Our goal: For a function \( f \in L^2_w(a, b) \), find \( p_n \in P_n \) such that:

\[
\| f - p_n \|_2^2 = \inf_{q \in P_n} \| f - q \|_2^2
\]

\[
= \inf_{q \in P_n} \int_a^b (f(x) - q(x))^2 w(x) \, dx.
\]

**Proof**

Let \( n = 0 \), \( f(x) = -2x^2 \) on \([-1, 1]\), \( w(x) = 1 \).

Find \( p_0 = c \) to minimize \( \| f - p_0 \|_2^2 \).

\[
\| f - p_0 \|_2^2 = \int_{-1}^{1} (-2x^2 - c)^2 \, dx
\]

\[
= 2c^2 + \frac{8c}{3} + \frac{8}{5}
\]

Find \( c \) to minimize this quantity.

\[
\frac{d}{dc} \| f - p_0 \|_2^2 = 4c + \frac{8}{3} = 0
\]

\[
\Rightarrow c = -\frac{2}{3}
\]

What would the best approximation in \( \| . \|_\infty \) be?

**Computing the best 2-norm approximation**

Recall: The least squares solution to \( Ax = b \) is obtained by solving \( A^\top A x = A^\top b \)

projection of \( b \) onto \( \text{col} A \), \( Q \) is obtained by applying the Gram-Schmidt process to the columns of \( A \).

To apply Gram-Schmidt, all that was needed was a vector space and an inner product.
We can do the same thing for polynomial least squares:

1. Trivially, $P_n$ is a vector space.

2. We can define an inner product on $P_n$ by:
   \[ (f, g) = \int_a^b f(x) g(x) w(x) \, dx. \]

Two functions are orthogonal if $(f, g) = 0$.

Let $p_0, p_1, p_2, \ldots, p_n$ be a basis for $P_n$.

The 2-norm approximation problem takes the form:

For $q(x) = \sum_{j=0}^{n} c_j p_j(x)$, $\| f - q \|^2_2$ is given by:

\[
A = \int_a^b \left( f(x) - \sum_{j=0}^{n} c_j p_j(x) \right)^2 \, dx
= \int_a^b (f(x))^2 \, dx - 2 \sum_{j=0}^{n} c_j \int_a^b f(x) p_j(x) \, dx + \sum_{j=0}^{n} \sum_{k=0}^{n} c_j c_k \int_a^b p_j(x) p_k(x) \, dx
= (f, f) - 2 \sum_{j=0}^{n} c_j (f, p_j) + \sum_{j=0}^{n} \sum_{k=0}^{n} c_j c_k (p_j, p_k)
\]

Writing down

\[ DA = 0, \]

we have that:

\[
\frac{DA}{dc_0} = 0, \quad \frac{DA}{dc_1} = 0, \quad \ldots, \quad \frac{DA}{dc_n} = 0
\]

\[ \Rightarrow \frac{DA}{dc_k} = -2 (f, p_k) + 2 \sum_{k=0}^{n} c_k (p_k, p_k) = 0 \quad \text{linear}
\]

\[ \Rightarrow \sum_{k=0}^{n} c_k (p_k, p_k) = (f, p_k) \quad \text{in n+1 equations, in n+1 variables} \quad c_0 \ldots c_n.
\]

In matrix form:

\[
\begin{pmatrix}
(p_0, p_0) & \ldots & (p_0, p_n) \\
(p_1, p_0) & \ldots & (p_1, p_n) \\
\vdots & & \vdots \\
(p_n, p_0) & \ldots & (p_n, p_n)
\end{pmatrix}
\begin{pmatrix}
c_0 \\
c_1 \\
\vdots \\
c_n
\end{pmatrix}
= \begin{pmatrix}
(f, p_0) \\
(f, p_1) \\
\vdots \\
(f, p_n)
\end{pmatrix}
\]
Therefore once this linear system has been solved, the best 2-norm approximation to \( f \) is:
\[
c^p = c_0 p_0 + c_1 p_1 + \ldots + c_n p_n.
\]

If the \( p_i \)'s were orthonormal, i.e. \((p_i, p_k) = \delta_{ik}\),
then the above system is of the form:
\[
I \ c = \text{RHS}
\]
so \( c^* = (f, p_i) \). (if \( p_i \)'s are orthonormal).