**ENGR 2013: Engineering Analysis and Applications**

**Laboratory No. 12: How Can We Optimally Compress Data and
Discover Underlying Relationships?**

**Goal:** The goal of this lab is to introduce you to the Gram-Schmidt Orthogonalization algorithm and discuss some of the implications of this algorithm.

**Preliminary Work:**

Let’s start with the notion of a span: The span of a set of vectors, also called linear span, is the linear space formed by all the vectors that can be written as linear combinations of the vectors belonging to the given set. For this lab, simply think of this as a set of vectors for which you want to find an underlying set of orthonormal vectors that can be used to represent the original vectors using linear combinations. “Orthonormal” simply means orthogonal and they have a norm, or magnitude, of $1$.

The Gram-Schmidt Orthogonalization algorithm is an algorithm to find this underlying set of orthonormal vectors (see Theorem 10.2.5 in the textbook, or this [wiki page](https://en.wikipedia.org/wiki/Gram%E2%80%93Schmidt_process)).

Next, review these nice web pages demonstrating the concept: [Mayorov: Implementing and visualizing Gram-Schmidt orthogonalization](https://zerobone.net/blog/cs/gram-schmidt-orthogonalization/) and [Harvey Mudd Calculus Online: Gram Schmidt Method](https://math.hmc.edu/calculus/hmc-mathematics-calculus-online-tutorials/linear-algebra/gram-schmidt-method/).

Finally, review the Python code here: */data/courses/engr\_2011/current/labs/lab\_12/picone\_joseph*.

**Tasks:**

1. Replicate the example here:

<https://byjus.com/maths/gram-schmidt-orthonormalization-process/>

1. Check whether the input vectors and output vectors are orthogonal.
2. Using Python, find an equation that describes each input vector as a linear combination of the vectors in the orthonormal set.

**Now the fun stuff:**



Figure 1. A simple binary signaling example

1. Using Python, find two orthogonal signals that can represent the signals shown in Figure 1. an equation that describes each input vector as a linear combination of the vectors in the orthonormal set. How could this be applied to digital communications?
2. Following (4), construct 8 sequences that are 4 bits long (e.g., “1100”). Pick these so that there is significant correlation between them (we have studied correlation in previous labs). Find a set of orthonormal vectors that can represent your 8-bit sequences. Represent each input sequence in terms of your orthonormal vectors. How could you use this to send this information over a communications link using less than the 8 bits required for the original signal.
3. Pick $10$ publicly traded stocks. Try to pick $8$ from the same market segment (e.g., automotive stocks) and $2$ from an unrelated sector (e.g., Amazon and Walmart). Represent these stocks by their stock price using the price on the first day of month for the last 8 months. Find a set of orthonormal vectors that represent your input vectors and write the input vectors in terms of the orthonormal vectors. Did you learn anything useful from this analysis. Explain?

**Summary:**

The Gram-Schmidt algorithm is used to find a set of orthonormal vectors for a set of data. There are many uses for these vectors. For example, we can use them to compress data by representing each data vector in terms of its coefficients used to represent the vector by linear combinations of these orthonormal vectors. This type of analysis is a large part of fields such as coding theory and digital communications. We can also examine these orthonormal vectors to discover hidden relationships in your data. This kind of analysis is a large part of the field of machine learning.