Assignment 3
Theoretical Neuroscience

TAs:
Kira Dusterwald ⟨kira.dusterwald.21@ucl.ac.uk⟩
Samo Hromadka ⟨samo.hromadka.21@ucl.ac.uk⟩
Dimitri Meunier ⟨dimitri.meunier.21@ucl.ac.uk⟩
Aaditya Singh ⟨aaditya.singh.21@ucl.ac.uk⟩

Due 10 February, 2023

1. **Stability of equilibria**

Consider Wilson-Cowan equations of the form

$$\tau \dot{\nu}_E = \phi_E(\nu_E, \nu_I) - \nu_E \tag{1a}$$
$$\tau \dot{\nu}_I = \phi_I(\nu_I, \nu_E) - \nu_I \tag{1b}$$

where the gain functions, ϕ_E and ϕ_I, are increasing functions of ν_E and decreasing functions of ν_I (e.g, $\phi_E(\nu_E, \nu_I) \sim 1 + \tanh(W_{EE}\nu_E - W_{EI}\nu_I + \theta_E)$).

Nullclines for Eq. (1) are sketched in the figure below. Show that equilibria A and C are stable, B is unstable, and D may or may not be stable. Give conditions for the stability of equilibrium D in terms of the derivatives of the gain functions evaluated at the equilibrium.

Hint: This problem is relatively hard, in the sense that it requires a somewhat deep understanding of nullclines and their construction, and also strong familiarity with linear stability analysis in two dimensions. On the other hand, the answer doesn't require a huge amount of algebra – only a few lines. The main insight you need is that you can compute the slopes of the nullclines in terms of derivatives of the gain functions. Once you do that, the rest should be easy (ish).

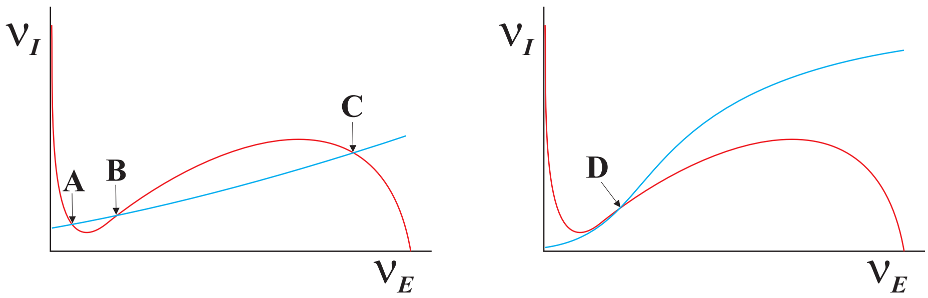

Figure 1: Two possible sets of nullclines. In both figures, the red curve is the excitatory nullcline and the blue curve is the inhibitory one.

2. **Adaptation**

   Consider a network of N analog neurons that obey the time-evolution equations

$$\tau \frac{dx_i}{dt} = \phi\left(\sum_{j=1}^{N} W_j x_j - \theta_i\right) - x_i \,. \tag{2}$$

   (a) Assume that $\theta_i = \theta \; \forall i$. Show that Eq. (2) can be effectively reduced to a one-variable model,

$$\tau \frac{dz}{dt} = \phi\left(Jz - \theta\right) - z \,. \tag{3}$$

   Write down expressions for z in terms of the W_i and x_i and J in terms of the W_i.

   (b) Let's go back to Eq. (2), where θ_i depends on i. Show that Eq. (2) can still be reduced to a one-variable model,

$$\tau \frac{dz}{dt} = \tilde{\phi}\left(Jz\right) - z \tag{4}$$

   where J is the same as in part (a). Write down an expressions for $\tilde{\phi}(\cdot)$ in terms of $\phi(\cdot)$ and W_i and θ_i.

   (c) Assume that both W_i and θ_i are correlated random variables with joint distribution $p(W, \theta)$. Assuming $N \to \infty$, write down an expression for $\tilde{\phi}(Jz)$ as an integral over this joint distribution.

   (d) Let's go back to the case in which $\theta_i = \theta$, so that z evolves according to Eq. (3). Let $\phi(y) = \tanh(y)$ (which isn't realistic because it allows negative firing rates, but it makes the analysis easier). To model spike frequency adaptation, let θ evolve according to

$$\tau_0 \dot{\theta} = -(\theta - \theta_0 z) \,, \tag{5}$$

   with $\tau_0 \gg \tau$. Assume that $\theta_0 > J - 1 > 0$. Sketch the nullclines.

(e) Show that the system exhibits bursting, and sketch $z(t)$ and $\theta(t)$ versus time. Here "bursting" just means a limit cycle in θ-z space. We call it bursting because $\tau_0 \gg \tau$, so z spends most of its time changing slowly, with only brief periods during which it changes very rapidly from positive to negative or back.

3. **Why we can ignore temporal correlations**

We're going to consider a randomly connected network of excitatory and inhibitory neurons with current-based synapses and linear integrate-and-fire (aka LIF) neurons,

$$\tau \frac{dV_{E,i}}{dt} = -(V_{E,i} - V_{\text{rest}}) + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} W_{EE,ij} g_{E,j}(t) - \frac{1}{\sqrt{n}} \sum_{j=1}^{n} W_{EI,ij} g_{I,j}(t) + \sqrt{n} h_E \tag{6a}$$

$$\tau \frac{dV_{I,i}}{dt} = -(V_{I,i} - V_{\text{rest}}) + \frac{1}{\sqrt{n}} \sum_{j=1}^{n} W_{IE,ij} g_{E,j}(t) - \frac{1}{\sqrt{n}} \sum_{j=1}^{n} W_{II,ij} g_{I,j}(t) + \sqrt{n} h_I \,. \tag{6b}$$

If a neuron exceeds threshold, it spikes and is reset to V_{rest}. This is a bit unrealistic: connectivity is all-all, and the number of excitatory and inhibitory neurons are the same. But making it more realistic would only complicate the analysis without adding any insight.

As usual, $g_{E,j}(t)$ and $g_{I,j}(t)$ are the conductance changes due to presynaptic spikes,

$$g_{E,j}(t) = \sum_k g(t - t_{E,j}^k) \tag{7a}$$

$$g_{I,j}(t) = \sum_k g(t - t_{I,j}^k) \tag{7b}$$

where $t_{E,j}^k$ is the time of the k^{th} spike on excitatory neuron j, and similarly for $t_{I,j}^k$. We'll choose $g(t)$ so that it integrates to 1,

$$\int_0^\infty dt\, g(t) = 1 \tag{8}$$

with $g(t)$ non-negative. And, of course, $g(t < 0) = 0$.

We can perform the usual manipulations,

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} W_{EE,ij} g_{E,j}(t) = \sqrt{n}\, W_{EE} \nu_E + \frac{1}{\sqrt{n}} \sum_j \delta W_{EE,ij} \nu_{E,j} \\ + \sqrt{n}\, W_{EE} G_E(t) + \frac{1}{\sqrt{n}} \sum_j \delta W_{EE,ij} \delta g_{E,j}(t) \end{aligned} \tag{9}$$

where most quantities have obvious definitions,

$$W_{EE} = \frac{1}{n^2} \sum_{ij} W_{EE,ij} \tag{10a}$$

$$\delta W_{EE,ij} = W_{EE,ij} - W_{EE} \tag{10b}$$

$$\nu_{E,j} = \langle g_{E,j}(t) \rangle_t \tag{10c}$$

$$\nu_E = \frac{1}{n} \sum_j \nu_{E,j} \tag{10d}$$

$$\delta g_{E,j}(t) = g_{E,j}(t) - \nu_{E,j} \tag{10e}$$

$$G_E(t) = \frac{1}{n} \sum_j \delta g_{E,j}(t) \,, \tag{10f}$$

and similarly for the inhibitory neurons. The difference between what I'm doing here and what I did in class was to separate out the term $G_E(t)$. Inserting this into Eq. (6a) gives us

$$\begin{aligned}\tau\frac{dV_{E,i}}{dt} = -(V_{E,i} - V_{\text{rest}}) &+ \sqrt{n}\left[W_{EE}\nu_E - W_{EI}\nu_I + h_E\right] \\ &+ \sqrt{n}\left[W_{EE}G_E(t) - W_{EI}G_I(t)\right] \\ &+ \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\delta W_{EE,ij}\nu_{E,j}(t) - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\delta W_{EI,ij}\nu_{I,j}(t) \\ &+ \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\delta W_{EE,ij}\delta g_{E,j}(t) - \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\delta W_{EI,ij}\delta g_{I,j}(t)\,.\end{aligned} \tag{11}$$

Assume, as usual, that $n \gg 1$. Then, the first term on the right hand side of Eq. (11) fixes the mean firing rates, to within $\mathcal{O}(1/\sqrt{n})$. The second term ensures that the average temporal fluctuations are $\mathcal{O}(1/\sqrt{n})$. It's this term that reduces correlations among neurons – or at least correlations among the whole population. There can, of course, be instabilities that lead to oscillations, in which case neurons do become highly correlated. But there are also regimes where oscillations are small.

(a) Finally, the actual homework problem. Define

$$\delta G_{E,i}(t) \equiv \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\delta W_{EE,ij}\delta g_{E,j}(t)\,, \tag{12}$$

and similarly for $\delta G_{E,i}(t)$. Assume the weights are random – the elements are draws *iid* from some distribution. Show that, in the large n limit,

$$\langle \delta G_{E,i}(t)\delta G_{E,i'}(t-\tau)\rangle_t = \delta_{ii'}C(\tau)\text{Var}[W_{EE,ij}] \tag{13a}$$

$$\langle \delta G_{E,i}(t)\delta G_{I,i'}(t-\tau)\rangle_t = 0 \tag{13b}$$

where

$$C_E(\tau) \equiv \frac{1}{n}\sum_j \langle \delta g_{E,j}(t) g_{E,j}(t-\tau)\rangle_t\,. \tag{14}$$

Thus, in the large n limit, we get to ignore temporal correlations among neurons! What's even better is that all excitatory neurons have the same temporal correlational structure, as do all inhibitory neurons. Of course, that correlational structure has to be found self-consistently, which is nontrivial. But it's nice to know that it exists.

4. **Continuous time Hopfield networks**

Consider a continuous time Hopfield network,

$$\frac{dx_i}{dt} = \phi\left(h_i\right) - x_i \tag{15}$$

where ϕ is the gain function (taken to be non-negative, more or less sigmoidal, and saturating), N is the number of neurons, and h_i is the synaptic drive,

$$h_i \equiv \sum_{j=1}^{N} J_{ij}x_j\,. \tag{16}$$

We'll let

$$J_{ij} = \frac{1}{Nf(1-f)} \sum_{\mu=1}^{p} \xi_i^\mu (\xi_j^\mu - f) \tag{17}$$

where the ξ_i^μ are random binary vectors, a fraction f of which are 1,

$$\xi_i^\mu = \begin{cases} 1 & \text{probability } f \\ 0 & \text{probability } 1-f\,. \end{cases} \tag{18}$$

There are several differences between this formulation and the one we used in class: the x_i are continuous rather than discrete; the gain function is smooth and non-negative (the latter ensuring that the x_i will be non-negative); the elements of the patterns are 0 and 1 rather than -1 and 1; and the probability of 1 is f rather than $1/2$. However, the analysis is nearly identical.

As usual, the goal is to find the equilibria. With this formulation, the equilibria aren't necessarily all that close to the patterns, ξ_i^μ. However, we still expect the equilibria to be at least related to the patterns. With that in mind, we define the overlaps, denoted m_μ, via

$$m_\mu = \frac{1}{Nf(1-f)} \sum_i (\xi_i^\mu - f) x_i. \tag{19}$$

If $x_i = \xi_i^\mu$, m_μ will be close to 1, whereas if x_i is independent of ξ_i^μ, m_μ will be close to zero. At an equilibrium, we expect one of the m_μ to be large and the rest to be small.

(a) For this question, we'll let J_{ij} be symmetric, but otherwise arbitrary. Define the "energy" E via

$$E \equiv \frac{1}{2} \sum_{ij} x_i J_{ij} x_j - \sum_i \psi(h_i) \tag{20}$$

where ψ is obeys

$$\frac{d\psi(x)}{dx} = \phi(x)\,. \tag{21}$$

Show that

$$\frac{dE}{dt} = -\sum_{ij} \frac{dx_i}{dt} J_{ij} \frac{dx_j}{dt}\,. \tag{22}$$

Thus, if J_{ij} is symmetric and positive definite (consistent with Eq. (17) if $f = 0$), then E is a non-increasing function of time. I tried, but could not find, a Lyapunov function when J_{ij} is symmetric but not positive definite. That does not mean one does not exist. For extra credit, find one!

For the rest of the questions, use the connection matrix given in Eq. (17).

(b) Show that

$$h_i = \sum_\mu \xi_i^\mu m_\mu. \tag{23}$$

(c) Show that the m_μ obey the equation

$$\frac{dm_\nu}{dt} = \frac{1}{N} \sum_i \frac{\xi_i^\nu}{f} \phi\Big(m_\nu + \sum_{\mu\neq\nu} \xi_i^\mu m_\mu\Big) - \frac{1}{N} \sum_i \frac{1-\xi_i^\nu}{1-f} \phi\Big(\sum_{\mu\neq\nu} \xi_i^\mu m_\mu\Big) - m_\nu. \tag{24}$$

(d) Define

$$\zeta_i \equiv \sum_{\mu \neq \nu} \xi_i^\mu m_\mu . \tag{25}$$

We're going to treat ζ_i as a random variable with respect to index, i. Because ξ_i^ν and ξ_i^μ are uncorrelated, it follows that ξ_i^ν is independent of ζ_i. Thus, in the large N limit, the equation for the m_ν becomes

$$\frac{dm_\nu}{dt} = \Phi(m_\nu) - m_\nu \tag{26}$$

where

$$\Phi(m) \equiv \int d\zeta \, p(\zeta) [\phi(m + \zeta) - \phi(\zeta)] \, . \tag{27}$$

Note that $\Phi(m)$ is just a smoothed, and offset, version of $\phi(m)$. In what follows, we'll assume that $p(\zeta)$ does not change with time, which is true only at equilibria.

Equation (26) has an equilibrium at $m_\nu = 0$. Under what conditions is this equilibrium is stable?

(e) Assume that x_i is independent of ξ_i^μ when $\mu \neq \nu$. Show that ζ_i is a zero mean Gaussian random variable with variance, denoted σ^2, given by

$$\sigma^2 = \frac{p-1}{N(1-f)} \langle x_i^2 \rangle \left[1 + \frac{1-2f}{Nf} \right] \approx \frac{p-1}{N(1-f)} \langle x_i^2 \rangle \tag{28}$$

If $x_i \propto \xi_i^\nu$, then $\langle x_i^2 \rangle \propto f$, and $\sigma^2 \propto f/(1-f)$. Thus, small f decreases the noise and, therefore, increases the capacity.

(f) This isn't a question, but there are a couple things to notice. For the system to have a "memory" – a fixed point for which m_μ is $\mathcal{O}(1)$ – the smoothed gain function, $\Phi(m)$, must be sufficiently steep. Thus, σ can't be too big (because the larger σ is, the more the gain function is smoothed; see Eq. (27)). Given Eq. (28), for small f the variance should scale as p/N, which would mean that the capacity shouldn't depend much on f (at least when f is small). However, I told you in class (and in the Hopfield writeup) that capacity scales as $1/f$. I have never been able to find a simple explanation for the $1/f$ scaling.

5. **Adding a little rigor to tacky math**

Consider an all-inhibitory network of N neurons for which the firing rate of neuron i, denoted ν_i, is given by

$$\nu_i = \phi \left(\sqrt{N} h_0 - \frac{1}{\sqrt{N}} \sum_j w_{ij} \nu_j \right) . \tag{29}$$

The gain function, ϕ, is sigmoidal, the external input, h_0, is positive, and the weights are non-negative, independent and identically distributed, and their mean and variance is give by $\overline{w}$ and σ_w^2, respectively. As usual, we write $w_{ij} = \overline{w} + \delta w_{ij}$, so that

$$\frac{1}{\sqrt{N}} \sum_j w_{ij} \nu_j = \sqrt{N} \, \overline{w} \, \overline{\nu} + \frac{1}{\sqrt{N}} \sum_j \delta w_{ij} \nu_j \tag{30}$$

where

$$\overline{\nu} \equiv \frac{1}{N} \sum_j \nu_j \, . \tag{31}$$

(a) Show that in the large N limit

$$\sigma^2 \equiv \sum_i \frac{1}{N} \left(\frac{1}{\sqrt{N}} \sum_j \delta w_{ij} \nu_j \right)^2 = \overline{\nu^2} \sigma_w^2 \tag{32}$$

where

$$\overline{\nu^2} \equiv \frac{1}{N} \sum_j \nu_j^2 \tag{33}$$

and corrections are $\mathcal{O}(1/\sqrt{N})$.

(b) We can now write the following equation for ν_i,

$$\nu_i = \phi\left(\sqrt{N}(h_0 - \overline{w}\,\overline{\nu}) + \sigma \xi_i\right) \tag{34}$$

where

$$\xi_i \equiv \frac{1}{\sqrt{\sigma N}} \sum_j \delta w_{ij} \nu_j \, . \tag{35}$$

In the large N limit, $\sum_i \xi_i^2/N = 1$, but that tells us nothing about its distribution. What we want to argue is that ξ_i is a Gaussian random variable. For that we need δw_{ij} and ν_j to be weakly correlated. To check if that's true, compute the empirical covariance (squared), denoted ρ_i^2,

$$\rho_i^2 = \frac{\left(\sum_j \delta w_{ij} \nu_j\right)^2}{\left(\sum_j w_{ij}^2\right)\left(\sum_j \nu_j^2\right)} \, . \tag{36}$$

Show that on average

$$\frac{1}{N} \sum_i \rho_i^2 \sim \mathcal{O}\left(\frac{1}{\sqrt{N}}\right) \, . \tag{37}$$

If we were mathematicians, this wouldn't mean much. But as physicists, we'll declare this to be sufficiently weakly correlated that we can treat ξ_i as a zero mean, unit variance Gaussian random variable.

We now need to show that ν_i and ξ_i are weakly correlated. You can do that using a minor extension of the above analysis.

6. **Networks with time-varying dynamics**

Consider a network of N neurons that evolves according to

$$\frac{dx_i}{dt} = \phi\left(\sum_j W_{ij} x_j + \sum_\mu J_{i\mu} z_\mu + \sum_\mu C_{i\mu} u_\mu(t) \right) - x_i \tag{38}$$

where $u_\mu(t)$ is a control signal, ϕ is the gain function (as usual, it's more or less sigmoidal), and $\mathbf{z}$ is related to $\mathbf{x}$ via

$$z_\mu = \sum_j A_{\mu j} x_j . \tag{39}$$

In this setting the dimensionality of both $\mathbf{z}$ and $\mathbf{u}$ is typically much less than N, but that's not necessary for the questions.

(a) Show that z_μ evolves according to

$$\frac{dz_\mu}{dt} = \sum_i A_{\mu i}\phi\left(\sum_j W_{ij}x_j + \sum_\nu J_{i\nu}z_\nu + \sum_\nu C_{i\nu}u_\nu(t)\right) - z_\mu \tag{40}$$

Thus, if $W_{ij} = 0$,

$$\frac{dz_\mu}{dt} = f_\mu\big(\mathbf{z}, \mathbf{u}(t)\big) \tag{41}$$

where the function f_μ is given by a neural network with one hidden layer.

(b) Assume the goal of the network is to produce as output the function $z^*_\mu(t)$. Show that under the learning rule

$$\Delta A_{\mu i} = \eta\big(z^*_\mu(t) - z_\mu(t)\big)x_i(t), \tag{42}$$

the instantaneous error, $\big(z^*_\mu(t) - z_\mu(t)\big)^2$, decreases. Assume that η, the learning rate, is small. Is there any guarantee that the total error, which is the time average of $\big(z^*_\mu(t) - z_\mu(t)\big)^2$, will decrease?

7. **Coupled line attractor**

Consider a coupled network of N neurons whose units evolve according to

$$\frac{dr_i}{dt} = \phi\Big(\sum_j W_{i-j}r_j + h_i\Big) - r_i \tag{43a}$$

$$\tau\frac{dh_i}{dt} = g(t)\sum_j A_{i-j}r_j - h_i. \tag{43b}$$

We'll take W to be symmetric: $W_{i-j} = W_{j-i}$. Assume that when $h_i = 0$, Eq. (43a) has a stable equilibrium given by $f(\theta_i - \theta)$,

$$f(\theta_i - \theta) = \phi\Big(\sum_j W_{i-j}f(\theta_j - \theta)\Big) \tag{44}$$

where the θ_i are equally spaced. Assume that this equation is satisfied for all θ, making it a true line attractor.

(a) In the limit that $g(t)$ is infinitesimally small, show that the position on the line attractor, θ, evolves according to

$$\tau\frac{d^2\theta}{dt^2} + \frac{d\theta}{dt} = g(t)\sum_{ij} v^\dagger_{0i}(\theta)\phi'_i A_{i-j}f(\theta_j - \theta) \tag{45}$$

where

$$\phi'_i \equiv \phi'\Big(\sum_j W_{i-j}f(\theta_j - \theta)\Big), \tag{46}$$

$\mathbf{v}^\dagger_0(\theta)$ is the adjoint eigenvalue of the linearized dynamics whose eigenvalue is 0,

$$\sum_j v^\dagger_{0j}(\theta)\phi'_j W_{j-i} = v^\dagger_{0i}(\theta) \tag{47}$$

and it's normalized so that

$$\sum_i v_{0i}^\dagger(\theta) f'(\theta_i - \theta) = 1. \tag{48}$$

(b) Recall that the adjoint eigenvector is related to $f(\theta_i - \theta)$ via

$$v_{0i}^\dagger(\theta) = \frac{f'(\theta_i - \theta)/\phi_i'(\theta)}{Z} \tag{49}$$

where

$$Z \equiv \sum_i \frac{f'(\theta_i - \theta)^2}{\phi_i'(\theta)}. \tag{50}$$

Consequently, θ evolves according to

$$\tau \frac{d^2\theta}{dt^2} + \frac{d\theta}{dt} = \frac{g(t)}{Z} \sum_{ij} f'(\theta_i - \theta) A_{i-j} f(\theta_j - \theta). \tag{51}$$

Show that in the large N limit, Z is independent of θ.

(c) Show that in the large N limit, the right hand side of Eq. (51) becomes independent of θ. Show also that if A_{i-j} is even ($A_{i-j} = A_{j-i}$), the right hand side is zero.

8. **Sparse connectivity**

Consider a network whose equilibrium is given by

$$\nu_i = \phi(h_i) \tag{52a}$$

$$h_i = \frac{1}{\sqrt{k}} \sum_{j=1}^{n} w_{ij} \nu_j \tag{52b}$$

where the weights are given by

$$w_{ij} = \begin{cases} \overline{w} + \xi_{ij} & \text{probability } k/n \\ 0 & \text{probability } 1 - k/n \,, \end{cases} \tag{53}$$

and the ξ_{ij} are independent zero-mean random variables with variance σ^2. The parameter k is the average number of connections/neurons. As usual, we're interested in the empirical mean and variance of h_i with respect to the index i.

Show that, in the large k limit,

$$h_i = \sqrt{k}\,\overline{w}\langle \nu \rangle + \eta_i \tag{54}$$

where η_i is a zero mean random variable with variance (with respect to index i) given by

$$\text{Var}[\eta_i] = \left[\sigma^2 + \overline{w}^2(1 - k/n)\right] \langle \nu^2 \rangle \,. \tag{55}$$

As usual,

$$\langle \nu^p \rangle = \frac{1}{n} \sum_i \nu_i^p \,. \tag{56}$$

9. **More fun with averages**

Assume w_{ij} are a set independent, zero-mean, random variables with variance σ^2, with both indices running from 1 to n. Assume $n \gg 1$. Define

$$z_i \equiv \frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_{ij} \nu_j \tag{57}$$

where the ν_j are a fixed set of firing rates. Define

$$z \equiv \frac{1}{n} \sum_i z_i \,. \tag{58}$$

We want to compute the variance of z with respect to the index, i (the mean is obviously zero). The variance, denoted σ_z^2, is given by

$$\sigma_z^2 = \frac{1}{n} \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} w_{ij} \nu_j \right)^2 \,. \tag{59}$$

As usual, we write this

$$\begin{aligned} \sigma_z^2 &= \frac{1}{n} \sum_{jk} \nu_j \nu_k \frac{1}{n} \sum_i w_{ij} w_{ik} \qquad (60) \\ &= \frac{1}{n} \sum_j \nu_j^2 \frac{1}{n} \sum_i w_{ij}^2 + \frac{1}{n} \sum_{j \neq k} \nu_j \nu_k \frac{1}{n} \sum_i w_{ij} w_{ik} \,. \end{aligned}$$

The first term is easy – in the large n limit it's $\langle \nu^2 \rangle \sigma^2$. In class we claimed that the second term was small. Your job is to verify that. Show that

$$\text{Var}\left[\frac{1}{n} \sum_{j \neq k} \nu_j \nu_k \frac{1}{n} \sum_i w_{ij} w_{ik} \right] \approx \frac{2 \langle \nu^2 \rangle^2 \sigma^4}{n} \,. \tag{61}$$