MIT Department of Brain and Cognitive Sciences
9.29J, Spring 2004 - Introduction to Computational Neuroscience
Instructor: Professor Sebastian Seung

Problem Set 1 (due Feb. 13)
Linear regression, convolution, and correlation

Feb. 5, 2003

This assignment is based on data from experiments described in R. Wessel, C. Koch, and F. Gabbiani, Coding of time-varying electric field amplitude modulations in a wave-type electric fish. *J Neurophysiol* 75:2280-93 (1996). The weakly electric fish *Eigenmannia* has a special organ that generates an oscillating electric field with a frequency of several hundred Hz. It also has an electrosensory organ, with which it is able to sense electric fields.

Download the data file: `fish.mat` (Available in the assignments section.)

Start up MATLAB and load the file by typing

```
>> load fish.mat
```

Typing `whos` will list the three vectors in the MATLAB workspace:

- `time` contains the sampling times in milliseconds.
- `rho` is a binary vector representing the spike train of an electrosensory neuron
- `stim` is the stimulus, a random amplitude modulation of an oscillating electric field that was applied to the fish.

Note: If your data link is too slow, you can use the abridged data file `fish10000.mat` for this assignment instead. (Available in the assignment section.)

1. For each of the following questions, give one or two MATLAB commands that answer the question, along with the answer itself. Try to make your MATLAB code short and elegant. No loops allowed!

   (a) How many spikes are in the whole experiment?

   (b) How long is the whole experiment, in seconds?

   (c) What is the firing rate in Hz, averaged over the whole experiment?

   (d) How many spikes are in the first half of the experiment?

   (e) What is the firing rate in Hz, averaged over the first half of the experiment?

   (f) What is the maximum value of the stimulus?

   (g) What is the minimum value of the stimulus?

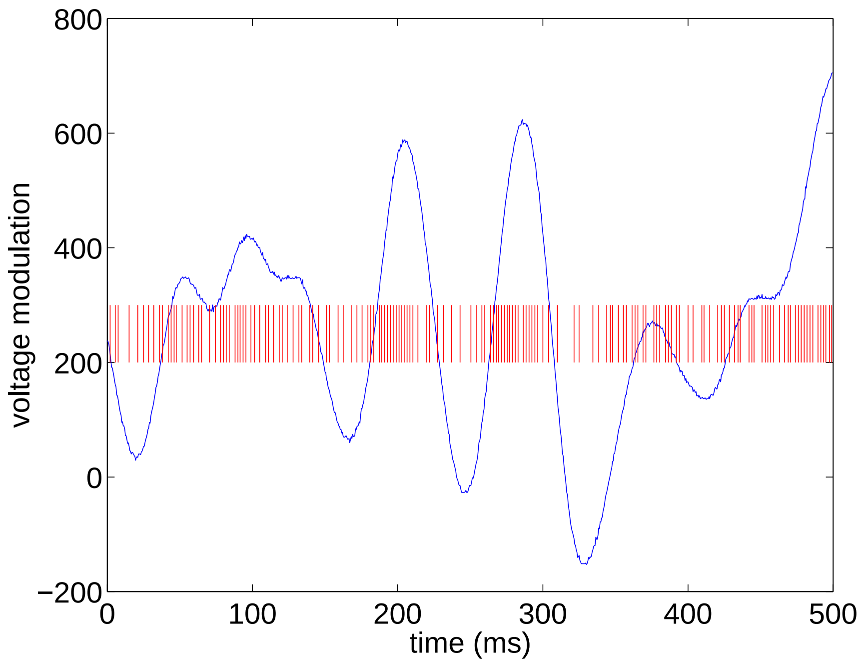
   (h) At what time (in milliseconds) did the hundredth spike occur?

   (i) What is the mean of the spike train?

   (j) What is the variance of the spike train?

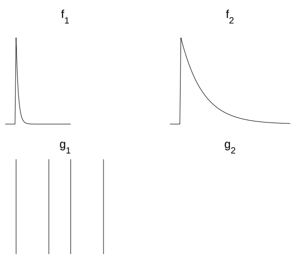
2. Note that the variance of the spike train was equal to the mean minus the square of the mean. Prove that this holds true for any binary signal.

3. Write a program that plots the first 1000 samples of the spike train and the stimulus like this:



Hint: use the `plot` command to graph the stimulus, and then use the `line` command to draw the spikes. Again, no loops allowed! Submit your code, along with the graph it produces.

4. Convolution with a spike train.



   (a) Sketch $f_1 * g_1$.
   (b) Sketch $f_2 * g_2$.

   ($*$ is the convolution operator; all x-axes are on the same scale.) The purpose of this question is to build your intuition about convolution. To that end, you may find the "Joy of Convolution" applets at
   `http://www.jhu.edu/~signals/convolve/` and
   `http://www.jhu.edu/~signals/discreteconv2/` useful to play with.

5. Estimating firing rate by convolution with a boxcar filter. In the following, use only the first 10000 samples of the data. Let's define the probability of firing at time t to be the number of spikes in a window centered at time t, divided by the length of the window. For example, in MATLAB this can be computed by convolving `rho` with `ones(101,1)/101`.

   Take the central section of the resulting vector (dropping the first 50 and last 50 elements) to obtain a vector of the same length as `rho`, and call this vector `prob`. (Even after omitting these elements, there are some suspect estimates of probability of firing at the edges, because the windows stick out past the edge of the spike train, but we'll ignore that problem.)

   With what number should `prob` be multiplied, to obtain an estimate of firing rate in Hz? Do the multiplication and call the result `rate`. Submit your code, along with a plot of firing rate and the stimulus versus time on the same graph.

6. Linear model of firing rate versus stimulus amplitude. Again, use only the first 10000 samples of the data in the following. Try approximating the firing rate as a linear function of the stimulus. Using the function `polyfit`, find the coefficients `a` and `b` such that `a*stim+b` best approximates `prob`. Submit your code, along with a plot of `a*stim+b` and `prob` versus `time` on the same graph. You should see a nice fit. Also plot `prob` versus `stim` (without connecting the data points) and `a*stim+b` versus `stim` on the same graph. This graph will show the straight line that you fit to the data points. Use the `corrcoef` command to compute the correlation coefficient.

   Now try the same thing, but calculate firing rate using windows of length 1001 samples and length 11 samples. Judging from the correlation coefficient, which of the three windows gives the best fit? Explain in words why.

7. Linear model of the spike train versus stimulus amplitude. Using the function `polyfit`, find the coefficients `a` and `b` such that `a*stim+b` best approximates `rho`. Use the `corrcoef` command to compute the correlation coefficient, which will indicate that the fit is bad. Clearly, no linear function of the stimulus can ever reproduce the spikiness of the spike train. Nevertheless, the coefficients a and b are very similar to the ones calculated by applying `polyfit` to `stim` and `prob`. Explain why.

8. Calculate the autocorrelation or autocovariance of the stimulus using the `xcorr` or `xcov` commands. Use a maximum lag of $M = 1000$. Plot your result. Roughly how wide is the central peak, in milliseconds?

   Now do the same thing for the spike train. Your result should look very different, with many peaks. Explain what property of the spike train causes this. Extra credit: by reading the Wessel et al. paper cited above and available on the class web site, can you figure out why the spike train has this property?

9. Calculate the cross-correlation or cross-covariance of the stimulus and the spike train. Use a maximum lag of $M = 1000$. Plot your result. This graphs tells you what the stimulus is like on average, near the time of a spike. On average, is the stimulus increasing or decreasing at the time of a spike? Explain your reasoning!