EXAMINATION PAPER

Examination Session:	Year:	Exam Code:
May/June	2024	MATH2051-WE01

Title: Numerical Analysis II

Time:	3 hours	
Additional Material provided:		
Materials Permitted:		
Calculators Permitted:	Yes	Models Permitted: Casio FX83 series or FX85 series.

Instructions to Candidates:	Answer all questions. Section A is worth 40% and Section B is worth 60%. Within each section, all questions carry equal marks. Students must use the mathematics specific answer book.

Revision:	

ED01/2024

SECTION A

Q1 (a) Define what is meant by a floating-point number having a *finite precision* and a *finite range*.

(b) A floating-point representation uses four *decimal* digits for the fraction part and two *decimal* digits for the exponent, allowing only normalised representation. Determine the ratio largest/smallest representable numbers. (Note: the exponent bias is irrelevant for this question.)

(c) Let $x_0 = 1$ and $x_{k+1} = x_k/3$. If one was to compute x_k using the usual Python `float` (regarded to have smallest representable 2^{-1022} and machine epsilon 2^{-52}), what would one get for x_{500} and x_{1000}? Justify your answers.

Q2 (a) Given $f \in C^\infty(\mathbb{R})$, i.e. f is continuously differentiable as many times as needed, consider the centred difference approximation

$$f'(x) = \frac{f(x+h) - f(x-h)}{2h} + \Phi(x;h).$$

Using Taylor's *theorem*, compute the remainder Φ up to (and including) h^4.

(b) Compute the Taylor expansion of Φ to arbitrary h^n.

(c) With $R_h^{(1)} := [f(x+h) - f(x-h)]/(2h)$, derive the general Richardson extrapolation formula $R_h^{(n)}$ suitable for centred difference.

(d) Given a small fixed h, can one keep on using Richardson iteratively to obtain arbitrarily accurate approximations of $f'(x)$? Justify your answer.

Q3 (a) Use Gaussian elimination on the following matrix

$$A = \begin{pmatrix} 2 & 4 & -2 & -2 \\ 1 & 2 & 4 & -3 \\ -3 & -3 & 8 & -2 \\ -1 & 1 & 6 & -3 \end{pmatrix}$$

to generate an upper triangular matrix U.

(b) The Gaussian elimination algorithm used to produce an upper triangular matrix U from a matrix A and reduce a system $Ax = b$ into an upper triangular form (assuming there is no need for pivoting) is as follows:

Let $A^{(1)} = A$ and $b^{(1)} = b$. Then for each k from 1 to $n-1$, compute a new matrix $A^{(k+1)}$ and right-hand side $b^{(k+1)}$ by the following procedure:

i Define the row multipliers

$$m_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}, \quad i = k+1, \ldots, n.$$

ii Use these to remove the unknown x_k from equations $k+1$ to n, leaving

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_{ik} a_{kj}^{(k)}, \quad b_i^{(k+1)} = b_i^{(k)} - m_{ik} b_k^{(k)}, \quad i, j = k+1, \ldots, n.$$

The final matrix $A^{(n)} = U$ will then be upper triangular.

Determine the number of unique mathematical operations required to perform this algorithm in the general case (assuming all operations are necessary and there is no need to perform operations for quantities which would be zero).

Q4 Consider the following matrix problem: to solve $Ax = b$ for an n-dimensional vector x given an n-dimensional vector b and some n by n matrix A. If we assume there is some floating point representation error δb for b, then it can be shown that for some arbitrary vector norm $\|.\|$

$$\frac{\|\delta x\|}{\|x\|} \leq \kappa(A)\frac{\|\delta b\|}{\|b\|},$$

where $\kappa(A) = \|A\|\|A^{-1}\|$ is the condition number of A and the matrix norms are all induced from the vector norm. Now consider the 1-norm $\|.\|_1$.

(a) Prove that:

$$\|A\|_1 = \max_{j=1,\ldots,n} \sum_{i=1}^{n} |a_{ij}|.$$

(b) A student wishes to solve for x for various b, such that the maximum value of the relative error of representing b is $\|\delta b\|/\|b\| = 1 \times 10^{-10}$. They require a unique solution and their maximum tolerance for a relative error on this solution is 1×10^{-8}. Which of the following three matrices (assumed exact) would meet the student's criteria?

$$A_1 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 1 & 1 & 0 \\ \epsilon & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad A_3 = \begin{pmatrix} 1 & \epsilon & 0 \\ \epsilon & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Here ϵ could be any arbitrarily small positive number (but assumed exactly represented). You should justify your answers.

SECTION B

Q5 (a) For $I \subset \mathbb{R}$, state what is meant by $f : I \to \mathbb{R}$ to be Lipschitz continuous.

(b) Prove that if f is continuously differentiable and I is bounded, then $|f|$ is Lipschitz continuous in I. (Hint: split the integral.)

(c) State (do not prove) the Contraction Mapping Theorem.

(d) Starting from part (c), state and prove the local convergence theorem for the iteration $x_{k+1} = f(x_k)$ with $f \in C^1(\mathbb{R})$.

(e) Prove that the fixed point iteration $x_{k+1} = |\cos x_k|$ converges for *all* $x_0 \in \mathbb{R}$.

Q6 One seeks to interpolate $f(x) = \sin(\pi x/6)$ with nodes $\{0, 1, 3, 5\}$.

(a) Compute a $p \in \mathcal{P}_3$ that agrees with f at those nodes.

(b) Compute explicitly $p(2)$ and $p(4)$, as well as the errors $f - p$ at these points.

(c) Is your p unique? Prove or give a counterexample.

(d) Given that $|(x-0)(x-1)(x-3)(x-5)| \leq 13$ for all $x \in [0, 5]$, compute a bound M for $|f(x) - p(x)|$ valid for all $x \in [0, 5]$.

Q7 Suppose we try to fit the data $(a, 0)$, $(-1, 2)$, $(-1, b)$, $(0, 3)$ with a quadratic polynomial $p_2(x) = c_0 + c_1 x + c_2 x^2$, for $a, b \in \mathbb{R}$.

(a) Write down the (generally) over-determined system of equations that results when we impose that p_2 passes through all four data points.

(b) State the values of (a, b) for which a unique least squares polynomial fit exists. You should use the normal equations $A^\top A\boldsymbol{x} = A^\top \mathbf{b}$ which determine this fit.

(c) If $a = 1$ state the value of b for which the least squares polynomial is symmetric about the y axis.

Q8 Consider bounded functions $u(x), v(x)$, $x \in [0,1]$ together with the inner product (u,v) defined by:

$$(u,v) = \int_0^1 u\, v\, \mathrm{d}x.$$

(a) Show that the functions

$$W_{11}(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1/2 \\ -1 & \text{if } 1/2 < x \leq 1. \end{cases}$$

and $H(x) = 1$. Are orthogonal with respect to this inner product.

(b) Now consider the extended family of functions

$$W_{ni}(x) = \begin{cases} 2^{(n-1)/2} & \text{if } (2i-2)2^{-n} \leq x \leq (2i-1)2^{-n} \\ -2^{(n-1)/2} & \text{if } (2i-1)2^{-n} < x \leq i2^{-(n-1)}, \\ 0 & \text{otherwise,} \end{cases}$$

defined for $i = 1, \ldots 2^{n-1}$. Show that

$$(W_{ni}, W_{mj}) = \delta_{ij}\delta_{mn}.$$

Hint: You should try sketching the functions, it will make the solution clear.

(c) We can define a representation:

$$f(x) = \sum_{n=1}^{\infty}\sum_{j=1}^{2^{n-1}} C_{nj}W_{nj} + C_0 H(x),$$

for constants C_{nj}, C_0. Show that this satisfies a Parseval identity.

(d) Using the representation of f given above (in (c)), we can define an approximation f_3 of f by truncating this sum as follows:

$$f_3(x) = \sum_{n=1}^{3}\sum_{j=1}^{2^{n-1}} C_{nj}W_{nj} + C_0 H(x). \tag{1}$$

Derive an expression for the approximation of the integral

$$I(f) = \int_0^1 f(x)\sin(4\pi x)\mathrm{d}x$$

using (1), you must give this expression in the simplest form possible.