Assignment 4

G4360 Introduction to Theoretical Neuroscience

DUE: October 20

I tend to write long problem sets, but most of it is informational, there is not that much for you to actually do. So please don't be put off by the length; in the long problems, things you actually have to do are indicated in red.

Notation: boldface small letters, like $\mathbf{r}$, represent column vectors; $\mathbf{r}^{\mathrm{T}}$ is a row vector, the transpose of $\mathbf{r}$; boldface capital letters, like $\mathbf{W}$, represent matrices; $\mathbf{W}^{\mathrm{T}}$ is the transpose of $\mathbf{W}$; non-boldface letters represent numbers, either scalars or the individual elements of vectors or matrices.

Problem 1: A little linear algebra.

a. Show that any matrix $\mathbf{M}$ that has a complete basis of eigenvectors can be written $\mathbf{M} = \sum_i \lambda_i \mathbf{r}_i \mathbf{l}_i^T$ where $\mathbf{r}_i$ and $\mathbf{l}_i$ are the i^{th} right and left eigenvectors of $\mathbf{M}$, respectively, and λ_i is the corresponding eigenvalue. To show this, express an arbitrary vector $\mathbf{v}$ in the eigenvector basis, apply $\mathbf{M}$ to it, and apply $\sum_i \lambda_i \mathbf{r}_i \mathbf{l}_i^T$ to it, and show they give the same result.[1]

b. Show that $\mathbf{l}_j^T M \mathbf{r}_i = \lambda_i \delta_{ij}$. When the eigenvectors are orthonormal, this means $\mathbf{r}_i^T M \mathbf{r}_i = \lambda_i$.

Don't forget that left eigenvectors and right eigenvectors satisfy $\mathbf{l}_i^T \mathbf{r}_j = \delta_{ij}$.

Problem 2: The inhibition-stabilized network (ISN). First we'll do an extensive setup. The problem will be to demonstrate the paradoxical effect using nullclines. Consider a two-population model of firing-rate neurons: one excitatory (E) population and one inhibitory (I) population. r_E and r_I are the firing rates of the excitatory and inhibitory populations, respectively, represented by the vector $\mathbf{r} = \begin{pmatrix} r_E \\ r_I \end{pmatrix}$. The matrix of connections between them is $\mathbf{W} = \begin{pmatrix} w_{EE} & -w_{EI} \\ w_{IE} & -w_{II} \end{pmatrix}$, where w_{XY} represents the (positive) strength of the connection from Y to X. We let the vector of external inputs to the two populations be $\mathbf{i}$. We let $f(\mathbf{v})$ be a nonlinear function applied element-wise to the elements of the vector $\mathbf{v}$, *i.e.* $f(\mathbf{v})$ is a vector with i^{th} element $f(\mathbf{v})_i \equiv f(v_i)$. We assume the

[1] We're assuming eigenvalues and eigenvectors are real here. For the general case where they may be complex, you simply replace the transpose, $\mathbf{v}^T$ or $\mathbf{M}^T$, with the adjoint or conjugate transpose, $\mathbf{v}^\dagger$ or $\mathbf{M}^\dagger$, meaning take the complex conjugate of every element and then take the transpose. This ensures the length-squared of a vector, $\mathbf{v} \cdot \mathbf{v} = \mathbf{v}^\dagger \mathbf{v}$, is real and positive and equal to $\sum_i |v_i|^2$. Then everything goes through exactly as before. For example, if $\mathbf{E}$ is the matrix whose columns are the eigenvectors, then we define the i^{th} row of $\mathbf{E}^{-1}$ to be $\mathbf{l}_i^\dagger$, the adjoint of the i^{th} left eigenvector, and the above expression becomes $\lambda_i \mathbf{r}_i \mathbf{l}_i^\dagger$.

steady-state firing rate $\mathbf{r}_{SS}$ for a given input is given by f applied to each unit's input: $\mathbf{r}_{SS} = f(\mathbf{W}\mathbf{r} + \mathbf{i})$. We assume the network approaches its instantaneous steady state with first-order dynamics: letting $\mathbf{T} = \begin{pmatrix} \tau_E & 0 \\ 0 & \tau_I \end{pmatrix}$ be the diagonal matrix of E and I time constants, we have

$$\mathbf{T}\frac{d}{dt}\mathbf{r} = -\mathbf{r} + f(\mathbf{W}\mathbf{r} + \mathbf{i}) \tag{1}$$

Suppose $\mathbf{r}_{SS}$ is a stable fixed point; we will linearize the dynamics about this fixed point. You know that, letting f'_E and f'_I be the derivative of f evaluated at the E and I components of $\mathbf{W}\mathbf{r}_{SS} + \mathbf{i}$, respectively, the linearized weights are $\begin{pmatrix} \partial f_E/\partial r_E & \partial f_E/\partial r_I \\ \partial f_I/\partial r_E & \partial f_I/\partial r_I \end{pmatrix} = \begin{pmatrix} f'_E w_{EE} & f'_E w_{EI} \\ f'_I w_{IE} & f'_I w_{II} \end{pmatrix}$; to make notation simpler, let's define this to be $\mathbf{J} = \begin{pmatrix} j_{EE} & -j_{EI} \\ j_{IE} & -j_{II} \end{pmatrix}$ (we'll assume $f(x)$ is a monotonically increasing function of x, so that all the f'_X's are positive and hence all the j_{XY}'s are positive). Let $\mathbf{i}_{SS}$ be the steady-state input that yields the fixed point $\mathbf{r}_{SS}$. If there is a deviation $\Delta\mathbf{i}$ from $\mathbf{i}_{SS}$, in the linearized equation this becomes $\delta\mathbf{i} \equiv \begin{pmatrix} f'_E \Delta i_E \\ f'_I \Delta i_I \end{pmatrix}$. Define small deviations in response from the steady state by $\mathbf{r} = \mathbf{r}_{SS} + \delta\mathbf{r}$. Then the equation for the dynamics linearized about the fixed point is

$$\mathbf{T}\frac{d}{dt}\delta\mathbf{r} = -\delta\mathbf{r} + \mathbf{J}\delta\mathbf{r} + \delta\mathbf{i} = -(\mathbf{1} - \mathbf{J})\delta\mathbf{r} + \delta\mathbf{i} \tag{2}$$

where $\mathbf{1}$ is the identity matrix. Note that, since $\delta\mathbf{r}$ is multiplied by $\mathbf{J} - \mathbf{1}$, both eigenvalues of $\mathbf{J} - \mathbf{1}$ must have negative real part for the fixed point to be stable, meaning both eigenvalues of $\mathbf{1} - \mathbf{J}$ must have positive real part. This should all be familiar to you, but if it's not, satisfy yourself that this is all true.

Recall what we did in class to show the ISN paradoxical response: for a steady-state input perturbation $\delta\mathbf{i}$, we wrote down the equation for the steady-state response $\delta\mathbf{r}$: $\delta\mathbf{r} = (\mathbf{1} - \mathbf{J})^{-1}\delta\mathbf{i}$. For a 2×2 matrix $\mathbf{M} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the inverse is given by $\mathbf{M}^{-1} = \frac{1}{\mathrm{Det}\,\mathbf{M}} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. So $(\mathbf{1} - \mathbf{J})^{-1} = \frac{1}{\mathrm{Det}\,(\mathbf{1}-\mathbf{J})} \begin{pmatrix} 1 + j_{II} & -j_{EI} \\ j_{IE} & 1 - j_{EE} \end{pmatrix}$. Recall that the determinant is the product of the eigenvalues, for the fixed point to be stable, we must have $\mathrm{Det}\,(\mathbf{1} - \mathbf{J}) > 0$. Thus, for a stable fixed point, if and only if $j_{EE} > 1$ (which means the E population alone would be unstable if I firing was frozen at its fixed point level; look at the equation for r_E with r_I fixed, to see why $j_{EE} > 1$ implies excitatory instability), the I cells show a "paradoxical" response. This means that, if an input is given only to I cells ($\delta\mathbf{i} \propto \begin{pmatrix} 0 \\ 1 \end{pmatrix}$), the steady-state response of the I cells is of opposite sign to the input, so

that adding excitation to I cells paradoxically lowers their firing rate in the new steady state.

Now show the same things using nullclines. Again assume that the function $f(x)$ is a monotonically increasing function of x. The equations for the E and I nullclines are the E and I components of the fixed-point equation: $\mathbf{r} = f(\mathbf{Wr} + \mathbf{i})$. We will draw the nullclines with r_E on the x axis and r_I on the y axis. The point where the nullclines cross – where both nullcline equations are satisfied – is the fixed point. Our linearization only applies in the vicinity of the fixed point; but we'll continue to use j_{XY} to mean $f'_X w_{XY}$ at any point. The derivatives and thus the j's will have different values from point to point.

a For the I nullcline, compute its slope, dr_I/dr_E; you should find that it is given by $\frac{j_{IE}}{1+j_{II}}$. This means that the nullcline always has positive slope.

b Now for the E nullcline, compute the inverse of its slope, dr_E/dr_I; you should find that this inverse slope is $\frac{j_{EI}}{j_{EE}-1}$. This means that the slope is positive if the E subnetwork is unstable, and negative if the E subnetwork is stable.

c Show that the condition that $\mathrm{Det}\,(\mathbf{J} - \mathbf{1}) > 0$ at the fixed point, which is necessary for stability of the fixed point, is equivalent to the I nullcline having a larger slope than the E nullcline at the fixed point. (Recall that the determinant of a 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is $(ad - bc)$). So for a fixed point to be stable, it is necessary that the I nullcline have a larger slope than the E nullcline at their crossing that defines the fixed point.

d So, we'll draw two versions of the nullclines: one that is an ISN, one that is not. We're not going to quantitatively determine the nullclines for particular parameter values, just qualitatively draw the structure of the nullclines, as follows: First draw the I nullcline, which will be the same for both versions. Imagine that, for r_E small, the I-nullcline solution for r_I should be small (*i.e.*, the I nullcline starts at the bottom left), while for r_E large, r_I is large (ends at top right); so the nullcline starts toward the bottom left, ends up at the top right, and always has a positive slope, for example it could have a sigmoidal shape.

e Now, draw the E nullclines, assuming a stable fixed point. Imagine that when r_I is high, r_E is low, so the nullcline starts in the upper left corner; while when r_I is low, r_E is high, so it ends up in the lower right corner. In the non-ISN version, it has a negative slope all the way. In the ISN version, it has a positive slope in a middle portion, so the nullcline looks like a sideways S (*i.e.* it goes down, then up, then down

again); and the fixed point is on the positive-sloping middle portion (and the necessary condition for stability on E and I nullcline slopes is obeyed).

f Draw the arrows indicating the direction of flow in the different regions of the nullcline plane. To do this, consider: if you are to the left or the right of the I nullcline (lower or greater r_E compared to the value that gives an r_I derivative of 0), is r_I going up or down? Similarly, if you are above or below the E nullcline (greater or lower r_I compared to the value that gives an r_E-derivative of 0), is r_E increasing or decreasing? The nullclines divide the plane into four regions, and based on these, you can determine the direction of flow in each region (*e.g.*, down & left, down & right, up & left, up & right), so draw an arrow in each region corresponding to the direction of flow. Show (by drawing arrows) that, in negative-sloping regions of the E nullcline, if r_I is kept fixed, small perturbations of r_E off the E nullcline will flow back to the nullcline; while in positive-sloping regions, it will flow away. This also tells you that in positive-sloping regions, the E subnetwork alone is unstable, while in negative-sloping regions it is stable.

g Now, suppose you add a positive input to the I cells. Show (using the equation for the I nullcline) that the resulting change in the I nullcline is to reduce r_E by the same amount for any given r_I, that is, to move the I nullcline leftward. There is no change in the E nullcline. Draw, as a dashed line, the new I nullcline after an input is added to I. Show (again, using drawings) that, for a stable fixed point, if the network is an ISN, then in moving from the old fixed point to the new fixed point, both r_E and r_I are decreased; while for a non-ISN, the result is to decrease r_E but increase r_I. (For the ISN, assume that the new fixed point, like the old one, is on the positive-sloping portion of the E nullcline.)

h In the ISN case, draw the dynamical path followed by r_E, r_I from the old fixed point to the new fixed point after adding the positive input to I. This addition of input instantaneously moves the I nullcline; the resulting derivative at the old fixed point (which is no longer on the I nullcline and so no longer a fixed point) has an upward component (it will go in the direction of the flow for the region it's in, given the new I nullcline), becoming horizontal as the flow crosses the I nullcline, and then going downward to the new fixed point (it might spiral into the fixed point if there are complex eigenvalues, or go straight down to it if eigenvalues are real). Note, regarding the old fixed point as a perturbation from the new fixed point, that, even though the new fixed point is stable, the dynamics move further away from the new fixed point (the upward movement) before ultimately flowing back to it; and that, after the addition of excitatory input to the I neuron, r_I transiently goes up before going down

in the new steady state. These are effects of non-normal dynamics. (Recall that biological weight matrices, of the form $\mathbf{J} = \begin{pmatrix} j_{EE} & -j_{EI} \\ j_{IE} & -j_{II} \end{pmatrix}$ with all j_{XY}'s positive, are non-normal, meaning that their eigenvectors are not orthogonal, because $\mathbf{J}\mathbf{J}^{\mathrm{T}} \neq \mathbf{J}^{\mathrm{T}}\mathbf{J}$, which is the necessary and sufficient condition for non-normality.)

Problem 3: Non-normal dynamics.

We consider a linear equation with constant input $\mathbf{i}$, obtained as a linearization about a fixed point as in Eq. 2. For simplicity, we'll use $\mathbf{r}$, $\mathbf{W}$, and $\mathbf{i}$ in place of δr, $\mathbf{f}'\mathbf{W}$, and $\mathbf{f}'\mathbf{i}$, respectively, so we'll study the linear equation:

$$\tau \frac{d\mathbf{r}}{dt} = -\mathbf{r} + \mathbf{W}\mathbf{r} + \mathbf{i} \tag{3}$$

(Note we're assuming the same τ for E and I.)

To further simplify, we're going to restrict to a class of connection matrices that has $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ as one of its eigenvectors: $\mathbf{W} = \begin{pmatrix} w_1 & -(w_1 + x) \\ w_2 & -(w_2 + x) \end{pmatrix}$ where $x > 0$.

(1) Verify that the (unnormalized) eigenvectors are $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$ with eigenvalue $\lambda_1 = -x$, and $\begin{pmatrix} \frac{w_1+x}{w_2} \\ 1 \end{pmatrix}$ with eigenvalue $\lambda_2 = w_1 - w_2$.

We assume the system is stable, that is $\lambda_1 < 1$ and $\lambda_2 < 1$. Note that the two eigenvectors are not orthogonal to each other, and that they are similar to one another in that both have all-positive entries, i.e. both point into the upper right quadrant.

We're going to use the Schur transformation, which is a transform to an orthogonal basis that makes the weight matrix as simple as possible for an orthogonal transform, namely upper triangular – all zeros below the diagonal (with a transformation to the non-orthogonal basis of the eigenvectors, the matrix can be made even simpler – diagonal; but upper triangular is as simple as we can make it with a transformation to an orthogonal basis). To do this, we choose one eigenvector as the first Schur basis vector, and choose the other vector orthogonal to this one (more generally, in higher dimensions, you take some ordering of the eigenvectors and then do Gram-Schmidt orthonormalization to produce an orthogonal Schur basis; the transformation is not unique, because each ordering produces a different Schur basis). We'll choose our (normalized) Schur basis vectors to be $\mathbf{s}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, and a vector orthogonal to it, $\mathbf{s}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}$. Recall that the component of a vector along an orthonormal basis vector is just given by the dot product of the vector with the basis vector.

(2) Show that $\mathbf{r}$ has components $\begin{pmatrix} r_1 \\ r_2 \end{pmatrix} = \begin{pmatrix} \frac{r_E+r_I}{\sqrt{2}} \\ \frac{r_E-r_I}{\sqrt{2}} \end{pmatrix}$ in the $\mathbf{s}_1, \mathbf{s}_2$ basis, that is, the components r_1 and r_2 represent the sum and difference of E and I activities, respectively. You already know that $\mathbf{W}\mathbf{s}_1 = -x\mathbf{s}_1$.

(3) Show that in the $\mathbf{s}_1, \mathbf{s}_2$ basis, $\mathbf{W}$ takes the form $\begin{pmatrix} \lambda_1 & w_{FF} \\ 0 & \lambda_2 \end{pmatrix}$ where the "feedforward weight" w_{FF} is given by $w_{FF} = w_1 + w_2 + x$. You can do this in one or both of two ways. First, you can show that $\mathbf{W}\mathbf{s}_2 = (w_1 + w_2 + x)\mathbf{s}_1 + (w_1 - w_2)\mathbf{s}_2$, which along with $\mathbf{W}\mathbf{s}_1 = \lambda_1\mathbf{s}_1$ gives the form of the matrix.[2] Second, you can explicitly transform to the new basis: compute $\mathbf{S}^T\mathbf{W}\mathbf{S}$ where $\mathbf{S}$ is the orthogonal matrix whose columns are $\mathbf{s}_1$ and $\mathbf{s}_2$. We call w_{FF} a feedforward weight because it is an effective feedforward weight between activity patterns; the pattern $\mathbf{s}_2$ projects to $\mathbf{s}_1$ with strength w_{FF}, and there is no projection back, it is a strictly feedforward connectivity between patterns, without loops. In addition, there are the self-loops $\mathbf{s}_1 \to \mathbf{s}_1$ with strength λ_1 and $\mathbf{s}_2 \to \mathbf{s}_2$ with strength λ_2.

(4) Show that $\mathbf{W}$ is non-normal if and only if $w_{FF} \neq 0$, in two ways. First, show this in the $\mathbf{s}_1, \mathbf{s}_2$ basis, by showing that $\mathbf{W}\mathbf{W}^{\mathrm{T}} = \mathbf{W}^{\mathrm{T}}\mathbf{W}$ if and only if $w_{FF} = 0$. Second, show this in the r_E, r_I basis: if and only if $w_{FF} = w_1 + w_2 + x = 0$, then the matrix $\mathbf{W}$ in this basis becomes a symmetric matrix, $\mathbf{W} = \mathbf{W}^{\mathrm{T}}$, and therefore is normal; and the eigenvectors become orthogonal, *i.e.* the second eigenvector becomes $\begin{pmatrix} -1 \\ 1 \end{pmatrix}$. Note that, for the matrix to be normal, the upper right entry of $\mathbf{W}$ in the original basis becomes positive and equal to w_2, so the matrix no longer describes an excitatory unit and an inhibitory unit.

(5) Solve the linear dynamics, Eq. 3, in the $\mathbf{s}_1, \mathbf{s}_2$ basis.

a. First, suppose $i_2(t)$ is time dependent. Show that the i_2-dependent part of the

[2] If this being the form of the matrix representation isn't clear to you: in a given basis of vectors $\mathbf{b}_1, \mathbf{b}_2$, we can write any vector $\mathbf{v}$ as $\mathbf{v} = v_1\mathbf{b}_1 + v_2\mathbf{b}_2 = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$. Suppose $\mathbf{W}\mathbf{b}_1 = w_{11}\mathbf{b}_1 + w_{21}\mathbf{b}_2$ and $\mathbf{W}\mathbf{b}_2 = w_{12}\mathbf{b}_1 + w_{22}\mathbf{b}_2$. Then $\mathbf{W}\mathbf{v} = (w_{11}v_1 + w_{12}v_2)\mathbf{b}_1 + (w_{21}v_1 + w_{22}v_2)\mathbf{b}_2 = \begin{pmatrix} w_{11}v_1 + w_{12}v_2 \\ w_{21}v_1 + w_{22}v_2 \end{pmatrix}$. This is what you get from applying $\begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}$ to $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$. Since this is true for an arbitrary vector $\mathbf{v}$, $\begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}$ is the representation of $\mathbf{W}$ in the $\mathbf{b}_1, \mathbf{b}_2$ basis.

A simpler way to see this is just to note that, in the $\mathbf{b}_1, \mathbf{b}_2$ basis, $\mathbf{b}_1 = (1, 0)^T$ and $\mathbf{b}_2 = (0, 1)^T$. Given this and $\mathbf{W}\mathbf{b}_1 = w_{11}\mathbf{b}_1 + w_{21}\mathbf{b}_2$ and $\mathbf{W}\mathbf{b}_2 = w_{12}\mathbf{b}_1 + w_{22}\mathbf{b}_2$, you can see that $\mathbf{W} = \begin{pmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{pmatrix}$.

response is

$$r_2(t) = \ldots + \frac{1}{\tau}\int_0^t dt'\, e^{-(1-\lambda_2)(t-t')/\tau} i_2(t') \quad (4)$$

$$r_1(t) = \ldots + \frac{w_{FF}}{\tau}\int_0^t dt'\, e^{-(1-\lambda_1)(t-t')/\tau} r_2(t') \quad (5)$$

$$= \ldots + \frac{w_{FF}}{\tau^2}\int_0^t dt' \int_0^{t'} dt''\, e^{-(1-\lambda_1)(t-t')/\tau} e^{-(1-\lambda_2)(t'-t'')/\tau} i_2(t'') \quad (6)$$

By using $\int_0^t dt' \int_0^{t'} dt'' = \int_0^t dt'' \int_{t''}^t dt'$ (show this; hint, draw the two-dimensional t'/t'' plane and sketch the region covered by the first double integral, and show that it is the same as the region covered by the second double integral) do the dt' integral to show that the double integral term becomes $\frac{w_{FF}}{\tau}\int_0^t dt''\, g_{\lambda_1,\lambda_2}(t-t'')i_2(t'')$ where $g_{\lambda_1,\lambda_2}(t) = \frac{e^{-(1-\lambda_1)t/\tau} - e^{-(1-\lambda_2)t/\tau}}{\lambda_1 - \lambda_2}$. This shows that, when i_2 is constant, the response at time t after its onset is just the integral from 0 to t of $g_{\lambda_1,\lambda_2}(t-t')$. This shows that g is the "impulse response function" telling the response of r_1 to an input to r_2; that is, if you give a δ pulse of input to r_2 at time 0, $i_2(t) = \delta(t)$,[3] then the response of r_1 at time t is proportional to $\int_0^t g_{\lambda_1,\lambda_2}(t-t')\delta(t') = g_{\lambda_1,\lambda_2}(t)$. If the integral of g is increased by becoming a pulse (rather than by becoming a slowed exponential, as in Hebbian amplification), this can give amplification without slowing (the amplification also arises from w_{FF} being large).

b. Now let's return to taking **i** to be constant, not varying in time. You will have to first solve for $r_2(t)$, then solve for $r_1(t)$ with $w_{FF}r_2(t)$ as one of the inputs. You should find

$$r_2(t) = r_2(0)e^{-(1-\lambda_2)t/\tau} + \frac{i_2}{1-\lambda_2}\left(1 - e^{-(1-\lambda_2)t/\tau}\right) \quad (7)$$

$$r_1(t) = r_1(0)e^{-(1-\lambda_1)t/\tau} + \frac{i_1 + w_{FF}i_2/(1-\lambda_2)}{1-\lambda_1}\left(1 - e^{-(1-\lambda_1)t/\tau}\right)$$
$$+ w_{FF}\left(r_2(0) - \frac{i_2}{1-\lambda_2}\right) g_{\lambda_1,\lambda_2}(t) \quad (8)$$

(6) Graph the function $g_{\lambda_1,\lambda_2}(t) = \frac{e^{-(1-\lambda_1)t/\tau} - e^{-(1-\lambda_2)t/\tau}}{\lambda_1 - \lambda_2}$ for some choices of λ_1 and λ_2 as real numbers less than 1. (You can take $\tau = 1$; this just sets the units of time.) Consider both positive, one positive and one negative, or both negative. How does this affect the time

[3] $\delta(t)$ is the Dirac delta function, defined by $\delta(t) = 0$ for $t \neq 0$ and $\int_{-\epsilon}^{\epsilon}\delta(t) = 1$ for any $\epsilon > 0$. It has the property that $\int dt' f(t')\delta(t'-t) = f(t)$ so long as t is within the integral limits. It is the continuous analogue of the discrete Kronecker delta function, $\delta_{ij} = 1$, $i = j$; $= 0$, otherwise, which satisfies the discrete analogue of that integral equation, $\sum_j x_j\delta_{ij} = x_i$. Examples of instantiations of the Dirac delta function include the limit, as $dt \to 0$, of a function equal to $1/dt$ on $-dt/2 \leq x \leq dt/2$ and equal to 0 otherwise; or $\lim_{\sigma\to 0}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-t^2/(2\sigma^2)}$.

course and height of the function?. Note that this is a transient pulse that rises with time constant $\min\left(\frac{\tau}{1-\lambda_1}, \frac{\tau}{1-\lambda_2}\right)$ and then falls with time constant given by the max of the two time constants. It can be quite large because it is multiplied by $w_{FF} = w_1 + w_2 + x$, which will be large if the weights are large. It is this transient that can cause the dynamics to transiently move far from the steady-state fixed point, although as $t \to \infty$ the dynamics will approach the fixed point.

(7) (a) For the case of both λ's negative, plot $g_{\lambda_1,\lambda_2}(t)$, $te^{-t/\tau}$ and $e^{-t/\tau}$ on the same plot, normalizing them all so (say) their peak is 1. How does the time course of g compare to the others? Note, $te^{-t/\tau}$ is the value of $g_{\lambda_1,\lambda_2}(t)$ for $\lambda_1 = \lambda_2 = 0$,[4] while $e^{-t/\tau}$ is the time course of decay in the absence of any recurrent connections.

(b) For the same choice of λ's: starting from the initial condition $r_E = 1$, $r_I = 0$ (translate this into $r_1(0)$ and $r_2(0)$), with no external input, compare the time course of $r_1(t)$ (Eq. 8, with $i_1 = i_2 = 0$) to the exponential time course $r_1(0)e^{-t/\tau}$ that would result just from the individual cell leaks in the absence of any recurrent connections. Also plot the timecourse of r_2, r_E, and r_I on the same plot. You can use (say) $w_{FF} = 10$. Make sure you understand why r_1 and r_2 behave as they do, in terms of the Schur weight matrix; and why this translates as it does into the behavior of r_E and r_I. Make the same graph starting from the initial condition of $r_E = 0$, $r_I = 1$ (note, this is a linear model, which arises from linearizing about a fixed point, so negative rates can arise and indicate rates that are negative relative to the fixed point firing rates).

(8) Now focus on the case of nonzero input.

(a) Let the initial condition be $r_1(0) = r_2(0) = 0$ and the steady input starting at time 0 be $i_E = 1$, $i_I = 0$. For the same parameters as in (7), plot the time course of response (over a long enough time to reach steady state by eye) of r_E. Also plot the response in the absence of recurrent connections, $r_E(t) = \frac{i_E}{\tau}\int_0^t dt'\, e^{-(t-t')/\tau} = i_E(1 - e^{-t/\tau})$. Also plot the latter curve scaled up to have the same steady state value as the former. How do their time courses compare? Can you get amplified response due to recurrence with time course faster than the time course without recurrence?

(b) Now consider the steady state for nonzero input, $r_2 = \frac{i_2}{1-\lambda_2}$, $r_1 = \frac{i_1 + w_{FF} i_2/(1-\lambda_2)}{1-\lambda_1}$ (from Eqs. 7-8). Note that small inputs i_2 to the difference of E and I can, for large w_{FF}, cause large steady state response of the sum of E and I, r_1. This is the effect underlying the paradoxical response: if i_2 is negative, representing tilting of the difference towards I, the result can be that I as well as E decreases in the new steady state. We'll study when precisely this occurs.

Consider an input only to I. Show that this corresponds to $i_2 = -i_1$, with the input to I

[4] To see this, compute $g_{\lambda_1,\lambda_1}(t)$ by letting $\lambda_2 = \lambda_1 + \epsilon$ and evaluate $\lim_{\epsilon\to 0} g_{\lambda_1,\lambda_2}(t) = \lim_{\epsilon\to 0} \frac{e^{-(1-\lambda_1)t/\tau}\left(1-e^{\epsilon t/\tau}\right)}{-\epsilon} = \lim_{\epsilon\to 0} \frac{e^{-(1-\lambda_1)t/\tau}\left(1-(1+\epsilon t/\tau+O(\epsilon^2))\right)}{-\epsilon} = \frac{t}{\tau}e^{(1-\lambda_1)t/\tau}$.

positive or negative according to whether i_1 is positive or negative. Show that for this case ($i_2 = -i_1$), the steady-state I response is $r_I = \frac{1}{\sqrt{2}}(r_1 - r_2) = \frac{i_1((1-\lambda_1)+(1-\lambda_2)-w_{FF})}{\sqrt{2}(1-\lambda_1)(1-\lambda_2)}$. Note that the nonnormal effect – the effect involving w_{FF} – has sign opposite to i_1; this represents a paradoxical effect (adding positive input to I cells causes their firing rates to decrease in the new steady state). The remaining terms represent the normal effect – the effect in a normal matrix in which the eigenvectors are the Schur basis vectors and are orthogonal, so $w_{FF} = 0$; these terms have the same sign as i_1, representing a non-paradoxical effect. Thus the paradoxical effect arises precisely when the effect of nonnormality exceeds the remaining effects. Intuitively, if you give negative input to r_2 and an equal but opposite positive input to r_1, this pushes r_2 down and pushes r_1 up, both representing a nonparadoxical increase in inhibition; but since r_2 has a feedforward connection to r_1, the pushing down of r_2 also pushes down r_1 by the nonnormal effect (the feedforward connection). When the nonnormal effect exceeds the normal effects, the result is a paradoxical change.
(9) Finally, show that the r_I response is paradoxical – negative for positive i_1 – if and only if $w_1 > 1$, which means if and only if the excitatory subnetwork by itself is unstable. To show this, you'll need to use the facts that $\lambda_2 < 1$ and $\lambda_1 < 1$.
In the general case – a weight matrix $\begin{pmatrix} w_{EE} & -w_{EI} \\ w_{IE} & -w_{II} \end{pmatrix}$ – if the eigenvectors and eigenvalues are real, the results are similar: each eigenvector has both of its entries of the same sign (which you can take to be positive), so they both represent weighted sums of E and I; and, taking one of them to be the first Schur vector, the 2nd Schur vector, which will make a feedforward connection to the first, must have its two entries of opposite signs, representing a weighted difference of E and I. So it remains true that differences in E and I are amplified, via w_{FF}, into sums of E and I. When the eigenvectors and eigenvalues are complex, it gets a little more complicated, but there is a sense in which the same picture continues to be true.