Numerical Analysis — FMN011 — 2017/05/30 – Solution

The exam lasts 5 hours and has 15 questions. A minimum of 35 points out of the total 70 are required to get a passing grade. These points will be added to those you 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 following procedure finds the simple root of $f(x) = 0$ in the interval $[a, b]$ using the bisection method so that the residual is less than the given positive quantity ϵ. Some parts are missing; fill in the blanks.

   1. Set $a^{(1)} = a$, $b^{(1)} = b$, and $i = 0$.
   2. $i = i + 1$.
   3. Calculate $x_m =$__________.
   4. If __________ $\leq \epsilon$, take the desired roots as $\hat{x} = x_m$ and stop. Otherwise, continue to next step.
   5. If __________, set $a^{(i+1)} = x_m$ and $b^{(i+1)} = b^{(i)}$, and go to step 2.
   6. If __________, set $b^{(i+1)} = x_m$ and $a^{(i+1)} = a^{(i)}$, and go to step 2.

   Solution:

   1. Set $a^{(1)} = a$, $b^{(1)} = b$, and $i = 0$.
   2. $i = i + 1$.
   3. Calculate $x_m = (a^{(i)} + b(i))/2$.
   4. If $|f(x_m)| \leq \epsilon$, take the desired roots as $\hat{x} = x_m$ and stop. Otherwise, continue to next step.
   5. If $f(a^{(i)}) \cdot f(x_m) > 0$, set $a^{(i+1)} = x_m$ and $b^{(i+1)} = b^{(i)}$, and go to step 2.
   6. If $f(a^{(i)}) \cdot f(x_m) < 0$, set $b^{(i+1)} = x_m$ and $a^{(i+1)} = a^{(i)}$, and go to step 2.

2. **(4p)** To find the root of $x^3 + 2x - 2 = 0$ we use a fixed-point iteration, $x_{n+1} = g(x_n)$, with $g(x) = 1 - 0.5x^3$.

   (a) If you know that the root is close to 0.77, can you be sure that the method will converge for some appropriate initial value, x_0?

   (b) Use the plot of $g(x)$ on the last page of this exam (Figure **??**) to illustrate that the iteration converges for $x_0 = -1$.

   (c) Explain why the iteration will not converge for $x_0 = -2$.

   (d) What can you say about the rate of convergence of this fixed-point iteration?

Solution:

(a) As $|g'(0.77)| \approx 0.89 < 1$, if we start close to root the method will converge.

(b) See Figure 3.

(c) $|g(-2)| = 6 > 1$.

(d) It is linear.

3. **(4p)** As Newton-Raphson's method,

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

is an iterative method, it needs some convergence criteria to stop the iterative procedure. Give two such criteria.
Solution:

$$\begin{aligned} &|f(x_{n+1})| < \epsilon \\ &|x_{n+1} - x_n| < \epsilon \\ &\frac{|x_n - x_{n-1}|}{|x_{n+1} - x_n|} < \epsilon \end{aligned}$$

4. **(4p)** Let A be an $n \times n$ matrix. Show that the function defined as the sum of all the entries of A is not a norm.
Solution: The following matrix is not the zero matrix, but the given function is 0. We know that $||A|| = 0 \Rightarrow A = 0$.

$$\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$$

5. **(4p)** The determinant of a product of matrices is the product of their determinants:

$$\det(AB) = \det(A) \cdot \det(B).$$

Suppose you have the LU factorization, $PA = LU$. How can you determine $|\det(A)|$?
Solution: As P only changes the order of the rows of A, it only changes the sign of the determinant. Therefore, $|\det(A)| = |\det(L)| \cdot |\det(U)| = |\det(U)|$, as L is lower triangular with ones on the diagonal.

6. **(5p)** Consider the system of equations

$$\begin{aligned} x_1 + x_2 &= 1 \\ x_1 - x_2 &= 3 \\ 4x_1 + x_2 &= 2 \end{aligned}$$

(a) Write the normal equations for this system.

(b) Give the least squares solution to the system.

(c) What is the residual vector?

(d) Is there a different solution that would give a residual with Euclidean norm equal to 1? Justify.

(e) Give a basis for the space that is orthogonal to the residual vector.

Solution:

(a)

$$\begin{pmatrix} 1 & 1 & 4 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \\ 4 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 4 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 3 \\ 2 \end{pmatrix}$$

(b) $x_1 = 0.9474, x_2 = -1.2632$.

(c) $(-1.3158,\ -0.7895,\ 0.5263)^T$.

(d) No. The least squares residual is the smallest in Eucledian norm, and is greater than 1.

(e) The columns of the matrix, $(1, 1, 4)^T$ and $(1, -1, 1)^T$.

7. **(5p)** The zeros of the Chebyshev polynomial T_n are

$$x_k = \cos\left(\frac{2k-1}{2n}\pi\right), k = 1, 2, \ldots, n.$$

Use the zeros of T_3 to construct an interpolating polynomial for $f(x) = \ln(x+2)$ on the interval $[-1, 1]$.
Solution:

$$\begin{aligned} f(\cos\pi/6) &= \log(\sqrt{3}/2 + 2) = 1.0529 \\ f(\cos 3\pi/6) &= \log(2) = 0.6931 \\ f(\cos 5\pi/6) &= \log(-\sqrt{3}/2) + 2) = 0.1257 \end{aligned}$$

$p(x) = -0.1384x^2 + 0.5353x + 0.6931$.

8. **(5p)** Figure 1 shows a bridge in a forest. I want to describe the 1-D curve of its walking surface by a cubic spline. After sampling the curve at 9 knots, $(x_1, y_1), \ldots, (x_9, y_9)$, what are the conditions that must be satisfied by the spline? Do not forget the end conditions. You do not not need to give the linear system that must be solved or construct the curve.

Solution: $S(x) = S_i(x) \in \Pi_3$ for $x \in [x_i, x_{i+1}],\ i = 1, \ldots, 8,$

$$\begin{aligned} S(x_i) &= y_i,\ i = 1, \ldots, 9 \\ S_i'(x_{i+1}) &= S_{i+1}'(x_{i+1}),\ i = 1, \ldots, 8 \\ S_i''(x_{i+1}) &= S_{i+1}''(x_{i+1}),\ i = 1, \ldots, 8 \\ S_1'(x_1) &= 0 \\ S_8'(x_9) &= 0 \end{aligned}$$

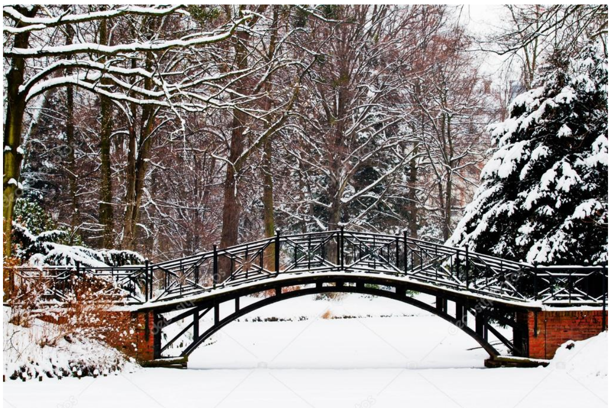


Figure 1: Bridge in a forest, for Problem 8.

9. **(5p)** Draw a sketch of the Bézier curve with control points (0,0), (0.5,0.5), (0.5,-0.5), (1,0). Mark the control points and the control polygon.
Solution: See Figure 4.

10. **(5p)** Which are the 12-th roots of unity? Which of them are primitive?
Solution: The roots of unity are $e^{-ik\pi/6}$, $k = 0, \dots, 11$. They are primitive for $k = 1, 5, 7, 11$. See Figure 5.

11. **(5p)** We wish to use the DFT interpolation theorem in $[-\pi, \pi]$ to interpolate $f(t) = |t|$ with 4 interpolation points.

    (a) Give the set of interpolation points $(t_j, f(t_j))$.

    (b) Knowing that

$$F_4 = \frac{1}{2}\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & -i & -1 & i \\ 1 & -1 & 1 & -1 \\ 1 & i & -1 & -i \end{pmatrix},$$

    calculate the Fourier transform of $x = [f(t_0), f(t_1), f(t_2), f(t_3)]^T$.

    (c) How many operations does it take to calculate the Fourier transform for a 4-dimensional real vector x? How many operations would it take if you used the fast Fourier transform?

    (d) Construct the interpolating polynomial given by the formula

$$P_n(t) = \frac{a_0}{\sqrt{n}} + \frac{2}{\sqrt{n}} \sum_{k=1}^{n/2-1} \left(a_k \cos \frac{2k\pi(t-c)}{d-c} - b_k \sin \frac{2k\pi(t-c)}{d-c} \right) + \frac{a_{n/2}}{\sqrt{n}} \cos \frac{n\pi(t-c)}{d-c}$$

    (e) Evaluate $P_n(-2\pi/3)$. What is the relative error (in %) of the interpolation at this point?

Solution:

(a) $(-\pi, \pi), (-\pi/2, \pi/2), (0, 0), (\pi/2, \pi/2)$.

(b)

$$F_4 x = \begin{pmatrix} \pi \\ \pi/2 \\ 0 \\ \pi/2 \end{pmatrix},$$

(c) 4*2(4+3)=56 operations to calculate the Fourier transform for a 4-dimensional real vector x. Aproximately $2N\log_2 N = 16$ operations with the fast Fourier transform.

(d)

$$P_4(t) = \frac{\pi}{2} + \frac{\pi}{2}\cos(t+\pi)$$

(e) $P_4(-2\pi/3) = 3\pi/4$. The relative error is

$$|\frac{2\pi/3 - 3\pi/4}{2\pi/3}| * 100\% = 12.5\%.$$

12. **(5p)** The DFT of a real vector is

```
-0.15
-0.25 + 3.59i
-0.35 - 0.92i
-0.45 - 0.22i
-0.55
-0.45 + 0.22i
-0.35 + 0.92i
-0.25 - 3.59i
```

Given that the DFT trigonometric interpolation polynomial is

$$P_n(t) = \frac{a_0}{\sqrt{n}} + \frac{2}{\sqrt{n}} \sum_{k=1}^{n/2-1} (a_k \cos(2\pi kt) - b_k \sin(2\pi kt)) + \frac{a_{n/2}}{\sqrt{n}} \cos(n\pi t),$$

show how to construct a low-pass filter that keeps frequencies up to $4\pi t$ by using least squares approximation.

Solution: We remove the tail of the formula. The coefficients we need are

$$\begin{aligned} a_0 &= -0.15 \\ a_1 &= -0.25,\ b_1 = 3.59 \\ a_2 &= -0.35 \end{aligned}$$

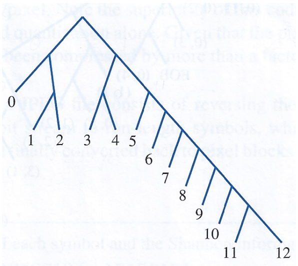
$$P_2(t) = \frac{-0.15}{\sqrt{8}} + \frac{2}{\sqrt{8}}(-0.25\cos(2\pi t) - 3.59\sin(2\pi t) - 0.35\cos(4\pi t))$$

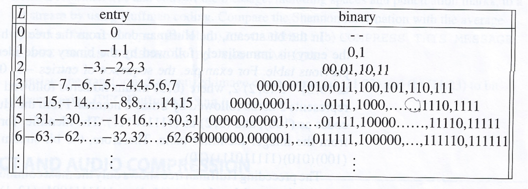
13. **(5p)** Find a two-term DCT least squares approximation of the function

$$f(t) = 3 + 2\cos t + \cos 2t.$$

Solution: As the function is already in the basis of DCT, the least squares approximation is

$$f(t) = 3 + 2\cos t.$$





L	entry	binary
0	0	- -
1	−1,1	0,1
2	−3,−2,2,3	*00,01,10,11*
3	−7,−6,−5,−4,4,5,6,7	000,001,010,011,100,101,110,111
4	−15,−14,...,−8,8,...,14,15	0000,0001,......0111,1000,......,1110,1111
5	−31,−30,...,−16,16,...,30,31	00000,00001,......,01111,10000......,11110,11111
6	−63,−62,...,−32,32,...,62,63	000000,000001,...,011111,100000,...,111110,111111
⋮	⋮	⋮

Figure 2: DPCM tree and integer identifying table for Problem 15.

14. **(5p)** Suppose you have a 40 × 40 intensity matrix corresponding to a black-and-white image. Enumerate the (five) steps you need to carry out to compress it using the JPEG standard.
Solution:

    (a) Subtract 128 for each entry.

    (b) Divide into 8 × 8 blocks.

    (c) Take 2-DCT.

    (d) Quantize (lossy compression) with JPEG standard matrices.

    (e) Use JPEG standard Huffman trees (one for the DC component and another for the AC components) to do a lossless compression.

    (f) Take the 2D-DCT inverse.

    (g) Recompose the matrix.

    (h) Add 128.

15. **(5p)** The DC component of an 8 × 8 transformed and quantized image matrix is given in JPEG code as 1 0 0 0 0 0. Decode this entry using Figure 2.
Solution: The first part of the code, 100, is read from the tree: $L = 3$. The second part, 000, is read from the table, and is -7.

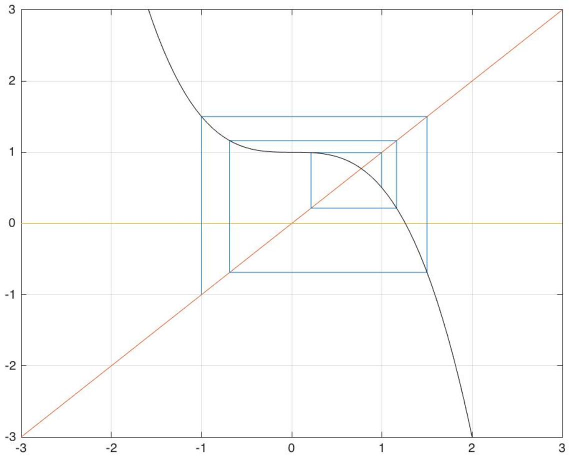


Figure 3: Plot of g vs x, for Problem 2.

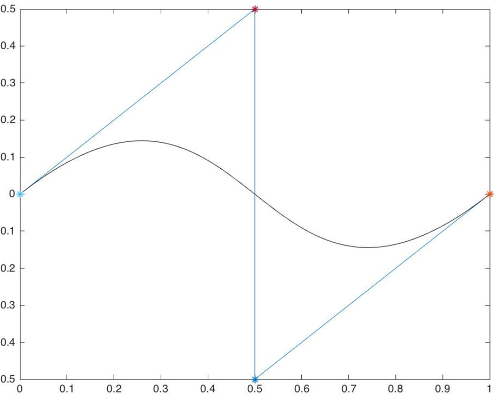


Figure 4: Bézier curve, for Problem 9.

C.Arévalo

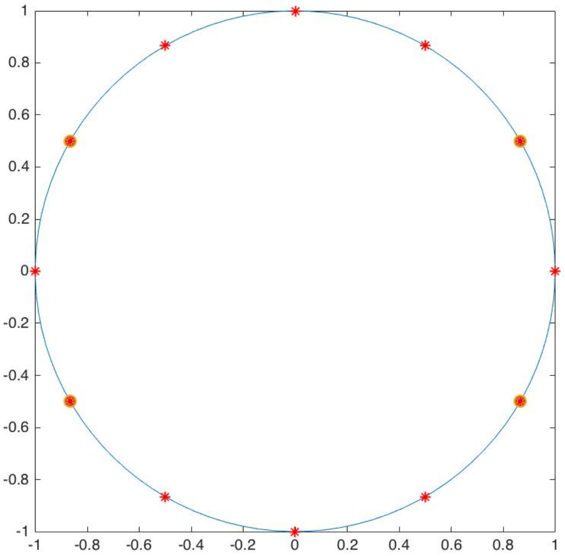


Figure 5: 12th roots of unity, for Problem 10.