

## Numerical Analysis — FMN011 — 110603 Solutions

The exam lasts 4 hours. A minimum of 35 points out of the total 70 are required to get a passing grade. These points will be added to those obtained in your two home assignments, and the final grade is based on your total score.

Justify all your answers and write down all important steps. Unsupported answers will be disregarded.

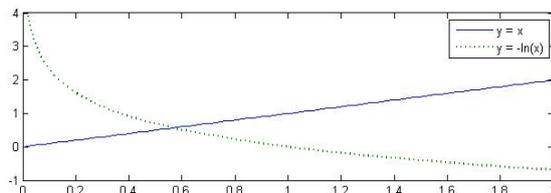
During the exam you are allowed a pocket calculator, but no textbook, lecture notes or any other electronic or written material.

1. (4p) Suppose matrix  $A$  has condition number (relative to the  $\infty$  norm) equal to  $10^{11}$ . The solution to the system  $Ax = b$  is  $x = [1, 1, \dots, 1]^T$ , but when solved in Matlab I get a relative error of magnitude  $10^{-6}$  in the  $\infty$ -norm. Is it reasonable to attribute this error to ill-conditioning? Justify your answer.

**Solution:**  $\|x - \tilde{x}\|_\infty = 1$ , so the absolute error is also  $10^{-6}$ . About 11 digits of accuracy may be lost because of the ill-conditioning, and Matlab has 16 digits of accuracy, so the result may have only 5 correct digits. This is consistent with an error of  $10^{-6}$ . The error may come from the ill-conditioning.

2. (4p) Consider the function  $f(x) = x + \ln x$ .
  - (a) Plot the functions  $y = x$  and  $y = -\ln x$  to show  $f$  has a unique root  $P$  in  $(0, \infty)$ .

**Solution:**



The function has a solution in  $(0, 1)$ .

- (b) Can the root  $P$  be found using a fixed point iteration  $x = g(x)$  with  $g(x) = -\ln x$ ? Justify.
- Solution:**  $|g'(x)| = 1/x$ , so  $|g'(x)| > 1$  in  $(0, 1)$ . Therefore the the solution cannot be found with this fixed point iteration.
3. (5p) If  $\|\cdot\|$  is a matrix norm, then  $\|A\| = 0 \Rightarrow A = 0$ . Find a  $2 \times 2$  nonzero matrix for which the spectral radius is zero, i.e.,  $\rho(A) = 0$ . Is the spectral radius a matrix norm? Justify. (The spectral radius is  $\rho(A) = \{\max_\lambda |\lambda|, Ax = \lambda x, x \neq 0\}$ .)

**Solution:**

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

$A$  is a nonzero matrix. If the spectral radius were a matrix norm, the spectral radius of  $A$  would be different from 0. Therefore, the spectral radius is not a matrix norm.

4. (5p) Consider the matrix

$$A = \begin{bmatrix} 0 & -8 & 8 & 3 \\ 0 & 4 & -4 & 3 \\ 3 & 4 & 1 & 4 \\ -4 & -3 & 8 & 3 \end{bmatrix}$$

What will the initial pivot in Gaussian elimination be if

- (a) No pivoting is used?

**Solution:** Cannot be done because the pivot would be 0.

- (b) Partial pivoting is used?

**Solution:** -4 (largest magnitude in column 1)

- (c) Scaled partial pivoting is used?

**Solution:** 3 (largest after dividing by largest element in the row,  $3/4 > 4/8$ )

5. (5p) Prove that Newton's method  $x_{n+1} = x_n - \partial F(x_n)^{-1}F(x_n)$  applied to the linear system  $Ax = b$  converges to the exact solution in one single step. ( $\partial F(x)$  is the Jacobian matrix, i.e., the matrix whose rows are the gradients of the rows of  $F(x)$ ).

**Solution:**  $F(x) = Ax - b$ , so  $\partial F(x_n) = A$ . Therefore,

$$x_{n+1} = x_n - A^{-1}(Ax - b) = A^{-1}b \text{ (exact solution)}$$

6. (5p) Select the most appropriate answer.

- (a) If a function is interpolated at  $n$  points  $\{x_1, x_2, \dots, x_n\}$ , the error of the interpolation at a point  $x \neq x_j$  contained in the interval of interpolation depends on

- i. the basis chosen for the interpolation
- ii. the value of the function at  $x$
- iii. the number of data points

**Solution:**(iii)

- (b) All previous computations can still be used when new data points are added in the following type of polynomial representations:

- i. Lagrange's
- ii. Bernstein's
- iii. Newton's

**Solution:**(iii)

- (c) With quadratic splines, which of the following are continuous at the knots?

- i. The function.
- ii. The function and its derivative.
- iii. The function and its first and second derivatives.

**Solution:** (ii)

- (d) What type of matrix is involved in the construction of a cubic spline?
- i. a symmetric matrix
  - ii. a diagonal matrix
  - iii. a diagonally dominant matrix

**Solution:** (iii)

- (e) The Jacobi or Gauss-Seidel methods applied to  $Ax = b$  will converge if
- i.  $A$  is strictly diagonally dominant
  - ii. the largest eigenvalue of  $A$  has absolute value equal to 0.5
  - iii. the initial guess is close enough to the exact solution

**Solution:** (i)

7. Consider the following polynomial bases:

- Monomial basis
- Lagrange basis
- Newton basis

- (a) (2p) Describe the pattern of nonzero entries in the basis matrix used to determine the coefficients for polynomial interpolation using each of the three bases.

**Solution:** Full; diagonal; triangular.

- (b) (2p) Rank the three bases according to the cost of determining the coefficients of the interpolating polynomial (from largest to smallest).

**Solution:** Monomial; Newton; Lagrange.

- (c) (2p) Which of the three methods has the best-conditioned basis matrix?

**Solution:** Lagrange

8. (6p)  $A = QD$  with

$$Q = \begin{pmatrix} -0.39074 & -0.55664 & 0.29336 & -0.67187 \\ -0.44012 & -0.45065 & 0.24592 & 0.7367 \\ -0.51035 & -0.04114 & -0.85787 & -0.043687 \\ -0.62703 & 0.69669 & 0.34281 & -0.062859 \end{pmatrix}, D = \begin{pmatrix} 1 & 0 \\ 0 & -1 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

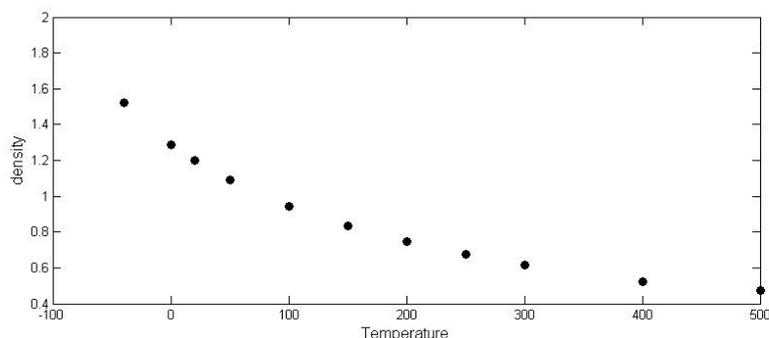
where  $Q$  is an orthogonal matrix. Solve the least squares problem  $Ax = b$  with  $b = [1\ 0\ 0]^T$  using the  $QR$  factorization of  $A$ . Show every step of the solution process.

**Solution:**  $A = QD$

$Ax = b \Rightarrow QDx = b \Rightarrow Dx = Q^T b$ . We need to solve taking only the first 2 rows, so  $\tilde{d} = [-0.39074, -0.44012]^T$ .

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} -0.39074 \\ -0.44012 \end{pmatrix} \Rightarrow x = \begin{pmatrix} -0.39074 \\ 0.44012 \end{pmatrix}$$

9. (5p) Air density as a function of temperature is given in tabular form and the data is plotted as follows:



Suppose you desire the density at temperatures -80 and 80 degrees Centigrade, and these temperatures are not included in the table. Discuss what methods you can use, and whether they would be adequate or not.

**Solution:** *Interpolation with a polynomial of degree 10 is not appropriate because of the numerical instability of calculating the coefficients and the Runge phenomenon. One can use splines or do a least squares with a quadratic or cubic polynomial.*

10. (4p) True or false:

- (a) If the  $n \times n$  matrix  $H$  is a Householder reflector, and  $x$  is an arbitrary  $n$ -vector, then the last  $k$  components of the vector  $Hx$  are zero for some  $k < n$ .

**Solution:** *False,  $Hx$  can be a reflection of  $x$  that does not have any zero components.*

- (b) The FFT algorithm can compute both the discrete Fourier transform and its inverse with the same efficiency.

**Solution:** *True. The inverse is obtained by a change of signs.*

11. (5p) A set of data was interpolated using the Fourier transform and the result was

$$\begin{aligned}
 P(t) = & 1.61 - 0.13 \cos 2\pi t - 0.50 \sin 2\pi t - 0.19 \cos 4\pi t - \\
 & 0.21 \sin 4\pi t - 0.20 \cos 6\pi t - 0.09 \sin 6\pi t - \\
 & 0.10 \cos 8\pi t
 \end{aligned}$$

Carry out the least squares fit of order 4.

**Solution:**

$$P_4(t) = 1.61 - 0.13 \cos 2\pi t - 0.50 \sin 2\pi t - 0.19 \cos 4\pi t$$

12. (a) **(2p)** The Shannon information formula is

$$I = - \sum_{i=1}^k p_i \log_2 p_i$$

Calculate the average least number of bits needed to code the matrix

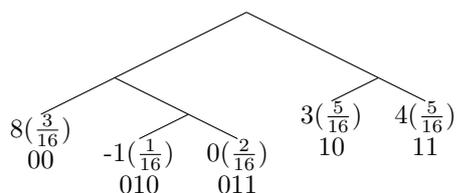
$$M = \begin{bmatrix} 0 & 8 & 8 & 3 \\ 0 & 4 & 4 & 3 \\ 3 & 4 & -1 & 4 \\ 4 & 3 & 8 & 3 \end{bmatrix}$$

**Solution:**

$$I = \frac{1}{16} \log_2 16 + \frac{2}{16} \log_2 \frac{16}{2} + \frac{10}{16} \log_2 \frac{16}{5} + \frac{3}{16} \log_2 \frac{16}{3} \approx 2.13$$

- (b) **(2p)** Construct a Huffman tree for  $M$ .

**Solution:**



- (c) **(2p)** What is the average for this coding? What is the average if the standard binary system is used for the matrix entries?

**Solution:**

$$3\left(\frac{1}{16}\right) + 3\left(\frac{2}{16}\right) + 2\left(\frac{5}{16}\right) + 2\left(\frac{5}{16}\right) + 2\left(\frac{3}{16}\right) \approx 2.19$$

For the standard binary system: 4

13. **(5p)**  $A$  is a real matrix and when the *Matlab* command

`[X,L]=eig(A)`

is executed, the following result is obtained:

$$\begin{aligned} X &= \begin{matrix} 0.3004 & -0.73463 & -0.73463 & -0.57735 \\ 0.042914 & 0.3025 + 0.25928i & 0.3025 - 0.25928i & -0.57735 \\ 0.55788 & 0.10803 - 0.06482i & 0.10803 + 0.06482i & -0.57735 \\ 0.77245 & -0.51856 - 0.12964i & -0.51856 + 0.12964i & -1.6726e-015 \end{matrix} \\ L &= \begin{matrix} 6 & 0 & 0 & 0 \\ 0 & -2.3592e-016 + 3i & 0 & 0 \\ 0 & 0 & -2.3592e-016 - 3i & 0 \\ 0 & 0 & 0 & -2.207e-015 \end{matrix} \end{aligned}$$

- (a) What algorithm was used by *Matlab*?

**Solution:** The QR algorithm

(b) What does matrix  $X$  contain?

**Solution:** *The eigenvectors of  $A$*

(c) What does matrix  $L$  contain?

**Solution:** *The eigenvalues of  $A$*

(d) Is it possible that matrix  $A$  is a symmetric matrix? Justify your answer.

**Solution:** *No, because a real symmetric matrix has real eigenvalues.*

(e) Is it possible to conclude that matrix  $A$  is invertible? Justify.

**Solution:** *No, because the eigenvalue of magnitude  $10^{-15}$  might represent a 0.*

14. (5p) The eigenvalues of

$$S = \begin{bmatrix} 5 & 6 & 8 & 11 \\ 6 & 7 & 9 & 12 \\ 8 & 9 & 10 & 13 \\ 11 & 12 & 13 & 14 \end{bmatrix}$$

are -3.319, -0.5106, -0.02962 and 39.86. Find the singular values of  $S$ .

**Solution:**  *$S$  is symmetric, so  $\lambda_i = |s_i|$ , therefore the singular values are 3.319, 0.5106, 0.02962 and 39.86.*