

Ch 8.2: Continuous Function Approximation, "L2"

Thursday, October 23, 2025 11:09 AM

Let's look at the 2nd scenario: rather than fitting $\hat{y}_i \approx y_i$, $i=1, \dots, m$
 we instead want to approximate a function f with
 another (simpler or nicer) function p , so that $f(x) \approx p(x)$
 $\forall x \in [a, b]$

As in the discrete case, we have options for how to make $f(x) \approx p(x)$ precise:

① (default) L^2 (continuous least squares) $E = \|f - p\|_2^2 := \int_a^b |f(x) - p(x)|^2 dx$
 or just write "2"
 ↪ a, b inferred from context

② minimax $E = \|f - p\|_\infty := \sup_{x \in [a, b]} |f(x) - p(x)|$

③ L^1 $E = \|f - p\|_1 := \int_a^b |f(x) - p(x)| dx$

We'll focus on ① L^2 as it's nicest in some ways

What kind of function is $p(x)$?

We'll write $p(x) = \sum_{k=0}^n a_k \cdot \phi_k(x)$

for some set of linearly independent functions $\{\phi_k\}_{k=0}^n$
 we discussed 2 weeks ago or so

Aside: If $f \in C([a, b])$,
 $\forall \varepsilon > 0 \exists$ a polynomial p

st. $\|f - p\|_\infty < \varepsilon$

"Weierstrass Approximation Thm"
 You can construct an explicit polynomial using Bernstein polynomials

While our loss function seems continuous or " ∞ -dimensional",
 what actually matters is that $n < \infty$. Our parameters we're
 looking for α 's the vector $\vec{\alpha} \in \mathbb{R}^{n+1}$, very finite dimensional.

Recall (from APPM 3310) the notion of an inner product \checkmark vectorspace

over \mathbb{R} (not \mathbb{C}), it is any bilinear function $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$

s.t. ① $\forall f, g \in V$, $\langle f, g \rangle = \langle g, f \rangle$ Symmetric

② $\forall f, g, h \in V$, $\forall \alpha \in \mathbb{R}$, $\langle f + \alpha g, h \rangle = \langle f, h \rangle + \alpha \langle g, h \rangle$ linear in 1st argument

③ $\forall f \in V$, $\langle f, f \rangle \geq 0$ and $\langle f, f \rangle = 0$ iff $f = \vec{0}$

Δ Notation warning: \vec{a}, \vec{b} are vectors w/ entries a_j, b_j . Not to be confused with our domain $[a, b]$. Sorry.

and any inner product induces a norm, $\|f\| = \sqrt{\langle f, f \rangle}$

In our case, $\langle f, g \rangle := \int_a^b f(x)g(x)dx$ so $\|f\|_2 = \sqrt{\int_a^b f(x)^2 dx}$

$$\begin{aligned} \text{So, } \|f - p\|_2^2 &= \langle f - p, f - p \rangle \\ &= \left\langle f - \sum_{k=0}^n a_k \phi_k, f - \sum_{j=0}^m a_j \phi_j \right\rangle \\ &= \langle f, f \rangle - 2 \langle f, \sum_{j=0}^m a_j \phi_j \rangle + \left\langle \sum_{k=0}^n a_k \phi_k, \sum_{j=0}^m a_j \phi_j \right\rangle \\ &= \|f\|_2^2 - 2 \sum_{j=0}^m a_j \underbrace{\langle f, \phi_j \rangle}_{b_j} + \sum_{k=0}^n \sum_{j=0}^m a_k a_j \underbrace{\langle \phi_k, \phi_j \rangle}_{G_{kj}} \\ &= \|f\|_2^2 - 2 \vec{b}^T \vec{a} + \vec{a}^T G \vec{a} \end{aligned}$$

$$\text{So } \min_{\vec{a} \in \mathbb{R}^{n+1}} E(\vec{a}) = \min_{\vec{a} \in \mathbb{R}^{n+1}} \|f\|_2^2 - 2 \vec{b}^T \vec{a} + \vec{a}^T G \vec{a}$$

Gram matrix.
Guaranteed to be positive definite if $\{\phi_j\}$ lin. independent

$$\text{so } \nabla E(\vec{a}) = -2 \vec{b} + 2 G \vec{a}$$

solve $\nabla E(\vec{a}) = 0$ means solve $G \vec{a} = \vec{b}$ for \vec{a} .
Finite dimensional.

The "no-dimensional part"

is calculating \vec{b} , \vec{G} in the 1st place: $b_j = \langle f, \phi_j \rangle = \int f(x) \phi_j(x) dx$

$$G_{kj} = \langle \phi_k, \phi_j \rangle = \int \phi_k(x) \phi_j(x) dx$$

This part you do by hand, not on a computer.

For a fixed basis $\{\phi_j\}_{j=0}^n$ and interval $[a, b]$, you can

compute G (i.e. it doesn't depend on f)

Ex: $n=2$, $\phi_0(x)=1$, $\phi_1(x)=x$, $\phi_2(x)=x^2$ so $\text{span}(\{\phi_j\}) = P_2$ monomial basis polynomials of degree 2 or less

$$G_{kj} = \langle x^k, x^j \rangle = \int_a^b x^{k+j} dx = \frac{b^{j+k+1} - a^{j+k+1}}{j+k+1} \quad \left. \begin{array}{l} \text{the Hilbert matrix} \\ (\text{notoriously ill-conditioned}) \end{array} \right\}$$

then if we have a function f , we can solve.

Ex: $f(x) = \sin(\pi \cdot x)$. Let's say $a=0, b=1$

$$\begin{aligned} b_j &= \int_0^1 \sin(\pi \cdot x) \cdot x^j dx = \begin{cases} \frac{2}{\pi} & j=0 \\ \frac{\pi}{2} & j=1 \\ \frac{\pi^2 - 4}{\pi^3} & j=2 \end{cases} \\ \text{so } \vec{a} &= \begin{pmatrix} 1 & \frac{2}{\pi} & \frac{\pi^2 - 4}{\pi^3} \\ \frac{1}{2} & \frac{\pi}{2} & \frac{\pi^2 - 4}{\pi^3} \\ \frac{1}{3} & \dots & \frac{\pi^2 - 4}{\pi^3} \end{pmatrix}^{-1} \begin{pmatrix} \frac{2}{\pi} \\ \frac{\pi}{2} \\ \frac{\pi^2 - 4}{\pi^3} \end{pmatrix}. \end{aligned}$$

$$\text{eg. } a=0, b=1 \text{ then } G = \begin{pmatrix} 1 & \frac{2}{\pi} & \frac{\pi^2 - 4}{\pi^3} \\ \frac{1}{2} & \frac{\pi}{2} & \frac{\pi^2 - 4}{\pi^3} \\ \frac{1}{3} & \dots & \frac{\pi^2 - 4}{\pi^3} \end{pmatrix}$$

Nicer bases

\hookrightarrow polynomials of degree $\leq n$

If we want $p \in P_n$, any basis $\{\phi_k\}_{k=0}^n$ works in principle,
but some are nicer than others, i.e. leading to faster + more stable computation.

In particular, if we have an **orthogonal basis** then $\langle \phi_k, \phi_j \rangle = 0$ if $k \neq j$
so G is a **diagonal matrix** (hence G^{-1} is easy)

What's an orthogonal basis for P_n ? let $[a, b] = [-1, 1]$

well, $\{\phi_k(x) = x^k\}_{k=0}^n$ is a basis, so we can orthogonalize it

$$\text{via Gram-Schmidt: } \tilde{\phi}_0 = \phi_0 = 1$$

$$\tilde{\phi}_1 = \phi_1 - \frac{\langle \phi_1, \tilde{\phi}_0 \rangle}{\|\tilde{\phi}_0\|^2} \cdot \tilde{\phi}_0 = x$$

$$\tilde{\phi}_2 = \phi_2 - \frac{\langle \phi_2, \tilde{\phi}_1 \rangle}{\|\tilde{\phi}_1\|^2} \tilde{\phi}_1 - \frac{\langle \phi_2, \tilde{\phi}_0 \rangle}{\|\tilde{\phi}_0\|^2} \tilde{\phi}_0 = x^2 - \gamma_3$$

etc.

$\{1, x, x^2 - \gamma_3, \dots\}$ are the **Legendre polynomials**

and orthogonal on $[-1, 1]$

(Often normalized in different ways, e.g. $\|\tilde{\phi}_k\| = 1$ or $\tilde{\phi}_k(1) = 1$ etc.)

Well-studied, easy to generate

In practice, use the 3-term recurrence:

$$\phi_{k+1}(x) = (x - \beta_k) \phi_k(x) - \gamma_k \phi_{k-1}(x)$$

(define $\phi_{-1} = 0$
 $\phi_0 = 1$)

$$\beta_k = \frac{\langle x \cdot \phi_k, \phi_k \rangle}{\|\phi_k\|^2} \quad \gamma_k = \frac{\|\phi_k\|^2}{\|\phi_{k-1}\|^2}$$

If $\{\phi_k\}$ orthogonal

and letting $\alpha_k = \|\phi_k\|^2$, then our L^2 approximation

of $f \in C[a, b]$ over $\text{span}(\{\phi_k\})$ is

$$p(x) = \sum_{j=0}^n a_j \phi_j(x) \quad \text{with} \quad a_j = \frac{\langle \phi_j, f \rangle}{\alpha_j}.$$

Standard Legendre polynomials are orthogonal on the domain $[-1, 1]$

$$\text{ex: } \phi_0(x) = 1, \phi_1(x) = x, \int_{-1}^1 1 \cdot x \, dx = 0 \quad \checkmark$$

If you want a domain $[a, b]$, you must adjust the Legendre polynomials accordingly (since, ex: $a=0, b=2$, $\int_0^2 1 \cdot x \, dx \neq 0$)

You don't have to recompute, just adjust them:

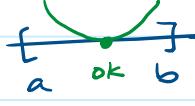
$$(k+j) \quad 0 = \int_{-1}^1 \phi_k(x) \phi_j(x) \, dx = \int_a^b \phi_k\left(\frac{2y-a+b}{b-a}\right) \phi_j\left(\frac{2y-a+b}{b-a}\right) \frac{2}{b-a} dy$$

change of variables: $y = \frac{b-a}{2}x + \frac{a+b}{2}$ so $x = -1 \Rightarrow y = a$
 $\text{so } dy = \frac{(b-a)}{2} dx \quad x = +1 \Rightarrow y = b$

$$\text{i.e. } x = \frac{2y}{b-a} - \frac{a+b}{b-a}$$

... so if $\tilde{\phi}_k(y) = \phi_k\left(\frac{2y}{b-a} - \frac{a+b}{b-a}\right)$ then $\{\phi_k\}$ orthog on $[-1, 1]$
means $\{\tilde{\phi}_k\}$ orthog. on $[a, b]$.

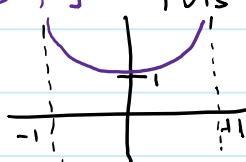
Weights and Chebyshev polynomials

We say $w(x)$ is a valid weight function on $[a, b]$ if it's integrable (a sufficient condition is being continuous) and it's non-negative $w(x) \geq 0$ and it's not 0 on any subinterval ( )
 $w(x)=0$ at a single point or two is ok

Then define $\langle f, g \rangle_w = \int_a^b f(x)g(x) w(x) \, dx, \|f\|_w = \sqrt{\langle f, f \rangle_w}$

and it's a valid inner product on $C[a, b]$ so all our theory still holds!

Main example $w(x) = \frac{1}{\sqrt{1-x^2}}$ on $[-1, 1]$ Puts more weight near boundaries



The Legendre polynomials are orthogonal for $w(x) = 1$ but not orthogonal for $w(x) = \frac{1}{\sqrt{1-x^2}}$. What is?

The Chebyshev polynomials (of the 1st kind), $\{T_k\}$

Very well-studied

$$T_0(x) = 1$$

$$T_2(x) = 2x^2 - 1$$

$$T_1(x) = x$$

$$T_3(x) = 4x^3 - 3x$$

$$T_{k+1}(x) = 2 \cdot x \cdot T_k(x) - T_{k-1}(x) \quad \text{"3-term recurrence"}$$

OR

$$T_k(\underbrace{\cos \theta}_x) = \cos(k \cdot \theta) \quad x = \cos^{-1}(\theta)$$

$$\text{i.e. } T_k(x) = \cos(k \cdot \cos^{-1}(x))$$

$$\text{so } k=0, T_0(x) = \cos(0) = 1 \quad \checkmark$$

$$T_1(x) = \cos(\cos^{-1}(x)) = x \quad \checkmark$$

To do a weighted least-squares fit...

$$\min_{\vec{a} \in \mathbb{R}^{n+1}} \|f - \sum_{k=0}^n a_k \phi_k\|_w^2 \quad \text{we have } G \cdot \vec{a} = \vec{b} \text{ with}$$

$$G_{kj} = \langle \phi_k, \phi_j \rangle_w, \quad b_j = \langle f, \phi_j \rangle_w$$

so if $\{\phi_k\}$ is orthogonal

on our domain, G is a diagonal matrix and $a_k = \frac{\int_a^b f(x) \phi_k(x) w(x) dx}{\int_a^b \phi_k^2(x) w(x) dx}$

Chebyshev polynomials are great

($n \geq 1$ since T_0 has no roots)

(Thm 8.9) Thm: The roots of $T_n(x)$ are all simple, in $[-1, 1]$ and

we have a formula for them: $r_k = \cos\left(\frac{2k-1}{2n} \cdot \pi\right), k=1, 2, \dots, n$

Proof

$$T_n(x) = \cos(n \cdot \cos^{-1}(x))$$

$$\text{so } T_n(r_k) = \cos(n \cdot \cos^{-1}(r_k)) = \cos(n \cdot \frac{2k-1}{2n} \pi) = \cos((k-\frac{1}{2})\pi) = \sin(k\pi) = 0$$

since $k \in \mathbb{N}$

See book (§8.3) and web for more

Notation: $\tilde{T}_n(x) := \frac{1}{2^{n-1}} T_n(x) \quad (n \geq 1)$ so leading coefficient is 1,
i.e. \tilde{T}_n is a monic polynomial

T_k has even symmetry if k is even,
odd symmetry if k is odd.

you can show Chebyshev nodes minimize bounds on interpolation error,
 See § 8.3 Thm 8.10 + Corollary 8.11

More examples of orthogonal polynomials

- Suppose we want a domain $(0, \infty)$ not $[a, b]$.

Can't use $w(x) = 1$ since most nice functions won't
 have a finite norm: $\int_0^\infty f(x)^2 dx = \infty$ is likely

so usually choose

$$w(x) = e^{-x}. \text{ Doing Gram-Schmidt on } \{1, x, x^2, \dots\}$$

with this weight gives **Laguerre polynomials**

- Or domain $(-\infty, \infty)$, use $w(x) = e^{-x^2}$, get **Hermite polynomials**

- To generalize Chebyshev, take $[-1, 1]$, let $w(x) = (1-x)^\alpha (1+x)^\beta$,
 get **Jacobi polynomials**, of which **Chebyshev** are a special case
 $(\alpha = \pm \frac{1}{2}, \beta = 0)$
 as are **Legendre** ($\alpha = \beta = 0$)

- On $[a, b]$, do change of variables to get to $[-1, 1]$

p. 7 (Bonus content, not covered in lecture nor on exam)

Friday, October 24, 2025 8:52 AM

Beyond L^2 ... L^∞ : minimax approximation

Harder to approximate. Goal: find $p \in P_n$ to minimize $\|f - p\|_\infty$

$$\|f - p\|_\infty = \sup_{x \in [a, b]} |f(x) - p(x)|$$

use:

Thm (Chebyshev Equioscillation Thm)

Let $f \in C[a, b]$, and $n \geq 0$ an integer.

(1) There's a unique polynomial $g \in P_n$ s.t.

$$\|f - g\|_\infty = \inf_{p \in P_n} \|f - p\|_\infty \quad \text{i.e. "inf" is a "min" and it's unique}$$

(2) This g is the only polynomial in P_n for which there are at least $n+2$ distinct points $a \leq x_0 < x_1 < \dots < x_{n+1} \leq b$ s.t. $f(x_j) - g(x_j) = \sigma \cdot (-1)^j \cdot \rho_n(f)$ where σ is either $+1$ or -1 . Oscillation

Suggests a method to find g : let $E(x) = f(x) - p(x)$,

parameterize g by coefficients $\vec{\alpha}$, then solve

$$\begin{aligned} \text{eq'n} \quad & E(x_j) = (-1)^j \rho \\ & E'(x_j) = 0 \end{aligned} \quad j = 0, 1, \dots, n+1$$

says it's a critical pt, i.e. max or min hopefully

for unknowns $\vec{\alpha}, \rho, \{x_j\}$
Nonlinear system
use, e.g., Newton...

or...

Remez Exchange method:

w/ above equation, guess nodes $\{x_j\}$, then solve for $\vec{\alpha}$ and ρ is linear. Do that, then find $n+2$ nodes where error is greatest, let these be next guess for nodes $\{x_j\}$, and iterate.

Extension to rational functions approximation is current research!