**ENGR 2013: Engineering Analysis and Applications**

**Laboratory No. 5: How Do You Find a Signal in Noise?**

**Goal:** The goal of this lab is to introduce you to a very important computation in signal processing – autocorrelation. This is used in a wide range of engineering and statistics applications. We are also going to build on what you learned in lab no.4.

**Preliminary Work:** In this lab we are going to explore the following equation:

$y\left(n\right)=g\_{1}\sin(\left(2πf\_{1}(n/f\_{s})\right))+g\_{2}\sin(\left(2πf\_{2}(n/f\_{s})\right)+g\_{3}w(n))$ .

Write a simple Python program that generates values of $y\left(n\right)$ for the range [0, 10,000]. Set $f\_{1}=100.0 Hz $and $f\_{2}=169.0 Hz$, and $f\_{s}=8000.0 Hz$. These are very common values in digital telephony. For $w\left(n\right)$, use the numpy function random to generate random values:

$w(n)= $np.random.normal(0,1,1) .

This generates what we refer to as zero mean Gaussian white noise. Your program should just print $y(n)$ to the terminal. To capture this data in a file, redirect the program output to a file using this command:

python gen\_signal.py > signal.dat

Your data should look something like this:

0.462001

2.116416

1.787632

…

In this lab, we are going to explore three fundamental mathematical/signal processing concepts. First, we introduce the autocorrelation function, which for this lab can be defined as:

$R\left[k\right]=(\frac{1}{N})\sum\_{n=0}^{N-1}x\left(n\right)x(n+k),   for k=[0,K]$ ,

where $N$ is the total number of samples used in the signal, which we refer to as the window size, and $K$ is the number of lags in your autocorrelation function. This function can be computing as a dot product between the vector $x$ and a version of $x$ that is shifted by $k$ samples.

Next, we introduce the concept of an [autocorrelation](https://www.geeksforgeeks.org/autocorrelation/) matrix (a special case of a [covariance matrix](https://www.geeksforgeeks.org/covariance-matrix/)):

$$R=\left[\begin{matrix}R[0]&R[1]&R[2]&…&R[K]\\R[1]&R[0]&R[1]&..&R[K-1]\\R[2]&R[1]&R[0]&…&R[K-2]\\…&…&…&…&…\\R[K]&R[K-1]&R[K-2]&…&R[0]\end{matrix}\right]$$

Note this is a symmetric matrix whose values are simply a function of the distance from the diagonal (or more precisely related to the difference of the row and column indices). Later in the course we will see this is a form of what is called a [Toeplitz matrix](https://www.geeksforgeeks.org/find-if-given-matrix-is-toeplitz-or-not/) and because of that, it has special properties.

Finally, we are going to explore the condition number of a matrix. You can learn about this [here](https://www.geeksforgeeks.org/compute-the-condition-number-of-a-given-matrix-using-numpy/). The [condition number](https://en.wikipedia.org/wiki/Condition_number) is a simple calculation in Python that gives you insight into whether a matrix is invertible (low is good). A good discussion of this can be found [here](https://math.stackexchange.com/questions/261295/to-invert-a-matrix-condition-number-should-be-less-than-what#:~:text=Mathematically%2C%20if%20the%20condition%20number,can%20be%20just%20as%20bad.).

**Tasks:**

1. Set $g\_{1}$ = 100, $g\_{2}$= 0, and $g\_{3}$=1. This generates a single sinewave a $100$ Hz which is sampled at $8,000$ Hz. Compute and plot the autocorrelation function for $k=[0, 100]$. Explain what integer value, or lag, this function peaks at and how that relates to the frequency of the sinewave and the sampling frequency.
2. Using the values in (1), plot the maximum value of the autocorrelation function as a function of $g\_{3}$ for $g\_{3}=1, 2, 5, 10, 20, 50, 100$. Plot this on a log-linear scale ($max\{R[K]\}$ vs. $log(g\_{3})$).
3. Set $g\_{1}$ = 100, $g\_{2}$= 100, while keeping $g\_{3}$=1. Plot the autocorrelation function. Explain what you observe, namely how the peaks relate to the frequencies of the sinewaves.
4. Construct the autocorrelation matrix, $R$, for $g\_{1}=100$, $g\_{2}=100$, while keeping $g\_{3}=1$. Create a table that shows the rank of the matrix as a function of $K$ for the range $K=4, 8, 12, 16, 32, 64$. What do you observe? Repeat this for $[g\_{1}=1, g\_{2}=1]$, and $[g\_{1}=0.1,g\_{2}=0.1]$. Does anything change?
5. Construct a table that shows the condition value of the autocorrelation matrix for $K=4, 8, 12, 16, 32, 64$ for $g\_{1}$= $g\_{2}=0.01, 0.1, 1, 10, 100$ (keep $g\_{1}=g\_{2}$ while you iterate over the values $0.01, 0.1, 1, 10, 100$ ). Explain what you observe.

For these tasks, place your code in a subdirectory t01, t02, …, t05. Also place a pdf file containing your plots (t01.pdf, t02,pdf, …) in the corresponding directory.

The TAs will ask you to explain your results, and you will be judged on your ability to explain the reasons why you are observing the outputs you generated.

**Summary:**

The sinewave signal you created has a dimensionality of $4$ (each sinewave requires two parameters). We also often say it has $4$ degrees of freedom. When you add noise, it makes the signal to appear more complex than it is. The degree to which the noise confuses the process depends on the signal to noise ratio – the power of the sinewaves divided by the power of the noise. In your statistics class, you will learn how autocorrelation and covariance functions can be used to remove noise from a signal and allow you to uncover its hidden structure.

Would this be useful to you in say, a place like a casino? ☺