Numerical analysis (2/7): Interpolation, approximation

University of Luxembourg

Philippe Marchner

Siemens Digital Industries Software, France

October 26th, 2023

Outline

Outline

Overview of the content

Polynomial interpolation is an important building block of numerical analysis.
We try to fit a polynomial P of degree $\leq n$ that exactly pass by $n+1$ given points

$$y_i = P(x_i), \ i \in \{0, n\}.$$

It is a special case of approximation: find a simpler function P (we focus on polynomials) that is the closest to f with respect to some **distance**

- We can study the interpolation error, where we have more liberty to choose the x_i
- We do not need P to pass through x_i, but only capture the main trend of f: least squares

If we have a lot of data $(x_i, y_i), \ i \in \{0, m\}$ with $m \gg n$, we can try to **fit the data** to a polynomial P of lower degree n, we talk about discrete least squares.

Outline

Outline

Polynomial evaluation

Let P be a polynomial of degree n

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_2 x^2 + a_1 x + a_0,$$

that we want to evaluate at a point $x = \xi$

- one algorithm consists in computing each product $a_n \xi^n$ and next summing.
- but this technique is generally not used
  - because of the errors that it generates (cf last week exercise)
  - the number of elementary operations is too large, most particularly when $n \gg 1$: $n(n+1)/2$ multiplications and n additions $\rightarrow$ $\mathcal{O}(n^2)$ operations.
- the Hörner algorithm is generally preferred

Polynomial evaluation

Let P a polynomial of degree n

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + ... + a_2 x^2 + a_1 x + a_0,$$

that we want to evaluate at a point $x = \xi$

Hörner algorithm

The Hörner algorithm is based on the following factorization

$$P(x) = a_0 + x(a_1 + x(a_2 + x(a_3 + \dots x(a_{n-2} + x(a_{n-1} + x a_n)) \dots)))$$

This algorithm requires n additions and n multiplications $\rightarrow$ $\mathcal{O}(n)$ operations
It has a convenient recursive form

Outline

Lagrange interpolation

The interpolation problem

Let (y_i, x_i) be $n+1$ distinct points $x_0, x_1, ...x_n$ and $y_0, y_1, ...y_n$ on the interval $[a, b]$.
The goal is to build a polynomial P of degree less or equal to n such that

$$P(x_i) = y_i \quad \forall i = 0, 1, \ldots, n$$

Theorem

There exists one and only one polynomial P_n of degree less or equal to n satisfying

$$P_n(x_i) = y_i \quad \forall i = 0, 1, \ldots, n$$

which writes

$$P_n(x) = \sum_{i=0}^{n} y_i \, L_i(x) \quad \text{with} \quad L_i(x) = \prod_{k=0 \quad k \neq i}^{n} \frac{(x - x_k)}{(x_i - x_k)}.$$

- This polynomial P_n is called Lagrange interpolation polynomial at the points $x_0, x_1, ..., x_n$.
- The polynomials $L_i(x)$ are the Lagrange basis functions associated to the points x_i.

Proof

- Existence

  The polynomial function P_n is a polynomial of degree n. Since $L_i(x_j) = \delta_{ij}$, it satisfies

$$P_n(x_i) = y_i \quad \forall i = 0, 1, \ldots, n$$

Proof

- Existence

  The polynomial function P_n is a polynomial of degree n. Since $L_i(x_j) = \delta_{ij}$, it satisfies

$$P_n(x_i) = y_i \quad \forall i = 0, 1, \ldots, n$$

- Uniqueness

  Let Q be another polynomial function solution to the problem. Then $\forall i = 0, 1, ...n$

$$Q(x_i) - P(x_i) = 0 \quad \forall i = 0, 1, \ldots, n.$$

  Therefore $Q - P$ is a polynomial function of degree less or equal to n which is zero at $n + 1$ points. As a consequence, this polynomial is identically zero.

Example

With two points $n = 1$

$$P_1(x) = y_0 \frac{(x - x_1)}{(x_0 - x_1)} + y_1 \frac{(x - x_0)}{(x_1 - x_0)}$$

which can also be written

$$P_1(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0} (x - x_0) .$$

Example

Find the interpolating polynomial for

$$(x_0, x_1, x_2) = (0, 2, 3), \quad (y_0, y_1, y_2) = (1, 5, -2)$$

Step 1: look for the 2nd order Lagrange basis (L_0, L_1, L_2) such as $L_i(x_j) = \delta_{ij}$.
Step 2: use the theorem to find P_2 by linear combination

Outline

Remark

Computing $P_n(x) = \sum_{i=0}^{n} y_i \, L_i(x)$ requires too much elementary evaluations.

Newton form of the interpolation polynomial

We prefer to use the Newton formula which consists in writing the interpolation polynomial P_n at points $x_0, x_1, \ldots x_n$ under the form

$$P_n(x) = a_0 + a_1 \, (x - x_0) + \ldots + a_n \, (x - x_0) \, (x - x_1) \ldots (x - x_{n-1})$$

Advantages

- we can use the stable Hörner algorithm to evaluate this polynomial
- If we know P_{n-1}, it is sufficient to compute a_n to determine P_n

$$P_n(x) = P_{n-1}(x) + a_n \, (x - x_0) \, (x - x_1) \ldots (x - x_{n-1})$$

(this is useful most particularly if one point is added!)

Newton interpolation formula - Divided differences

The coefficients a_n can be expressed explicitly by *divided differences*

$$P_n(x) = \sum_{j=0}^{n} \left(a_j \prod_{i=0}^{j-1} (x - x_i) \right), \quad a_j = f[x_0, x_1, \ldots, x_j]$$

where $f[.]$ are the divided differences of f such as

$$f[x_0, x_1, \ldots, x_j] = \frac{f[x_1, \ldots, x_j] - f[x_0, \ldots, x_{j-1}]}{x_j - x_0}.$$

The recursion begins with $f[x_j] = f(x_j)$, $\forall j \in \{0, n\}$ (see implementation exercise)

Remark

Although the interpolating polynomial is unique, it can be written in different forms

Outline

Error measure for polynomial interpolation

Since one of the goals of the interpolation is to replace the evaluation of $f(x)$ by $P_n(x)$, an important point is to measure the error

$$E_n(x) = f(x) - P_n(x), \quad x \in [a, b]$$

Can we approach a continuous function f by its Lagrange polynomial interpolant ?

If f has $n+1$ bounded derivatives,

$$E_n(x) = \frac{f^{(n+1)}(\xi_x)}{(n+1)!} \prod_{i=0}^{n} (x - x_i), \; \xi_x \in [a, b]$$

Example

- Sinusoïdal function on $[-\pi, \pi]$, equidistant points
- $x \mapsto 1/(1 + x^2)$ on $[-a, a]$, equidistant points

Error measure for the polynomial interpolation

Theorem

Let $f : [a, b] \longrightarrow \mathbb{R}$, $n+1$ times continuously differentiable and P_n the Lagrange interpolation polynomial at points $x_0, x_1, ..., x_n$ in $[a, b]$. Then

$$|f(x) - P_n(x)| \leq \frac{M_{n+1}}{(n+1)!} |\pi_n(x)|$$

where

$$M_{n+1} = \max_{a \leq x \leq b} \left| f^{(n+1)}(x) \right|$$

and

$$\pi_n(x) = \prod_{i=0}^{n} (x - x_i).$$

Example

$f(x) = \sin(x)$ for $x \in [a, b]$

- we have $M_{n+1} \leq 1$
- we also can easily obtain the upper bound $|\pi_n(x)| \leq (b-a)^{n+1}$ (for any choice of the $n+1$ points x_i)
- then

$$\forall x \in [a, b]\,, \quad |f(x) - P_n(x)| \leq \frac{(b-a)^{n+1}}{(n+1)!}$$

- and, consequently

$$\lim_{n\to\infty} |f(x) - P_n(x)| = 0$$

Example

$f(x) = 1/(1+x^2)$ for $x \in [-a, a]$

- if the points x_i are equispaced in $[-a, a]$, we can show that

$$\max_{-a \leq x \leq a} |E_n(x)| < C \frac{e^{-n}}{\sqrt{n} \log n} (2a)^{n+1}$$

- so if $2a < e$, $E_n(x)$ converges towards 0 with n.
- in our simulation $a = 5$ and in this case, the upper bound tends to infinity. We cannot say anything about $E_n(x)$.
- our computation lets to think that E_n does not tends to 0
- and $\max_{|x| \leq 5} |E_n(x)|$ did not converge towards 0 (Runge phenomena)

Choice of the interpolation points

- the error estimates of the theorem shows that the error depends on
  - the $(n+1)$-th derivative of f
  - the maximum of the function π_n, which only depends on the x_i
- We can try to minimize $\max_{a \leq x \leq b} |\pi_n(x)|$ by a better choice of the points x_i
- choosing equally distributed points is far from being optimal!
- the optimal choice is obtained by the Chebyshev points

$$x_i = \frac{a+b}{2} + \frac{b-a}{2} \cos\left(\frac{(2i+1)\pi}{2n+2}\right), i = 0, 1, \ldots, n$$

Another way to avoid Runge phenomena is by piecewise interpolation

- linear, quadratic, spline, nearest neighbourhood, etc.

Outline

Outline

Least squares problem

Continuous polynomial least squares approximation

Let f be a continuous function in $[a, b]$
Goal: find $\min_{P \in \mathbb{R}_n[X]} \|f - P\|_{L_2}^2$, with $P = \sum_{i=0}^{N} a_i \phi_i(x)$, $\{\phi_i\}$ a basis of $\mathbb{R}_n[X]$.
To do so we study the error $\langle f - P, f - P \rangle_{L_2}$

$$\mathcal{E}(a_0, \dots, a_n) = \int_a^b \left[f(x) - \sum_{i=0}^{N} a_i \phi_i(x) \right]^2 dx$$

and solve $\frac{\partial \mathcal{E}}{\partial a_j} = 0, \quad \forall j \in \{0, n\}$

We obtain a linear system for the $(a_j)_{j \in \{0,n\}}$

$$\langle f, \phi_j \rangle = \sum_{i=0}^{n} a_i \langle \phi_i, \phi_j \rangle$$

Outline

Orthogonal polynomials

Definition

A sequence of orthogonal polynomials (finite or infinite) $\phi_0(x), \phi_1(x), \ldots$ is such that

$$\phi_i \text{ is of degree } i$$

$$\langle \phi_i, \phi_j \rangle = 0, \quad \text{if } i \neq j.$$

We study the sequences of orthogonal polynomials for a scalar product like

$$\langle \phi, \psi \rangle_{L_2} = \int_a^b \phi(x)\psi(x)w(x)dx$$

where w ($=$ weight) is a strictly positive continuous function on $]a, b[$

Remark: orthogonal polynomials ($w = 1$) solves the least squares system: $a_j = \frac{\langle f, \phi_j \rangle}{\langle \phi_j, \phi_j \rangle}$

Properties of orthogonal polynomials

Let (ϕ_n) be a sequence of orthogonal polynomials; then

- Any polynomial P of degree less or equal to k writes in a unique way like

$$P(x) = d_0\phi_0(x) + d_1\phi_1(x) + \ldots + d_k\phi_k(x) \quad \text{with } d_i = \frac{\langle \phi_i, P \rangle}{\langle \phi_i, \phi_i \rangle}$$

- If P is of degree $< k$, then $\langle P, \phi_k \rangle = 0$
- If A_i designates the coefficient of the term of highest degree of ϕ_i, then we have the three terms short recurrence

$$\widehat{\phi}_{i+1}(x) = (x - B_i)\widehat{\phi}_i(x) - C_i\widehat{\phi}_{i-1}(x) \text{ where } \widehat{\phi}_i(x) := \frac{\phi_i(x)}{A_i} \quad \text{(normalized polyn.)}$$

$$B_i = \frac{\left\langle x\widehat{\phi}_i(x), \widehat{\phi}_i(x) \right\rangle}{\left\langle \widehat{\phi}_i(x), \widehat{\phi}_i(x) \right\rangle}, \; C_i = \frac{\left\langle \widehat{\phi}_i(x), \widehat{\phi}_i(x) \right\rangle}{\left\langle \widehat{\phi}_i(x), \widehat{\phi}_{i-1}(x) \right\rangle}.$$

- ϕ_i has exactly i real-valued distinct zeroes

Standard examples of orthogonal polynomials

Jacobi polynomials

$$[a, b] = [-1, 1] \quad \text{and} \quad w(x) = (1-x)^{\alpha}(1+x)^{\beta}$$

with $\alpha > -1$ and $\beta > -1$

- $\alpha = \beta = 0$: Legendre polynomials.

$$(n+1)L_{n+1}(x) = (2n+1)xL_n(x) - nL_{n-1}(x).$$

- $\alpha = \beta = -\frac{1}{2}$: First-kind Chebyshev polynomials defined by

$$T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x).$$

- $\alpha = \beta = \frac{1}{2}$: Second-kind Chebyshev polynomials

Standard examples of orthogonal polynomials

Laguerre polynomials

$$[a, b[= [0, +\infty[\quad \text{and} \quad w(x) = e^{-x}$$

The recursive relation writes:

$$\mathcal{L}_{n+1}(x) = -\frac{1}{n+1}(x - 2n - 1)\mathcal{L}_n(x) - n\mathcal{L}_{n-1}(x)$$

Hermite polynomials

$$]a, b[=]-\infty, +\infty[\quad \text{and} \quad w(x) = e^{-x^2}$$

The recursive relation writes:

$$H_{n+1}(x) = 2xH_n(x) - 2nH_{n-1}(x)$$

Orthogonal polynomials - Best least squares approximation

Theorem

Let $f : [a, b] \to \mathbb{R}$ be such that $\langle f, f \rangle < \infty$ Then, for any $1 \leq k \leq (N-1)$, if $\phi_0, \phi_1, ..., \phi_k$ is a sequence of orthogonal polynomials for the inner product $\langle . \rangle$, there is a unique polynomial function P of the form

$$P(x) = a_0\phi_0(x) + a_1\phi_1(x) + ... + a_k\phi_k(x)$$

minimizing $\langle f - P, f - P \rangle$. The coefficients are given by

$$a_i = \frac{\langle \phi_i, f \rangle}{\langle \phi_i, \phi_i \rangle}, \quad i = 0, ..., k.$$

Example

Find the polynomial P of degree less or equal to 3 that minimizes

$$\int_{-1}^{1} \left(e^x - P(x)\right)^2 dx.$$

Legendre polynomials in $[-1, 1]$

We use the basis of Legendre polynomials

$$L_0(x) = 1,\ L_1(x) = x,\ L_2(x) = \frac{3}{2}\left(x^2 - \frac{1}{3}\right),\ L_3(x) = \frac{5}{2}\left(x^3 - \frac{3}{5}x\right).$$

Determining the coefficients

For $f(x) = e^x$, we compute

$$\langle f, L_0 \rangle = \int_{-1}^{1} e^x dx = e - \frac{1}{e} \quad \langle f, L_1 \rangle = \int_{-1}^{1} xe^x dx = \frac{2}{e}$$

$$\langle f, L_2 \rangle = \frac{3}{2} \int_{-1}^{1} e^x \left(x^2 - \frac{1}{3} \right) dx = e - \frac{7}{e} \quad \langle f, L_3 \rangle = \frac{5}{2} \int_{-1}^{1} e^x \left(x^3 - \frac{3}{5} \right) dx = -5e + \frac{37}{e}.$$

In addition, we have: $\langle L_i, L_i \rangle = \frac{2}{2i+1}$. Hence, we deduce the polynomial solution

$$P(x) = 1.175201194 L_0(x) + 1.10363824 L_1(x) + 0.3578143506 L_2(x) + 0.07045563367 L_3(x).$$

Remarks

- orthogonal polynomials have various mathematical properties
- they are useful for developing highly accurate numerical methods (numerical integration, spectral methods, etc.)
- beyond polynomials: trigonometric basis $(1, \cos(kx), \sin(kx))_{k>0}$ leads to Fourier series

Outline

Polynomial least square data fit

We are given a set of data $(y_i)_{1\le i\le m}$, with possibly $m \gg n$, and want to solve the minimization problem

$$\min_c \sum_{i=1}^{m} \left(y_i - \hat{f}(x_i, c)\right)^2$$

where the unknowns are the coefficients $(c_j)_{1\le j\le n}$ from a polynomial basis

$$\hat{f}(x, c) = \sum_{j=1}^{n} c_j \phi_j(x)$$

We try the monomial basis such that $\phi_j(x) = x^{(j-1)}$, $j \in [1, n]$. The polynomial degree n should be large enough to contain information but not too large to avoid spurious noise.

Example

Linear regression with $\hat{f}(x, c) = c_1 x + c_2$

Remarks

- if $m = n$, we have interpolation
- we study the discrete analogue of the continuous least square problem
- there are other choices to measure the error, based on different distances
  - the maximum norm:
$$\max_{0 \leq i \leq m} |y_i - \sum_{1 \leq j \leq n} c_j \phi_j(x_i)|.$$
  - the absolute value norm:
$$\sum_{0 \leq i \leq m} |y_i - \sum_{1 \leq j \leq n} c_j \phi_j(x_i)|.$$

  This will lead different solutions for the $(c_j)_{1 \leq j \leq n}$. We focus on the Euclidean distance because it leads a linear problem.

The problem for the p-norm writes

$$\min_c \|Ac - y\|_p^2$$

with $(A_{ij}) = \phi_j(x_i)$ is a $m \times n$ matrix, sampling the data on the chosen polynomial basis.

Normal equations

$$E(c) = \sum_{0 \le i \le m} (y_i - \sum_{1 \le j \le n} c_j \phi_j(x_i))^2.$$

with $c = (c_1, c_2, ..., c_n)$. if E has a minimum in c, then

$$\frac{\partial E}{\partial c_i}(c) = 0, \quad \forall\ i = 1, ..., m$$

which writes

$$2 \sum_{0 \le i \le m} \phi_j(x_i) \left(y_i - (c_1 \phi_1(x_i) + ... + c_n \phi_n(x_i))\right) = 0, \forall\ i = 1, ..., m$$

which are called the normal equations. In matrix form, we have

$$2A^T(Ac - y) = 0 \Leftrightarrow A^T A c = A^T y$$

Remarks

- the matrix $A^T A$ is of size $n \times n$
- if A has full column rank, $A^T A$ is symmetric positive definite and we have a unique minimizer
- A is usually badly conditioned
- The number of data m might be very large and/or noisy
- The method is not restricted to a polynomial basis
- There are various ways to solve the normal equation (see Chapter on linear systems)

Summary of the contents

Polynomial interpolation :

1. Global polynomial interpolation (Lagrange interpolant)
   Error study: Runge phenomena, Chebychev points
2. Local interpolation follows the same ideas (see exercises)
   Higher degree local interpolation requires more data (new point, derivative(s), ...)
   $\Rightarrow$ more unknowns to be solved

Approximation theory :

1. Orthogonal polynomials are an important building block for advanced numerical methods
2. Discrete least-squares is an example of convex optimization, and is a starting point for advanced topics (non-linearity, machine learning, , etc.)