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. $x \to \frac{x}{\|x\|}$, 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 P_2.

So we start with a non-orthonormal basis. In this case $X = \overbrace{\{1, t, t^2\}}^{X_i}$, i.e. anything can be written as a linear combination and these are linearly independent. This means that an orthonormal basis must be writable in terms of linear combinations of these. The orthonormal set we will call Y.

Start w/, e.g., $y_0 = \frac{x_0}{\|x_0\|} = 1$, then $y_1 = \frac{\overbrace{(x_1 - \alpha_0 y_0)}^{\text{removes from } x_1 \text{ any components along } y_0}}{\|x_1 - \alpha_0 y_0\|}$ where we need to find α_0.

But we need $(y_0, y_1) = 0 \Rightarrow \frac{1}{\|x_1 - \alpha_0 y_0\|}\left[\underbrace{(1, x_1)}_{x_1} - \underbrace{\alpha_0}_{\alpha_0} \underbrace{(1, y_0)}_{1}\right] = 0 \Rightarrow x_1 = \alpha_0$

Then $y_1 = \frac{(t-t)}{\|x_1 - \alpha_0 y_0\|} = 0$, but wait that can't be!

The problem is $(1, x_1) = \int_0^1 t\,dt = \frac{1}{2} \neq x_1$ while $(1, y_0) = \int_0^1 1\,dt = 1$

So this leads to $\frac{1}{2} - \alpha_0 = 0 \Rightarrow \alpha_0 = \frac{1}{2}$ so $y_1 = \frac{(t - \frac{1}{2})}{\underbrace{\|t - \frac{1}{2}\|}_{\left(\int_0^1 (t - \frac{1}{2})^2 dt\right)^{1/2}}} = \sqrt{12}\,(t - \frac{1}{2})$

So we have $Y = \{1, \sqrt{12}(t - \frac{1}{2}), ?\}$

To finish up:

$$y_2 = \frac{x_2 - (\alpha_0 y_0 + \alpha_1 y_1)}{\|x_2 - (\alpha_0 y_0 + \alpha_1 y_1)\|}$$

s.t. $(y_0, y_2) = 0 \Rightarrow \frac{1}{\|\ \|}\left[\underbrace{(1, t^2)}_{\frac{1}{3}} - \alpha_0 \underbrace{(1, 1)}_{1} - \alpha_1 \sqrt{12}\underbrace{(1, t - \frac{1}{2})}_{0}\right] = \frac{1}{\|\ \|}\left(\frac{1}{3} - \alpha_0\right) = 0$

and $\alpha_0 = \frac{1}{3}$

$(y_1, y_2) = 0 \Rightarrow \frac{1}{\|\ \|}\left[\underbrace{(\sqrt{12}(t - \frac{1}{2}), t^2)}_{\sqrt{12}/12} - \frac{1}{3}\underbrace{(\sqrt{12}(t - \frac{1}{2}), 1)}_{0} - \alpha_1 12 \underbrace{(t - \frac{1}{2}, t - \frac{1}{2})}_{\frac{1}{12}}\right] = 0$

$-\alpha_1 12 \quad \Rightarrow \alpha_1 = \frac{1}{\sqrt{12}}$

$y_2 = \sqrt{180}\,(t^2 - t + \frac{1}{6})$ Notice the pattern: $\alpha_0 = (y_0, x_2)$, $\alpha_1 = (y_1, x_2) \Rightarrow \alpha_\ell = (y_\ell, x_n)$ for y_n

So $Y = \{1, \sqrt{12}\,(t - \frac{1}{2}), \sqrt{180}\,(t^2 - t + \frac{1}{6})\}$ is one orthonormal basis.

There 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 w/ a non-orthonormal basis (a complete set of linearly independent vectors) $X = \{x_1, x_2, \dots, x_n\}$ and forms an orthonormal basis $Y = \{y_1, y_2, \dots, y_n\}$ by selecting one of the X's, say x_1 and forming $y_1 = \frac{x_1}{\|x_1\|}$

then $y_{m+1} = \dfrac{x_{m+1} - [\langle y_1, x_{m+1}\rangle y_1 + \langle y_2, x_{m+1}\rangle y_2 + \dots + \langle y_m, x_{m+1}\rangle y_m]}{\|x_{m+1} - [\langle y_1, x_{m+1}\rangle y_1 + \langle y_2, x_{m+1}\rangle y_2 + \dots + \langle y_m, x_{m+1}\rangle y_m]\|}$

So $y_1 = \frac{x_1}{\|x_1\|}$, $y_2 = \dfrac{x_2 - \langle y_1, x_2\rangle y_1}{\|x_2 - \langle y_1, x_2\rangle y_1\|}$, $y_3 = \dfrac{x_3 - \langle y_1, x_3\rangle y_1 - \langle y_2, x_3\rangle y_2}{\|x_3 - \langle y_1, x_3\rangle y_1 - \langle y_2, x_3\rangle y_2\|}$, etc.

s.t. $\langle y_i, y_j\rangle = \delta_{ij}$

In words, we pick one 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 orthormal basis is that it gives us a means of figuring out the matrix elements of a linear transformation w.r.t. the basis.

This arises because $Ax_j = \sum_k a_{kj} x_k \Rightarrow \underline{\langle x_i, Ax_j\rangle} = \langle x_i, \sum_k a_{kj} x_k\rangle = \sum_k a_{kj} \langle x_i, x_k\rangle$
$= \sum_k a_{kj} \delta_{ik} = \underline{a_{ij}}$

Let's see this in action. Going back to D on P_1, we have worked w/ the a non-orthonormal basis $\underbrace{\{\overset{x_1}{1}, \overset{x_2}{t}\} \Rightarrow D = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}}$ (we already found the matrix form of D, so let's check $\langle x_i, Dx_j\rangle = d_{ij}$

$\langle x_1, Dx_1\rangle = \langle 1, 0\rangle = 0 = d_{11}$

$\langle x_1, Dx_2\rangle = \langle 1, 1\rangle = 1 = d_{12}$

$\langle x_2, Dx_1\rangle = \langle t, 0\rangle = 0 = d_1$

$\langle x_2, Dx_2\rangle = \langle t, 1\rangle = \frac{1}{2} \neq d_{22}$

If instead we worked w/ an orthonormal basis (from Gram-Schmidt): $\{\overset{x_1}{1}, \overset{x_2}{\sqrt{3}(2t-1)}\}$ then

$$\left.\begin{aligned}
&\langle x_1, Dx_1\rangle = \langle 1, 0\rangle = 0 \\
&\langle x_1, Dx_2\rangle = \langle 1, 2\sqrt{3}\rangle = \int_0^1 2\sqrt{3}\, dt = 2\sqrt{3} \\
&\langle x_2, Dx_1\rangle = \langle \sqrt{3}(2t-1), 0\rangle = 0 \\
&\langle x_2, Dx_2\rangle = \langle \sqrt{3}(2t-1), 2\sqrt{3}\rangle = \int_0^1 6(2t-1)dt = 6t^2 - 6t\Big|_0^1 = 0
\end{aligned}\right\} D = \begin{pmatrix} 0 & 2\sqrt{3} \\ 0 & 0 \end{pmatrix}$$

Checking the result: $D(\alpha_1 x_1 + \alpha_2 x_2) = D(\alpha_1 + \alpha_2\sqrt{3}(2t-1)) = 0 + \alpha_2\sqrt{3}\,2 = \alpha_2\sqrt{3}\,2\, x_1$

so as a matrix

$D\begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} = \begin{pmatrix} \sqrt{3}\,2\,\alpha_2 \\ 0 \end{pmatrix}$ 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.

Lets 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 $A = \begin{pmatrix} a_{11} & a_{12} & \cdots \\ a_{21} & a_{22} & \\ \vdots & & \ddots \end{pmatrix}$ where each row is a vector orthogonal to the rest of the rows.

so we might call it $A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \end{pmatrix}$ where each a_i has n components (or according to book $A = \{a_1, a_2, \ldots\}$)

Clearly $(AA^{+})_{ik} = \underbrace{(AA^{*T})}_{}{}_{ik} = \sum_j a_{ij} a^{*T}_{jk} = \sum_j a^{*}_{kj} a_{ij} = \langle a_k, a_i \rangle = \delta_{ki} \|a_i\|^2$

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

So $AA^{+} = \begin{pmatrix} \|a_1\|^2 & 0 & 0 \\ 0 & \|a_2\|^2 & \\ 0 & & \ddots \end{pmatrix} \Rightarrow \det(AA^{+}) = \|a_1\|^2 \|a_2\|^2 \cdots \|a_n\|^2$

Recall that: $\det A = (\det A^{+})^{*} \Rightarrow \det A^{*} = \det A^{+}$ and $\det(AB) = \det A \det B$

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

Now A in the above was special. Suppose we start w/ an arbitrary matrix $B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \end{pmatrix}$

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

Now let's Graham-Schmidt the shit out of it: $B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \end{pmatrix} \Rightarrow B' = \begin{pmatrix} b_1 / \|b_1\| \\ \dfrac{b_2 - \langle \frac{b_1}{\|b_1\|}, b_2 \rangle \frac{b_1}{\|b_1\|}}{\|b_2 - \langle \frac{b_1}{\|b_1\|}, b_2 \rangle \frac{b_1}{\|b_1\|}\|} \\ \vdots \end{pmatrix}$

Now let's form $C = \begin{pmatrix} \|b_1\| b_1' \\ \|b_2 - \langle \frac{b_1}{\|b_1\|}, b_2 \rangle \frac{b_1}{\|b_1\|}\| b_2' \\ \vdots \end{pmatrix} = \underbrace{\begin{pmatrix} b_1 \\ b_2 - \langle \frac{b_1}{\|b_1\|}, b_2 \rangle \frac{b_1}{\|b_1\|} \\ \vdots \end{pmatrix}}_{}$

This is special, since each row is formed by adding multiples of other rows to it, e.g. $b_2 \rightarrow b_2 - f b_1$

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

But since C is comprised of orthonormal vectors $\{b_1', b_2', \ldots, b_n'\}$ times scalar multiples $\underbrace{\{\beta_1, \beta_2, \ldots, \beta_n\} = \{\|b_1\|, \|b_2 - \langle \frac{b_1}{\|b_1\|}, b_2 \rangle \frac{b_1}{\|b_1\|}\|}_{}, \ldots, \}$ then $\det C = |\beta_1||\beta_2| \cdots |\beta_n| \, \|\cancel{b_1'}\| \|\cancel{b_2'}\| \cdots \|\cancel{b_n'}\|$

where each $|\beta_i| \leq \|b_i\|$, therefore we get [Hadamard's inequality: $|\det B| \leq \|b_1\| \|b_2\| \cdots$