Gatsby Computational Neuroscience Unit
Neuroscience Candidacy 2011

Written Examination
26-27 Jan 2011

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.

It should take no more than 3 hours.

1. A passive neuron has a membrane resistance of 100 MΩ and a time constant of 10 ms. At time $t = 0$ the neuron is 1 mV above its resting membrane potential. If there is no other input after time $t = 0$, what's the total charge that flows out of the cell as it relaxes back to rest (i.e., in the limit $t \to \infty$)?

2. As the specific membrane resistance of a dendrite increases, what happens to the electrotonic length? To the time constant?

3. Consider a standard Hodgkin-Huxley neuron that receives constant input sufficient to make it fire at 5 Hz. Plot firing rate versus the time constant of the m-channel (recall that the active sodium current is proportional to m^3h).

4. Consider a standard Hodgkin-Huxley neuron that receives an additional current of the form $I = -xy(V - V_0)$ where x and y are gating variables, V is the membrane potential and V_0 is the reversal potential. What value of V_0, and what behavior of x and y, would lead to an after-hyperpolarization?

5. A simple model for a network of N neurons is

$$x_i(t+1) = \text{sign}\left(\sum_{j=1}^{N} w_{ij}x_j(t) + I_i(t)\right) .$$

   A friend computes the population averaged value of x ($\equiv (1/N)\sum_i x_i$), and finds that it's stationary and equal to -0.8. What is the population averaged variance of x?

6. Consider the standard Hopfield model,

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

   where $\xi_i^\mu = \pm 1$, both with probability 1/2. Assume that the ξ_i^μ are correlated: $p(\xi_i^\mu = \xi_i^\nu) = (1+\epsilon)/2$, $\mu \neq \nu$ and $\epsilon > 0$. Show that as $N \to \infty$, perfect recall is possible if $(p-1)\epsilon < 1$ and impossible if $(p-1)\epsilon > 1$.

   This is wrong: I thought the mean cross-talk was $(p-1)\epsilon$; in fact, it's zero (because of the sum over memories; e.g., the sum on μ). It's variance, though, is $(p-1)\epsilon^2$. Thus, as $N \to \infty$, the capacity should scale, at least naively, as ϵ^{-2}.

7. You are planning to fit a linear model to relate the local field potential recorded in V1 to a natural visual stimulus. You intend to use simple regularised linear regression, but a colleague argues that as the stimulus was not Gaussian you must use MID (or something similar) instead to be unbiased. Explain why he's wrong.

8. Zhang and Sejnowski argued that for a population encoding a single 2D stimulus using circularly symmetric tuning curves, the Fisher information was broadly independent of the tuning width. How (and why) do you think this result would change if the tuning curves were allowed to be elliptical and non-uniform, with separate parameters for the major and minor axes and orientation? [You won't have time to derive the answer—argue instead from intuition.]

9. The accuracy of single-neuron-derived choices is reported sometimes as $d' = Z(\text{hit rate}) - Z(\text{false alarm rate})$, where $Z(p)$ is the inverse of the cumulative Gaussian distribution, and sometimes as the area under the ROC curve. Which measure is preferable, and why?

10. Give two biophyical properties of realistic neuronal firing that violate the assumptions of the spike-response GLM model:

$$\lambda(t|a[0,t), N[0,t)) = \exp\left[\sum_\tau \alpha_\tau\, a(t-\tau) + \sum_\tau \beta_\tau\, dN(t-\tau) + \gamma\right]$$

    where $a(t)$ is stimulus, $N(t)$ is the discretised point-process counting function (so that dN is 1 in bins with a spike and 0 otherwise) and α, β and γ are regression coefficients.

11. Suggest at least two different computational uses that might be served by the apparent variability of sensory neuronal responses.

12. A randomly connected network is operating in the balanced regime. What happens to the average firing rate of the excitatory and inhibitory populations if the inhibitory-inhibitory connections increase? Justify your answer.

13. Consider two brains: one in which the NMDA receptors are much faster than normal (say on the order of 10 ms) and one in which they are normal (on the order of 100 ms). Assuming both work reasonably well, could you say something about the relative importance of spike timing for learning in the two brains?

14. Define the Jeffress model for sound localization. If one was endowed with a small head and an undue fondness for high frequency sounds, how would it break?

15. What is the BCM learning rule? What is needed to make it more competitive than conventional Hebbian learning, and why?

16. Can a credible prior over the environment generate a form of recency in memory recall? How about primacy? Justify your answers.

17. Why does simultaneous contrast seem to be an anti-Bayesian effect?

18. Consider building a Hopfield-like network, to report reasonably faithfully on the familiarity of a stimulus. Roughly how much larger might the capacity be for this than for recall? Assume that the connections are dense and the patterns are *not* sparse.

19. Why would naive Hebbian learning not lead to the formation of ocular dominance over development?

20. Define the conditioning paradigms called forwards and backwards blocking. If a subject was suitably Bayesian, how would her inferences differ in the two cases?

Gatsby Computational Neuroscience Unit
Neuroscience Candidacy 2011

Written Examination
28 Jan 2011

Part II

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

It should take no more than 5 hours.

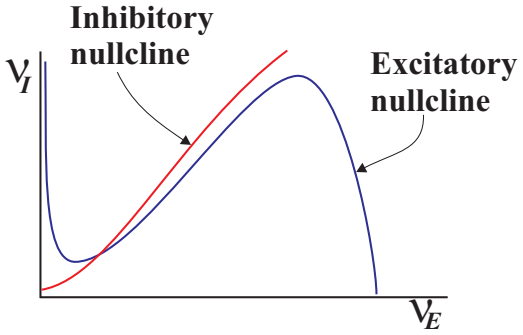
1. An infinitely long passive cable receives a delta-function pulse of current, in both space and time, at time $t = 0$: $I(t) \propto \delta(t)\delta(x)$. What is the effect of doubling the axial resistance? To answer this, sketch voltage versus time at $x = 0$, $x = \lambda$ and $x = 2\lambda$, where λ is the electrotonic length, for two axial resistances, one twice the size of the other.

2. A neuron is firing repetitively and regularly without input. Explain, based on ionic currents, how an *excitatory* current applied after a spike can *delay* the time of the next spike.

3. Consider a firing rate model of the usual form,

$$\begin{aligned} \tau \frac{d\nu_E}{dt} &= \phi_E \left(W_{EE}\nu_E - W_{EI}\nu_I \right) - \nu_E \\ \tau \frac{d\nu_I}{dt} &= \phi_I \left(W_{IE}\nu_E - W_{II}\nu_I \right) - \nu_I \,, \end{aligned}$$

   with the usual nullclines,



   Plot the equilibrium firing rate versus W_{EE} as W_{EE} decreases (all the way to zero). Sketch the nullclines just after the excitatory nullcline breaks into two pieces, and when $W_{EE} = 0$.

4. Consider the standard Hopfield model,

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

   where $\xi_i^\mu = \pm 1$, both with probability $1/2$. The capacity of such a network is about $0.14N$. When that capacity is exceeded ($p > 0.14N$), blackout occurs, and none of the memories are correctly recalled. To avoid blackout, the following online learning rule has been proposed: when pattern p is presented, the weights are updated according to

$$J_{ij}^p = \text{Trunc}\left(J_{ij}^{p-1} + \frac{1}{N} \xi_i^p \xi_j^p, \theta \right)$$

   where the function $\text{Trunc}(x, \theta)$ truncates at $\pm\theta$:

$$\text{Trunc}(x,\theta) = \begin{cases} -\theta & x < -\theta \\ x & -\theta \le x \le \theta \\ \theta & x > \theta \,. \end{cases}$$

   Why does this learning rule avoid blackout? What is a reasonable setting for θ? ("Reasonable" = large enough so that memories are not quickly forgotten but small enough that the probability of blackout is small.)

5. The latency with which V1 neurons fire in response to the sudden onset of a visual stimulus on a blank screen depends on the stimulus, and various authors have proposed that the resulting *pattern* of spikes across the area might be exploited as a neural code. The simplest form of this idea uses $\{t_1^{(i)}\}$—the times of the first post-stimulus spike in each neuron labelled by by i—as the information-carrying element. Discuss at least three challenges posed to this simple idea, and sketch ways in which they might possibly be overcome.

6. The figure below is taken from a paper that appeared in *J Neuronphysiol* in the mid-90s. The authors studied the activity of M1 cells while monkeys reached from a fixed central position to one of 48 target locations, arranged in evenly-spaced rings of 8 targets (angular position θ) at each of 6 different radial distances d. The monkey performed each of the 48 different reaching movements multiple times whilst they recorded the activity of a number of cells. For each cell, the authors averaged together all the firing rates measured during reaches to a particular target in short time bins defined relative to the start of the movement. This produced a set of average firing rates $f(t, \theta, d)$ for each cell, corresponding to the tth bin of the movement to a target at distance d and angle θ.

   The authors then fit a series of regression models to attempt to predict these firing rates from the target locations. Their model, for each neuron at the tth time bin was:

$$f(t,\theta,d) = k_0(t) + \underbrace{k_1(t)\sin\theta + k_2(t)\cos\theta}_{\textsf{direction terms}} + \underbrace{k_3(t)d + k_4(t)d^2}_{\textsf{distance terms}} + \underbrace{k_5(t)[d\cdot\sin\theta] + k_6(t)[d\cdot\cos\theta]}_{\textsf{target terms}}$$

   Once this model was fit, they looked at the fraction of the variance over different target locations in $f(t, \theta, d)$ for a fixed cell and time that could be explained by their complete model (R^2) and by reduced models which used a subset of the coefficients, *freezing their values at those found by fitting the complete model*. There were three subsets of terms identified by the braces above, each combined with k_0. Thus, for example, to obtain $R^2_{\textsf{tar}}$ they looked at the variance explained by the model

$$f(t,\theta,d) = k_0(t) + k_5(t)[\sin\theta\cdot d] + k_6(t)[\cos\theta\cdot d].$$

   The figure illustrates the results for a single neuron. The top line shows the average hand velocity over all repetitions of all 48 movement types. The remaining 4 graphs show the fractions of variance explained by the full model, and each of the 3 reduced versions.

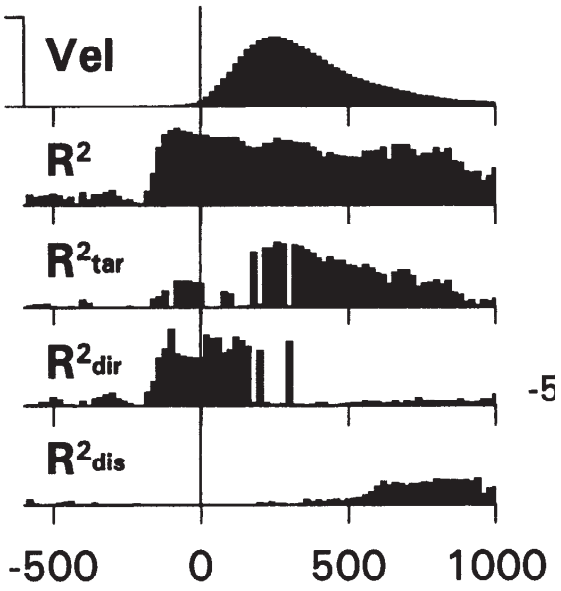


FIG. 3. Example of a unit with sequential correlation of its discharge with direction, target position, and distance. Switchlike changes in the partial R^2 between direction and target position occur, but a gradual transition is observed from the target position to distance. Same conventions as in Fig. 1. Unit recorded in primary motor cortex from *monkey B*.

   The authors pointed to two phenomena that they thought remarkable—the rapid switching between "target" and "direction" coding around 200 ms into the movement, and the gradual rise in "distance" coding towards the end of the period. Both of these are likely to be artifacts of the experimental design and analysis. Explain why.

7. 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:

   (a) 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.

   (b) this implies that

   $$I(\theta;\{n_{ia}\}) \to \log\pi - \tfrac{1}{2}\log 2\pi e/J$$

   where J is the population Fisher information.

8. Discuss, giving real examples, the relative merits of population codes based on the following strategies: scalar coding; labelled-line coding; distributed coding; cartesian coding. (Note that these are not all mutually exclusive). What properties would you expect to hold for the quantities most likely to be encoded by each of the strategies? How might each scheme handle the encoding of uncertainty about the represented quantity?

9. Glimcher had monkeys playing the inspection game by making saccades. He defined by the following payoff matrix (upper right numbers, reward to computer; lower left numbers, reward to monkey):

	computer L	computer R
monkey L	1.7 1	2 1
monkey R	-0.3 0	-2 2

   If monkey and computer play at a Nash equilibrium, what policy will the monkey adopt.

10. An undergraduate suggests that the activity of dopamine cells at the time of a reward is proportional to:

$$DA = \begin{cases} \log P(R|\text{model}) & \text{if } R < \langle R\rangle \\ -\log P(R|\text{model}) & \text{if } R \geq \langle R\rangle \end{cases}$$

    where $\langle R\rangle$ is the average reward and the probabilities are computed by a model of the reward. With what data on the activity of dopamine cells is this consistent and inconsistent?

11. Consider the case that stimulus vector $\mathbf{s}$ is mapped deterministically to output vector $\mathbf{v} = \mathbf{f}(W \cdot \mathbf{s}$ by a square matrix W and a point-wise non-linear function $\mathbf{f}(z_1, z_2, \ldots) = [f(z_1)\ f(z_2)\ \ldots]$. Write down the differential entropy of the density $p[\mathbf{v}]$ as a function of the density $p[\mathbf{s}]$ and thereby derive an iterative, stochastic gradient ascent learning rule, for maximizing it.

12. It has been suggested that one way to derive hyperbolic discounting is to average a collection of exponential discounting terms. Show how this might work, and comment on its biological credibility.