CSCC37 Review Seminar

Hold by AMACSS - Tuesday, PM 6-8

QUESTION 1:

Consider the normalized floating-point system $\mathbb{R}_{6(7,2)}$ with limited exponent range $-(55)_6 \leq \epsilon \leq (55)_6$.

1. What is the smallest positive (nonzero) representable? Give your answer in both base-6.
2. What is the largest positive number representable? Give your answer in both base-6.
3. Assuming round-to-nearest, what is the tightest upper bound on the relative error $\frac{|f(x)-x|}{|x|}$ when $x \in \mathcal{R}$ is stored as $f(x) \in \mathbb{R}_{6(7,2)}$ in this floating-point system? Give your answer in base-10.
4. What is the floating-point representation of $(407)_{10}$ in this system?

Additional Question;

A terminating decimal fraction with n digits must have a terminating base-6 fraction representation?

QUESTION 2:

Consider the linear system $Ax = b$, $A \in \mathbb{R}^{3\times 3}$, $x, b \in \mathbb{R}^3$ with

$$A = \begin{bmatrix} 2 & 6 & 6 \\ 1 & 7 & 6 \\ 4 & 12 & 12 \end{bmatrix}, \quad b = \begin{bmatrix} 20 \\ 25 \\ 40 \end{bmatrix}.$$

a.

Compute the $PA = LU$ factorization of A. Use exact arithmetic. Show all intermediate calculations, including Gauss transforms and permutation matrices.

b.

Use the factorization computed in (a) to solve the system, if a solution exists. If there is no solution, you must justify why there is no solution.

QUESTION 3:

We know from linear algebra that the determinant of the product of two square matrices is equal to the product of the determinants; i.e., for $A, B \in \mathbb{R}^{n \times n}$, $\det(AB) = \det(A)\det(B)$. Using this fact, derive an *optimally efficient formula* for computing the determinant of $A \in \mathbb{R}^{n \times n}$ based on the $PA = LU$ factorization of A.

QUESTION 4:

A floating-point operation, or flop, is an operation of the form $mx + b$. Show how to convert a $(k+1)$-digit base b $(b \neq 10)$ positive integer $d_k d_{k-1} ... d_1 d_0$ into its base 10 equivalent in k flops or less.



QUESTION 5:

What is the purpose of the following fixed point iteration?

$g(x)$: $x_{k+1} = 2x_k - x_k^2 y$ ← constant

Hint: Identify it as the Newton iteration for a certain function. $F(x)$

$$x_{k+1} = 2x_k - x_k^2 y \iff y = \frac{2x_k - x_{k+1}}{x_k} = \frac{2}{x_k} - \frac{x_{k+1}}{x_k^2}$$

As $k \to \infty$, $x_{k+1} = x_k = \bar{x}$ when coverge happen

$$y = \frac{2}{x} - \frac{x^2}{x^2} = \frac{1}{x} \Rightarrow x = \frac{1}{y}$$

Therefore, $\frac{1}{y}$ is the fixed point of this iteration

Since this a Newton iteration, all the FPs are the roots of $F(x)$.

Recall: $g(x) = x - \frac{F(x)}{F'(x)}$ Newton Method

↳ To find $F(x)$:

Given $g(x) = 2x - x^2 y$

$= x - (x^2 y - x)$ equ

C37
- A37 (L'Hopital)
A22 (Taylor Series)
(Invertible Matrix)

$x^2 y - x = \frac{F(x)}{F'(x)} = \frac{x+yx}{A37} \frac{x^2 x}{\frac{1}{x}}$

$x(xy-1)$ $F'(x)$

$F(x) \Rightarrow F'(x)$

$F(x) = \frac{1}{x} - y$ ✓

$F'(x) = -x^{-2}$

$\frac{F(x)}{F'(x)} = \frac{\frac{1}{x} - y}{-x^{-2}}$

$= x^2 y - x$ ✓

∴ The purpose of this FP iter

is to find the root of $\frac{1}{x} - y$ which is $\frac{1}{y}$

QUESTION 6:

Consider the data points $\{(-1,4),(0,6),(1,12)\}$.

1. Derive the divided-difference form of the quadratic interpolant. Show all your work, including the divided-difference table.
2. Derive the Lagrange form of the quadratic interpolant.
3. Apply a theorem discussed in lecture to verify that the two polynomials derived above are the same polynomial.
4. Append (2,16) to the data and construct a single cubic polynomial interpolant over all four data points. What is the method that requires that least amount of extra work given the quadratic interpolant you already computed?

$P_a, P_b \in P_2$ x^2

def of Lagrange

def Vandermonde

def divided-diff

P_a

P_b

x^3

1. using divided-diff form

$P_c = ax^3 + \cdots$

a_4

x_i $y[x_i] = a_1$ $y[x_{i+1}, x_i] = a_2$ $y[x_{i+2}, x_{i+1}, x_i] = a_3$

x_1	-1	4
x_2	0	6
x_3	1	12
x_4	2	16

$$\frac{6-4}{0-(-1)} = 2 = y[x_2, x_1] = a_2$$

$$\frac{12-6}{1-0} = 6 = y[x_3, x_2]$$

$$\frac{16-12}{2-1} = 4$$

$$\frac{6-2}{1-(-1)} = 2$$

$$\frac{6-4}{2-0} = 1$$

$y[x_3, x_2, x_1]$

$y[x_4, x_3, x_2, x_1]$

By Newton Polynomial thm : 2-1

$$P_a(x) = y[-1] + (x-(-1))\, y[0,-1] + (x-0)(x-(-1))\, y[1,0,-1]$$

$$= 4 + 2(x+1) + 2x(x+1)$$

2. $l_0(x) = \prod_{\substack{j=0 \\ (j \neq 0)}}^{2} \frac{x - x_j}{x_0 - x_j} = (\quad)\cdot(\quad) = \frac{x(x-1)}{2}$

$$l_1(x) = \prod_{\substack{j=0 \\ (j \neq 1)}}^{2} \frac{x - x_j}{x_1 - x_j} = \frac{(x+1)(x-1)}{-1}$$

$l_2(x)$

$$= \frac{(x+1)x}{2}$$

by Layrange method of interpolation : $P(x) = \sum_{i=0}^{n} l_i(x) y$

$$P_b(x) = 4 \cdot \frac{x(x-1)}{2} + 6 \cdot \frac{(x+1)(x-1)}{-1} + 12 \cdot \frac{(x+1)x}{2}$$

compare p_a and p_b

Existence & polynomial uniqueness thm

$\{x_i : i = 0, 1, \dots n\}$ and $\{y_i : i = 0 \dots n\}$

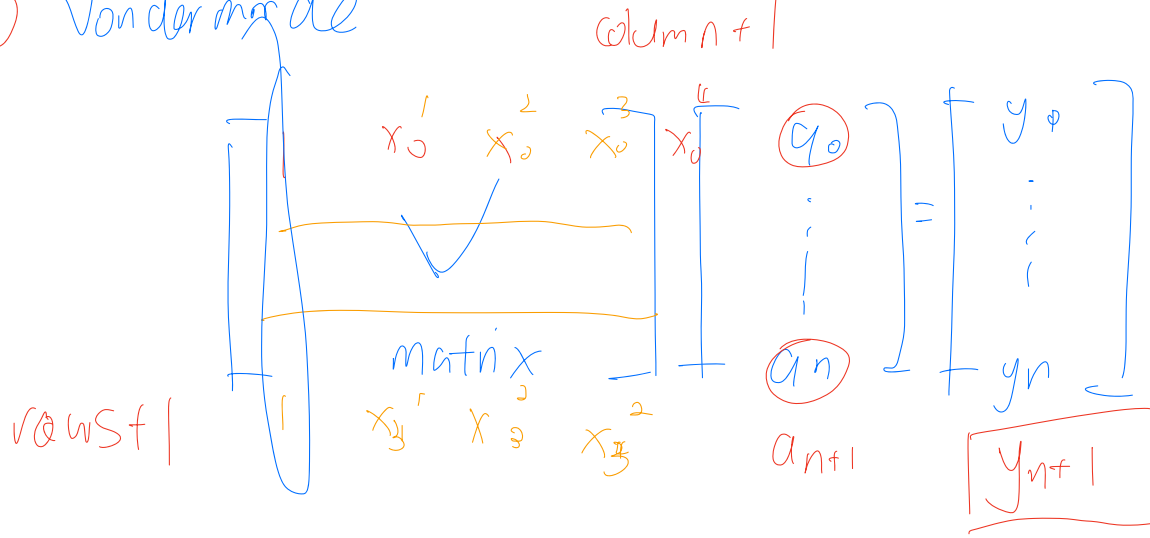
where x_i is distinct, y_i not necessarily.

$\Rightarrow \exists$ a unique poly $p \in P_n$ s.t $p(x_i) = y_i$

highest order is n

(x^n)

4. (1) Vandermonde



$V = 4 \times 4 \Rightarrow 5 \times 5$

(2) Lagrange

$l_0(x) = \prod_{\substack{j=0 \\ j \neq 0}} \dots \times \square \times \square \times \square \times _$

$l_i(x) =$

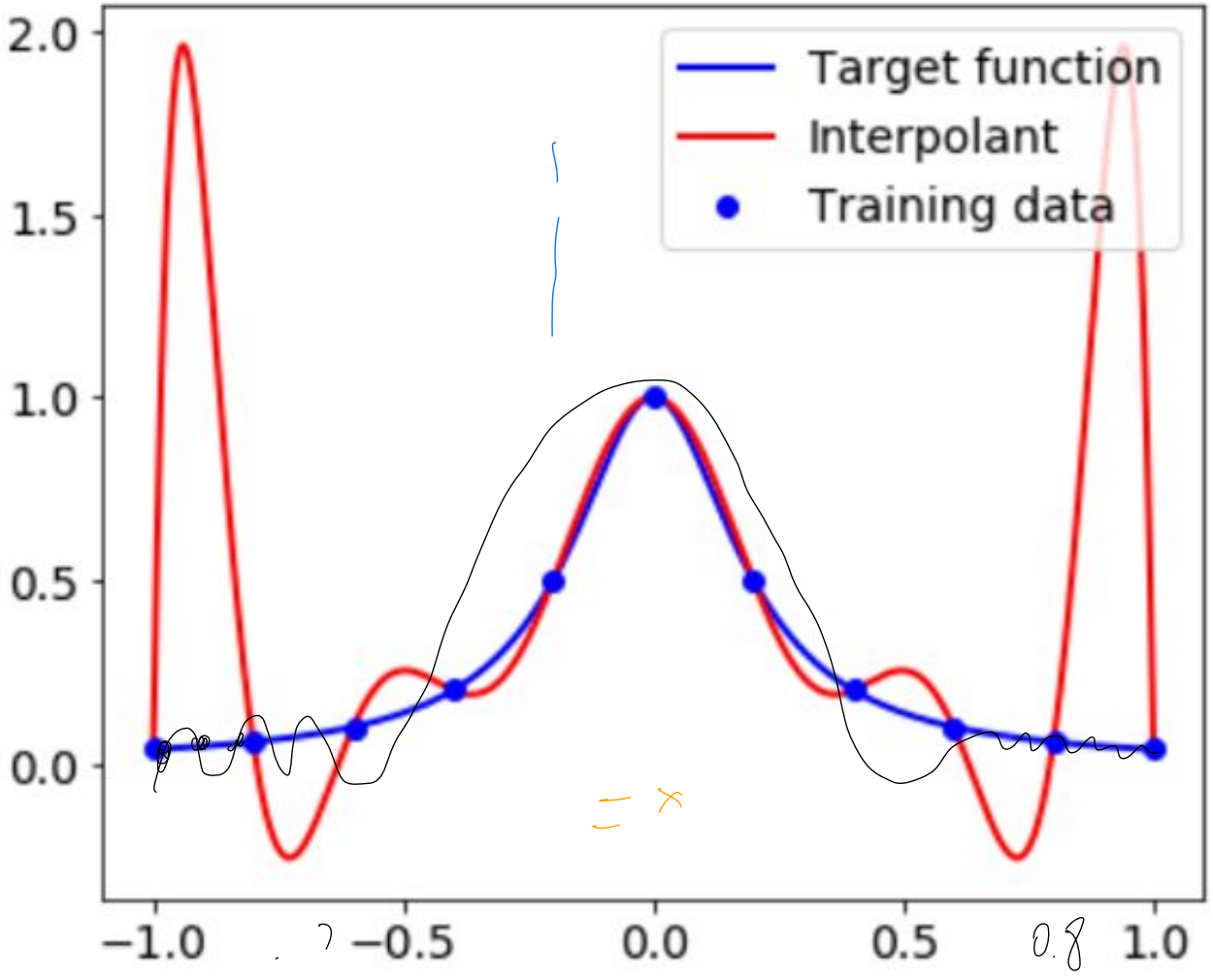
$l'(x) =$

$l'(x) =$

QUESTION 7:

A classic example of an underlying function which results in the upper bound for the polynomial interpolation error to increase rapidly is the Runge function $y(x) = \frac{1}{1+25x^2}$. For $p(x) \in \mathcal{P}_n$ interpolating equally spaced points in $[-1, 1]$, sketch a graph illustrating the Runge phenomenon. Propose a solution to interpolate the Runge function without the issue stated by the Runge phenomenon.

what is a runge phenomenon?


Target function
Interpolant
Training data
2.0
1.5
1.0
0.5
0.0
−1.0
−0.5
0.0
0.5
1.0

Solutions to interpolate the Runge function:

- Restrict ourselves to a strung of lower degree polynomials each of which are only applied over one or two subintervals - use piecewise polynomial interpolation.
- Smooth the wrinkles in the interpolating polynomial by fitting a lower degree polynomial.
- Choose the interpolation points more judiciously.

⇒ Chebyshev interpolant

more data points on the left/right side

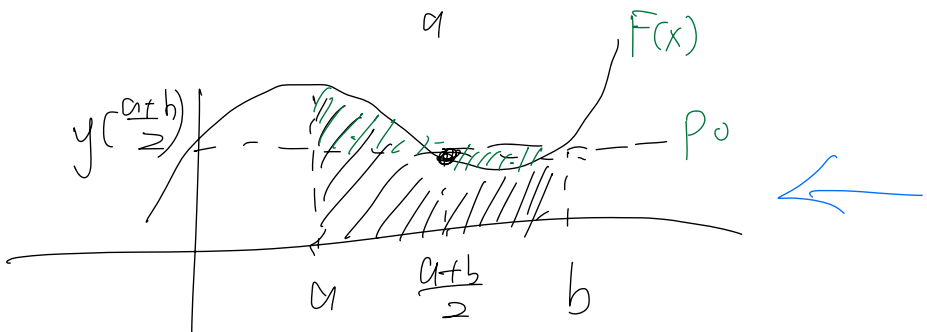
and less data point in the middle.

QUESTION 8:

State the midpoint rule, the trapezoidal rule and the Simpson's rule for numerical integration. What is the precision for each rule?

Consider $I(f) = \int_a^b f(x)\,dx$

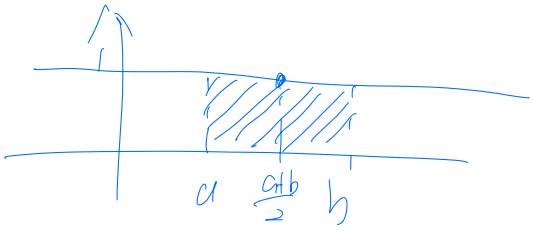
A37 Riemann Sum



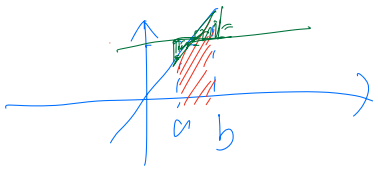
$$I(f) \approx M(f) = I(p_0) = \underbrace{(b-a)}_{\text{weight}} \cdot \underbrace{f\left(\frac{a+b}{2}\right)}_{\text{node } x_0}$$

precision : $m = 1$

when $f = x^0 = 1$ $\quad M(x^0) = M(1) = (b-a)\cdot 1 = I(x^0) = \int_a^b 1\,dx$



when $f = x^1$ $\quad M(x^1) = (b-a)\cdot\frac{a+b}{2} = \frac{b^2-a^2}{2}$



$= I(x^1) = \int_a^b x\,dx$

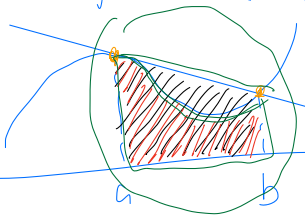
when $f = x^2$

$M(x^2) = (b-a)\left(\frac{a+b}{2}\right)^2 \neq I(x^2) = \int_a^b x^2\,dx$

precision = 1 $\quad \neq \frac{b^3-a^3}{3}$

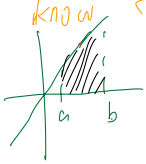
$A_0, A_1 = \frac{b-a}{2}$ $\quad x_0 = a$ $\quad x_1 = b$

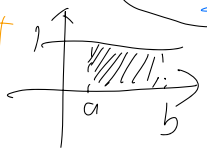
Trapezoid rule : $I(f) \approx T(f) = I(p_1) = (b-a)\left[\frac{f(a)+f(b)}{2}\right]$



p_1

don't know f





precision = 1

when $f = x^0 = 1$ $\quad T(x^0) = (b-a)\cdot\frac{1+1}{2} = b-a$ $\quad = I(x^0) = \int_a^b x^0\,dx$

$= b-a$

$f = x^1$ $\quad T(x^1) = (b-a)\cdot\frac{a+b}{2} = \frac{b^2-a^2}{2}$ $\quad = I(x^1) = \int_a^b x^1\,dx$

$f = x^2$ $\quad T(x^2) = (b-a)\cdot\frac{a^2+b^2}{2} \neq$ $\quad \neq I(x^2) = \int_a^b x^2\,dx$

Precision: $m = 3$

$$S(x^0) = \frac{b-a}{6}(1+4+1) = b-a = I(x^0)$$

$$S(x^1) = \frac{b-a}{6}\left(a + 4\frac{a+b}{2} + b\right) = \frac{b-a}{6}(3a+3b) = \frac{b^2-a^2}{2} = I(x^1)$$

$$S(x^2) = \frac{b-a}{6}\left(a^2 + 4\left(\frac{a+b}{2}\right)^2 + b^2\right) = \frac{b-a}{6}(a^2 + (a+b)^2 + b^2)$$

$$= \frac{b-a}{6}(a^2 + a^2 + 2ab + b^2 + b^2) = \frac{b-a}{6}(2a^2 + 2ab + 2b^2)$$

$$= \frac{b-a}{3}(b^2 + ab + a^2) = \frac{b^3-a^3}{3} = I(x^2)$$

$$S(x^3) = \frac{b-a}{6}\left(a^3 + 4\left(\frac{a+b}{2}\right)^3 + b^3\right) = \frac{b-a}{6}\left((a+b)(a^2-ab+b^2) + \frac{1}{2}(a+b)^3\right)$$

$$= \frac{b-a}{6}(a+b)\left(a^2 - ab + b^2 + \frac{1}{2}(a+b)^2\right) = \frac{b-a}{6}(a+b)\left(\frac{3a^2}{2} + \frac{3b^2}{2}\right)$$

$$= \frac{(b-a)(a+b)(a^2+b^2)}{4} = \frac{b^4-a^4}{4} = I(x^3)$$

Simpson's Rule: $I(f) \approx S(f) = I(p_2) = \frac{b-a}{6}\left(f(a) + 4f\left(\frac{a+b}{2}\right) + f(b)\right)$

precision: $m = 3$

$A_0 = \frac{b-a}{6}$ $\quad x_0 = a$

$A_1 = \frac{2}{3}(b-a)$ $\quad x_1 = \frac{a+b}{2}$

$x_2 = b$

} 3 node

$p(x) = x^3 + 4x^2 + 6x + 6$

$A_2 = \frac{b-a}{6}$

$f = x^4$ $\quad$ precision = 3

$S(x^4) \neq I(x^4) = \int_a^b x^4 \, dx$

You should be familiar with ...

- Fixed Point Theorem (FPT)

process of proof

If there is an interval $[a, b]$ such that $g(x) \in [a, b]$ for all $x \in [a, b]$ (1)
And if $\|g'(x)\| \leq L < 1$, $\forall x \in [a, b]$ (2)
Then $g(x)$ has a unique fixed point in $[a, b]$

- Rate of Convergence Theorem (RCT)

For the Fixed Point Iteration $x_{k+1} = g(x_k)$, if $g'(\hat{x}) = g''(\hat{x}) \cdots = g^{(p-1)}(\hat{x}) = 0$ but $g^p(\hat{x}) \neq 0$, then we have p-th order convergence.

- Theorem: Existence of Interpolatory Quadrature
- Weierstrass' Theorem (existence of polynomial interpolation)

Given any set of $n+1$ distinct nodes, choose weights A_i s.t. the quadrature rule is exact for all polynomials of degree $\leq n$.

Moreover, the weights are unique. $\Rightarrow$ {Fundamental Theorem of Algebra

Good Luck on Your Exam!

$p \in P_3$ P_2, P_1

given 4 nodes : $A_1, A_2, A_3 \cdots A_n$

Recall your

x^3

$x^5 \rightarrow$ run test

less than nodes numbe

Vandermonde

Theorem :



What is the precision of Quature rule?

(n+1 nodes) $\geq n$

at least n.

n+1 ?

n+2 ?

$\Rightarrow$ Note : m may be greater than n. $f = x^{n+1}$

at least precision is n