So let's now turn to how to generate an orthonormal basis in a vector space.

While normalizing things may seem easy enough, i.e. \( x \rightarrow \frac{x}{\|x\|} \), finding an orthogonal set can be tricky. Luckily, we have a process due to Gram-Schmidt.

While the book takes you through it in general, we will apply it to \( \mathbb{R}^3 \).

So we start with a non-orthonormal basis. In this case \( X = \{ 1, t, t^2 \} \), i.e. anything can be written as a linear combination and these are linearly independent.

This means that an orthonormal basis must be written in terms of linear combinations of these. The orthonormal set we will call \( \{ \gamma_1, \gamma_2, \gamma_3 \} \).

Start with, e.g., \( \gamma_2 = \frac{x_2}{\|x_2\|} = 1 \), then \( \gamma_1 = \frac{x_1 - \alpha_2 \gamma_2}{\|x_1 - \alpha_2 \gamma_2\|} \) where we need to find \( \alpha_2 \).

But we need \( \langle \gamma_2, \gamma_1 \rangle = 0 \) \( \Leftrightarrow \frac{1}{\|x_2\|} \int_{-\frac{1}{2}}^{\frac{1}{2}} (1, 1, t) \cdot (1, t, \frac{t}{2}) dt = 0 \) \( \Leftrightarrow \int_{-\frac{1}{2}}^{\frac{1}{2}} \left[ \left( 1, 1, 1 \right) \cdot \left( 1, t, \frac{t}{2} \right) \right] dt = 0 \) \( \Leftrightarrow \int_{-\frac{1}{2}}^{\frac{1}{2}} \left( \frac{1}{2} - \frac{t^2}{4} \right) dt = 0 \).

\( \alpha_2 = \frac{1}{2} \).

Then \( \gamma_1 = \left( \frac{t + \frac{1}{2}}{\|x_1 - \alpha_2 \gamma_2\|} \right) \), but wait that can't be!

The problem is \( \langle 1, x_1 \rangle = \int_{-\frac{1}{2}}^{\frac{1}{2}} t dt = \frac{1}{2} \neq 0 \), while \( \langle 1, x_2 \rangle = \int_{-\frac{1}{2}}^{\frac{1}{2}} t dt = 0 \).

So this leads to \( \frac{1}{2} - \alpha_2 = 0 \) \( \Leftrightarrow \alpha_2 = \frac{1}{2} \). So \( \gamma_1 = \left( \frac{t - \frac{1}{2}}{\|x_1 - \alpha_2 \gamma_2\|} \right) \).

So we have \( \gamma_1 = \left( \frac{t - \frac{1}{2}}{\|x_1 - \alpha_2 \gamma_2\|} \right) \).

To finish up:

\( \gamma_2 = \frac{x_2 - \langle x_2, \gamma_1 \rangle \gamma_1}{\|x_2 - \langle x_2, \gamma_1 \rangle \gamma_1\|} \)

and \( \gamma_2 \cdot \gamma_1 = 0 \) \( \Leftrightarrow \frac{1}{\|x_2\|} \int_{-\frac{1}{2}}^{\frac{1}{2}} (1, t, t^2) \cdot (1, t, \frac{t}{2}) dt = 0 \)

\( \gamma_2 = \left( \frac{t^2 - \frac{1}{2}}{\|x_2\|} \right) \).

Notice the pattern: \( \alpha_0 = \langle x_0, \gamma_1 \rangle, \alpha_1 = \langle x_1, \gamma_2 \rangle \Rightarrow \alpha_2 = \langle x_2, \gamma_2 \rangle \) for \( \gamma_2 \).

So \( \gamma = \left\{ \gamma_0, \gamma_1, \gamma_2 \right\} \) is an orthonormal basis.

Then are obviously others which differ by choosing a different vector for the starting point, or using a different linearly independent set.
So in summary, Gram-Schmidt starts us a non-orthonormal basis (a complete set of linearly independent vectors) \( X = \{x_1, x_2, \ldots, x_n\} \) and forms an orthonormal basis \( Y = \{y_1, y_2, \ldots, y_n\} \) by selecting one of the \( x_i \), say \( x_1 \), and forming \( y_1 = \frac{x_1}{\|x_1\|} \)

then \( y_{n+1} = \frac{x_{n+1} - \sum_{i=1}^{n} \langle x_{n+1}, y_i \rangle y_i}{\|x_{n+1} - \sum_{i=1}^{n} \langle x_{n+1}, y_i \rangle y_i\|} \)

So \( y_i \cdot y_j = \delta_{ij} \).

In words, we pick an vector to start with. Then with our second choice we subtract out of it any components it has along the first choice, then normalize. Then for our third we remove any components along the first two, then normalize. And so on...

Another advantage of an orthonormal basis is that it gives us a means of figuring out the matrix elements of a linear transformation w.r.t. the basis.

This works because \( A x_j = \sum_k a_{kj} x_k \) and \( \langle x_i, \sum_k a_{kj} x_k \rangle = \sum_k \langle x_i, a_{kj} x_k \rangle = \sum_k a_{kj} \langle x_i, x_k \rangle = \sum_k a_{kj} \delta_{ik} = \sum_k a_{ki} \delta_{ij} = a_{ij} \).

Let's see this in action. Going back to \( D \) on \( P_1 \), we have worked w.r.t. the non-orthonormal basis \( \{x_1, x_2\} \) where \( D = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \), so let's check \( \langle x_i, D x_j \rangle = a_{ij} \).

\[
\begin{array}{l}
(x_1, D x_1) = \langle x_1, 0 \rangle = 0 = d_{11} \\
(x_1, D x_2) = \langle x_1, D x_1 \rangle = 1 = d_{12} \\
(x_2, D x_1) = \langle x_2, 0 \rangle = 0 = d_{21} \\
(x_2, D x_2) = \langle x_2, D x_2 \rangle = \frac{3}{2} = d_{22} \\
\end{array}
\]

If instead we worked w.r.t. an orthonormal basis (from Gram-Schmidt): \( \{x_1, \sqrt{2} \mathbf{e}_2 \} \) then

\[
\begin{pmatrix} x_1, D x_1 \end{pmatrix} = \begin{pmatrix} 0, 1 \end{pmatrix} = O
\]

\[
\begin{pmatrix} x_1, D x_2 \end{pmatrix} = \begin{pmatrix} \sqrt{2}, \sqrt{2} \end{pmatrix} = \begin{pmatrix} 0 & 2 \sqrt{2} \\ \sqrt{2} & 0 \end{pmatrix}
\]

\[
\begin{pmatrix} x_2, D x_1 \end{pmatrix} = \begin{pmatrix} \sqrt{2} \mathbf{e}_2, \sqrt{2} \mathbf{e}_2 \end{pmatrix} = \begin{pmatrix} 0 & 2 \sqrt{2} \\ \sqrt{2} & 0 \end{pmatrix}
\]

\[
\begin{pmatrix} x_2, D x_2 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ \sqrt{2} & 0 \end{pmatrix}
\]

Ordering the result: \( D(x_1, x_2) = \begin{pmatrix} 0 & 2 \sqrt{2} \\ \sqrt{2} & 0 \end{pmatrix} \). So as a matrix

\[
D = \begin{pmatrix} 0 & 2 \sqrt{2} \\ \sqrt{2} & 0 \end{pmatrix}
\]

which is what the \( D \) from above does!
Let's continue our discussion of how an inner product (and orthogonality and normalization that stem from it) impacts our study of linear operators acting on a vector space.

Let's start by stating what may be obvious, but will nonetheless be useful later on:

A linear transformation $A$ on an inner-product space is the zero transformation if and only if $\langle x, Ay \rangle = 0$ for all $x$ and $y$. 
Now we know that the determinant of the matrix form of any linear operator tells us quite a bit (is it invertible, the eigenvalues, etc.). So we ask, “What does adding orthogonality and normalization do to finding the determinant?”

Consider a matrix $A = \left( \begin{array}{cc} a_{11} & a_{12} \\ a_{21} & a_{22} \end{array} \right)$ where each row is a vector orthogonal to the rest of the rows. So we might call it $A = \left( \begin{array}{c} a_1 \\ a_2 \end{array} \right)$ where each $a_i$ has $n$ components (or according to book $A = \{a_1, a_2, \ldots \}$).

Clearly $(AA^\top)_{i,j} = \Sigma a_{i,k} a_{k,j} = \Sigma a_{i,k} a_{j,k} = a_{i,j} = (a_i, a_j) = \|a_i\| \|a_j\|$

This just takes each row and forms the inner product in all the others.

So $AA^\top = \left( \begin{array}{ccc} ||a_1||^2 & \langle a_1, a_2 \rangle & \cdots \\ \langle a_2, a_1 \rangle & ||a_2||^2 & \cdots \\ \vdots & \vdots & \ddots \end{array} \right) \Rightarrow \det(AA^\top) = ||a_1||^2 ||a_2||^2 \cdots ||a_n||^2$

Recall that: $\det A = (\det A^\top)^* = \det A^* = \det A^\top$ and $\det(AB) = \det A \det B$

Together these give: $\det(AA^\top) = \det A \det A^\top = \det A^2 \Rightarrow \det A = \sqrt{\det(AA^\top)} = ||a_1|| ||a_2|| \cdots$

Now $A$ in the above was special. Suppose we start with an arbitrary matrix $B = \left( \begin{array}{c} b_1 \\ b_2 \end{array} \right)$.

This time the rows of the matrix need not be orthogonal.

Now let’s Gram-Schmidt the shit out of it: $B = \left( \begin{array}{c} b_1 \\ b_2 \end{array} \right) \Rightarrow B' = \left( \begin{array}{c} b_1/\|b_1\| \\ b_2 - \frac{\langle b_2, b_1 \rangle}{\|b_1\|^2} b_1/\|b_1\| \end{array} \right)$

Now let’s form $C = \left( \begin{array}{c} b_1/\|b_1\| \\ b_2 - \frac{\langle b_2, b_1 \rangle}{\|b_1\|^2} b_1/\|b_1\| \end{array} \right)$

This is special, since each row is formed by adding multiples of other rows to it, e.g. $b_2 \rightarrow b_2 - \frac{\langle b_2, b_1 \rangle}{\|b_1\|^2} b_1$

But this means $\det B' = \det C$!

But since $C$ is comprised of orthogonal vectors $\{b_1, b_2, \ldots, b_n\}$ times scalar multiples, $\left( b_1, b_2, \ldots, b_n \right) = \left( \begin{array}{c} b_1/\|b_1\| \\ b_2 - \frac{\langle b_2, b_1 \rangle}{\|b_1\|^2} b_1/\|b_1\| \end{array} \right)$... Then $\det C = 1/\|b_1\| \|b_2 - \frac{\langle b_2, b_1 \rangle}{\|b_1\|^2} b_1\| \|b_3\| \cdots$

where each $|b_i| \leq \|b_i\|$, therefore we get $\text{Hadamard's inequality: } \det B \leq \|b_1\| \|b_2\| \|b_3\| \cdots$