Math 514

November 2, 2020

- Instructor: Chenxi Wu cwu367@wisc.edu
- Section 1: 9:55-10:45 am Section 2: 12:05-12:55 pm
- Office hours: 10:45am-noon Monday, Wednesday or by appointment

Recall that in the first half of the semester, we covered the following topics:

(i) Methods for root finding: Newton's method etc.

(ii) Numerical Linear Algebra: LU decomposition, QR algorithm etc.

The following are the new topics we will cover in the second half of the semester:

(i) Interpolation and approximation: how to get the formula of a function using discrete date.

(ii) Numerical integration: how to integrate a function knowing only its value on a discrete set.

(iii) Numerical solution for differential equations: numerical solution for ODE and PDE.

The first topic will be the foundation of the second and third topic.

1 Polynomial interpolation (Chapter 6)

Definition 1.1. *The* **polynomial interpolation** *of a function f at points $x_0, \ldots x_n$, is a polynomial p that shares some properties of f at those points, e.g. has the same value or same derivatives.*

We will focus on single variable functions, and discuss two kinds of polynomial interpolation problems:

(i) Lagrange interpolation: find a polynomial p of degree at most n, such that $p(x_i) = f(x_i)$.

(ii) Hermite interpolation: find a polynomial p of degree at most $2n+1$, such that $p(x_i) = f(x_i)$, $p'(x_i) = f'(x_i)$.

An application for Hermite interpolation is for approximating smooth curves where the direction of the curve at certain points are also specified.

1.1 Lagrange interpolation

1.1.1 Existence

Theorem 1.2. *(Lemma 6.1 in textbook) The Lagrange interpolation polynomial exists. In other words, for any $n+1$ distinct real numbers $x_0, \ldots, x_n$, and $n+1$ real numbers $y_0, \ldots, y_n$, there is a polynomial p of degree at most n such that $p(x_i) = y_i$.*

How do we find such a p?

Firstly, we observe that the map

$$T : p \mapsto [p(x_0), \ldots p(x_n)]^T \in \mathbb{R}^{n+1}$$

is linear. In other words, $(cp + dq)(x_i) = cp(x_i) + dq(x_i)$ for all i. Hence, finding p is like solving a system of non homogenous linear equations. Recall from linear algebra, let e_i be the standard basis vector of $\mathbb{R}^{n+1}$ corresponding to $y_i = 1$ and $y_j = 0$ for all $j \neq i$, then, if we can find some p_i of degree at most n such that $T(p_i) = e_i$, (i.e. $p_i(x_j) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$) then

$$T(\sum_i y_i p_i) = \sum_i y_i e_i = [y_0, \ldots y_n]^T \ .$$

Now we try and find the p_i: $p_i(x_j) = 0$ for all $j \neq i$, so $(x - x_j)$ must be a factor of p_i. So p_i must be something times

$$\prod_{j \neq i} (x - x_j) \ .$$

On the other hand, $p_i(x_i) = 1$, so the "something" should be

$$\frac{1}{\prod_{j \neq i} (x_i - x_j)} \ .$$

Now we have a proof of Theorem 1:

Proof. Let

$$p(x) = \sum_i \left(y_i \cdot \frac{\prod_{j \neq i} (x - x_j)}{\prod_{j \neq i} (x_i - x_j)} \right)$$

Then by calculation, $p(x_i) = y_i$. □

1.1.2 Uniqueness

The problem of finding Lagrange interpolation polynomial is one with $n+1$ conditions and $n+1$ unknowns, so intuitively there should be a discrete set of solutions. Actually the solution can be shown to be unique:

Theorem 1.3. *(Theorem 6.1 in textbook) The Lagrange interpolation polynomial is unique. In other words, given $x_0, \dots, x_n$ and $y_0, \dots, y_n$ with x_i distinct, there is a single polynomial p of degree at most n such that $p(x_i) = y_i$*

Proof. The first step of the proof is to reduce the problem to the case where $y_i = 0$. Suppose p and q are two such polynomials, then $(p - q)(x_i) = 0$ for all i. So, to prove the theorem, we only need to show that if a polynomial $r = p - q$ of degree at most n vanishes at $n+1$ distinct points, then $r = 0$. This fact follows from the "fundamental theorem of algebra" and can be proved using long division. Here we provide two other alternative proofs:

(i) Approach I: use the mean value theorem in calculus. Firstly we show the following fact:

Lemma 1.4. *If $f \in C^{m-1}$ (f is $m-1$-th order differentiable with $m-1$-th derivative continuous), and $f = 0$ at m distinct points, then for any $0 \leq k \leq m-1$, $f^{(k)} = 0$ at at least $m-k$ points.*

Proof. By mean value theorem, between two consecutive zeros of f there must be a zero of f'. Hence f' vanishes at at least $m-1$ points. Now let f' take the role of f and continue the process, we get f'' vanishes at at least $m-2$ points, etc. $\square$

Suppose r vanishes at $x_0, \ldots, x_n$, and r is of degree $d > 0$. Then $r^{(d)}$ is a non zero constant. Apply Lemma **??** with $m = n+1$ and $k = d$, we see a contradiction. Hence $r = const$, which implies $r = 0$.

(ii) Approach II: Let $r = \sum_j a_j x^j$, then a_j are solutions of a system of linear equation $\sum_j a_j x_i^j = 0$. However from linear algebra,

$$\begin{vmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_{n+1} \\ \dots & \dots & \dots & \dots \\ x_1^n & x_2^n & \dots & x_{n+1}^n \end{vmatrix} = \prod_{i<j}(x_j - x_i) \neq 0$$

Hence $a_j = 0$ for all j, which implies that $r = 0$.

$\square$

1.1.3 Error Estimate

We know that $f^{(n+1)} = 0$ iff f is a polynomial of degree at most n, so one may guess that if $f^{(n+1)}$ is small, f should be close to a polynomial of degree at most n, hence probably close to its Lagrange interpolation polynomial at $n+1$ points. To make this more precise, we have the following theorem on error estimate of Lagrange interpolation:

Theorem 1.5. *(Theorem 6.2 in textbook) If $f \in C^{n+1}$, p is the Lagrange interpolation of f at $n+1$ distinct points $x_0, \ldots x_n$. Then for any x, there is some $s \in [\min\{x_i, x\}, \max\{x_i, x\}]$, such that*

$$f(x) - p(x) = \frac{f^{(n+1)}(s) \prod_i (x - x_i)}{(n+1)!}$$

Proof. When $x = x_i$ it's obvious. Now suppose x is distinct from all x_i. Consider the auxiliary function:

$$G(t) = f(t) - p(t) - (f(x) - p(x)) \cdot \frac{\prod_i (t - x_i)}{\prod_i (x - x_i)}$$

Then $G = 0$ at x_i and x, hence by Lemma **??** (let $m = n + 2$, $k = n + 1$), there must be some point $s \in [\min\{x_i, x\}, \max\{x_i, x\}]$ where

$$G^{(n+1)}(s) = f^{(n+1)}(s) - \frac{(f(x) - p(x))(n+1)!}{\prod_i (x - x_i)} = 0$$

Hence

$$f(x) - p(x) = \frac{f^{(n+1)}(s) \prod_i (x - x_i)}{(n+1)!}$$

$\square$

When the set $\{x_i\}$ becomes denser, $\prod_i (x - x_i)$ decreases, and $(n+1)!$ increases. However, when $n \to \infty$, the Lagrange interpolation polynomial may not converge to f even if f is smooth, if $f^{(n)}$ increases too fast.

Example 1.6. $f(x) = \cos(x)$, $x_i = 5i/n$, $i = 0, 1, 2, \ldots n$

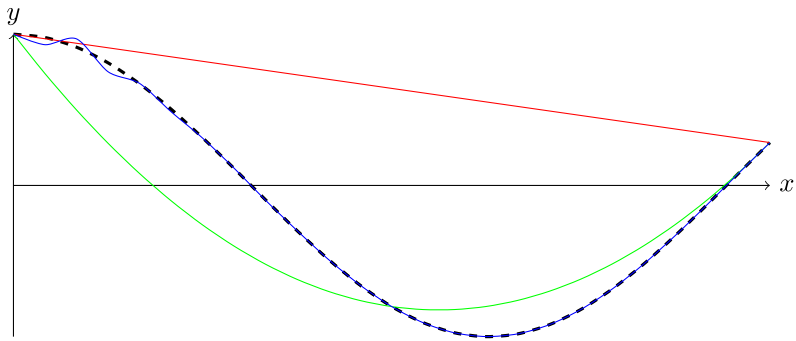


Figure 1: Black dashed line: $y = \cos(x)$. Red line: Lagrange interpolation with 2 points. Green line: Lagrange interpolation with 3 points. Blue line: Lagrange interpolation with 11 points.

Example 1.7. $f(x) = 1/(1+2(x-2)^2)$, $x_i = 5i/n$, $i = 0, 1, 2, \dots n$.

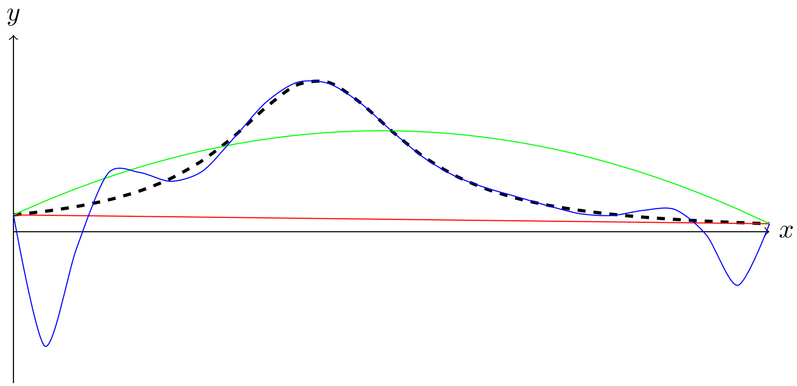


Figure 2: Black dashed line: $y = 1/(1+2(x-2)^2)$. Red line: Lagrange interpolation with 2 points. Green line: Lagrange interpolation with 3 points. Blue line: Lagrange interpolation with 11 points.

The reason that the Lagrange interpolation polynomials in Example **??** converges but those in Example **??** don't, is that the higher order derivatives of cos is $\pm\sin$, $\pm\cos$ hence all bounded, while it is not true for the function in Example **??**. As a practice, calculate the k-th derivative of $1/(1+2(x-2)^2)$ at $x = 2$.

1.2 Hermite interpolation polynomial

1.2.1 Existence

Similar to the Lagrange case, we can construct the Hermite interpolation polynomial as follows:

Theorem 1.8. *(Existence part of Theorem 6.3 in textbook) There is a polynomial p of degree at most $2n+1$, such that $p(x_i) = y_i$, $p'(x_i) = z_i$, $i = 0, \dots n$, where x_i are distinct.*

Use the same strategy as the Lagrange case, possibly via a few trials and errors, one can find the formula of p as below:

Proof. Let

$$p(x) = \sum_i \left(z_i \cdot \frac{(x - x_i) \prod_{j \neq i} (x - x_j)^2}{\prod_{j \neq i} (x_i - x_j)^2} \right.$$

$$\left. + y_i \cdot \left(1 - (x - x_i) \sum_{j \neq i} \frac{2}{x_i - x_j} \right) \cdot \frac{\prod_{j \neq i} (x - x_j)^2}{\prod_{j \neq i} (x_i - x_j)^2} \right)$$

Then by calculation, $p(x_i) = y_i$, $p'(x_i) = z_i$. $\square$

1.2.2 Uniqueness and Error Estimate

The mean value theorem argument (i.e. Lemma **??**) can also be used to show the uniqueness and error estimate for Hermite interpolation polynomials:

Theorem 1.9. *(Uniqueness part of Theorem 6.3 in textbook) The Hermite interpolation polynomial is unique. In other words, there is a unique p of degree at most $2n+1$ such that $p(x_i) = y_i$, $p'(x_i) = z_i$, $i = 0, \ldots n$, where x_i are distinct.*

Proof. Similar to the proof of Theorem **??**, if we have two Hermite interpolation polynomials p and q, then $r = p - q$ satisfies $r(x_i) = r'(x_i) = 0$ and r has degree at most $2n + 1$. However, if r is non zero, it can not have $n+1$ distincts roots x_i with multiplicity at least 2 each, hence $r = 0$.

We can also prove $r = 0$ using analysis like in Theorem **??**. If r has degree at most $2n+1$, $r' = r = 0$ at $n+1$ points, then there must be n other points where $r' = 0$. Now suppose r has degree $d > 0$. Apply Lemma **??**, let $m = 2n+2$, $k = d$, then we get $r^{(d)}$ vanishes at $2n + 2 - d$ points, which contradicts with the fact that $r^{(d)}$ is a non zero constant. Hence $r = const$ which implies that $r = 0$. $\square$

Theorem 1.10. *(Theorem 6.4 in textbook) If $f \in C^{(2n+2)}$, there is $s \in [\min\{x_i, x\}, \max\{x_i, x\}]$, such that*

$$f(x) - p(x) = \frac{f^{(2n+2)}(s) \prod_i (x - x_i)^2}{(2n+2)!}$$

Proof. If $x = x_i$ then it is trivially true. Now assume x is not in $\{x_i\}$. Let

$$G(t) = f(t) - p(t) - \frac{(f(x) - p(x)) \prod_i (t - x_i)^2}{\prod_i (x - x_i)^2}$$

Then G vanishes at the $n+2$ points $x, x_0, \ldots x_n$, and G' vanishes at $n+1$ of them $x_0, \ldots, x_n$. By the same argument as above, G' vanishes at $n+1$ more points, hence it is zero at at least $2n+2$ points. Now use Lemma **??** on G' for $m = 2n+2$, $k = 2n+1$. □

1.3 Applications

1.3.1 Numerical Differentiation

Suppose p is the Lagrange interpolation of f at $n+1$ points. By mean value theorem, $f'-p'$ is zero at n points $d_1, \dots d_n$, so p' can be seen as the Lagrange interpolation polynomial with condition $p'(d_i) = f'(d_i)$ (see Theorem 6.5 in textbook). Now one can get an estimate for $f'(x) - p'(x)$ using Theorem **??**.

Example 1.11. *For example, if we know the value of f at $x + ih$ for $i = -1, 0, 1$ as y_{-1}, y_0 and y_1, then the Lagrange interpolation polynomial is:*

$$p(x+t) = y_{-1}t(t-h)/(2h^2) - y_0(t+h)(t-h)/h^2 + y_1 t(t+h)/(2h^2)$$

So

$$p'(0) = \frac{y_1 - y_{-1}}{2h} = \frac{f(x+h) - f(x-h)}{2h}$$

As $h \to 0$ this indeed converges to $f'(x)$.

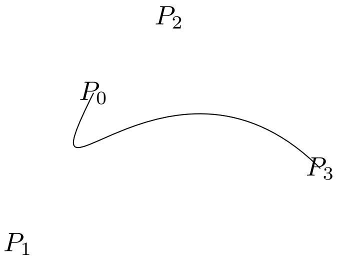
However, this approach is generally unstable. If f is complex analytic one can use complex analysis to do it which is stable, which we will not cover in this class.

Numerical differentiation is useful in optimization or root finding via Newton's method.

1.3.2 Cubic Bézier curves

The **cubic Bézier curve** is a curve parametrized by cubic functions: $\gamma : [0,1] \to \mathbb{R}^2$, $\gamma(t) = (\gamma_1(t), \gamma_2(t))$, where γ_1 and γ_2 are both of degree at most 3, and $\gamma(0) = P_0$, $\gamma'(0) = 3(P_1 - P_0)$, $\gamma(1) = P_3$, $\gamma'(1) = 3(P_3 - P_2)$, where P_0, P_1, P_2 and P_3 are the four "control points".

To find the formula for cubic Bézier curve, we can apply the formula for Hermite interpolation polynomial for $n = 1$. Bézier curves has many applications in computer graphics and font design, and you might have already used it in applications that generate or edit vector graphics. Below is an example (drawn using LaTeX/TikZ):



1.3.3 Linear and Hermite splines (Chapter 11)

From the Example 2 above we see that polynomial interpolation with high degree is not guaranteed to work well. Hence, in practice, we often try to keep the degree of the polynomial low, which means that we will need to use piecewise functions for interpolation. We will discuss two kinds of piecewise polynomial interpolation: **linear spline** and **Hermite cubic spline**. The textbook also covered the **natural cubic spline**.

Linear Spline

Definition 1.12. *Let f be a single variable function on $[a, b]$, $a = x_0 < x_1 \cdots < x_n = b$ $n+1$ distinct points. The* **Linear Spline** s_L *with* **knots** *at x_i is defined as*

$$s_L(x) = \frac{x_i - x}{x_i - x_{i-1}} f(x_{i-1}) + \frac{x - x_{i-1}}{x_i - x_{i-1}} f(x_i), \ \textit{where } x_{i-1} \le x \le x_i$$

In other words, use the 2-point Lagrange interpolation for each interval $[x_{i-1}, x_i]$.

Theorem 1.13. *(Theorem 11.1 in textbook) Let $f \in C^2$, $h = \max\{x_i - x_{i-1}\}$, $M = \max|f''|$, then for any $x \in [a,b]$, $|f(x) - s_L(x)| \leq \frac{1}{8}h^2 M$.*

Proof. Suppose x is between x_{i-1} and x_i. Theorem **??** implies that

$$f(x) - s_L(x) = \frac{f''(c)(x - x_{i-1})(x - x_i)}{2!}$$

for some $c \in [x_{i-1}, x_i]$. From assumption, $|f''(c)| < M$ and

$$|(x - x_{i-1})(x - x_i)| \leq |(x_i - x_{i-1})/2|^2 \leq h^2/4 \ .$$

□

The linear spline formula can be alternatively written as $s_L = \sum_i f(x_i)\phi_i$, where ϕ_i are the "hat functions", where, if $i = 1, \ldots, n-1$,

$$\phi_i(x) = \begin{cases} (x - x_{i-1})/(x_i - x_{i-1}) & x \in [x_{i-1}, x_i] \\ (x - x_{i+1})/(x_i - x_{i+1}) & x \in [x_i, x_{i+1}] \\ 0 & otherwise \end{cases}$$

ϕ_0 and ϕ_n can be written down similarly. As a consequence, s_L lies in the span of ϕ_n.

Hermite cubic spline

Definition 1.14. *Let $f \in C^1[a,b]$, $a = x_0 < x_1 \cdots < x_n = b$ $n+1$ distinct points. The* **Hermite Cubic Spline** *s_H with* **knots** *at x_i is defined as $s_H(x) = p_i(x)$ for $x \in [x_{i-1}, x_i]$, where p_i is the Hermite interpolation polynomial defined using $\{x_{i-1}, x_i\}$.*

Theorem 1.15. *(Theorem 11.4 in textbook) Let $f \in C^2$, $h = \max\{x_i - x_{i-1}\}$, $M = \max|f^{(4)}|$, then for any $x \in [a,b]$, $|f(x) - s_H(x)| \leq \frac{1}{384}h^4 M$.*

The proof is similar to Theorem **??**. Note that $384 = 4!2^4$.

One can also find a set of basis functions for s_H.

1.4 Review

- Definition and formula of Lagrange/Hermite interpolation polynomials.
- Uniqueness.
- Error estimate.

2 Approximation Theory (Chapter 8, 9)

We can see that the Lagrange interpolation polynomial, Hermite interpolation polynomial, and the splines all lie in a vector space spanned by finitely many functions. In other words, all these algorithms can be seen as a way to **approximate a function using the linear combination of simpler functions**.

Recall that a set of functions form a vector space if it is closed under addition and scalar multiplication.

2.1 Approximation in normed vector space

Definition 2.1. *Let* V *be a vector space. A* **norm** *on* V *is a function:* $\|\cdot\| : V \to \mathbb{R}_{\geq 0}$ *such that:*

- $\|x\| = 0$ *iff* $x = 0$
- $\|x + y\| \leq \|x\| + \|y\|$
- $\|cx\| = |c|\|x\|$.

Example 2.2. *(i)* $V = C([a, b])$ *(continuous functions on* $[a, b]$*),* $\|f\|_\infty = \max|f|$*. This is called the* L^∞ *norm.*

(ii) $V = L^2([a, b])$, $\|f\|_2 = (\int_a^b |f(x)|^2 dx)^{1/2}$*. This is called the* L^2 *norm.*

(iii) Replace 2 *with* $p \geq 1$ *we get* L^p *norm. If* $p < 1$*, the triangle inequality is no longer satisfied.*

Definition 2.3. *Let $L = span\{x_1, \ldots x_m\}$ be a m-dimensional subspace of V, $x \in V$. The* **best approximation** *of x is the element $x' \in L$ that minimizes $\|x - x'\|$.*

Theorem 2.4. *(Theorem 8.2 in textbook) The best approximation always exists.*

The proof has two steps:

(i) $\| \cdot - x\|$ is continuous on L.

(ii) $\| \cdot - x\|$ goes to infinity at infinity.

Key idea: if a function is defined on a finite dimensional vector space, continuous, and goes to infinity at infinity, then it has a minimum.

Please ignore the proof below if you are not interested.

Proof. Let $x_1, \dots x_m$ be a basis of L. Consider a function $F_x : \mathbb{R}^m \to \mathbb{R}$ defined as

$$F_x((t_1, \dots, t_m)) = \|x - \sum_i t_i x_i\|$$

The first step of the proof is to show that F is continuous:

Lemma 2.5. *F_x is continuous.*

Proof. Suppose $t' \in \mathbb{R}^m$ satisfies that $|t'_i| < \epsilon$ for all i, then by triangle inequality,

$$|F_x(t + t') - F_x(t)| \leq |\sum_i t'_i x_i| \leq \epsilon \sum_i |x_i|$$

This implies that if t' is sufficiently small, $F_x(t+t')$ can be arbitrarily close to $F_x(t)$, hence F_x is continuous. $\square$

Now, let $D_R \subset \mathbb{R}^m = \{t : |t_i| \leq R \text{ for all } i\}$. It is a closed set, hence compact (recall the definition of compactness in your analysis class), hence a continuous function F_x takes minimum at some point x_R^* on D_R. We just need to show that if R is large enough, x_R^* is also the minimum of F_x.

Let $g > 0$ be the minimum of F_0 on the set D_1. Now we set $R_0 = (2|x| + 1)/g$. Then for any y outside D_{R_0}, then $F_x(y) = \|y - x\| \geq \|x\| + 1 > F_x(0) \geq F_x(x_{R_0}^*)$ $\square$

2.2 Stone-Weiersterass theorem

Theorem 2.6. *(Theorem 8.1 in textbook) For continuous function $f \in C([a,b])$, any $\epsilon > 0$, there is some polynomial p such that $\|f - p\|_\infty < \epsilon$.*

There are many proofs, some work for more general settings. An easy proof is first use linear spline to approximate f, then use polynomials to approximate the basis function (which is a linear combination of absolute values, which can be approximated by $(x^2 + \epsilon')^{1/2}$, which can be approximated using Taylor expansion).

2.3 Approximation in inner product space

Sometimes the norm on a vector space arises from a inner product (a symmetric, positive definite, bilinear form) $(\cdot,\cdot) : V \times V \to \mathbb{R}$, by $||x|| = \sqrt{(x,x)}$. If so, we call it an **inner product space**.

Example 2.7. *The L^2 norm on $L^2([a,b])$ arises from inner product $(f,g) = \int_a^b fg dx$. Let w be a non negative, continuous and integrable "weight function" on $[a,b]$, we can also defined the "weighted L^2 norm" which is from $(f,g)_w = \int_a^b wfg dx$.*

It's easy to see that the L^2_w norm satisfies:

$$\|f\|_w \leq (\int_a^b w dx)^{1/2} \|f\|_\infty$$

Example 2.8. *On $C^1([a,b])$ we can define the $(1,2)$ Sobolev norm $\|f\|_{1,2} = (\int_a^b |f(x)|^2 + |f'(x)|^2 dx)^{1/2}$. This norm also come from an inner product*

$$(f,g)_{1,2} = \int_a^b f(x)g(x) + f'(x)g'(x) dx$$

Let $L = span\{x_1, \dots, x_m\}$, then we can use Gram-Schmidt process to get an orthonormal basis of L under $(\cdot, \cdot)$, called $\{e_1, \dots, e_m\}$. Then we have:

Theorem 2.9. *The best approximation of $x \in V$ by an element of L is unique, and it is*

$$x^* = \sum_i (x, e_i) e_i$$

Proof. For any other $x' = \sum_i t_i e_i \in L$,

$$\|x' - x\|^2 = ((x' - x^*) + (x^* - x), (x' - x^*) + (x^* - x))$$

$$= \|x' - x^*\|^2 + \|x^* - x\|^2 + 2(\sum_i (t_i - (x, e_i)) e_i, \sum_i (x, e_i) e_i - x)$$

$$= \|x' - x^*\|^2 + \|x^* - x\|^2 + 2 \sum_j (t_i - (x, e_j))(e_j, \sum_i (x, e_i) e_i - x)$$

$$= \|x' - x^*\|^2 + \|x^* - x\|^2 + 2 \sum_j (t_i - (x, e_j))(\sum_i (x, e_j)(e_j, e_j) - (x, e_j))$$

$$= \|x' - x^*\|^2 + \|x^* - x\|^2 \geq \|x^* - x\|^2$$

And equality is reached only when $x' = x^*$. $\square$

When the inner product is the L^2_w inner product, the integrals in the formula for best approximation will often be calculated numerically (cf. next Section).

The proof is the same as the finite dimensional case you have seen in linear algebra.

If x_i are only orthogonal and not orthonormal, the formula becomes

$$x^* = \sum_i \frac{(x, x_i)}{(x_i, x_i)} x_i$$

If x_i are not known to be orthogonal either, the formula becomes

$$x^* = \sum_i (\sum_j (x, x_j)(A^{-1})_{i,j}) x_i$$

Where

$$A_{i,j} = (x_i, x_j)$$

2.4 Orthogonal polynomials

Definition 2.10. *We call ϕ_j, $j = 0, 1, 2, \ldots$ a system of* **orthogonal polynomials** *with weight w, if*

(i) ϕ_j *is of degree* j.

(ii) ϕ_j *are orthogonal to each other in* L^2_w *norm.*

Theorem 2.11. *If w is positive, continuous and integrable on (a, b) then a system of* **orthogonal polynomials** *with weight w exists.*

Proof. This is Gram-Schmidt:

$$\phi_0 = 1$$

$$\phi_j = x^j - \sum_{i=0}^{j-1} \frac{\int_a^b w t^j \phi_i(t) dt}{\int_a^b w \phi_i^2(t) dt} \phi_i(x)$$

□

From linear algebra, we know that the system of orthogonal polynomials is unique up to scaling, since ϕ_j is the basis vector of the orthogonal complement of $span\{1, x, \ldots, x^{j-1})$ in $span\{1, x, \ldots, x^j\}$.

Remark 2.12. *Stone-Weiersterass theorem implies that as degree increases, optimal approximation in $L^2_w([a, b])$ can become arbitrarily accurate. In other words, the orthogonal polynomials form an orthonormal basis of $L^2_w([a, b])$. (Which is NOT a basis in the sense of linear algebra. In algebra there is only finite sum.)*

Example 2.13. *Let* $(a, b) = (-1, 1)$.

- *If* $w = 1$, *the resulting orthogonal polynomials are called the* **Legendre polynomials** L_j.
- *If* $w(x) = (1-x^2)^{-1/2}$, *the resulting orthogonal polynomials are called the* **Chebyshev polynomials** T_j.

Remark 2.14. *The Chebyshev polynomials have a particularly nice formula:*

$$T_j = \cos(j \cos^{-1} x) \ .$$

They are polynomials because

$$T_0 = 1, T_1 = x, T_2 = 2x^2 - 1$$

$$\begin{aligned} T_j = \cos(j \cos^{-1} x) &= x \cos((j-1) \cos^{-1} x) - \sin(\cos^{-1} x) \sin((j-1) \cos^{-1} x) \\ &= x \cos((j-1) \cos^{-1} x) - \sin^2(\cos^{-1}(x)) \cos((j-2) \cos^{-1} x) \\ &\quad - \sin(\cos^{-1} x) \sin((j-2) \cos^{-1} x) \cos(\cos^{-1} x) \\ &= xT_{j-1} - (1 - x^2)T_{j-2} - x(T_{j-3} - T_{j-1})/2 \ . \end{aligned}$$

They are othogonal, because $j \neq j'$,

$$\int_{-1}^{1} \cos(j \cos^{-1} x) \cos(j' \cos^{-1} x)(1 - x^2)^{-1/2} dx$$

$$= -\int_{-1}^{1} \cos(j \cos^{-1} x) \cos(j' \cos^{-1} x) d \cos^{-1}(x)$$

$$= \int_{0}^{\pi} \cos(jt) \cos(j't) dt = 0$$

Remark 2.15. *Furthermore, if $T_j = \cos(j \cos^{-1} x)$, $2^{-j} T_{j+1}$ is the degree $j + 1$ monic (leading coefficient being 1) polynomial with the smallest L^∞ norm. This tells us that the term $\prod_i (x - x_i)$ in Theorem* **??** *can be minimized (in L^∞) if x_i are chosen as the roots of Chebyshev polynomials, or, in other words, if $\prod_i (x - x_i) = 2^{-n} T_{n+1}$. This is proved in Chapter 8 of the textbook.*

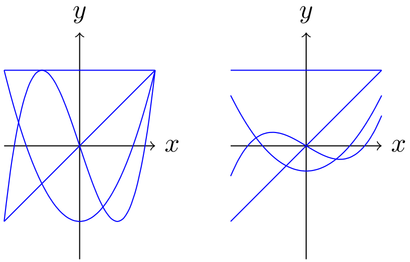


Figure 3: Chebyshev polynomials and Lagrange polynomials

- First Legendre polynomials (which I calculated using Gram-Schmidt, another alternative calculation can be found in the exercises, and also HW 4)

$$L_0 = 1, L_1 = x, L_2 = x^2 - \frac{1}{3}$$

$$L_3 = x^3 - \frac{3}{5}x$$

- First Chebyshev polynomials:

$$T_0 = 1, T_1 = x, T_2 = 2x^2 - 1, T_3 = 4x^3 - 3x$$

Theorem 2.16. *If the weight function w is positive, continuous and integrable on (a, b), then ϕ_j has j distinct real roots in (a, b).*

Proof. Suppose not, then ϕ_j switches sign fewer than j times in (a, b). Suppose $x_1, \ldots, x_k$ are the points in (a, b) where ϕ_j changes sign, then $(\phi_j, \prod_{i=1}^{k}(x - x_i))$ is non zero. However $\prod_{i=1}^{k}(x - x_i) \in span\{\phi_0, \ldots \phi_{j-1}\}$, hence a contradiction. □

This Theorem will be used in the next section when we discuss Gauss's method for numerical integration.

2.5 Review

- Normed vector space, inner product space, L^∞, L^2 and L_w^2 norms.
- Existence of optimal approximation. Calculation of optimal approximation for inner product space.
- Concept of orthogonal polynomials.

Example 2.17. *Consider the function $y = e^x$ on $[-1, 1]$.*

- *Find the Lagrange interpolation polynomial, interpolating at* 0, ± 1.
- *Find the Hermite interpolation polynomial, interpolating at* ± 1.
- *Find the best approximation via a polynomial of degree at most* 2, *under the L^2 norm.*

Answer:

- Use formula $p_L = \sum_i y_i \prod_{j\neq i}(x - x_j) / \prod_{j \neq i}(x_i - x_j)$:

$$p_L(x) = e^{-1} \cdot \frac{x(x-1)}{-1 \cdot -2} + 1 \cdot \frac{(x+1)(x-1)}{1 \cdot -1} + e \cdot \frac{x(x+1)}{1 \cdot 2}$$
$$= (e^{-1}/2 + e/2 - 1)x^2 + (e/2 - e^{-1}/2)x + 1$$

- Use formula $p_H = \sum_i z_i(x - x_i)\prod_{j\neq i}(x - x_j)^2 / \prod_{j\neq i}(x_i - x_j)^2 + \sum_i y_i(1 - (x - x_i)\sum_{j \neq i}(2/(x_i - x_j)))\prod_{j\neq i}(x - x_j)^2 / \prod_{j \neq i}(x_i - x_j)^2$:

$$p_H(x) = e^{-1} \cdot \frac{(x+1)(x-1)^2}{(-1-1)^2} + e \cdot \frac{(x-1)(x+1)^2}{(1+1)^2}$$
$$+ e^{-1} \cdot (1 + (x+1)) \cdot \frac{(x-1)^2}{(-1-1)^2} + e \cdot (1 - (x-1)) \cdot \frac{(x+1)^2}{(1+1)^2}$$
$$= (e^{-1}/2)x^3 + (e/4 - e^{-1}/4)x^2 + (e/2 - e^{-1})x + e/4 + 3e^{-1}/4$$

- Use formula $x^* = \sum_i ((x, x_i)/(x_i, x_i))x_i$:

$$p_2(x) = \frac{\int_{-1}^{1} e^t dt}{\int_{-1}^{1} 1^2 dt} \cdot 1 + \frac{\int_{-1}^{1} te^t dt}{\int_{-1}^{1} t^2} \cdot x + \frac{\int_{-1}^{1} (t^2 - 1/3)e^t dt}{\int_{-1}^{1} (t^2 - 1/3)^2 dt} \cdot (x^2 - 1/3)$$

$$= \frac{(e - e^{-1})}{2} \cdot 1 + \frac{2e^{-1}}{2/3} \cdot x + \frac{2e/3 - 14e^{-1}/3}{8/45}(x^2 - 1/3)$$

$$= \frac{15e - 105e^{-1}}{4} x^2 + 3e^{-1}x + \frac{-3e + 33e^{-1}}{4}$$

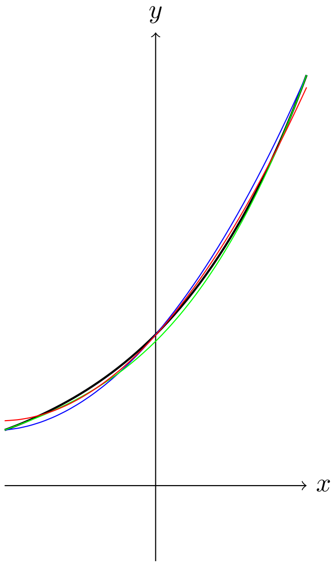


Figure 4: Black line: $y = e^x$. Blue line: Lagrange interpolation. Green line: Hermite interpolation. Red line: L^2 best approximation

3 Numerical Integration (Chapter 7, 10)

3.1 Quadrature rule

Question: Estimate $\int_a^b f(x)dx$.

Let $x_0 = a < x_1 < \cdots < x_n = b$ be $n+1$ distinct points in $[a, b]$, then we can use the Lagrange interpolation polynomial to estimate f, and hence

$$\int_a^b f(x)dx \approx \sum_k w_k f(x_k)$$

Where

$$w_k = \int_a^b \frac{\prod_{j \neq k}(x - x_j)}{\prod_{j \neq k}(x_k - x_j)} dx$$

To estimate the integration.

The points x_i are called **quadrature points**, and w_i called **quadrature weights**. The formula still works if some x_i is outside $[a, b]$.

3.2 Newton-Cotes method

Definition 3.1. *When* $a = x_0 < x_1 < \cdots < x_n = b$ *are* $n+1$ *evenly spaced points, the formula above is called the* **Newton-Cotes formula***, the evenly spaced* x_i *the* **Newton-Cotes quadrature***.*

Example 3.2. *When* $n = 1$, $w_0 = w_1 = \frac{b-a}{2}$*, this is called the* **Trapezium rule** *(as it's like calculating the area of a collection of trapeziums). When* $n = 2$, $w_0 = w_2 = \frac{b-a}{6}$, $w_1 = \frac{2(b-a)}{3}$*. This is called* **Simpson's rule***.*

Example 3.3. $\int_0^1 \sin(x)dx$*. Using Trapezium rule, the estimate is*

$$\frac{\sin(0) + \sin(1)}{2} = 0.4207$$

Using Simpson's rule, the estimate is

$$\sin(0)/6 + \sin(0.5) * 2/3 + \sin(1)/6 = 0.4599$$

The true value is $1 - \cos(1) = 0.4597$*.*

Theorem 3.4. *(Theorem 7.1 in the textbook) The error for the quadrature is bounded by*

$$\frac{\max|f^{(n+1)}|}{(n+1)!}\int_a^b \prod_i |x - x_i| dx$$

The proof is follows immediately from the error estimate of Lagrange interpolation. When x_i are Newton-Cotes quadrature, the error bound is $O(\max|f^{(n+1)}|(b-a)^{n+2})$, because when we scale the interval $[a, b]$ by C, the function being integrated is scaled by C^{n+1} while the range of the integration is also scaled by C.

For Newton-Cotes, when n is even (in order words when we have odd number of quadrature points), the error bound can be improved to $\max|f^{(n+2)}| \cdot O((b-a)^{n+3})$ provided $f \in C^{n+2}$:

Theorem 3.5. *Let n be an even number, $f \in C^{n+2}([a,b])$, $I_n(f)$ the Newton-Cotes formula using $n+1$ evenly spaced points on $[a,b]$, then there is some C_n (depending on n) such that*

$$|\int_a^b f(x)dx - I_n(f)| \le C_n \max|f^{(n+2)}|(b-a)^{n+3}$$

Proof. The uniqueness of Lagrange interpolation implies that if f is a polynomial of degree at most n, $\int_a^b f(x)dx = I_n(f)$. Now consider the polynomial $g = \prod_i (x - x_i)$. Because x_i are evenly spaced, the graph of $\prod_i (x - x_i)$ is symmetric with respect to the point $(x_{n/2}, 0)$ where $x_{n/2} = (a+b)/2$. So $\int_a^b g dx = 0 = I_n(g)$. However, any polynomial of degree at most $n+1$ can be written in the form $cg + h$, where h is a polynomial of degree at most n. Hence, if f is any polynomial of degree at most $n+1$, $\int_a^b f(x)dx = I_n(f)$.

Now suppose $f \in C^{n+2}$. Let x_{n+1} be the midpoint of $[x_0, x_1]$, let p' be the Lagrange interpolation polynomial of f, then $f - p'$ vanishes at $x_0, \ldots, x_n, x_{n+1}$, hence the quadrature formula using $x_0, \ldots x_{n+1}$ is 0. Apply Theorem **??** we get

$$|\int_a^b (f - p')dx| \leq C_n \max |f^{n+2}|(b-a)^{n+3}$$

However, because $p'(x_i) = f(x_i)$ for $i = 0, 1, 2, \ldots, n$,

$$\int_a^b p' dx = I_n(p') = I_n(f)$$

Which proves the theorem. $\square$

We can also use mean value theorem to get finer bounds. As an example, when $n = 2$, we have

Theorem 3.6. *(Theorem 7.2 in the textbook) If $f \in C^4$, there is some $c \in [a, b]$ such that*

$$\int_a^b f(x)dx - (b-a)\cdot(f(a)/6 + 2f((a+b)/2)/3 + f(b)/6) = -\frac{f^{(4)}(c)(b-a)^5}{2880}$$

Proof. Let

$$G_1(t) = \int_{(a+b)/2-t}^{(a+b)/2+t} f(s)ds - 2t\cdot(f((a+b)/2-t)/6$$

$$+2f((a+b)/2)/3 + f((a+b)/2+t)/6)$$

$$G(t) = G_1(t) - (\frac{t}{(b-a)/2})^5 G_1((b-a)/2)$$

Then $G(0) = G((b-a)/2) = G'(0) = G''(0) = 0$. So there is $0 < c_1 < (b-a)/2$ such that $G'(c_1) = 0$, $0 < c_2 < c_1$ such that $G''(c_2) = 0$, $0 < c_3 < c_2$ such that $G'''(c_3) = 0$.

By calculation, $G_1'''(c_3) = \frac{c_3}{3} \cdot (f'''((a+b)/2 - c_3) - f'''((a+b)/2 + c_3))$, so

$$\frac{c_3}{3} \cdot (f'''((a+b)/2 - c_3) - f'''((a+b)/2 + c_3)) - \frac{1920}{(b-a)^5} c_3^2 G_1((b-a)/2) = 0$$

Hence

$$\frac{\cdot(f'''((a+b)/2 + c_3) - f'''((a+b)/2 - c_3))}{2c_3} = -\frac{2880}{(b-a)^5} G_1((b-a)/2)$$

Now apply mean value theorem for f''' on $[(b+a)/2 - c_3, (b+a)/2 + c_3]$, we get the c. $\square$

3.3 Composite Method

For the same reason as in Example **??**, when $n \to \infty$ the error can not be guaranteed to decay to 0. So we often evenly decompose the interval $[a, b]$ into subintervals then carry out low order Newton-Cotes. For example, if we cut it into n subintervals and apply trapezium rule on each we get

$$\frac{b-a}{n}(f(x_0)/2 + \sum_{i=1}^{n-1} f(x_i) + f(x_n)/2)$$

If we cut it into $n/2$ subintervals, and apply simpson's rule on each, we get

$$\frac{b-a}{3n}(f(x_0) + 4\sum_{i=1}^{n/2} f(x_{2i-1}) + 2\sum_{i=1}^{n/2-1} f(x_{2i}) + f(x_n))$$

The error estimate can be calculated as before. When f is smooth the error decay like $O(n^{-3})$ and $O(n^{-5})$ respectively.

Example 3.7. $\int_{-0.5}^{0.5} \sqrt{1-x^2}dx$. *The right answer should be*

$$\sqrt{3}/4 + \pi/6 = 0.9566114774905181$$

(i) Trapezium rule:

$$(\sqrt{3/4} + \sqrt{3/4})/2 = 0.8660254037844386$$

(ii) Simpson's rule:

$$\sqrt{3/4}/6 + 1 \times 2/3 + \sqrt{3/4}/6 = 0.9553418012614795$$

Now we do composite rules:

```
from math import *
f=lambda x : (1-x*x)**0.5
true_value=0.5*0.75**0.5+pi/6
def composite_trapezium(n, a, b, f):
    r=0
    r+=0.5*(f(a)+f(b))
    for i in range(1, n):
        r+=f(((n-i)*a+i*b)/n)
    return r*(b-a)/n
def composite_simpsons(n, a, b, f):
    r=0
    r+=f(a)+f(b)
    for i in range(1, n, 2):
        r+=4*f(((n-i)*a+i*b)/n)
    for i in range(2, n, 2):
        r+=2*f(((n-i)*a+i*b)/n)
    return r*(b-a)/3/n
for n in range(2, 11):
    result=composite_trapezium(n, -0.5, 0.5, f)
    print(n, result, log(abs(true_value-result))/log(n))
for n in range(4, 21, 2):
    result=composite_simpsons(n, -0.5, 0.5, f)
    print(n, result, log(abs(true_value-result))/log(n))
```

Output:

```
2 0.9330127018922193 -5.405144181208228
3 0.9460173327169924 -4.139270839443451
4 0.9506292692220368 -3.692553073204491
5 0.952775736166314 -3.456730150933653
6 0.9539450135525174 -3.3079228288736666
7 0.9546512047359057 -3.2039874286754015
8 0.9551100227450633 -3.1264744308706844
9 0.9554248044017013 -3.0659595201887435
10 0.9556500753040448 -3.017094894510029
4 0.9565014583319759 -6.574978802776244
6 0.9565875738310257 -5.939122508722478
8 0.9566036072527387 -5.651720448514391
10 0.9566081883499551 -5.482917566334656
12 0.9566098729686051 -5.369491332969493
14 0.9566106052007204 -5.286790845669449
16 0.9566109637450405 -5.223110718777519
18 0.9566111557047383 -5.172130454182419
20 0.9566112658643311 -5.13011276845817
```

If n gets even higher the rounding error will have more prominent effect.