EE 16B   Designing Information Devices and Systems II
Fall 2015
Note 5

Lecture given by Babak Ayazifar (9/10)
Notes by: Ankit Mathur

Spectral Leakage Example

Compute the length-8 and length-6 DFT for the following $x(n)$ –

$$x(n) = e^{i\frac{\pi}{2}n}$$

Consider first the period of $x(n)$ here, which is clearly $P_x = \frac{2\pi}{\frac{\pi}{2}} = 4$. First, the length-8 DFT, which should work out nicely, since we know that $\omega_0 = \frac{2\pi}{8} = \frac{\pi}{4}$. Since, 8 is an integer multiple of P_x, there is no complication. The signal contains only a single frequency, and it will be reflected that way in the DFT. It will reflect itself in the second multiple of the harmonic. That means –

$$X_2 \neq 0$$
$$X_k = 0 \text{ when } k \neq 2$$

The more complicated case is the length-6 DFT. This is because $N = 6 \pmod 4 \neq 0$. What happens here is that you have a function that does not go through a complete set of cycles in the window of interest, which means that it is cut off. Since your calculations have an assumed periodicity, the reconstructed function from the DFT is not an accurate representation of the original function.
An example of this can be seen Let's walk through the example, noting that, in this example, $\omega_0 = \frac{2\pi}{6} = \frac{\pi}{3}$ –

$$X_k = \sum_{n=0}^{5} x(n)e^{-ik\frac{\pi}{3}}$$
$$= \sum_{n=0}^{5} e^{i\frac{\pi}{2}n}e^{-ik\frac{\pi}{3}}$$

A quick aside –
We need an equation that can help us take a shortcut for geometric series, and this is its derivation –
Assuming that $\alpha \neq 1$ (because this is a trivial case, in which case the answer is $B - A + 1$) –

$$S = \sum_{n=A}^{B} \alpha^n$$

$$S = \alpha^A + \alpha A + 1 + \cdots + \alpha^B$$

$$\alpha \cdot S = \alpha A + 1 + \alpha A + 2 + \cdots + \alpha^{B+1}$$

$$(\alpha - 1) \cdot S = \alpha^{B+1} - \alpha^A$$

$$S = \frac{\alpha^{B+1} - \alpha^A}{\alpha - 1}$$

$$\implies \sum_{n=A}^{B} \alpha^n = \boxed{\frac{\alpha^{B+1} - \alpha^A}{\alpha - 1}}$$

We manipulate the X_k formula to use the geometric series.

$$
\begin{aligned}
X_k &= \sum_{n=0}^{5} \left(e^{i(\frac{\pi}{2} - k \cdot \frac{\pi}{3})} \right)^n \\
&= \frac{e^{i(\frac{\pi}{2} - k \cdot \frac{\pi}{3}) \cdot 6} - 1}{e^{i(\frac{\pi}{2} - k \cdot \frac{\pi}{3})} - 1} \neq 0 \text{ for any } k
\end{aligned}
$$

However, if we use the synthesis equation $x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{ik\omega_0 n}$ to construct the original period, we will construct a discontinuous graph/incorrect graph because of our incorrect period.

Length-2 DFT Example

Consider a signal $\mathbf{x} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$. Now, consider a $\Psi_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ and $\Psi_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. Clearly, we can construct $\mathbf{x}$ as a linear combination of Φ_0 and Φ_1.

$$4\Psi_0 + 2 \cdot \Psi_1 = 4 \cdot \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 2 \end{bmatrix}$$

Now, we move to a slightly more difficult version of this problem. Say we have two mutually orthogonal vectors that span the 2-dimensional space –

$$\Psi_0(n) = \frac{1}{2} e^{i0\pi n} = \frac{1}{2}$$

$$\Psi_1(n) = \frac{1}{2} e^{i\pi n} = \frac{1}{2}(-1)^n$$

$$\Psi_k(n) = \frac{1}{N} e^{ik\omega_0 n}$$

In the lollipop plot form, $\Psi_0(n)$ would look like a graph that has just a constant lollipop with magnitude $\frac{1}{2}$ for every integer.

In the lollipop plot form, $\Psi_1(n)$ would look like a graph that has just a lollipop with magnitude $\frac{1}{2}$, alternating between negative for even integers and negative for odd integers.

Remember that, here, we are dealing with the set of 2-periodic functions, which is why we are considering only $\Psi_0(n)$ and $\Psi_1(n)$. We can then conclude –

$$x(n) = 6 \cdot \Psi_0(n) + 2 \cdot \Psi_1(n)$$

This example simply shows that the DFT is simply a change of basis into a coordinate system that tells us useful information about the frequency domain. In the coordinate system for 2 periodic functions (the one where we use these two Ψ functions), we get that change of basis, which tells us about the frequency content of $x(n)$.

BMI and PCA

We initially started with the following diagram detailing the different levels at which signals can be accumulated for patients regarding brain data.
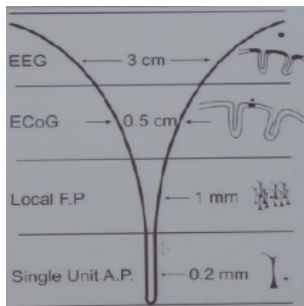

In this lecture, we discussed the data received from probes all the way at the neuron level. The basic case is when you only have 1 neuron whose data you are receiving, along with, obviously, some noise. However, the more difficult case is when –
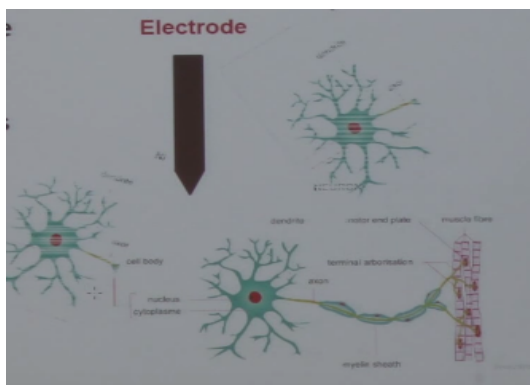
The challenge is to figure out how each of these will contribute to the overall signal that is being received, which is really a superposition of those. Unfortunately, these neurons will fire at around the same frequency

range. Techniques we used last semester involved using filters to cut down the frequencies we did not want (the ones that were too high or low).

This problem is formalized as something called "Spike Sorting" (`http://www.scholarpedia.org/article/Spike_sorting`), and the following diagram shows a workflow for that –
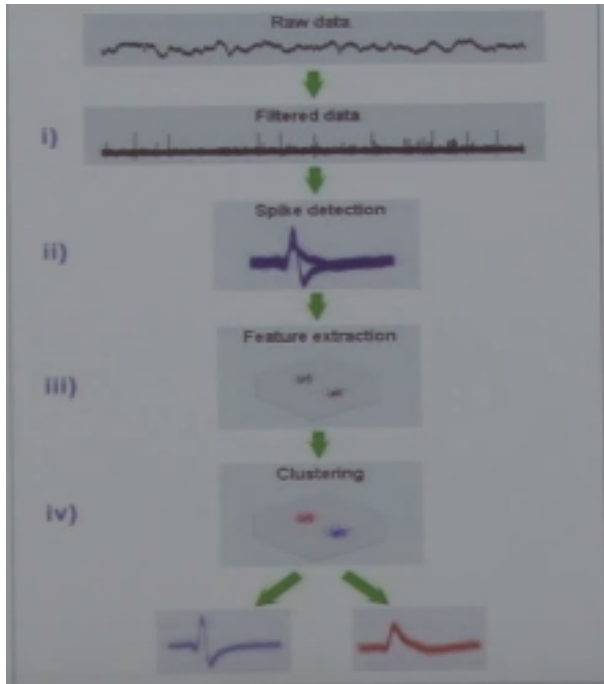
The goal is to map the fuzzy diagram in the third row to a set of features that we can extract and individualize to specific neurons. We want to capture the directions that allow us to separate the segments that we want. Usually, 2 or 3 principal directions will capture what you want, and the method that we want to use to do this kind of clustering is a singular value decomposition. Singular values, which are similar to eigenvalues, and their corresponding singular vectors are what help us to project our data onto those to generate information about where information in our graphs are coming from.

Modeling Neuron Data

Consider a sample with m spikes and n samples. We use the $m \times n$ matrix $A = \begin{bmatrix} \dots & a_1^T & \dots \\ \dots & a_2^T & \dots \\ \vdots & \vdots & \vdots \\ \dots & a_m^T & \dots \end{bmatrix}$ to represent this data.

The following equation allows us to average across each spike, whose data we collected at n time sample
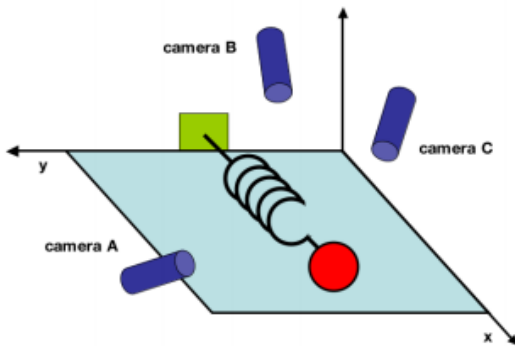
points.

$$\bar{a} = \frac{1}{m} \sum_{i=1}^{m} a_i^T$$

Then, using this, we can write a matrix $X_{m \times n}$, which has the mean removed from it.

$$X_{m \times n} = \begin{bmatrix} \cdots & a_1^T - avg_1^T & \cdots \\ \cdots & a_2^T - avg_2^T & \cdots \\ \vdots & \vdots & \vdots \\ \cdots & a_m^T - avg_m^T & \cdots \end{bmatrix} = A - I\bar{a}^T$$

We are going to consider the covariance matrix $C_x = \frac{1}{m-1} X^T X$, which will help us do some statistical analysis when we are looking for the singular values.

Consider an example where you have a spring moving about in a system –

We will be looking for the camera that tells us the most about the movement of the spring. As such, we will be looking for the camera that shows us the largest variation, which is where variance begins to play into this equation. The multidimensional version of that story is embedded in the covariance matrix.

We want to then determine a unit vector q such that $var(q) = q^T C_x q$ is maximized. In the next lecture, we will see how to determine that from the eigenvalue-eigenvector decomposition.