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.

In a finite-dimensional vector space, an orthonormal set is complete if it is not contained in any larger orthonormal set.

Now that we have definitions, let's look at some results.

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 $\sum \lambda_i \mathbf{x}_i = 0$ and $\lambda_i = 0$ then $\mathbf{x}_i$ is linearly independent.

We consider $(\mathbf{x}_j, \mathbf{0}) = 0 = (\mathbf{x}_j, \sum \lambda_i \mathbf{x}_i) = \sum \lambda_i (\mathbf{x}_j, \mathbf{x}_i) = \sum \lambda_i \delta_{ij} = \lambda_j = 0$

So if a set is orthonormal then $\sum \lambda_i \mathbf{x}_i = 0 \Rightarrow \lambda_i = 0$, hence linearly independent.

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

Furthermore the vector $\mathbf{x}' = \mathbf{x} - \sum \lambda_i \mathbf{x}_i$ is orthogonal to each $\mathbf{x}_i$.

Now, wait, why is there an inequality? Is that $\|\mathbf{x}\|^2 = \sum \|\mathbf{x}_i\|^2$ for an orthonormal basis? $\mathbf{x} = \sum \mathbf{x}_i \Rightarrow (\mathbf{x}, \mathbf{x}) = \sum \lambda_i (\mathbf{x}_i, \mathbf{x}) = \sum \|\mathbf{x}_i\|^2$

The answer is that in this $\mathbf{x}_i$ is an orthonormal set, not necessarily a complete basis. So it might be that $\mathbf{x} = \sum \mathbf{x}_i + \sum \lambda_i \mathbf{x}_i$, which leads to $\|\mathbf{x}\|^2 > \sum \|\mathbf{x}_i\|^2$.

Notice this is $\mathbf{x}'$.

Then of course $(\mathbf{x}, \mathbf{x}') = (\mathbf{x}, \mathbf{x}) - (\mathbf{x}, \sum \lambda_i \mathbf{x}_i) = (\mathbf{x}, \mathbf{x}) - \sum \lambda_i (\mathbf{x}_i, \mathbf{x})$

$= \mathbf{x}_i - \mathbf{x}_i = 0$  [buses have $\mathbf{x}_i - \mathbf{x}_i = 0$]

so they do $(\mathbf{x}', \mathbf{x})$.
The question is "when do we know if our orthonormal set $X$ is a complete basis of $V$?"

Any of these say yes:

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

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

I won't go through the proof of their connections, but will point out that a couple of them replace $Q \Rightarrow y \in X$ not ($S \Rightarrow$ not $C$). But this is what you should expect. E.g., if $x$ is blue $\Rightarrow x \notin color$, but if $y \in color \Rightarrow y = blue$, rather $y \notin color \Rightarrow y \neq blue$.

Some examples of failures:

\begin{itemize}
  \item $(x_1, x_2) \in \{x_1, x_2, x_3 \mid y\}$
  \item $\langle y_1, x_i \rangle = 0, \langle y_1, x_4 \rangle = 0$, but $y \neq 0$
  \item $y \neq x_1 x_3 + x_2 x_4$
  \item $y \neq (x_1, y) x_3 + (x_2, y) x_4 = 0$
  \item $(y_1, y) = \langle y_1, x_1 \rangle \langle x_1, y \rangle + \langle y_1, x_2 \rangle \langle x_2, y \rangle = 0$ but $(y_1, y) \neq 0$
  \item Remember Bessel?
\end{itemize}

\[ \text{Schwarz's inequality: For vectors } x \text{ and } y \Rightarrow \|x\| \|y\| \geq \langle x, y \rangle \]

\[ \text{To prove this we just say if } y = 0 \text{ both sides vanish, but if } y \neq 0 \text{ we form } \frac{y}{\|y\|} \text{ which by itself is orthonormal, then use Bessel's inequality } \|x\| \|\frac{y}{\|y\|}\| \geq \langle x, \frac{y}{\|y\|} \rangle \]

\[ \text{No sum since } \|\frac{y}{\|y\|}\| \text{ is the only one in the set} \]
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 trickier. Luckily, we have a process due to Gram-Schmidt.

While the book takes you through it in general, we will apply it to \( P_1 \).

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 \( Y \), remains from \( x_i \) on concentulating along \( y_0 \).

Let's start with \( t \), e.g., \( y_0 = \frac{x_1}{\|x_1\|} = 1 \), then \( y_1 = \frac{x_1 - \alpha_0 y_0}{\|x_1 - \alpha_0 y_0\|} \) where we need to find \( \alpha_0 \). But we need \( (y_0, y_1) = 0 \) \( \Rightarrow \frac{1}{\|x_1 - \alpha_0 y_0\|} \left[ (1, x_1) - \alpha_0 (1, y_0) \right] = 0 \)

\( x_1 = \alpha_0 \) \( y_1 = \frac{1}{\|x_1 - \alpha_0 y_0\|} \left( \frac{1}{\|x_1 - \alpha_0 y_0\|} \right) = 1 \)

Then \( y_1 = \frac{1}{\|x_1 - \alpha_0 y_0\|} \left( \frac{1}{\|x_1 - \alpha_0 y_0\|} \right) = 1 \), but wait that can't be!

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

So this leads to \( \frac{1}{2} - \alpha_0 = 0 \) \( \Rightarrow \alpha_0 = \frac{1}{2} \) so \( y_1 = \frac{t - \frac{1}{2}}{\frac{1}{2}} = \sqrt{2} (t - \frac{1}{2}) \)

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

To finish up:

\[ y_2 = \frac{x_3 - (\alpha_0 y_0 + \alpha_1 y_1)}{\|x_3 - (\alpha_0 y_0 + \alpha_1 y_1)\|} \]

So \( (y_0, y_2) = 0 \) \( \Rightarrow \frac{1}{\|x_3 - (\alpha_0 y_0 + \alpha_1 y_1)\|} \left[ (1, x_3) - \alpha_0 (1, y_0) - \alpha_1 (1, y_1) \right] = 0 \)

\( (y_1, y_2) = 0 \) \( \Rightarrow \frac{1}{\|x_3 - (\alpha_0 y_0 + \alpha_1 y_1)\|} \left[ (\sqrt{2} (t - \frac{1}{2}), t^2) - \frac{1}{2} (\sqrt{2} (t - \frac{1}{2}), 1) - \alpha_1 (1, \sqrt{2} (t - \frac{1}{2})) \right] = 0 \)

\( y_2 = \sqrt{3/4} (t - \frac{1}{2}) + \frac{1}{6} \) Notice the pattern: \( \alpha_0 = (y_0, x_2) \), \( \alpha_1 = (y_1, x_2) \) \( \Rightarrow \alpha_2 = (y_2, x_2) \) for \( y_2 \)

\( y = \{1, \sqrt{2} (t - \frac{1}{2}), \sqrt{3/4} (t - \frac{1}{2}) + \frac{1}{6} \} \) is our 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) \( \mathbf{X} = \{ \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n \} \) and forms an orthonormal basis \( \mathbf{Y} = \{ \mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_n \} \) by selecting one of the \( \mathbf{x}_i \)'s say \( \mathbf{x}_1 \) and forming \( \mathbf{y}_1 = \frac{\mathbf{x}_1}{\| \mathbf{x}_1 \|} \) then

\[
\mathbf{y}_{n+1} = \frac{\mathbf{x}_{n+1} - \langle \mathbf{y}_1, \mathbf{x}_{n+1} \rangle \mathbf{y}_1}{\| \mathbf{x}_{n+1} - \langle \mathbf{y}_1, \mathbf{x}_{n+1} \rangle \mathbf{y}_1 \|}
\]

So \( \mathbf{y}_1 = \frac{\mathbf{x}_1}{\| \mathbf{x}_1 \|} \), \( \mathbf{y}_2 = \frac{\mathbf{x}_2 - \langle \mathbf{y}_1, \mathbf{x}_2 \rangle \mathbf{y}_1}{\| \mathbf{x}_2 - \langle \mathbf{y}_1, \mathbf{x}_2 \rangle \mathbf{y}_1 \|} \), \( \mathbf{y}_3 = \frac{\mathbf{x}_3 - \langle \mathbf{y}_1, \mathbf{x}_3 \rangle \mathbf{y}_1 - \langle \mathbf{y}_2, \mathbf{x}_3 \rangle \mathbf{y}_2}{\| \mathbf{x}_3 - \langle \mathbf{y}_1, \mathbf{x}_3 \rangle \mathbf{y}_1 - \langle \mathbf{y}_2, \mathbf{x}_3 \rangle \mathbf{y}_2 \|} \), etc.

so \( \langle \mathbf{y}_i, \mathbf{y}_j \rangle = \delta_{i,j} \).

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 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 arises because \( \mathbf{A} \mathbf{x}_j = \sum_k a_{kj} \mathbf{x}_k \Rightarrow \mathbf{A} \mathbf{x}_j = \sum_k a_{kj} \mathbf{x}_k \mathbf{y}_k = \langle \mathbf{x}_j, \mathbf{A} \mathbf{x}_k \rangle \mathbf{y}_k = \sum_k a_{kj} \langle \mathbf{x}_i, \mathbf{x}_k \rangle \mathbf{y}_k = \sum_k a_{kj} \delta_{ik} \mathbf{y}_k = a_{ij} \mathbf{y}_i \).

Let’s see this in action. Going back to \( \mathbf{D} \) on \( \mathbb{F}_2 \), we have worked with the non-orthonormal basis \( \{ 1, 1 \sqrt{2} (2t - 1) \} \) (we already found the matrix form of \( \mathbf{D} \)), so let’s check \( \langle \mathbf{x}_i, \mathbf{D} \mathbf{x}_j \rangle = a_{ij} \).

\[
\begin{align*}
\langle \mathbf{x}_1, \mathbf{D} \mathbf{x}_1 \rangle &= \langle (1, 0), (0, 1) \rangle = 0 = d_{11} \\
\langle \mathbf{x}_1, \mathbf{D} \mathbf{x}_2 \rangle &= \langle (1, 0), (\sqrt{2}, 1) \rangle = \sqrt{2} = d_{12} \\
\langle \mathbf{x}_2, \mathbf{D} \mathbf{x}_1 \rangle &= \langle (\sqrt{2}, 1), (1, 0) \rangle = 0 = d_{21} \\
\langle \mathbf{x}_2, \mathbf{D} \mathbf{x}_2 \rangle &= \langle (\sqrt{2}, 1), (\sqrt{2}, 1) \rangle = 2 = d_{22}
\end{align*}
\]

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

\[
\begin{align*}
\langle \mathbf{x}_1, \mathbf{D} \mathbf{x}_1 \rangle &= \langle (1, 0), (0, 1) \rangle = 0 = d_{11} \\
\langle \mathbf{x}_1, \mathbf{D} \mathbf{x}_2 \rangle &= \langle (1, 0), (\sqrt{2}, 1) \rangle = 0 = d_{12} \\
\langle \mathbf{x}_2, \mathbf{D} \mathbf{x}_1 \rangle &= \langle (\sqrt{2}, 1), (1, 0) \rangle = 0 = d_{21} \\
\langle \mathbf{x}_2, \mathbf{D} \mathbf{x}_2 \rangle &= \langle (\sqrt{2}, 1), (\sqrt{2}, 1) \rangle = 2 = d_{22}
\end{align*}
\]

Checking the result:

\[
\mathbf{D} (\mathbf{x}_0 + \mathbf{x}_1) = \mathbf{D} (\mathbf{x}_0 + \mathbf{x}_1, 1 \sqrt{2} (2t - 1)) = 0 + \mathbf{x}_1, 1 \sqrt{2} \mathbf{x}_0 = \mathbf{x}_1, 1 \sqrt{2} \mathbf{x}_0
\]

so as a matrix

\[
\mathbf{D} = \begin{pmatrix} 0 & 1 \\ 1 \sqrt{2} (2t - 1) & 0 \end{pmatrix}
\]

which is what the \( \mathbf{D} \) from above does!