Linear response theory and Optimal kernel Method

Quan Wen

Sep 28 2023

Functional Derivative

The response of a sensory neuron in the early sensory system at time t can be viewed as a functional r given the sensory stimulus $\mathcal{S}$. What I mean more precisely is $r : \mathcal{S} \mapsto \Re$, namely the response of a neuron at time t maps the space of stimulus function $\mathcal{S}$ to a real number (e.g., the firing rate). This is because in general the response of the neuron evaluated at t could depend on the entire stimulus history of $s(t-\tau)$, where $\tau \in [0, \infty]$.

We can discretize time into bins with bin size Δt, whose centers are at positions $\tau_i, i = 1, 2, ...$, and define $s_i = s(t - \tau_i)$. Now $\tilde{r}(s_1, s_2, ...)$ is a multivariable function depending on s_i. By Taylor series expansion, we have

$$\tilde{r}(\mathbf{s} + \delta\mathbf{s}) = \tilde{r}(\mathbf{s}) + \sum_i \frac{\partial \tilde{r}}{\partial s_i}\Delta s_i + \frac{1}{2}\sum_{i,j}\frac{\partial^2 \tilde{r}}{\partial s_i \partial s_j}\Delta s_i \Delta s_j + ..., \tag{1}$$

To take off the hat and regain the functional r, we can replace $\sum_i \to \int dt$ when $\lim_{\Delta t \to 0}$. This can be done in the following way

$$\begin{aligned} r[s + \delta s] = r(s) + \int d\tau \frac{\delta r}{\delta s(\tau)}\delta s(\tau) \\ + \frac{1}{2}\int d\tau d\tau' \frac{\delta^2 r}{\delta s(\tau)\delta s(\tau')}\delta s(\tau)\delta s(\tau') + ... \end{aligned} \tag{2}$$

Here for clarity, I eliminate t in the expression, but one should remember that $s(\tau) \equiv s(t - \tau)$. Note that

$$\begin{aligned} \frac{\delta r}{\delta s(\tau)} &= \lim_{\Delta t \to 0}\frac{\partial \tilde{r}}{\partial s_i}\frac{1}{\Delta t} \\ \frac{\delta^2 r}{\delta s(\tau)\delta s(\tau')} &= \lim_{\Delta t \to 0}\frac{\partial^2 \tilde{r}}{\partial s_i \partial s_j}\frac{1}{\Delta t^2}. \end{aligned} \tag{3}$$

A heuristic definition of the functional derivative can now be written down as

$$\begin{aligned}\frac{\delta r}{\delta s(\tau)} &= \lim_{\Delta t\to 0}\lim_{\epsilon\to 0}\frac{\tilde{r}(s_1,...,s_i+\epsilon,...)-\tilde{r}(s_1,...,s_i,...)}{\Delta t\epsilon} \\ &= \lim_{\Delta t\to 0}\lim_{\epsilon\to 0}\frac{\tilde{r}(s_1,...,s_i+\epsilon\frac{1}{\Delta t},...)-\tilde{r}(s_1,...,s_i,...)}{\epsilon}\end{aligned} \tag{4}$$

In the physics literature, we formally define the functional derivative by considering a test function

$$\frac{\delta r}{\delta s(\tau)} = \lim_{\epsilon\to 0}\frac{r[s(t)+\epsilon\delta(t-\tau)]-r[s(t)]}{\epsilon} \tag{5}$$

where $\delta(t-\tau)$ is the delta function. To avoid confusion, I would like to emphasize that $r[s(t)]$ is a functional, not a function of $s(t)$ at a particular time t.

Functional derivative is very similar to function derivative, and there are convenient rules we would like to follow. For example, if h and f are both functions, we have

$$\begin{aligned}\frac{\delta f(x)}{\delta f(y)} &= \delta(x-y) \\ \frac{\delta h(f(x))}{\delta f(y)} = h'\frac{\delta f(x)}{\delta f(y)} &= h'\delta(x-y)\end{aligned} \tag{6}$$

Another familiar example is what you have learned in classical mechanics.

$$S(q,\dot{q},t) = \int_a^b L(q,\dot{q},t)dt \tag{7}$$

where $\dot{q} = dq/dt$ and the values of $q(a)$ and $q(b)$ are specified. We would like to identify the function q that minimizes or maximizes S. Just like in the multivariate calculus, the necessary condition is

$$\frac{\delta S}{\delta q(t')} = 0 \tag{8}$$

To compute

$$\begin{aligned}\frac{\delta S}{\delta q(t')} &= \int_a^b dt\left[\frac{\partial L(q,\dot{q},t)}{\partial q}\delta(t-t') + \frac{\partial L(q,\dot{q},t)}{\partial \dot{q}}\dot{\delta}(t-t')\right] \\ = \frac{\partial L(q,\dot{q},t')}{\partial q} + \int_a^b dt\left[\frac{d}{dt}\left(\frac{\partial L(q,\dot{q},t)}{\partial \dot{q}}\delta(t-t')\right) - \delta(t-t')\frac{d}{dt}\left(\frac{\partial L(q,\dot{q},t)}{\partial \dot{q}}\right)\right] \\ &= \frac{\partial L(q,\dot{q},t')}{\partial q} - \frac{d}{dt'}\left(\frac{\partial L(q,\dot{q},t')}{\partial \dot{q}}\right)\end{aligned}$$

As a result, we obtain the famous Euler-Lagrange equation

$$\frac{\partial L(q,\dot{q},t)}{\partial q} - \frac{d}{dt}\left(\frac{\partial L(q,\dot{q},t)}{\partial \dot{q}}\right) = 0 \tag{9}$$

Optimal Wiener Kernel

Now, let us go back to computational neuroscience. Consider the problem that from the history of the stimulus, we would like to estimate or predict the response of this sensory neuron $r_{est}(t)$. Let us consider the simplest functional model, namely

$$r_{est}(t) = r_0 + \int_0^\infty D(\tau)s(t-\tau)d\tau \tag{10}$$

r_0 is considered to be a constant. As we can see, if the functional derivative of r with respect to s does not depend on the choice of t, then

$$\frac{\delta r}{\delta s(t-\tau)} = D(\tau), \tag{11}$$

$D(\tau)$ is also called the first Wiener Kernel in the literature.

Now let us define an error function,

$$E = \int_0^T dt \left(r_{est}(t) - r(t)\right)^2 \tag{12}$$

Our goal is to minimize the error function and find the optimal kernel.

Therefore, we have

$$\begin{aligned} \frac{\delta E}{\delta D} = 0 &= \int_0^T dt(r_{est}(t) - r(t))\frac{\delta r_{est}}{\delta D}. \\ \frac{\delta r_{est}}{\delta D(\tau)} &= \int_0^T d\tau' s(t-\tau')\delta(\tau'-\tau) = s(t-\tau). \end{aligned} \tag{13}$$

As a result, we have

$$\int_0^T dt \int_0^\infty d\tau' D(\tau')s(t-\tau')s(t-\tau) = \int_0^T dt(r(t)-r_0)s(t-\tau) \tag{14}$$

Rearranging the integral on the left, we have

$$\int_0^\infty D(\tau')d\tau' \int_0^T dts(t-\tau')s(t-\tau) = \int_0^T dt(r(t)-r_0)s(t-\tau) \tag{15}$$

Now, we can define the autocorrelation function of the stimulus as

$$Q_{ss}(\tau) = \frac{1}{T}\int_0^T s(t)s(t+\tau)dt, \tag{16}$$

We can also define the correlation function between stimulus and response as

$$Q_{rs}(\tau) = \frac{1}{T}\int_0^T r(t)s(t+\tau)dt. \tag{17}$$

Then,

$$\int_0^\infty D(\tau')d\tau' Q_{ss}(\tau - \tau') = Q_{rs}(-\tau) \tag{18}$$

Now consider the simplest case in which the stimuli is white noise. In this case $Q_{ss}(\tau) = \sigma^2\delta(\tau)$. Then

$$\int_0^\infty D(\tau')\sigma^2\delta(\tau - \tau')d\tau' = D(\tau)\sigma^2 \tag{19}$$

As a result, we find the optimal kernel has an explicit expression

$$D(\tau) = \frac{Q_{rs}(-\tau)}{\sigma^2} \tag{20}$$

When the stimuli is not white noise, it is hard to calculate the kernel explicitly. However, if we ignore the causality, and make the assumption that response not only depends on the past but also the future of the stimulus, then things become much easier. performing Fourier transform, we found that in the frequency domain

$$\tilde{D}(\omega)\tilde{Q}_{ss}(\omega) = \tilde{Q}_{rs}(-\omega) \tag{21}$$

The Spike-Triggered Average

Now let me introduce the concept of spike triggered average, which is defined as

$$C(\tau) = \Big\langle \frac{1}{n}\sum_{i=1}^{n} s(t_i - \tau) \Big\rangle \tag{22}$$

Here t_i is the time when a spike occurs. The average $\langle ... \rangle$ is over different trials during which we are presenting exactly the same stimulus. The timing of the spike, however, can be variable. Below I would like to show that the spike triggered average has a direct connection the stimulus response correlation function $Q_{rs}(-\tau)$, provided that r is defined as the firing rate of a neuron. Given the density function

$$\rho(t) = \sum_i \delta(t - t_i) \tag{23}$$

It is easy to see that the

$$\sum_{i=1}^{n} s(t_i - \tau) = \int_0^T \rho(t)s(t - \tau)dt$$

Thus,

$$C(\tau) = \Big\langle \frac{1}{n} \int_0^T \rho(t) s(t-\tau) dt \Big\rangle$$

Formally, the firing rate of a neuron $r(t)$ is defined as

$$r(t) = \frac{1}{\Delta t} \int_t^{t+\Delta t} d\tau \langle \rho(\tau) \rangle, \tag{24}$$

where the average $\langle ... \rangle$ is also over different trials. Next, we want to make an important approximation by replacing the number of spikes n within a single trial with the mean number of spikes across all trials $\langle n \rangle$ during the period T. Then, because we are using exactly the same stimulus on each trial, the spike triggered average can be reduced to

$$\begin{aligned} C(\tau) &\approx \frac{1}{\langle n \rangle} \int_0^T \langle \rho(t) \rangle s(t-\tau) dt \\ &= \frac{1}{\langle n \rangle} \int_0^T r(t) s(t-\tau) dt \\ &= \frac{T}{\langle n \rangle} Q_{rs}(-\tau) \\ &= \frac{1}{\langle r \rangle} Q_{rs}(-\tau) \end{aligned} \tag{25}$$

Linear nonlinear model

The optimal kernel derived from a linear model has two problems. First, there is nothing to prevent the predicted response to become negative, and the predicted response does not saturate. As the magnitude of the response increases, the response would also increase without bound. If we use L to represent the linear term we have been discussing thus far:

$$L(t) = \int_0^\infty d\tau D(\tau) s(t-\tau) \tag{26}$$

The modification is to replace the linear prediction $r_{est}(t) = r_0 + L(t)$ with the generalization

$$r_{est}(t) = r_0 + F[L(t)] \tag{27}$$

Reverse-Correlation Methods and the Receptive Field of a Simple Cell

With the above preparations, we are now at a position to map the receptive field a sensory neuron in the visual cortex. The spike-triggered average for visual stimuli is defined, as the average over trials of stimuli evaluated at times

$t_i - \tau$, where t_i is the spike times. Because the light intensity of a visual image depends on locations as well as time, the spike-triggered average stimulus is a function of three variables,

$$C(x, y, \tau) \approx \frac{1}{\langle n \rangle} \Big\langle \sum_{i=1}^{n} s(x, y, t_i - \tau) \Big\rangle \tag{28}$$

The visual stimulus we are presenting is Gaussian white noise that is uncorrelated in both space and time,

$$\langle s(x, y, t) s(x', y', t') \rangle = \sigma_s^2 \delta(t - t') \delta(x - x') \delta(y - y') \tag{29}$$

We thus have

$$L(t) = \int_0^\infty d\tau \int dx dy D(x, y, \tau) s(x, y, t - \tau), \tag{30}$$

where the optimal kernel can be found through spike-triggered average

$$D(x, y, \tau) = \frac{\langle r \rangle C(x, y, \tau)}{\sigma_s^2} \tag{31}$$

The kernel $D(x, y, \tau)$ defines the space-time receptive field of a neuron. For some neurons, the kernel can be written as a product of two functions, one that describes the spatial and the other describes the temporal receptive field, namely

$$D(x, y, \tau) = D_s(x, y) D_t(\tau) \tag{32}$$

And the linear response of a neuron can also be viewed as a product of two parts

$$L(t) = L_s L_t \tag{33}$$

To derive and understand the spatial filter, let us consider an array of retinal ganglion cells with receptive field covering a small patch of the retina. We assume that the statistics of the visual inputs on that small patch are stationary and translational invariant, namely

$$\langle s(x) \rangle = \langle s(x + u) \rangle \tag{34}$$

$$\langle s(x) s(x + u) \rangle = f(u) \tag{35}$$

This means all locations and directions (as well as all times) in that patch are the same, which is equivalent for all the neurons in that patch to have the same receptive field or kernel. Under this assumption, the response of the neuron that localizes at (x_0, y_0) can be viewed as the convolution of the kernel with the visual stimulus, namely

$$L_s(x_0, y_0) = \int \int dx dy D_s(x - x_0, y - y_0) s(x, y) \tag{36}$$

$$L_t(t) = \int_0^\infty d\tau D_t(\tau) s(t - \tau) \tag{37}$$

The above network now has a very sexy name, CNN.