Best Approximation: Least-squares Theory

We have seen that finding the minimax approximation is complicated. Here we discuss best approximation in the least-squares sense. We shall see that the problem reduces to solving a system of linear equations.

Let \( f \in C[a, b] \) and we want to approximate \( f \) by \( p \in \Pi_n \). Note that \( C[a, b] \) is an inner product space. Suppose we measure the difference by the

\[
\|f - p\|^2 = \langle f - p, f - p \rangle = \int_a^b \left( f(x) - p(x) \right)^2 \omega(x) \, dx
\]

where \( \omega \) is a continuous positive function (weight function). Such a polynomial is called the least squares approximation to \( f \) by a polynomial of degree \( \leq n \).

The following theorem holds for a subspace \( W \) of an inner product space \( V \). Hence it will also hold if we take \( V = C[a, b] \) and \( W = \Pi_n \).

**Theorem:** Let \( W \) be a subspace of an inner product space \( V \). For \( f \in V \) and \( p \in W \), the following properties are equivalent.

1. \( p \) is best approximation to \( f \) in \( W \)
2. \( f - p \perp W \)

**Proof:** Let \( f - p \perp W \). If \( q \in W \), then (using \( p - q \perp f - p \))

\[
\|f - q\|^2 = \|f - p + p - q\|^2 = \|f - p\|^2 + \|p - q\|^2 \geq \|f - p\|^2
\]

Hence the result follows.

Conversely, let \( p \) is the best approximation and \( \epsilon > 0 \). Now for \( q \in W \), we must have (since \( p - \epsilon q \in W \))

\[
\|f - p\|^2 \leq \|f - p + \epsilon q\|^2
\]

This on simplification gives

\[
\epsilon \left( 2 \langle f - p, q \rangle + \epsilon \|q\|^2 \right) \geq 0
\]

Now we make \( \epsilon \to 0^+ \), which gives \( \langle f - p, q \rangle \geq 0 \). Similarly, taking \( \epsilon < 0 \), we get \( \langle f - p, q \rangle \leq 0 \). Combining both and noting that \( q \) is an arbitrary element of \( W \), we get \( f - p \perp W \).

**Example:** Consider a function \( f \in C[0, 1] \). We want least-squares approximate to it by \( p \in \Pi_n \). We know that \( \{1, x, \cdots, x^n\} \) is a basis of \( \Pi_n \). Hence, by the above theorem \( f - p \) must be orthogonal to \( x^i \) for \( i = 0, 1, 2, \cdots, n \). Let

\[
p = \sum_{i=0}^{n} a_i x^i
\]

Then we arrive at a linear system \( AX = b \), where \( X = (a_0, a_1, \cdots, a_n)^T \), \( b = (b_0, b_1, \cdots, b_n)^T \) and \( A = (A_{i,j})_{n+1 \times n+1} \) and 0 \leq i, j \leq n. Further,

\[
b_i = \int_0^1 f(x)x^i \, dx, \quad A_{i,j} = \int_0^1 x^{i+j} \, dx = \frac{1}{i+j+1}
\]

The matrix is called Hilbert matrix which is ill-conditioned as \( n \) becomes larger. Thus the basis chosen for the least-squares approximation is poor.
1 Orthonormal system

A set of vectors (here they are functions) $f_1, f_2, \cdots$ is called orthonormal if $< f_i, f_j > = \delta_{ij}$. Let us consider the basis of the approximation space that consists of orthonormal vectors.

**Theorem:** Let $g_1, g_2, \cdots, g_n$ be an orthonormal basis for $W$. The best least-squares approximation to $f \in V$ by $g = \sum_{i=1}^{n} a_i g_i \in W$ is obtained iff $a_i = < f, g_i >$.

**Proof:** By the previous theorem, the best approximation satisfy $f - g \perp W$. Hence $f - \sum_{i=1}^{n} a_i g_i$ must be perpendicular to each $g_i$:

$$< f - \sum_{i=1}^{n} a_i g_i, g_k >= 0 \implies c_k = < f, g_k >$$

Hence, once we know the orthonormal basis, finding the best least-squares approximation is trivial. Note that a set of orthonormal vectors can be found by the Gram-Schmidt process.

**Theorem:** The orthogonal polynomial $p_0, p_1, \cdots, p_n$ satisfy a three term recurrence relation as follows:

$$p_n(x) = (x - a_n)p_{n-1}(x) - b_n p_{n-2}(x), \quad n \geq 2$$

with $p_0(x) = 1$, $p_1(x) = x - a_1$ and

$$a_n = < xp_{n-1}, p_{n-1} > / < p_{n-1}, p_{n-1} >, \quad b_n = < xp_{n-1}, p_{n-2} > / < p_{n-2}, p_{n-2} >$$

**Proof:** It is clear that each polynomial $p_i$ is a monic polynomial of degree $i$. Now we show by induction that $< p_i, p_i > = 0$ for $i = 0, 1, 2, \cdots, n - 1$. For $n = 0$, nothing to prove. For $n = 1$, we find

$$< p_1, p_0 > = < xp_0 - a_1 p_0, p_0 > = < xp_0, p_0 > - a_1 < p_0, p_0 > = 0$$

Assume that the result is true for $n = k$ and hence $< p_k, p_i > = 0$ for $k = 0, 1, 2, \cdots, k - 1$. Now for $n = k + 1$, we find $p_{k+1}(x) = (x - a_{k+1}) p_k(x) - b_{k+1} p_{k-1}(x)$. We need to show that $< p_{k+1}, p_i > = 0$ for $i = 0, 1, 2, \cdots, k$. Now

$$< p_{k+1}, p_k > = < xp_{k+1}, p_k > - a_{k+1} < p_k, p_k > - b_{k+1} < p_{k-1}, p_k > = 0$$

$$< p_{k+1}, p_{k-1} > = < xp_{k+1}, p_{k-1} > - a_{k+1} < p_k, p_{k-1} > - b_{k+1} < p_{k-1}, p_{k-1} >$$

$$= < xp_k, p_{k-1} > - b_{k+1} < p_{k-1}, p_{k-1} > = 0$$

For $0 \leq i \leq k - 2$, we find (using $< x f, g > = < f, x h >$)

$$< p_{k+1}, p_i > = < xp_{k+1}, p_i > - a_{k+1} < p_k, p_i > - b_{k+1} < p_{k-1}, p_i >$$

$$= < xp_k, p_i > - a_{k+1} < p_{k+1}, p_i > - b_{k+1} < p_{k-1}, p_i >$$

$$= < xp_k, p_i > - a_{k+1} < p_{k+1}, p_i > - b_{k+1} < p_{k-1}, p_i > - a_i p_i + b_i + b_{i+1} p_i - a_{i+1} p_i = 0$$

where for $i = 0$, we use $x p_0 = p_1 + a_1 p_0$. We have used recurrence relation in the last step.

**Theorem:** The orthogonal polynomial $p_n$ derived in the previous theorem is monic polynomial of degree $n$ and it has the lowest (square) norm among all monic polynomial of degree $n$.

**Proof:** Let $q$ be an arbitrary monic polynomial of degree $n$. Then we can write $q = p_n - \sum_{i=0}^{n-1} a_i p_i$. The norm of $q$ will be minimum if $q \perp \Pi_{n-1}$. Now $p_n \perp \Pi_{n-1}$ and hence this implies $a_i = 0$ for $i = 0, 1, 2, \cdots, n - 1$. 
Theorem: An orthogonal polynomial $p_n(x)$ ($n \geq 1$) has $n$ simple real zeros which lie in the interval $(a,b)$.

Proof: Let $x_{1,n}, x_{2,n}, \ldots, x_{r,n}$ that lie in $(a,b)$ are the points where $p_n(x)$ changes sign. We claim that $r \geq n$. If $r < n$, then take $\bar{x} \in (\max x_{i,n}, b)$ and consider

$$q(x) = p_n(\bar{x})(x - x_{1,n})(x - x_{2,n}) \cdots (x - x_{r,n})$$

is a polynomial of degree $< n$ which has the same sign as that of $p_n$ in whole $(a,b)$. Since $q$ has degree $< n$, we have $< p, p_n > = 0$. But because $p$ and $p_n$ has same sign, we must have $p(x)p_n(x)\omega(x) > 0$ at all $x \in (a,b)$ except $x_{1,n}, x_{2,n}, \ldots, x_{r,n}$ where it is zero. Hence $< p, p_n > > 0$. This contradiction implies that $r \geq n$. Hence $p_n$ has at least $n$ zeros. Since $p_n$ does not have more than $n$ zeros, we must have $r = n$.