CTMS-MAT-13: Numerical Methods

Wednesday 27 May 2026: Exam & Solutions

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.

All trigonometric values are in radians.

Question 1:

(a) What three operations form the elementary row operations used in Gaussian elimination? [3]

(b) Given $a = 3$, find the solution to the linear system $A\vec{x} = \vec{b}$ with

$$A = \begin{pmatrix} 1 & 4 & 5 \\ 3 & -1 & 3 \\ 0 & a & 1 \end{pmatrix} \quad \text{and} \quad \vec{b} = \begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix}.$$

[7]

(c) What is a diagonally dominant matrix? [3]

(d) For what value of a does the linear system have no solution. [7]

(a) The three elementary row operations are:

- scaling by a non-zero number
- addition of one row to another
- swapping two rows

(b) With $a = 3$, the system $A\vec{x} = \vec{b}$ is

$$A = \begin{pmatrix} 1 & 4 & 5 \\ 3 & -1 & 3 \\ 0 & 3 & 1 \end{pmatrix}, \qquad \vec{b} = \begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix}.$$

Write the augmented matrix and perform forward elimination.

Step 1. Eliminate x_1 from row 2 using row 1, with multiplier $m_{21} = a_{21}/a_{11} = 3/1 = 3$:

$$R_2 \leftarrow R_2 - 3R_1 = \left(0,\ -1-12,\ 3-15 \mid 2-3\right) = \left(0,\ -13,\ -12 \mid -1\right).$$

Row 3 already has $a_{31} = 0$, so no operation is needed.

After Step 1:

$$\left(\begin{array}{ccc|c} 1 & 4 & 5 & 1 \\ 0 & -13 & -12 & -1 \\ 0 & 3 & 1 & 5 \end{array}\right).$$

Step 2. Eliminate x_2 from row 3 using row 2, with multiplier $m_{32} = a_{32}/a_{22} = 3/(-13) = -3/13$:

$$R_3 \leftarrow R_3 - \left(-\frac{3}{13}\right) R_2 = R_3 + \frac{3}{13} R_2.$$

$$\text{coefficient of } x_3 : \quad 1 + \frac{3}{13}(-12) = 1 - \frac{36}{13} = -\frac{23}{13},$$

DRAFT

$$\text{right-hand side:}\quad 5+\frac{3}{13}(-1)=5-\frac{3}{13}=\frac{62}{13}.$$

After Step 2, the upper triangular system is

$$\left(\begin{array}{ccc|c} 1 & 4 & 5 & 1 \\ 0 & -13 & -12 & -1 \\ 0 & 0 & -\frac{23}{13} & \frac{62}{13} \end{array}\right).$$

Back substitution.

From row 3:

$$-\frac{23}{13}x_3=\frac{62}{13}\quad\Longrightarrow\quad x_3=-\frac{62}{23}.$$

From row 2:

$$-13x_2-12x_3=-1\quad\Longrightarrow\quad x_2=\frac{-1+12x_3}{-13}=\frac{-1+12\left(-\frac{62}{23}\right)}{-13}=\frac{-\frac{767}{23}}{-13}=\frac{767}{299}=\frac{59}{23}.$$

From row 1:

$$x_1+4x_2+5x_3=1\quad\Longrightarrow\quad x_1=1-4x_2-5x_3=1-\frac{236}{23}+\frac{310}{23}=\frac{97}{23}.$$

Hence

$$\boxed{\vec{x}=\left(\frac{97}{23},\ \frac{59}{23},\ -\frac{62}{23}\right)^T}.$$

Check: $x_1+4x_2+5x_3=\frac{97+236-310}{23}=\frac{23}{23}=1$; $3x_1-x_2+3x_3=\frac{291-59-186}{23}=\frac{46}{23}=2$; $3x_2+x_3=\frac{177-62}{23}=\frac{115}{23}=5$.

(c) A square matrix is diagonally dominant if, for every row of the matrix, the magnitude of the diagonal entry in a row is greater than or equal to the sum of the magnitudes of all the other entries in that row. More precisely, the $(n\times n)$-matrix A is diagonally dominant if

$$|a_{ii}|\geq\sum_{j\neq i}^{n}|a_{ij}|\quad\forall i$$

where a_{ij} denotes the entry in the i-th row and j-th column.

(d) The system will have no solution when the determinant of the matrix is zero, or (equivalently) when one row is a linear combination of the others. Repeat the elimination keeping a general:

$$A=\begin{pmatrix} 1 & 4 & 5 \\ 3 & -1 & 3 \\ 0 & a & 1 \end{pmatrix},\qquad \vec{b}=\begin{pmatrix} 1 \\ 2 \\ 5 \end{pmatrix}.$$

Step 1. As before, $R_2\leftarrow R_2-3R_1$ gives $\left(0,\ -13,\ -12\,|\,-1\right)$. Row 3 is unaffected since $a_{31}=0$.

Step 2. Eliminate x_2 from row 3 using row 2, with multiplier $m_{32}=a/(-13)=-a/13$:

$$R_3\leftarrow R_3+\frac{a}{13}R_2.$$

$$\text{coefficient of } x_3:\quad 1+\frac{a}{13}(-12)=\frac{13-12a}{13},$$

$$\text{right-hand side:}\quad 5+\frac{a}{13}(-1)=\frac{65-a}{13}.$$

DRAFT

The final equation is

$$\left(\frac{13 - 12a}{13}\right) x_3 = \frac{65 - a}{13} \quad \leftrightarrow \quad (13 - 12a)\, x_3 = 65 - a.$$

For the system to have no solution, this equation must be inconsistent, i.e. the coefficient of x_3 vanishes while the right-hand side does not:

$$13 - 12a = 0 \quad \rightarrow \quad a = \frac{13}{12}.$$

Checking the right-hand side at $a = \frac{13}{12}$:

$$65 - \frac{13}{12} = \frac{780 - 13}{12} = \frac{767}{12} \neq 0,$$

so the equation becomes $0 \cdot x_3 = \dfrac{767}{12}$, which has no solution.

Hence the linear system has no solution when $\boxed{a = \dfrac{13}{12}}$.

DRAFT

Question 2:

(**a**) For the function

$$f(x) = x^3 - \sinh(x) + 4x^2 + 6x + 9$$

with initial guesses $x_0 = 7$ and $x_1 = 8$, compute the next three iterations of the secant method

$$x_{k+1} = x_k - f(x_k)\,\frac{x_{k-1} - x_k}{f(x_{k-1}) - f(x_k)}.$$

[7]

(**b**) State how the secant method is derived from the Newton method. [3]

(**c**) Compute the first three approximations with Newton method with initial guess $x_0 = 7$. [7]

(**d**) Is the rate of convergence of the secant method sublinear, linear or superlinear? [3]

(**a**) Evaluate f at the two starting points:

$$f(7) = 7^3 - \sinh(7) + 4(7)^2 + 6(7) + 9 = 590 - \sinh(7) \approx 41.683877,$$

$$f(8) = 8^3 - \sinh(8) + 4(8)^2 + 6(8) + 9 = 825 - \sinh(8) \approx -665.478826.$$

Iteration 1 ($k = 1$):

$$x_2 = x_1 - f(x_1)\frac{x_0 - x_1}{f(x_0) - f(x_1)} = 8 - (-665.478826)\frac{7-8}{41.683877 - (-665.478826)} \approx 7.058945.$$

$$f(x_2) = f(7.058945) \approx 20.798251.$$

Iteration 2 ($k = 2$):

$$x_3 = x_2 - f(x_2)\frac{x_1 - x_2}{f(x_1) - f(x_2)} = 7.058945 - (20.798251)\frac{8 - 7.058945}{-665.478826 - 20.798251} \approx 7.087465.$$

$$f(x_3) = f(7.087465) \approx 10.037673.$$

Iteration 3 ($k = 3$):

$$x_4 = x_3 - f(x_3)\frac{x_2 - x_3}{f(x_2) - f(x_3)} = 7.087465 - (10.037673)\frac{7.058945 - 7.087465}{20.798251 - 10.037673} \approx 7.114068.$$

$$f(x_4) = f(7.114068) \approx -0.401497.$$

Summary:

n	x_n	$f(x_n)$
0	7.000000	41.683877
1	8.000000	−665.478826
2	7.058945	20.798251
3	7.087465	10.037673
4	7.114068	−0.401497

The iterates are converging towards the root $r \approx 7.1131$.

(**b**) Newton's method,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)},$$

requires the derivative $f'(x_n)$. The secant method replaces this derivative by the slope of the line through the two most recently computed points, i.e. a finite-difference (divided-difference) approximation

$$f'(x_n) \approx \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}.$$

DRAFT

Substituting this approximation into Newton's formula gives

$$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})} = x_n - f(x_n)\frac{x_{n-1} - x_n}{f(x_{n-1}) - f(x_n)},$$

which is the secant method. Thus the secant method is Newton's method with the derivative replaced by a secant-line (two-point) approximation, removing the need to evaluate f'.

(c) Newton's method is $x_{n+1} = x_n - f(x_n)/f'(x_n)$ with

$$f'(x) = 3x^2 - \cosh(x) + 8x + 6.$$

Start: $x_0 = 7$, with $f(x_0) \approx 41.683877$ and $f'(x_0) \approx -339.317035$.

Approximation 1:

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} = 7 - \frac{41.683877}{-339.317035} \approx 7.122846.$$

$$f(x_1) \approx -3.933005, \qquad f'(x_1) \approx -404.800250.$$

Approximation 2:

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)} = 7.122846 - \frac{-3.933005}{-404.800250} \approx 7.113130.$$

$$f(x_2) \approx -0.026775, \qquad f'(x_2) \approx -399.298351.$$

Approximation 3:

$$x_3 = x_2 - \frac{f(x_2)}{f'(x_2)} = 7.113130 - \frac{-0.026775}{-399.298351} \approx 7.113063.$$

$$f(x_3) \approx -0.000001 \approx 0.$$

The first three Newton approximations (after x_0) are

$$x_1 \approx 7.122846, \qquad x_2 \approx 7.113130, \qquad x_3 \approx 7.113063,$$

converging rapidly (quadratically) to the root $r \approx 7.113063$.

(d) The secant method converges **superlinearly**: there exists $\mu > 0$ such that

$$\lim_{n\to\infty} \frac{|x_{n+1} - r|}{|x_n - r|^\alpha} = \mu, \qquad \alpha = \frac{1+\sqrt{5}}{2} \approx 1.618,$$

i.e. the order of convergence is the golden ratio $\alpha \approx 1.618$, which lies strictly between 1 (linear) and 2 (quadratic, as for Newton's method).

DRAFT

Question 3:

(a) Define the Kronecker delta, δ_{ij}, as used to define the Lagrange interpolating polynomials. [3]

(b) With pairs of data (0,1), (1,2) and (4,2) show that the Lagrange polynomials are

$$\begin{aligned} l_1(x) &= \tfrac{1}{4}(x-1)(x-4) \\ l_2(x) &= \tfrac{1}{3}x(4-x) \\ l_3(x) &= \tfrac{1}{12}x(x-1) \end{aligned}$$

[7]

(c) Either using Lagrange or Newton interpolation, show the interpolating polynomial for the data given above is given by

$$p(x) = -\frac{1}{4}x^2 + \frac{5}{4}x + 1$$

[6]

(d) Given the error of the n-degree polynomial interpolant of a function $f(x)$ through $n+1$ equidistant spaced points is given by

$$f(x) - p_n(x) = \frac{f^{(n+1)}(x)}{(n+1)!} \prod_{i=0}^{n} (x - x_i),$$

describe the Runge phenomena, and how it can be suppressed. [4]

(a) The Kronecker delta is defined by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases}$$

The Lagrange basis polynomials $l_i(x)$ associated with nodes $x_1, \ldots, x_n$ are the unique polynomials of degree $n-1$ satisfying

$$l_i(x_j) = \delta_{ij} \qquad \text{for all } i, j,$$

i.e. $l_i(x_i) = 1$ and $l_i(x_j) = 0$ for $j \neq i$.

(b) Label the nodes $x_1 = 0$, $x_2 = 1$, $x_3 = 4$. The Lagrange basis polynomial for node x_i is

$$l_i(x) = \prod_{\substack{j=1 \\ j \neq i}}^{3} \frac{x - x_j}{x_i - x_j},$$

constructed so that the numerator vanishes at the other nodes and the denominator normalises the value at x_i to 1.

For $i = 1$ ($x_1 = 0$):

$$l_1(x) = \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} = \frac{(x-1)(x-4)}{(0-1)(0-4)} = \frac{(x-1)(x-4)}{4} = \frac{1}{4}(x-1)(x-4).$$

For $i = 2$ ($x_2 = 1$):

$$l_2(x) = \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} = \frac{x(x-4)}{(1-0)(1-4)} = \frac{x(x-4)}{-3} = \frac{1}{3}x(4-x).$$

For $i = 3$ ($x_3 = 4$):

$$l_3(x) = \frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)} = \frac{x(x-1)}{(4-0)(4-1)} = \frac{x(x-1)}{12} = \frac{1}{12}x(x-1).$$

As a check, each l_i takes the value 1 at its own node and 0 at the others, e.g. $l_1(0) = \frac{1}{4}(-1)(-4) = 1$, $l_1(1) = 0$, $l_1(4) = 0$.

DRAFT

(c) The interpolating polynomial is

$$p(x) = \sum_{i=1}^{3} y_i \, l_i(x) = 1 \cdot l_1(x) + 2 \cdot l_2(x) + 2 \cdot l_3(x),$$

using $y_1 = 1$, $y_2 = 2$, $y_3 = 2$. Substituting the polynomials from part (b):

$$p(x) = \frac{1}{4}(x-1)(x-4) + \frac{2}{3}x(4-x) + \frac{1}{6}x(x-1).$$

Expand each term:

$$\begin{aligned} \frac{1}{4}(x-1)(x-4) &= \frac{1}{4}\left(x^2 - 5x + 4\right) = \frac{1}{4}x^2 - \frac{5}{4}x + 1, \\ \frac{2}{3}x(4-x) &= \frac{2}{3}\left(4x - x^2\right) = -\frac{2}{3}x^2 + \frac{8}{3}x, \\ \frac{1}{6}x(x-1) &= \frac{1}{6}\left(x^2 - x\right) = \frac{1}{6}x^2 - \frac{1}{6}x. \end{aligned}$$

Collecting coefficients of x^2:

$$\frac{1}{4} - \frac{2}{3} + \frac{1}{6} = \frac{3}{12} - \frac{8}{12} + \frac{2}{12} = -\frac{3}{12} = -\frac{1}{4}.$$

Collecting coefficients of x:

$$-\frac{5}{4} + \frac{8}{3} - \frac{1}{6} = -\frac{15}{12} + \frac{32}{12} - \frac{2}{12} = \frac{15}{12} = \frac{5}{4}.$$

The constant term is 1 (the other two terms contribute no constant). Hence

$$p(x) = -\frac{1}{4}x^2 + \frac{5}{4}x + 1,$$

as required. (As a check: $p(0) = 1$, $p(1) = -\frac{1}{4} + \frac{5}{4} + 1 = 2$, $p(4) = -4 + 5 + 1 = 2$, matching the data.)

(d) The **Runge phenomenon** is the observation that, for a fixed interval $[a, b]$, interpolating a smooth function at $n+1$ *equally spaced* points with a single polynomial of degree n does not necessarily converge to the function as $n \to \infty$; instead, large oscillations develop near the ends of the interval, and the maximum interpolation error can grow without bound as n increases.

This occurs because, although $f^{(n+1)}(x)/(n+1)!$ may stay bounded (or grow only slowly), the product

$$\prod_{i=0}^{n}(x - x_i)$$

in the error formula becomes very large near the endpoints of $[a, b]$ for equally spaced nodes, so the error term in

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{i=0}^{n}(x - x_i)$$

is not controlled and the error grows with n, particularly near $x = a$ and $x = b$.

The Runge phenomenon can be suppressed by:

- using **non-equally spaced nodes**, in particular **Chebyshev nodes**, which cluster near the endpoints of $[a, b]$ and minimise $\max_x |\prod_{i=0}^{n}(x - x_i)|$;
- using **piecewise low-degree interpolation** (e.g. cubic splines) instead of a single high-degree global polynomial;
- using a **least-squares** polynomial fit of fixed (low) degree rather than exact interpolation through all points.

DRAFT

Question 4:

(a) The differential equation

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

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

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

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 $h = 0.1$, show that to compute u_1, the k coefficients are given by

$$k_1 = -1, \quad k_2 = -0.9, \quad k_3 = -0.91 \quad \text{and} \quad k_4 = -0.818.$$ [7]

(b) What condition on the Butcher array makes a Runge-Kutta scheme implicit? [3]

(c) For an s-stage Runge-Kutta scheme, what condition on the coefficients b_i is necessary for stability? [3]

(d) Find the global truncation error after two steps, i.e. $|y(2h) - u_2|$. [7]

(a) Here $f(y,t) = 1 - 2y$ does not depend on t, $u_0 = y(0) = 1$ and $h = 0.1$. Reading off the coefficients $a_{i,j}$ and c_i from the Butcher array, the stage values are

$$k_1 = f\left(u_0, t_0\right), \quad k_2 = f\left(u_0 + \frac{h}{2}k_1, t_0 + \frac{h}{2}\right), \quad k_3 = f\left(u_0 + \frac{h}{2}k_2, t_0 + \frac{h}{2}\right), \quad k_4 = f\left(u_0 + hk_3, t_0 + h\right).$$

Since f is independent of t, $f(y,t) = 1 - 2y$ for every stage.

Stage 1:

$$k_1 = f(u_0) = 1 - 2(1) = -1.$$

Stage 2:

$$u_0 + \frac{h}{2}k_1 = 1 + (0.05)(-1) = 0.95, \qquad k_2 = f(0.95) = 1 - 2(0.95) = -0.9.$$

Stage 3:

$$u_0 + \frac{h}{2}k_2 = 1 + (0.05)(-0.9) = 0.955, \qquad k_3 = f(0.955) = 1 - 2(0.955) = -0.91.$$

Stage 4:

$$u_0 + hk_3 = 1 + (0.1)(-0.91) = 0.909, \qquad k_4 = f(0.909) = 1 - 2(0.909) = -0.818.$$

Hence

$$k_1 = -1, \qquad k_2 = -0.9, \qquad k_3 = -0.91, \qquad k_4 = -0.818,$$

as required. The update is then

$$u_1 = u_0 + \frac{h}{6}\left(k_1 + 2k_2 + 2k_3 + k_4\right) = 1 + \frac{0.1}{6}\big(-1 - 1.8 - 1.82 - 0.818\big) = 1 - \frac{0.5438}{6} \approx 0.909367.$$

DRAFT

(b) A Runge-Kutta scheme is **implicit** if the coefficient matrix $A = (a_{i,j})$ in the Butcher array is *not* strictly lower triangular, i.e. if there exists some $a_{i,j} \neq 0$ with $j \geq i$. In that case the definition of stage k_i,

$$k_i = f\left(u_n + h\sum_{j=1}^{s} a_{i,j}k_j,\ t_n + c_i h\right),$$

involves k_i itself (when $j = i$) or stages k_j with $j > i$ that have not yet been computed, so each k_i must be found by solving a system of (generally nonlinear) equations. If A is strictly lower triangular ($a_{i,j} = 0$ for $j \geq i$), each k_i can be evaluated directly from previously computed stages and the scheme is **explicit** – as is the case for the RK4 scheme in part (a).

(c) A necessary condition is

$$\sum_{i=1}^{s} b_i = 1.$$

This is precisely the condition under which the local truncation error $\tau(h) = \max_n |\tau_{n+1}(h)| \to 0$ as $h \to 0$, i.e. the method is (at least first-order) consistent. By the Lax equivalence theorem, consistency together with (zero-)stability is necessary and sufficient for convergence, so $\sum_{i=1}^{s} b_i = 1$ is a necessary condition for the scheme to be a stable, convergent numerical method.

(d) From part (a), $u_1 \approx 0.909367$. Apply the RK4 step again with u_1 in place of u_0 to obtain u_2.

Stage values for the second step:

$$k_1 = f(u_1) = 1 - 2(0.909367) = -0.818733,$$

$$u_1 + \frac{h}{2}k_1 = 0.909367 + (0.05)(-0.818733) = 0.868430, \qquad k_2 = 1 - 2(0.868430) = -0.736860,$$

$$u_1 + \frac{h}{2}k_2 = 0.909367 + (0.05)(-0.736860) = 0.872524, \qquad k_3 = 1 - 2(0.872524) = -0.745047,$$

$$u_1 + hk_3 = 0.909367 + (0.1)(-0.745047) = 0.834862, \qquad k_4 = 1 - 2(0.834862) = -0.669724.$$

Then

$$u_2 = u_1 + \frac{h}{6}(k_1 + 2k_2 + 2k_3 + k_4) = 0.909367 + \frac{0.1}{6}\big(-4.452272\big) \approx 0.835162.$$

The exact solution at $t = 2h = 0.2$ is

$$y(2h) = \frac{1}{2}\big(e^{-0.4} + 1\big) \approx \frac{1}{2}(0.670320 + 1) \approx 0.835160.$$

Hence the global truncation error after two steps is

$$|y(2h) - u_2| \approx |0.835160 - 0.835162| \approx 2.11 \times 10^{-6}.$$

This is consistent with RK4 being a fourth-order method: the global error is $\mathcal{O}(h^4)$, and with $h = 0.1$ we expect errors of order 10^{-5}–10^{-6} after a small number of steps.

DRAFT

Question 5:

Using Gaussian elimination, or otherwise, what is the solution to the linear system given by $\boldsymbol{Ax} = \boldsymbol{b}$ with

$$A = \begin{pmatrix} 1 & 0 & 5 \\ 2 & 2 & -3 \\ 0 & 4 & 4 \end{pmatrix} \quad \text{and} \quad \boldsymbol{b} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}?$$

- ○ $\left(\frac{1}{6}, \frac{25}{12}, \frac{1}{6}\right)^T$
- ○ $\left(\frac{1}{6}, 1, \frac{1}{6}\right)^T$
- ○ $(1, 1, 1)^T$
- ○ $\left(\frac{1}{2}, \frac{13}{20}, \frac{1}{10}\right)^T$
- ○ $\left(\frac{1}{2}, 1, \frac{1}{2}\right)^T$
- ○ $\left(\frac{25}{12}, \frac{13}{20}, \frac{25}{12}\right)^T$

Write the augmented matrix and perform forward elimination:

$$\left(\begin{array}{ccc|c} 1 & 0 & 5 & 1 \\ 2 & 2 & -3 & 2 \\ 0 & 4 & 4 & 3 \end{array}\right).$$

Step 1. Eliminate x_1 from row 2 using row 1, with multiplier $m_{21} = a_{21}/a_{11} = 2/1 = 2$:

$$R_2 \leftarrow R_2 - 2R_1 = \big(0,\ 2-0,\ -3-10 \,|\, 2-2\big) = \big(0,\ 2,\ -13 \,|\, 0\big).$$

Row 3 already has $a_{31} = 0$, so no operation is needed. After Step 1:

$$\left(\begin{array}{ccc|c} 1 & 0 & 5 & 1 \\ 0 & 2 & -13 & 0 \\ 0 & 4 & 4 & 3 \end{array}\right).$$

Step 2. Eliminate x_2 from row 3 using row 2, with multiplier $m_{32} = a_{32}/a_{22} = 4/2 = 2$:

$$R_3 \leftarrow R_3 - 2R_2 = \big(0,\ 4-4,\ 4+26 \,|\, 3-0\big) = \big(0,\ 0,\ 30 \,|\, 3\big).$$

After Step 2, the upper triangular system is

$$\left(\begin{array}{ccc|c} 1 & 0 & 5 & 1 \\ 0 & 2 & -13 & 0 \\ 0 & 0 & 30 & 3 \end{array}\right).$$

Back substitution.

From row 3:

$$30x_3 = 3 \quad \Longrightarrow \quad x_3 = \frac{1}{10}.$$

From row 2:

$$2x_2 - 13x_3 = 0 \quad \Longrightarrow \quad x_2 = \frac{13x_3}{2} = \frac{13}{20}.$$

From row 1:

$$x_1 + 5x_3 = 1 \quad \Longrightarrow \quad x_1 = 1 - 5x_3 = 1 - \frac{1}{2} = \frac{1}{2}.$$

Hence

$$\boxed{\vec{x} = \left(\frac{1}{2},\ \frac{13}{20},\ \frac{1}{10}\right)^T}.$$

Check: $x_1 + 5x_3 = \frac{1}{2} + \frac{1}{2} = 1$; $\quad 2x_1 + 2x_2 - 3x_3 = 1 + \frac{13}{10} - \frac{3}{10} = 2$; $\quad 4x_2 + 4x_3 = \frac{13}{5} + \frac{2}{5} = 3$.

DRAFT

Question 6:

Given the integral

$$I = \int_1^2 \frac{1}{4} + \sin(\pi x)\, \mathrm{d}x$$

what is the error of the approximate integral for the composite Simpson's 1/3 rule

$$S_n = \frac{h}{3}\left(f(x_0) + 4\sum_{i=1}^{n/2} f(x_{2i-1}) + 2\sum_{i=1}^{n/2-1} f(x_{2i}) + f(x_n) \right)$$

when using four subintervals?

- ◯ 1.06e-4
- ◯ 2.67e-6
- ◯ 8.44e-7
- ◯ 0.38817
- ◯ 0.38807
- ◯ 8.30e-2

There is an error in the question: the correct answer does not appear. The question was disregarded in the marking for those who attempted it.

Exact value. With $f(x) = \frac{1}{4} + \sin(\pi x)$,

$$I = \int_1^2 \left(\frac{1}{4} + \sin(\pi x)\right) \mathrm{d}x = \left[\frac{x}{4} - \frac{\cos(\pi x)}{\pi}\right]_1^2 = \left(\frac{1}{2} - \frac{1}{\pi}\right) - \left(\frac{1}{4} + \frac{1}{\pi}\right) = \frac{1}{4} - \frac{2}{\pi} \approx -0.38661977.$$

Composite Simpson's rule with $n = 4$. The step size is $h = (2-1)/4 = 0.25$, with nodes $x_0 = 1,\ x_1 = 1.25,\ x_2 = 1.5,\ x_3 = 1.75,\ x_4 = 2$. Evaluating $f(x) = \frac{1}{4} + \sin(\pi x)$:

$$f(x_0) = \frac{1}{4}, \quad f(x_1) = \frac{1}{4} - \frac{\sqrt{2}}{2}, \quad f(x_2) = -\frac{3}{4}, \quad f(x_3) = \frac{1}{4} - \frac{\sqrt{2}}{2}, \quad f(x_4) = \frac{1}{4}.$$

Hence

$$S_4 = \frac{h}{3}\Big(f(x_0) + 4\big(f(x_1) + f(x_3)\big) + 2f(x_2) + f(x_4)\Big) = \frac{0.25}{3}\left(\frac{1}{2} + 8\left(\frac{1}{4} - \frac{\sqrt{2}}{2}\right) - \frac{3}{2}\right) = \frac{1}{12} - \frac{\sqrt{2}}{3} \approx -0.38807119.$$

Error. The error of the approximation is

$$I - S_4 = \left(\frac{1}{4} - \frac{2}{\pi}\right) - \left(\frac{1}{12} - \frac{\sqrt{2}}{3}\right) = \frac{1}{6} - \frac{2}{\pi} + \frac{\sqrt{2}}{3}.$$

Hence

$$\boxed{I - S_4 = \frac{1}{6} - \frac{2}{\pi} + \frac{\sqrt{2}}{3} \approx 1.45 \times 10^{-3}}.$$

DRAFT

Question 7:

The Gauss-Seidel scheme solves the linear equation $A\vec{x} = \vec{b}$ iteratively as

$$(D + L)\vec{x}_{k+1} = -U\vec{x}_k + \vec{b}.$$

where D is the diagonal matrix of A and L is the strictly lower triangular part of A. Find the residual, $|\vec{x}^* - \vec{x}_2|_\infty$ between the exact solution, $\vec{x}^*$, and the Gauss-Seidel solution after two iterates with initial guess $\vec{x}_0 = (1, -1)^T$, where

$$A = \begin{pmatrix} 2 & 1 \\ -1 & 2 \end{pmatrix} \quad \text{and} \quad \vec{b} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

- ○ 0
- ○ 1/5
- ○ 2/5
- ○ 1/2
- ○ 3/4
- ○ 1/4

Split $A = D + L + U$ with

$$D = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \qquad L = \begin{pmatrix} 0 & 0 \\ -1 & 0 \end{pmatrix}, \qquad U = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Then

$$D + L = \begin{pmatrix} 2 & 0 \\ -1 & 2 \end{pmatrix}, \qquad (D + L)^{-1} = \begin{pmatrix} \frac{1}{2} & 0 \\ \frac{1}{4} & \frac{1}{2} \end{pmatrix},$$

and the iteration is $\vec{x}_{k+1} = (D + L)^{-1} \left(-U\vec{x}_k + \vec{b}\right)$.

Iteration 1. With $\vec{x}_0 = (1, -1)^T$:

$$-U\vec{x}_0 + \vec{b} = -\begin{pmatrix} -1 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2 \end{pmatrix},$$

$$\vec{x}_1 = (D + L)^{-1} \begin{pmatrix} 2 \\ 2 \end{pmatrix} = \left(1,\ \frac{3}{2}\right)^T.$$

Iteration 2. With $\vec{x}_1 = (1, 3/2)^T$:

$$-U\vec{x}_1 + \vec{b} = -\begin{pmatrix} 3/2 \\ 0 \end{pmatrix} + \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 2 \end{pmatrix},$$

$$\vec{x}_2 = (D + L)^{-1} \begin{pmatrix} -1/2 \\ 2 \end{pmatrix} = \left(-\frac{1}{4},\ \frac{7}{8}\right)^T.$$

Exact solution. $\det A = 4 + 1 = 5$, so

$$\vec{x}^* = A^{-1}\vec{b} = \frac{1}{5} \begin{pmatrix} 2 & -1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = (0,\ 1)^T.$$

Residual.

$$\vec{x}^* - \vec{x}_2 = \left(0 - \left(-\frac{1}{4}\right),\ 1 - \frac{7}{8}\right)^T = \left(\frac{1}{4},\ \frac{1}{8}\right)^T,$$

so the (infinity-norm) residual is

$$\boxed{|\vec{x}^* - \vec{x}_2|_\infty = \max\left(\frac{1}{4}, \frac{1}{8}\right) = \frac{1}{4}}.$$

DRAFT

Question 8:

For the vector-valued function

$$f(x,y) = \begin{pmatrix} 4x^2 - 20x + \dfrac{1}{4}y^2 - 8 \\ \dfrac{1}{2}xy^2 - 2x - 5y + 8 \end{pmatrix}$$

has the Jacobian matrix $J = J(x,y)$

$$J = \begin{pmatrix} 4(2x-5) & \dfrac{1}{2}y \\ \dfrac{1}{2}y^2 - 2 & xy - 5 \end{pmatrix}.$$

Then, let $\vec{u}_n = (x_n, y_n)^T$, and using Newton's method, $\vec{u}_{n+1} = \vec{u}_n - J^{-1}(\vec{u}_n) f(\vec{u}_n)$, with an initial guess $\vec{u}_0 = (0,0)^T$, show that the first iteration of Newton's method is

- ○ $(0.3, 2.33)^T$
- ○ $(0.0, -1.44)^T$
- ○ $(-0.1, -2.33)^T$
- ○ $(-0.4, 1.76)^T$
- ○ $(-0.1, 3.22)^T$
- ○ $(0.2, 1.22)^T$

Evaluate f and J at $\vec{u}_0 = (0,0)^T$:

$$f(0,0) = \begin{pmatrix} -8 \\ 8 \end{pmatrix}, \qquad J(0,0) = \begin{pmatrix} -20 & 0 \\ -2 & -5 \end{pmatrix}.$$

The Jacobian has determinant $\det J(0,0) = (-20)(-5) - (0)(-2) = 100$, so

$$J^{-1}(0,0) = \frac{1}{100}\begin{pmatrix} -5 & 0 \\ 2 & -20 \end{pmatrix} = \begin{pmatrix} -\dfrac{1}{20} & 0 \\ \dfrac{1}{50} & -\dfrac{1}{5} \end{pmatrix}.$$

Then

$$J^{-1}(0,0)f(0,0) = \begin{pmatrix} -\dfrac{1}{20} & 0 \\ \dfrac{1}{50} & -\dfrac{1}{5} \end{pmatrix}\begin{pmatrix} -8 \\ 8 \end{pmatrix} = \begin{pmatrix} \dfrac{8}{20} \\ -\dfrac{8}{50} - \dfrac{8}{5} \end{pmatrix} = \begin{pmatrix} \dfrac{2}{5} \\ -\dfrac{44}{25} \end{pmatrix}.$$

Hence the first Newton iterate is

$$\vec{u}_1 = \vec{u}_0 - J^{-1}(0,0)f(0,0) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} \dfrac{2}{5} \\ -\dfrac{44}{25} \end{pmatrix} = \begin{pmatrix} -\dfrac{2}{5} \\ \dfrac{44}{25} \end{pmatrix},$$

i.e.

$$\boxed{\vec{u}_1 = (-0.4,\ 1.76)^T}.$$

DRAFT