Continuous Least Squares

- We have already discussed the idea in the past, but for our purposes today let’s reintroduce it in its simplest form.

- Suppose \( f \) is continuous on \([a, b]\) and we want to approximate it by a linear combination of basis functions

\[
f(x) \approx p(x) = \sum_{i=1}^{n} \alpha_i \phi_i(x)
\]

(1)

- The basis functions \( \phi_i \) could be polynomials, for example,

\[
\phi_i(x) = x^{i-1},
\]

but they may not be.

- For example, we could use periodic, exponential, or rational functions depending on what we know about \( f \).

- We pick the coefficients \( \alpha_j \) such that

\[
F(\alpha_1, \ldots, \alpha_n) = \int_{a}^{b} \left( \sum_{j=1}^{n} \alpha_j \phi_j(x) - f(x) \right)^{2} \, dx = \min.
\]

(3)
• This approach is called **Continuous Linear Least Squares**
  
  - **Continuous** because we are approximating a function, not a vector (which would lead to a discrete least squares problem).
  
  - **Linear** because we approximate by a linear combination of basis functions, which will give rise to a linear system of equations. Linear does not mean that the approximating function \( p \) is linear!
  
  - **Least** because we minimize.
  
  - **Squares** because we minimize the integral of a square.

• We can find the linear system by the standard Calculus approach of computing partial derivatives of \( F \) and setting them to zero:

\[
\frac{\partial}{\partial \alpha_i} F(\alpha_1, \ldots, \alpha_n) = \int_a^b 2 \left( \sum_{j=1}^n \alpha_j \phi_j(x) - f(x) \right) \phi_i(x) = 0.
\]

(4)

• This can be rewritten as the linear system

\[
\sum_{j=1}^n \alpha_j \int_a^b \phi_j(x) \phi_i(x)dx = \int_a^b \phi_i(x) f(x)dx, \quad i = 1, 2, \ldots, n
\]

(5)

• We have actually seen this before in the special case that

\[
[a, b] = [0, 1] \quad \text{and} \quad \phi_i(x) = x^{i-1}.
\]

(6)
• The coefficient matrix of that linear system is the infamous **Hilbert Matrix**

\[
\left[ \int_0^1 x^{i-1} x^{j-1} dx \right] = \left[ \frac{1}{i+j-1} \right]_{i,j,1,...,n} \tag{7}
\]

• of course, we know that the Hilbert Matrix is extremely ill-conditioned, so we need to be careful with our choice of basis functions.

• Moreover, the integral \( \int_a^b f(x)g(x)dx \) looks much like the **dot product**

\[
\mathbf{f} \cdot \mathbf{g} = \mathbf{f}^T \mathbf{g} = \sum_{i=1}^{n} f_i g_i \tag{8}
\]

of two **vectors** \( \mathbf{f} \) and \( \mathbf{g} \).

• It’s a powerful simplifying principle that integrals behave just like sums.

• There must be a way to exploit the similarity.

• The relevant concept here is that of an **inner product**, which is a generalization of the familiar **dot product**”.

• The basic generalization procedure (we saw this earlier for norms) is to ask what properties of some concept are important, and then what other objects have the same properties.

• The relevant important properties of the dot
product are:

\[ u^T v = v^T u \]
\[ u^T u \geq 0 \]
\[ u^T u = 0 \iff u = 0 \]  \hspace{1cm} (9)
\[ (\alpha u + \beta v)^T w = \alpha u^T w + \beta v^T w \]

- We also use the dot product to define the 2-norm of vectors:
  \[ \|u\|_2 = \sqrt{u^T u}. \]  \hspace{1cm} (10)

- All these properties carry over when we replace the vectors with functions.

- An inner product is a function that associates a number \((f, g)\) with two continuous functions \(f\) and \(g\) defined on an interval \([a, b]\) such that these properties hold:
  \[(f, g) = (g, f)\]
  \[ (f, f) \geq 0 \]
  \[ (f, f) = 0 \iff f = 0 \]  \hspace{1cm} (11)
  \[ (\alpha f + \beta g, h) = \alpha (f, h) + \beta (g, h) \]

- moreover, we define the inner product norm
  \[ \|f\| = \sqrt{(f, f)}. \]  \hspace{1cm} (12)
• The requirement of continuity can be relaxed.

• As in the case of the standard inner products of vectors, we say that two functions $f$ and $g$ are orthogonal (with respect to the inner product $(\cdot, \cdot)$) if

$$ (f, g) = 0. \quad (13) $$
Examples of Inner Products

- For **vectors** there are possible inner products other than the dot product. For example if $A$ is symmetric and positive definite then

$$ (u, v) = u^T Av \quad (14) $$

$A = I \Rightarrow (u, v) = u^T v$

defines an inner product (exercise).

- For functions inner products can be used, for example, to give extra weight to the end points or other specific points, or to incorporate derivatives (provided the functions involved are differentiable).

- For example, the following define inner products (exercise):

$$(f, g)_1 = \int_{a}^{b} f(x)g(x)dx$$

$$(f, g)_2 = \int_{a}^{b} w(x)f(x)g(x)dx, \quad w > 0$$

$$(f, g)_3 = \int_{a}^{b} w(x)f(x)g(x) + \hat{w}(x)f'(x)g'(x)dx, \quad w > 0, \quad \hat{w} \geq 0$$

$$(f, g)_4 = \int_{a}^{b} w(x)f(x)g(x)dx + f(c)g(c), \quad w > 0$$

(15)

- Proceeding just as before one can show easily (exercise) that for any inner product $(\cdot, \cdot)$ the
solution of

\[ F(\alpha_1, \ldots \alpha_n) = (f - \sum_{i=1}^{n} \alpha_i \phi_i, f - \sum_{i=1}^{n} \alpha_i \phi_i) = \min \]  

(16)
is given by the solution of the linear system

\[ Ha = b \]  

(17)

where

\[ H = [(\phi_i, \phi_j)], \quad a = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}, \quad \text{and} \quad b = \begin{bmatrix} (f, \phi_1) \\ \vdots \\ (f, \phi_n) \end{bmatrix}. \]  

(18)

• The matrix \( H \) is obviously symmetric, and it is easy to see that it is positive definite: Let

\[ p = \sum_{i=1}^{n} x_i \phi_i \quad \text{and} \quad x = [x_1, \ldots, x_n]^T. \]  

(19)

Then

\[ x^T H x = \left( \sum_{i=1}^{n} x_i \phi_i, \sum_{i=1}^{n} x_i \phi_i \right) = (p, p) > 0. \]  

(20)

• Of course, it would be nice if the matrix \( H \) was diagonal, i.e., the basis functions \( \phi_i \) were orthogonal.
• We can always get an orthogonal basis by the **Gram-Schmidt Process**. It constructs basis functions

$$\psi_1, \psi_2, \ldots$$  \hspace{1cm} (21)

such that

$$\text{span} \{\phi_1, \ldots, \phi_k\} = \text{span} \{\psi_1, \ldots, \psi_k\}, \quad k = 1, 2, \ldots$$  \hspace{1cm} (22)

and

$$(\psi_i, \psi_j) = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$  \hspace{1cm} (23)

• as we discussed in the past (when discussing the $QR$ factorization) it proceeds as follows:

1. Let

$$\psi_1 = \frac{\phi_1}{\|\phi_1\|}.$$  \hspace{1cm} (24)

2. For $k = 2, 3, 4, \ldots$, let

$$\eta_k = \phi_k - \sum_{j=1}^{k-1} (\phi_k, \psi_j) \psi_j \quad (\eta_k, \psi_i) = (\phi_k, \psi_i) -$$

$$\psi_k = \frac{\eta_k}{\|\eta_k\|}.$$

• For the special case that

$$\phi_k(x) = x^{k-1}$$  \hspace{1cm} (26)

and

$$(f, g) = \int_a^b w(x)f(x)g(x)dx$$  \hspace{1cm} (27)
the Gram Schmidt algorithm can be modified and simplified.


• Suppose the inner product $(\cdot, \cdot)$ is defined by (27).

• **Theorem:** The sequence of polynomials defined inductively in the following way is orthogonal:

\[
Q_n = (x - a_n)Q_{n-1} - b_nQ_{n-2}
\]  

(28)

with

\[
Q_0 = 1,
\]
\[
Q_1 = x - a_1,
\]
\[
a_n = \frac{(xQ_{n-1}, Q_{n-1})}{(Q_{n-1}, Q_{n-1})}
\]

(29)

and

\[
b_n = \frac{(xQ_{n-1}, Q_{n-2})}{(Q_{n-2}, Q_{n-2})}
\]

\[
= \int_a^b w(x) B_{n-i}(x) (x \phi_i(x)) \, dx
\]

• \(i \leq n-2, \deg x \phi_i \leq n-1\)

**Proof:** We see from the formulas that for each \(n\), \(Q_n\) is a polynomial of degree \(n\) with leading coefficient 1 and is therefore not zero. Hence
the denominators in the formulas for \( a_n \) and \( b_n \) are not zero. We now show by induction on \( n \) that

\[
(Q_n, Q_i) = 0 \quad \text{for} \quad i < n. \tag{30}
\]

For \( n = 0 \) there is nothing to show. For \( n = 1 \) we have

\[
(Q_1, Q_0) = (x - a_1 Q_0, Q_0) \\
= (xQ_0, Q_0) - a_1 (Q_0, Q_0) \\
= (xQ_0, Q_0) - \frac{(xQ_0, Q_0)}{(Q_0, Q_0)} (Q_0, Q_0) \\
= 0.
\]

(31)

Now assume that our assertion is true for \( n - 1 \). Then we have

\[
(Q_{n-1}, Q_i) = 0 \quad \text{for} \quad i < n - 1.
\]

Then we have

\[
(Q_n, Q_{n-1}) = ((x - a_n)Q_{n-1} - b_n Q_{n-2}, Q_{n-1}) \\
= ((xQ_{n-1}, Q_{n-1}) - a_n (Q_{n-1}, Q_{n-1}) - b_n \frac{(Q_{n-2}, Q_{n-1})}{(Q_{n-1}, Q_{n-1})} (Q_{n-1}, Q_{n-1}) = 0 \\
= (xQ_{n-1}, Q_{n-1}) - \frac{(xQ_{n-1}, Q_{n-1})}{(Q_{n-1}, Q_{n-1})} (Q_{n-1}, Q_{n-1}) \\
= 0
\]

(32)
Similarly,

\[(Q_n, Q_{n-2}) = ((x - a_n)Q_{n-1} - b_nQ_{n-2}, Q_{n-2})
\]

\[= (xQ_{n-1}, Q_{n-2}) - a_n\left(\frac{Q_{n-1}, Q_{n-2}}{Q_{n-2}, Q_{n-2}}\right) - b_n\left(\frac{Q_{n-2}, Q_{n-2}}{Q_{n-2}, Q_{n-2}}\right) = 0\]

\[= (xQ_{n-1}, Q_{n-2}) - \left(\frac{xQ_{n-1}, Q_{n-2}}{Q_{n-2}, Q_{n-2}}\right)(Q_{n-2}, Q_{n-2}) = 0\]  \hspace{1cm} (33)

Now, if

\[i < n - 2\]  \hspace{1cm} (34)

we have

\[(Q_n, Q_i) = ((x - a_n)Q_{n-1} - b_nQ_{n-2}, Q_i)
\]

\[= (xQ_{n-1}, Q_i) - a_n\left(\frac{Q_{n-1}, Q_i}{Q_{n-2}, Q_i}\right) - b_n\left(\frac{Q_{n-2}, Q_i}{Q_{n-2}, Q_i}\right) = 0\]

\[= (xQ_{n-1}, Q_i) = (Q_{n-1}, xQ_i)\]

\[\times Q_i = \sum_{j=0}^{n-2} z_i(Q_i(x))\]  \hspace{1cm} (35)

Here we have used the fact that for the inner product defined by (27) we have the additional property that

\[(xf, g) = (f, xg)\]

The polynomial \(xQ_i\) is a polynomial of degree

\[i + 1 < n - 1\]  \hspace{1cm} (37)
which can be written as a linear combination of $Q_0, \ldots, Q_{i+1}$. By the induction hypothesis, $Q_{n-1}$ is orthogonal to all of these and so

$$(Q_n, Q_i) = (Q_{n-1}, xQ_i) = 0. \quad (38)$$

We used the crucial property (36) which is not part of the definition of an inner product. It turns out that in general if (36) does not hold then the recurrence relation (28) does not hold.

Can you come up with an inner product for which (36) does not hold?