Numerical Analysis Hilary Term 2018
Lecture 11: Least-Squares Approximation

For the problem of least-squares approximation, $\langle f,g\rangle = \int_a^b w(x)f(x)g(x)\,\mathrm{d}x$ and $\|f\|_2^2 = \langle f,f\rangle$ where $w(x)>0$ on (a,b).

Theorem. If $f\in \mathrm{L}^2_w(a,b)$ and $p_n\in\Pi_n$ is such that

$$\langle f-p_n, r\rangle = 0 \qquad \forall r\in\Pi_n, \tag{1}$$

then

$$\|f-p_n\|_2 \le \|f-r\|_2 \qquad \forall r\in\Pi_n,$$

i.e., p_n is a best (weighted) least-squares approximation to f on $[a,b]$.

Proof.

$$\begin{aligned}
\|f-p_n\|_2^2 &= \langle f-p_n, f-p_n\rangle \\
&= \langle f-p_n, f-r\rangle + \langle f-p_n, r-p_n\rangle \qquad \forall r\in\Pi_n \\
&\quad \text{Since } r-p_n\in\Pi_n \text{ the assumption (1) implies that} \\
&= \langle f-p_n, f-r\rangle \\
&\le \|f-p_n\|_2\|f-r\|_2 \ \text{ by the Cauchy–Schwarz inequality.}
\end{aligned}$$

Dividing both sides by $\|f-p_n\|_2$ gives the required result. □

Remark: the converse is true too (see Problem Sheet 6, Q9).

This gives a direct way to calculate a best approximation: we want to find $p_n(x)=\sum_{k=0}^n \alpha_k x^k$ such that

$$\int_a^b w(x)\left(f-\sum_{k=0}^n \alpha_k x^k\right)x^i\,\mathrm{d}x = 0 \ \text{ for } \ i=0,1,\ldots,n. \tag{2}$$

[Note that (2) holds if, and only if,

$$\int_a^b w(x)\left(f-\sum_{k=0}^n \alpha_k x^k\right)\left(\sum_{i=0}^n \beta_i x^i\right)\mathrm{d}x = 0 \qquad \forall q=\sum_{i=0}^n \beta_i x^i\in\Pi_n.]$$

However, (2) implies that

$$\sum_{k=0}^n\left(\int_a^b w(x)x^{k+i}\,\mathrm{d}x\right)\alpha_k = \int_a^b w(x)f(x)x^i\,\mathrm{d}x \ \text{ for } \ i=0,1,\ldots,n$$

which is the component-wise statement of a matrix equation

$$A\alpha=\varphi, \tag{3}$$

to determine the coefficients $\alpha=(\alpha_0,\alpha_1,\ldots,\alpha_n)^{\mathrm{T}}$, where $A=\{a_{i,k}, i,k=0,1,\ldots,n\}$, $\varphi=(f_0,f_1,\ldots,f_n)^{\mathrm{T}}$,

$$a_{i,k}=\int_a^b w(x)x^{k+i}\,\mathrm{d}x \ \text{ and } \ f_i=\int_a^b w(x)f(x)x^i\,\mathrm{d}x.$$

The system (3) are called the **normal equations**.

Example: the best least-squares approximation to e^x on $[0,1]$ from Π_1 in $\langle f,g\rangle = \int_a^b f(x)g(x)\,dx$. We want

$$\int_0^1 [e^x - (\alpha_0 1 + \alpha_1 x)]1\,dx = 0 \text{ and } \int_0^1 [e^x - (\alpha_0 1 + \alpha_1 x)]x\,dx = 0.$$

$\iff$

$$\begin{aligned} \alpha_0 \int_0^1 dx + \alpha_1 \int_0^1 x\,dx &= \int_0^1 e^x\,dx \\ \alpha_0 \int_0^1 x\,dx + \alpha_1 \int_0^1 x^2\,dx &= \int_0^1 e^x x\,dx \end{aligned}$$

i.e.,

$$\begin{bmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} e-1 \\ 1 \end{bmatrix}$$

$\implies \alpha_0 = 4e - 10$ and $\alpha_1 = 18 - 6e$, so $p_1(x) := (18-6e)x + (4e-10)$ is the best approximation.

Proof that the coefficient matrix A is nonsingular will now establish existence and uniqueness of (weighted) $\|\cdot\|_2$ best-approximation.

Theorem. The coefficient matrix A is nonsingular.

Proof. Suppose not $\implies \exists\alpha \neq 0$ with $A\alpha = 0 \implies \alpha^T A\alpha = 0$

$$\iff \sum_{i=0}^{n} \alpha_i (A\alpha)_i = 0 \iff \sum_{i=0}^{n} \alpha_i \sum_{k=0}^{n} a_{ik}\alpha_k = 0,$$

and using the definition $a_{ik} = \int_a^b w(x)x^k x^i\,dx$,

$$\iff \sum_{i=0}^{n} \alpha_i \sum_{k=0}^{n} \left(\int_a^b w(x)x^k x^i\,dx \right) \alpha_k = 0.$$

Rearranging gives

$$\int_a^b w(x) \left(\sum_{i=0}^{n} \alpha_i x^i \right) \left(\sum_{k=0}^{n} \alpha_k x^k \right) dx = 0 \text{ or } \int_a^b w(x) \left(\sum_{i=0}^{n} \alpha_i x^i \right)^2 dx = 0$$

which implies that $\sum_{i=0}^{n} \alpha_i x^i = 0$ and thus $\alpha_i = 0$ for $i = 0, 1, \ldots, n$. This contradicts the initial supposition, and thus A is nonsingular. □

Remark: This result does not imply that the normal equations are usable in practice: the method would need to be stable with respect to small perturbations. In fact, difficulties arise from the "ill-conditioning" of the matrix A as n increases. The next lecture looks at a fix.