Gatsby Computational Neuroscience Unit
Neuroscience Candidacy 2009

Written Examination
26 Jan 2009

Part I

This part has 20 short questions. Answer all of them, to the best of your ability. Each is worth 4 marks. No reference materials are allowed.

This part should take 1 hour. You may continue to work for another 30 minutes once that time is up, but indicate clearly which answers (or parts of answers) were written afterward.

1. For a Hodgkin-Huxley neuron, the sodium current is given by

$$I_{Na} = g_{Na} m^3 h (V - \mathcal{E}_{Na}) \,.$$

Consider a very simplified model of a spike,

$$\begin{aligned} t < 0 \text{ ms}: &\qquad m = 0, h = 1 \\ t = 0 \text{ ms}: &\qquad m \to 1 \\ t = 1 \text{ ms}: &\qquad h \to 0 \,. \end{aligned}$$

Assume that during the spike, $V(t)$ rises linearly from V_{rest} to $\mathcal{E}_{Na}$ in 1 ms.

Write down an expression for the total current that flows into the cell between the times 0 and 1 ms.

2. The propagation speed of an action potential in axons is linear in the radius if they are myelinated and proportional to the square root of the radius if they are unmyelinated. In what sense does this "explain" the fact that thin axons tend to be unmyelinated and thick ones tend to be myelinated?

3. For the following neurotransmitters, give the approximate reversal potential and time constant: AMPA, NMDA, GABA_A, GABA_B.

4. Provide a biophysical explanation for synaptic facilitation.

5. A particular synapse invariably fails to release a vesicle for every **second** presynaptic spike. Assume that the presynaptic spike-train is drawn from a homogeneous Poisson process. What is the coefficient of variation (CV) of the interval between synaptic release events?

6. A self-exciting point process with counting process $N(t)$ and event times t_i has intensity

$$\lambda(t) = \lambda_0 (1 - e^{-(t - t_{N(t)})/\tau})$$

Give a function $r(t)$ (in terms of λ_0, $t_{N(t)}$ and $r(t_{N(t)})$) such that the $r_i = r(t_i)$ appear Poisson distributed.

7. Suggest an experimental setting in which we might choose to fit a multidimensional LNP model by maximising mutual information rather than by diagonalising the spike-triggered covariance.

8. Does a population code with broad tuning curves always represent values less accurately than one with narrower tuning? Explain why (not).

9. Suppose that h_i, the steady state synaptic drive to a cell, is given by

$$h_i = \frac{1}{K^{1/2}} \sum_{j=1}^{N} c_{ij} w_{ij} \nu_j \,.$$

Here ν_j is the firing rate of neuron j, w_{ij} is the connection strength from neuron j to neuron i, c_{ij} is a Bernoulli random variable,

$$c_{ij} = \begin{cases} 1 & \text{with probability } c \\ 0 & \text{with probability } 1 - c \,, \end{cases}$$

and $K \equiv cN$ is the average number of connections per neuron. Assume that c_{ij}, w_{ij} and ν_j are uncorrelated and drawn iid. Write down an expression for the mean and variance of h_i (with respect to index, i) in terms of: K, the first and second moments of the firing rates, and the mean and variance of w_{ij}. As usual, assume that $N \gg 1$.

10. Consider a standard Hopfield model with sparse connectivity,

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

where c_{ij} is the same Bernoulli random variable as in the previous short question ("sparse" means $c < 1$; typically it's about 0.1) and the ξ_i^μ are, as usual, uncorrelated random variables that take on two values, $+1$ and -1, each with probability $1/2$. Use simple arguments about signal and noise to show that the capacity of the network (the maximum number of memories, p, that can be stored) is proportional to K ($\equiv cN$).

11. In our mean field analysis, we always assume that neurons operate in the asynchronous regime, meaning they fire with more or less Poisson statistics, and their firing times are uncorrelated. Why is this assumption important?

12. Consider the following differential equation, written in polar coordinates,

$$\begin{aligned}\frac{dr}{dt} &= r - 2r^2 + 1.01r^3\\ \frac{d\theta}{dt} &= \omega\,.\end{aligned}$$

Sketch the trajectories in the x-y plane, with initial conditions $r \approx 0.1$ and $\theta = 0$. Note that r and θ are related to x and y via $x = r\cos\theta$ and $y = r\sin\theta$.

13. What forms of learning rules are called covariance rules? Show why.

14. What is the baseline firing rate of dopamine neurons? Discuss two implications this has for the representation of negative prediction errors.

15. How does Oja's rule determine the norm of the weights? What learning rule would force the norm to be constant throughout learning?

16. How should uncertainty affect a Bayesian learner?

17. A simple cell with a receptive field that is separable in space and time cannot report the direction of motion of a bar. What about a population of such cells? [Support your answer. A simple "yes" or "no" will not receive any marks.]

18. Ungerleider and Mishkin (1982) and Goodale and Milner (1992) have each proposed different accounts for the function of the dorsal and ventral visual processing streams. Say briefly what these proposals are.

19. What is whole cell recording? How might the sample of cells measured by whole cell methods be biased, and how would this bias differ from that of conventional extracellular recording?

20. List 4 different neuromodulatory systems, and a hypothesised computational role for each.

Gatsby Computational Neuroscience Unit
Neuroscience Candidacy 2009

Written Examination
27 Jan 2009

Part II

This part contains 12 questions, of which you need to answer 10. Each is worth 12 marks.

This part should take 2 hours. You may continue to work for another 2 hours once this time is up, but indicate clearly which answers (or parts of answers) were written afterward.

1. When we analyzed two-dimensional models of both single neurons and networks, we always encountered bistability. This bistability led to N-shaped nullclines (recall, for example, the excitatory nullcline associated with a randomly connected network of excitatory and inhibitory neurons).

   The question we are interested in here is: is this bistability important? To answer this, consider a two variable model of a neuron that has the form

$$\begin{aligned} \frac{dv}{dt} &= f(v,w) \\ \frac{dw}{dt} &= g(v,w)\,. \end{aligned}$$

   Suppose that both nullclines, when plotted in the u-v plane, are monotonic functions of u. Can such a system exhibit repetitive spiking? Can it exhibit all-or-none spiking?

   The second question, can it exhibit all-or-none spiking, is hard, so even if you don't answer it fully, you may get full credit for telling us how you would approach it.

2. Consider a synapse that exhibits both depression and facilitation: whenever there is a spike the release probability gets a positive kick (because of a buildup of calcium), and whenever there is a spike *and* neurotransmitter is released, the release probability gets, in addition to the positive kick, a negative one (because of resource depletion). Letting p be the release distribution, we can put this all into one equation,

$$\frac{dp}{dt} = \frac{p_0 - p}{\tau} + [f_F(1-p) - (1-f_D)xp] \sum_i \delta(t - t_i)$$

   where the t_i are spike times and, when updating the right hand side, you should use the value of p immediately before a spike. What's new is the factor of x that appears in the depressing piece. That variable signals release: it is 1 if neurotransmitter is released and 0 otherwise. The probability that $x = 1$ is p.

   Suppose I told you the probability distribution of p immediately before a spike (call it $P(p)$). Write down an expression for the probability distribution a time t after the spike, assuming no new spikes intervene.

3. Consider the time-dependent synaptic drive to a neuron,

$$h_i(t) = \frac{1}{N^{1/2}} \sum_{j=1}^{N} w_{ij} s_j(t)\,.$$

   Here $s_j(t)$ is a set of PSPs associated with spikes on the presynaptic neuron: letting t_j^k be the time of the k^{th} spike on neuron j, $s_j(t)$ is given by

$$s_j(t) = \sum_k v(t - t_j^k)\,,$$

   where $v(t)$ is nonzero for $t < 0$, integrates to 1, and is local in time (e.g., $v(t) = \Theta(t)\exp(-t/\tau)/\tau$). Assuming that spiking is Poisson, spike times are uncorrelated, and the w_{ij} are $\mathcal{O}(1)$, show that the temporal and spatial fluctuations in h_i are both $\mathcal{O}(1)$. (Spatial fluctuations refer to fluctuations with respect to the index, i.)

4. Consider a Hopfield network in which the neurons obey Dale's law,

$$\begin{aligned} x_i(t+1) &= \text{sign}\left[\frac{1}{N}\sum_{j=1}^{N} J_{ij}x_j(t)\right] \\ J_{ij} &= f_j\left(\sum_{\mu=1}^{p}\xi_i^\mu\xi_j^\mu\right) \end{aligned}$$

where, as usual, the ξ_i^μ are uncorrelated random variables that take on two values, $+1$ and -1, each with probability $1/2$. What's new is the function f_j. It's a threshold linear function that picks out either the positive or negative part of the weights. Specifically, for $j = 1, ..., N/2$,

$$f_j(w) = \max(0, w)$$

and for $j = N/2 + 1, ..., N$,

$$f_j(w) = \min(0, w)\,.$$

This ensures that the sign of J_{ij} is independent of i.

Show, based on arguments about signal and noise, that the capacity of the network should be about 2 times smaller than that of the usual Hopfield network (in which $f_j(w) = w$). Assume $N \gg 1$.

5. Discuss the ways in which receptive fields and tuning properties change at each stage as you move along the ventral visual stream from retina to inferior temporal cortex. Do neurons become more selective or more invariant as you move up the hierarchy?

6. What is a generalised linear model (GLM)? Why have they been useful for modelling spike trains?

   Write down the form of a GLM and sketch (or give expressions) for parameters that could model

   (a) a bursty neuron

   (b) an oscillatory neuron

   (c) an off-responsive neuron

7. It is common to use Fisher Information as a measure of the "efficiency" of a population code. Give three different reasons why it might be an inappropriate tool for this purpose. In each case, describe a population code with high Fisher Information but poor representational properties, or vice versa. Intuitive arguments are sufficient — you don't need to explicitly calculate the FI in each case.

8. Describe intuitively Atick and Redlich's functional explanation for why retinal ganglion cells lose their inhibitory surrounds at low light levels. How do you think that the theory might need to change to account for correlated neural noise?

9. What are the SARSA and Q-learning? Discuss their differences. How have they been told apart using the activity of dopamine neurons?

10. Consider a case for immediate reinforcement learning with two actions $a \in \{0, 1\}$ leading to stochastic rewards r^a with means $\overline{r}^0 = \langle r^0 \rangle; \overline{r}^1 = \langle r^1 \rangle$ and a probabilistic choice rule $P[a = 1] = \sigma(\beta(m^1 - m^0))$ based on propensities m^a, where β is an inverse temperature and $\sigma(z) = 1/(1 + \exp(-z))$ is the logistic sigmoid function. Two different actor learning rules are:

$$\textbf{A:} \quad \Delta m^a = \begin{cases} \epsilon(1 - P[a])(r^a - \psi) & \text{if } a \text{ is chosen} \\ -\epsilon P[a]\ (r^{1-a} - \psi) & \text{if } 1 - a \text{ is chosen,} \end{cases}$$

and equivalently for Δm^{1-a}, where ψ is an arbitrary reinforcement comparison term, and

$$\textbf{B:} \quad \Delta m^a = \begin{cases} \epsilon(r^a - E[r; \mathbf{m}]) & \text{if } a \text{ is chosen} \\ 0 & \text{otherwise} \end{cases}$$

where $E[r; \mathbf{m}]$ is the expected reward using the current propensities.

Show that A and B both implement stochastic gradient ascent on $E[r; \mathbf{m}]$. For rule B, assuming $E[r; \mathbf{m}]$ is constant, derive the kernel relating the reward on trial t to the choice on trial $t + \tau$. How would this rule perform in the face of changing reward contingencies?

11. Define advantage learning. Given two actions $a \in \{0, 1\}$ with mean rewards $E[r^0] > E[r^1] > 0$, sketch the trajectory of the advantage values for the actions assuming they start at 0. Is there any way that the activity of dopamine neurons in reinforcement learning contexts can be seen as being consistent with this? Advantage learning has been suggested as working better than Q-learning for a case with very small timesteps. Can you suggest why?

12. Describe the topological structure of the representation of ocular and angular preference over primary visual cortex of the macaque monkey. What happens to this structure near the fovea? The V1-V2 border? The blind spot? Is this map the smoothest representation of these variables - justify your answer? What are cytochrome oxidase blobs? Where do they tend to be located on the map? What have cells in the blobs been considered to code?