CSC2800 Numerical Computation

Tutorial 8
Midterm Exam Solution

By Tilen

Distribution of Midterm Answer Book

- Once you leave the classroom with your answer book, then your mark will not have any changes
- Do not help your classmates to collect the midterm answer book
- If you need to leave early, you may return it to tutors
- For those who don’t come to tutorial, they may collect their midterm answer books on 24/3 at 2:00-4:00pm at Rm1026

Q1

- A fictional machine uses 4-digit decimal mantissa to represent floating-point number
- Round-off is used
- (a) Create an example for A,B and C s.t. $(A+B)+C \neq A+(B+C)$
- (b) $x_T=123.36$, $y_T=2.3114$. Calculate relative errors of (x^2+y) on this machine.

Q1 Solution

- (a) Let A=0.1111, B=0.00002, C=0.00004.
- (A+B)+C
  - =0.1111+0.00004 (0.11112 has been rounded down)
  - =0.1111 (0.11114 has been rounded down)
- A+(B+C)
  - =0.1111+0.00006
  - =0.1112 (0.11116 has been rounded up)

Q1 Solution(2)

- (b) True value=15220.001
  - x_A=123.4, y_A=2.311
  - x_A^2=123.4*123.4=15227.56=15230(rounded up)
  - $x_A^2+y_A$=15230+2.311=15232.311=15230 (rounded down)
  - Relative error=(15220.001-15230)/15220.001=-0.000657 or -0.0657%

Q2

- For each of the following functions, propose a method to compute its value that could minimize the effect of round-off errors.

$$(a)\ \ f(x)=3x^4+21x^3+3.2x^2+2x-1\ ,\ for\ 1<x<2.$$

$$(b)\ \ f(x)=x-\sqrt{x^2+1}\ when\ x\ is\ \mathrm{large}$$

$$(c)\ \ f(x)=\frac{e^{2x}-1}{e^x}\ when\ x\ is\ close\ to\ 0.$$

Q2 Solution

$$(a)\quad f(x) = -1 + x(2 + x(3.2 + x(21 + 3x)))$$

$$(b)\quad f(x) = x - \sqrt{x^2+1} \times \frac{x+\sqrt{x^2+1}}{x+\sqrt{x^2+1}} = \frac{-1}{x+\sqrt{x^2+1}}$$

$$(c)\quad f(x) = \frac{e^{2x}-1}{e^x} = e^x - e^{-x} = \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots\right) - \left(1 - x + \frac{x^2}{2!} - \frac{x^3}{3!} + \ldots\right)$$

$$¿\,2\left(x + \frac{x^3}{3!} + \frac{x^5}{5!} + \ldots\right) = 2\sum_{n=1}^{\infty} \frac{x^{2n-1}}{(2n-1)!}$$

Q3

- How many terms are needed to approximate

$$f(x) = x^3 + 3x + 0.1 + \cos x \quad for \ |x| < 0.4$$

so that the truncation error is less than or equal to 10^{-6}

Q3 Solution

- Polynomials(x^3+3x+0.1) can be calculated with no truncation error (refer to Homework1 Q6)
- Only need to approximate truncation error of cos(x)

$$\cos x = 1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\ldots=\sum_{n=0}^{\infty}\frac{(-1)^n x^{2n}}{2n!}$$

Q3 Solution(2)

- cos(x) is an alternating convergent series
- Truncation error is bounded by the next term excluded from the approximation series

$$Truncation\ error = \frac{x^{2n}}{2n!} \leq \frac{0.4^{2n}}{2n!}$$

$$\frac{0.4^{2n}}{2n!} \leq 10^{-6}$$

$$=> n \geq 4$$

Q3 Solution(3)

- How many terms needed?

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cancel{\frac{x^8}{8!}} - \ldots + \ldots$$

$$Polynomial\ Part = x^3 + 3x + 0.1$$

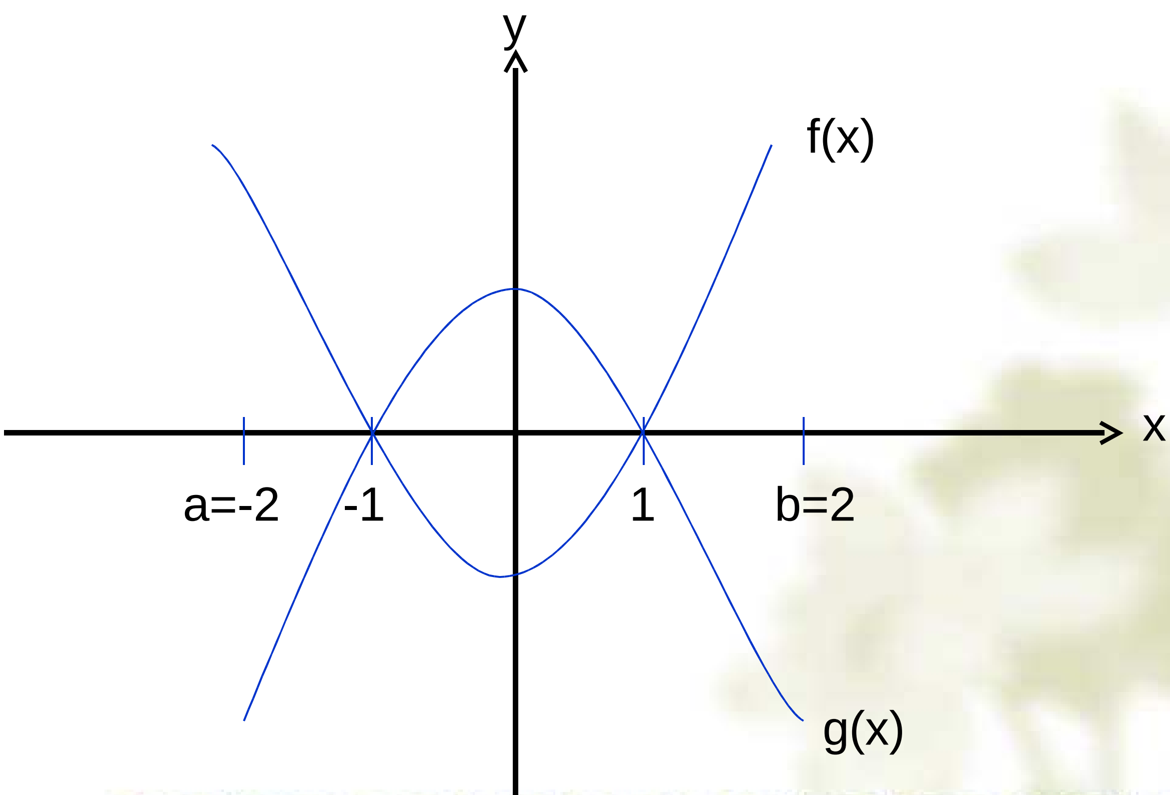
- Total number of terms needed = 4+3 = 7

Q4

- Jack claimed that any two functions, $f(x)$ and $g(x)$, that satisfy the following conditions will not intersect in the interval $[a,b]$.
- (a) They are continuous in the interval $[a,b]$
- (b) $f(a)$ and $f(b)$ are both positive
- (c) $g(a)$ and $g(b)$ are both negative

Q4 Solution

- Give a counter-example: $f(x)=x^2-1$ and $g(x) = -f(x)$



Q5

- Suppose an iteration $x_i+1 = g(x_i)$ is converging to the solution α, and we know that $\delta_{99}=0.25X10^{-19}$,

  $\delta_{100}=0.5X10^{-20}$ and

  $\delta_{101}=0.99X10^{-21}$.

Estimate the convergent rate of iteration

Q5

Definition: Let the sequence $\{r_n\}$ converge to r. Denote the difference between r_n and r by e_n; i.e. $e_n = r_n - r$. If there exists a positive number $p \geq 1$ and a constant $c \neq 0$ such that

$$\lim_{n\to\infty}\frac{|r_{n+1}-r|}{|r_n-r|^p}=\lim_{n\to\infty}\frac{|e_{n+1}|}{|e_n|^p}=c$$

then p is called the order of convergence of the sequence. The constant c is called the rate of convergence.

Q5 Solution1

- Trial an error: Guess value of p
- Guess p=1

$$\frac{\delta_{100}}{(\delta_{99})^1} = 0.2 \quad \frac{\delta_{101}}{(\delta_{100})^1} = 0.198$$

$$\Rightarrow c \approx 0.2 \ or \ 0.199, \ p = 1 (linear)$$

Q5 Solution2

- Formal method: Solve for p

$$\frac{\delta_{100}}{(\delta_{99})^p} = c \wedge \frac{\delta_{101}}{(\delta_{100})^p} = c$$

$$\Rightarrow \frac{\delta_{100}}{(\delta_{99})^p} = \frac{\delta_{101}}{(\delta_{100})^p} \Rightarrow \left(\frac{\delta_{100}}{\delta_{99}}\right)^p = \frac{\delta_{101}}{\delta_{100}}$$

$$\Rightarrow p \ln\left(\frac{\delta_{100}}{\delta_{99}}\right) = \ln\left(\frac{\delta_{101}}{\delta_{100}}\right)$$

$$\Rightarrow p = \frac{\ln\left(\frac{\delta_{101}}{\delta_{100}}\right)}{\ln\left(\frac{\delta_{100}}{\delta_{99}}\right)}$$

Q6

- Suppose you want to find the zeroes of
  - $f(x)=x^3-e^x$
- (a) Using the Newton-Raphson method, construct an updating formula for finding the zeroes of f(x)
- (b) Assuming the iteration based on your updating formula converges to the solution. What is the expecting convergent rate of the iteration?

Q6 Solution

$$(a)\ x_{i+1} = x_i - \frac{f(x_i)}{f'(x_i)} = x_i - \frac{x_i^3 - e^{x_i}}{3x_i - e^{x_i}}$$

- (b) Newton-Raphson iteration is of order 2

Q7

- Given the following system of linear equations

$$
\begin{aligned}
10x_1 + 2x_2 + 14x_3 &= 1\\
-24x_1 + 2x_2 + 14x_3 &= 1\\
6x_1 - 30x_2 + 13x_3 &= 1
\end{aligned}
$$

- Construct an update formula that is guaranteed to converge
- Explain why it will converge

Q7 Solution

- updating formula diagonal dominant that is guaranteed to converge

$$\begin{aligned}10x_1 + 2x_2 + 14x_3 &= 1\\ -24x_1 + 2x_2 + 14x_3 &= 1\\ 6x_1 - 30x_2 + 13x_3 &= 1\end{aligned} \quad \longrightarrow \quad \begin{aligned}-24x_1 + 2x_2 + 14x_3 &= 1\\ 6x_1 - 30x_2 + 13x_3 &= 1\\ 10x_1 + 2x_2 + 14x_3 &= 1\end{aligned}$$

- updating formula

$$x_{1(i+1)} = (1 - 2x_{2(i)} - 14x_{3(i)})/-24$$

$$x_{2(i+1)} = (1 - 6x_{1(i+1)} - 13x_{3(i)})/-30$$

$$x_{3(i+1)} = (1 - 10x_{1(i+1)} - 2x_{2(i+1)})/14$$

Q8

$$Given\ \ matrix\ \ A = \begin{bmatrix} 4 & 2 & 4 \\ -4 & -2.5 & -3 \\ 16 & 7 & 17.5 \end{bmatrix}$$

- Derive the matrices L and U in which A=LU and L is a lower triangular matrix and U is an upper triangular matrix.

Q8 Solution

- Perform Gauss elimination on A

$$A=\begin{bmatrix} 4 & 2 & 4 \\ -4 & -2.5 & -3 \\ 16 & 7 & 17.5 \end{bmatrix}$$

$$\begin{bmatrix} 4 & 2 & 4 \\ 0 & -0.5 & 1 \\ 0 & -1 & 1.5 \end{bmatrix}$$

$\rightarrow Row\,2 - Row\,1 \times f_{21},\ f_{21} = \frac{-4}{4} = -1$

$\rightarrow Row\,3 - Row\,1 \times f_{31},\ f_{31} = \frac{16}{4} = 4$

$$\begin{bmatrix} 4 & 2 & 4 \\ 0 & -0.5 & 1 \\ 0 & 0 & -0.5 \end{bmatrix}$$

$\rightarrow Row\,3 - Row\,2 \times f_{32},\ f_{32} = \frac{-1}{-0.5} = 2$

Q8 Solution(2)

$$A = LU \text{ , where}$$

$$U = \begin{bmatrix} 4 & 2 & 4 \\ 0 & -0.5 & 1 \\ 0 & 0 & -0.5 \end{bmatrix}$$

$$L = \begin{bmatrix} 1 & 0 & 0 \\ f_{21} & 1 & 0 \\ f_{31} & f_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 1 & 0 \\ 4 & 2 & 1 \end{bmatrix}$$

Q9

$$B = \begin{bmatrix} & & A & & 0 \\ & & & & 0 \\ & & & & \vdots \\ & & & & 0 \\ b_{n,1} & b_{n,2} & \cdots & b_{n,n-1} & b_{n,n} \end{bmatrix}, b_{ij} = a_{ij} \ for\ 1 \le i, j \le n-1$$

- Write the pseudocode of an algorithm to compute L_B and U_B such that $B=L_BU_B$.
- Given L_A and U_A such that $A=L_AU_A$.

$$L_A = \begin{bmatrix} 1 & & & & \\ l_{2,1} & 1 & & & \\ l_{3,1} & l_{3,2} & 1 & & \\ \vdots & \vdots & \cdots & \ddots & \\ l_{n-1,1} & l_{n-1,2} & \cdots & l_{n-1,n-2} & 1 \end{bmatrix} \quad U_A = \begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} & \cdots & u_{1,n-1} \\ & u_{2,2} & u_{2,3} & \cdots & u_{2,n-1} \\ & & u_{3,3} & \cdots & u_{3,n-1} \\ & & & \ddots & \vdots \\ & & & & u_{n-1,n-1} \end{bmatrix}$$

Q9 Solution

- Make use of L_A and U_A

$$L_B=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ l_{2,1} & 1 & 0 & 0 & 0 & 0 \\ l_{3,1} & l_{3,2} & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \cdots & \ddots & 0 & 0 \\ l_{n-1,1} & l_{n-1,2} & \cdots & l_{n-1,n-2} & 1 & 0 \\ ? & ? & ? & ? & ? & 1 \end{bmatrix} \quad U_B=\begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} & \cdots & u_{1,n-1} & ? \\ 0 & u_{2,2} & u_{2,3} & \cdots & u_{2,n-1} & ? \\ 0 & 0 & u_{3,3} & \cdots & u_{3,n-1} & ? \\ 0 & 0 & 0 & \ddots & \vdots & ? \\ 0 & 0 & 0 & 0 & u_{n-1,n-1} & ? \\ 0 & 0 & 0 & 0 & 0 & ? \end{bmatrix}$$

- By definition of triangular matrix

Q9 Solution(2)

- Last column of b is all 0(up to i=n-1)
- => Last column of U_B is all 0

$$B=\begin{bmatrix} & & & & 0 \\ & A & & & 0 \\ & & & & \vdots \\ & & & & 0 \\ b_{n,1} & b_{n,2} & \cdots & b_{n,n-1} & b_{n,n} \end{bmatrix}$$

$$L_B=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ l_{2,1} & 1 & 0 & 0 & 0 & 0 \\ l_{3,1} & l_{3,2} & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \cdots & \ddots & 0 & 0 \\ l_{n-1,1} & l_{n-1,2} & \cdots & l_{n-1,n-2} & 1 & 0 \\ ? & ? & ? & ? & ? & 1 \end{bmatrix} \quad U_B=\begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} & \cdots & u_{1,n-1} & 0 \\ 0 & u_{2,2} & u_{2,3} & \cdots & u_{2,n-1} & 0 \\ 0 & 0 & u_{3,3} & \cdots & u_{3,n-1} & 0 \\ 0 & 0 & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & u_{n-1,n-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & ? \end{bmatrix}$$

$$b_{n,n}=?\times 0+?\times 0+?\times 0\ldots+1\times u_{n,n}\Rightarrow u_{n,n}=b_{n,n}$$

Q9 Solution(3)

$$L_B=\begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ l_{2,1} & 1 & 0 & 0 & 0 & 0 \\ l_{3,1} & l_{3,2} & 1 & 0 & 0 & 0 \\ \vdots & \vdots & \cdots & \ddots & 0 & 0 \\ l_{n-1,1} & l_{n-1,2} & \cdots & l_{n-1,n-2} & 1 & 0 \\ ? & ? & ? & ? & ? & 1 \end{bmatrix} \quad U_B=\begin{bmatrix} u_{1,1} & u_{1,2} & u_{1,3} & \cdots & u_{1,n-1} & 0 \\ 0 & u_{2,2} & u_{2,3} & \cdots & u_{2,n-1} & 0 \\ 0 & 0 & u_{3,3} & \cdots & u_{3,n-1} & 0 \\ 0 & 0 & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & u_{n-1,n-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & b_{n,n} \end{bmatrix} \quad B=\begin{bmatrix} & & A & & 0 \\ & & & & 0 \\ & & & & \vdots \\ & & & & 0 \\ b_{n,1} & b_{n,2} & \cdots & b_{n,n-1} & b_{n,n} \end{bmatrix}$$

$$l_{n,1}\times u_{1,1}+?\times 0+?\times 0+\ldots+1\times 0=b_{n,1} \quad \Rightarrow l_{n,1}=\frac{b_{n,1}}{u_{1,1}}$$

$$l_{n,1}\times u_{1,2}+l_{n,2}\times u_{2,2}+?\times 0+?\times 0+\ldots+1\times 0=b_{n,2} \quad \Rightarrow l_{n,2}=\frac{b_{n,2}-l_{n,1}\times u_{1,2}}{u_{2,2}}$$

$$l_{n,1}\times u_{1,3}+l_{n,2}\times u_{2,3}+l_{n,3}\times u_{3,3}+?\times 0+\ldots+1\times 0=b_{n,3} \quad \Rightarrow l_{n,3}=\frac{b_{n,3}-l_{n,1}\times u_{1,3}-l_{n,2}\times u_{2,3}}{u_{3,3}}$$

$$l_{n,j}=\frac{b_{n,j}-l_{n,1}\times u_{1,j}-l_{n,2}\times u_{2,j}-\ldots-l_{n,j-1}\times u_{j-1,j}}{u_{j,j}}$$

Q9 Solution(4)

- Pseudocode

```
for i = 1 to n-1{                                    //Copy LA and LA
        for j = 1 to n-1{
                LB[i][j]=LA[i][j];
                UB[i][j]=UA[i][j];
        }
        UB[i][n]=0;
        UB[n][i]=0;
        LB[i][n]=0;
}
UB[n][n] = 1;
LB[n][n] = 1;
for i = 1 to n-1{
        LB[n][i] = B[n][i]/UB[i][i];         //Compute eliminating factor
        for j = i+1 to n-1
                B[n][j]=B[n][j] – LB[n][i]*B[i][j];
}
```

Q & A