}{dt} = \langle \mathbf{y}(\mathbf{x} - \mathbf{y} \cdot \mathbf{w}) \rangle_{P(\mathbf{x})} \tag{2}$$

where two vectors back to back indicate an outer product: $(\mathbf{yx})_{ij} = y_i x_j$ and $(\mathbf{yy} \cdot \mathbf{w})_{ij} = y_i (\mathbf{y} \cdot \mathbf{w})_j$. Use $\mathbf{\Sigma}$ to denote the (uncentered) covariance of the data,

$$\langle \mathbf{xx} \rangle_{P(\mathbf{x})} = \mathbf{\Sigma} \, . \tag{3}$$

1. Show that $\mathbf{w}$ evolves according to

$$\frac{d\mathbf{w}}{dt} = \mathbf{w} \cdot \mathbf{\Sigma} - \mathbf{w} \cdot \mathbf{\Sigma} \cdot \mathbf{w}^T \cdot \mathbf{w} \tag{4}$$

   where T denotes transpose.

   (5 marks)

   Solution

   Inserting Eq. (1) into (2) gives us

$$\begin{aligned} \frac{d\mathbf{w}}{dt} &= \left\langle \mathbf{w} \cdot \mathbf{x}(\mathbf{x} - \mathbf{x} \cdot \mathbf{w}^T \cdot \mathbf{w}) \right\rangle_{P(\mathbf{x})} \\ &= \mathbf{w} \cdot \langle \mathbf{xx} \rangle_{P(\mathbf{x})} - \mathbf{w} \cdot \langle \mathbf{xx} \rangle_{P(\mathbf{x})} \cdot \mathbf{w}^T \cdot \mathbf{w} \, . \end{aligned} \tag{5}$$

   Using Eq. (3) for the averages over $P(\mathbf{x})$, we recover Eq. (4).

2. Assuming all eigenvalues of $\mathbf{\Sigma}$ are different, what is the **stable** steady state solution of $\mathbf{w}$? Stable is emphasized because there are many steady state solutions.

   (20 marks)

   Solution

   Although we're ultimately interested in steady state, we'll start with the full dynamics, as we'll need that. As usual, we let

$$\mathbf{w}(t) = \sum_k \mathbf{a}_k(t) \mathbf{v}_k \tag{6}$$

   where $\mathbf{v}_k$ is the k^{th} eigenvector of $\mathbf{\Sigma}$, with corresponding eigenvalue λ_k,

$$\mathbf{\Sigma} \cdot \mathbf{v}_k = \lambda_k \mathbf{v}_k \, . \tag{7}$$

Because $\mathbf{\Sigma}$ is symmetric, the $\mathbf{v}_k$ are orthogonal; we'll choose them to be orthonormal,

$$\mathbf{v}_k \cdot \mathbf{v}_l = \delta_{kl} \tag{8}$$

Inserting Eq. (6) into Eq. (4) gives us

$$\sum_k \frac{d\mathbf{a}_k}{dt} \mathbf{v}_k = \sum_k \lambda_k \mathbf{a}_k \mathbf{v}_k - \left(\sum_l \lambda_l \mathbf{a}_l \mathbf{a}_l \right) \cdot \sum_k \mathbf{a}_k \mathbf{v}_k \,. \tag{9}$$

Using the orthogonality condition, Eq. (8) (dot both sides with $\mathbf{v}_m$ and then let $m \to k$), we have

$$\frac{d\mathbf{a}_k}{dt} = \lambda_k \mathbf{a}_k - \mathbf{C} \cdot \mathbf{a}_k \tag{10}$$

where

$$\mathbf{C} \equiv \sum_l \lambda_l \mathbf{a}_l \mathbf{a}_l \,. \tag{11}$$

In steady state, $\mathbf{a}_k$ must be an eigenvector of $\mathbf{C}$ with eigenvalue λ_k. This is a nonlinear equation, because $\mathbf{C}$ depends on the $\mathbf{a}_k$. But let's say we found a solution; we'll figure out what the solution is by demanding that it's stable. Because $\mathbf{C}$ is $m \times m$, there are at most m eigenvectors; we'll call these, in a slight abuse of notation, the set of "nonzero" eigenvectors. That leaves another (at least) $(n - m)$ vectors $\mathbf{a}_k$; these must all be set to zero. We'll denote them $\mathbf{a}_{0l}$ with l *not* one of the nonzero eigenvectors.

Now we'll look at stability. Focus first on the $\mathbf{a}_{0l}$. Letting $\mathbf{a}_{0l} \to \mathbf{a}_{0l} + \delta \mathbf{a}_{0l}$, with, recall $\mathbf{a}_{0l} = 0$, and inserting that into Eq. (10), we have

$$\frac{d\delta\mathbf{a}_{0l}}{dt} = \lambda_l \delta \mathbf{a}_{0l} - \mathbf{C} \cdot \delta \mathbf{a}_{0l} \,. \tag{12}$$

We can let $\delta \mathbf{a}_{0l}$ point in a direction whose eigenvalue is one of the λ_k from the nonzero eigenvalues. In that case, the right hand side is proportional to $(\lambda_l - \lambda_k)$. For stability, we must have $\lambda_l < \lambda_k$. Thus, the nonzero set must consist of the largest, and only the largest, eigenvalue. Since all the eigenvalues are different, there's only one largest, which we denote λ_1. Thus, $\mathbf{C} = \lambda_{k_1} \mathbf{v}_1 \mathbf{v}_1$, and Eq. (10) becomes

$$\frac{d\mathbf{a}_{k_1}}{dt} = \lambda_{k_1} (1 - \mathbf{a}_{k_1} \cdot \mathbf{a}_{k_1}) \mathbf{a}_{k_1} \,. \tag{13}$$

In steady state, $\mathbf{a}_{k_1} \cdot \mathbf{a}_{k_1} = 1$. We therefore have

$$\mathbf{w} = a_{k_1} \mathbf{v}_{k_1} \tag{14}$$

with $\mathbf{a}_{k_1} \cdot \mathbf{a}_{k_1} = 1$. It's quite easy to show that this solution is stable.

3. If all went well in the previous question, you would have found that in steady state $\mathbf{w}$ was highly degenerate: all rows were proportional to the eigenvector of $\mathbf{\Sigma}$ with the largest eigenvalue. This means all components of $\mathbf{y}$ are extracting more or less the same features from the

data, which is not all that useful. To fix that we might try adding mutual inhibition between the y_i. We'll treat them as neurons which obey the equation

$$\tau \frac{d\mathbf{y}}{dt} = -\mathbf{M} \cdot \mathbf{y} - \mathbf{y} + \mathbf{w} \cdot \mathbf{x} \tag{15}$$

where $\mathbf{M}$ is a matrix all of whose elements are positive. Consider the simplest possible scenario: the weights are updated according to Eq. (2) with y given by the steady state solution to Eq. (15). Show that the weights evolve according to

$$\frac{d\mathbf{w}}{dt} = (\mathbf{I} + \mathbf{M})^{-1} \cdot \mathbf{w} \cdot \mathbf{\Sigma} - (\mathbf{I} + \mathbf{M})^{-1} \cdot \mathbf{w} \cdot \mathbf{\Sigma} \cdot \mathbf{w}^T \cdot (\mathbf{I} + \mathbf{M})^{-T} \cdot \mathbf{w} \tag{16}$$

where the superscript "$-T$" indicates the inverse and transpose.

(5 marks)

Solution

At equilibrium, $\mathbf{y}$ is given by

$$\mathbf{y} = (\mathbf{I} + \mathbf{M})^{-1} \cdot \mathbf{w} \cdot \mathbf{x} \,. \tag{17}$$

Inserting this into Eq. (2) then gives us

$$\frac{d\mathbf{w}}{dt} = \left\langle (\mathbf{I} + \mathbf{M})^{-1} \cdot \mathbf{w} \cdot \mathbf{x} \left(\mathbf{x} - \mathbf{x} \cdot \mathbf{w}^T \cdot (\mathbf{I} + \mathbf{M})^{-T} \cdot \mathbf{w} \right) \right\rangle_{P(\mathbf{x})} \,. \tag{18}$$

Averaging over $\mathbf{x}$ and using Eq. (3) gives us Eq. (16).

4. Ignoring the nonlinearity in Eq. (16), $\mathbf{w}$ obeys a linear equation. We want to figure out what the fastest growing mode looks like. Because matrices are difficult to deal with, define the vector $\mathbf{m}_k \equiv \mathbf{w} \cdot \mathbf{v}_k$ where $\mathbf{v}_k$ is the k^{th} eigenvector of $\mathbf{\Sigma}$ whose corresponding eigenvalue is λ_k,

$$\mathbf{\Sigma} \cdot \mathbf{v}_k = \lambda_k \mathbf{v}_k \,. \tag{19}$$

Using the $\mathbf{m}_k$, find an equation for the fastest growing mode.

(10 marks)

Solution

The $\mathbf{m}_k$ obey the equation

$$\frac{d\mathbf{m}_k}{dt} = \lambda_k (\mathbf{I} + \mathbf{M})^{-1} \cdot \mathbf{m}_k \,. \tag{20}$$

All the $\mathbf{m}_k$ evolve according to the same matrix, but with amplitude scaled by the eigenvalues of $\mathbf{\Sigma}$. Since we're after the fastest growing mode, we only need consider the largest eigenvalue, which we'll take to be the one with $k = 1$. The direction of the fastest growing mode is

$$\mathbf{m}_1 = \mathbf{u}_1 \,, \tag{21}$$

where $\mathbf{u}_1$ is the eigenvector of $(\mathbf{I} + \mathbf{M})^{-1}$ with largest eigenvalue.