Gatsby Computational Neuroscience Unit
Theoretical Neuroscience

Final examination, theoretical neuroscience
12 May 2020

Part II – long questions

There are four questions, one from each main section of the course. Please answer three out of the four, starting the answers for each question on a new page. Don't forget to write your name at the top of the answer to each question.

You have a maximum of 10 hours for this exam. You shouldn't need all of it, but we don't want you to stress about time.

Good luck!

1 Biophysics

Consider a simplified Hodgkin-Huxley type model, with membrane potential relative to the leak reversal potential,

$$\tau \frac{dV}{dt} = -V - m_\infty(V) h (V - \mathcal{E}) \tag{1a}$$

$$\tau_h \frac{dh}{dt} = h_\infty(V) - h \tag{1b}$$

where

$$m_\infty(V) = \Theta(V - V_t)$$

$$h_\infty(V) = \frac{1}{1 + \exp((V - V_h)/\epsilon_h)}$$

As usual, $\Theta(V - V_t)$ is the Heaviside step function: it's equal to 1 if $V > V_t$ and zero otherwise. The parameters are

$$\begin{aligned} \mathcal{E} &= 80 \text{ mV} \\ V_t &= 20 \text{ mV} \\ \epsilon_h &= 10 \text{ mV}\,. \end{aligned}$$

The remaining parameters, V_h, τ and τ_h, will be specified as needed.

1. Sketch $\tau dV/dt$ for h ranging from 0 to 1. At what value of V_h (typo: this should have been h, not V_h) is there a transition from zero to three fixed points of the membrane potential dynamics?

   (5 marks)

2. Sketch the nullclines in V-h space, with V on the x-axis and h on the y-axis, with $V_h = 25$. At what value of V_h is there a transition from one to three fixed points?

   (10 marks)

3. Assume that V_h is such that there are three fixed points. Consider the initial condition $h(t = 0) = 1$ and $V(t = 0)$ just slightly above V_t. The system evolves under the dynamics given in Eq. (1). Show that if $\tau \gg \tau_h$ or $\tau \ll \tau_h$, as $t \to \infty$ the voltage will asymptote to a value greater than V_t.

   (15 marks)

4. As in the previous question, assume that V_h is such that there are three fixed points, the initial conditions are $h(t=0)=1$ and $V(t=0)$ just slightly above V_t, and the system evolves under the dynamics given in Eq. (1). Show (graphically) that the amplitude of the "spike" (i.e., the maximum voltage) is an increasing function of τ_h.

   (10 marks)

Solutions

1. Using the fact that $m_\infty(V)$ turns on suddenly at $V = V_t = 20$ mV, we have

$$\tau \frac{dV}{dt} = \begin{cases} -V & V < V_t \\ h\mathcal{E} - (1+h)V & V \geq V_t\,. \end{cases} \tag{2}$$

   Those curves are plotted in Fig. 1 for $h = 0, 0.5$ and 1.

   The transition from one to three fixed points occurs when $h\mathcal{E} - (1+h)V_t = 0$, which happens at $h = V_t/(\mathcal{E} - V_t) = 1/3$.

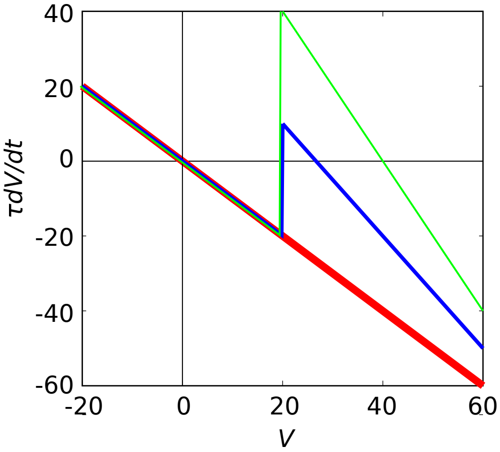

Figure 1: $\tau dV/dt$ versus V for $h = 0$ (red), 0.5 (blue) and 1 (green).

2. The h-nullcline is easy; it's just Eq. (1b). That's shown in red in Fig. 2. For the V-nullcline, we'll use Eq. (2). When $V < V_t$, one of the nullclines is at $V = 0$; that's the blue line on the left edge of Fig. 2. There's also a nullcline formed by the line $V = V_t$, $h \geq 1/3$ (see answer to question 1). That's the other vertical line in Fig. 2. Finally, for $V > V_t$, the nullcline is given by $h\mathcal{E} = (1+h)V$; solving for h in terms of V gives $h = V/(\mathcal{E} - V)$. That's the curved blue line in Fig. 2.

   The transition from one to three fixed points occurs when the red nullcline hits the lower tip of the blue nullcline. The tip is at $(V, h) = (V_t, 1/3)$, so that happens when $1/3 = h_\infty(V_t)$. Using Eq. (2) for $h_\infty(V)$ and doing a bunch of algebra, I get

$$\frac{V_t - V_h}{\epsilon_h} = \ln 2\,.$$

   Solving for V_h gives $V_h = V_t - \epsilon_h \ln 2 \approx 13$ mV.

3. If $\tau \gg \tau_h$, then the dynamics relaxes very rapidly almost to the h-nullcline (the red curve); after that it drifts along the red curve to the fixed point. That's the brown trajectory in Fig. 2. If, on the other hand, $\tau_h \gg \tau$, then the dynamics relaxes very rapidly almost to the V-nullcline (blue curve); after that it drifts down the blue curve (which it quickly crosses) to the fixed point. That's the green trajectory in Fig. 2.

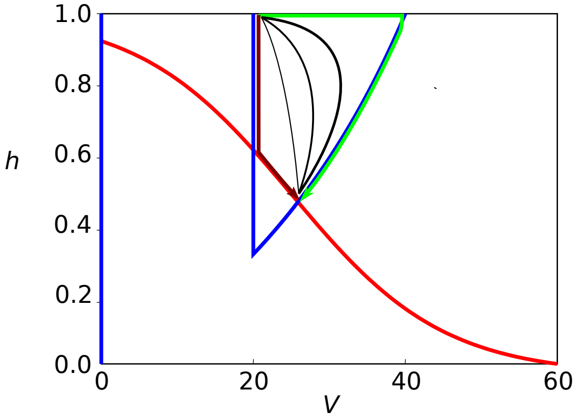

Figure 2: Red: h-nullcline; blue: V-nullcline. Brown: trajectory when $\tau \gg \tau_h$; green: trajectory when $\tau_h \gg \tau$. The black trajectories corresponding to increasing τ_h (thin to thick = smaller to larger).

4. As τ_h goes from small to large values, the curves interpolate smoothly between the brown and green ones in Fig. 2. Several such curves are shown in black, with thinner lines corresponding to smaller τ_h. Thus, the amplitude of the "spike" grows.

2 Networks

Consider a network of excitatory and inhibitory neurons in which the average excitatory and inhibitory firing rates evolve according to

$$
\begin{aligned}
\tau \frac{d\nu_E}{dt} &= \phi\big(J_{EE}\,\nu_E - J_{EI}\nu_I + h_E\big) - \nu_E \\
\tau \frac{d\nu_I}{dt} &= \phi\big(J_{IE}\,\nu_E - J_{II}\nu_I + h_I\big) - \nu_I
\end{aligned}
$$

where $\phi(x)$ is threshold-linear-saturating,

$$
\phi(x) = \begin{cases} 0 & x < 0 \\ \Gamma x & 0 \le x < \nu_{\max}/\Gamma \\ \nu_0 & x \ge \nu_{\max}/\Gamma\,. \end{cases}
$$

ν_0 should be $\nu_{\max}$.

1. Sketch the nullclines in the limit $\Gamma \to \infty$. Put ν_E on the x-axis and ν_I on the y-axis.

   (10 marks)

2. What are the conditions on J_{QR} and h_Q $(Q, R \in E, I)$ that guarantee an equilibrium with both ν_E and ν_I greater than 0 and less than $\nu_{\max}$?

   (10 marks)

3. What are the conditions on J_{QR} and h_Q that guarantee a periodic orbit?

   (10 marks)

4. Assume you're in a regime with a stable fixed point, and both h_E and h_I are positive. Add spike-frequency adaptation,

   $$\tau_h \frac{dh_E}{dt} = -\nu_E h_E\,.$$

   Consider the limits $\tau_h \gg \tau$ and $\Gamma \to \infty$. Write down a differential equation for $h_E(t)$. Assuming that $h_E(t=0) > 0$, find the equilibrium of that equation, and write down an explicit expression for its behavior near the equilibrium.

   (10 marks)

Solutions

1. In the limit $\Gamma \rightarrow$, the argument of ϕ must be zero. This tells us that

$$\nu_I = \frac{J_{EE}\nu_E}{J_{EI}} + \frac{h_E}{J_{EI}} \tag{1a}$$

$$\nu_I = \frac{J_{IE}\nu_E}{J_{II}} + \frac{h_I}{J_{II}} \tag{1b}$$

where the first equation is for the excitatory nullcline and the second is for the inhibitory nullcline. In addition, of course, the firing rates must lie between 0 and $\nu_{\max}$. The above expressions are plotted in Fig. 1, with red for the excitatory nullcline and blue for the inhibitory one.

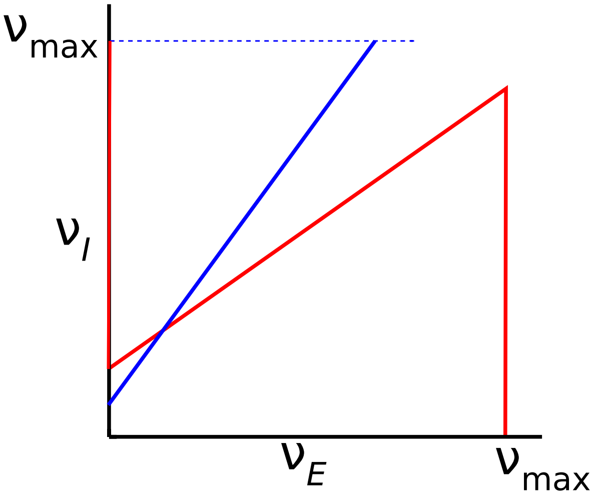

Figure 1: Red: excitatory nullcline; blue: inhibitory nullcline.

2. It's straightforward – but tedious – to actually answer this question: just solve for the fixed points, and find conditions under which they're positive but less than $\nu_{\max}$. However, I won't do that; instead, I'll answer the question I should have asked, which is: under what conditions do the nullclines look like they do in Fig. 1? To simplify things, let $J_{EE} \rightarrow J_{EE} - 1/\Gamma$ and $J_{II} \rightarrow J_{II} + 1/\Gamma$. And I'm not going to worry about exceeding $\nu_{\max}$.

To satisfy the conditions on the firing rates, we only need to examine Fig 1, which tells us two things: the excitatory y-intercept must be above the inhibitory one, and the inhibitory slope must be larger than the excitatory one. To satisfy the first condition, we must have

$$\frac{h_E}{J_{EI}} > \frac{h_I}{J_{II}} \, .$$

And to satisfy the second,

$$\frac{J_{IE}}{J_{II}} > \frac{J_{EE}}{J_{EI}} \, ,$$

which can be written succinctly as $D > 0$ where D is the determinant of the linearized dynamics (with the factor of Γ taken out) in the non-saturating regime,

$$D \equiv J_{IE} J_{EI} - J_{II} J_{EE} \, .$$

Finally, for stability (which I didn't ask for, but I meant to), we need both eigenvalues of the linearized dynamics to be positive. These are given by the usual expressions,

$$\lambda_{\pm} \propto \frac{T \pm \sqrt{T^2 - 4D}}{2}$$

where (again taking out the factor of Γ) $T \equiv J_{EE} - J_{II}$ is the trace. For both eigenvalues to be negative we need $T < 0$ and $D > 0$. We already have $D > 0$, so we just need $T < 0$, which means

$$J_{II} > J_{EE} \, .$$

3. As above, to simplify things, I'm going to let $J_{EE} \to J_{EE} - 1/\Gamma$ and $J_{II} \to J_{II} + 1/\Gamma$, and I'm not going to worry about exceeding $\nu_{\max}$.

   For a periodic orbit, we just need the equilibrium to be unstable via a Hopf bifurcation, which mean the trace must become positive. So we keep all conditions from the previous question, but now we have

$$J_{II} < J_{EE} \, .$$

4. Because $\tau_h \gg \tau$, the firing rates are always at their equilibrium given h_E, and as h_E evolves that equilibrium changes slowly. Solving Eq. (1) for ν_E in terms of h_E gives us

$$\nu_E = \frac{J_{II} h_E - J_{EI} h_I}{D} \, .$$

   Inserting this into the equation for h_E yields

$$D \tau_h \frac{d h_E}{dt} = -h_E (J_{II} h_E - J_{EI} h_I) \, .$$

   Since $h_E(t = 0) > 0$, this asymptotes to $h_E = J_{EI} h_I / J_{II}$. Near that equilibrium, the effective time constant is $D\tau_h / J_{EI} h_I$. Thus, asymptotically,

$$h_E - \frac{J_{EI} h_I}{J_{II}} \sim e^{-t J_{EI} h_I / (\tau_h D)} \, .$$

3 Coding

Suppose that the early stages of visual processing in a model animal are captured by the process of inference in an overcomplete linear factor generative model with Gabor basis functions. That is, neural activities in V1 reflect latent variables $\mathbf{z}$ that relate to the retinal ganglion cell (RGC) activities $\mathbf{x}$ according to:

$$\mathbf{x} \sim \mathcal{N}\left(\boldsymbol{\Phi}\mathbf{z}, \boldsymbol{\Psi}\right)$$
$$\mathbf{z} \sim \pi(\mathbf{z})$$

with $\mathsf{dim}(\mathbf{z}) > \mathsf{dim}(\mathbf{x})$, and each column of $\boldsymbol{\Phi}$ corresponding to a vectorised Gabor wavelet.

Suppose that π is a standard Normal distribution $\mathcal{N}(0, I)$, that the noise term in the model is isotropic ($\boldsymbol{\Psi} = \psi I$) and that neural firing rates $\mathbf{r}$ reflect the posterior mode of $P(\mathbf{z}|\mathbf{x})$.

1. What is the relationship expected between RGC activities $\mathbf{x}$ and V1 firing rates $\mathbf{r}$.

   ANSWER: The answer is given by linear regression under the expected covariances. Write

   $$\mathsf{W} = \left\langle \mathbf{z}\mathbf{x}^\mathsf{T} \right\rangle \left\langle \mathbf{x}\mathbf{x}^\mathsf{T} \right\rangle^{-1} = \Phi^\mathsf{T}(\Phi\Phi^\mathsf{T} + \psi I)^{-1} = (I + \Phi^\mathsf{T}\psi^{-1}\Phi)^{-1}\Phi^\mathsf{T}\psi^{-1} = (\psi I + \Phi^\mathsf{T}\Phi)^{-1}\Phi^\mathsf{T}$$

   Then $\mathbf{r} = \mathsf{W}\mathbf{x}$.

In an experiment, an image $\mathbf{I}_0$ corresponding to a single Gabor wavelet is presented to the animal. Noise in transduction means that RGC activities are drawn from $\mathcal{N}(\mathbf{I}_0, \Sigma_n)$.

2. Assuming that the parameters $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ of the generative model have been adapted to RGC activity during development (and that the wavelets in $\boldsymbol{\Phi}$ are constrained to spatial frequencies well below Nyquist), what relationship would you expect to hold between Ψ and Σ_n?

   ANSWER: We expect that $\Psi \succ \Sigma_n$. That is, that the model will capture the transduction noise, plus further "noise" that describes model mismatch.

3. The experiment is repeated many times using the same stimulus. What is the expected covariance observed in V1 activities? Does it depend on $\mathbf{I}_0$?

   ANSWER: The covariance is $\mathsf{W}\Sigma_n\mathsf{W}^\mathsf{T}$, so independent of $\mathbf{I}_0$.

Now suppose the experiment uses a set $\{\mathbf{I}(\theta)\}$ of images of Gabor wavelets, all with the same envelope, spatial frequency and phase, but with varying orientations θ. Consider only neurons that encode latents associated with Gabor basis functions with same envelope, spatial frequency and phase.

4. Compute the Fisher information available in this population about θ. Is there any difference between this basis-function based view and the tuning curve approach to Fisher information which is more commonly employed?

4 Learning

Consider a run-of-the-mill recurrent neural network,

$$\frac{d\mathbf{x}}{dt} = \phi(\mathbf{w}\cdot\mathbf{x}) \tag{1}$$

where, as usual, ϕ is monotonic non-decreasing. We want to minimize the loss, S, given by

$$S(\mathbf{w}) = \int_0^T dt\, L(\mathbf{x}(t))$$

where $\mathbf{x}(t)$ in this expression is a solution to Eq. (1) Typically there's an average over initial conditions, but that won't come up here.

In class we used Lagrange multipliers to derive learning rules. But we should be able to derive them using standard backprop, which is what you're going to do now.

Our strategy is to let $\mathbf{w} \to \mathbf{w} + \delta\mathbf{w}$ with $\delta\mathbf{w}$ infinitesimally small, and determine how much the loss, $S(\mathbf{w})$, changes. That is, we'll compute $S(\mathbf{w}+\delta\mathbf{w}) - S(\mathbf{w})$. That in turn will allow us to compute $\partial S(\mathbf{w})/\partial\mathbf{w}$.

1. Let $\mathbf{w} \to \mathbf{w} + \delta\mathbf{w}$, with $\delta\mathbf{w}$ infinitesimally small. Let $\mathbf{x}(t)$ be the solution to Eq. (1) when the weights are set to $\mathbf{w}$, and $\mathbf{x}(t) + \delta\mathbf{x}(t)$ the solution when the weights are set to $\mathbf{w} + \delta\mathbf{w}$. Show that $\delta\mathbf{x}(t)$ evolves according to

$$\frac{d\delta\mathbf{x}}{dt} = \mathbf{\Phi}\cdot\mathbf{w}\cdot\delta\mathbf{x} + \mathbf{\Phi}\cdot\delta\mathbf{w}\cdot\mathbf{x} \tag{2}$$

   where $\mathbf{\Phi}$ is a diagonal matrix with diagonal elements

$$\Phi_{ii} = \phi'\left(\sum_j w_{ij}x_j\right) . \tag{3}$$

   Note that $\mathbf{\Phi}$ depends on time, a dependence we (usually) drop to reduce clutter. What are the initial conditions?

   (5 marks)

2. Given the above, we could compute the change in loss associated with a weight change, $\delta\mathbf{w}$, simply by using the definition of the loss function,

$$S(\mathbf{w}+\delta\mathbf{w}) - S(\mathbf{w}) = \int_0^T dt\, [L(\mathbf{x}(t) + \delta\mathbf{x}(t,\delta\mathbf{w})) - L(\mathbf{x})] = \int_0^T dt\, \frac{\partial L(\mathbf{x}(t))}{\partial\mathbf{x}(t)}\cdot\delta\mathbf{x}(t,\delta\mathbf{w})\, . \tag{4}$$

   The problem, of course, is that we can't solve Eq. (2) analytically for $\delta\mathbf{x}(t,\delta\mathbf{w})$, and a numerical solution would be extremely expensive: we would have to solve the equation for each value of δw_{ij}. So, for reasons that will become obvious soon, let's define a new variable, $\delta\mathbf{x}_\tau(t)$, which evolves according to

$$\frac{d\mathbf{x}_\tau}{dt} = \mathbf{\Phi}\cdot\mathbf{w}\cdot\mathbf{x}_\tau + \mathbf{\Phi}\cdot\delta\mathbf{w}\cdot\mathbf{x}\,\delta(t-\tau)\, , \tag{5}$$

where $\delta(t-\tau)$ is the Dirac delta function, $\delta\mathbf{w}$ is, as above, infinitesimally small, and, for clarity, we dropped the explicit dependence on $\delta\mathbf{w}$ (although it really is there). The initial conditions are $\mathbf{x}_\tau(t) = 0$.

Show that

$$S(\mathbf{w} + \delta\mathbf{w}) - S(\delta\mathbf{w}) = \int_0^T d\tau \int_\tau^T dt\, \frac{\partial L(\mathbf{x}(t))}{\partial \mathbf{x}(t)} \cdot \mathbf{x}_\tau(t)\,. \tag{6}$$

(5 marks)

3. Define $\boldsymbol{\Gamma}_\tau(t)$ via

$$\mathbf{x}_\tau(t) = \boldsymbol{\Gamma}_\tau(t) \cdot \mathbf{x}_\tau(\tau)$$

(which implies that $\boldsymbol{\Gamma}_\tau(\tau)$ is the identity). Show that

$$\mathbf{x}_\tau(\tau) = \boldsymbol{\Phi}(\tau) \cdot \delta\mathbf{w} \cdot \mathbf{x}(\tau)\,.$$

This question is a bit misleading, since you don't need to know anything about $\boldsymbol{\Gamma}_\tau$ to answer it. Also, it should really read $\mathbf{x}_\tau(\tau^+)$, where τ^+ means after the delta function.

(5 marks)

4. Show that when Δt is infinitesimal,

$$\boldsymbol{\Gamma}_\tau(t) = \boldsymbol{\Gamma}_{\tau+\Delta t}(t)[\mathbf{I} + \Delta t \boldsymbol{\Phi}(\tau) \cdot \mathbf{w}]\,.$$

(5 marks)

5. Define

$$\mathbf{z}(\tau) \equiv \int_\tau^T dt\, \frac{\partial L(\mathbf{x}(t))}{\partial \mathbf{x}(t)} \cdot \boldsymbol{\Gamma}_\tau(t)\,. \tag{7}$$

Show that

$$\frac{d\mathbf{z}(\tau)}{d\tau} = -\frac{\partial L(\mathbf{x}(\tau))}{\partial \mathbf{x}(\tau)} - \mathbf{z}(\tau) \cdot \boldsymbol{\Phi}(\tau) \cdot \mathbf{w}\,, \tag{8}$$

and it has final conditions $\mathbf{z}(T) = 0$ (which follows from its definition). The fact that $\mathbf{z}(t)$ has to be integrated backwards in time is no surprise, since we're really doing backprop.

(10 marks)

Hint: first compute $\mathbf{z}(\tau) - \mathbf{z}(\tau + \Delta\tau)$; then let $\Delta\tau$ go to zero.

6. Use the above analysis to compute

$$\frac{\partial S(\mathbf{w})}{\partial \mathbf{w}}$$

in terms of quantities that are (relatively) easy to compute, such as $\mathbf{z}(\tau)$ and $\mathbf{x}(t)$.

(10 marks)

Solutions

1. We want the evolution equations when $\mathbf{w} \to \mathbf{w} + \delta\mathbf{w}$ with $\delta\mathbf{w}$ infinitesimally small. Letting $\mathbf{x}(t) \to \mathbf{x}(t) + \delta\mathbf{x}(t)$ where $\mathbf{x}(t)$ is the solution to Eq. (1), we have

$$\frac{d\mathbf{x}}{dt} + \frac{d\delta\mathbf{x}}{dt} = \phi(\mathbf{w}\cdot\mathbf{x} + \mathbf{w}\cdot\delta\mathbf{x} + \delta\mathbf{w}\cdot\mathbf{x})\,.$$

   Taylor expanding and using Eq. (1) gives us

$$\frac{d\delta\mathbf{x}}{dt} = \boldsymbol{\Phi}\cdot(\mathbf{w}\cdot\delta\mathbf{x} + \delta\mathbf{w}\cdot\mathbf{x})$$

   where $\boldsymbol{\Phi}$ is given in Eq. (3).

   Because $\delta\mathbf{x}$ must be identically zero when $\delta\mathbf{w} = 0$, the initial conditions must be $\delta\mathbf{x}(t=0) = 0$.

2. Comparing Eqs. (4) and (6), we see that we need to show that

$$\int_0^T d\tau\, \mathbf{x}_\tau(t) = \delta\mathbf{x}(t) \tag{9}$$

   where we dropped the $\delta\mathbf{w}$ dependence. Note that

$$\int_0^T d\tau\, \frac{d\mathbf{x}_\tau}{dt} = \int_0^T d\tau\, \boldsymbol{\Phi}(t)\cdot\mathbf{w}\cdot\mathbf{x}_\tau(t) + \boldsymbol{\Phi}\cdot\delta\mathbf{w}\cdot\mathbf{x}(t)\delta(t-\tau)\,.$$

   Now use the following: the integral over τ commutes with the derivative with respect to t, $\boldsymbol{\Phi}(t)$ doesn't depend on τ, and the integral over the Dirac delta function yields 1. We thus have

$$\frac{d}{dt}\int_0^T d\tau\, \mathbf{x}_\tau(t) = \boldsymbol{\Phi}(t)\cdot\mathbf{w}\cdot\int_0^T d\tau\, \mathbf{x}_\tau(t) + \boldsymbol{\Phi}\cdot\delta\mathbf{w}\cdot\mathbf{x}(t)\,.$$

   Consequently, $\int_0^T d\tau\, \mathbf{x}_\tau(t)$ obeys the same equation as $\delta\mathbf{x}(t)$. It also has the same initial condition. Since the solution to ODEs are unique (except in pathological cases), Eq. (9) must be satisfied.

3. $\mathbf{x}_\tau(\tau)$ is the value of $\mathbf{x}_\tau$ right after the jump associated with the delta function (remember that $\mathbf{x}_\tau = 0$ before the delta function). The size of this jump is, from Eq. (5), $\boldsymbol{\Phi}(\tau)\cdot\delta\mathbf{w}\cdot\mathbf{x}(\tau)$.

4. By definition, $\boldsymbol{\Gamma}_\tau(t)$ is an operator that pushes vectors along the trajectory defined by Eq. (5), starting immediately after $t = \tau$. We can do this in two steps: first push the vector from τ to $\tau + \Delta t$, then push it from $\tau + \Delta t$ to t. The latter push is done by $\boldsymbol{\Gamma}_{\tau+\Delta t}(t)$; the former is done by a matrix that we'll call $\mathbf{I} + d\boldsymbol{\Gamma}_\tau$ where $\mathbf{I}$ is the identity matrix. This gives us

$$\boldsymbol{\Gamma}_\tau(t) = \boldsymbol{\Gamma}_{\tau+\Delta t}(t)\cdot(\mathbf{I} + d\boldsymbol{\Gamma}_\tau)\,,$$

   valid so long at $t \geq \tau + \Delta t$. Setting $t = \tau + \Delta t$, at which point $\boldsymbol{\Gamma}_{\tau+\Delta t}(\tau + \Delta t) = \mathbf{I}$, we have

$$\boldsymbol{\Gamma}_\tau(\tau + \Delta t) = (\mathbf{I} + d\boldsymbol{\Gamma}_\tau)\,.$$

Now note that for a vector $\mathbf{x}$ that obeys Eq. (5), we have, for small Δt,

$$\boldsymbol{\Gamma}_\tau(\tau + dt) \cdot \mathbf{x}(\tau) = \mathbf{x}(\tau) + \Delta t \boldsymbol{\Phi}(\tau) \cdot \mathbf{w} \cdot \mathbf{x}(\tau) \,.$$

Comparing this to the equation above it, we see that

$$d\boldsymbol{\Gamma}_\tau = \Delta t \boldsymbol{\Phi}(\tau) \cdot \mathbf{w} \,.$$

5. This is a straightforward, if somewhat tedious, calculation. Start with

$$\begin{aligned}\mathbf{z}(\tau) - \mathbf{z}(\tau + \Delta t) &= \int_\tau^T dt\, \frac{\partial L(\mathbf{x}(t))}{\partial \mathbf{x}(t)} \cdot \boldsymbol{\Gamma}_\tau(t) - \int_{\tau+\Delta t}^T dt\, \frac{\partial L(\mathbf{x}(t))}{\partial \mathbf{x}(t)} \cdot \boldsymbol{\Gamma}_{\tau+\Delta t}(t) \\ &= \int_\tau^{\tau+\Delta t} dt\, \frac{\partial L(\mathbf{x}(t))}{\partial \mathbf{x}(t)} \cdot \boldsymbol{\Gamma}_\tau(t) + \int_{\tau+\Delta t}^T dt\, \frac{\partial L(\mathbf{x}(t))}{\partial \mathbf{x}(t)} \cdot \big(\boldsymbol{\Gamma}_\tau(t) - \boldsymbol{\Gamma}_{\tau+\Delta t}(t)\big) \,.\end{aligned}$$

Because $\boldsymbol{\Gamma}_\tau(\tau) = \mathbf{I}$, to lowest nonvanishing order in Δt the first term in the second line is $\Delta t \partial L(\mathbf{x}(\tau))/\partial \mathbf{x}(\tau)$. Using the solution to the previous problem, the second term in the second line is, again to lowest nonvanishing order in Δt

$$\Delta t \int_\tau^T dt \frac{\partial L(\mathbf{x}(t))}{\partial \mathbf{x}(t)} \cdot \boldsymbol{\Gamma}_\tau(t) \cdot \boldsymbol{\Phi}(\tau) \cdot \mathbf{w} = \Delta t \mathbf{z}(\tau) \cdot \boldsymbol{\Phi}(\tau) \cdot \mathbf{w} \,.$$

We thus have

$$\frac{\mathbf{z}(\tau) - \mathbf{z}(\tau + \Delta t)}{\Delta t} = \frac{\partial L(\mathbf{x}(\tau))}{\partial \mathbf{x}(\tau)} + \mathbf{z}(\tau) \cdot \boldsymbol{\Phi}(\tau) \cdot \mathbf{w} \,.$$

Taking the limit $\Delta t \to 0$ gives the desired result.

6. First use $\mathbf{x}_\tau(t) = \boldsymbol{\Gamma}_\tau(t) \cdot \mathbf{x}_\tau(\tau)$, and insert that into Eq. (6); then use Eq. (7). That gives

$$S(\mathbf{w} + \delta \mathbf{w}) - S(\mathbf{w}) = \int_0^T d\tau\, \mathbf{z}(\tau) \cdot \mathbf{x}_\tau(\tau) \,.$$

Now use the fact that $\mathbf{x}_\tau(\tau) = \boldsymbol{\Phi}(\tau) \cdot \delta \mathbf{w} \cdot \mathbf{x}(\tau)$ to write

$$S(\mathbf{w} + \delta \mathbf{w}) - S(\mathbf{w}) = \int_0^T d\tau\, \mathbf{z}(\tau) \cdot \boldsymbol{\Phi}(\tau) \cdot \delta \mathbf{w} \cdot \mathbf{x}(\tau) \,.$$

Taking a derivative with respect to $\delta \mathbf{w}$ yields

$$\frac{\partial S(\mathbf{w})}{\partial \mathbf{w}} = \int_0^T d\tau\, \mathbf{z}(\tau) \cdot \boldsymbol{\Phi}(\tau)\, \mathbf{x}(\tau)$$

where I'm using the notation that two adjacent vectors is an outer product. This is the famous backprop through time update rule.