Gatsby Computational Neuroscience Unit
Combined Qualifying Examination

19 December 2016

There are 10 questions. Answer all of them, to the best of your ability. Each is worth 10 marks. No reference materials are allowed.

It should take no more than 6 hours

Good luck!

1. Consider N neurons that encode a scalar stimulus value x using a distributed population code with (not necessarily identical) tuning curves $h_n(x)$ and Poisson noise. We follow Ganguli and Simoncelli, seeking the "optimal" code representing a stimulus drawn from a distribution $p(x)$.

   Suppose the tuning curve of the nth neuron can be written in terms of a "standard" symmetric zero-centred tuning curve $h_0()$ defined on a warped version of the stimulus:

   $$h_n(x) = g(x_n) h_0(D(x) - n\delta)$$

   where $D : x \to \xi$ is the monotonic warping function that transforms x to the standard variate ξ; δ is the spacing of tuning curves in ξ-space; and $g(x_n)$ is a gain function which depends on the preferred stimulus for the neuron $x_n = D^{-1}(n\delta)$.

   - First, take $g(x_n) = \mathit{const}$ and $D(x) = x$, and derive the Fisher Information about x carried by the population response. Show that in the limit $\delta \to 0$ with $g(x_n) = c\delta$, this information is independent of x [assuming that the coding space is either unbounded or cyclic]. Call this value J_0.
   - Now, for the general case, show that the Fisher Information is given by

     $$J(x) \approx D'(x)^2 g(x) J_0$$

     under the assumption that $g(x)$ varies slowly.
   - Using this expression, find the functions $D(x)$ and $g(x)$, under the constraints $D(\infty) = N\delta$ and $\int dx\, p(x) \sum_n h_n(x) = R$, that maximise:
     - the average Fisher information

       $$\langle J \rangle = \int dx\, p(x) J(x)$$

     - the Fisher information bound on the (Shannon) mutual information

       $$\langle I \rangle = \int dx\, p(x) \log J(x)$$

   - Interpret the difference between the two results.

2. Consider a log-linear population code for probability distributions (such as a PPC). That is, suppose that a vector of firing rates $\vec{r} = [r_n]$ encodes a probability distribution $q(x)$ on a scalar variable x according to:

   $$\log q(x) = \sum_n r_n h_n(x) - \Phi(\vec{r})\,.$$

   - Relate this encoding scheme to the exponential family of distributions. Identify the natural parameters and the sufficient statistics of the encoded distribution over x.

   The PPC is usually defined by identifying the encoded distribution over x with the result of Bayes' rule:

   $$q(x) \propto p(\vec{r}|x) p(x)\,.$$

   - Derive the form that this identification implies for the distribution $p(\vec{r}|x)$ given the log-linear encoding above.
   - Describe an experiment you might use to test whether the distribution $p(\vec{r}|x)$ takes the correct form. Pay particular attention to the specification of the variable x and the origin of the variability in $\vec{r}$ given x.
   - Log-linear codes facilitate computations such as cue combination. Show how information about the same variable x from two different pathways could be combined efficiently using this representation of uncertainty.

3. Suppose that the (noisy) firing rates $r_n(t)$ of a recorded population of neurons are linearly sensitive to a one-dimensional stimulus time-series $s(t)$, so that

$$E[r_n(t)|\vec{s}(t)] = \lambda_n(t) = \vec{k}_n \cdot \vec{s}(t)\,,$$

where $\vec{s}(t) = [s(t-\tau)\dots s(t)]$ is a windowed vector of length τ extracted from the stimulus at time t.

We seek to model the population using a Poisson GLM with exponential nonlinearity, one time-step self-coupling and instantaneous cross-coupling:

$$\hat{\lambda}_n(t) = \exp\left(\vec{w}_n \cdot \vec{s}(t) + h_n r_n(t-1) + \sum_m h_{nm} r_m(t)\right), \qquad r_n(t) \sim \mathsf{Poisson}[\hat{\lambda}_n(t)]$$

using maximum-likelihood estimation.

Assume that the stimulus is drawn from a stationary zero-mean Gaussian process, and write $E[\vec{s}(t)\vec{s}(t)^T] = \Sigma_0$, $E[\vec{s}(t)\vec{s}(t-1)^T] = \Sigma_{-1}$.

- Show that, in the limit of an infinitely long stimulus, the ML parameters will satisfy the conditions:

$$\left\langle \lambda_n(t)\frac{\partial \log \hat{\lambda}_n(t)}{\partial \theta}\right\rangle_{\vec{s}} = \left\langle \hat{\lambda}_n(t)\frac{\partial \log \hat{\lambda}_n(t)}{\partial \theta}\right\rangle_{\vec{s}}$$

  for each $\theta \in \{\vec{w}_n, h_n, h_{nm}\}$, where the angle brackets indicate expectations over the stimulus distribution.
- Calculate the expectations in the above expression to find conditions that link $\vec{w}_n$, h_n, h_{nm}, $\vec{k}_n$ and the stimulus covariances.
- Interpret your results – how well do the model fits reflect the true generative process?

4. consider a feedforward linear mapping from n-d input $\mathbf{x}$ to m-d output $\mathbf{y}$ via $m \times n$ weight matrix W, so $\mathbf{y} = W\mathbf{x}$. Williams' symmetric error correcting learning rule is a form of predictive coding-based learning, suggesting that

$$\Delta W_{ij} \propto y_i(x_j - [U\mathbf{y}]_j)$$

where U is a $n \times m$ top-down matrix.

- write down the equation for the evolution of the weights, averaging over the statistics of the input patterns $\{\mathbf{x}\}$.
- in the symmetrized version of the rule $U = W^T$. In this case, write down the condition that must be satisfied for W to stop changing.
- writing $P = W^T W$, show that these equilibrium points satisfy $PR = RP = PRP$ for a symmetric matrix R (which you should define)
- if the (orthonormal) eigenvectors $\{\mathbf{e}_\alpha\}$ and eigenvalues $\{\lambda_\alpha\}$ of R are such that $\{\lambda_\alpha\}$ are all different, show that P has the property that $P\mathbf{e}_\alpha = \mathbf{e}_\alpha$ or $P\mathbf{e}_\alpha = 0$.
- interpret what this implies as an end-point of learning for the rule.

5. consider a reward machine that has two states: empty (0) and full (1), with a probability of transitioning from 0 → 1 of $p = 0.2$ per step. An agent can inspect the machine (at cost a); if it finds it to be full, then it gets a reward of $r = 1$ (and the machine goes back to state 0). Otherwise, the machine remains empty. If it is full and isn't inspected, then it remains full.

- if the agent inspects with probability π, what is the long run probability that the machine is in state 0 or 1?
- derive an expression for the inspection frequency π^* that optimizes the long run average net reward per step as a function of a. Sketch how this long run average reward varies with π for indicative values of a, both smaller and greater than 1.
- if the agent instead inspects every n steps on a deterministic schedule, what is the long-run probability that it will find the machine full when it inspects?
- will optimal deterministic inspection lead to greater, equal or lesser long run reward per step than optimal stochastic inspection. Justify your answer.

6. consider an axon growth cone growing in an environment in which the gradient of a guidance molecule points at angle ψ relative to the origin. The cone takes discrete steps of length 1 unit, starting from an anchoring point at the origin, and pointing at an angle $\theta(t)$. Axons like to grow straight, and so, ignoring noise, a reasonable model of the change in $\theta(t)$ over time is:

$$\theta(t+1) = \theta(t) + \alpha\angle(\phi(t), \theta(t)) + \beta\angle(\psi, \theta(t))$$

where $\angle(\alpha, \beta)$ denotes the signed angle between the unit vectors with angles α, β, and constrains the resultant angle to be between $-\pi$ and π, $\phi(t)$ is the angle joining the tip of the axon to the origin, $\alpha > 0$ quantifies the force of straightness (making θ align with ϕ), and $\beta > 0$ quantifies the force of the gradient (making θ align with ψ). Let $\psi = 0$.

   (a) if the axon starts off growing at $\theta(0) = \psi = 0$, write down a formula for $\theta(t)$ as a function of t.

   (b) take the case that the axon starts with $\theta(0) = \pi/2$, and in which the axon is quite long, so that $\psi(t)$ changes much more slowly than $\theta(t)$.

      - at what value does $\theta(t)$ equilibrate (as a function of $\psi(t)$)?
      - assuming that equilibrium is maintained as $\phi(t)$ changes, sketch the form of the evolution of the axon
      - and by writing down an equation for the evolution of $\phi(t)$ and doing a first order approximation, write down a closed-form expression for its long-run value.
      - if the axon can put down more anchor points as it grows, changing the origin to which $\phi(t)$ refers, need this form remain true? why?

7. Why do inhibitory neurons tend to have higher firing rates than excitatory neurons?

   Consider a neuron deep inside a recurrent network. Input from other neurons will produce, on average, a total normalized conductance, g_{tot}, and a total reversal potential, $\mathcal{E}_{tot}$. Under these conditions, a leaky integrate and fire neuron obeys the equation

$$\tau \frac{dV}{dt} = -(V - \mathcal{E}_L) - g_{tot}(V - \mathcal{E}_{tot}) + \text{ zero mean fluctuations} \tag{1}$$

   where τ is the membrane time constant and $\mathcal{E}_L$ is the leak reversal potential.

   Show that if the synaptic time constant is very small compared to the membrane time constant, τ, then for a randomly connected network of excitatory and inhibitory neurons,

$$g_{tot}(V - \mathcal{E}_{tot}) \approx \frac{K_E \tau \nu_E V_{EPSP}}{\mathcal{E}_E - \mathcal{E}_L}(V - \mathcal{E}_E) + \frac{K_I \tau \nu_I V_{IPSP}}{\mathcal{E}_L - \mathcal{E}_I}(V - \mathcal{E}_I) \tag{2}$$

   where K_E and K_I are the average number of excitatory and inhibitory connections *received* by each neuron, V_{EPSP} and V_{IPSP} are the average excitatory and inhibitory post-synaptic potentials in response to a single presynaptic spike (so all other neurons are silent), $\mathcal{E}_E$ and $\mathcal{E}_I$ are the excitatory and inhibitory reversal potentials, and ν_E and ν_I are the population averaged excitatory and inhibitory firing rates.

   Using reasonable values for the above parameters, argue that for the network to operate in the asynchronous regime at a few Hz, the average inhibitory firing rate, ν_I, should be several times larger than the average excitatory rate, ν_E.

8. A leaky integrate and fire neuron with one extra current evolves according to

$$\tau \frac{dV}{dt} = -(V - \mathcal{E}_L) - xg(V - \mathcal{E}) + V_{ext}(t) \tag{3}$$

   where g is constant, x obeys the standard equation

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

   and a spike is emitted when V reaches V_{th}, at which point the voltage is reset to $\mathcal{E}_L$.

   You find, experimentally, the following. First you introduce a step increase in voltage,

$$V_{ext}(t) = V_0 \Theta(t),$$

   and no spike is emitted. Then you inject negative current into the cell for a long time, and when you stop injecting the current the neuron emits a burst of spikes for about 1 second.

   Explain, qualitatively, how this could happen. By "qualitatively," I mean choose approximate values for the parameters, and provide a sketch of $x_\infty(V)$, that will produce the observed experimental results.

9. Consider the standard Wilson-Cowan model,

$$\tau \frac{d\nu_E}{dt} = \phi_E(W_{EE}\nu_E - W_{EI}\nu_I - \theta_E) - \nu_E$$
$$\tau \frac{d\nu_I}{dt} = \phi_I(W_{IE}\nu_E - W_{II}\nu_I - \theta_I) - \nu_I$$

Both ϕ_E and ϕ_I are sigmoidal, and when $W_{EE} = W_0$ and $\theta_E = 0$, this set of equations has the usual nullclines (a plot is shown below).

Consider two kinds of adaptation: spike-frequency adaptation,

$$\tau_a \frac{d\theta_E}{dt} = \alpha\nu_E(\theta_0 - \theta_E) - \theta_E$$

($\theta_0 > 0$) and synaptic adaptation,

$$\tau_a \frac{dW_{EE}}{dt} = (W_0 - W_{EE}) - \alpha\nu_E W_{EE}.$$

In both cases, $\tau_a \ll \tau$. Show that for sufficiently large α, spike frequency adaptation can lead to up-down states, but no matter how large α is, synaptic adaptation cannot.

The first one is pretty easy. The second one is harder: you need to show that as W_{EE} decreases, the amount the excitatory nullcline drops at fixed ν_E is an increasing function of ν_E.

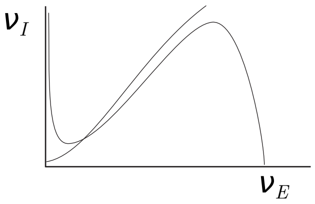

Figure 1: The usual nullclines.

10. Consider the standard Hopfield model,

$$x_i(t+1) = \text{sign}\left(\frac{1}{N}\sum_{j,\mu} \xi_i^\mu \xi_j^\mu x_j(t)\right)$$

where the sum on j runs from 1 to N (there are N neurons) and the sum on μ runs from 1 to p (there are p memories). At time t,

$$x_i = \begin{cases} \xi_i^\mu & \text{with probability } 1-q \\ -\xi_i^\mu & \text{with probability } q\,. \end{cases}$$

If you studied hard for the qualifier, you would know that in the large N and p limit,

$$P\big(x_i(t+1) = \xi_i^\mu\big) = \Phi\left(\frac{1-2q}{(p/N)^{1/2}}\right)$$

where Φ is the cumulative normal function. Assuming errors are totally random, write down a self-consistent expression for q. Under the totally random assumption, solve for the capacity of the network. If you did that correctly, you would find a capacity larger than the usual one, $p/N = 0.138$. Why is that?