Numerical Analysis Hilary Term 2023
Lecture 9: Best Approximation in Inner-Product Spaces

Best approximation of functions: given a function f on $[a, b]$, find the "closest" polynomial/piecewise polynomial (see later sections)/ trigonometric polynomial (truncated Fourier series).

Norms: are used to measure the size of/distance between elements of a vector space. Given a vector space V over the field $\mathbb{R}$ of real numbers, the mapping $\|\cdot\| : V \to \mathbb{R}$ is a **norm** on V if it satisfies the following axioms:
(i) $\|f\| \geq 0$ for all $f \in V$, with $\|f\| = 0$ if, and only if, $f = 0 \in V$;
(ii) $\|\lambda f\| = |\lambda|\|f\|$ for all $\lambda \in \mathbb{R}$ and all $f \in V$; and
(iii) $\|f + g\| \leq \|f\| + \|g\|$ for all $f, g \in V$ (the **triangle inequality**).

Examples: 1. For vectors $x \in \mathbb{R}^n$, with $x = (x_1, x_2, \ldots, x_n)^{\mathrm{T}}$,

$$\|x\| \equiv \|x\|_2 = (x_1^2 + x_2^2 + \cdots + x_n^2)^{\frac{1}{2}} = \sqrt{x^{\mathrm{T}} x}$$

is the ℓ^2- or vector two-norm.
2. For continuous functions on $[a, b]$,

$$\|f\| \equiv \|f\|_\infty = \max_{x \in [a,b]} |f(x)|$$

is the L^∞- or ∞-norm.
3. For integrable functions on (a, b),

$$\|f\| \equiv \|f\|_1 = \int_a^b |f(x)| \, \mathrm{d}x$$

is the L^1- or one-norm.
4. For functions in

$$V = \mathrm{L}^2_w(a, b) \equiv \{f : [a, b] \to \mathbb{R} \mid \int_a^b w(x)[f(x)]^2 \, \mathrm{d}x < \infty\}$$

for some given **weight** function $w(x) > 0$ (this certainly includes continuous functions on $[a, b]$, and piecewise continuous functions on $[a, b]$ with a finite number of jump-discontinuities),

$$\|f\| \equiv \|f\|_2 = \left(\int_a^b w(x)[f(x)]^2 \, \mathrm{d}x\right)^{\frac{1}{2}}$$

is the L^2- or two-norm—the space $\mathrm{L}^2(a, b)$ is a common abbreviation for $\mathrm{L}^2_w(a, b)$ for the case $w(x) \equiv 1$.
Note: $\|f\|_2 = 0 \implies f = 0$ almost everywhere on $[a, b]$. We say that a certain property P holds *almost everywhere* (a.e.) on $[a, b]$ if property P holds at each point of $[a, b]$ except perhaps on a subset $S \subset [a, b]$ of zero measure. We say that a set $S \subset \mathbb{R}$ has *zero measure* (or that it is of *measure zero*) if for any $\varepsilon > 0$ there exists a sequence $\{(\alpha_i, \beta_i)\}_{i=1}^\infty$ of subintervals of $\mathbb{R}$ such that

$S \subset \cup_{i=1}^{\infty}(\alpha_i, \beta_i)$ and $\sum_{i=1}^{\infty}(\beta_i - \alpha_i) < \varepsilon$. Trivially, the empty set $\emptyset(\subset \mathbb{R})$ has zero measure. Any finite subset of $\mathbb{R}$ has zero measure. Any countable subset of $\mathbb{R}$, such as the set of all natural numbers $\mathbb{N}$, the set of all integers $\mathbb{Z}$, or the set of all rational numbers $\mathbb{Q}$, is of measure zero.

Least-squares polynomial approximation: aim to find the best polynomial approximation to $f \in \mathrm{L}^2_w(a,b)$, i.e., find $p_n \in \Pi_n$ for which

$$\|f - p_n\|_2 \leq \|f - q\|_2 \qquad \forall q \in \Pi_n.$$

Seeking p_n in the form $p_n(x) = \displaystyle\sum_{k=0}^{n} \alpha_k x^k$ then results in the minimization problem

$$\min_{(\alpha_0,\ldots,\alpha_n)} \int_a^b w(x) \left[f(x) - \sum_{k=0}^{n} \alpha_k x^k \right]^2 \mathrm{d}x.$$

The unique minimizer can be found from the (linear) system

$$\frac{\partial}{\partial \alpha_j} \int_a^b w(x) \left[f(x) - \sum_{k=0}^{n} \alpha_k x^k \right]^2 \mathrm{d}x = 0 \;\text{ for each }\; j = 0, 1, \ldots, n,$$

but there is important additional structure here.

Inner-product spaces: a real **inner-product space** is a vector space V over $\mathbb{R}$ with a mapping $\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}$ (the **inner product**) for which

(i) $\langle v, v \rangle \geq 0$ for all $v \in V$ and $\langle v, v \rangle = 0$ if, and only if $v = 0$;

(ii) $\langle u, v \rangle = \langle v, u \rangle$ for all $u, v \in V$; and

(iii) $\langle \alpha u + \beta v, z \rangle = \alpha \langle u, z \rangle + \beta \langle v, z \rangle$ for all $u, v, z \in V$ and all $\alpha, \beta \in \mathbb{R}$.

Examples: **1.** $V = \mathbb{R}^n$,

$$\langle x, y \rangle = x^{\mathrm{T}} y = \sum_{i=1}^{n} x_i y_i,$$

where $x = (x_1, \ldots, x_n)^{\mathrm{T}}$ and $y = (y_1, \ldots, y_n)^{\mathrm{T}}$.

2. $V = \mathrm{L}^2_w(a,b) = \{f : (a,b) \to \mathbb{R} \mid \displaystyle\int_a^b w(x)[f(x)]^2 \,\mathrm{d}x < \infty\}$,

$$\langle f, g \rangle = \int_a^b w(x) f(x) g(x) \,\mathrm{d}x,$$

where $f, g \in \mathrm{L}^2_w(a,b)$ and w is a weight-function, defined, positive and integrable on (a,b).

Notes: **1.** Suppose that V is an inner product space, with inner product $\langle \cdot, \cdot \rangle$. Then $\langle v, v \rangle^{\frac{1}{2}}$ defines a norm on V (see the final paragraph on the last page for a proof). In Example 2 above, the norm defined by the inner product is the (weighted) L^2-norm.

2. Suppose that V is an inner product space, with inner product $\langle \cdot, \cdot \rangle$, and let $\|\cdot\|$ denote the norm defined by the inner product via $\|v\| = \langle v, v \rangle^{\frac{1}{2}}$, for $v \in V$. The angle θ between $u, v \in V$ is

$$\theta = \cos^{-1}\left(\frac{\langle u, v \rangle}{\|u\| \|v\|} \right).$$

Thus u and v are orthogonal in $V \iff \langle u, v\rangle = 0$.

E.g., x^2 and $\frac{3}{4} - x$ are orthogonal in $\mathrm{L}^2(0,1)$ with inner product $\langle f, g\rangle = \int_0^1 f(x)g(x)\,\mathrm{d}x$ as

$$\int_0^1 x^2\left(\tfrac{3}{4} - x\right)\,\mathrm{d}x = \tfrac{1}{4} - \tfrac{1}{4} = 0.$$

3. Pythagoras Theorem: Suppose that V is an inner-product space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$ defined by this inner product. For any $u, v \in V$ such that $\langle u, v\rangle = 0$ we have

$$\|u \pm v\|^2 = \|u\|^2 + \|v\|^2.$$

Proof.

$$\begin{aligned}
\|u \pm v\|^2 = {} & \langle u \pm v, u \pm v\rangle = \langle u, u \pm v\rangle \pm \langle v, u \pm v\rangle && \text{[axiom (iii)]} \\
= {} & \langle u, u \pm v\rangle \pm \langle u \pm v, v\rangle && \text{[axiom (ii)]} \\
= {} & \langle u, u\rangle \pm \langle u, v\rangle \pm \langle u, v\rangle + \langle v, v\rangle && \\
= {} & \langle u, u\rangle + \langle v, v\rangle && \text{[orthogonality]} \\
= {} & \|u\|^2 + \|v\|^2. &&
\end{aligned}$$

4. The **Cauchy–Schwarz inequality**: Suppose that V is an inner-product space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$ defined by this inner product. For any $u, v \in V$,

$$|\langle u, v\rangle| \le \|u\|\|v\|.$$

Proof. For every $\lambda \in \mathbb{R}$,

$$0 \le \langle u - \lambda v, u - \lambda v\rangle = \|u\|^2 - 2\lambda\langle u, v\rangle + \lambda^2\|v\|^2 = \phi(\lambda),$$

which is a quadratic in λ. The minimizer of ϕ is at $\lambda_* = \langle u, v\rangle/\|v\|^2$, and thus since $\phi(\lambda_*) \ge 0$, $\|u\|^2 - \langle u, v\rangle^2/\|v\|^2 \ge 0$, which gives the required inequality. □

5. The **triangle inequality**: Suppose that V is an inner-product space with inner product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$ defined by this inner product. For any $u, v \in V$,

$$\|u + v\| \le \|u\| + \|v\|.$$

Proof. Note that

$$\|u + v\|^2 = \langle u + v, u + v\rangle = \|u\|^2 + 2\langle u, v\rangle + \|v\|^2.$$

Hence, by the Cauchy–Schwarz inequality,

$$\|u + v\|^2 \le \|u\|^2 + 2\|u\|\|v\| + \|v\|^2 = (\|u\| + \|v\|)^2.$$

Taking square-roots yields

$$\|u + v\| \le \|u\| + \|v\|.$$

□

Note: The function $\|\cdot\| : V \to \mathbb{R}$ defined by $\|v\| := \langle v, v\rangle^{\frac{1}{2}}$ on the inner-product space V, with inner product $\langle\cdot,\cdot\rangle$, trivially satisfies the first two axioms of norm on V; this is a consequence of $\langle\cdot,\cdot\rangle$ being an inner product on V. Result 5 above implies that $\|\cdot\|$ also satisfies the third axiom of norm, the triangle inequality.

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.

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

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

Remark: the converse is true too (see problem sheet 3).

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 $\langle f,g\rangle = \int_a^b f(x)g(x)\,\mathrm{d}x$. We want

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

$\Longleftrightarrow$

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

i.e.,

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

$\Longrightarrow$ $\alpha_0 = 4\mathrm{e} - 10$ and $\alpha_1 = 18 - 6\mathrm{e}$, so $p_1(x) := (18-6\mathrm{e})x + (4\mathrm{e}-10)$ 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 $\Longrightarrow \exists \alpha \neq 0$ with $A\alpha = 0 \Longrightarrow \alpha^{\mathrm{T}} A \alpha = 0$

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

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

$$\Longleftrightarrow \quad \sum_{i=0}^n \alpha_i \sum_{k=0}^n \left(\int_a^b w(x) x^k x^i\,\mathrm{d}x \right) \alpha_k = 0.$$

Rearranging gives

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

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:

- Note in the simplest least-squares approximation problem $\min_x \|Ax - b\|_2$ that we dealt with in lecture 4, the theorem gives the solution $A^T(Ax - b) = 0$, that is, $x = (A^T A)^{-1} A^T b$. This coincides with the QR-based solution derived in lecture 4.
- The above theorem 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.