UNIVERSITY OF OSLO

Faculty of mathematics and natural sciences

Exam in:	MAT3110/MAT4110 — Introduction to numerical analysis
Day of examination:	26 November 2020
Examination hours:	15:00 – 19:00

This problem set consists of 8 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 14 subproblems (1a, 1b, 2a, etc.), and you can get between 5 and 10 points for each sub-problem.
- All answers must be justified.

Problem 1 QR factorization

Let $\delta > 0$ and let

$$A = \begin{pmatrix} 3 & 3 & 0 \\ 0 & \delta & 1 \\ 4 & 4 & 0 \end{pmatrix}.$$

1a

Compute the QR factorization of A using the Gram–Schmidt algorithm.

1b

If δ is very small, what can go wrong if we run this algorithm on a computer? Be as specific as you can.

Solution:

1a

We run through the algorithm:

$n = 1$**:** $w \leftarrow \begin{pmatrix} 3 \\ 0 \\ 4 \end{pmatrix}$, $r_1 \leftarrow \begin{pmatrix} 5 \\ 0 \\ 0 \end{pmatrix}$, $q_1 \leftarrow \begin{pmatrix} 3/5 \\ 0 \\ 4/5 \end{pmatrix}$, $k \leftarrow 1$.

(Continued on page 2.)

$n = 2$: $w \leftarrow \begin{pmatrix} 0 \\ \delta \\ 0 \end{pmatrix}$, $r_2 \leftarrow \begin{pmatrix} 5 \\ \delta \\ 0 \end{pmatrix}$, $q_2 \leftarrow \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$, $k \leftarrow 2$.

$n = 3$: $w \leftarrow \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$, $r_3 \leftarrow \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$.

Since $k < n$ at the end of the algorithm, we add a unit vector q_3 which is orthogonal to q_1, q_2, for instance $q_3 = \begin{pmatrix} -4/5 \\ 0 \\ 3/5 \end{pmatrix}$. (Only $\pm q_3$ are correct answers.) We get

$$A = QR, \qquad Q = \begin{pmatrix} 3/5 & 0 & -4/5 \\ 0 & 1 & 0 \\ 4/5 & 0 & 3/5 \end{pmatrix}, \qquad R = \begin{pmatrix} 5 & 5 & 0 \\ 0 & \delta & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

1b

If $\delta \approx 0$ then we might run into roundoff errors when we compute w in the second step:

$$w = \begin{pmatrix} 3 \\ \delta \\ 4 \end{pmatrix} - \left(\tfrac{9}{5} + \tfrac{16}{5}\right) \begin{pmatrix} 3/5 \\ 0 \\ 4/5 \end{pmatrix} = \begin{pmatrix} \widetilde{0} \\ \delta \\ \widehat{0} \end{pmatrix}$$

where $\widetilde{0}, \widehat{0} \approx 0$. If the round off error is of the same order as δ, then the relative error is large, and $q_2 = w/\|w\|$ will be incorrect; in particular, q_2 will not be orthogonal to q_1. As a consequence, the computation of w in the third step will be incorrect, leading to incorrect r_3 and q_3.

Problem 2

Let $A \in \mathbb{R}^{n \times n}$ be a given matrix and define

$$f(x) = \frac{\|Ax\|}{\|x\|} \qquad \text{for } x \in \mathbb{R}^n,\ x \neq 0$$

(where $\|\cdot\|$ is the Euclidean norm, $\|x\| = \sqrt{\sum_{i=1}^n (x_i)^2}$.)

2a

In what way does $f(x)$ tell us how sensitive A is to x?

2b

Assume that $n = 2$ and that A can be decomposed as

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 8 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3/5 & 4/5 \\ -4/5 & 3/5 \end{pmatrix}.$$

What type of decomposition is this?

(Continued on page 3.)

2c

Let A be as in problem **2b**. Find nonzero vectors $y, z \in \mathbb{R}^2$ such that

$$f(y) = \max_{\substack{x \in \mathbb{R}^2 \\ x \neq 0}} f(x), \qquad f(z) = \min_{\substack{x \in \mathbb{R}^2 \\ x \neq 0}} f(x).$$

Solution:

2a

Let $e \in \mathbb{R}^n$. Then $f(x)$ is the relative error when we evaluate $A(e+x)$ instead of Ae:

$$\frac{\|A(e+x) - Ae\|}{\|(e+x) - e\|} = \frac{\|Ax\|}{\|x\|} = f(x).$$

Put another way: It tells us how sensitive the map $e \mapsto Ae$ is to changes in the input.

2b

This is an SVD decomposition, $A = USV^\mathsf{T}$, with

$$U = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \quad S = \begin{pmatrix} 8 & 0 \\ 0 & 1 \end{pmatrix}, \quad V = \begin{pmatrix} 3/5 & -4/5 \\ 4/5 & 3/5 \end{pmatrix}.$$

Indeed, S is clearly a diagonal matrix with nonnegative entries, and both U and V are orthogonal matrices: $UU^\mathsf{T} = VV^\mathsf{T} = I$.

2c

Write $U = (u_1 \; u_2)$, $V = (v_1 \; v_2)$, $S = \operatorname{diag}(\sigma_1, \sigma_2)$. Then $Ax = u_1\sigma_1\langle v_1, x\rangle + u_2\sigma_2\langle v_2, x\rangle$. Since u_1, u_2 are orthonormal, the norm of Ax is

$$\|Ax\| = \sqrt{\sigma_1^2\langle v_1, x\rangle^2 + \sigma_2^2\langle v_2, x\rangle^2}.$$

If, say, $\|y\| = 1$ and $f(y)$ is as large as possible, then necessarily we need $\langle v_1, y\rangle = 1$ and $\langle v_2, y\rangle = 0$ — that is, $y = v_1$. Likewise, we need $z = v_2$. Thus:

$$y = \begin{pmatrix} 3/5 \\ 4/5 \end{pmatrix}, \quad z = \begin{pmatrix} -4/5 \\ 3/5 \end{pmatrix}.$$

Problem 3 Newton's method

Consider the system of equations

$$\begin{aligned} x^3 - y^3 &= 3 \\ x^2 + y^2 &= 4. \end{aligned} \tag{1}$$

3a

Write down Newton's method for solving (1), and perform one iteration when starting at $(x_0, y_0) = (1, 1)$.

(Continued on page 4.)

3b

What can go wrong if we start at, or close to, the x- or y-axes?

Solution:

3a

Let

$$f(x,y) = \begin{pmatrix} x^3 - y^3 - 3 \\ x^2 + y^2 - 4 \end{pmatrix}.$$

Newton's method is

$$\begin{pmatrix} x_{n+1} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} x_n \\ y_n \end{pmatrix} - \nabla f(x_n, y_n)^{-1} f(x_n, y_n).$$

We compute

$$\nabla f(x,y) = \begin{pmatrix} 3x^2 & -3y^2 \\ 2x & 2y \end{pmatrix} \quad \Rightarrow \quad \nabla f(x,y)^{-1} = \frac{1}{6x^2y + 6xy^2} \begin{pmatrix} 2y & 3y^2 \\ -2x & 3x^2 \end{pmatrix}.$$

This gives

$$\begin{aligned} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{12} \begin{pmatrix} 2 & 3 \\ -2 & 3 \end{pmatrix} \begin{pmatrix} -3 \\ -2 \end{pmatrix} \\ &= \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \frac{1}{12} \begin{pmatrix} -12 \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 2 \\ 1 \end{pmatrix}. \end{aligned}$$

3b

The Jacobian $\nabla f(x,y)$ is not invertible when $x = 0$ or $y = 0$ (the first or second column becomes zero, respectively). As a consequence the Newton method breaks down.

This happens because near, say, the x-axis (i.e., when $y \approx 0$), the function f changes little or nothing at all when moving in the y-direction. As a consequence, Newton's method dictates moving very far (or infinitely far) in the y-direction.

Problem 4 Polynomial interpolation

Let $f : \mathbb{R} \to \mathbb{R}$ be a function satisfying

$$f(0) = 2, \quad f(1) = 1, \quad f(3) = 5.$$

4a

Find the polynomial p of the lowest possible order which interpolates f through these points.

(Continued on page 5.)

4b

Assume that $f \in C^3([0,4])$ satisfies

$$\|f^{(3)}\|_\infty \leqslant M, \qquad \text{where } \big\|f^{(3)}\big\|_\infty = \sup_{x\in[0,4]} \big|f^{(3)}(x)\big|$$

for some $M > 0$. Estimate the error $\|f - p\|_\infty$ in terms of M.

*Hint: You don't need the solution from problem **4a** in order to solve problem **4b**.*

Solution:

4a

We have 3 interpolation points, hence there exists a unique interpolating polynomial $p \in \mathbb{P}_n$ when $n = 2$. We use the Lagrange form for p:

$$\begin{aligned} p(x) &= \sum_{i=0}^{2} f(x_i)L_i(x) = 2\frac{x-1}{0-1}\frac{x-3}{0-3} + 1\frac{x-0}{1-0}\frac{x-3}{1-3} + 5\frac{x-0}{3-0}\frac{x-1}{3-1} \\ &= \frac{2}{3}(x-1)(x-3) - \frac{1}{2}x(x-3) + \frac{5}{6}x(x-1) \\ &= x^2 - 2x + 2. \end{aligned}$$

4b

The standard error estimate is

$$|f(x) - p(x)| = \frac{|f^{(n+1)}(\xi)|}{(n+1)!}\prod_{i=0}^{n}|x - x_i| \leqslant \frac{M}{6}\big|x(x-1)(x-3)\big|$$

where $n = 2$, for some $\xi \in (0,4)$. The last term can be estimated in various ways, for instance by upper bounding each of the terms x, $x-1$ and $x-3$ by 4, 3 and 3, resp. (since $x \in [0,4]$). This gives:

$$\|f - p\|_\infty \leqslant 6M.$$

(The function $x \mapsto x(x-1)(x-3)$ attains its largest absolute value at $x = 4$, with the value 12; thus, the sharpest upper bound is $2M$.)

Problem 5 Numerical quadrature

Consider the quadrature rule $I(f) \approx J(f)$, where

$$I(f) = \int_0^4 f(x)\,dx \quad \text{and} \quad J(f) = f(0)w_0 + f(x_1)w_1 + f(4)w_2$$

where $w_0, w_1, w_2 \in \mathbb{R}$ and $x_1 \in (0,4)$.

5a

Let $x_1 \in (0,4)$ be fixed. Let m denote the largest integer such that the quadrature rule is exact for all $f \in \mathbb{P}_m$. Show that w_0, w_1, w_2 can be chosen such that $m \geqslant 2$.

(Continued on page 6.)

Hint: You need to find expressions for w_0, w_1, w_2*, and to show that* $I(f) = J(f)$ *for all* $f \in \mathbb{P}_2$*.*

5b

Is there a choice of $x_1 \in (0, 4)$ which makes the order m of the quadrature rule greater than 2? Explain why, or why not.

Solution:

5a

Let $f \in \mathbb{P}_2$. Then $f = f(0)L_0 + f(x_1)L_1 + f(4)L_2$, where L_i is the i-th Lagrange function. Hence,

$$\int_0^4 f(x)\,dx = f(0)\underbrace{\int_0^4 L_0(x)\,dx}_{=w_0} + f(x_1)\underbrace{\int_0^4 L_1(x)\,dx}_{=w_1} + f(4)\underbrace{\int_0^4 L_2(x)\,dx}_{=w_2}$$
$$= w_0 f(0) + w_1 f(x_1) + w_2 f(4).$$

From the definition it is clear that the method is exact for all $f \in \mathbb{P}_2$.

5b

Yes, the midpoint $x_1 = 2$ would yield Simpson's rule,

$$J(f) = \frac{4}{6}f(0) + \frac{16}{6}f(2) + \frac{4}{6}f(4).$$

We only need to check that J integrates $f(x) := x^3$ exactly:

$$I(f) = 64,$$
$$J(f) = \frac{16}{6}2^3 + \frac{4}{6}4^3 = \frac{16 \cdot 8 + 4 \cdot 4^3}{6} = 64.$$

An alternative solution is to recall the formula $f(x) - p(x) = \frac{f^{(3)}(\xi)}{6}(x-0)(x-x_1)(x-4)$, for some $\xi \in (0, 4)$, where $p \in \mathbb{P}_2$ interpolates f through $0,\ x_1,\ 4$. If $f \in \mathbb{P}_3$ then $f^{(3)} \equiv c$ for some $c \in \mathbb{R}$, so

$$I(f) - J(f) = \int_0^4 f(x) - p(x)\,dx = \frac{c}{6}\int_0^4 (x-0)(x-x_1)(x-4)\,dx.$$

We see that the choice $x_1 = 2$ would make the integral vanish, and hence $I(f) = J(f)$.

Problem 6 Numerical methods for ODEs

Consider the ordinary differential equation (ODE)

$$\begin{cases} \dot{x}(t) = f(x(t), t) & \text{for } t > 0 \\ x(0) = x_0 \end{cases} \tag{2}$$

(Continued on page 7.)

We partition the time domain $t \in [0, T]$ into subintervals with endpoints $t_n = hn$, where $h = \frac{T}{N}$, for some $N \in \mathbb{N}$. The solution is approximated as $y_n \approx x(t_n)$.

6a

Consider the Runge–Kutta method with the Butcher tableau

$$\begin{array}{c|cc} \alpha & 0 & 0 \\ 2/3 & 1/3 & \beta \\ \hline & 1/4 & \gamma \end{array}$$

for numbers $\alpha, \beta, \gamma \in \mathbb{R}$. Determine α, β, γ such that the method is consistent. Is the method explicit or implicit?

6b

For the Runge–Kutta method from problem **6a**, write down a formula for y_{n+1} in terms of y_n.

6c

We consider the linear problem (2) with $f(x, t) = \lambda x$ for some $\lambda \in \mathbb{C}$ with $\operatorname{Re} \lambda < 0$. We solve the problem using the numerical method

$$\begin{cases} y_{n+1} = y_n + hf\big(y_n + \frac{h}{2} f(y_n, t_n),\ t_n + \frac{h}{2}\big) & \text{for } n = 0, 1, \ldots, \\ y_0 = x_0. \end{cases} \tag{3}$$

Find the stability function for (3). Is the method unconditionally stable?

Solution:

6a

We need $\alpha = 0 + 0 = 0$, $2/3 = 1/3 + \beta$, and $1/4 + \gamma = 1$, so

$$\alpha = 0,\ \beta = 1/3,\ \gamma = 3/4.$$

The method is implicit since the Runge–Kutta matrix has nonzero entries along the diagonal.

6b

We get

$$\begin{aligned} k_1 &= f(y_n, t_n) \\ k_2 &= f(y_n + h(k_1 + k_2)/3, t_n + 2/3) \\ y_{n+1} &= y_n + h(k_1 + 3k_2)/4. \end{aligned}$$

(Continued on page 8.)

6c

If we insert $f(x, \lambda) = \lambda x$ then we get

$$y_{n+1} = y_n + h\lambda\big(y_n + \tfrac{h}{2}\lambda y_n\big) = y_n\big(1 + h\lambda + \tfrac{1}{2}(h\lambda)^2\big).$$

Hence, $y_{n+1} = R(h\lambda)y_n$ with

$$R(z) = 1 + z + \tfrac{z^2}{2}.$$

If, say, $\lambda \in \mathbb{R}$ and $\lambda \ll 0$ then $|R(\lambda h)|$ is a large number (larger than 1) unless h is proportionally small. In other words, the method is *not* unconditionally stable.