This is meant as brief review of Gram-Schmidt Orthogonalization, meant primarily for students who are not EEs and have not taken Linear Algebra. This is definitely not a complete discussion of the topic, I'm not going to prove any theorems here; those who wish to read more are directed to either Prof. Fontaine's book "Signal Processing: A Unified Approach" and "Linear Algebra" by Friedberg.

We start with the definitions of linear combination, span and linear dependence:

Definition 1. A ***linear combination*** of a set of vectors $\{v_1, v_2, ..., v_n\}$ is an expression constructed from finitely many vector sums and scalar multiplies within the set, i.e, $a_1 v_1 + a_2 v_2 + ... + a_n v_n$ where the $a_i's$, $1 \leq i \leq n$, are real or complex scalars.

Definition 2. Let S be a set of vectors. We define ***span(S)*** to be the set of linear combinations of vectors in S. $span(\emptyset) = \{0\}$ by definition.

Definition 3. A set of vectors is said to be ***linearly dependent*** if any of the vectors of the set can be expressed as a linear combination of the other vectors. If this is not the case then the set of vectors is said to be ***linearly independent***.

So, for example, in R^2, the vectors $\{(1, 0), (0, 1), (1, 1)\}$ are linearly dependent since $(1, 1) = 1(1, 0) + 1(0, 1)$. In fact, if the number of vectors in the set is greater than the dimension of the "space" we are working in, the set is automatically linearly dependent. Here we can say the dimension of R^2 is 2 since there are two coordinates, though the precise definition of "dimension" is more formal. On the other hand, the set $\{(1, 0), (0, 1)\}$ is linearly independent since it is impossible to express either one as a scalar multiple of the other. In fact, that set is an *orthogonal set* of vectors, more on that later. One remark is that if a square matrix's columns (or rows, turns out this is if and only if) are linearly independent, then the matrix is invertible.

Definition 4. A set of vectors $\{v_1, v_2, ..., v_n\}$ are said to be ***orthogonal*** if $\langle v_i, v_j \rangle = 0$ whenever $i \neq j$. $\langle,\rangle$ represents the inner product (or dot product in R^n and C^n). The vectors are ***orthonormal*** if they are orthogonal and for $1 \leq i \leq n$, $\|v_i\| = 1$, where $\|\cdot\|$ represents the norm.

It can be proved that if a set of vectors are orthogonal, then they are linearly independent too.

Theorem 5. *Let $S = \{v_1, v_2, ..., v_n\}$ be a set of orthonormal vectors and let x be any vector. Then there exists a unique vector $u \in span(S)$ which is "closest" to x, i.e $u = \min_{u' \in S} \|u' - x\|$. This vector is called the **orthogonal projection of x on S**. Moreover,*

$$u = \sum_{j=1}^{n} \langle x, v_j \rangle v_j.$$

This is an extremely important theorem as it says that we can always approximate a vector with a set of orthogonal vectors and that approximation will be unique and can be easily identified through the inner product.

Finally, we discuss the Gram-Schmidt Orthogonalization process. What this process does for us is, given a set of *linearly independent* vectors, return an equal sized set of orthogonal vectors which span the same set of vectors.

Theorem 6. *Given $S = \{v_1, v_2, ..., v_n\}$, a set of linearly independent vectors, the set $S' = \{w_1, w_2, ..., w_n\}$ is an orthogonal set of vectors, where the vectors in S' were constructed in the following process:*

$$w_1 = v_1$$

$$w_i = v_i - \sum_{j=1}^{i-1} \frac{\langle v_i, w_j \rangle}{\|w_j\|^2} w_j, \ 2 \le i \le n$$

Additionally, $span(S) = span(S')$.

I want to point out an intuitive reasoning to why this works. First, note that if I was to start a set of orthogonal vectors, I could take whichever vector of S I wanted first. So the first step of this process is to just add the first vector of S to S'.

Next, take the second vector in S. We wish to project it onto the space orthogonal to the first vector and then add that vector to our set S'. Well let's think about the opposite task: projecting the vector onto the first vector. By Theorem 5 we know how to do that and that is the summation of Theorem 6. But what we want to do is get rid of that part of the second vector and only include the part of the second vector orthogonal to the first vector. So what we do is subtract that sum off of the vector and take the resulting vector in our set S'. Similarly, the rest of the vectors are handled the same way.

Finally, we can create an orthonormal set from S' by normalizing all the vectors.

Note the requirement of linear independence. If the set is not linearly independent, this process will break down as there will come a point where one vector can be "perfectly" be expressed as the orthogonal projection on the other vectors. Thus you will be left with the zero vector which pretty much causes the process to break down.