Problem Set 7, Computational Neuroscience

Due: 04/23/2004

7.1 Optimal Coding in Early Visual System

Suppose you receive a noisy signal \( x \) and you transmit it over a noisy linear channel:

\[
y = A(x + n) + u
\]

\( x \) is our original signal, \( n \) the noise in it, \( A \) a constant representing the linear channel, and \( u \) the noise in the channel. (Remember that we need a scale to make sense of things— the questions we are about to ask wouldn’t make sense without \( u \).)

The general kind of question we will be trying to answer in this section is, “How do we set \( A \) so that \( I(y; x) \) is maximized?” This is one popular meaning of “coding well”: finding the coding (here, \( A \)) that maximizes the information that the output carries about the input signal that we want to transmit in the presence of noise.

Given \( x, n, \) and \( u \) to be gaussians with mean zero and variances \( \sigma^2, \eta^2, \) and \( \nu^2, \) respectively, we have shown in the lecture that the mutual information

\[
I = \frac{1}{2} \log_e \frac{A^2(\sigma^2 + \eta^2) + \nu^2}{A^2\eta^2 + \nu^2}
\]

(2)

- Does \( A \) matter at all in the case \( \nu = 0? \) If \( \nu \neq 0, \) what’s the maximum amount of information \( y \) can carry about \( x? \) How large is \( A \) in that case? So, if somebody asked you to set \( A \) so as to maximize \( I(y; x) \), how would you set it?

Clearly, \( A \) sets the scale of the output signal \( y \) with respect to the noise in it, \( u \). We want to make \( A \) as large as possible, so that \( u \) is negligible (infinite signal-to-noise ratio). But when considering neural transmission, we can’t assume that neurons have an infinite signal-to-noise ratio. Rather, we want to minimize the the cost of information transmission, a given constraint within which we must find the best \( A \). We will formulate this constraint by demanding that \( C \equiv \langle y^2 \rangle \), the typical output signal magnitude or the capacity, is minimized. It is straightforward to show that

\[
C = A^2(\sigma^2 + \eta^2) + \nu^2
\]

(3)

Now we can construct an energy function

\[
E = C - \frac{k^2}{2} I
\]

(4)

in which \( k \) is a constant to set the relative weight for minimizing \( C \) and maximizing \( I \).

- By taking \( \frac{dE}{dA} = 0 \), show that

\[
|A| = \begin{cases} 
\sqrt{\frac{\nu^2}{\eta^2}} \sqrt{\frac{1 + \sqrt{1 + k^2 \eta^2 / \sigma^2 \nu^2}}{2(1 + \nu^2 / \sigma^2)}} - 1 & \text{if} \sqrt{\ldots} \text{is not real} \\
0 & \text{if} \sqrt{\ldots} \text{is not real}
\end{cases}
\]

(5)

- For \( \nu^2 = 1, \eta^2 = 0.001, \text{ and } k^2 = 10 \), plot \( |A| \) as a function of \( \sigma \), for \( \sigma \) from 0.1 to 100 in double-log plot. What is the approximate behaviour? Is it consistent with the decorrelation theory?
7.1.1 Multi-Dimensional Solution

Things get much more interesting if \( x \) is not a scalar, but an \( N \)-dimensional vector. As an example, \( x \) might be the vector representing light intensity at each of 1,000,000 photoreceptors. Some of the retina’s output cells (retinal ganglion cells) seem to pool inputs from several photoreceptors in a roughly linear fashion. In addition to that, the pooling could also include integration over time, say, at LGN level (a stage between retina and visual cortex) — thus spatiotemporal coding. This pooling would be represented by \( A \), our linear channel: the retinal ganglion cells’ (or LGN) outputs would be \( y \). A re-phrasing of the basic question, then, would be “given a probability distribution on \( x \), noise models \( n \) and \( u \), and a scaling constraint on \( y \), what is the pooling strategy that would result in the retinal ganglion cells’ (or LGN) output being most informative about the light signals \( x \) coming into the eye?”

As we will show later (in a two-dimensional case) that if we look at the Fourier transform of the coding, we get the same solution, except that \(|A|, |\sigma^2|\) are now functions of spatiotemporal frequencies. Specifically, \(|\sigma^2(f, w)|\) is now the power spectrum of natural time-varying images, or the Fourier transform of the covariance matrix of natural time-varying images. \(|A(f, w)|\) is a cell’s response to sine wave pattern of spatial frequency \( f \) and modulated at temporal frequency \( w \).

- Given \( k^2 > 2 \), show that when \( \sigma^2 \gg \eta^2 \),
  \[ |A(f, w)| \sim \frac{1}{\sqrt{\sigma^2(f, w)}} \]  
  \[(6)\]

You can either show this analytically or illustrate this numerically.

- For certain given spatial frequency \( f \), the power spectrum \( \sigma^2 \sim \frac{1}{w^2} \), substitute this into equation (??), show how the \(|A|\) depends on the temporal frequency. Without losing generality, we can assume \( \nu^2 = 1 \), i.e., the unit of everything. Now for \( k^2 = 10 \) and \( \eta^2 = 0.02 \), plot \(|A|\) as a function of the temporal frequency \( w \) for \( w = 0.2Hz \) to \( w = 20Hz \). Especially, Comments on the behavior at low and high temporal frequency regions, explain intuitively why?

- As we mentioned in the class, that we sometimes approximate the above solution by
  \[ |A(w)| \sim \frac{w}{(1 + w^2/\eta^2)^{3/2}} \]  
  \[(7)\]

for \( w_0 \sim 5Hz \). Show that with some choices of \( k^2 \) and \( \eta^2 \), the rigorous solution in the previous bullet is close to this, especially when signal-to-noise ratio is high. You should hand in a plot with two curves and indicate clearly the parameters used.