Numerical Analysis Hilary Term 2025
Lecture 9: Best Approximation in Inner-Product Spaces

Best approximation: Given an element f of an inner-product space, we aim to find the "closest"/"best" approximation to f in a finite-dimensional subspace.

Norm: 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:

- For vectors $x \in \mathbb{R}^n$, with $x = (x_1, x_2, \ldots, x_n)^\top$,

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

  is the ℓ^2- or vector two-norm.

- For continuous functions on $[a, b]$,

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

  is the L^∞- or ∞-norm.

- For integrable functions on (a, b),

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

  is the L^1- or one-norm.

- For functions in

$$L_w^2(a, b) := \left\{ f : [a, b] \to \mathbb{R} : \int_a^b w(x)[f(x)]^2 \, \mathrm{d}x < \infty \right\}$$

  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),

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

  is the L^2- or two-norm—the space $L^2(a, b)$ is a common abbreviation for $L_w^2(a, b)$ 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.

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:

- $V = \mathbb{R}^n$,

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

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

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

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

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

Notes:

- 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, as we will show in the next section. We will abuse notation and refer to $\| \cdot \| := \langle \cdot, \cdot \rangle^{\frac{1}{2}}$ as the norm induced by $\langle \cdot, \cdot \rangle$ before we actually prove that $\| \cdot \|$ is a norm. In second example above, the norm defined by the inner product is the (weighted) L^2-norm.

- 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 $L^2(0,1)$ with inner product $\langle f, g \rangle = \int_0^1 f(x)g(x)\,\mathrm{d}x$ as

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

Properties of inner-product spaces

Theorem. (**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. Let $u, v \in V$. Then, there holds

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

Theorem. (**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 $u, v \in V$,

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

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

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

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

Theorem. (**triangle inequality**) Suppose that V is an inner-product space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$ induced by this inner product. For any $u, v \in V$, there holds

$$\|u + v\| \leq \|u\| + \|v\|.$$

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,

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

Taking square-roots yields $\|u + v\| \leq \|u\| + \|v\|$. □

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. The above result implies that $\|\cdot\|$ also satisfies the third axiom of norms, the triangle inequality.

Least-Squares Approximation
We now consider the abstract least squares approximation problem. Let V be an inner product space and let $W\subset V$ be a finite-dimensional subspace of V. The least squares approximation problem is: Given $f\in V$, find $p\in W$ such that

$$\|f-p\|\le\|f-r\|\qquad\forall r\in W,\tag{1}$$

where $\|\cdot\|$ is the norm induced by the inner-product $\langle\cdot,\cdot\rangle$ on V.

Example: Polynomial Approximation. An important example of best approximation is polynomial approximations to functions in $L^2_w(a,b)$. That is, given $f\in L^2_w(a,b)$, we seek $p_n\in\Pi_n$ for which

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

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

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

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

$$\frac{\partial}{\partial\alpha_j}\int_a^b w(x)\left|f(x)-\sum_{k=0}^n\alpha_kx^k\right|^2\,\mathrm{d}x=0\qquad\text{for each}j=0,1,\dots,n.$$

However, we seek an alternative approach that exploits the inner-product structure on $L^2_w(a,b)$.

Theorem. If $f\in V$ and $p\in W$ is such that

$$\langle f-p,r\rangle=0\qquad\forall r\in W,\tag{2}$$

then p satisfies (1); i.e., p is a best least-squares approximation to f in W.

Proof. Let $f\in V$ and suppose that $p\in W$ satisfies (2). Then, for all $r\in W$, there holds

$$\begin{aligned}\|f-p\|^2&=\langle f-p,f-p\rangle\\&=\langle f-p,f-r\rangle+\langle f-p,r-p\rangle\\&=\langle f-p,f-r\rangle&&[\text{by (2) since } r-p\in W]\\&\le\|f-p\|\|f-r\|&&[\text{by the Cauchy–Schwarz inequality}].\end{aligned}$$

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

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

Matrix form of least-squares approximation

Condition (2) give a direct way to calculate a best approximation. Let $n = \dim W$ and let $\{\phi_k\}_{k=1}^n$ be a basis for W. Then, we want to find $p = \sum_{k=1}^n \alpha_k \phi_k$ such that

$$\left\langle f - \sum_{k=0}^{n} \alpha_k \phi_k, \sum_{i=0}^{n} \beta_i \phi_i \right\rangle = 0 \qquad \forall \beta_1, \beta_2, \ldots, \beta_n \in \mathbb{R},$$

which holds if and only if

$$\left\langle f - \sum_{k=1}^{n} \alpha_k \phi_k, \phi_i \right\rangle = 0 \qquad \text{for } i = 1, 2, \ldots, n. \tag{3}$$

However, (3) implies that

$$\sum_{k=0}^{n} \langle \phi_k, \phi_i \rangle \alpha_k = \langle f, \phi_i \rangle \qquad \text{for } i = 1, 2, \ldots, n,$$

which is the component-wise statement of a matrix equation

$$A\alpha = \varphi, \tag{4}$$

to determine the coefficients $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)^\top$, where $A = \{a_{i,k}, i, k = 1, 2, \ldots, n\}$, $\varphi = (f_1, f_2, \ldots, f_n)^\top$,

$$a_{i,k} = \langle \phi_k, \phi_i \rangle \quad \text{and} \quad f_i = \langle f, \phi_i \rangle .$$

The system (4) is called the **normal equations**.

Theorem. The coefficient matrix A is nonsingular.

Proof. Suppose that A is singular. Then, there exists $\alpha \neq 0$ with $A\alpha = 0$, and so $\alpha^\top A\alpha = 0$. In component form, this reads

$$0 = \sum_{i=1}^{n} \alpha_i (A\alpha)_i = 0 = \sum_{i=1}^{n} \alpha_i \sum_{k=0}^{n} a_{ik} \alpha_k,$$

and using the definition $a_{ik} = \langle \phi_k, \phi_i \rangle$, we obtain

$$\sum_{i,k=1}^{n} \alpha_i \langle \phi_k, \phi_i \rangle \alpha_k = 0.$$

Rearranging gives

$$0 = \left\langle \sum_{i=1}^{n} \alpha_i \phi_i, \sum_{k=1}^{n} \alpha_k \phi_k \right\rangle = \left\| \sum_{i=1}^{n} \alpha_i \phi_i \right\|^2$$

which implies that $\sum_{i=1}^n \alpha_i x^i = 0$ and thus $\alpha_i = 0$ for $i = 1, 2, \ldots, n$. This contradicts the initial supposition, and thus A is nonsingular. $\square$

Examples

Approximation of functions in $L^2_w(a, b)$: Recall that given $f \in L^2_w(a, b)$, we seek $p_n \in \Pi_n$ for which

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

We will use monomials as a basis for Π_n: $\phi_i = x^i$ for $i = 0, 1, \ldots, n$ (note that, for convenience, we are indexing from 0 rather than 1 for the basis). Then, the orthogonality conditions (3) read

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

or equivalently

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

Consider 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

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

which holds if and only if

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

In matrix form, we obtain

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

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

Linear least squares from lecture 4: Recall the following least squares problem from lecture 4. Given $A \in \mathbb{R}^{m \times n}$ with $m \geq n$ and $b \in \mathbb{R}^m$, find the minimizer $x \in \mathbb{R}^n$ to

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2. \tag{5}$$

As before, we assume that A has full rank. We can recast this problem as a best approximation problem. Let $V = \mathbb{R}^m$ be equipped with the standard inner product $\langle u, v \rangle = u^\top v$ and $W = \mathrm{col}(A) \subset \mathbb{R}^m$, the column space of A. Then, problem (5) is equivalent to

$$\min_{y \in \mathrm{col}(A)} \|y - b\|_2.$$

Since A has full rank, the columns of A, denoted $a_1, a_2, \ldots, a_n$, form a basis for $\mathrm{col}(A)$. Choosing $\phi_i = a_i$, we have $y = \sum_{i=1}^n x_i a_i$ and (3) reads

$$\left(\sum_{k=1}^n x_k a_k \right)^\top a_i = b^\top a_i, \qquad i = 1, 2, \ldots, n.$$

In matrix form (4), we have

$$A^\top A x = A^\top b \implies x = (A^\top A)^{-1} A^\top b.$$

If $A = QR$ is the thin QR factorization of A, then

$$(A^\top A)^{-1} A^\top b = (R^\top Q^\top Q R)^{-1} R^\top Q^\top b = (R^\top R)^{-1} R^\top Q^\top b = R^{-1} R^{-\top} R^\top Q^\top b = R^{-1} Q^\top b,$$

which is the same solution derived in lecture 4.

Remark: The above results do 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.