MATH 266 June 2000

NUMERICAL ANALYSIS, SOLUTION OF LINEAR EQUATIONS

TIME ALLOWED: TWO HOURS AND A HALF

Instructions to candidates

Full marks may be obtained for FIVE complete answers.

All questions will be marked but only the best **five** counted

This examination contributes 70% towards the final mark. The balance comes from coursework which consists of 9 mini projects each of which contains some theoretical work and some computer practical to illustrate the properties of the methods used.

1. a) Describe the standard computer representation of floating point numbers in binary form. Explain the significance of the parameters E_{max} and E_{min}. How would the computer represent a number too large to be represented normally? What would happen if such a number arose in some calculation?

b) Write down the binary representation of a the smallest representable single precision number greater than 1 and b the largest representable single precision number smaller than 1. [You may assume that in single precision, the mantissa has 24 bits.] Write down the values of $a - 1$ and $1 - b$ in binary and decimal representations.

c) Find the binary and decimal representations of the hexadecimal number 2.3C6EF37 and show that it is approximately equal to $\sqrt{5}$.

Write down the binary, hexadecimal and decimal representations of the number which represents $\sqrt{5}$ stored in single precision in the computer memory. Evaluate the relative error of this stored number. [You may assume that in the single precision representation the mantissa has 24 bits.]

d) Two numbers a and b have associated absolute errors of $\pm\epsilon$ and $\pm\eta$ respectively. Write down the absolute and relative errors in $a + b$, $a - b$, $a * b$ and a/b.

e) Use the standard formula for the solutions of a quadratic equation to find the exact solutions for the equation :

$$x^2 - 17.25x + 74.39 = 0.$$

Now repeat the calculation rounding to four decimal digits at every stage. Work out the absolute and relative errors of each of these approximate solutions. Comment on the size of these errors and explain how they arose.

[20 marks]

2. a) Describe the method of Simple Iteration for finding a solution to the equation $x = g(x)$. Show that the iterates will converge to the solution if $|g'(x)| < 1$ in some region which includes the root. What determines whether the convergence is monotonic or oscillatory?

Show that the equation $x = -\ln x$ has a solution for $0 < x < 1$. Show that the method of Simple Iteration diverges for this root. Show that the equation $x = e^{-x}$ has the same solution as the equation $x = -\ln x$. Can the method of Simple Iteration be used to find this root of $x = e^{-x}$?

b) If the error e_n at the nth stage of the Newton Raphson method for finding a root of the equation $f(x) = 0$ is small enough, show that the error at the $(n+1)$th stage is proportional to e_n^2 provided that at the root, $x = \alpha$, $f'(\alpha) \neq 0$ and $f''(\alpha) \neq 0$.

Show that if $f'(\alpha) = 0$, but $f''(\alpha) \neq 0$, e_{n+1} is proportional to e_n.

[20 marks]

3. a) Describe Euler's Method for finding an approximation to the solution of the differential equation $dy/dt = f(t, y)$ with the initial condition $y(a) = \alpha$. Show that the error in the value of $y(b)$, $b > a$ is proportional to $h = (b - a)/N$, where h is the step length and N is the number of steps between a and b, provided that the function $f(t, y)$ is suitably well behaved.

Show that

$$\frac{d^2y}{dt^2} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial y} f.$$

Describe the Second Order Taylor Series method for finding an approximate solution to the differential equation $dy/dt = f(t, y)$. Deduce that the error in $y(b)$ is proportional to h^2, where h is the step length.

Describe the Modified Euler Method for finding an approximate solution to $dy/dt = f(t, y)$. Show that when this method is applied over one step, it agrees with the Second Order Taylor method to order h^2.

b) Describe the procedure you would adopt in using a shooting method to find a solution to the boundary value problem of the sort:

$$\frac{d^2y}{dx^2} = 7\frac{dy}{dx} + \frac{y}{1 + x^2}, \qquad y(a) = \alpha, \qquad y(b) = \beta$$

[Note: there is no analytic solution to this equation.]

[20 marks]

4. Find a lower triangular matrix L with all its diagonal elements equal to 1 and zeros everywhere above the leading diagonal and an upper triangular matrix U with zeros everywhere below the leading diagonal such that the matrix A given below can be written as the product $A = L\,U$.

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 & 0 \\ 0 & 1 & 3 & 1 & 0 \\ 0 & 0 & 1 & 2 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{pmatrix}.$$

Show that the matrix U can be written as the product $D\,M$, where D is a diagonal matrix whose diagonal elements are 1, 1, 2, 3/2 and 1/3 and M is the transpose of the matrix L.

Find the inverses of the matrices L, D and M and hence or otherwise show that the inverse of the matrix A is

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

Explain the significance of the condition number of a matrix. Using whichever norm you like, find the condition number for the matrix A.

[20 marks]

5. State Gerschgorin's circle theorem on the eigenvalues of a matrix.

Use Gerschgorin's theorem to show that the matrix A below has three distinct eigenvalues which are real and two which could be complex. Find the regions in which these eigenvalues lie.

$$A = \begin{pmatrix} 55.9 & 12.3 & -17.7 & 1.2 & 0.0 \\ 35.3 & 325.5 & -6.5 & 4.2 & 1.1 \\ -7.6 & -5.5 & 67.7 & 21.9 & 2.4 \\ -1.3 & -2.2 & 1.7 & 1.9 & 2.1 \\ 2.2 & 5.3 & 1.7 & -3.5 & -135.5 \end{pmatrix}$$

Describe the power method for obtaining the largest eigenvalue of a matrix and explain the theory of how it works. Using the smallest possible value for the largest sized eigenvalue and the largest possible value for the next largest sized eigenvalue obtained from Gerschgorin's theorem, estimate the rate of convergence of the power method.

Describe the method of inverse iteration for finding eigenvalues of a matrix. Explain how you would use it to evaluate the negative eigenvalue of largest size of the matrix A.

[20 marks]

6. Describe the Jacobi method and the Gauss-Siedel method for finding a solution to the set of linear equations $A\,\mathbf{x} = \mathbf{b}$. What conditions will ensure that these methods converge?

Show that both the Jacobi and Gauss-Siedel methods will converge for the matrix A and column vector $\mathbf{b}$, where

$$A = \begin{pmatrix} 125.0 & 3.0 & 1.5 \\ 2.5 & 75.0 & 5.0 \\ 1.5 & -7.5 & 50.0 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} 100.0 \\ 80.0 \\ 60.0 \end{pmatrix}.$$

Find the first two iterates for the solution of the equation $A.\mathbf{x} = \mathbf{b}$ using the Gauss-Siedel Method rounding to at least 6 decimal digits starting from the vector $\mathbf{x}_0 = (0.8, 1.0, 1.2)^T$.

[20 marks]

7. a) The two point Lagrange Interpolation formula with remainder term for a function $f(x)$ is:

$$f(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1) + \frac{(x - x_0)(x - x_1)}{2} f''(\xi).$$

Use this formula to derive the Trapezoidal Rule for the approximation to the integral:

$$\int_0^h f(x)dx = \frac{h}{2}\Big(f(0) + f(h)\Big) - \frac{h^3}{12} f''(\xi).$$

Describe how you would use the Trapezoidal Rule for an interval divided up into n strips of width h. What is the error in this case?

The Trapezoidal Rule is to be used to find an approximation to the value of the integral:

$$\int_0^{0.1} \frac{dx}{\sqrt{1 - x^4}}$$

with an absolute error not greater than 2×10^{-7}. Show that the greatest value of $f''(x)$ for $0 < x < 0.1$ is approximately 0.06. Use this estimate to find the minimum number of strips to evaluate the integral to the required accuracy.

b) Determine the constants α, β, x_1 and x_2, such that the quadrature rule for the integral

$$\int_{-1}^{1} f(x)\, dx = \alpha f(x_1) + \beta f(x_2)$$

is exact for $f(x) = 1$, $f(x) = x$, $f(x) = x^2$ and $f(x) = x^3$.

How would you adapt this quadrature rule to find an approximate value for $\int_a^b f(x)\, dx$?

[20 marks]