The inner product defined on functions and the convenience of working in an orthonormal basis is one of the ideas that we will be developing in detail. Of course you have already been using them, e.g. Fourier nodes, Legendre polynomials, Bessel functions, etc. We are just going to dive a bit deeper into their construction, hopefully to provide an underlying theme which "unifies" the various results.

Also, clearly we will want to extend these finite dimensional results to infinite ones so that we can cover more functions.

For example, suppose we want to realize $e^t$ as a polynomial vector. Clearly to actually get $e^t$, the Taylor series includes all $(\infty)$ terms.
So let's get back to developing inner-products in general.

An orthonormal set is linearly independent (though not the reverse), hence provides a basis (and a complete one if it is complete).

Recall that if \( \sum \alpha_i x_i = 0 \) \( \forall \alpha_i = 0 \) then \( x_i \) is linearly independent.

Well consider \( (x_j, 0) = 0 = (x_j, \sum \alpha_i x_i) = \sum \alpha_i (x_j, x_i) = \sum \alpha_i \delta_{ij} = \alpha_j = 0 \)

So if a set is orthonormal then \( \sum \alpha_i x_i = 0 \Rightarrow \alpha_i = 0 \), hence linearly independent.

Bessel's inequality: If \( x_i \) is a finite orthonormal set in an inner-product space and \( x \) is any vector, then \( ||x||^2 \geq \sum ||x_i||^2 \) where \( ||x||^2 = (x, x) \).

Furthermore the vector \( x' = x - \sum x_i x_j \) is orthogonal to each \( x_j \).

Now why, why is there an inequality? Is it \( ||x|| = \sum ||x_i|| \) for an orthonormal basis? \( x = \sum x_i x_i \) \( \forall (x_i, x_j) = \delta_{ij} \) \( \Rightarrow ||x||^2 = \sum ||x_i||^2 \)

The answer is that in this \( x_i \) is an orthonormal set, not necessarily a complete basis. So it might be that \( x = \sum \alpha_i x_i + \sum \beta_j y_j \) which leads to \( ||x||^2 > \sum ||y_j||^2 \)

Then of course \( (x_i, x') = (x_i, x) - (x_i, \sum x_j x_j) = (x_i, x) - \sum \beta_j (x_i, y_j) \)

\( = \alpha_i - \alpha_i = 0 \) [by orthogonality \( \alpha_i - \alpha_i = 0 \)]
The question is "when do we know if our orthonormal set $X$ is a complete basis of $U$?"

Any of these say yes:

1. $X$ is complete (if not contained in larger set)
2. if $(x_i, x) = 0$ for all $i$ then $x = 0$
3. $X$ spans $V$ (every vector in $V$ is a linear combination of elements of $X$)
4. if $x \in V$ then $x = \sum (x_i, x)x_i$
5. if $x, y \in V$ then $(y, x) = \sum (y, x_i)(x_i, x)$ "Parseval's equation"
6. if $x \in V$ then $||x||^2 = \sum ||x_i||^2$

All of these provide different tests and definitions which identify a complete orthonormal set.

I won't go through the proofs of their connections, but will point out that a couple of them replace $Q \Rightarrow Y$ with $\neg Q \Rightarrow \neg \neg Y$. But this is what you should expect. E.g., if $x$ is blue $\Rightarrow x$ is color, but if $y$ is color $\Rightarrow y$ is blue, rather $y$ is color $\Rightarrow y \neq$ blue.

\[ x, y \rightarrow x \in \mathbb{R}^2, \text{ is set a complete orthonormal set} \]

\[ (1) \quad \{x_1, x_2\} \subseteq \{x_1, x_2, x_3 \Rightarrow y\} \]
\[ (2) \quad (\gamma, x_i) = 0, (\gamma, x) = 0, \text{ but } y \neq 0 \]
\[ (3) \quad y \neq x + \alpha x_2 \]
\[ (4) \quad y \neq (x, y) x_2 + (x_2, y) x_3 \]
\[ (5) \quad (x, y) = (x, x) (x, y)^{(\ast)} + (x, x_3) (x, y_3) = 0 \quad \text{but} \quad (x, y) \neq 0 \]
\[ (6) \quad \text{Remember Bessel?} \]

\[ \text{Schwarz's inequality: } \quad \text{For vectors } x \text{ and } y \Rightarrow \|x, y\| \leq \|x\| \|y\| \]
\[ \text{Needed since } (\gamma) \text{ can be negative} \]

To prove this we just say if $y = 0$ both sides vanish, but if $y \neq 0$ we form $\frac{x}{y^\ast}$ which by itself is orthonormal, then use Bessel's inequality $\|x\|^2 \geq \left( \frac{x}{y^\ast} \right)^\ast \Rightarrow \|x\| \|y\|^2 \geq (x, y)^2$

as seen since $\|y\|$ is the only one in the set.
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 $P_2$.

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 writable in terms of linear combinations of these. The orthonormal set we will call $Y$.

Start with, e.g., $y_0 = \frac{x_0}{\|x_0\|} = 1$, then $y_1 = \frac{(x_1 - x_0 y_0)}{\|x_1 - x_0 y_0\|}$ where we need to find $x_0$.

But we need $(y_0, y_1) = 0 \Rightarrow \frac{1}{\|x_1 - x_0 y_0\|} \left[ (1, x_1) - x_0 \left(1, y_0\right) \right] = 0$.

Now, $x_1 = \alpha_0$

Then $y_1 = \frac{(t-t)}{\|x_1 - x_0 y_0\|} = 0$, but wait that can't be!

The problem is $(1, x_1) = \int_0^1 t \, dt = 0 \neq x_1$ while $(1, y_0) = \int_0^1 \, dt = 1$.

So this leads to $\frac{1}{3} - \alpha_0 = 0 \Rightarrow \alpha_0 = \frac{1}{3}$. So $y_1 = \frac{(t - \frac{1}{3})}{\|t - \frac{1}{3}\|} = \sqrt{3}(t - \frac{1}{3})

So we have $Y = \{1, \sqrt{3}(t - \frac{1}{3})\}$?

To finish up:

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

$s.t. (y_0, y_2) = 0 \Rightarrow \frac{1}{\|x_2 - (\alpha_0 y_0 + \alpha_1 y_1)\|} \left[ (1, x_2) - \alpha_0 \left(1, y_0\right) - \alpha_1 \left(1, y_1\right) \right] = 0$.

So $y_2 = \frac{1}{\|x_2 - (\alpha_0 y_0 + \alpha_1 y_1)\|} \left[ (1, x_2) - \frac{1}{3} (1, y_0) - \frac{1}{\sqrt{3}} (1, y_1) \right] = 0$.

$y_2 = \frac{1}{\|x_2 - (\alpha_0 y_0 + \alpha_1 y_1)\|}$

Notice the pattern: $\alpha_0 = (y_0, x_0)$, $\alpha_1 = (y_1, x_1) \Rightarrow \alpha \in \langle y_n, x_n \rangle$ for $Y^n$.

So $Y = \{1, \sqrt{3}(t - \frac{1}{3}), \sqrt{180} \left(t^2 - t + \frac{1}{3}\right)\}$ is an 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 with 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 \( x_i \)s say \( x_1 \) and forming \( y_1 = \frac{x_1}{\|x_1\|} \).

Then \( y_{n+1} = \frac{x_{n+1} - \sum_{j=1}^{n} \langle x_{n+1}, y_j \rangle y_j}{\|x_{n+1} - \sum_{j=1}^{n} \langle x_{n+1}, y_j \rangle y_j\|} \).

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 component 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 \( A x_j = \sum_k a_{kj} x_k = \langle x_i, A x_j \rangle = \langle x_i, \sum_k a_{kj} x_k \rangle = \sum_k a_{kj} \langle x_i, x_k \rangle = \sum_k a_{kj} \delta_{ik} = a_{ij} \).

Let's see this in action. Going back to \( D \) on \( P_1 \), we have worked out the non-orthonormal basis \( \{ 1, x \} \Rightarrow D = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \) (we already found the matrix form of \( D \), so let's check \( \langle x, 0 \rangle \rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \),

\( \langle x, x \rangle = \langle 1, 1 \rangle = 1 \),

\( \langle x, 0 \rangle \rangle = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \rangle = 0 \),

\( \langle x, 0 \rangle \rangle = \langle 1, 1 \rangle = 1 \).

\[ \begin{align*}
(x_1, 0 x_1) &= \langle 1, 0 \rangle = (0, 0) \\
(x_1, 1 x_1) &= \langle 1, 1 \rangle = (1, 0) = D_{11}.
\end{align*} \]

\( D = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \)

If instead we worked with an orthonormal basis (from Gram-Schmidt): \( \{ 1, \sqrt{3} (4t-1) \} \) then

\[ \begin{align*}
(x_1, 0 x_1) &= \langle 1, 0 \rangle = (1, 0) \rangle = 0 \\
(x_1, 1 x_1) &= \langle 1, \sqrt{3} (4t-1) \rangle = \begin{pmatrix} 1 \\ \sqrt{3} (4t-1) \end{pmatrix} = (0, 1) = D_{11}.
\end{align*} \]

\[ (x_1, 0 x_1) = \langle 1, 0 \rangle = (0, 0) = D_{11}.
\]

\[ (x_1, 1 x_1) = \langle 1, \sqrt{3} (4t-1) \rangle = \begin{pmatrix} 1 \\ \sqrt{3} (4t-1) \end{pmatrix} = (0, 1) = D_{11}.
\]

\[ C = \begin{pmatrix} 1 \\ \sqrt{3} (4t-1) \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} = D_{11}.
\]

Checking the result: \( D (x_0 + \sqrt{3} (4t-1)) = \begin{pmatrix} 1 & \sqrt{3} (4t-1) \end{pmatrix} = \begin{pmatrix} 0 & 0 \end{pmatrix} = D \begin{pmatrix} 1 \\ 0 \end{pmatrix} \).

So as a matrix

\[ D = \begin{pmatrix} 1 & \sqrt{3} (4t-1) \\ 0 & 0 \end{pmatrix} \]

which is what we found above does!