UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in:	MAT3110/MAT4110 — Introduction to numerical analysis
Day of examination:	19 January 2021
Examination hours:	09:00 – 13:00

This problem set consists of 7 pages.

Appendices:	None
Permitted aids:	All written aids

Please make sure that your copy of the problem set is complete before you attempt to answer anything.

Note:

- There are in total 11 subproblems (1, 2a, 2b, ...), and you can get 5–10 points for each sub-problem, for a total of 100 points.
- All answers must be justified.

Problem 1 Root finding

Let $f(x) = \cos(x) - x$. This function has a single root x_0 somewhere in $[0, 1]$, and we wish to compute it.

1a (10 points)

Perform two steps with both the bisection method and Newton's method. Justify your choice of starting values.

1b (10 points)

Which of the two methods can we expect to be the most accurate after several iterations? Justify your answer.

Solution:

1a

For the bisection method we choose $x_0 = 0, x_1 = 1$. Then $f(x_0) = 1 > 0$ and $f(x_1) = \cos(1) - 1 < 0$. Since f is continuous, it has a zero in (x_0, x_1), and the bisection method will be able to find it. We compute

$$x_2 = \frac{x_0 + x_1}{2} = \frac{1}{2}$$

(Continued on page 2.)

and note that $f(x_2) = \cos(1/2) - 1/2 > 0$, and therefore the new interval will be (x_2, x_1). We finally get

$$x_3 = \frac{x_1 + x_2}{2} = \frac{3}{4}.$$

For Newton's method, we note that $f'(x) = -\sin(x) - 1$, which is nonzero in $[0, 1]$. Hence, as long as the iteration stays in $[0, 1]$, the method will converge. We set e.g. $x_0 = 0$ and get

$$\begin{aligned} x_1 &= x_0 - \frac{f(x_0)}{f'(x_0)} = -\frac{1}{-1} = 1, \\ x_2 &= x_1 - \frac{f(x_1)}{f'(x_1)} = 1 - \frac{\cos(1) - 1}{-\sin(1) - 1} \approx 0.7504. \end{aligned}$$

1b

The bisection method converges linearly, while Newton's method converges quadratically, so we can expect Newton's method to be the most accurate.

Problem 2 Polynomial interpolation (10 points)

Let $f : [0, 2] \to \mathbb{R}$ be a given function and let $n \in \mathbb{N}$. We wish to interpolate f using an n-th order polynomial p.

- Explain how we should do this in order to minimize the maximal error $\|f - p\|_{C([0,2])} = \sup_{x\in[0,2]} |f(x) - p(x)|$.
- Give an estimate of $\|f - p\|_{C([0,2])}$.

Solution: Assume first that we are on the interval $[-1, 1]$, and let $x_0, \dots, x_n \in [-1, 1]$ be distinct interpolation points. We let p interpolate f over these points:

$$p(x) = \sum_{k=0}^{n} f(x_k) \prod_{\substack{l=0,\dots,n \\ l\neq k}} \frac{x - x_l}{x_k - x_l}.$$

The basic error estimate is

$$\|f - p\|_{C([-1,1])} \leqslant \frac{\|f^{(n+1)}\|_{C([-1,1])}}{(n + 1)!} \|w_n\|_{C([-1,1])}$$

where $w_n(x) = \prod_{k=0}^{n}(x - x_k)$. The term $\|w_n\|_{C([-1,1])}$ is the least possible when $x_0, \dots, x_n$ are chosen as the Chebysheff points, yielding $\|w_n\|_{C([-1,1])} = 2^{-n}$, and therefore

$$\|f - p\|_{C([-1,1])} \leqslant \frac{\|f^{(n+1)}\|_{C([-1,1])}}{2^n(n + 1)!}.$$

(Continued on page 3.)

Any other choice of interpolation points will yield a larger right-hand side.

To transform this analysis to the interval $[0,2]$, it is enough to note that the two intervals are of the same length, and that translating a function does not change its norm, so the same results apply:

$$\|f-p\|_{C([0,2])} \leqslant \frac{\|f^{(n+1)}\|_{C([0,2])}}{2^n(n+1)!}.$$

Problem 3 Polynomial interpolation

Let $f:[0,1]\to\mathbb{R}$ be the function $f(x)=\cos(2x)-e^x$. For some $n\in\mathbb{N}$, let p be the n-th order polynomial which interpolates f over the uniform grid $0, 1/n, \dots, 1$.

3a (10 points)

Prove that $\|f-p\|_{C([0,1])}\to 0$ as $n\to\infty$.

(Here, $\|f-p\|_{C([0,1])} = \sup_{x\in[0,1]}|f(x)-p(x)|$.)

3b (10 points)

How large must n be in order to guarantee that $\|f-p\|_{C([0,1])}\leqslant 10^{-10}$?

Hint: In this problem you might (or might not) need Stirling's approximation:

$$m! \geqslant m^m e^{-m}.$$

Solution: The basic error estimate is: For every $x\in[0,1]$ there is some $\xi\in[0,1]$ such that

$$|f(x)-p(x)| \leqslant \frac{|f^{(n+1)}(\xi)|}{(n+1)!}\prod_{k=0}^{n}|x-x_k|.$$

We have

$$|f^{(m)}(\xi)| \leqslant \left|\tfrac{d^m}{d\xi^m}\cos(2\xi)\right| + \left|\tfrac{d^m}{d\xi^m}e^{\xi}\right| \leqslant 2^m + e.$$

Moreover, $|x-x_k|\leqslant 1$, so we get

$$|f(x)-p(x)| \leqslant \frac{2^{n+1}+1}{(n+1)!} \qquad \forall\, x\in[0,1].$$

Using Stirling's formula we get

$$\|f-p\|_{C([0,1])} \leqslant \frac{2^{n+1}+1}{(n+1)!} \leqslant \frac{e^{n+1}(2^{n+1}+1)}{(n+1)^{n+1}}.$$

3a

It is clear that the expression above converges to zero as $n\to\infty$.

(Continued on page 4.)

3b

Testing different values of n shows that $n = 29$ gives an upper bound of $\approx 4.3 \times 10^{-11}$.

Alternative solution: If, say, $x \in [x_m, x_{m+1}]$ then

$$\prod_{k=0}^{n} |x - x_k| = \frac{1}{n^{n+1}} \prod_{k=0}^{n} |xn - k| \leqslant \frac{(m+1)!(n-m)!}{n^{n+1}} \leqslant \frac{(n+1)!}{n^{n+1}},$$

where the last inequality follows from $1 \leqslant \binom{n+1}{m+1} = \frac{(n+1)!}{(m+1)!(n-m)!}$. We get

$$|f(x) - p(x)| \leqslant \frac{\|f^{(n+1)}\|}{(n+1)!} \frac{(n+1)!}{n^{n+1}} \leqslant \frac{2^{n+1}+1}{n^{n+1}}.$$

3a

It is clear that $\frac{2^{n+1}+1}{n^{n+1}} \to 0$ as $n \to \infty$.

3b

With this improved estimate we find that $n = 12$ gives an upper bound $\sim 7.66 \times 10^{-11}$.

Problem 4 QR factorization

Let A, Q and R be the matrices

$$A = \begin{pmatrix} 0 & 1 \\ \sqrt{2} & 3\sqrt{2} \\ 0 & 1 \end{pmatrix}, \quad R = \sqrt{2}\begin{pmatrix} 1 & 3 \\ 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad Q = \frac{1}{\sqrt{2}}\begin{pmatrix} 0 & 1 & 1 \\ \sqrt{2} & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}.$$

Note that $A = QR$ (you don't have to show this).

4a (5 points)

Explain what it means that QR is the QR factorization of A. Justify your answer.

4b (10 points)

Find the least squares solution of the equation

$$Ax = b, \qquad \text{where } b = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Solution:

(Continued on page 5.)

4a

A QR factorization consists of an orthogonal matrix Q and an upper triangular matrix R (with 1's as its first nonzero element in each row, if the factorization is in normal form). It is straightforward to see that $Q^\top Q = I$, so Q is orthogonal, and that R is upper triangular.

4b

We wish to minimize $\|Ax - b\| = \|Rx - Q^\top b\|$. Write

$$R = \begin{pmatrix} R_1 \\ 0 \end{pmatrix}, \quad Q^\top b = \begin{pmatrix} c_1 \\ c_2 \end{pmatrix}, \quad R_1 = \sqrt{2}\begin{pmatrix} 1 & 3 \\ 0 & 1 \end{pmatrix}, \quad c_1 \in \mathbb{R}^2, c_2 \in \mathbb{R}.$$

Then $\|Ax - b\|^2 = \|Rx - Q^\top b\|^2 = \|R_1 x - c_1\|^2 + \|c_2\|^2$, so we need to minimize the first term; to this end, we solve $R_1 x = c_1$. We compute

$$c_1 = \begin{pmatrix} 2 \\ 2\sqrt{2} \end{pmatrix} \qquad \Rightarrow \qquad x = R_1^{-1} c_1 = \begin{pmatrix} \sqrt{2} - 6 \\ 2 \end{pmatrix}.$$

Problem 5 SVD (10 points)

Compute the singular value decomposition (SVD) of

$$A = \begin{pmatrix} 1 & 2 \\ 3 & 6 \end{pmatrix}.$$

Hint: You may use the fact that one of the eigenpairs of the normal matrix $A^\top A$ *is* $\lambda_1 = 50$, $\mathbf{v}_1 = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

Solution: We see that A is non-invertible, so $A^\top A$ must also be non-invertible, whence the second eigenvalue is $\lambda_2 = 0$. The second eigenvector is chosen such that $V = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{pmatrix}$ is orthogonal; this is achieved by letting $\mathbf{v}_2 = \frac{1}{\sqrt{5}}\begin{pmatrix} 2 \\ -1 \end{pmatrix}$. We get the two singular values

$$\sigma_1 = \sqrt{50} = 5\sqrt{2}, \qquad \sigma_2 = 0.$$

Setting $S = \operatorname{diag}(\sigma_1, \sigma_2)$ we want to find an orthogonal matrix U such that $A = USV^\top$, or $US = AV$. Writing $U = \begin{pmatrix} \mathbf{u}_1 & \mathbf{u}_2 \end{pmatrix}$, we have

$$US = \begin{pmatrix} \sigma_1 \mathbf{u}_1 & 0 \end{pmatrix} \qquad \Rightarrow \qquad \mathbf{u}_1 = \frac{1}{\sigma_1} A\mathbf{v}_1 = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

Finally, we let $\mathbf{u}_2$ be such that U is orthogonal: $\mathbf{u}_2 = \frac{1}{\sqrt{10}}\begin{pmatrix} 3 \\ -1 \end{pmatrix}$. Thus, $A = USV^\top$ with

$$U = \frac{1}{\sqrt{10}}\begin{pmatrix} 1 & 3 \\ 3 & -1 \end{pmatrix}, \qquad S = \begin{pmatrix} 5\sqrt{2} & \\ & 0 \end{pmatrix}, \qquad V = \frac{1}{\sqrt{5}}\begin{pmatrix} 1 & 2 \\ 2 & -1 \end{pmatrix}.$$

(Continued on page 6.)

Problem 6

We wish to approximate the integral $I(f) = \int_0^{20} f(x)\,dx$ of a function f.

6a (5 points)

If we wish to approximate $I(f)$ using an 5-point quadrature rule, which quadrature rule should we choose to make the error as small as possible? Justify your answer.

6b (10 points)

Recall that the Gauss quadrature of order 3 on the interval $[-1, 1]$ is

$$\int_{-1}^{1} g(x)\,dx \approx f\big(-\sqrt{1/3}\big) + f\big(\sqrt{1/3}\big). \tag{1}$$

Write down the composite integration rule over $N = 2$ subintervals which approximates $I(f)$. Use the quadrature rule (1) in the composite method.

Solution:

6a

The n-point Gauss quadrature rule has order $2n-1$, which is the largest possible. We should therefore use the 3-point Gauss quadrature rule.

6b

Translating the interval $[-1, 1]$ to $[0, 10]$ gives quad. points and weights

$$x_0 = 5 - 5/\sqrt{3}, \quad x_1 = 5 + 5/\sqrt{3}, \quad w_0 = w_1 = 5$$

and on the interval $[10, 20]$

$$x_2 = 15 - 5/\sqrt{3}, \quad x_3 = 15 + 5/\sqrt{3}, \quad w_2 = w_3 = 5.$$

Thus, the composite quadrature rule is

$$I(f) \approx 5\Big(f(5 - 5/\sqrt{3}) + f(5 + 5/\sqrt{3}) + f(15 + 5/\sqrt{3}) + f(15 + 5/\sqrt{3})\Big).$$

Problem 7 Runge–Kutta method (10 points)

Consider the ODE

$$\begin{cases} x'(t) = f(x(t), t) \\ x(0) = x_0 \end{cases}$$

where f is a given smooth function, and the Runge–Kutta method

$$\begin{aligned} k &= f(y_n + hk/2, t_n + h/2) \\ y_{n+1} &= y_n + hk. \end{aligned}$$

(Continued on page 7.)

Set $f(x,t) = \lambda x$ for some $\lambda \in \mathbb{C}$ with $\mathrm{Re}(\lambda) < 0$. Find the stability function of this method, and determine whether the method is unconditionally stable or not.

Hint: If you are unable to determine stability, it's enough to insert $h\lambda = 1, 10, 100$ *in the stability function and conclude based on that.*

Solution: We insert $f(x,t) = \lambda x$ and get

$$k = \lambda(y_n + hk/2) \qquad \Leftrightarrow \qquad k = \frac{h\lambda}{1 - h\lambda/2} y_n$$

$$y_{n+1} = y_n + hk = y_n\Big(1 + \tfrac{h\lambda}{1-h\lambda/2}\Big) = y_n R(h\lambda)$$

where $R(z) = 1 + \frac{z}{1-z/2} = \frac{1+z/2}{1-z/2} = \frac{2+z}{2-z}$. If $\mathrm{Re}(z) < 0$ then $|2+z| < |2-z|$, and therefore $|R(z)| < 1$ for all such z. We conclude that the method is unconditionally stable.