Numerical Analysis — FMN011 — 130603 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)** The fixed point iteration converges for both $g_1(x) = \cos(x)$ (to x=0.7391) and $g_2(x) = \cos^2(x)$ (to x=0.6417), but while the iteration converges to the fixed point of g_1 with 4 correct decimal digits in 23 iterations, it needs 206 iterations to achieve the same accuracy for g_2. Explain why fixed point iteration converges in both cases, and why it converges faster for g_1.
   Solution: $|g_1'(0.7391)| = 0.6736$ *and* $|g_2'(0.6417)| = 0.9590$. *They both converge because the magnitudes of their derivatives at the fixed points are less than 1, but the closer to 1, the slower the convergence.*

2. **(6p)** Consider Newton-Raphson's method, $x_{i+1} = x_i - \dfrac{f(x_i)}{f'(x_i)}$

   (a) What is the convergence rate for Newton-Raphson's method for finding the root $x = 2$ of each of the following equations?

      i. $f(x) = (x-1)(x-2)^2$

      ii. $f(x) = (x-1)^2(x-2)$

      Solution: *i. linear, because 2 is a double root; ii. quadratic, because 2 is a simple root.*

   (b) How should the formula be modified for multiple roots and what will be its rate of convergence?

      Solution: $x_{i+1} = x_i - m\dfrac{f(x_i)}{f'(x_i)}$ *will have quadratic convergence for a root of multiplicity m.*

   (c) Write out explicitly the formula for the Newton-Raphson iteration you would use for solving $(x-1)(x-2)^2 = 0$

      Solution: $x_{i+1} = x_i - 2\,\dfrac{(x_i-1)\,(x_i-2)}{3x_i-4}$. *Also accepted:* $x_{i+1} = x_i - \dfrac{(x_i-1)\,(x_i-2)}{3x_i-4}$, *or any further simplification of any of the two formulas.*

3. **(4p)** Describe how the LU factorization with pivoting is used to solve $Ax = b$, and mention the computational complexity of each main step.
   Solution:

   (a) Find P, L, and U, such that PA=LU ($\mathcal{O}(2n^3/3)$)

(b) *Solve Ly=Pb ($\mathcal{O}(n^2)$)*

(c) *Solve Ux=y ($\mathcal{O}(n^2)$)*

4. **(6p)** Consider the matrix

$$A = \begin{pmatrix} 1 & 1 \\ 1.0001 & 1 \end{pmatrix}$$

(a) Find the infinity-norm condition number of A.
Solution: $\kappa_\infty(A) = \|A\|_\infty \|A^{-1}\|_\infty = (2.0001) \cdot (2.0001 \times 10^4) = 40004$

(b) Find the error and the residual for the approximate solution $\begin{pmatrix} -1 & 3 \end{pmatrix}^T$ when $b = \begin{pmatrix} 2 & 2.0001 \end{pmatrix}^T$
Solution: *Error is (2, -2), residual is (0, 0.0002)*

(c) Find a vector x satisfying $\|A\|_\infty = \|Ax\|_\infty / \|x\|_\infty$
Solution: $x = \begin{pmatrix} 1 & 1 \end{pmatrix}^T$

5. **(6p)** The system $Ax = b$ was solved using the Gauss-Seidel iterative method, whose iteration matrix has the form $-(L+D)^{-1}U$. Matrix A is a strictly diagonally dominant $10^5 \times 10^5$ matrix that has 4×10^5 non-zero entries (4 entries per row).

(a) Approximately how many operations would one iteration step of the method require?
Solution: *7 operations per row,* $\times 10^5$ *operations per step.*

(b) What is the order of the number of operations required by Gaussian elimination?
Solution: $2n^3/3 = 2 \cdot 10^{15}/3$

(c) If 100 iteration steps were needed by the Gauss-Seidel method to compute the solution with the required accuracy, how do the number of operations required by Gaussian elimination compare to the number required by the Gauss-Seidel method?
Solution: *Gauss elimination needs approximately* $(2 \cdot 10^{15}/3)/(100 \times 7 \times 10^5) \approx 10^7$ *times more operations.*

6. **(4p)** The interpolation error function is given by

$$e(x) = \frac{f^{(n)}(\theta)}{n!}(x - x_1)(x - x_2)\cdots(x - x_n)$$

What function is minimized if the x_i are chosen to be the Chebyshev points?
Solution: $f(x) = \max_x |(x - x_1)(x - x_2)\cdots(x - x_n)|$

7. **(5p)** A cubic spline has the form

$$S_i(x) = y_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \quad x \in [x_i, x_{i+1}]$$

and its coefficients can be obtained by solving a system of the form

$$\begin{pmatrix} & & & & & & \\ \delta_1 & 2(\delta_1+\delta_2) & \delta_2 & \ddots & & & \\ 0 & \delta_2 & 2(\delta_2+\delta_3) & \delta_3 & & & \\ & \ddots & \ddots & \ddots & \ddots & & \\ & & & \delta_{n-2} & 2(\delta_{n-2}+\delta_{n-1}) & \delta_{n-1} & \\ & & & & & & \end{pmatrix} \begin{pmatrix} c_1 \\ \\ \vdots \\ \\ c_n \end{pmatrix} = \begin{pmatrix} \\ 3(\gamma_2-\gamma_1) \\ \\ \vdots \\ 3(\gamma_{n-1}-\gamma_n) \\ \\ \end{pmatrix}$$

where the first and last rows of the matrix and the right-hand side vector are missing. Construct the first and last rows of the matrix and of the right-hand side vector so that the boundary conditions $c_1 = c_2$ and $c_{n-i} = c_n$ are satisfied.
Solution: *row 1:* $[1 \ -1 \ 0 \ \ldots \ 0 \mid 0]$ *and row n:* $[0 \ 0 \ 0 \ \ldots \ -1 \ 1 \mid 0]$

8. **(5p)** The QR factorization of A is

$$Q = \begin{pmatrix} -0.4743 & 0.2927 & -0.5741 & -0.2224 & 0.5377 & -0.1455 \\ -0.3162 & 0.0197 & -0.4939 & 0.0623 & -0.8056 & -0.0527 \\ -0.7906 & -0.2138 & 0.5404 & -0.0801 & -0.0390 & -0.0857 \\ -0.1581 & 0.1414 & -0.0225 & 0.9645 & 0.1548 & -0.0133 \\ 0 & -0.9209 & -0.3273 & 0.0972 & 0.1869 & -0.0190 \\ -0.1581 & 0.0099 & -0.0692 & -0.0216 & 0.0387 & 0.9839 \end{pmatrix}, R = \begin{pmatrix} -6.3246 & -6.7989 \\ 0 & -7.6010 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Write the system that must be solved to find the least squares solution of $Ax = b$ where

$$b = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}^T$$

and solve.
Solution:

$$\begin{pmatrix} -6.3246 & -6.7989 \\ 0 & -7.6010 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} =$$

$$\begin{pmatrix} -0.4743 & -0.3162 & -0.7906 & -0.1581 & 0 & -0.1581 \\ 0.2927 & 0.0197 & -0.2138 & 0.1414 & -0.9209 & 0.0099 \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$

$$x = \begin{pmatrix} -0.0138 \\ 0.0826 \end{pmatrix}$$

9. **(6p)** True or false:

(a) The QR factorization of matrix is not unique.
Solution: *True. We can have $A = QDDR$ where $DD = I$ and QR is orthogonal and DR upper triangular (set elements in D as ± 1)*

(b) If M is an orthogonal matrix, then $\|x\|_p = \|Mx\|_p$ for all positive integers p.
Solution: *False. Only for the 2-norm.*

(c) A Householder reflector is a matrix that is both symmetric and orthogonal.
Solution: *True.* $H = I - 2P$ *where* $P^2 = P = \frac{vv^T}{v^Tv}$ *gives* $H^T = (I-2P)^T = I - 2P^T = I - 2P = H$ *and* $H^TH = (I-2P)(I-2P) = I - 4P + 4P^2 = I$

10. **(4p)** Given a general square matrix A, what method would you use to compute the following?

(a) Only the smallest eigenvalue of A (in magnitude) and its corresponding eigenvector
Solution: *Inverse power iteration (if the method converges), otherwise QR algorithm*

(b) Only the largest eigenvalue of A (in magnitude) and its corresponding eigenvector
Solution: *Power iteration (if the method converges), otherwise QR algorithm*

(c) The eigenvalue of A closest to some specified scalar β
Solution: *Shifted inverse power iteration with* $s = \beta$ *(if the method converges), otherwise QR algorithm*

(d) All of the eigenvalues and eigenvectors of A
Solution: *QR algorithm*

11. **(5p)** The following computations were done in Matlab:

```
>> A=[10 -12 -6;5 -5 -4; -1 0 3];
>> [U,S,V]=svd(A)
U =
   -0.8966   -0.2617    0.3574
   -0.4324    0.3420   -0.8343
    0.0961   -0.9025   -0.4198
S =
   18.6448         0         0
         0    2.8850         0
         0         0    0.2231
V =
   -0.6020   -0.0016   -0.7985
    0.6930    0.4958   -0.5234
    0.3967   -0.8684   -0.2974
```

How would you compute the rank-1 approximation of A?
Hint: $A = \sum_{i=1}^{r} s_i u_i v_i^T$
Solution:

$$A_1 = 18.6448 \begin{pmatrix} -0.8966 \\ -0.4324 \\ 0.0961 \end{pmatrix} \begin{pmatrix} -0.6020 & 0.6930 & 0.3967 \end{pmatrix} = \begin{pmatrix} 10.0640 & -11.5850 & -6.6316 \\ 4.8533 & -5.5870 & -3.1982 \\ -1.0786 & 1.2417 & 0.7108 \end{pmatrix}$$

12. **(5p)** Explain or illustrate how you can plot the trigonometric polynomial

$$P(t) = \frac{a_0}{\sqrt{8}} + \frac{2}{\sqrt{8}} \sum_{k=1}^{3} \left(a_k \cos \frac{2\pi kt}{8} - b_k \sin \frac{2\pi kt}{8} \right) + \frac{a_4}{8} \cos \pi t,$$

that interpolates the points $(j, x_j), j = 0, \ldots, 7; \quad x_j \in \mathbb{R}$, using the discrete Fourier transform. ***Solution:***

(a) *Apply DFT to x to get* $\begin{pmatrix} a_0 \\ a_1 + ib_1 \\ a_2 + ib_2 \\ a_3 + ib_3 \\ a_4 \\ a_3 - ib_3 \\ a_2 - ib_2 \\ a_1 - ib_1 \end{pmatrix}$

(b) *Expand the vector, for example to 128 elements, by adding 120 zeros between a_4 and $a_5 = a_3 - ib_3$*

(c) *Apply the inverse DFT to get a vector X of 128 elements (or enough to plot)*

(d) *Plot X vs $T = [0, 1, \ldots, 127]$*

13. **(5p)** Describe how quantization together with the discrete cosine transform is used for image compression.
Solution: *A pixel matrix is centered around 0 by subtracting 128 from each element. The DCT is applied to it and then it is quantized, multiplying by a given quantization matrix and rounding to integers. To recover the image, the resulting matrix is dequantized by multiplying by the quantization matrix and the inverse DCT is applied. Finally, 128 is added to each element.*

14. **(5p)** The Shannon information is $I = -\sum_{i=1}^{k} p_i \log_2 p_i$. Draw a Huffman tree and convert the message

COY KNOWS PSEUDONOISE CODES

to a bit stream, using Huffman coding. What is the average number of bits needed per symbol?

Solution: *One possible tree gives:*

O	*11*
S	*011*
-	*100*
E	*101*
N	*0001*
C	*0011*
D	*0101*
K	*00001*
W	*00100*
P	*00101*
Y	*00000*
U	*01000*
I	*01001*

$$\textit{Average number of bits per symbol} = 94/27 \approx 3.48$$