Numerical Analysis Hilary Term 2026
Lecture 1: Lagrange Interpolation

Numerical analysis is the study of computational algorithms for solving problems in scientific computing. It combines mathematical beauty, rigor and numerous applications; we hope you'll enjoy it! In this course we will cover the basics of three key fields in the subject:

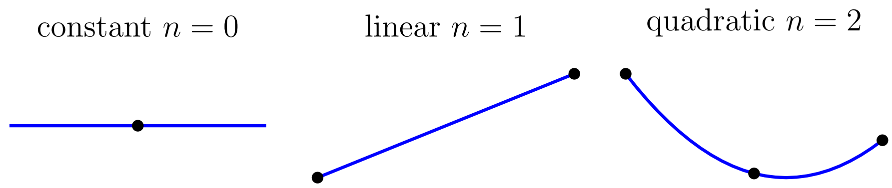
- Approximation Theory (lectures 1, 9–11); recommended reading: L. N. Trefethen, Approximation Theory and Approximation Practice, and E. Süli and D. F. Mayers, An Introduction to Numerical Analysis.
- Numerical Linear Algebra (lectures 2–8); recommended reading: L. N. Trefethen and D. Bau, Numerical Linear Algebra.
- Numerical Solution of Differential Equations (lectures 12–16); recommended reading: E. Süli and D. F. Mayers, An Introduction to Numerical Analysis.

This first lecture comes from Chapter 6 of Süli and Mayers, and Chapter 10+15 of Trefethen.

Notation: $\Pi_n = \{\text{real polynomials of degree} \leq n\}$

Setup: Given data f_i at distinct x_i, $i = 0, 1, \ldots, n$, with $x_0 < x_1 < \cdots < x_n$, can we find a polynomial p_n such that $p_n(x_i) = f_i$? Such a polynomial is said to **interpolate** the data, and (as we shall see) can approximate f at other values of x if f is smooth enough. This is the most basic question in approximation theory.

E.g.:



Theorem. Let $x_0 < x_1 < \cdots < x_n$, and $f_i \in \mathbb{R}$ for $i = 0, \ldots, n$. There exists a polynomial interpolant $p_n \in \Pi_n$ such that $p_n(x_i) = f_i$ for $i = 0, 1, \ldots, n$.

Proof. Consider, for $k = 0, 1, \ldots, n$, the "cardinal polynomial" (aka Lagrange basis polynomial)

$$L_{n,k}(x) = \frac{(x - x_0)\cdots(x - x_{k-1})(x - x_{k+1})\cdots(x - x_n)}{(x_k - x_0)\cdots(x_k - x_{k-1})(x_k - x_{k+1})\cdots(x_k - x_n)} \in \Pi_n. \tag{1}$$

Then $L_{n,k}(x_i) = \delta_{ik}$, that is,

$$L_{n,k}(x_i) = 0 \;\text{ for }\; i = 0, \ldots, k-1, k+1, \ldots, n \;\text{ and }\; L_{n,k}(x_k) = 1.$$

So now define

$$p_n(x) = \sum_{k=0}^{n} f_k L_{n,k}(x) \in \Pi_n \tag{2}$$

$\Longrightarrow$

$$p_n(x_i) = \sum_{k=0}^{n} f_k L_{n,k}(x_i) = f_i \text{ for } i = 0, 1, \ldots, n. \qquad \square$$

The polynomial (2) is the **Lagrange interpolating polynomial**.

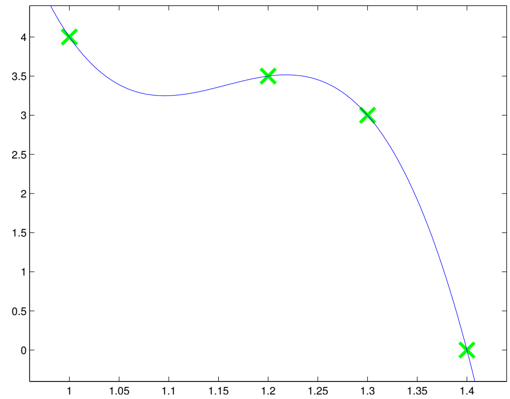
Theorem. The interpolating polynomial p_n of degree $\leq n$ is unique.

Proof. Consider two interpolating polynomials $p_n, q_n \in \Pi_n$. Their difference $d_n = p_n - q_n \in \Pi_n$ satisfies $d_n(x_k) = 0$ for $k = 0, 1, \ldots, n$. i.e., d_n is a polynomial of degree at most n but has at least $n + 1$ distinct roots. Fundamental theorem of algebra $\Longrightarrow$ $d_n \equiv 0 \Longrightarrow p_n = q_n$. $\square$

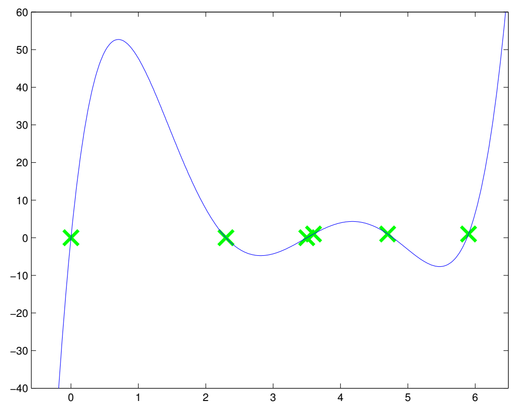
Matlab:

```
>> help lagrange
 LAGRANGE  Plots the Lagrange polynomial interpolant for the
 given DATA at the given NODES/interpolation points

>> lagrange([1,1.2,1.3,1.4],[4,3.5,3,0]);
```




```
>> lagrange([0,2.3,3.5,3.6,4.7,5.9],[0,0,0,1,1,1]);
```



Data from an underlying smooth function: Suppose that $f(x)$ has at least $n+1$ smooth derivatives in the interval (x_0, x_n). Let $f_k = f(x_k)$ for $k = 0, 1, \ldots, n$, and let p_n be the Lagrange interpolating polynomial for the data (x_k, f_k), $k = 0, 1, \ldots, n$.

Error: How large can the error $f(x) - p_n(x)$ be on the interval $[x_0, x_n]$?

Theorem. Suppose that f is $(n+1)$-times continuously differentiable. Let p_n be the polynomial interpolant at $\{x_i\}_{i=0}^n$. For every $x \in [x_0, x_n]$ there exists $\xi = \xi(x) \in (x_0, x_n)$ such that

$$e(x) \stackrel{\text{def}}{=} f(x) - p_n(x) = (x - x_0)(x - x_1) \cdots (x - x_n) \frac{f^{(n+1)}(\xi)}{(n+1)!}, \tag{3}$$

where $f^{(n+1)}$ is the $(n+1)$-st derivative of f.

Proof. Trivial for $x = x_k$, $k = 0, 1, \ldots, n$ as $e(x) = 0$ by construction. So suppose $x \neq x_k$. Let

$$\phi(t) \stackrel{\text{def}}{=} e(t) - \frac{e(x)}{\pi(x)} \pi(t),$$

where

$$\begin{aligned} \pi(t) &\stackrel{\text{def}}{=} (t - x_0)(t - x_1) \cdots (t - x_n) \\ &= t^{n+1} - \left(\sum_{i=0}^{n} x_i \right) t^n + \cdots (-1)^{n+1} x_0 x_1 \cdots x_n \\ &\in \Pi_{n+1}. \end{aligned}$$

Now note that ϕ vanishes at $n+2$ points x and x_k, $k = 0, 1, \ldots, n$. $\Longrightarrow$ ϕ' vanishes at $n+1$ points $\xi_0, \ldots, \xi_n$ between these points $\Longrightarrow$ ϕ'' vanishes at n points between these new points, and so on until $\phi^{(n+1)}$ vanishes at an (unknown) point ξ in (x_0, x_n). But

$$\phi^{(n+1)}(t) = e^{(n+1)}(t) - \frac{e(x)}{\pi(x)} \pi^{(n+1)}(t) = f^{(n+1)}(t) - \frac{e(x)}{\pi(x)} (n+1)!$$

since $p_n^{(n+1)}(t) \equiv 0$ and because $\pi(t)$ is a monic polynomial of degree $n+1$. The result then follows immediately from this identity since $\phi^{(n+1)}(\xi) = 0$.

□

The above proof may seem ingenious/mysterious. It is perhaps helpful to observe the connections and similarity to Taylor's theorem with remainder, and its proof. Indeed the latter can be seen as a special case of the theorem when x_i all tend to a single point $x_i \to x_*$ and one interpolates high-order derivatives at a single point.

Example: $f(x) = \log(1+x)$ on $[0, 1]$. Here, $|f^{(n+1)}(\xi)| = n!/(1+\xi)^{n+1} < n!$ on $(0, 1)$. So $|e(x)| < |\pi(x)| n!/(n+1)! \le 1/(n+1)$ since $|x - x_k| \le 1$ for each x, x_k, $k = 0, 1, \ldots, n$, in $[0, 1] \Longrightarrow |\pi(x)| \le 1$. This is probably pessimistic for many x, e.g. for $x = \frac{1}{2}$, $\pi(\frac{1}{2}) \le 2^{-(n+1)}$ as $|\frac{1}{2} - x_k| \le \frac{1}{2}$.

This shows the important fact that the error can be large at the end points when the interpolation points $\{x_k\}$ are equispaced, an effect known as the "Runge phenomena" (Carl Runge, 1901), which we see below and return to in lecture 4. More generally, as the expression (3) suggests, the location of the samples $\{x_k\}$ is very important, as we see at the end.

Generalisation: Given data f_i and g_i at distinct x_i, $i = 0, 1, \ldots, n$, with $x_0 < x_1 < \cdots < x_n$, can we find a polynomial p such that $p(x_i) = f_i$ and $p'(x_i) = g_i$? (i.e., interpolate derivatives in addition to values)

Theorem. There is a unique polynomial $p_{2n+1} \in \Pi_{2n+1}$ such that $p_{2n+1}(x_i) = f_i$ and $p'_{2n+1}(x_i) = g_i$ for $i = 0, 1, \ldots, n$.

Construction: Given $L_{n,k}(x)$ in (1), let

$$\begin{aligned} H_{n,k}(x) &= [L_{n,k}(x)]^2(1 - 2(x - x_k)L'_{n,k}(x_k)) \\ \text{and} \quad K_{n,k}(x) &= [L_{n,k}(x)]^2(x - x_k). \end{aligned}$$

Then

$$p_{2n+1}(x) = \sum_{k=0}^{n} [f_k H_{n,k}(x) + g_k K_{n,k}(x)] \tag{4}$$

interpolates the data as required. The polynomial (4) is called the **Hermite interpolating polynomial**. Note that $H_{n,k}(x_i) = \delta_{ik}$ and $H'_{n,k}(x_i) = 0$, and $K_{n,k}(x_i) = 0$, $K'_{n,k}(x_i) = \delta_{ik}$.

Theorem. Let p_{2n+1} be the Hermite interpolating polynomial in the case where $f_i = f(x_i)$ and $g_i = f'(x_i)$ and f has at least $2n{+}2$ smooth derivatives. Then, for every $x \in [x_0, x_n]$,

$$f(x) - p_{2n+1}(x) = [(x - x_0)(x - x_1) \cdots (x - x_n)]^2 \frac{f^{(2n+2)}(\xi)}{(2n+2)!},$$

where $\xi \in (x_0, x_n)$ and $f^{(2n+2)}$ is the $(2n+2)$nd derivative of f.

Proof (non-examinable): see Süli and Mayers, Theorem 6.4. □

Chebyshev interpolation (New 2026:) Let us return to interpolating the function values f_i and not the derivative. The error expression (3) is exact, but does not tell us how best (or near-best) to choose the interpolation points $\{x_i\}_{i=0}^n$ (assuming we get to choose them). One might expect the choice depends strongly on f in a complicated fashion. Fortunately, this is far from the truth! We now introduce Chebyshev interpolation, a cornerstone tool in numerical analysis. In brief, taking $\{x_i\}_{i=0}^n$ to be the *Chebyshev points*[1]

$$x_i = \cos\theta_i, \quad \theta_i = \frac{(2i+1)\pi}{2(n+1)}, \qquad i = 0, \ldots, n, \tag{5}$$

results in p_n being a near-optimal approximation to f in the sense of (8) with $\Lambda_n = O(\log n)$, *regardless* of f. The Chebyshev points x_i are displayed below for $n = 50$. Note the points cluster near the endpoints ± 1.

[1] These are the roots of T_{n+1}, the $(n{+}1)$st Chebyshev polynomial, to be introduced in the lecture on orthogonal polynomials. Another common choice is $\theta_i = \frac{i\pi}{n}$, also called Chebyshev points. This set gives essentially the same Lebesgue constant (7), i.e., $\Lambda_n \approx \log n$. The crucial point is that they cluster x_i near the endpoints ± 1 following a cos-distribution; we will see this effect again with the nodes in Gauss quadrature, Lecture 11.

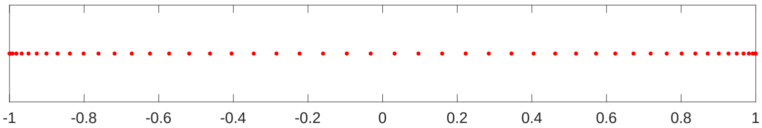


Figure 1: Chebyshev points (5) for $n = 50$.

To make this precise, we prepare two definitions.

Best polynomial approximation: Given a continuous function[2] $f \in C([-1,1])$, a natural question is: what is the optimal approximation $p \in \Pi_n$ to f ? The answer depends on the norm (or metric) used. In the L_∞ norm $\|\cdot\|_\infty$ defined by $\|f\|_\infty = \sup_{x\in[-1,1]} |f|$, the solution exists and is unique (see Trefethen Thm. 10.1) for any $f \in C([-1,1])$. We denote it by p_n^*, that is,

$$\|f - p_n^*\|_\infty = \min_{p_n \in \Pi_n} \|f - p_n\|_\infty. \tag{6}$$

In Sheet 3 you'll explore some pretty properties of p_n^*. For now, note that $\|f - p_n^*\|_\infty \to 0$ as $n \to \infty$ for any continuous f, and be aware that the smoother the f is, the faster the convergence; e.g. if f is holomorphic=analytic on $[-1,1]$, the convergence is exponential $\|f - p_n^*\|_\infty \le C\exp(-cn)$ for some positive constants c and C.

Lebesgue constant: Let $\mathcal{L}_n$ be the interpolation operator that outputs the interpolant p_n given f; that is, $\mathcal{L}_n f = p_n$. Note that $\mathcal{L}_n$ is well-defined once $\{x_i\}_{i=0}^n$ are chosen.

The Lebesgue constant Λ_n is defined[3] as the ∞-norm of the operator $\mathcal{L}_n$:

$$\Lambda_n = \sup_{f\in C([-1,1])} \frac{\|\mathcal{L}_n f\|_\infty}{\|f\|_\infty} \left(= \sup_{f\in C([-1,1])} \frac{\|p_n\|_\infty}{\|f\|_\infty} \right) \tag{7}$$

Here is the interpretation: Given data values on an $(n+1)$-point grid coming from sampling a function with $\|f\|_\infty = 1$, Λ_n is the largest possible value of the polynomial interpolant.

The significance of Λ_n is the following:

Theorem. Let f be continuous $f \in C([-1,1])$[4], and let p_n be the polynomial interpolant of f at $\{x_i\}_{i=0}^n$. With Λ_n as defined in (7),

$$\|f - p_n\|_\infty \le (\Lambda_n + 1)\|f - p_n^*\|_\infty. \tag{8}$$

Proof. We have $\|f - p_n\|_\infty \le \|f - p_n^*\|_\infty + \|p_n^* - p_n\|_\infty$. The key step is to note that $p_n^* - p_n = \mathcal{L}_n(p_n^* - p_n) = \mathcal{L}_n(p_n^* - f)$, because $p_n^* - p_n \in \Pi_n$, so $\mathcal{L}_n(p_n^* - p_n) = p_n^* - p_n$. Also $\mathcal{L}_n f = p_n$ by definition. Hence $\|p_n^* - p_n\|_\infty \le \|\mathcal{L}_n(p_n^* - f)\|_\infty \le \Lambda_n\|f - p_n^*\|_\infty$ by the definition of Λ_n.

[2] We work with the interval $[-1,1]$ for simplicity. One can use an affine transformation to deal with a general interval $[a,b]$.

[3] Let us repeat that Λ_n depends on $\{x_i\}_{i=0}^n$, not just n.

[4] We actually don't even need f to be continuous for the proof to hold, but then $f - p_n^*$ wouldn't go to 0 as $n \to \infty$, so this is a sensible assumption.

□

Also, there exist functions $f \in C[-1,1]$ such that $\|f-p_n\|_\infty > (\Lambda_n - \epsilon - 1)\|f-p_n^*\|_\infty$ for any $\epsilon > 0$ (consider an example where $\|p_n\|_\infty > (\Lambda_n - \epsilon)\|f\|_\infty$). The crux is that iff the Lebesgue constant is modest (say $O(1)$; note that $\Lambda_n \geq 1$ trivially), the interpolant p_n is guaranteed to be near best.

The Lebesgue constant can be characterized using Lagrange basis polynomials $L_{n,k}(x)$ as (Sheet 1)

$$\Lambda_n = \max_{x\in[-1,1]} \sum_{k=0}^{n} |L_{n,k}(x)|. \tag{9}$$

$\sum_{k=0}^{n} |L_{n,k}(x)|$ is called the Lebesgue function associated with the set of interpolation points $\{x_i\}_{i=0}^n$.

The remainder is non-examinable.

Equispaced points Figure 2 shows the Lebesgue function $\sum_{k=0}^{n} |L_{n,k}(x)|$ for $n = 30$.

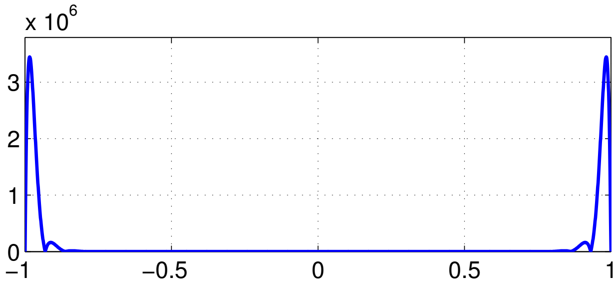


Figure 2: Lebesgue function for 30 equispaced points. $\Lambda_n \approx 2^n$.

Observe how large it is near $x = \pm 1$; in fact Λ_n grows like 2^n (more precisely $\frac{2^n}{n \log n}$). A famous example where interpolation in equispaced points results in $\|p_n - f\|_\infty$ diverging as $n \to \infty$, even though f is analytic, is $f(x) = \frac{1}{1+25x^2}$. This is Runge's phenomenon; see Figure 5.

Chebyshev points Now with Chebyshev points, the Lebesgue constant is fantastically small with $\Lambda_n = O(\log(n))$. Here is $\sum_{k=0}^{n} |L_{n,k}(x)|$ for $n = 30$.

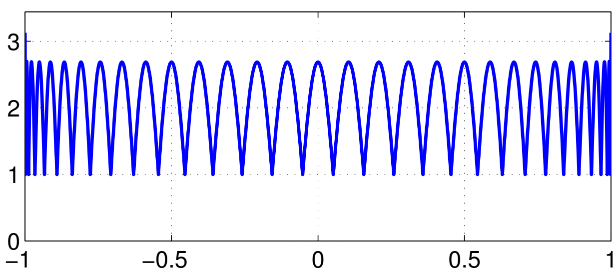


Figure 3: Lebesgue function for $n = 30$ Chebyshev points. $\Lambda_n \approx \frac{2}{\pi} \log n$.

Observe how small Λ_n is with Chebyshev interpolation and the equioscillation-like behavior of the Lebesgue function. Note that $\sum_{k=0}^{n} |L_{n,k}(x)| = 1$ at $x = x_i$.

Examples We compare four methods for polynomial approximation: (i) piecewise-linear approximation, (ii) equispaced interpolation, (iii) Chebyshev interpolation, and (iv) best polynomial approximation[5]. We take two representative functions: the exponential function $f(x) = \exp(x)$ (which is entire), and the Runge function $f(x) = \frac{1}{1+25x^2}$.

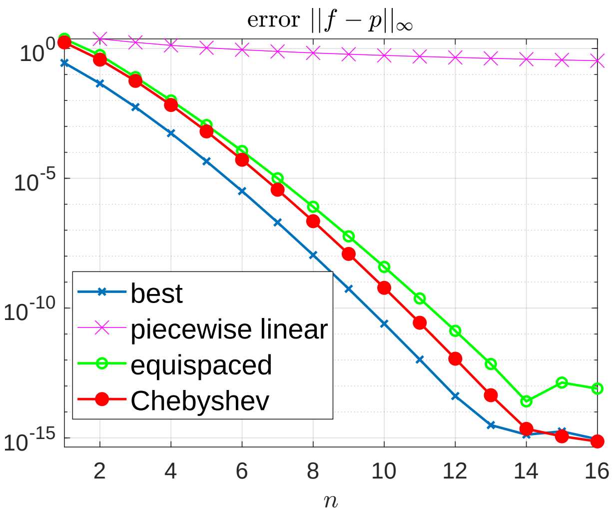


Figure 4: Exponential function $f(x) = \exp(x)$.

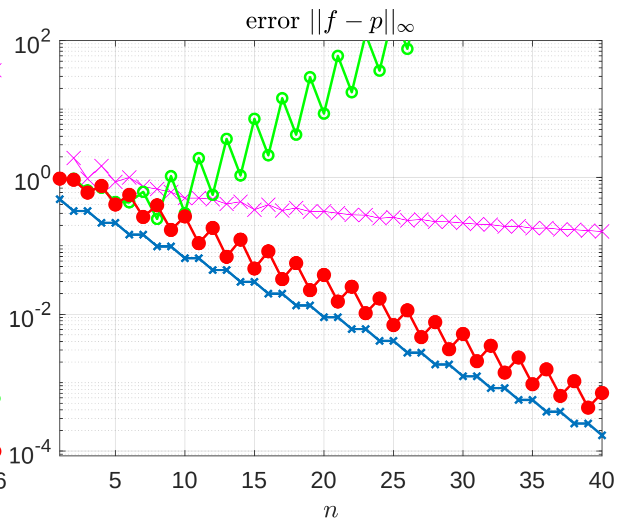


Figure 5: Runge function $f(x) = \frac{1}{1+25x^2}$.

Unlike the best approximation p_n^*, Chebyshev interpolation is linear in f (as is any interpolation once $\{x_i\}_{i=0}^n$ are fixed), and can be computed quickly in $O(n \log n)$ operations, using the FFT. The combination of near-optimal speed and convergence (and numerical stability) makes Chebyshev interpolation a truly remarkable algorithm.

Chebyshev interpolation (where one represents p_n in the Chebyshev polynomial basis) is actually equivalent to Fourier series for functions that are even and periodic on $[-\pi, \pi]$ under the change of variables $x = \cos(\theta)$.

[5]These are easily computed in MATLAB using the Chebfun toolbox `https://www.chebfun.org/`, namely `p=chebfun(f,n+1)` for Chebyshev interpolation, and `p=minimax(f,n)` for best approximation.