Gatsby Computational Neuroscience Unit
Theoretical Neuroscience

Final Examination
2 May 2019

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 new question on a new page. Don't forget to write your name at the top of the answer to each question.

You have a maximum of 7 hours for this exam.

Good luck!

1 Biophysics

Release probability is governed by two quantities. One is the amount of calcium in the presynaptic terminal, with higher calcium implying higher release probability. The other is release itself: every time a vescicle is released, the probability of subsequent release drops; then it decays exponentially back to baseline (which is calcium dependent).

1. Given this description, write down a set of equations describing release probability. Assume that presynaptic spikes arrive at Poisson rate ν, and that when a presynaptic spike arrives the calcium concentration in the presynaptic terminal jumps instantenously and then decause back to baseline with time constant τ_c. Use τ_r to denote the time constant at which release probability decays to its baseline.

   (15 marks)

2. Does average release probability increase or decrease as τ_c increases? Explain your reasoning.

   (5 marks)

3. Does average release probability increase or decrease as τ_r increases? Explain your reasoning.

   (5 marks)

4. Plot, qualitatively, the release probability versus the presynaptic firing rate, ν, in the regime $\tau_c \gg \tau_r$

   (15 marks)

Solutions

1. There is a fair amount of leeway here, but the following equations will work,

$$\begin{aligned} \frac{dc}{dt} &= \frac{c_0 - c}{\tau_c} + \sum_i \delta(t - t_i) \\ \frac{dp}{dt} &= \frac{p_0 \sigma(c) - p}{\tau_r} - (1 - f_D) p \sum_i \delta(t - t_i) \xi_i(t_i) \end{aligned}$$

   where c is calcium concentration in the presynaptic terminal, p is the release probability, t_i is the time of the i^{th} presynaptic spike, $\delta(\cdot)$ is the usual Dirac delta function, $\sigma(\cdot)$ is the sigmoid function, and ξ_i is an indicator function that tells us whether or not a vesicle was released,

$$\xi_i(t_i) = \begin{cases} 1 & \text{vesicle was released} \\ 0 & \text{otherwise}\,. \end{cases}$$

   We have to be a little bit careful with the delta functions. First, concentration is updated after release probability – so it doesn't affect release probability until the next spike. Second, the delta function in the equation for p is shorthand for the following rule: whenever there is a spike,

$$p \to p - (1 - f_D) p \xi_i(t_i) = p(1 - \xi_i(t_i)) + f_D p \xi_i(t_i) \, .$$

This makes a certain amount of sense: if $\xi_i(t_i) = 0$ then nothing happens; if $\xi_i(t_i) = 1$ then p is reduced by a factor of f_D.

2. It increases: the larger τ_c is, the smaller the effect of the first term in the equation for the concentration. Since the second term is positive, this means more calcium on average. More calcium increases the release probability via the term $\sigma(c)$.

3. It decreases: again the larger τ_r is, the smaller the effect of the first term in the equation for the release probability. Since the second term is negative, this means a smaller release probability on average.

4. For this question, we'll need

$$\left\langle \sum_i \delta(t - t_i) \right\rangle = \nu \, .$$

Consequently, we can write down approximate equations for the average calcium concentration and release probability,

$$\begin{aligned} \tau_c \frac{d\overline{c}}{dt} &\approx c_0 - \overline{c} + \nu \tau_c \\ \tau_r \frac{d\overline{p}}{dt} &\approx p_0 \sigma(\overline{c}) - \overline{p} - (1 - f_D) \overline{p}^2 \nu \tau_r \end{aligned}$$

where we used $\langle \xi_i(t_i) \rangle = \overline{p}$. When $\nu \tau_c \sim \mathcal{O}(1)$, $\overline{c}$ will increase, which will raise the release probability. Because $\tau_c \gg \tau_r$, at this value of the firing rate, the drive to the release probability, $\nu \tau_r$, will be negligible. Eventually, the firing rate will be high enough that $\nu \tau_r$ is not negligible, at which point the release probability will start to go down. Because of the factor $\sigma(c)$, it's not clear exactly when it will go down, but eventually that term will saturate, and we'll see a decrease in the release probability. Thus, the general shape is: first an increase, around $\nu \tau_c \sim \mathcal{O}(1)$, then a decrease, around $\nu \tau_r \sim \mathcal{O}(1)$.

2 Networks

Consider the usual time-dependent Hopfield network, but with input,

$$\frac{ds_i}{dt} + s_i = \tanh\left(\frac{\beta}{N}\sum_{j=1}^{N} J_{ij}s_j + h_i\right)$$

where

$$J_{ij} = \sum_{\mu=1}^{p} \xi_i^\mu \xi_j^\mu$$

with

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

For all questions, assume $1 \ll p \ll N$, $\beta \sim \mathcal{O}(1)$ and $\beta > 0$.

1. Define the overlaps in the usual way,

   $$m_\mu = \frac{1}{N}\sum_i \xi_i s_i.$$

   Show that the m_μ evolve according to

   $$\frac{dm_\mu}{dt} + m_\mu = \frac{1}{N}\sum_i \tanh\left(\beta m_\mu + \beta \sum_{\nu\neq\mu} \xi_i^\mu \xi_i^\nu m_\nu + \xi_i^\mu h_i\right).$$

   (5 marks)

2. Let

   $$h_i = \beta\gamma\xi_i^\rho$$

   with $\gamma > 0$. Show that when $\mu \neq \rho$, m_μ evoloves according to

   $$\begin{aligned}\frac{dm_\mu}{dt} + m_\mu = {}& \frac{1}{2N}\sum_i \tanh\left(\beta m_\mu + \beta \sum_{\nu\neq\mu,\rho} \xi_i^\nu m_\nu + \beta(m_\rho + \gamma)\right) \\ &+ \frac{1}{2N}\sum_i \tanh\left(\beta m_\mu - \beta \sum_{\nu\neq\mu,\rho} \xi_i^\nu m_\nu - \beta(m_\rho + \gamma)\right),\end{aligned}$$

   and when $\mu = \rho$,

   $$\frac{dm_\rho}{dt} + m_\rho = \frac{1}{N}\sum_i \tanh\left(\beta(m_\rho + \gamma) + \beta \sum_{\nu\neq\rho} \xi_i^\rho \xi_i^\nu m_\nu\right).$$

   (10 marks)

3. Assume that m_1 and m_ρ are both $\mathcal{O}(1)$ and the rest of the m_μ are negligibly small. Show that m_1 and m_ρ evolve according to

$$\begin{aligned}\frac{dm_1}{dt} + m_1 &= \frac{1}{2}\big[\tanh\big(\beta(m_1+m_\rho+\gamma)\big) + \tanh\big(\beta(m_1-m_\rho-\gamma)\big)\big]\\ \frac{dm_\rho}{dt} + m_\rho &= \frac{1}{2}\big[\tanh\big(\beta(m_\rho+\gamma+m_1)\big) + \tanh\big(\beta(m_\rho+\gamma-m_1)\big)\big]\,.\end{aligned}$$

(5 marks)

4. Let $\beta \to \infty$, so that $\tanh\beta \to \text{sign}$.

   (a) Show that m_1 and m_ρ can't both be zero or both be nonzero: if one is zero the other is active, and vice-versa.
   (5 marks)

   (b) Show that when $0 < \gamma < 1$, then either m_1 or m_ρ (but, of course, not both) can be zero.
   (5 marks)

   (c) Show that when $\gamma > 1$, m_1 must be zero and m_ρ must be $+1$.
   (5 marks)

5. Justify the assumption that m_ν can be negligibly small when $\nu \neq 1, \rho$.

   (5 marks)

Solutions

1. Multiply both sides of the equation for s_i by ξ_i^μ and sum over i, and take advantage of the definition of the overlaps; this gives

$$\frac{dm_\mu}{dt} + m_\mu = \frac{1}{N}\sum_i \xi_i^\mu \tanh\left(\beta\sum_\nu \xi_i^\nu m_\nu + h_i\right)\,.$$

Using the fact that when $\xi_i = \pm 1$, $\xi_i\tanh(x) = \tanh(\xi_i x)$, we can bring the factor of ξ_i^μ inside the tanh, giving us

$$\frac{dm_\mu}{dt} + m_\mu = \frac{1}{N}\sum_i \tanh\left(\beta\sum_\nu \xi_i^\mu\xi_i^\nu m_\nu + \xi_i^\mu h_i\right)\,.$$

Finally, separating out the $\nu = \mu$ and $\nu \neq \mu$ terms, we get the desired result.

2. When $h_i = \beta\gamma\xi_i^\rho$, we have

$$\frac{dm_\mu}{dt} + m_\mu = \frac{1}{N}\sum_i \tanh\left(\beta\xi_i^\mu\sum_{\nu\neq\rho}\xi_i^\nu m_\nu + \beta\xi_i^\mu\xi_i^\rho(m_\rho+\gamma)\right)\,.$$

Note that we now include the $\nu = \mu$ term in the sum over ν; we'll separate it out later.

When $\mu \neq \rho$, the term $\xi_i^\mu\xi_i^\rho$ is either $+1$ or -1, both with probability 1/2. When $\mu = \rho$, the term $\xi_i^\mu\xi_i^\rho$ is just $+1$. In either case, we get the desired result. With a small amount of algebra.

3. When $\mu = 1$, we just throw out all the terms with $\nu \neq 1, \rho$. When $\mu = \rho$, it's only slighlty harder: we just need to realize that $\xi_i^\rho \xi_i^1$ is either $+1$ or -1, both with probability $1/2$.

4. When $\beta \to \infty$, the equations for m_1 and m_ρ can be written

$$\begin{aligned}
\frac{dm_1}{dt} + m_1 &= \frac{1}{2}\big[\text{sign}\big(m_1 + m_\rho + \gamma\big) + \text{sign}\big(m_1 - m_\rho - \gamma\big)\big] \\
\frac{dm_\rho}{dt} + m_\rho &= \frac{1}{2}\big[\text{sign}\big(m_\rho + \gamma + m_1\big) - \text{sign}\big(m_1 - m_\rho - \gamma\big)\,.
\end{aligned}$$

   (a) Because of the $+$ and $-$ signs that I've highlighted in red, one of the terms in brackets must be zero and one must be ± 1.

   (b) If $0 < \gamma < 1$, then $m_1 \pm \gamma$ has the same sign as m_1, and similarly for m_ρ. As far as the eqilibria go, then, we can eliminate γ from the equations, and it becomes clear that either m_1 or m_ρ can be zero, and if one is zero then the other isn't.

   (c) If $\gamma > 1$, then $m_1 + \gamma$ and $m_1 - \gamma$ have different signs (as is easy to see just by checking the two possibilities for m_1). Consequently, m_1 must be zero. We also have $m_\rho + \gamma > 0$, which forces m_ρ to be $+1$.

5. Let's consider, for definiteness, the equation for m_2,

$$\begin{aligned}
\frac{dm_2}{dt} + m_2 = &\frac{1}{2N}\sum_i \tanh\left(\beta m_2 + \beta \sum_{\nu\neq 1,2,\rho} \xi_i^2 \xi_i^\nu m_\nu + \beta(\xi_i^2\xi_i^1 m_1 + m_\rho + \gamma)\right) \\
&+ \frac{1}{2N}\sum_i \tanh\left(\beta m_2 + \beta \sum_{\nu\neq 1,2,\rho} \xi_i^2 \xi_i^\nu m_\nu - \beta(\xi_i^2\xi_i^1 m_1 + m_\rho + \gamma)\right) \\
&= \frac{1}{4N}\sum_i \tanh\left(\beta m_2 + \beta \sum_{\nu\neq 1,2,\rho} \xi_i^2 \xi_i^\nu m_\nu + \beta(m_1 + m_\rho + \gamma)\right) \\
&+ \frac{1}{4N}\sum_i \tanh\left(\beta m_2 + \beta \sum_{\nu\neq 1,2,\rho} \xi_i^2 \xi_i^\nu m_\nu + \beta(-m_1 + m_\rho + \gamma)\right) \\
&+ \frac{1}{4N}\sum_i \tanh\left(\beta m_2 + \beta \sum_{\nu\neq 1,2,\rho} \xi_i^2 \xi_i^\nu m_\nu - \beta(m_1 + m_\rho + \gamma)\right) \\
&+ \frac{1}{4N}\sum_i \tanh\left(\beta m_2 + \beta \sum_{\nu\neq 1,2,\rho} \xi_i^2 \xi_i^\nu m_\nu - \beta(-m_1 + m_\rho + \gamma)\right)\,.
\end{aligned}$$

   In the absence of fluctuations, a self-consistent solution is $m_\nu = 0$ when $\nu \neq 1, \rho$. There are, of course, fluctuaitons, but they'll be $\mathcal{O}(\sqrt{p/N})$. So long as $p \ll N$, as assumed here, those fluctuations will be negligible.

3 Coding

Consider a population of neurons which encode the orientation θ of a bar. The population can be broken down into A subpopulations of cells, each of size N_a and each tuned to a common orientation θ_a. Suppose that each cell fires n_{ai} spikes with a probability distribution given by $P(n_{ai}|f(\theta - \theta_a))$ where f is a common tuning curve shape.

Assuming that all values of θ are equally likely *a priori* and that the population tuning is dense enough (that is, A is large enough) to encode all values of θ equally well, derive an asymptotic relationship between the mutual information $I(\theta; \{n_{ia}\})$ and the Fisher information in the population.

Specifically, show that:

1. as the population size for each a grows, the posterior distribution on θ approaches
$$P(\theta|\{n_{ia}\}) \to \mathcal{N}(\theta^*, 1/\sum_a N_a J_a(\theta^*)),$$
where θ^* is the maximum-likelihood orientation and $J_a(\theta^*)$ is the Fisher information conveyed by the sub-population tuned to θ_a;
2. and thus that
$$I(\theta; \{n_{ia}\}) \to \log \pi - \tfrac{1}{2} \log 2\pi e/J$$
where J is the population Fisher information.

Now assume that $f = [f_{max} - \rho|\theta - \theta_a|]^+$ (where $[x]^+ = \max(x, 0)$), and that $P(n|f)$ is Poisson with mean f.

3. Using the result above, detemine the value of ρ that would maximize the mutual information.

Finally,

4. discuss two circumstances in which the above relationship might break down, even for very large populations.

4 Learning

Consider a neuron whose membrane potential, V, is the sum of (very simplified) PSPs,

$$V(t) = \sum_i w_i(t) x_i(t)$$

where we'll work in discrete time ($t = 1, 2, ...$) and $x_i(t)$ tells us whether or not there's a spike on presynaptic neuron i,

$$x_i(t) = \begin{cases} 1 & \text{neuron } i \text{ spikes on timestep } t \\ 0 & \text{otherwise}\,. \end{cases}$$

We'll assume that the true input-output mapping is of the form

$$V^*(t) = \sum_i w_i^* x_i(t).$$

Finally, at each timestep the neuron gets feedback, $\delta(t)$, given by

$$\delta(t) = V^*(t) - V(t) + \xi(t)$$

where $\xi(t)$ is a zero mean Gaussian random variable with variance σ_δ^2. It is independent from one time step to the next.

1. The goal of the neuron is to compute the posterior distribution over the true weights, $\mathbf{w}^*$, given the data, where the data, $d(t)$, that the neuron receives on each time step is

   $$d(t) = \{\mathbf{x}(t), \mathbf{w}(t), V(t), \delta(t)\}\,.$$

   The neuron has to find this posterior online, for which the hidden Markov formalism is ideal. Show that the posterior distribution at time t is given in terms of the posterior distribution at time $t-1$ via

   $$p\big(\mathbf{w}^* | d(1), d(2), ..., d(t)\big) \propto p\big(d(t)|\mathbf{w}^*\big) p\big(\mathbf{w}^* | d(1), d(2), ..., d(t-1)\big)\,.$$

   (10 marks)

2. Show that the likelihood, $p\big(d(t)|\mathbf{w}^*\big)$, can be simplified,

   $$p\big(\mathbf{w}^* | d(1), d(2), ..., d(t)\big) \propto p\big(\delta(t)|\mathbf{w}^*, \mathbf{x}(t), \mathbf{w}(t)\big) p\big(\mathbf{w}^* | d(1), d(2), ..., d(t-1)\big)\,.$$

   (10 marks)

3. The likelihood is a Gaussian function of $\mathbf{w}^*$; consequently, if $p\big(\mathbf{w}^* | d(1), d(2), ..., d(t-1)\big)$ is Gaussian, then $p\big(\mathbf{w}^* | d(1), d(2), ..., d(t)\big)$ will also be Gaussian. Let

   $$p\big(\mathbf{w}^* | d(1), d(2), ..., d(t)\big) \propto \exp\left[-\frac{1}{2}(\mathbf{w}^* - \boldsymbol{\mu}(t)) \cdot \boldsymbol{\Sigma}^{-1}(t) \cdot (\mathbf{w}^* - \boldsymbol{\mu}(t))\right]\,.$$

   Write down an expression that relates $\boldsymbol{\mu}(t)$ and $\boldsymbol{\Sigma}(t)$ to $\boldsymbol{\mu}(t-1)$ and $\boldsymbol{\Sigma}(t-1)$. It's not necessary to solve for these variables; just explain how you would solve for them.

   (10 marks)

4. So far we have assumed that V^* was a linear function of $\mathbf{x}$. However, far more typical is for $V^*(\mathbf{x})$ to be a nonlinear function of $\mathbf{x}$. In this case the error signal, $\delta(t)$, is

$$\delta(t) = V^*(\mathbf{x}(t)) - \mathbf{w}(t) \cdot \mathbf{x}(t) + \xi \, . \tag{5}$$

How would you proceed? This is an open-ended question with no clear answer that I'm aware of. So pick an approach and discuss its pros and cons.

(10 marks)

Solutions

1. Using Bayes' theorem, we have

$$\begin{aligned} p\big(\mathbf{w}^* | d(1), d(2), ..., d(t)\big) &\propto p\big(d(t) | \mathbf{w}^*, d(1), d(2), ..., d(t-1)\big) p\big(\mathbf{w}^* | d(1), d(2), ..., d(t-1)\big) \\ &= p\big(d(t) | \mathbf{w}^*\big) p\big(\mathbf{w}^* | d(1), d(2), ..., d(t-1)\big) \, . \end{aligned}$$

The second equality follows from the Markov property: conditioned on the true weights, the current data does not depend on past data.

2.

$$\begin{aligned} p\big(d(t) | \mathbf{w}^*\big) &= p\big(\delta(t) | \mathbf{w}^*, \mathbf{x}(t), \mathbf{w}(t)\big) \\ &= p\big(\mathbf{x}(t), \mathbf{w}(t), V(t), \delta(t) | \mathbf{w}^*\big) \\ &= p\big(\delta(t) | \mathbf{w}^*, \mathbf{x}(t), \mathbf{w}(t), V(t)\big) p\big(\mathbf{x}(t), \mathbf{w}(t), V(t) | \mathbf{w}^*\big) \, . \end{aligned}$$

Without an error signal, neither $\mathbf{x}(t)$, $\mathbf{w}(t)$ nor $V(t)$ provide any information about the true weights, $\mathbf{w}^*$. Thus, the term $p\big(\mathbf{x}(t), \mathbf{w}(t), V(t) | \mathbf{w}^*\big)$ is independent of $\mathbf{w}^*$, and so, because we are after the posterior over $\mathbf{w}^*$, it can be dropped.

3. Taking the logs of both sides of our update rule, we have

$$\begin{aligned} -\frac{1}{2}(\mathbf{w}^* - \boldsymbol{\mu}(t)) \cdot \boldsymbol{\Sigma}^{-1}(t) \cdot (\mathbf{w}^* - \boldsymbol{\mu}(t)) \sim& \log p\big(\delta(t) | \mathbf{w}^*, \mathbf{x}(t), \mathbf{w}(t), V(t)\big) \\ & - \frac{1}{2}(\mathbf{w}^* - \boldsymbol{\mu}(t-1)) \cdot \boldsymbol{\Sigma}^{-1}(t-1) \cdot (\mathbf{w}^* - \boldsymbol{\mu}(t-1)) \, . \end{aligned}$$

The log likelihood is Gaussian, giving us

$$\begin{aligned} -\frac{1}{2}(\mathbf{w}^* - \boldsymbol{\mu}(t)) \cdot \boldsymbol{\Sigma}^{-1}(t) \cdot (\mathbf{w}^* - \boldsymbol{\mu}(t)) \sim& -\frac{1}{2} \frac{(\delta(t) - (\mathbf{w}^* - \mathbf{w}(t)) \cdot \mathbf{x}(t))^2}{\sigma_\delta^2} \\ & - \frac{1}{2}(\mathbf{w}^* - \boldsymbol{\mu}(t-1)) \cdot \boldsymbol{\Sigma}^{-1}(t-1) \cdot (\mathbf{w}^* - \boldsymbol{\mu}(t-1)) \, . \end{aligned}$$

To find $\boldsymbol{\mu}(t)$ and $\boldsymbol{\Sigma}(t)$ in terms of $\boldsymbol{\mu}(t-1)$ and $\boldsymbol{\Sigma}(t-1)$, we just have to complete the square on the right hand side. Tedious, but straightforward.

4. There is no really clear way forward, since we don't have "true" weights any more. We could *define* the true weights to be the ones that minimize some distance measure between $V^*(\mathbf{x})$ and $\mathbf{w}^* \cdot \mathbf{x}$. However, this doesn't tell us how to compute the likelihood of the data, $p(\delta(t)|\mathbf{w}^*)$. We might try a hack, and write

$$\begin{aligned}\delta(t) &= V^*(\mathbf{x}(t)) - \mathbf{w}(t) \cdot \mathbf{x}(t) + \xi \\ &= (\mathbf{w}^* - \mathbf{w}(t)) \cdot \mathbf{x}(t) + V^*(\mathbf{x}(t)) - \mathbf{w} * \cdot \mathbf{x}(t) + \xi \\ &\equiv (\mathbf{w}^* - \mathbf{w}(t)) \cdot \mathbf{x}(t) + \xi' \, .\end{aligned}$$

   We could then treat ξ' as Gaussian noise, and proceed as before. This might work, but there are no guarantees.

   This is an open area of research!