Let \([a, b]\) be a closed interval. Consider for \(f, g\) continuous functions on the interval an inner product of the form
\[
\langle f, g \rangle = \int_a^b f(x)g(x)w(x) \, dx,
\]
where \(w(x)\) is a positive weight function. We have the following algebraic properties of an inner product. Let \(c\) be a constant, and let \(f, g, h\) be functions on \([a, b]\). Then
\[
\langle f, g \rangle = \langle g, f \rangle
\]
\[
\langle cf, g \rangle = c \langle f, g \rangle
\]
\[
\langle f + g, h \rangle = \langle f, h \rangle + \langle g, h \rangle.
\]
All these properties follow immediately from basic integral formulas.

The inner product notation provides a natural setting for the error in a continuous least squares approximation on an interval \([a, b]\) with weight function \(w(x)\). If \(f\) is a function, and \(P\) is an approximation (polynomial or otherwise) of \(f\), then the error
\[
E = \int_a^b [f(x) - P(x)]^2 w(x) \, dx = \langle f - P, f - P \rangle.
\]
The algebraic formulas above show that
\[
E = \langle f, f \rangle - 2\langle f, P \rangle + \langle P, P \rangle.
\]
A set of functions \(\{\phi_0, \ldots, \phi_n\}\) is said to be orthogonal if we have
\[
\langle \phi_i, \phi_j \rangle = \begin{cases} 
0 & \text{if } i \neq j \\
\alpha_j > 0 & \text{if } i = j 
\end{cases}
\]
A function is a linear combination of \(\phi_0, \ldots, \phi_n\) if it is of the form
\[
\sum_{i=0}^n a_i \phi_i
\]
for constants \(a_i\). Given an orthogonal set of functions, we have the following formula for least squares approximations:

**Proposition 1.** Consider an interval \([a, b]\) and a weight function \(w(x)\). Let \(\{\phi_0, \ldots, \phi_n\}\) be an orthogonal set of functions, and let \(f\) be a continuous function on \([a, b]\). Then among all linear combinations of the
\( \phi_i \), the least squares approximation is given by

\[
\sum_{i=0}^{n} a_i \phi_i \quad \text{for} \quad a_i = \frac{\langle \phi_i, f \rangle}{\langle \phi_i, \phi_i \rangle} = \alpha_i^{-1} \langle \phi_i, f \rangle
\]

**Proof.** Let \( a_0, \ldots, a_n \) be constants and consider the approximation function \( \sum_{i=0}^{n} a_i \phi_i - f \). So the error is

\[
E = \left\langle \sum_{i=0}^{n} a_i \phi_i - f, \sum_{i=0}^{n} a_i \phi_i - f \right\rangle.
\]

\( E \) is a quadratic function of the \( n + 1 \) variables \( a_0, \ldots, a_n \), and we will compute the critical point to find the values of \( a_0, \ldots, a_n \) which determine the minimizer.

Now we compute \( E \) using the algebra of inner products above

\[
E = \left\langle \sum_{i=0}^{n} a_i \phi_i - f, \sum_{i=0}^{n} a_i \phi_i - f \right\rangle
= \left\langle \sum_{i=0}^{n} a_i \phi_i, \sum_{i=0}^{n} a_i \phi_i \right\rangle - 2 \left\langle \sum_{i=0}^{n} a_i \phi_i, f \right\rangle + \langle f, f \rangle
\]

Now the middle term computes to be

\[
\left\langle \sum_{i=0}^{n} a_i \phi_i, f \right\rangle = \sum_{i=0}^{n} \langle a_i \phi_i, f \rangle = \sum_{i=0}^{n} a_i \langle \phi_i, f \rangle.
\]

To compute the first term in the error, we will find

\[
\left\langle \sum_{i=0}^{n} a_i \phi_i, \sum_{i=0}^{n} a_i \phi_i \right\rangle = \sum_{i=0}^{n} a_i^2 \langle \phi_i, \phi_i \rangle.
\]

To illustrate the role of orthogonality, we will just compute this in the simple case \( n = 1 \):

\[
\langle a_0 \phi_0 + a_1 \phi_1, a_0 \phi_0 + a_1 \phi_1 \rangle = a_0^2 \langle \phi_0, \phi_0 \rangle + 2a_0a_1 \langle \phi_0, \phi_1 \rangle + a_1^2 \langle \phi_1, \phi_1 \rangle.
\]

The orthogonality condition says that \( \langle \phi_0, \phi_1 \rangle = 0 \), and so the cross term vanishes and we have

\[
\langle a_0 \phi_0 + a_1 \phi_1, a_0 \phi_0 + a_1 \phi_1 \rangle = a_0^2 \langle \phi_0, \phi_0 \rangle + a_1^2 \langle \phi_1, \phi_1 \rangle.
\]

For the case of larger values of \( n \), there are many more cross terms, which will all vanish with \( \langle \phi_i, \phi_j \rangle = 0 \) for \( i \neq j \). So all together, we
have computed
\[
E = \sum_{i=0}^{n} a_i^2 \langle \phi_i, \phi_i \rangle - 2 \sum_{i=0}^{n} a_i \langle \phi_i, f \rangle + \langle f, f \rangle
\]
\[
= \sum_{i=0}^{n} (a_i^2 \langle \phi_i, \phi_i \rangle - 2a_i \langle \phi_i, f \rangle) + \langle f, f \rangle.
\]

The critical point for \( E \) satisfy \( \frac{\partial E}{\partial a_k} = 0 \) for all values of \( k = 0, \ldots, n \). To compute the partial derivatives,
\[
\frac{\partial E}{\partial a_k} = \frac{\partial}{\partial a_k} \left[ \sum_{i=0}^{n} (a_i^2 \langle \phi_i, \phi_i \rangle - 2a_i \langle \phi_i, f \rangle) + \langle f, f \rangle \right]
\]
\[
= \sum_{i=0}^{n} \left( \frac{\partial}{\partial a_k} a_i^2 \langle \phi_i, \phi_i \rangle - 2 \frac{\partial}{\partial a_k} a_i \langle \phi_i, f \rangle \right) + 0
\]
\[
= 2a_k \langle \phi_k, \phi_k \rangle - 2 \langle \phi_k, f \rangle
\]
The reason for the last equality is that \( \frac{\partial}{\partial a_k} a_i \) is 0 for \( i \neq k \) and is 1 for \( i = k \). Therefore, only the term \( i = k \) remains in the sum.

So then we can solve the equation \( \frac{\partial E}{\partial a_k} = 0 \) for \( a_k \) to find that
\[
a_k = \frac{\langle \phi_k, f \rangle}{\langle \phi_k, \phi_k \rangle}
\]
for all \( k = 0, \ldots, n \). \( \square \)