FAMILY NAME: ....Solutions........
OTHER NAME(S): ....
STUDENT NUMBER: ....
SIGNATURE: ....

THE UNIVERSITY OF NEW SOUTH WALES

SCHOOL OF MATHEMATICS AND STATISTICS

Example Class Test 2

MATH2089
Numerical Methods Example Class Test 2

(1) TIME ALLOWED – 50 minutes

(2) TOTAL NUMBER OF QUESTIONS – 4

(3) ANSWER ALL QUESTIONS

(4) THE QUESTIONS ARE OF EQUAL VALUE

(5) THIS PAPER MAY **NOT** BE RETAINED BY THE CANDIDATE

(6) **ONLY** CALCULATORS WITH AN AFFIXED "UNSW APPROVED" STICKER MAY BE USED

(7) Write your answers on this test paper in the space provided.
Ask your tutor if you need more paper.

All answers must be written in ink. Except where they are expressly required pencils may only be used for drawing, sketching or graphical work.

1. a) **[3 marks]** Give the results of the following MATLAB commands when executed on a computer:

i)
```
h = 1e-12;
z = 2 + h > 2
```
Answer:

$h > 2\varepsilon = 4.4 \times 10^{-16}$, so $2+h$ will be greater than 2 when stored on the computer

$\therefore\ z = 1$ (representing logical true)

ii)
```
u = [0  1];
v = u./(u.^2-u)
```
Answer:

$v = [0\ \ 1]./([0\ \ 1].\wedge 2 - [0\ \ 1])$

$= [0\ \ 1]./([0\ \ 1] - [0\ \ 1])$

$= [0\ \ 1]./[0\ \ 0]$

$= [\text{NaN}\ \ \text{Inf}]$

b) **[3 marks]** A technician claims the amount of energy used in a chemical reaction (in appropriate units) is

$$E = 1201.469380194205,$$

and that measurements were made to 5 significant figures.

i) Give an estimate of the relative error in E.
Answer:

5 significant figures $\Rightarrow$ relative error in $E < \frac{1}{2} \times 10^{-5}$

ii) Give an estimate of the absolute error in E.
Answer:

Absolute error $= |E| \times$ Relative error in E

$\approx 1200 \times \frac{1}{2} \times 10^{-5}$

$= 6 \times 10^{-3}$ (or 0.6×10^{-2})

iii) Give the correctly rounded value for E.
Answer:

5 significant figures, so correctly rounded value is

1201.5

c) **[4 marks]** Estimate the size n of the largest n by n matrix that can be stored in 2Gb RAM using double precision floating point arithmetic.

Answer:

Using double precision floating point arithmetic, each element requires 8 bytes of storage

$\therefore$ $n \times n$ matrix requires $8n^2$ bytes

Either: Using 1 Gb = 2^{30} bytes

$$8n^2 = 2 \times 2^{30} = 2^{31}$$

$$\Rightarrow n^2 = 2^{28}$$

$$\Rightarrow n = 2^{14} = 16,384$$

Or: Using 1 Gb = 10^9 bytes

$$8n^2 = 2 \times 10^9$$

$$n^2 = 0.25 \times 10^9$$

$$n = 15,811$$

Notes n = size of matrix must be an integer

2. a) **[6 marks]** If $f \in C^3(\mathbb{R})$ then

$$f'(x) = \frac{-f(x+2h) + 4f(x+h) - 3f(x)}{2h} - \frac{h^2}{3} f'''(\zeta), \quad \zeta \in [x, x+2h]$$

You are **not** required to derive this.

i) Give an expression for the truncation error as a function of the step-size h using "Big-O" $O()$ notation.
Answer:

Truncation error $= -\frac{h^2}{3} f'''(\xi) = O(h^2)$

ii) The rounding error in calculating the finite difference approximation is $O(\frac{\epsilon}{h})$. Estimate the optimal stepsize h^* in terms of the relative machine precision ϵ.
Answer:

Total error = Rounding error + Truncation Error

$= O(\frac{\varepsilon}{h}) + O(h^2)$

Minimize the total error by balancing the two terms

$\frac{\varepsilon}{h} = h^2 \Rightarrow h^3 = \varepsilon \qquad h^* = O(\varepsilon^{1/3})$

iii) When using MATLAB and double precision floating point arithmetic, give an estimate for the optimal stepsize h^*.
Answer:

In double precision $\varepsilon = 2.2 \times 10^{-16}$

$h^* \approx \varepsilon^{1/3} \approx 6 \times 10^{-6}$

b) **[4 marks]** Give MATLAB commands for **EITHER** an anonymous function `myf` **OR** a function M-file `myf.m` to calculate

$$f(x) = xe^{x^2}.$$

Your function should work for an array of inputs `x`, producing an array of output values of the same size.

Answer:

Anonymous function

```
myf = @(x) x.* exp(x.^2)
```

or Matlab function M-file myf.m

```
function f = myf(x)
f = x.* exp(x.^2);
```

Notes 1) Function name must be myf as requested

2) Note the use of .* and .^ element by element operators, so an array x of input values will produce an array of the same size as output.

3. Consider the problem of finding the fourth root $a^{1/4}$ of a real number $a > 1$.

a) **[1 mark]** Convert this into a problem of finding the zero x^* of a **polynomial** $p(x)$.
Answer:

$x = a^{1/4} \iff x^4 = a \quad (a > 1, \text{ real})$

$\iff x^4 - a = 0$

$\therefore p(x) = x^4 - a$

b) **[2 marks]** Prove that p has at least one zero in the interval $(1, a)$
Answer:

p is a polynomial (quartic) so is continuous on $\mathbb{R}$

$p(1) = 1 - a < 0$ as $a > 1$

$p(a) = a^4 - a = a(a^3 - 1) > 0$ as $a > 1$

$\therefore p$ has at least one zero on $(1, a)$

c) **[2 marks]** Prove that p has at most one zero in the interval $(1, a)$
Answer:

$p'(x) = 4x^3 > 4 > 0$ for all $x \in (1, a)$

$\therefore p$ is strictly increasing on $(1, a)$

$\therefore p$ has at most one zero in $(1, a)$.

d) **[4 marks]** Show that Newton's method for finding a zero of $p(x)$ can be written as

$$x_{k+1} = \frac{1}{4}\left(3x_k + \frac{a}{x_k^3}\right).$$

Answer:

Newtons method for $p(x) = 0$ is

$$x_{k+1} = x_k - \frac{p(x_k)}{p'(x_k)}$$

$$= x_k - \frac{(x_k^4 - a)}{4x_k^3} \qquad x_k \neq 0$$

$$= x_k - \frac{1}{4}x_k + \frac{a}{4x_k^3}$$

$$= \frac{1}{4}\left(3x_k + \frac{a}{x_k^3}\right)$$

e) **[1 mark]** Let $e_k = |x_k - x^*|$ for $k = 0, 1, \ldots$ be the errors produced by Newton's method. If $e_4 = 3 \times 10^{-6}$, estimate e_5.

Answer:

As $p \in C^2(\mathbb{R})$ and $p'(x) > 0$ for all $x \in (1, a]$

$\Rightarrow$ $p'(x^*) > 0$ for zero $x^* \in (1, a)$

$\therefore$ x^* is a simple zero

$\therefore$ expect Newton's method to have a quadratic (second order $\nu = 2$) rate of convergence

using $e_5 \approx e_4^2 \approx (3 \times 10^{-6})^2 = 9 \times 10^{-12}$

4. Consider the data values y_j measured at the times t_j for $j = 1, 2, 3, 4, 5$, given in Table 4.1. The data, which is in column vectors `tdat` and `ydat` produces

j	1	2	3	4	5
t_j	0	0.5	1	1.5	2
y_j	8.1	7.2	4.9	3.1	0.2

Table 4.1: Data

the approximation obtained with the following MATLAB commands

```
A = [ones(size(tdat)) tdat tdat.^2];
[m, n] = size(A);
x = A \ ydat
x =
     8.1800
    -1.9800
    -1.0000
```

The data and the approximation are plotted in Figure 4.1.

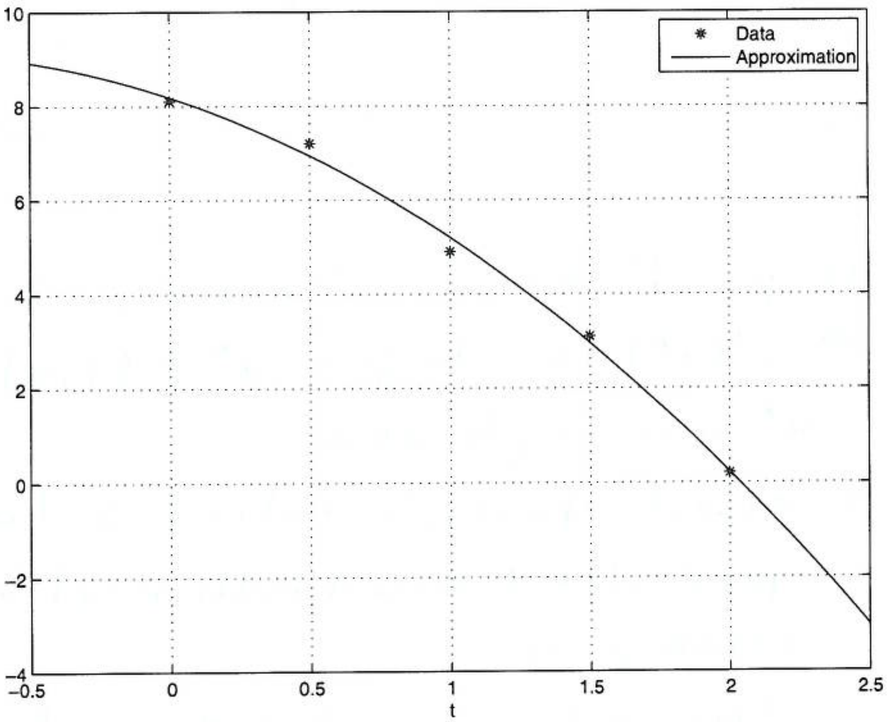

Figure 4.1: Data and approximation

a) **[2 marks]** What are the values of `m` and `n` for this example?
Answer:

m = number of rows in A = number of data values = 5

n = number of columns in A = number of parameters = 3

b) **[1 mark]** It is claimed that the solution $\mathbf{x}$ to the linear system $A\mathbf{x} = \mathbf{y}$ is given by $\mathbf{x} = A^{-1}\mathbf{y}$. Why is this not correct for this example?
Answer:

A is a 5×3 matrix

The inverse A^{-1} only exists for square nonsingular matrices, so A^{-1} is not defined for this matrix A.

c) **[1 mark]** What do the MATLAB commands above calculate?
Answer:

x = A \ ydat

calculates the least squares solution to the linear system $A\underset{\sim}{x} = \underset{\sim}{ydat}$

d) **[2 marks]** Write down the approximation obtained.
Answer:

$$\underset{\sim}{x} = \begin{bmatrix} 8.18 \\ -1.98 \\ 1.00 \end{bmatrix}$$

The columns of A correspond to the basis functions $1, t, t^2$, so the approximating quadratic is

$$y = 8.18 - 1.98t - t^2$$

(which agrees with approximation in Figure 4.1)

e) **[2 marks]** The results of the following MATLAB commands are

```
[m, n] = size(A);
[Q, R] = qr(A);
norm(Q'*Q-eye(m), 1)
ans =
   6.5226e-16
```

It is claimed that this implies that the matrix Q is not orthogonal. Justify or refute this claim.

Answer:

Q is orthogonal $\Leftrightarrow Q^T Q = I \Leftrightarrow \| Q^T Q - I \| = 0$

Here $\| Q^T Q - I \|_1 = 6.5 \times 10^{-16} \approx 3\varepsilon$

As this is a small multiple of the relative machine precision ε, we have to accept that $\| Q^T Q - I \|_1 \approx 0$. Thus Q is orthogonal within the limits of computer arithmetic.

f) **[2 marks]** Show that a square orthogonal matrix Q has condition number $\kappa_2(Q) = 1$.

Answer:

$\kappa_2(Q) = \| Q \|_2 \, \| Q^{-1} \|_2$

For a square orthogonal matrix $Q^{-1} = Q^T$ as $Q^T Q = I = Q Q^T$

$\| Q \|_2 = \max_i \sqrt{\lambda_i(Q^T Q)} = \max_i \sqrt{\lambda_i(I)} = 1$

$\| Q^{-1} \|_2 = \| Q^T \|_2 = \max_i \sqrt{\lambda_i(Q Q^T)} = \max_i \sqrt{\lambda_i(I)} = 1$

$\therefore \kappa_2(Q) = 1 \times 1 = 1.$