Derivation of the Robinson et al. (2002) transfer function

(Dated: August 24, 2022)

A. Our end goal

We want to derive Eq. (8) of Robinson et al. (2002), which is, in slightly updated notation:

$$\frac{\phi_e(\omega)}{\phi_n(\omega)} = \frac{G_{es}L}{1 - G_{ei}L} \frac{G_{sn}Le^{i\omega t_0/2}}{1 - G_{srs}L^2} \frac{1}{q^2 r_e^2}, \tag{1}$$

where

$$q^2 r_e^2 = \left(1 - \frac{i\omega}{\gamma_e}\right)^2 - \frac{L}{1 - G_{ei}L} \left[G_{ee} + \frac{(G_{ese} + G_{esre}L)L}{1 - G_{srs}L^2} e^{i\omega t_0}\right] \tag{2}$$

and

$$L = \frac{1}{(1 - i\omega/\alpha)(1 - i\omega/\beta)} \tag{3}$$

are both functions of ω. Constants $G_{ab} = \rho_a \nu_{ab}$ are gain parameters, where $\rho_a = S'\left(V_a^{(0)}\right)$ is the slope of the sigmoid $S(V)$ evaluated at steady state $V_a^{(0)}$. Some of the gains appear in combinations: $G_{ese} = G_{es}G_{se}$, $G_{esre} = G_{es}G_{sr}G_{re}$, and $G_{srs} = G_{sr}G_{rs}$.

For everything else, the reader should see the Robinson paper(s).

B. The general recipe

In a nutshell, what we want is to understand the response to linear perturbations around a fixed point. The process is *extremely* general. We want to:

1. find the fixed points;

2. linearize around the fixed points;

3. move to the frequency domain;

4. derive the transfer function.

The transfer function is powerful. It lets you calculate the power spectrum for a given input drive with known frequency content, the temporal response to an impulsive stimulus, and stability of the system.

So, we begin.

C. Fixed points

The 2002 version of the Robinson et al. corticothalamic model obeys the following system of 2nd order delay differential equations:

$$D_e\phi_e(t) = S[V_e(t)], \tag{4}$$

$$D_\alpha V_e(t) = \nu_{ee}\phi_e(t) + \nu_{ei}\phi_i(t) + \nu_{es}\phi_s(t - t_0/2), \tag{5}$$

$$D_\alpha V_s(t) = \nu_{se}\phi_e(t - t_0/2) + \nu_{sr}\phi_r(t) + \nu_{sn}\phi_n(t), \tag{6}$$

$$D_\alpha V_r(t) = \nu_{re}\phi_e(t - t_0/2) + \nu_{rs}\phi_s(t), \tag{7}$$

where $S(V)$ is the so-called sigmoidal function that converts from a cell-body potential V to a firing rate $Q = S(V)$ (in the paper this was $\Sigma(V)$),

$$S(V) = \frac{Q_{\max}}{1 + \exp[-(V - \theta)/\sigma']}, \tag{8}$$

and the differential operators D_e and D_α are defined as

$$D_e = \frac{1}{\gamma_e^2}\frac{d^2}{dt^2} + \frac{2}{\gamma_e}\frac{d}{dt} + 1, \tag{9}$$

$$D_\alpha = \frac{1}{\alpha\beta}\frac{d^2}{dt^2} + \left(\frac{1}{\alpha} + \frac{1}{\beta}\right)\frac{d}{dt} + 1. \tag{10}$$

These both describe a damped oscillator (actually, D_e is critically damped and D_α is overdamped). Note here compared to many of the other Robinson et al. papers we've already assumed spatially uniform solutions and hence have no spatial dependence. Turns out the linear analysis is also doable in that case and is only a tad more complicated (D_e then describes damped wave propagation and has a term $-r_e^2\nabla^2$, and all the temporal derivatives become $\frac{\partial}{\partial t}$). But we'll stick with temporal-only here.

Also if you compare back to the paper's Eqs (2), (6), and (7), we've already:

- expanded out the dummy subscripts;
- zeroed out the ν_{ab} terms that aren't in the model (cf. the paper's Fig. 1);
- made the approximation $V_e = V_i$ (this is a big one!), which also sets $\nu_{ie} = \nu_{ee}$ and $\nu_{ii} = \nu_{ei}$;
- made the approximations $\gamma_i \approx \gamma_s \approx \gamma_r \approx \infty$ (so there's only the D_e wave propagation operator), which further implies that $\phi_i = S(V_i)$, $\phi_s = S(V_s)$, and $\phi_r = S(V_r)$.

Now, onto stuff to do!

1. Equations (4)–(7) have only been partly simplified in light of the above assumptions. Use the above assumptions to re-write them such that the only dynamical variables are ϕ_e, V_e, V_s, V_r, and ϕ_n:

$$
\begin{aligned}
D_e\phi_e(t) &= S[V_e(t)], && (11)\\
D_\alpha V_e(t) &= \nu_{ee}\phi_e(t) + \nu_{ei}S[V_e(t)] + \nu_{es}S[V_s(t - t_0/2)], && (12)\\
D_\alpha V_s(t) &= \nu_{se}\phi_e(t - t_0/2) + \nu_{sr}S[V_r(t)] + \nu_{sn}\phi_n(t), && (13)\\
D_\alpha V_r(t) &= \nu_{re}\phi_e(t - t_0/2) + \nu_{rs}S[V_s(t)]. && (14)
\end{aligned}
$$

2. Next we find steady state (i.e., fixed point) solutions. Assuming there is no time dependence, show that at a steady state V_e satisfies

$$
f(V_e) = S\left\{\nu_{se}S(V_e) + \nu_{sr}S\left[\nu_{re}S(V_e) + \frac{\nu_{rs}}{\nu_{es}}[V_e - (\nu_{ee} + \nu_{ei})S(V_e)]\right] + \nu_{sn}\phi_n\right\} + \frac{1}{\nu_{es}}\left[(\nu_{ee} + \nu_{ei})S(V_e) - V_e\right] = 0. \quad (15)
$$

3. For a given set of 8 ν_{ab} values (and $\nu_{sn}\phi_n$ can be lumped together as a single parameter) you can plot $f(V_e)$ in Eq. (15) as a function of V_e and see how many zeros there are. A common scenario is that there's three, two stable and one unstable, where we're most interested in the lowest stable root, while the highest (stable) one has $S(V_e) \approx Q_{\max}$ and is somewhat unphysiological.

D. Linearize

Next we move to the linear approximation.

1. Assume that the dynamics is well described by small perturbations (1st-order terms with superscript $^{(1)}$) around steady state values (the 0th-order constant terms with superscript $^{(0)}$):

$$
\begin{aligned}
\phi_e(t) &= \phi_e^{(0)} + \phi_e^{(1)}(t), && (16)\\
V_a(t) &= V_a^{(0)} + V_a^{(1)}(t), \quad \text{for } a = e, s, r, && (17)\\
\phi_n(t) &= \phi_n^{(0)} + \phi_n^{(1)}(t), && (18)
\end{aligned}
$$

where $\phi_n^{(0)}$ is essentially a parameter for the mean level of input ϕ_n. Let $\rho_a = S'\left(V_a^{(0)}\right)$ denote the sigmoid slope evaluated at the steady state. Show that Equations (11)–(14) expanded to first order yield

$$D_e\phi_e^{(1)}(t) = \rho_e V_e^{(1)(t)} \tag{19}$$

$$D_\alpha V_e^{(1)}(t) = \nu_{ee}\phi_e^{(1)}(t) + \nu_{ei}\rho_e V_e^{(1)}(t) + \nu_{es}\rho_s V_s^{(1)}(t - t_0/2), \tag{20}$$

$$D_\alpha V_s^{(1)}(t) = \nu_{se}\phi_e^{(1)}(t - t_0/2) + \nu_{sr}\rho_r V_r^{(1)}(t) + \nu_{sn}\phi_n^{(1)}(t), \tag{21}$$

$$D_\alpha V_r^{(1)}(t) = \nu_{re}\phi_e^{(1)}(t - t_0/2) + \nu_{rs}\rho_s V_s^{(1)}(t). \tag{22}$$

You'll note the 0th-order terms have all cancelled out (this is a handy check that you've got no algebraic errors).

E. Fourier transform

Next, we want to understand how this linear system behaves. One path from here you might have seen in an applied maths course is to note that Equations (19)–(22) allow you to calculate the Jacobian, from which you can calculate eigenvalues and eigenvectors and work out e.g. whether your fixed point is stable. But, and it's a big but, our system has delays, so unlike the nice ODEs you might have seen it's actually quite tricky because the characteristic equation ends up being transcendental. Also, we're interested in what happens when you drive this system with noise. For that, moving to the Fourier domain is helpful because lots of noise sources have a relatively simple description there.

1. First, a digression into some Fourier transform identities, because integrals are fun. Let $\mathcal{F}\{y(t)\} = Y(\omega) = \int_{-\infty}^{\infty} y(t)e^{i\omega t}dt$ be the Fourier transform of a signal $y(t)$. (There's many conventions here, and this is a common physics one that we will use.) Show that $\mathcal{F}\{\frac{dy}{dt}\} = -i\omega Y(\omega)$.

2. Show that $\mathcal{F}\{y(t-\tau)\} = e^{i\omega t}Y(\omega)$.

3. Show that $\mathcal{F}\{D_\alpha V_e\} = (1 - \frac{i\omega}{\alpha})(1 - \frac{i\omega}{\beta})V_e(\omega) = L^{-1}(\omega)V_e(\omega)$, which defines $L(\omega)$. By the way, the inverse Fourier transform of $L(\omega)$ gives the temporal response to a delta-function impulsive stimulus—this is the Green function for the damped harmonic oscillator! (cf. debate over whether it's a Green function or a Green's function.)

4. Show that $\mathcal{F}\{D_e\phi_e\} = \left(1 - \frac{i\omega}{\gamma_e}\right)^2 \phi_e(\omega)$.

F. Derive the transfer function

What we want is to know how $\phi_e(\omega)$ responds to a given input $\phi_n(\omega)$. To do this, we want the transfer function $\dfrac{\phi_e(\omega)}{\phi_n(\omega)}$. There will be some algebra.

1. Fourier transform both sides of Equations (19)–(22) (drop the superscripts!).

2. Derive Equation (1). It's...messy. Proceed by eliminating the $V_a(\omega)$ terms (similar to when you calculated the fixed point), and defining $G_{ab} = \rho_a \nu_{ab}$ for convenience because as you go along you'll see these parameters pair up. Defining $q^2 r_e^2$ [Equation (2)] as you go seems a bit obscure but if you move things around in that direction you'll see it start to fall out neatly. The reason for the r_e^2 is that it makes sense when you're in the full spatial case; here it plays no role since $q^2 r_e^2$ only appear together.

And there you have it! Now you can calculate the analytic power spectrum using

$$P(\omega) = |\phi_e(\omega)|^2, \tag{23}$$

$$= \left|\frac{\phi_e(\omega)}{\phi_n(\omega)}\right|^2 |\phi_n(\omega)|^2, \tag{24}$$

for example by assuming the input is white noise such that $|\phi_n(\omega)|^2$ is constant.