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) 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.

d) Show that the equations (1) and (2) below have the exact solution $x = 2$ and $y = 3$.

$$1.4x + 119.3y = 360.7, \tag{1}$$

$$151.5x + 7.906y = 326.718. \tag{2}$$

Solve these two equations by the method of Gaussian Elimination using equation (1) to eliminate x from equation (2). At *every* stage of the calculation round the numbers to **six** decimal digits. Describe what happens. Explain how one can use Gaussian Elimnation to produce an accurate solution to this problem.

[20 marks]

2. a) Describe the method of Simple Iteration for finding a solution to the equation $x = g(x)$. If α is a solution and $e_n = x_n - \alpha$ is the error of the nth iterate, find a relation between e_{n+1} and e_n which is accurate to first order in e_n when e_n is small. What does this tell us about the convergence of the process?

Show that the equation $x = 5 + \ln x$ has two solutions, one for $x < 1$ and the other for $x > 1$. Show that the method of Simple Iteration converges for one of these roots and diverges for the other. Write the equation in a different form so that Simple Iteration now converges to the root for which it diverged in the original form.

b) Describe the bisection method for finding a solution to the equation $f(x) = 0$. How would you use the method to find the two solutions for the problem in (a)? For each solution, which of the two methods would converge more quickly?

[20 marks]

3. 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 determine the values of a, b, c, d f, g, h and j such that the inverse of the matrix A is

$$A^{-1} = \begin{pmatrix} a & b & 1 & 1 & 1 \\ c & d & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & f & g \\ 1 & 1 & 1 & h & j \end{pmatrix}.$$

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]

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

Use Gerschgorin's theorem to show that the matrix A below has five distinct eigenvalues which are real. Find the regions in which these eigenvalues lie.

$$A = \begin{pmatrix} 175.9 & 22.3 & -16.9 & 3.1 & 0.0 \\ 16.1 & 485.5 & -6.3 & 13.1 & 5.6 \\ -9.5 & -4.5 & 48.5 & 11.7 & 3.6 \\ -2.6 & -1.3 & 3.7 & -3.7 & 4.3 \\ 4.1 & 6.2 & -7.7 & -2.1 & -202.2 \end{pmatrix}$$

Describe the power method for obtaining the largest eigenvalue of a matrix and explain the theory of how it works. What would happen if the largest sized eigenvalues were complex?

Consider the application of the Power Method to the matrix A above. 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, show that the difference between the power method vector and the eigenvector for the largest eigenvalue decreases by a factor of approxmately 2 for each iteration.

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

[20 marks]

5. Describe the Jacobi method and the Gauss-Seidel 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-Seidel methods will converge for the matrix A and column vector $\mathbf{b}$, where

$$A = \begin{pmatrix} 215.3 & -11.3 & 9.1 \\ 5.3 & 101.7 & 4.7 \\ 15.1 & -13.5 & -314.9 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} 111.0 \\ 97.0 \\ -77.0 \end{pmatrix}.$$

Find the matrix T and column vector $\mathbf{c}$ such that the Jacobi iteration method can be written in the form:

$$\mathbf{x}^{n+1} = T.\mathbf{x}^{n} + \mathbf{c}.$$

Find a norm for T, and hence estimate the number of iterations needed to reduce the difference between $\mathbf{x}^n$ and the solution $\mathbf{x}$ by a factor of 10^5.

[20 marks]

6. 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.

b) To use the Finite Difference method to find an approximate solution to the second order equation

$$\frac{d^2y}{dx^2} = p(x)\frac{dy}{dx} + q(x)y + r(x),$$

with boundary conditions $y(x) = \alpha$ and $y(b) = \beta$, where $a < b$, we divide the interval (a, b) up into $N + 1$ equal strips of width $h = (b - a)/(N + 1)$. We denote $x_i = a + ih$, where $0 \le i \le N + 1$ and $y_i = y(x_i)$. Find approximate expressions for d^2y/dx^2 and dy/dx in terms of y_{i+1}, y_i, y_{i-1} and h.

Write down the form of the matrices A and $\mathbf{b}$ when the finite difference method approximation to the boundary value problem

$$\frac{d^2y}{dx^2} = y\sin x + \frac{1}{1+x^2} \qquad y(-1) = 0, \quad y(1) = 1$$

is expressed as the linear equation

$$A.\mathbf{y} = \mathbf{b}.$$

[20 marks]

7.

a) Determine the constants α, β, γ, such that the quadrature rule for the integral

$$\int_{-1}^{1} f(x)\,dx = \alpha f(-1) + \beta f(0) + \gamma f(1)$$

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

What is the name usually given to this rule? How would you adapt this quadrature rule to find an approximate value for $\int_a^b f(x)\,dx$ using two strips? How would you adapt this formula for a larger number of strips?

The error in using this rule to approximate $\int_a^b f(x)\,dx$ for strips of width h is $(b-a)h^4 f^{(4)}(\xi)/180$, where $a < \xi < b$. How many strips would you use to evaluate $\ln 2$ with an error of less than $2 * 10^{-5}$ using the formula

$$\ln 2 = \int_1^2 \frac{dx}{x}.$$

b) The two point Lagrange Interpolation formula for a function $f(x)$ is:

$$f(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1)$$

Use this formula to find the values of α and β for the approximation to the integral by the quadrature rule:

$$\int_0^h f(x)dx = \alpha f(h/4) + \beta f(2h/3)$$

[20 marks]