Spike Train Analysis:
BIO 450/IBS 534 HW #1

Due before 11:59pm, Tuesday, January 28th

For this homework set, we will use the function `calculateSpikeTriggered Average.m` to calculate the spike triggered average (STA) for data from H1 neurons in the fruit fly Lobula Plate (see the Group #2 exercises from the in-class exercises for more details, data is also provided there). For all questions, you can use `averagingWindow` = 100.

Part I:

1. Use the Matlab built-in function `randperm` to plot STA curves using N =100, 500, 1,000, 5,000, 10,000, and 25,000 spikes. Also calculate using all 53,601 spikes. Make sure to save the function's output, as you will need the saved STA curves for subsequent questions.

2. What happens to the curve as N becomes larger?

3. Now let's try to quantify how the average changes as N becomes large. If $s_k(t)$ is the STA curve using k spikes, calculate

$$d(k) = \sqrt{\sum_t \left[s_k(t) - s_{53,601}(t) \right]^2} \tag{1}$$

   for each of the values of k you previously calculated. Plot $d(k)$ vs. k on a log-log plot (`loglog` in Matlab).

4. If a function is of the form $y(x) = Ax^b$, then the slope on a log-log plot provides an estimate of b. What is (roughly) the slope of the line you see? What does this say about how the accuracy of the measured STA improves as one watches the data for increasingly long periods of time?

Part II:

A common model in computational neuroscience is called the Linear-Nonlinear Model, and it is a way of combining data from spiking neurons with patterns of stimulus to predict a probability of spiking. Just like it sounds, this model has two parts: a linear part to translate a stimulus into a "score" and a nonlinear part to translate that score into a probability of spiking.

1. The linear potion of these types of models builds-off of the STA (for all spikes) that you found in the previous part of the homework set. We will call this STA $s(t)$. To determine the score, or the similarity of match to the filter, we can compute the *convolution* between the data set, $x(t)$, and $s(t)$ using

$$\chi(t) = \sum_{t'} x(t-t')s(t')dt, \tag{2}$$

where $\chi(t)$ is the value of the convolution at time t and dt is the sampling time of the data (here, .001 seconds). Use `convolutionOutput = convolveDataWithSTA(data,STA,dt)` to calculate this sum for all time points. Plot the found convolution for the first second of the data set.

2. Use `plotSpikesOnData(H1_times,convolutionOutput,H1_spikes,startTime,endTime)` to plot the first second of the convolution superimposed with the locations of the spikes (the red vertical lines). Here, `startTime` is 0 and `endTime` is 1. What happens to the convolution function when a lot of spikes occur?

3. Use the `hist` function in Matlab to make a histogram of the values of $\chi(t)$ you previously calculated. Use 20 bins, and make sure to save the output to your workspace, as we will use this in a moment.

4. Now, we would like to map these convolution scores, $\chi(t)$ onto a probability of firing, $p(spike|\chi)$. In other words, we want to find a function, $f(\chi)$ that takes the convolution value to something between zero and one, but we will let the data tell us what that function is! To do this, we will calculate the conditional probability of observing a spike, given that a value of $\chi(t)$ is observed (i.e. the number of spikes when $\chi(t) \approx y$ divided by the total number of time points where $\chi(t) \approx y$). The loops needed to achieve this are a little tricky, so I have provided code to find this: `[binLocations,pSpike,numSpikes] = findSpikingProbabilitiesFromConvolution(convData,spikes,numBins)`, where `numBins` is the number of bins (use 20), `convData` is the convolution from before, and `spikes` are the spike locations. Display this plot.

5. What does the curve on the right ($p(spike|\chi)$ vs. χ) look like? Does this agree with your intuition from the plot in question 2 of part II?

6. Given a new stimulus presentation to the same neuron, describe how would you predict the location of spikes using the STA and the curve you derived in this homework set?