

Approximation of Functions

Dangxing Chen

Duke Kunshan University

Minimax problem

Least-squares Approximation

Orthogonal polynomials

Revisit Chebyshev polynomial

Comparison with Machine Learning Models

Weierstrass Theorem

- ▶ Worse-case scenario error measure: Inf norm
- ▶ **Inf norm:** $\|f(x) - p(x)\|_\infty = \max_{a \leq x \leq b} |f(x) - p(x)|$
- ▶ **Weierstrass THM:** Let $f(x)$ be continuous on $[a, b]$ and let $\epsilon > 0$. Then \exists a polynomial $p(x)$ s.t.

$$\|f(x) - p(x)\|_\infty \leq \epsilon, a \leq x \leq b.$$

- ▶ **Question:** Why the polynomial fails in the Runge phenomenon?

Weierstrass Theorem

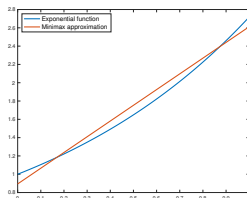
- ▶ Worse-case scenario error measure: Inf norm
- ▶ **Inf norm:** $\|f(x) - p(x)\|_\infty = \max_{a \leq x \leq b} |f(x) - p(x)|$
- ▶ **Weierstrass THM:** Let $f(x)$ be continuous on $[a, b]$ and let $\epsilon > 0$. Then \exists a polynomial $p(x)$ s.t.

$$\|f(x) - p(x)\|_\infty \leq \epsilon, a \leq x \leq b.$$

- ▶ **Question:** Why the polynomial fails in the Runge phenomenon?
- ▶ $\|f^{(n)}(x)\|_\infty$ increases fast w.r.t n
- ▶ Equispaced points error:
$$\|(x - x_0) \dots (x - x_n)\|_\infty \leq n! \left(\frac{b-a}{n}\right)^{n+1}$$
- ▶ Polynomial is very powerful!
- ▶ **Question:** What is and how to find the best polynomial?

Minimax problem

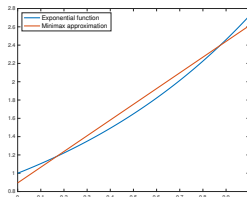
- ▶ Minimax problem $\min_{p_n} \|f - p_n\|_\infty$
 - ▶ Given p_n , maximize $|f(x) - p_n(x)|$
 - ▶ Find p_n , minimize $\|f - p_n\|_\infty$
- ▶ E.g., $f(x) = e^x$, $p_n(x) = a + bx$, $x \in [0, 1]$



- ▶ $\arg \max_x |f(x) - p_n(x)| =$

Minimax problem

- ▶ Minimax problem $\min_{p_n} \|f - p_n\|_\infty$
 - ▶ Given p_n , maximize $|f(x) - p_n(x)|$
 - ▶ Find p_n , minimize $\|f - p_n\|_\infty$
- ▶ E.g., $f(x) = e^x$, $p_n(x) = a + bx$, $x \in [0, 1]$



- ▶ $\arg \max_x |f(x) - p_n(x)| = 0, 1, \log(b)$, $\|f - g\|_\infty = \rho$
 - ▶ $e_1(0) = 1 - a = \rho$
 - ▶ $e_1(\log(b)) = a + b \log(b) - b = \rho = 1 - a \Rightarrow a = \frac{e - (e-1) \log(e-1)}{2}$
 - ▶ $e_1(1) = e - a - b = \rho = 1 - a \Rightarrow b = e - 1$
- ▶ **THM:** p_n minimizes $\|f - p_n\|_\infty$ iff $\exists x_0 < \dots < x_{n+1}$ s.t. $f(x_i) - p_n(x_i) = \sigma(-1)^i \|f - p_n\|_\infty$, where σ is 1 or -1.

Near Minimax problem

- ▶ Minimax is too difficult, depends on f
- ▶ Interpolation error: $e_n = \left| \frac{f^{(n+1)}(\eta)}{(n+1)!} (x - x_0) \dots (x - x_n) \right|$
- ▶ Cannot control $f^{(n+1)}(x)$

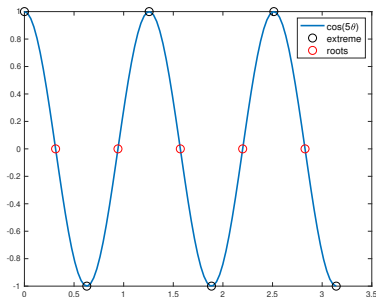
Near Minimax problem

- ▶ Minimax is too difficult, depends on f
- ▶ Interpolation error: $e_n = \left| \frac{f^{(n+1)}(\eta)}{(n+1)!} (x - x_0) \dots (x - x_n) \right|$
- ▶ Cannot control $f^{(n+1)}(x)$
- ▶ Near minimax: $\min_{x_0, \dots, x_n} |(x - x_0) \dots (x - x_n)|$
- ▶ Let $w_{n+1}(x) = (x - x_0) \dots (x - x_n)$
- ▶ **THM:** $w_{n+1}^*(x)$ is the solution iff $\exists a \leq \tilde{x}_0 < \dots < \tilde{x}_{n+1} \leq b$
s.t. $w_{n+1}^*(\tilde{x}_i) = \sigma(-1)^i \|w_{n+1}\|_\infty$, where σ is 1 or -1.

- ▶ **THM:** $w_{n+1}^*(x)$ is the solution iff $\exists a \leq \tilde{x}_0 < \cdots < \tilde{x}_{n+1} \leq b$ s.t. $w_{n+1}^*(\tilde{x}_i) = \sigma(-1)^i \|w_{n+1}\|_\infty$, where σ is 1 or -1.
- ▶ \Leftarrow
 - ▶ Suppose q_n is a monic polynomial with $\|q_{n+1}\|_\infty < \|w_{n+1}^*\|_\infty$
 - ▶ WLOG, assume $\sigma = 1$
 - ▶ $q_{n+1}(\tilde{x}_i) \begin{cases} < w_{n+1}^*(\tilde{x}_i), & \text{if } i \text{ even,} \\ > w_{n+1}^*(\tilde{x}_i), & \text{if } i \text{ odd.} \end{cases}$
 - ▶ $w_{n+1}^*(x) - q_{n+1}(x)$ flips sign at least $n + 1$ times
 - ▶ $w_{n+1}^*(x) - q_{n+1}(x)$ is a degree n polynomial (dominant term cancels)
 - ▶ IVT $\Rightarrow n + 1$ roots, contradiction
- ▶ **Question:** What function has such property?

Cosine function

- $\cos((n+1)\theta)$ for $\theta \in [0, \pi]$



- Extreme points: $\theta_k = \frac{k\pi}{n+1}$, $0 \leq k \leq n+1$, $|\cos(\theta_k)| = 1$
- Roots: $\theta_k = \frac{(2k+1)\pi}{2n+2}$, $0 \leq k \leq n$

Find the polynomial

- ▶ $\cos((n+1)\theta)$ for $\theta \in [0, \pi]$
- ▶ Extreme points: $\theta_k = \frac{k\pi}{n+1}$, $0 \leq k \leq n+1$
- ▶ Roots: $\theta_k = \frac{(2k+1)\pi}{2n+2}$, $0 \leq k \leq n$
- ▶ But $\cos(n\theta)$ is not a polynomial
- ▶ Suppose we focus on $f(x)$ with $x \in [-1, 1]$
- ▶ Goal: find a function $g(x)$ s.t.
 - ▶ $\cos(n\theta)$ is a polynomial
 - ▶ $\theta = g(x)$ s.t. $\cos(n\theta)$ has alternating extreme values
- ▶ Magic trick:

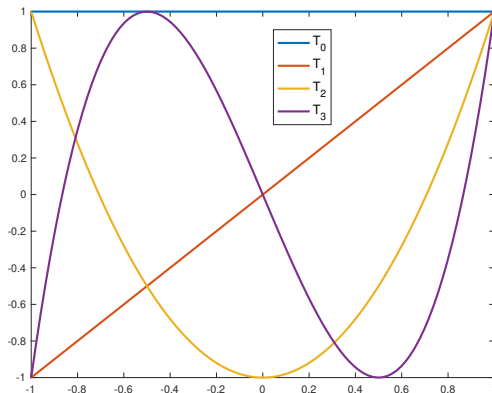
Find the polynomial

- ▶ $\cos((n+1)\theta)$ for $\theta \in [0, \pi]$
- ▶ Extreme points: $\theta_k = \frac{k\pi}{n+1}$, $0 \leq k \leq n+1$
- ▶ Roots: $\theta_k = \frac{(2k+1)\pi}{2n+2}$, $0 \leq k \leq n$
- ▶ But $\cos(n\theta)$ is not a polynomial
- ▶ Suppose we focus on $f(x)$ with $x \in [-1, 1]$
- ▶ Goal: find a function $g(x)$ s.t.
 - ▶ $\cos(n\theta)$ is a polynomial
 - ▶ $\theta = g(x)$ s.t. $\cos(n\theta)$ has alternating extreme values
- ▶ Magic trick: $g(x) = \cos^{-1}(x)$
- ▶ $T_n(x) = \cos(n \cos^{-1}(x))$
 - ▶ $T_0(x) = \cos(0) = 1$
 - ▶ $T_1(x) = x$
 - ▶ $T_{n\pm 1}(x) = \cos((n \pm 1)\theta) = \cos(n\theta) \cos(\theta) \mp \sin(n\theta) \sin(\theta)$
 - ▶ $T_{n+1}(x) + T_{n-1}(x) = 2 \cos(n\theta) \cos(\theta) = 2T_n(x)x$
 - ▶ $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$

Chebyshev polynomial

► Chebyshev polynomial:

$$p_n(x) = \cos(n\theta) = \cos(n \cos^{-1}(x)), x \in [-1, 1]$$



Comparison

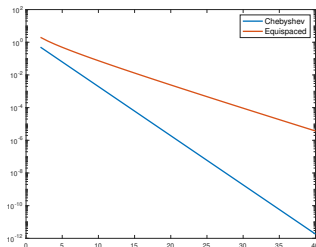
- ▶ Chebyshev polynomial:

$$p_{n+1}(x) = \cos((n+1)\theta) = \cos((n+1)\cos^{-1}(x)), x \in [-1, 1]$$

- ▶ Extreme points: $\theta_k = \frac{k\pi}{n+1}$, $0 \leq k \leq n+1$
- ▶ Error bound: $\|(x - x_0) \dots (x - x_n)\|_\infty \leq$

Comparison

- ▶ Chebyshev polynomial:
 $p_{n+1}(x) = \cos((n+1)\theta) = \cos((n+1)\cos^{-1}(x)), x \in [-1, 1]$
- ▶ Extreme points: $\theta_k = \frac{k\pi}{n+1}, 0 \leq k \leq n+1$
- ▶ Error bound: $\|(x - x_0) \dots (x - x_n)\|_\infty \leq \frac{1}{2^{n+1}}$
 - ▶ Extreme values: ± 1
 - ▶ $T_0(x) = 1, T_1(x) = x$
 - ▶ Recurrence relationship: $T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x)$
- ▶ Equispaced points: $\|(x - x_0) \dots (x - x_n)\|_\infty \leq n! \left(\frac{2}{n}\right)^{n+1}$
- ▶ Comparison:



Interpolation

- ▶ Chebyshev: $\cos((n+1)\theta) = \cos((n+1)\cos^{-1}(x)), x \in [-1, 1]$
- ▶ **Question:** how to interpolate a Chebyshev polynomial?

Interpolation

- ▶ Chebyshev: $\cos((n+1)\theta) = \cos((n+1)\cos^{-1}(x)), x \in [-1, 1]$
- ▶ **Question:** how to interpolate a Chebyshev polynomial?
- ▶ $f(x) \approx p_n(x) = c_0 + c_1x + \dots + c_nx^n$
- ▶ Roots: $\theta_k = \frac{(2k+1)\pi}{2n+2}$, $x_k = \cos(\theta_k)$, $0 \leq k \leq n$
- ▶ Same way with different points: $\mathbf{A}\mathbf{c} = \mathbf{f}$

$$\begin{bmatrix} 1 & x_0 & \dots & x_0^n \\ 1 & x_1 & \dots & x_1^n \\ \vdots & \ddots & \ddots & \vdots \\ 1 & x_n & \dots & x_n^n \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} f(x_0) \\ f(x_1) \\ \vdots \\ f(x_n) \end{bmatrix}.$$

- ▶ Differentiation: same, $\mathbf{D} = \mathbf{A}\mathbf{B}\mathbf{A}^{-1}$
- ▶ Integration: same, $\mathbf{S} = \tilde{\mathbf{A}}\mathbf{B}\mathbf{A}^{-1}$

Minimax problem

Least-squares Approximation

Orthogonal polynomials

Revisit Chebyshev polynomial

Comparison with Machine Learning Models

The least squares approximation problem

- ▶ Average error measurement: 2 norm
- ▶ Notation: $\|g\|_2 = \sqrt{\int_a^b |g(x)|^2 dx}$
- ▶ Least squares approximation $\min_{p_n} \|f - p_n\|_2^2$
- ▶ E.g. $f(x) = e^x, -1 \leq x \leq 1$, let $p_1(x) = c_0 + c_1x$
- ▶ **Question:** How to minimize $r = \|f - p_1(x)\|_2^2$?

The least squares approximation problem

- ▶ Average error measurement: 2 norm
- ▶ Notation: $\|g\|_2 = \sqrt{\int_a^b |g(x)|^2 dx}$
- ▶ Least squares approximation $\min_{p_n} \|f - p_n\|_2^2$
- ▶ E.g. $f(x) = e^x, -1 \leq x \leq 1$, let $p_1(x) = c_0 + c_1x$
- ▶ **Question:** How to minimize $r = \|f - p_1(x)\|_2^2$?
- ▶ $\frac{\partial r}{\partial c_0} = \frac{\partial r}{\partial c_1} = 0$
- ▶ $0 = \frac{\partial r}{\partial c_0} = -2 \int_{-1}^1 e^x - c_0 - c_1x \, dx$
- ▶ $0 = \frac{\partial r}{\partial c_1} = -2 \int_{-1}^1 (e^x - c_0 - c_1x) x \, dx$

The least squares approximation problem

- ▶ Average error measurement: 2 norm
- ▶ Notation: $\|g\|_2 = \sqrt{\int_a^b |g(x)|^2 dx}$
- ▶ Least squares approximation $\min_{p_n} \|f - p_n\|_2^2$
- ▶ E.g. $f(x) = e^x$, $-1 \leq x \leq 1$, let $p_1(x) = c_0 + c_1x$
- ▶ **Question:** How to minimize $r = \|f - p_1(x)\|_2^2$?
- ▶ $\frac{\partial r}{\partial c_0} = \frac{\partial r}{\partial c_1} = 0$
- ▶ $0 = \frac{\partial r}{\partial c_0} = -2 \int_{-1}^1 e^x - c_0 - c_1x \, dx$
- ▶ $0 = \frac{\partial r}{\partial c_1} = -2 \int_{-1}^1 (e^x - c_0 - c_1x) x \, dx$
- ▶
$$\begin{bmatrix} 2 & \int_{-1}^1 x \, dx \\ \int_{-1}^1 x \, dx & \int_{-1}^1 x^2 \, dx \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \begin{bmatrix} \int_{-1}^1 e^x \, dx \\ \int_{-1}^1 x e^x \, dx \end{bmatrix}$$
- ▶ $c_0 = \frac{1}{2} \int_{-1}^1 e^x \, dx$
- ▶ $c_1 = \frac{3}{2} \int_{-1}^1 x e^x \, dx$

General least squares problem

- ▶ $r = \int_a^b w(x)[f(x) - p_n(x)]^2 dx$
- ▶ Examples:
 - ▶ $w(x) = 1, a \leq x \leq b$
 - ▶ $w(x) = \frac{1}{\sqrt{1-x^2}}, -1 \leq x \leq 1$
 - ▶ $w(x) = e^{-x^2}, -\infty < x < \infty$
- ▶ $w(x) = 1$ not necessarily always a good choice
- ▶ $w(x)$ allows different degrees of importance to different points
- ▶ **Question:** What are equations need to be solved?

General least squares problem

- ▶ $r = \int_a^b w(x)[f(x) - p_n(x)]^2 dx$
- ▶ Examples:
 - ▶ $w(x) = 1, a \leq x \leq b$
 - ▶ $w(x) = \frac{1}{\sqrt{1-x^2}}, -1 \leq x \leq 1$
 - ▶ $w(x) = e^{-x^2}, -\infty < x < \infty$
- ▶ $w(x) = 1$ not necessarily always a good choice
- ▶ $w(x)$ allows different degrees of importance to different points
- ▶ **Question:** What are equations need to be solved?
- ▶ $r(c_0, \dots, c_n) = \int_a^b w(x) \left[f(x) - \sum_{j=0}^n c_j x^j \right]^2 dx$
- ▶ $\frac{\partial r}{\partial c_k} = 0$
- ▶ $\int_a^b w(x) \left[f(x) - \sum_{j=0}^n c_j x^j \right] (-2)x^k dx = 0$
- ▶ $\sum_{j=0}^n c_j \int_a^b w(x) x^{k+j} dx = \int_a^b w(x) f(x) x^k dx$
- ▶ **Question:** How to solve this?

Solve the general least squares problem

- Form the linear system

$$\begin{bmatrix} \int_a^b w(x) dx & \int_a^b w(x)x dx & \dots & \int_a^b w(x)x^n dx \\ \int_a^b w(x)x dx & \int_a^b w(x)x^2 dx & \dots & \int_a^b w(x)x^{n+1} dx \\ \vdots & \ddots & \ddots & \vdots \\ \int_a^b w(x)x^n dx & \int_a^b w(x)x^{n+1} dx & \dots & \int_a^b w(x)x^{2n} dx \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} \int_a^b w(x)f(x) dx \\ \int_a^b w(x)f(x)x dx \\ \vdots \\ \int_a^b w(x)f(x)x^n dx \end{bmatrix}.$$

- Setting up the linear system requires a lot of integrals
- **Question:** Is there an alternative convenient way?

Minimax problem

Least-squares Approximation

Orthogonal polynomials

Revisit Chebyshev polynomial

Comparison with Machine Learning Models

Inner product

- ▶ Problem: Basis $\{x^n\}$ are very linearly dependent
- ▶ **Def: Inner product** of f and g :

$$(f, g) = \int_a^b w(x) f(x) g(x) dx.$$

- ▶ Properties:

- ▶ $(\alpha f, g) = (f, \alpha g) = \alpha(f, g), \forall \alpha \in \mathbb{R}$
- ▶ $(f_1 + f_2, g) = (f_1, g) + (f_2, g)$ and $(f, g_1 + g_2) = (f, g_1) + (f, g_2)$
- ▶ $(f, g) = (g, f)$
- ▶ $(f, f) \geq 0$ and $(f, f) = 0$ iff $f(x) = 0$

- ▶ **Def: Two norm:** $\|f\|_2 = \sqrt{\int_a^b w(x) [f(x)]^2 dx} = \sqrt{(f, f)}$
- ▶ **Lemma (Cauchy-Schwartz):** $|(f, g)| \leq \|f\|_2 \|g\|_2$
- ▶ Many properties in the textbook

Create orthogonal polynomials

- ▶ **Def:** f and g are **orthogonal** if $(f, g) = 0$
- ▶ **THM:** \exists a sequence of polynomials $\{\phi_n(x) | n \geq 0\}$ with $\deg(\phi_n) = n, \forall n$, and

$$(\phi_n, \phi_m) = 0, \quad \forall n \neq m, \quad n, m \geq 0.$$

In addition, if we require $(\phi_n, \phi_n) = 1$ and coefficients of x^n is positive, then the sequence $\{\phi_n\}$ is unique.

- ▶ **Question:** Start with $1, x, \dots, x^n$, how to calculate it?

Create orthogonal polynomials

- ▶ **Def:** f and g are **orthogonal** if $(f, g) = 0$
- ▶ **THM:** \exists a sequence of polynomials $\{\phi_n(x) | n \geq 0\}$ with $\deg(\phi_n) = n$, $\forall n$, and

$$(\phi_n, \phi_m) = 0, \quad \forall n \neq m, \quad n, m \geq 0.$$

In addition, if we require $(\phi_n, \phi_n) = 1$ and coefficients of x^n is positive, then the sequence $\{\phi_n\}$ is unique.

- ▶ **Question:** Start with $1, x, \dots, x^n$, how to calculate it?
- ▶ **Gram-Schmidt process:**

$$\phi_0 = \frac{1}{\|1\|_2},$$

$$\phi_1 = \frac{x - (\phi_0, x)\phi_0}{\|x - (\phi_0, x)\phi_0\|_2},$$

$$\phi_2 = \frac{x^2 - (\phi_0, x^2)\phi_0 - (\phi_1, x^2)\phi_1}{\|x^2 - (\phi_0, x^2)\phi_0 - (\phi_1, x^2)\phi_1\|_2},$$

\vdots

Example

- ▶ Suppose $w(x) = 1, [a, b] = [-1, 1]$
- ▶ **Question:** How to calculate the first three orthogonal polynomials ?

Example

- ▶ Suppose $w(x) = 1$, $[a, b] = [-1, 1]$
- ▶ **Question:** How to calculate the first three orthogonal polynomials ?
- ▶ $\phi_0 = \frac{1}{(1,1)} = \frac{1}{\sqrt{\int_{-1}^1 1 \, dx}} = \frac{1}{\sqrt{2}}$
- ▶ $(x, \phi_0)\phi_0 = \frac{1}{2} \int_{-1}^1 x \, dx = 0$
- ▶ $\|x - (x, \phi_0)\phi_0\|_2^2 = \int_{-1}^1 x^2 \, dx = \frac{2}{3}$
- ▶ $\phi_1 = \frac{x - (x, \phi_0)\phi_0}{\|x - (x, \phi_0)\phi_0\|_2} = \sqrt{\frac{3}{2}}x$
- ▶ $(x^2, \phi_0)\phi_0 = \frac{1}{2} \int_{-1}^1 x^2 \, dx = \frac{1}{3}$
- ▶ $(x^2, \phi_1)\phi_1 = \frac{3}{2}x \int_{-1}^1 x^3 \, dx = 0$
- ▶ $\|x^2 - (x^2, \phi_0)\phi_0 - (x^2, \phi_1)\phi_1\|_2^2 = \int_{-1}^1 (x^2 - \frac{1}{3})^2 \, dx = \frac{8}{45}$
- ▶ $\phi_2 = \frac{x^2 - (x^2, \phi_0)\phi_0 - (x^2, \phi_1)\phi_1}{\|x^2 - (x^2, \phi_0)\phi_0 - (x^2, \phi_1)\phi_1\|_2} = \frac{x^2 - \frac{1}{3}}{\sqrt{\frac{8}{45}}} = \sqrt{\frac{5}{8}}(3x^2 - 1)$

QR decomposition

- ▶ Function version of QR ?

QR decomposition

- ▶ Function version of QR ?
- ▶ Gram-Schmidt: $\{1, x, \dots, x^n\} \Rightarrow \{\phi_0, \phi_1, \dots, \phi_n\}$
- ▶ $\mathbf{x}^T(x) = [1 \quad \dots \quad x^n]$, $\mathbf{q}^T(x) = [\phi_0(x) \quad \dots \quad \phi_n(x)]$
- ▶ $\mathbf{x}^T(x) = [1 \quad \dots \quad x^n] = \mathbf{q}^T(x)\mathbf{R}$
- ▶ $x^k = c_0\phi_0 + \dots + c_n\phi_n \Rightarrow c_i = (x^k, \phi_i)$
- ▶ $(x^k, \phi_i) = 0$ for $i > k$
- ▶ $\mathbf{R} = \begin{bmatrix} (1, \phi_0) & (x, \phi_0) & \dots & (x^n, \phi_0) \\ 0 & (x, \phi_1) & \dots & (x^n, \phi_1) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & (x^n, \phi_n) \end{bmatrix}$

Relationship between two polynomials

- ▶ **Question:** Relationships of coefficients?

Relationship between two polynomials

- ▶ **Question:** Relationships of coefficients?

- ▶ $p_n(x) = c_0 + \cdots + c_n x^n = \begin{bmatrix} 1 & \cdots & x^n \end{bmatrix} \begin{bmatrix} c_0 \\ \vdots \\ c_n \end{bmatrix} = \mathbf{x}^T(x) \mathbf{c}$

- ▶ $q_n(x) = d_0 \phi_0 + \cdots + d_n \phi_n = \begin{bmatrix} \phi_0 & \cdots & \phi_n \end{bmatrix} \begin{bmatrix} d_0 \\ \vdots \\ d_n \end{bmatrix} = \mathbf{q}^T(x) \mathbf{d}$

- ▶ $\mathbf{x}^T(x) = \mathbf{q}^T(x) \mathbf{R}$

- ▶ $p_n(x) = \mathbf{q}^T(x) \mathbf{R} \mathbf{c}$

- ▶ $q_n(x) = \mathbf{x}^T(x) \mathbf{R}^{-1} \mathbf{d}$

Minimax problem

Least-squares Approximation

Orthogonal polynomials

Revisit Chebyshev polynomial

Comparison with Machine Learning Models

Chebyshev polynomial

- ▶ **Def: Chebyshev polynomial**, $-1 \leq x \leq 1$

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

- ▶ Chebyshev polynomial is orthogonal with $w(x) = \frac{1}{\sqrt{1-x^2}}$
- ▶ $(T_m, T_n) = \int_{-1}^1 \frac{\cos(m \cos^{-1}(x)) \cos(n \cos^{-1}(x))}{\sqrt{1-x^2}} dx$

Chebyshev polynomial

- ▶ **Def: Chebyshev polynomial**, $-1 \leq x \leq 1$

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

- ▶ Chebyshev polynomial is orthogonal with $w(x) = \frac{1}{\sqrt{1-x^2}}$
- ▶ $(T_m, T_n) = \int_{-1}^1 \frac{\cos(m \cos^{-1}(x)) \cos(n \cos^{-1}(x))}{\sqrt{1-x^2}} dx$
- ▶ Change of variable: $x = \cos(\theta)$, $dx = -\sin(\theta) d\theta$
- ▶ $(T_m, T_n) = \int_{\pi}^0 \frac{\cos(m\theta) \cos(n\theta)}{\sqrt{1-\cos^2(\theta)}} (-\sin(\theta)) d\theta = \int_0^{\pi} \cos(m\theta) \cos(n\theta) d\theta$
- ▶ Recall: $2 \cos(\theta) \cos(\eta) = \cos(\theta - \eta) + \cos(\theta + \eta)$
- ▶ $(T_m, T_n) = \begin{cases} 0, & n \neq m, \\ \pi, & m = n = 0, \\ \frac{\pi}{2}, & n = m > 0. \end{cases}$

Chebyshev coefficients

- ▶ $f(x) = \frac{a_0}{2} + \sum_{i=1}^n a_i T_i(x)$
- ▶ **Question:** How are $\{a_i\}$ determined?

Chebyshev coefficients

- ▶ $f(x) = \frac{a_0}{2} + \sum_{i=1}^n a_i T_i(x)$
- ▶ **Question:** How are $\{a_i\}$ determined?
- ▶ $(T_j, f) = \frac{a_0}{2}(T_j, 1) + \sum_{i=1}^n a_i(T_j, T_i)$
- ▶ $j = 0$: $(T_0, f(x)) = \frac{a_0}{2}(T_0, T_0) = \frac{\pi}{2} \Rightarrow a_0 = \frac{2}{\pi}(T_0, f)$
- ▶ $j \neq 0$: $(T_j, f(x)) = a_j(T_j, T_j) \Rightarrow a_j = \frac{2}{\pi}(T_j, f)$
- ▶ $a_j = \frac{2}{\pi} \int_{-1}^1 \frac{T_j(x)f(x)}{\sqrt{1-x^2}} dx = \frac{2}{\pi} \int_{-1}^1 \frac{\cos(j \cos^{-1}(x))f(x)}{\sqrt{1-x^2}} dx$
- ▶ $x = \cos(\theta), 0 \leq \theta \leq \pi$
- ▶ **Cosine transform:** $a_j = \frac{2}{\pi} \int_0^\pi f(\cos(\theta)) \cos(j\theta) d\theta$
- ▶ **Question:** How to discretize it?

Chebyshev coefficients

- ▶ $f(x) = \frac{a_0}{2} + \sum_{i=1}^n a_i T_i(x)$
- ▶ **Question:** How are $\{a_i\}$ determined?
- ▶ $(T_j, f) = \frac{a_0}{2}(T_j, 1) + \sum_{i=1}^n a_i (T_j, T_i)$
- ▶ $j = 0$: $(T_0, f(x)) = \frac{a_0}{2}(T_0, T_0) = \frac{\pi}{2} \Rightarrow a_0 = \frac{2}{\pi}(T_0, f)$
- ▶ $j \neq 0$: $(T_j, f(x)) = a_j (T_j, T_j) \Rightarrow a_j = \frac{2}{\pi}(T_j, f)$
- ▶ $a_j = \frac{2}{\pi} \int_{-1}^1 \frac{T_j(x)f(x)}{\sqrt{1-x^2}} dx = \frac{2}{\pi} \int_{-1}^1 \frac{\cos(j \cos^{-1}(x))f(x)}{\sqrt{1-x^2}} dx$
- ▶ $x = \cos(\theta), 0 \leq \theta \leq \pi$
- ▶ **Cosine transform:** $a_j = \frac{2}{\pi} \int_0^\pi f(\cos(\theta)) \cos(j\theta) d\theta$
- ▶ **Question:** How to discretize it?
- ▶ **Chebyshev nodes:** $x_k = \cos\left(\frac{2k+1}{2n+2}\pi\right), 0 \leq k \leq n$
- ▶ **Discrete orthogonality:** $\sum_{k=0}^n T_i(x_k)T_j(x_k) = K_i \delta_{i,j}$
- ▶ $K_0 = n+1$ and $K_i = \frac{n+1}{2}$ for $1 \leq i \leq n$

Least squares

- ▶ Consider orthogonal polynomials $T_j(x)$
- ▶ $p_n(x) = a_0T_0(x) + \cdots + a_nT_n(x)$
- ▶ $r = \|f - p_n\|_2^2 = \int_{-1}^1 w(x) \left[f(x) - \sum_{j=0}^n a_j T_j(x) \right]^2 dx$
- ▶ $\min_{p_n} r \Rightarrow \frac{\partial r}{\partial a_k} = 0$
- ▶ $\frac{\partial r}{\partial a_k} = 0 \Rightarrow \int_{-1}^1 w(x) (f(x) - \sum_{j=0}^n a_j T_j(x)) T_k(x) dx = 0 \Rightarrow$
 $\sum_{j=0}^n a_j \int_{-1}^1 w(x) T_j(x) T_k(x) dx = \int_{-1}^1 w(x) f(x) T_k(x) dx$
- ▶ Inner product notation: $\sum_{j=0}^n a_j (T_j, T_k) = (f, T_k)$
- ▶ Orthogonality: $(T_j, T_k) = 0$ for $j \neq k$
- ▶ $a_k = \frac{(f, T_k)}{\|T_k\|_2^2}$
- ▶ $p_n^*(x) = \sum_{j=0}^n \frac{(f, T_j)}{\|T_j\|_2^2} T_j(x)$

Discrete cosine transform

▶ $p_n(x) = \frac{a_0}{2} + \sum_{j=1}^n a_j T_j(x)$

▶ $x_k = \cos\left(\frac{2k+1}{2n+2}\pi\right)$

▶ $a_k = \frac{2}{n+1} \sum_{j=0}^n f(x_j) T_k(x_j)$

▶ $\theta_j = \frac{2j+1}{2N+2}\pi$

▶ **Discrete cosine transform:**

$$a_k = \frac{2}{n+1} \sum_{j=0}^n f(\cos(\theta_j)) \cos(n\theta_j)$$

▶ **Question:** Matrix form

▶
$$\begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \frac{2}{n+1} \begin{bmatrix} 1 & 1 & \dots & 1 \\ \cos(\theta_0) & \cos(\theta_1) & \dots & \cos(\theta_n) \\ \vdots & \ddots & \ddots & \vdots \\ \cos(n\theta_0) & \cos(n\theta_1) & \dots & \cos(n\theta_n) \end{bmatrix}$$

Connection with polynomial interpolation approach

- ▶ $f(x) \approx p_n(x) = \frac{a_0}{2} + \sum_{j=1}^n a_j T_j(x)$
- ▶ **Chebyshev nodes:** $x_k = \cos\left(\frac{2k+1}{2n+2}\pi\right) = \cos(\theta_k)$, $0 \leq k \leq n$
- ▶ $f(x_k) = p_n(x_k)$, $\mathbf{A}\mathbf{a} = \mathbf{f}$

$$\begin{bmatrix} \frac{1}{2} & \cos(\theta_0) & \dots & \cos(n\theta_0) \\ \frac{1}{2} & \cos(\theta_1) & \dots & \cos(n\theta_1) \\ \vdots & \ddots & \ddots & \vdots \\ \frac{1}{2} & \cos(\theta_n) & \dots & \cos(n\theta_n) \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} f(\cos(\theta_0)) \\ f(\cos(\theta_1)) \\ \vdots \\ f(\cos(\theta_n)) \end{bmatrix}$$

- ▶ Orthogonality:

$$\mathbf{A}^T \mathbf{A} = \begin{bmatrix} \frac{n+1}{4} & 0 & \dots & 0 \\ 0 & \frac{n+1}{2} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{n+1}{2} \end{bmatrix}.$$

Differentiation and integration

- ▶ $f(x) = \frac{a_0}{2} + \sum_n a_n T_n(x)$
- ▶ $f'(x) = \frac{b_0}{2} + \sum_n b_n T_n(x)$
- ▶ $\int_{-1}^x f(t) dt = \frac{c_0}{2} + \sum_n c_n T_n(x)$
- ▶ Analytical computation:

- ▶ Differentiation:

$$b_k = 2 \sum_{\substack{j=k+1 \\ j+k \text{ odd}}} j \cdot a_j$$

- ▶ Integration:

$$c_k = \frac{1}{2k}(a_{k-1} - a_{k+1}), \quad k \geq 1,$$
$$c_0 = 2(c_1 - c_2 + c_3 - \dots)$$

- ▶ Matrix version:
 - ▶ \mathbf{A} as the discrete cosine matrix
 - ▶ Differentiation matrix: $\mathbf{D} = \mathbf{A}^{-1}\mathbf{B}\mathbf{A}$
 - ▶ Integration matrix: $\mathbf{S} = \mathbf{A}^{-1}\mathbf{C}\mathbf{A}$

Minimax problem

Least-squares Approximation

Orthogonal polynomials

Revisit Chebyshev polynomial

Comparison with Machine Learning Models

Comparison

- ▶ Theoretical supports: both have universal approximation property
- ▶ Advantages:
 - ▶ Chebyshev is a near minimax approximation and nice analytical properties
 - ▶ ML empirically work well for high-dimensional problem
- ▶ Disadvantages:
 - ▶ Polynomial interpolations face with curse of dimensionality
 - ▶ Polynomial not good with complex geometries
 - ▶ ML models convergence slow