Dr. D. Sinden

CTMS-MAT-13: Numerical Methods

Exam & Solutions: Saturday 29 August 2026

All questions carry equal marks. Only five questions will be marked. Please use the booklet provided, clearly indicating in the inside cover which questions are to be marked.

Note that all trigonometric values should be expressed in radians.

Question 1:

(a) Give one criteria for a square matrix to be *singular*. [2]

(b) If a matrix is singular, what does this imply about the number of solutions of $A\vec{x} = \vec{b}$? [2]

(c) Given the matrix

$$A = \begin{pmatrix} 2 & 1 & -1 \\ -3 & -1 & 2 \\ -2 & 1 & 2 \end{pmatrix},$$

use Gaussian elimination to reduce A to row echelon (upper triangular) form U. [7]

(d) By applying Gaussian elimination, or any other method, show that the solution to the linear system $A\vec{x} = \vec{b}$ with

$$\vec{b} = \begin{pmatrix} 8 \\ -11 \\ -3 \end{pmatrix} \quad \text{is} \quad \vec{x} = \begin{pmatrix} 2 \\ 3 \\ -1 \end{pmatrix}.$$

[7]

(e) For an $n \times n$ invertible matrix, state the order (in n) of the number of arithmetic operations required to reduce the matrix to upper triangular form by Gaussian elimination. [2]

(a) A square matrix A is singular if it is not invertible, equivalently if its determinant is zero, $\det(A) = 0$. This means its rows (or columns) are linearly dependent, and it has at least one zero eigenvalue.

(b) If A is singular then $A\vec{x} = \vec{b}$ does not have a unique solution: depending on $\vec{b}$ it has either no solution (if $\vec{b}$ is not in the column space of A) or infinitely many solutions.

(c) Work on the augmented system so that the same operations may be reused in part (d). The augmented matrix is

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ -3 & -1 & 2 & -11 \\ -2 & 1 & 2 & -3 \end{array}\right).$$

Eliminate the first column below the pivot $a_{1,1} = 2$ using $R_2 \to R_2 + \frac{3}{2}R_1$ and $R_3 \to R_3 + R_1$:

$$\left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & \frac{1}{2} & \frac{1}{2} & 1 \\ 0 & 2 & 1 & 5 \end{array}\right).$$

Now eliminate the entry below the second pivot $a_{2,2} = \frac{1}{2}$ using $R_3 \to R_3 - 4R_2$:

$$U = \left(\begin{array}{ccc|c} 2 & 1 & -1 & 8 \\ 0 & \frac{1}{2} & \frac{1}{2} & 1 \\ 0 & 0 & -1 & 1 \end{array}\right), \qquad U = \begin{pmatrix} 2 & 1 & -1 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & -1 \end{pmatrix}.$$

(d) From part (c) the transformed right-hand side is $\vec{b}' = (8, 1, 1)^T$. Back substitution on $U\vec{x} = \vec{b}'$ gives

$$\begin{aligned} -x_3 = 1 \ &\Rightarrow\ x_3 = -1, \\ \tfrac{1}{2}x_2 + \tfrac{1}{2}(-1) = 1 \ &\Rightarrow\ x_2 = 3, \\ 2x_1 + (3) - (-1) = 8 \ &\Rightarrow\ x_1 = 2. \end{aligned}$$

Hence $\vec{x} = (2, 3, -1)^T$, as required.

The alternative of performing matrix-vector multiplication was also accepted.

(e) Reducing an $n \times n$ matrix to upper triangular form by Gaussian elimination requires $\mathcal{O}(n^3)$ arithmetic operations (approximately $\frac{2}{3}n^3$).

Question 2:

(**a**) Find the Jacobian matrix $J = J(x, y)$ for the vector-valued function

$$\vec{f}(x, y) = \begin{pmatrix} x^2 + y^2 - 4 \\ xy - 1 \end{pmatrix}.$$

[4]

(**b**) Show that the inverse of the Jacobian is

$$J^{-1} = \frac{1}{|J|} \begin{pmatrix} x & -2y \\ -y & 2x \end{pmatrix}, \qquad \text{where} \quad |J| = 2x^2 - 2y^2.$$

[6]

(**c**) Let $\vec{u}_n = (x_n, y_n)^T$. Using Newton's method, $\vec{u}_{n+1} = \vec{u}_n - J^{-1}(\vec{u}_n)\,\vec{f}(\vec{u}_n)$, with initial guess $\vec{u}_0 = (2, 1)^T$, show that the first iterate is $\vec{u}_1 = (2, \frac{1}{2})^T$ and compute the second iterate $\vec{u}_2$. [10]

(**a**) Writing $\vec{f} = (f_1, f_2)^T$ with $f_1 = x^2 + y^2 - 4$ and $f_2 = xy - 1$, the partial derivatives are

$$\frac{\partial f_1}{\partial x} = 2x, \quad \frac{\partial f_1}{\partial y} = 2y, \quad \frac{\partial f_2}{\partial x} = y, \quad \frac{\partial f_2}{\partial y} = x,$$

so that

$$J = \begin{pmatrix} 2x & 2y \\ y & x \end{pmatrix}.$$

(**b**) The determinant is

$$|J| = (2x)(x) - (2y)(y) = 2x^2 - 2y^2.$$

For a 2×2 matrix $\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ the inverse is $\frac{1}{ad-bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$, hence

$$J^{-1} = \frac{1}{2x^2 - 2y^2} \begin{pmatrix} x & -2y \\ -y & 2x \end{pmatrix},$$

as required.

(**c**) First iterate: At $\vec{u}_0 = (2, 1)^T$,

$$\vec{f}(\vec{u}_0) = \begin{pmatrix} 4 + 1 - 4 \\ 2 - 1 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \qquad |J| = 2(4) - 2(1) = 6,$$

$$J^{-1}(\vec{u}_0) = \frac{1}{6} \begin{pmatrix} 2 & -2 \\ -1 & 4 \end{pmatrix}, \qquad J^{-1}\vec{f} = \frac{1}{6} \begin{pmatrix} 2 - 2 \\ -1 + 4 \end{pmatrix} = \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix}.$$

Therefore

$$\vec{u}_1 = \vec{u}_0 - J^{-1}(\vec{u}_0)\vec{f}(\vec{u}_0) = \begin{pmatrix} 2 \\ 1 \end{pmatrix} - \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} 2 \\ \frac{1}{2} \end{pmatrix}.$$

Second iterate: At $\vec{u}_1 = (2, \frac{1}{2})^T$,

$$\vec{f}(\vec{u}_1) = \begin{pmatrix} 4 + \frac{1}{4} - 4 \\ 1 - 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{4} \\ 0 \end{pmatrix}, \qquad |J| = 2(4) - 2\left(\tfrac{1}{4}\right) = \tfrac{15}{2},$$

$$J^{-1}(\vec{u}_1) = \frac{1}{15/2}\begin{pmatrix} 2 & -1 \\ -\frac{1}{2} & 4 \end{pmatrix}, \qquad J^{-1}\vec{f} = \frac{2}{15}\begin{pmatrix} 2\cdot\frac{1}{4} \\ -\frac{1}{2}\cdot\frac{1}{4} \end{pmatrix} = \frac{2}{15}\begin{pmatrix} \frac{1}{2} \\ -\frac{1}{8} \end{pmatrix} = \begin{pmatrix} \frac{1}{15} \\ -\frac{1}{60} \end{pmatrix}.$$

Hence

$$\vec{u}_2 = \vec{u}_1 - J^{-1}(\vec{u}_1)\vec{f}(\vec{u}_1) = \begin{pmatrix} 2 \\ \frac{1}{2} \end{pmatrix} - \begin{pmatrix} \frac{1}{15} \\ -\frac{1}{60} \end{pmatrix} = \boxed{\begin{pmatrix} \frac{29}{15} \\ \frac{31}{60} \end{pmatrix} \approx \begin{pmatrix} 1.93333 \\ 0.51667 \end{pmatrix}}.$$

(The iteration is converging to the exact root $\left(\sqrt{2+\sqrt{3}},\ \sqrt{2-\sqrt{3}}\right) \approx (1.93185, 0.51764)$.)

Question 3:

Consider the integral

$$I = \int_0^1 e^x \, \mathrm{d}x = e - 1 = 1.718281828459045.$$

(a) Given the Trapezium rule

$$I_n = \frac{h}{2}\left(f(x_0) + f(x_n) + 2\sum_{i=1}^{n-1} f(x_i)\right), \qquad h = \frac{b-a}{n},$$

show that the approximations for $n = 2^k$, $k = 0, 1, 2$, are

k	n	I_n
0	1	1.85914091
1	2	1.75393109
2	4	1.72722190

[8]

(b) Noting the Trapezium rule has error behaviour $I = I_n + a_1 h^2 + a_2 h^4 + \cdots$, and considering the difference of the errors for h and $h/2$, derive the Romberg formula

$$R_k^1 = \frac{1}{3}\left(4R_k^0 - R_{k-1}^0\right), \qquad R_0^0 = I_1,\ R_1^0 = I_2,\ \ldots$$

[6]

(c) Using the values above as $R_k^0 = I_{2^k}$, compute R_1^1, R_2^1 and hence, using $R_2^2 = \frac{1}{15}\left(16R_2^1 - R_1^1\right)$, show the second Romberg extrapolation is given by $R_2^2 = 1.718283$. [6]

(a) Here $f(x) = e^x$, $a = 0$, $b = 1$.

For $k = 0$, $n = 1$, $h = 1$, the nodes are $x_0 = 0$, $x_1 = 1$:

$$I_1 = \frac{1}{2}\left(f(0) + f(1)\right) = \frac{1}{2}(1 + e) = 1.85914091.$$

For $k = 1$, $n = 2$, $h = \frac{1}{2}$, the nodes are $x_0 = 0,\ x_1 = \frac{1}{2},\ x_2 = 1$:

$$I_2 = \frac{1/2}{2}\left(f(0) + 2f\left(\tfrac{1}{2}\right) + f(1)\right) = \frac{1}{4}\left(1 + 2e^{1/2} + e\right) = 1.75393109.$$

For $k = 2$, $n = 4$, $h = \frac{1}{4}$, the nodes are $x_i = \frac{i}{4}$, $i = 0, \ldots, 4$:

$$I_4 = \frac{1/4}{2}\left(f(0) + 2\left(f\left(\tfrac{1}{4}\right) + f\left(\tfrac{1}{2}\right) + f\left(\tfrac{3}{4}\right)\right) + f(1)\right) = \frac{1}{8}\left(1 + 2\left(e^{1/4} + e^{1/2} + e^{3/4}\right) + e\right) = 1.72722190.$$

(b) Let I_h denote the Trapezium approximation with step h. The error expansion is

$$I = I_h + a_1 h^2 + a_2 h^4 + \cdots,$$

and halving the step,

$$I = I_{h/2} + a_1\left(\frac{h}{2}\right)^2 + a_2\left(\frac{h}{2}\right)^4 + \cdots = I_{h/2} + \frac{a_1}{4}h^2 + \cdots.$$

Multiplying the second expression by 4 and subtracting the first eliminates the h^2 term:

$$4I - I = \left(4I_{h/2} - I_h\right) + \left(4 \cdot \tfrac{a_1}{4} - a_1\right)h^2 + \mathcal{O}(h^4) = \left(4I_{h/2} - I_h\right) + \mathcal{O}(h^4).$$

Hence

$$I = \frac{1}{3}\left(4I_{h/2} - I_h\right) + \mathcal{O}(h^4).$$

Writing $R_k^0 = I_{2^k}$ (so that a larger k corresponds to the halved step $h/2$), this is

$$R_k^1 = \frac{1}{3}\left(4R_k^0 - R_{k-1}^0\right).$$

(c) Using $R_0^0 = 1.85914091$, $R_1^0 = 1.75393109$ and $R_2^0 = 1.72722190$,

$$R_1^1 = \frac{1}{3}\,(4(1.75393109) - 1.85914091) = 1.71886115,$$
$$R_2^1 = \frac{1}{3}\,(4(1.72722190) - 1.75393109) = 1.71831884.$$

The second extrapolation removes the h^4 term:

$$R_2^2 = \frac{1}{15}\left(16R_2^1 - R_1^1\right) = \frac{1}{15}\,(16(1.71831884) - 1.71886115) = \boxed{1.718283}\,.$$

This agrees with the exact value $e - 1 = 1.718282$ to six significant figures (error $\approx 8.6 \times 10^{-7}$).

Question 4:

(a) For the numerical solution of ordinary differential equations $y' = f(y,t)$, explain the difference between an *explicit* and an *implicit* method. [3]

(b) For the second-order linear ordinary differential equation

$$y''(t) = -3y'(t) - 2y(t) \qquad \text{with} \quad y(0) = 1 \quad \text{and} \quad y'(0) = 0,$$

write the equation as a first-order system in $\vec{u} = (y, y')^T$, i.e. $\vec{u}' = A\vec{u}$, and give the matrix A. [4]

(c) Show that the forward Euler scheme can be written $\vec{u}_{n+1} = (I + hA)\vec{u}_n$, and hence compute the first two steps with $h = 0.1$, i.e. $\vec{u}_1$ and $\vec{u}_2$. [7]

(d) Show that the backward Euler scheme yields $\vec{u}_{n+1} = (I - hA)^{-1}\vec{u}_n$, and compute the first step $\vec{u}_1$ with $h = 0.1$. [6]

(a) An *explicit* method computes the new value $\vec{u}_{n+1}$ directly from quantities already known at t_n (e.g. forward Euler, $\vec{u}_{n+1} = \vec{u}_n + hf(\vec{u}_n, t_n)$); no equation needs to be solved. An *implicit* method defines $\vec{u}_{n+1}$ through an equation in which $\vec{u}_{n+1}$ itself appears on the right-hand side (e.g. backward Euler, $\vec{u}_{n+1} = \vec{u}_n + hf(\vec{u}_{n+1}, t_{n+1})$), so a (generally nonlinear) system must be solved at each step.

(b) Set $u_1 = y$ and $u_2 = y'$. Then $u_1' = u_2$ and $u_2' = y'' = -3y' - 2y = -2u_1 - 3u_2$. In vector form $\vec{u} = (y, y')^T$ satisfies $\vec{u}' = A\vec{u}$ with

$$A = \begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix}, \qquad \vec{u}(0) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

(c) Forward Euler applied to $\vec{u}' = A\vec{u}$ is

$$\vec{u}_{n+1} = \vec{u}_n + h\,A\vec{u}_n = (I + hA)\vec{u}_n.$$

With $h = 0.1$,

$$I + hA = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + 0.1\begin{pmatrix} 0 & 1 \\ -2 & -3 \end{pmatrix} = \begin{pmatrix} 1 & 0.1 \\ -0.2 & 0.7 \end{pmatrix}.$$

Starting from $\vec{u}_0 = (1, 0)^T$,

$$\begin{aligned} \vec{u}_1 &= \begin{pmatrix} 1 & 0.1 \\ -0.2 & 0.7 \end{pmatrix}\begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ -0.2 \end{pmatrix}, \\ \vec{u}_2 &= \begin{pmatrix} 1 & 0.1 \\ -0.2 & 0.7 \end{pmatrix}\begin{pmatrix} 1 \\ -0.2 \end{pmatrix} \\ &= \begin{pmatrix} 1 - 0.02 \\ -0.2 - 0.14 \end{pmatrix} \\ &= \boxed{\begin{pmatrix} 0.98 \\ -0.34 \end{pmatrix}}. \end{aligned}$$

(d) Backward Euler applied to $\vec{u}' = A\vec{u}$ evaluates the right-hand side at t_{n+1}:

$$\vec{u}_{n+1} = \vec{u}_n + h\,A\vec{u}_{n+1} \;\Rightarrow\; (I - hA)\vec{u}_{n+1} = \vec{u}_n \;\Rightarrow\; \vec{u}_{n+1} = (I - hA)^{-1}\vec{u}_n.$$

With $h = 0.1$,

$$I - hA = \begin{pmatrix} 1 & -0.1 \\ 0.2 & 1.3 \end{pmatrix}, \qquad \det(I - hA) = 1.3 - (-0.1)(0.2) = 1.32,$$

$$(I - hA)^{-1} = \frac{1}{1.32}\begin{pmatrix} 1.3 & 0.1 \\ -0.2 & 1 \end{pmatrix} = \begin{pmatrix} 0.984848 & 0.075758 \\ -0.151515 & 0.757576 \end{pmatrix}.$$

Starting from $\vec{u}_0 = (1, 0)^T$,

$$\vec{u}_1 = (I - hA)^{-1}\vec{u}_0 = \boxed{\begin{pmatrix} 0.984848 \\ -0.151515 \end{pmatrix}}.$$

Question 5:

Given the matrix

$$A = \begin{pmatrix} 16 & 4 & 8 \\ 4 & 5 & -4 \\ 8 & -4 & 22 \end{pmatrix},$$

what is the Cholesky matrix L (with $A = LL^T$) for the matrix A?

- ☒ $\begin{pmatrix} 4 & 0 & 0 \\ 1 & 2 & 0 \\ 2 & -3 & 3 \end{pmatrix}$
- ○ $\begin{pmatrix} 4 & 1 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 3 \end{pmatrix}$
- ○ $\begin{pmatrix} 2 & 0 & 0 \\ 1 & 2 & 0 \\ 2 & -3 & 4 \end{pmatrix}$
- ○ $\begin{pmatrix} 4 & 0 & 0 \\ 1 & 2 & 0 \\ 2 & 3 & 3 \end{pmatrix}$
- ○ $\begin{pmatrix} 8 & 0 & 0 \\ 1 & 2 & 0 \\ 2 & -3 & 3 \end{pmatrix}$
- ○ $\begin{pmatrix} 4 & 0 & 0 \\ 2 & 1 & 0 \\ 2 & -3 & 3 \end{pmatrix}$

Seek the lower-triangular factor $L = (l_{i,j})$ with $A = LL^T$. Matching entries of A column by column:

$$\begin{aligned}
a_{1,1} = 16 = l_{1,1}^2 \quad & \Rightarrow \; l_{1,1} = 4, \\
a_{2,1} = 4 = l_{2,1} l_{1,1} \quad & \Rightarrow \; l_{2,1} = 4/4 = 1, \\
a_{3,1} = 8 = l_{3,1} l_{1,1} \quad & \Rightarrow \; l_{3,1} = 8/4 = 2, \\
a_{2,2} = 5 = l_{2,1}^2 + l_{2,2}^2 \quad & \Rightarrow \; l_{2,2} = \sqrt{5-1} = 2, \\
a_{3,2} = -4 = l_{3,1} l_{2,1} + l_{3,2} l_{2,2} \quad & \Rightarrow \; l_{3,2} = \frac{-4-(2)(1)}{2} = -3, \\
a_{3,3} = 22 = l_{3,1}^2 + l_{3,2}^2 + l_{3,3}^2 \quad & \Rightarrow \; l_{3,3} = \sqrt{22-4-9} = 3.
\end{aligned}$$

Hence

$$L = \boxed{\begin{pmatrix} 4 & 0 & 0 \\ 1 & 2 & 0 \\ 2 & -3 & 3 \end{pmatrix}}$$

which is the first option.

A quick check confirms $LL^T = A$.

$$\begin{aligned}
LL^T &= \begin{pmatrix} 4 & 0 & 0 \\ 1 & 2 & 0 \\ 2 & -3 & 3 \end{pmatrix} \begin{pmatrix} 4 & 1 & 2 \\ 0 & 2 & -3 \\ 0 & 0 & 3 \end{pmatrix} \\
&= \begin{pmatrix} 4\cdot 4 & 4\cdot 1 & 4\cdot 2 \\ 1\cdot 4 & 1\cdot 1 + 2\cdot 2 & 1\cdot 2 - 2\cdot 3 \\ 2\cdot 4 & 2\cdot 1 - 3\cdot 2 & 2\cdot 2 + 3\cdot 3 + 3\cdot 3 \end{pmatrix} \\
&= \begin{pmatrix} 16 & 4 & 8 \\ 4 & 1+4 & 2-6 \\ 8 & 2-6 & 4+9+9 \end{pmatrix} & = A.
\end{aligned}$$

Question 6:

Given the following data:

i	0	1	2
x_i	0	1	3
y_i	1	2	1

Using polynomial interpolation, such as Lagrange or Newton interpolation, what is the value of $y(2)$?

- ◯ 3/2
- ◯ 7/3
- ◯ 11/6
- ☒ 2
- ◯ 5/2
- ◯ 1

Using Newton's divided differences with $x_0 = 0,\ x_1 = 1,\ x_2 = 3$ and $y_0 = 1,\ y_1 = 2,\ y_2 = 1$:

$$
\begin{aligned}
f[x_0] &= 1, \\
f[x_0, x_1] &= \frac{y_1 - y_0}{x_1 - x_0} = \frac{2-1}{1-0} = 1, \\
f[x_1, x_2] &= \frac{y_2 - y_1}{x_2 - x_1} = \frac{1-2}{3-1} = -\frac{1}{2}, \\
f[x_0, x_1, x_2] &= \frac{f[x_1, x_2] - f[x_0, x_1]}{x_2 - x_0} = \frac{-\frac{1}{2} - 1}{3-0} = -\frac{1}{2}.
\end{aligned}
$$

The interpolating polynomial is

$$
\begin{aligned}
p(x) &= f[x_0] + f[x_0, x_1](x - x_0) + f[x_0, x_1, x_2](x - x_0)(x - x_1) \\
&= 1 + x - \frac{1}{2}x(x-1) \\
&= 1 + \frac{3}{2}x - \frac{1}{2}x^2.
\end{aligned}
$$

Evaluating at $x = 2$,

$$
\begin{aligned}
p(2) &= 1 + 2 - \frac{1}{2}(2)(1) \\
&= 3 - 1 \\
&= \boxed{2.}
\end{aligned}
$$

The Lagrange basis polynomials for $x_0 = 0,\ x_1 = 1,\ x_2 = 3$ are

$$
\begin{aligned}
l_0(x) &= \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} = \frac{(x-1)(x-3)}{(0-1)(0-3)} = \frac{(x-1)(x-3)}{3}, \\
l_1(x) &= \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} = \frac{x(x-3)}{(1-0)(1-3)} = -\frac{x(x-3)}{2}, \\
l_2(x) &= \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} = \frac{x(x-1)}{(3-0)(3-1)} = \frac{x(x-1)}{6}.
\end{aligned}
$$

The interpolating polynomial is

$$\begin{aligned} p(x) &= y_0 l_0(x) + y_1 l_1(x) + y_2 l_2(x) \\ &= l_0(x) + 2l_1(x) + l_2(x) \\ &= \frac{(x-1)(x-3)}{3} - x(x-3) + \frac{x(x-1)}{6} \\ &= \frac{1}{3}x^2 - \frac{4}{3}x + 1 - x^2 + 3x + \frac{1}{6}x^2 - \frac{1}{6}x \\ &= 1 + \left(3 - \frac{4}{3} - \frac{1}{6}\right) x + \left(\frac{1}{3} + \frac{1}{6} - 1\right) x^2 \\ &= 1 + \frac{3}{2}x - \frac{1}{2}x^2. \end{aligned}$$

Evaluating the basis polynomials at $x = 2$,

$$l_0(2) = \frac{(1)(-1)}{3} = -\frac{1}{3}, \qquad l_1(2) = -\frac{(2)(-1)}{2} = 1, \qquad l_2(2) = \frac{(2)(1)}{6} = \frac{1}{3},$$

so that

$$\begin{aligned} p(2) &= 1 \cdot \left(-\frac{1}{3}\right) + 2 \cdot (1) + 1 \cdot \frac{1}{3} \\ &= -\frac{1}{3} + 2 + \frac{1}{3} \\ &= 2 \end{aligned}$$

in agreement with Newton interpolation.

Question 7:

Using the Jacobi scheme $\vec{x}_{n+1} = (I - D^{-1}A)\vec{x}_n + D^{-1}\vec{b}$, where D is the diagonal of A, find the residual $\|\vec{x}^* - \vec{x}_2\|_2$ (the Euclidean, or L^2, norm) between the exact solution $\vec{x}^*$ and the second Jacobi iterate $\vec{x}_2$, starting from $\vec{x}_0 = (0, 0)^T$, where

$$A = \begin{pmatrix} 3 & 1 \\ 1 & 3 \end{pmatrix} \qquad \text{and} \qquad \vec{b} = \begin{pmatrix} 5 \\ 5 \end{pmatrix}.$$

- [x] $\dfrac{5\sqrt{2}}{36}$
- [ ] $\dfrac{5\sqrt{2}}{18}$
- [ ] $\dfrac{25\sqrt{2}}{36}$
- [ ] $\dfrac{5}{36}$
- [ ] $\dfrac{5}{18}$
- [ ] $\dfrac{5}{9}$

By symmetry $\vec{x}^* = A^{-1}\vec{b}$ solves $3x_1 + x_2 = 5$, $x_1 + 3x_2 = 5$, giving $x_1 = x_2 = \frac{5}{4}$, i.e. $\vec{x}^* = \left(\frac{5}{4}, \frac{5}{4}\right)^T$. The diagonal of A is $D = 3I$, so $D^{-1} = \frac{1}{3}I$ and

$$I - D^{-1}A = \begin{pmatrix} 0 & -\frac{1}{3} \\ -\frac{1}{3} & 0 \end{pmatrix}, \qquad D^{-1}\vec{b} = \begin{pmatrix} \frac{5}{3} \\ \frac{5}{3} \end{pmatrix}.$$

Starting from $\vec{x}_0 = (0, 0)^T$,

$$\vec{x}_1 = (I - D^{-1}A)\vec{x}_0 + D^{-1}\vec{b} = \begin{pmatrix} \frac{5}{3} \\ \frac{5}{3} \end{pmatrix},$$

$$\vec{x}_2 = \begin{pmatrix} 0 & -\frac{1}{3} \\ -\frac{1}{3} & 0 \end{pmatrix} \begin{pmatrix} \frac{5}{3} \\ \frac{5}{3} \end{pmatrix} + \begin{pmatrix} \frac{5}{3} \\ \frac{5}{3} \end{pmatrix} = \begin{pmatrix} -\frac{5}{9} + \frac{5}{3} \\ -\frac{5}{9} + \frac{5}{3} \end{pmatrix} = \begin{pmatrix} \frac{10}{9} \\ \frac{10}{9} \end{pmatrix}.$$

The error vector is

$$\vec{x}^* - \vec{x}_2 = \begin{pmatrix} \frac{5}{4} - \frac{10}{9} \\ \frac{5}{4} - \frac{10}{9} \end{pmatrix} = \begin{pmatrix} \frac{5}{36} \\ \frac{5}{36} \end{pmatrix},$$

so its Euclidean norm is

$$\|\vec{x}^* - \vec{x}_2\|_2 = \sqrt{\left(\frac{5}{36}\right)^2 + \left(\frac{5}{36}\right)^2} = \frac{5}{36}\sqrt{2} = \boxed{\frac{5\sqrt{2}}{36}} \approx 0.1964.$$

Question 8:

The differential equation

$$y'(t) = y - t^2 + 1 \qquad \text{with} \quad y(0) = \tfrac{1}{2}$$

has the exact solution $y(t) = (t+1)^2 - \frac{1}{2}e^t$. From the classical explicit Runge–Kutta scheme given by the Butcher array

$$\begin{array}{c|cccc} 0 & & & & \\ \frac{1}{2} & \frac{1}{2} & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & \\ 1 & 0 & 0 & 1 & \\ \hline & \frac{1}{6} & \frac{1}{3} & \frac{1}{3} & \frac{1}{6} \end{array}$$

where

$$u_{k+1} = u_k + h\sum_{i=1}^{4} b_i f\left(u_k + h\sum_{j=1}^{4} a_{i,j}k_j, t_k + c_i h\right) = u_k + h\sum_{i=1}^{4} b_i k_i$$

and using step size $h = 0.2$, what is $|y(2h) - u_2|$, i.e. the global truncation error after two steps?

- [x] 1.14×10^{-5}
- [ ] 1.40×10^{-2}
- [ ] 5.83×10^{-4}
- [ ] 2.36×10^{-3}
- [ ] 3.77×10^{-1}
- [ ] 0

Here $f(t, y) = y - t^2 + 1$, $h = 0.2$, $u_0 = 0.5$. The stages, from reading the Butcher array, are

$$\begin{aligned} k_1 &= f(t_k,\, u_k), \\ k_2 &= f(t_k + \tfrac{h}{2},\, u_k + \tfrac{h}{2}k_1), \\ k_3 &= f(t_k + \tfrac{h}{2},\, u_k + \tfrac{h}{2}k_2), \\ k_4 &= f(t_k + h,\, u_k + hk_3). \end{aligned}$$

Step 1 ($t_0 = 0$, $u_0 = 0.5$):

$$\begin{aligned} k_1 &= f(0,\, 0.5) \\ &= 0.5 - 0 + 1 \\ &= 1.50000000, \\ k_2 &= f(0.1,\, 0.5 + 0.1 \cdot 1.5) \\ &= f(0.1,\, 0.65) \\ &= 0.65 - 0.01 + 1 \\ &= 1.64000000, \\ k_3 &= f(0.1,\, 0.5 + 0.1 \cdot 1.64) \\ &= 0.664 - 0.01 + 1 \\ &= 1.65400000, \\ k_4 &= f(0.2,\, 0.5 + 0.2 \cdot 1.654) \\ &= 0.8308 - 0.04 + 1 \\ &= 1.79080000, \\ u_1 &= 0.5 + \tfrac{0.2}{6}\,(k_1 + 2k_2 + 2k_3 + k_4) \\ &= 0.5 + 0.2(1.5 + 2 \cdot 1.64 + 2 \cdot 1.654 + 1.7908)/6 \\ &= 0.82929333. \end{aligned}$$

Step 2 ($t_1 = 0.2$, $u_1 = 0.82929333$):

$$
\begin{aligned}
k_1 &= f(0.2,\, 0.82929333) = 1.78929333,\\
k_2 &= f(0.3,\, 0.82929333 + 0.1 \cdot k_1) = 1.91822267,\\
k_3 &= f(0.3,\, 0.82929333 + 0.1 \cdot k_2) = 1.93111560,\\
k_4 &= f(0.4,\, 0.82929333 + 0.2 \cdot k_3) = 2.05551645,\\
u_2 &= 0.82929333 + \tfrac{0.2}{6}\,(k_1 + 2k_2 + 2k_3 + k_4) = 1.21407621.
\end{aligned}
$$

The exact value at $t = 2h = 0.4$ is

$$y(0.4) = (1.4)^2 - \tfrac{1}{2}e^{0.4} = 1.96 - 0.745912 = 1.21408765.$$

Thus the global truncation error after two steps is

$$|y(0.4) - u_2| = |1.21408765 - 1.21407621| = 1.14 \times 10^{-5}.$$