Gatsby Computational Neuroscience Unit
Theoretical Neuroscience

Final Examination
25 Apr 2018

Part I – short questions

There are five sections with four questions each. Please answer three out of each four, starting the answers for each new section on a new page. Don't forget to write your name at the top of each block of answers.

You have a maximum of 6 hours for this exam.

Good luck!

1 Biophysics

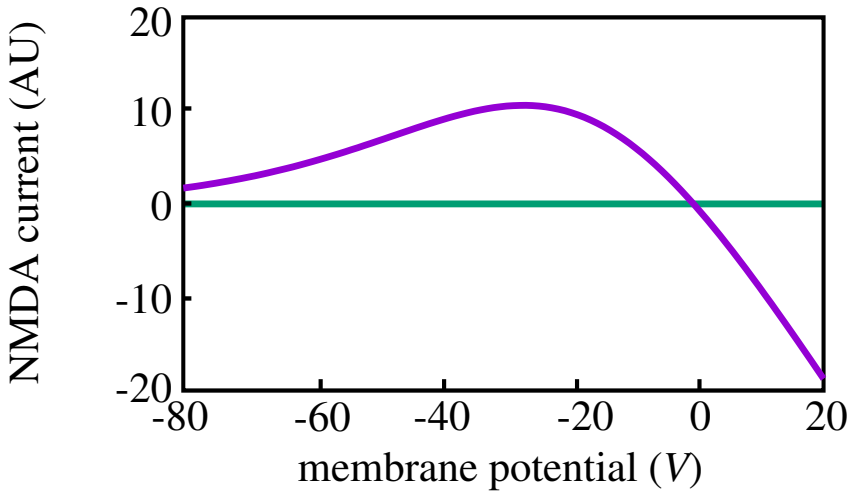
1. Plot the steady state I-V curve of an NMDA channel. In other words: assuming there is glutamate in the synaptic cleft, plot the current that flows into the postsynaptic bouton as a function of voltage. Comment on the relevance of this curve for synaptic plasticity.

   Solution

   The NMDA current is given by

   $$I_{NMDA} \propto -\frac{x(V-\mathcal{E}_{NMDA})}{1+([Mg^{++}]/3.57)\exp[-V/16.1]}$$

   where x is the fraction of open channels, $\mathcal{E}_{NMDA}$ is the NMDA channel reversal potential, $[Mg^{++}]$ is the magnesium concentration in mM, and V is in mV. We'll use $[Mg^{++}] = 1$ mM and $\mathcal{E}_{NMDA} = 0$ mV. A plot of the right hand side (with x nonzero, which is the case if there is glutamate in the synaptic cleft), is shown here,

   The precise values of the parameters aren't especially important; what is important is that the NMDA current is highest when the membrane potential is large. Essentially all learning rules involve a coincidence between pre and postsynaptic activity. If plasticity is triggered by ion flow through NMDA channels, then plasticity will be maximum when the membrane potential is high, which is the case when a postsynaptic spike occurs. Thus – at least in the simplest possible picture – plasticity occurs preferentially when there is both pre and postsynaptic activity.

2. Consider a quadratic integrate and fire neuron coupled to an I_h type current,

   $$\tau_m \frac{dV}{dt} = \frac{(V-V_{th})(V-V_{rest})}{V_{th}-V_{rest}} - gx(V-\mathcal{E}).$$

   A spike is emitted when V reached $+\infty$, at which point it is reset to $-\infty$. The gating variable, x, obeys the equation,

   $$\tau_x \frac{dx}{dt} = -(x - x_\infty(V)).$$

   As usual, we'll let $x_\infty(V)$ lie in the range $[0,1]$, and we'll assume that $x_\infty(V)$ is a monotonic function of V. Use $V_{th} = -50$, $V_{rest} = -70$, $g = 0.2$ and $\mathcal{E} = 0$.

   Plot the V-nullclines in x-V space, with V on the x-axis and x on the y-axis.

   Assuming $\tau_x \gg \tau_m$, sketch a curve $x_\infty(V)$ that will ensure the following:

   - There are two fixed points.
   - One is stable and the other is unstable.
   - The neuron can either go to the stable fixed point or exhibit repetitive spiking.

Solution

The V-nullcline is the curve (using actual numbers)

$$x = \frac{(V+70)(V+50)}{4V}\,. \tag{1}$$

For $V < 0$, in the regime where x is positive ($-70 < V < -50$), x is more or less an upside down parabola: setting V to -60 in the denominator, we get the approximate expression

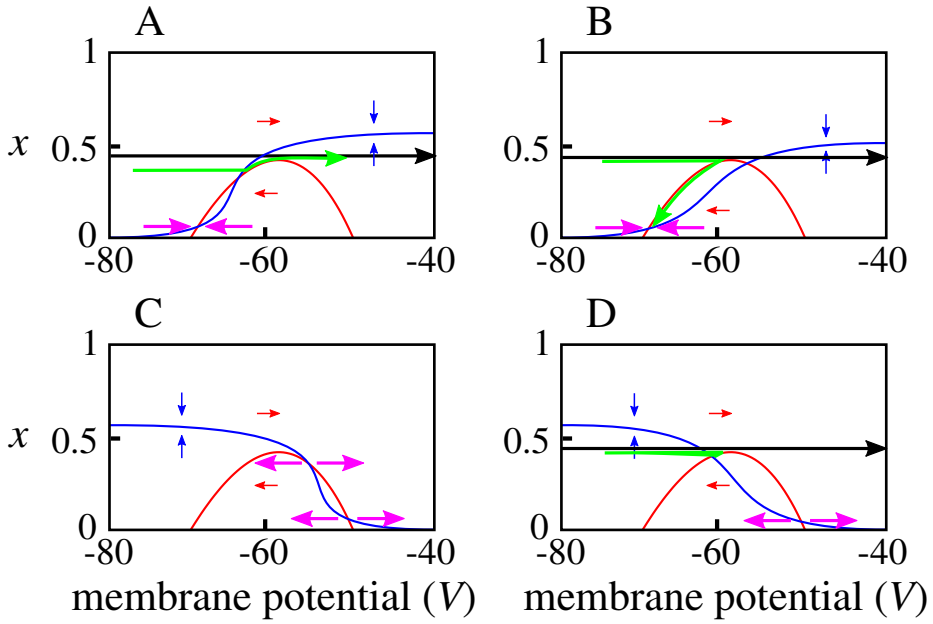
$$x \approx -\frac{(V+70)(V+50)}{240}\,. \tag{2}$$

This has a maximum at $V = -60$, yielding $x_{\max} = 1/2.4$, which is less than 1. A more careful analysis indicates that the maximum (with V negative) occurs at $V = -\sqrt{3500}$, yielding

$$x_{\max} = 2.5\,\frac{(7-\sqrt{35})(\sqrt{35}-5)}{\sqrt{35}}\,. \tag{3}$$

which is also less than 1 (and close to 1/2.4; it's about 1/2.38).

Now it's just a matter of drawing $x_\infty(V)$ and checking for fixed points and repetitive firing. The four main shapes are drawn here,



The V-nullclines are in red; the x-nullclines in blue. The red arrows show how the voltage evolves at fixed x; the blue arrows show how x evolves at fixed voltage. Black, green and magenta indicate trajectories.

In panel A, we're guaranteed to have one unstable and one stable fixed point, and repetitive firing. The black trajectory shows repetitive firing in the regime $\tau_x \gg \tau_m$ (we don't, of course, show all of it, as it goes to $+\infty$ and returns from $-\infty$). Note that the trajectories are pointed slightly down when above the x-nullcline and slightly up when below the nullcline. The net effect is, probably, a slow drift downward. However, if the trajectory were to come in from the left side below the top of the V-nullcline, it will just crawl back up, as shown by the green trajectory. It's likely that the black trajectory comes very close to the top of the V-nullcline, although it would take more careful analysis to show that.

In panel B there is also one unstable and one stable fixed point. There might be repetitive firing, as shown by the black trajectory. However, this time if the trajectory comes in from the left side below the V-nullcline, it will *cross* the V-nullcline and be pulled toward the fixed point, as shown by the green trajectory. If this were to happen for any initial condition above the peak of the V-nullcline, then repetitive firing would not be possible. Whether or not this is the case will require more analysis.

In panel C, there are two unstable fixed points, so that one is ruled out.

Panel D is much like panel B: there is one unstable and one stable fixed point, and there might repetitive firing, as shown by the black trajectory. However, again if the trajectory comes in below the peak of the V-nullcline, it will be pulled toward the stable fixed point (green trajectory; note that the arrowhead is not visible). For this scenario, as in panel B, more analysis is needed to determine what actually happens.

3. When "deriving" firing rate equations, we assumed that if we knew the steady state firing rates of presynaptic neurons, we would know the firing rate of the postsynaptic neurons. Why is this not true? Give examples.

   Solution

   The problem is that the firing rate statistics matter. Here are three examples.

   - There are a small number of excitatory presynaptic neurons – say around 20 – all firing asynchronously at 1 Hz. The resulting synaptic drive is very small; for your typical neuron, so small that the postsynaptic neuron will almost never first. Now imagine that the neurons fire perfectly synchronously. If 20 PSPs are enough to make the postsynaptic neuron fire, then it will fire at 1 Hz.
   - There are a large number of excitatory presynaptic neurons – say 10s of thousands – all firing synchronously at 1 Hz. This will make the postsynaptic neuron fire at 1 Hz. However, if the neurons fire asynchronously, the firing rate of the postsynaptic neuron is likely to be different. And, in fact, if there are enough presynaptic neurons, it could be much higher.
   - Presynaptic neurons are both excitatory and inhibitory. If the inhibitory ones tend to fire before the excitatory ones, then the postsynaptic firing rate will be lower than if the inhibitory neurons tend to fire after the excitatory ones.

4. Consider an LIF neuron with spike frequency adaptation. The LIF part is

$$\tau_m \frac{dV}{dt} = -V + h$$

   where V is voltage relative to rest and h is constant input. A spike is emitted when $V = \theta$, at which point the voltage is reset to 0.

   Now the spike frequency adaptation: whenever there's a spike, the threshold increases by Δ, and then decays back to 0,

$$\frac{d\theta}{dt} = -\frac{\theta}{\tau} + \Delta \sum_i \delta(t - t_i)\,.$$

   Here the t_i are the spike times of the neuron, an δ is the Dirac delta function.

   Show that in the limit $\tau \gg \tau_m$ [this used to say $\tau \gg \tau_m \log(1 + \Delta/h)$, but I'm not sure why; it's not needed], the firing rate, which we'll call ν^*, is given approximately by

$$\nu^* \approx \frac{1}{\tau \log(1 + \Delta/h)}\,.$$

   In hindsight this is not surprising: if the firing rate were much higher than $1/\tau$, the threshold would end up very high; if it were much lower than $1/\tau$, the threshold would end up very low.

   Hint: plot both the voltage, V and threshold, θ, versus time for several periods. In the steady state regime, the threshold should be exactly the same every time the neuron spikes.

   Solution

   The key idea is to realize that in steady state, the threshold jumps up by Δ when a spike occurs and decays by exactly Δ by the time the next spike occurs. Letting the first spike occur at $t = 0$, we can quantify this by writing

$$(\theta + \Delta)e^{-t^*/\tau} = \theta$$

   where t^* is the time of the second spike. This implies

$$t^* = \tau \log(1 + \Delta/\theta)\,. \tag{4}$$

   After a spike, the voltage evolves according to

$$V(t) = h\left(1 - e^{-t/\tau_m}\right)\,.$$

   The next spike occurs when $V(t^*) = \theta$. Using Eq. (4) for t^*, this gives us

$$h\left(1 - e^{-(\tau/\tau_m)\log(1+\Delta/\theta)}\right) = \theta\,. \tag{5}$$

To find t^* exactly, we would have to solve this equation for θ, which cannot be done analytically. However, we can make an approximation: Eq. (5) implies that $\theta < h$, which in turn tells us that

$$\frac{\tau}{\tau_m} \log\left(1 + \frac{\Delta}{\theta}\right) > \frac{\tau}{\tau_m} \log\left(1 + \frac{\Delta}{h}\right) .$$

The second term is much larger than 1 (because $\tau \gg \tau_m$), and thus so is the first. Consequently, the second term in parentheses in Eq. (5) is exponentially small, and so $\theta \approx h$. Inserting this into Eq. (4), we arrive at

$$t^* \approx \tau \log(1 + \Delta/h).$$

The inverse of this is the firing rate, giving us the desired result.

2 Networks

1. Consider a one memory probabilistic Hopfield network, which has the update rule

$$P(s_i(t+1)|m(t)) \propto e^{s_i(t+1)\beta\xi_i m(t)} \tag{6}$$

where s_i is either $+1$ or -1, $m(t)$ is the "overlap,"

$$m(t) \equiv \frac{1}{N}\sum_{j=1}^{N} \xi_j s_j(t)\,,$$

and ξ_i is a random binary vector,

$$\xi_i = \begin{cases} +1 & \text{probability } 1/2 \\ -1 & \text{probability } 1/2\,. \end{cases}$$

Show that for synchronous update,

$$\langle m(t+1)\rangle = \tanh \beta m(t)\,.$$

where the angle brackets indicate an average with respect to the probability distribution given in Eq. (6). Show that the variance of $m(t)$ is proportional to $1\sqrt{N}$. Consequently, this quantity self averages, and so we may write

$$m(t) = \tanh \beta m(t)\,. \tag{7}$$

Explain why this result is valid only if $\beta > 1$.

Solution

The properly normalized probability distribution over $s_i(t+1)$ is

$$P(s_i(t+1)|m(t)) = \frac{e^{s_i(t+1)\beta\xi_i m(t)}}{e^{\beta\xi_i m(t)} + e^{-\beta\xi_i m(t)}}\,.$$

consequently,

$$\langle s_i(t+1)\rangle = (+1)\times P(s_i(t+1)=1) + (-1)\times P(s_i(t+1)=-1) = \tanh(\xi_i\beta m(t)).$$

The average overlap is, then, given by

$$\langle m(t+1)\rangle = \frac{1}{N}\sum_j \xi_j \tanh(\xi_i\beta m(t))\,.$$

Use the fact that $\tanh(\xi_i\beta m(t)) = \xi_i \tanh(\beta m(t))$, which follows because ξ_i is either $+1$ or -1. Thus,

$$\langle m(t+1)\rangle = \frac{1}{N}\sum_j \xi_j^2 \tanh(\beta m(t)) = \tanh \beta m(t)\,.$$

The terms in the expression for $m(t)$ are independent. Thus, the variance of m(t) is given by

$$\mathrm{Var}[m(t)] = \frac{1}{N^2}\sum_j \xi_j^2 \mathrm{Var}[s_j(t)]\,.$$

We have already seen that the mean of $s_j(t)$ is $\tanh(\xi_i\beta m(t))$. The second moment of $s_j(t)$ is 1. Therefore,

$$\mathrm{Var}[s_j(t)] = 1 - \tanh^2(\xi_i\beta m(t)) = 1 - \tanh^2 \beta m(t) = \frac{1}{\cosh^2 \beta m(t)}\,.$$

Thus,

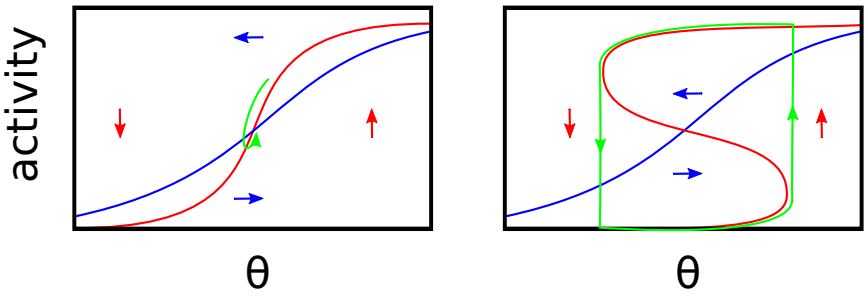
$$\mathrm{Var}[m(t)] = \frac{1}{N\cosh^2 \beta m(t)} \le \frac{1}{N}\,.$$

This implies that fluctuations in $m(t)$ are small compared to 1. Thus, so long as $m(t)$ is $\mathcal{O}(1)$, we may replace $m(t)$ by its average. However, if $\beta < 1$, then the solution to Eq. (7) is $m(t) = 0$. In that regime, the variance is larger than the mean, so Eq. (7) breaks down. This case we requires a more sophisticated analysis.

2. A friend has been doing experiments in a slice preparation, and she tells you that she often observes bursting: the slice will be completely quiet for a few hundred ms, then it will fire at a few Hz for another few hundred ms, then go back to silence, and so on. She asks you to explain it. You hypothesize that there's a variable that controls excitability in the network – synaptic strength or threshold for the generation of a spike are two obvious examples. Let's call that variable θ, and, for definiteness, define things so that high θ makes the network more excitable and low θ makes the network less excitable. You assume that θ depends on activity: it slowly increases when activity is low, and slowly decreases when activity is high. To complete your hypothesis, you claim that the network must be bistable; that is, for some range of θ, the network has two states, one silent and the other active. Why is bistability needed to get bursting?

   Solution

   A figure showing activity versus θ, both with and without bistability, is shown here.



   The blue lines are the θ nullclines and the red ones are the activity nullclines, with blue and red arrows indicating flow direction. In the left hand panel, the activity is not bistable. In this case, because θ changes slowly, the intersection is a fixed point (green trajectory). And that's pretty much the only option. In the right panel, there is bistability. In this case, trajectories show bursting: sudden transition from low to high activity, and then back again (green trajectory).

   This is known as a relaxation oscillator. Note that if the timescale for θ is very long, you see very sudden jumps in the state of the system even though the underlying parameters change slowly. This is probably a good model for climate change. Unfortunately.

3. Consider a network consisting of two excitatory populations coupled to one inhibitory one. The equilibrium equations for the two excitatory populations are

$$\boldsymbol{\nu}_1 = \phi\big(\mathbf{W}_1 \cdot \boldsymbol{\nu}_1 - \mathbf{J}_1 \cdot \boldsymbol{\nu}_I + h\mathbf{1} + \boldsymbol{\xi}_1\big)$$
$$\boldsymbol{\nu}_2 = \phi\big(\mathbf{W}_2 \cdot \boldsymbol{\nu}_2 - \mathbf{J}_2 \cdot \boldsymbol{\nu}_I - h\mathbf{1} + \boldsymbol{\xi}_2\big)$$

   where $\boldsymbol{\nu}_1$, $\boldsymbol{\nu}_2$ and $\boldsymbol{\nu}_I$ are vectors, ϕ is a pointwise nonlinearity (as usual, it's more or less sigmoidal), $\mathbf{1}$ is a vector containing all ones, and $\boldsymbol{\xi}_k$ is noise, which varies from trial to trial, and is independent across neurons. The equilibrium equation for the inhibitory neurons is

$$\boldsymbol{\nu}_I = \phi\big(\mathbf{B}_1 \cdot \boldsymbol{\nu}_1 + \mathbf{B}_2 \cdot \boldsymbol{\nu}_2 - \mathbf{J}_I \cdot \boldsymbol{\nu}_I + \boldsymbol{\xi}_I\big)\,.$$

   We'll assume that all weights are random and with elements drawn *i.i.d.* from a non-negative distribution, with weights of similar type drawn from the same distribution (i.e., $\mathbf{W}_1$ and $\mathbf{W}_2$ are drawn from the same distribution, $\mathbf{J}_1$ and $\mathbf{J}_2$ are drawn from the same distribution, etc.).

   There are two input conditions: $h = +h_0$ and $h = -h_0$. It should be relatively easy to decode the condition (that is, determine whether h is $+h_0$ or $-h_0$) from the excitatory neurons: if $h = +h_0$, population 1 should have higher average firing rate than population 2; if $h = -h_0$, population 2 should have higher average firing rate than population 1.

   One might think that the inhibitory neurons are not very informative. Certainly their population averaged firing rate will be approximately independent of condition. But by using a more sophisticated readout, is it be possible to decode condition from them? If so, how would decoding accuracy depend on the overall size of the weights?

Solution

The inhibitory neurons will clearly carry *some* information about the condition. That's because condition affects the excitatory firing rates, and different excitatory firing rates imply different inhibitory firing rates. To see this explicitly, consider a single inhibitory neuron, ν_{Ii}. Its activity is given by

$$\nu_{I,i} = \phi\left(\sum_j B_{1,ij}\nu_{1,j} + \sum_j B_{2,ij}\nu_{2,j} - \sum_j J_{I,ij}\nu_{I,j} + \xi_{I,i}\right) .$$

If $\boldsymbol{\nu}_1$ is active, the input will be different than if $\boldsymbol{\nu}_2$ is active.

How different depends on the fluctuation in the sums over j. That we know how to calculate: in general, if the terms in a sum are independent, then

$$\mathrm{Var}\left[\sum_{j=1}^{n} A_{ij}\nu_j\right] = n\mathrm{Var}[A]\langle\nu_j^2\rangle .$$

Thus, for the fluctuations to be large, we want the weights to be large. Particularly bad is $1/n$ scaling, for which the variance is $\mathcal{O}(1/n)$. Our standard $1/\sqrt{n}$ scaling, however, would be fine, because it would produce $\mathcal{O}(1)$ scaling.

4. Consider a randomly connected network of excitatory and inhibitory neurons operating in the balanced regime, with an equilibrium given by

$$\nu_{E,i} = \phi\left(K^{1/2}\left[W^{EE}\nu_E - W^{EI}\nu_I + h_E\right] + \xi_{I,i} + h_{ext}\right)$$
$$\nu_{I,i} = \phi\left(K^{1/2}\left[W^{IE}\nu_E - W^{II}\nu_I + h_I\right] + \xi_{E,i}\right)$$

where ϕ is the usual sigmoidal gain function, K is the average number of connections per neuron, ν_E and ν_I are the population averaged firing rates of the excitatory and inhibitory neurons, respectively, ξ is $\mathcal{O}(1)$ quenched noise, and h_{ext} is external input. Assume that the weights, W, and the static drive, h, are $\mathcal{O}(1)$.

Imagine that the external input, h_{ext} (which is applied to all the excitatory neurons), changes from trial to trial, with a variance equal to σ^2.

Show that if σ^2 is $\mathcal{O}(1)$ and K is large, then the trial to trial firing rate correlation between any two neurons is proportional to $1/K$.

Solution

The key idea is that the terms in brackets in the equation for the excitatory neurons must be $\mathcal{O}(1/\sqrt{K})$; otherwise, the neurons would either be firing at maximum rate or completely quiet. Thus, so long as the determinant of the weights is not zero (the standard case), fluctuations in the average firing rates are $\mathcal{O}(1/\sqrt{K})$. This means the average covariance among the firing rates of the individual neurons is $\mathcal{O}(1/K)$. For instance,

$$\mathrm{Var}[\nu_E] = \frac{1}{N^2}\sum_{ij}\mathrm{Covar}[\nu_{E,i},\nu_{E,j}] \sim \frac{1}{K} . \tag{10}$$

Similar relationships hold for the variance of ν_I and the covariance between ν_E and ν_I.

If the covariance between individual neurons could have different signs, then Eq. (10) wouldn't impose much of a constraint on the covariance between any two pairs of neurons. However, because h_{ext} causes all the neurons of a particular type to change by the same sign (since it affects the synaptic drive of all neurons by the same amount), it follows that the covariance between neurons of the same type must be positive, and the covariance between neurons of opposite type must all have the same sign (also positive, I believe, but that doesn't matter for this discussion). Thus, the covariance between any two pairs of neurons must be $\mathcal{O}(1/K)$.

3 Coding

1. Consider spikes on an axon generated according to a homogeneous Poisson process at a rate of 10 Hz. Show that if every other spike fails to cause a vesicle release, then the distribution of PSP times follows a Gamma-interval renewal process (of what order?). What is the distribution if the probability of vesicle release has an iid probability of 0.5?

2. A neural population responds to a stimulus with orientation θ with firing rates on the ith experimental trial given by

$$r_n^i(\theta) = (r_0 + \rho^i)e^{\kappa\cos(\theta-\theta_n-\theta^i)}$$

where ρ^i and θ^i reflect trial-to-trial random noise. Using a Gaussian approximation for r_n^i, find the Fisher information that this population conveys about θ.

3. A neuron tends to fire doublets of spikes, separated by about 2 ms, with the intervals between doublets dictated by a sensory stimulus. Describe a point-process GLM model suitable to describe this neuron.

4. Describe the different circumstances under which the spike-triggered covariance (STC) method would fail to consistently identify stimulus dimensions to which a neuron is sensitive.

4 Systems

1. Contrast the first few stages of the olfactory sensory pathway with those of vision, audition and somatosensation.

2. Describe two ways in which the cochlear output (expressed as a mean rate) differs from the magnitude output of a short-term Fourier transform.

3. Discuss the different roles played by the motor cortex, basal ganglia, cerebellum and spinal cord in controlling movements in mammals.

4. Describe two ways in which neural population activity could represent uncertainty about inferred features in the environment.

5 Learning

1. Consider a spike-timing dependent learning rule with kernel $H(\tau)$. Given pre- and post-synaptic spike times $\{t_i^{\rm pre}\}$ and $\{t_j^{\rm post}\}$ respectively, we write the net change in a synaptic weight as

$$\tau_w \frac{dw}{dt} = \sum_j \delta(t - t_j^{\rm post}) \sum_i H(t_i^{\rm pre} - t)\,.$$

   If the pre- and post-synaptic spikes are drawn from Poisson processes with rates $u(t)$ and $v(t)$ respectively, and both vary more slowly than $H(\tau)$ show that the expected change in the weight depends only on $H^1 = \int_{-\infty}^{\infty} d\tau H(\tau)$.

2. Consider the BCM learning rule

$$\tau_w \frac{d\vec{w}}{dt} = v\vec{u}(v - \theta)$$

   with a sliding threshold:

$$\tau_\theta \frac{d\theta}{dt} = v^p - \theta\,.$$

   Show that this rule avoids the usual instability of Hebbian learning provided that $p > 1$.

3. Consider the discount factor

$$g(t) = \gamma^{\log(1+kt)} \tag{11}$$

   where $0 < \gamma < 1$ and $k > 0$. This is reasonable if one works in log time – which is more or less how we view time, in many situations. For what value of γ is this hyperbolic discounting?

   Solution

   This can be written

$$g(t) = (1 + kt)^{\log \gamma}\,. \tag{12}$$

   For $\gamma = 1/e$, $\log \gamma = -1$, and we recover hyperbolic discounting.

4. Name one advantage and one disadvantages of model-based learning over model free learning.

   Solution

   The advantage of model-based learning is that it uses data very efficiently.

   The disadvantage is that it is computationally expensive. That's why we develop habits: we don't want to spend a lot of time and energy planning. At least I don't.