Lab 5: Neural Feature Extraction

Introduction

The first four labs introduced computational approaches to studying electrophysiology. In this lab, we will begin to use and examine tools that are used to interpret experimentally acquired data: a model for characterizing tissue response, rudimentary approaches for analyzing local field potentials, and a tool for dimensionality reduction that can be used in spike sorting.

Software

This lab must be completed using MATLAB. (Sorry, this is likely the end of the road for Python since we have a lot of tips available for MATLAB only.)

Part 1) Estimating effects of tissue response from impedance spectra

The insertion and presence of conventional recording electrodes in nervous tissue causes physical damage that leads to tissue responses, such as gliosis and edema. The resulting encapsulation of the electrode is known to significantly affect recording properties. Although it can be impractical to directly observe the tissue response during recording, the electrode boundary model can be used to estimate a number of parameters using readily obtainable data.

In Figure 3D of 'Complex impedance spectroscopy for monitoring tissue responses to inserted neural implants', by Justin C Williams, Joseph A Hippensteel, John Dilgen,William Shain, and Daryl R Kipke (2007), there are two impedance spectra of an electrode—one taken immediately after implantation and one seven days later. Using information from the graph, determine the following:

1. Estimate the values of $R_{en} + R_{ex}$ for both lines by extrapolating the first few points at the top of each plot. If $R_{en} + R_{ex} < 0$, treat it as 0.
2. Estimate α for each plot using the same few points at the top of each plot.
3. Calculate K for both lines. Use frequency values associated with the topmost point of each plot. (You can figure out the frequency from information in the paper.) Note that Z_{tot} (a complex value) can be read directly from the plots.
4. Using the parameters you estimated earlier, calculate the magnitude of the tissue related response and constant phase element at 1 kHz (that is, calculate $|Z_{total}|$ for $f = 1000\ Hz$) for both plots. Remember that $\omega = 2\pi f$.
5. What conclusions can you make about the electrode performance and the tissue response at the time of implantation and seven days later?

Part 2) Feature extraction from ECoG

Interpretation of field potentials is not straightforward. Analysis of even one channel from an ECoG array requires several layers of processing before arriving at meaningful features that can be used for decoding or interpretation. Here, we will examine one channel of data from a 128-channel ECoG array.

1. Load ecogdatasnippet.mat into MATLAB. The data is sampled at 1000 Hz.
2. Apply common average referencing to channel 29. Only use channels that show up in refChannels.
3. For channel 29, extract the three frequency bands analyzed in Pistohl et al (2011), 'Decoding natural grasp types from human ECoG'. MATLAB has a number of filter design tools that can all be used for this task. Use your favorite bandpass filter.
4. Calculate the power of these signals over time by squaring the voltage at each time point. Use MATLAB's smooth() function to smooth the features over a 100 ms window and plot over time. This will make task-related activity easier to see.

Part 3) Dimensionality reduction of spike recordings

Spike waveforms can be treated as high-dimensional data, with each sample along the voltage trace treated as a separate variable describing the waveform. (For example, a 3 ms snippet recorded at 10 kHz would have 30 samples and can be treated as 30-dimensional data.) However, many of the sampled points are often strongly correlated with each other and it is possible to significantly reduce the dimensionality of the data while losing very little information. This can be useful for visualization, spike sorting, and data compression.

One of the simplest and most widely utilized approaches to dimensionality reduction is principal components analysis (PCA). PCA identifies features of the data that capture the greatest variance, which can then be used to approximate the data, such that

$$\vec{x}_i \cong \sum_{j}^{k} w_{ij}\vec{u}_j$$

Where $\vec{x}_i$ is a normalized data point on the spike waveform, $\vec{u}_j$ is an eigenvector of data matrix X, w_{ij} is the weight of eigenvector j for sample i, and k is the number of eigenvectors used to reconstruct the data. You can think of $\vec{u}_j$ as a vector that captures a feature in the data, and w_{ij} as the strength of that feature in a given data point i. We call w_{ij} the jth *principal component* of data point i, and $\vec{u}_j$ the jth *principal eigenvector*. Note that for each eigenvector $\vec{u}_j$, there is a corresponding eigenvalue λ_j. Oftentimes, it is possible to very accurately reconstruct the data with $k \ll D$, where the D is the original dimensionality of the data.

1. Load spikes.mat into MATLAB. The file contains 41568 snippets of spikes, each 32 samples long. Plot a few to get a sense of what the data looks like.
2. Normalize the spike traces over the data set such that every point along the trace has a mean of 0 and standard deviation of 1. We do this so that each point along the trace has equal importance in the PCA calculation.
3. Use the MATLAB built-in pca() to calculate the data's principal components. Use MATLAB's help function and documentation of pca() to determine the which outputs correspond to w, u, and λ. You will need all three in the following steps.

Note: the MATLAB outputs will be in matrix form, while the equation above is written in vector form for ease of interpretation. (That is, MATLAB will provide W, U, and Λ.)

4. Each eigenvalue λ_j is proportional to the amount of variance captured by eigenvector $\vec{u}_j$. Determine the number of principal components, k, necessary to capture 90% of the variance in the data. (Hint: the cumsum() function might be of help here)
5. Pick a representative spike and plot the top k principal components of the data (corresponding to the k largest eigenvalues). Comment on which features are captured by each eigenvector.

Now that you've deconstructed the data into principal components, let's try to reconstruct the original spikes with lower-dimensional data.

6. Using the equation introduced earlier, approximate spike #5, with $k = 32$. That is, reconstruct the spike using every principal component. Un-normalize the reconstructed spike so that it looks like the original waveform. Compare this to the original waveform to make sure everything is working.
7. Now approximate spike #5 with only six principal components. Then four. Plot your results overlaid with the original waveform and comment on the resulting waveforms. Calculate the mean squared error of the approximations.

PCA can also be a useful tool for data clustering and visualization. Since a small number of principal components can capture a large portion of the data's variance, it is possible to represent differences in the data in a low-dimensional space.

8. Plot the first principal component (the one associated with the largest eigenvalue) of each data point against the second principal component of each. What do you see?

In the next lab, we will use clustering algorithms to sort this data.

Guidelines for Lab Report (on Labs 5 and 6 together)

Introduction: The introduction should be one paragraph long summarizing the motivation for developing the tools used in this lab and what they can be used for, along with a brief summary of everything you will show in this lab report.

Methods: From Lab 5, there should be methods paragraphs (and diagrams where necessary) on:

1. Assumptions of the models used and how the models were designed
2. Any signal conditioning methods used prior to analysis
3. The software tools used to implement signals analysis

Include the code as an Appendix to your report. Cite sources for any values used in your models.

Results: You should include the following in your Results:

1. The outcome of electrode interface calculations
2. Plot of the features (3 power bands) extracted from the ECoG signal
3. The accuracy and limits of dimensionality reduction by PCA (show plot of 9 PC's, plot of reconstruction with different numbers of PC's, and scatter plot of PC1 vs. PC2)

Include any others figures produced by MATLAB that could help explain and illustrate your findings.

Discussion: Should be ~2 paragraphs long describing what you could use these models or techniques for in the future.

This report will be combined with Lab 6, to create one cohesive report. The report (not including Appendix) should be no longer than 4 pages. Use 12 pt. font and 1.15-1.5 line spacing. If your text is over the 4-page limit with figures, you can move your figures to an appendix section that goes beyond the 4-page limit. Please upload your report to Canvas.