Linnaeus University
Department of Mathematics
Sofia Eriksson

Written Exam on Numerical Methods, 2MA903, 1 hp (5 hp)
Saturday 24th of March 2021, 12.00–17.00.

The solutions should be complete, correct, motivated, well structured and easy to follow.
Aids: Calculator (you may use a scientific calculator but *not* with internet connection).
Please begin each question on a new paper.
Preliminary grades: 15p-17p⇒E; 18p-20p⇒D; 21p-23p⇒C; 24p-26p⇒B; 27p-30p⇒A.

1. (a) For which positive integers k can the the number $7+2^{-k}$ be represented exactly (with no rounding error) in double precision floating point arithmetic?

   (b) Reformulate the function $g(x) = \frac{1-\cos(x)}{\sin(x)^2}$ to avoid cancellation problems for x very close to zero. (5p)

   (a)

$$\begin{aligned} 7+2^{-k} &= 4+2+1+2^{-k} \\ &= 1.11 * 2^2 + 2^{-k} \end{aligned}$$

The largest k we can use without rounding is 50, since we can write

$$\begin{aligned} 7+2^{-50} &= 1.1100000000000000000000000000000000000000000000000000 * 2^2 \\ &+ 0.0000000000000000000000000000000000000000000000000001 * 2^2 \\ &= 1.11 * 2^2 + 2^{2-52} = 1.11 * 2^2 + 2^{-50} \end{aligned}$$

The smallest positive integer k is 1, and

$$\begin{aligned} 7+2^{-1} &= 4+2+1+2^{-1} \\ &= 1.11 * 2^2 + 0.001 * 2^2 = 1.111 * 2^2 \end{aligned}$$

can also be represented without rounding.

Answer: for $1 \leq k \leq 50$

(b) Use the conjugate and trigonometric identities:

$$g(x) = \frac{1-\cos(x)}{\sin(x)^2} = \frac{1-\cos^2(x)}{\sin(x)^2(1+\cos(x))} = \frac{1}{1+\cos(x)}$$

2. Use the Newton-Raphson method to find approximations of all solutions of the equation $x^2 = e^x + 1$ with 4 correct decimals. (5p)

The Newton-Rapson method is written $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$ where $f(x) = 0$.

Let $f(x) = e^x + 1 - x^2$. For Newton-Raphson we need $f'(x) = e^x - 2x$.

For example a graphical solution reveals that this equation only has one roots, close to $x = -1$.

$$
\begin{aligned}
x_0 &= -1 & & \\
x_1 &= -1.155362403496964 & \quad \Delta x &= -0.155362403496964 \\
x_2 &= -1.147776180729494 & \quad \Delta x &= 0.007586222767470 \\
x_3 &= -1.147757632255525 & \quad \Delta x &= 1.854847396898229 \cdot 10^{-5}
\end{aligned}
$$

the latest $\Delta x < 0.5 \cdot 10^{-4}$

Answer: $x = -1.1478$

3. (a) Why is it beneficial to use LU-factorization when solving systems of linear equations $Ax = b$ where b varies?

(b) Find the forward and backward errors for the approximate solution $\tilde{x} = (-1, 3.0001)^T$ of the system

$$\begin{pmatrix} 1 & 1 \\ 1.0001 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 2 \\ 2.0001 \end{pmatrix}.$$

(5p)

(a)

Because Gaussian elimination has complexity $\mathcal{O}(n^3)$ while forward- or backward substitution only requires $\mathcal{O}(n^2)$. When solving $Ax = b$ many times with different right-hand-sides b the costly LU-decomposition can be done once but the repeated solving can be done with the cheaper substitutions.

(b)

p. 86-87 Sauer

4. If A is a 6×6 matrix with eigenvalues -6, -3, 1, 2, 5 and 7, which eigenvalues of A will the following algorithms find?

(a) Power iteration?

(b) Inverse Power Iteration with shift $s = 4$?

(c) Find the linear convergence rates of the two computations in (a) and (b). Which converges faster? (5p)

(a) To the eigenvalue with biggest absolute value: Answer: 7

(b) To the eigenvalue closest to s, that is, the method will converge to 5

(c) The convergence rates are the ratios between first (dominating) and second eigenvalue. Short motivation:

$A^k v = A^k \sum c_i x_i = \sum c_i \lambda_i^k x_i = \lambda_1 \sum c_i \left(\frac{\lambda_i}{\lambda_1}\right)^k x_i$

In (a) $|-6|/|7| = 6/7$ and in (b)$|5-4|/|2-4| = 1/2$. The IPI in (b) converges faster.

5. (a) Consider the following approximation of $f'(x)$,

$$D(h) = \frac{1}{12h}\left(f(x-2h) - 8f(x-h) + 8f(x+h) - f(x+2h)\right).$$

Show that the truncation error $D(h) - f'(x)$ is of the form

$$D(h) - f'(x) = c_1h^4 + c_2h^6 + c_3h^8 + \ldots.$$

(b) Draw a figure showing the behaviour of the error $|D(h) - f'(x)|$ as a function of h. The axes should be in log-log scale.

(5p)

Taylor:

$$f(x+h) = f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f^{(3)}(x) + \frac{h^4}{24}f^{(4)}(x) + \frac{h^5}{120}f^{(5)}(x) + \frac{h^6}{720}f^{(3)}(x) + \ldots$$
$$f(x-h) = f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{6}f^{(3)}(x) + \frac{h^4}{24}f^{(4)}(x) - \frac{h^5}{120}f^{(5)}(x) + \frac{h^6}{720}f^{(3)}(x) + \ldots$$
$$f(x+h) - f(x-h) = 2hf'(x) + 2\frac{h^3}{6}f^{(3)}(x) + 2\frac{h^5}{120}f^{(5)}(x) + 2\frac{h^7}{5040}f^{(7)}(x) + \ldots$$
$$f(x+2h) - f(x-2h) = 2\cdot 2hf'(x) + 2\cdot 2^3\frac{h^3}{6}f^{(3)}(x) + 2\cdot 2^5\frac{h^5}{120}f^{(5)}(x) + 2\cdot 2^7\frac{h^7}{5040}f^{(7)}(x) + \ldots$$

where we note that every second term is cancelled. The last two rows give

$$\begin{aligned}
D(h) &= \frac{1}{12h}\left(f(x-2h) - 8f(x-h) + 8f(x+h) - f(x+2h)\right) \\
&= \frac{8}{12h}\left(2hf'(x) + 2\frac{h^3}{6}f^{(3)}(x) + 2\frac{h^5}{120}f^{(5)}(x) + 2\frac{h^7}{5040}f^{(7)}(x) + \ldots\right) \\
&- \frac{1}{12h}\left(2\cdot 2hf'(x) + 2\cdot 2^3\frac{h^3}{6}f^{(3)}(x) + 2\cdot 2^5\frac{h^5}{120}f^{(5)}(x) + 2\cdot 2^7\frac{h^7}{5040}f^{(7)}(x) + \ldots\right) \\
&= \frac{1}{12h}\left((16-4)hf'(x) + (16-16)\frac{h^3}{6}f^{(3)}(x) + (16-64)\frac{h^5}{120}f^{(5)}(x) + (16-256)\frac{h^7}{5040}f^{(7)}(x) + \ldots\right) \\
&= f'(x) - 4\frac{h^4}{120}f^{(5)}(x) - 20\frac{h^6}{5040}f^{(7)}(x) + \ldots
\end{aligned}$$

(b)

As h decreases, $D(h) - f'(x)$ will scale as roughly as $\sim \mathcal{O}(h^4)$. However, in the computer, when $h \approx \epsilon$, $D(h) = 0$ due to cancellation. See graphs from lecture notes.

6. Consider the boundary value problem

$$\frac{d^2y}{dx^2} = x^2, \qquad x \in [0,\ 10]$$
$$y(x=0) = -30$$
$$y(x=10) = 200$$

(a) Approximate the boundary value problem described above as a finite difference problem with step size $\Delta x = h = 2$, and present the resulting system of equations in matrix form.

(You don't have to solve the system of equations.)

(b) Reformulate the problem as an initial value problem (such that it can be solved using the Shooting method). *(You don't have to solve the reformulated problem.)* (5)

(a) Let $y_i \approx y(x_i)$ where $x_i = ih$ with $i = 0, 1, \ldots, n$. This gives

$$\frac{y_{i+1} - 2y_i + y_{i-1}}{h^2} = x_i^2, \qquad i = 1, 2, 3, \ldots, n-1$$
$$y_0 = -30$$
$$y_n = 200.$$

In matrix form, for $h = 2$ (that is $n = 5$), we have $x_0 = 0$, $x_1 = 2$, ..., $x_5 = 10$.

$$\begin{bmatrix} -2 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 1 & -2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} 30 \\ 0 \\ 0 \\ -200 \end{bmatrix} + h^2 \begin{bmatrix} 4 \\ 16 \\ 36 \\ 64 \end{bmatrix} = \begin{bmatrix} 46 \\ 64 \\ 144 \\ 56 \end{bmatrix}.$$

(b) Introduce a new variable $v = y'$. Now $y'' = x^2$ is rewritten as $v' = x^2$ and we obtain a system of differential equations

$$\begin{cases} y' = v \\ v' = x^2 \end{cases} \qquad \begin{cases} y(0) = -30 \\ y(10) = 200 \end{cases}$$

To use a solver for IVPs we need to replace the boundary condition $y(10) = 200$ by an initial condition $v(0) = s$, yielding

$$\begin{cases} y' = v \\ v' = x^2 \end{cases} \qquad \begin{cases} y(0) = -30 \\ v(0) = s \end{cases}$$

Good luck!

Linnaeus University
Department of Mathematics
Sofia Eriksson

List of formulas for the exam in Numerical Methods, 2021

These formulas will be attached to the exam. The list is not guaranteed to be complete, and the use, meaning, conditions and assumptions of the formulas are purposely left out.

- **Taylor's formula**

$$f(x) = f(a) + \tfrac{f'(a)}{1!}(x-a) + \tfrac{f''(a)}{2!}(x-a)^2 + \ldots + \tfrac{f^{(n)}(a)}{n!}(x-a)^n + \tfrac{f^{(n+1)}(\xi)}{(n+1)!}(x-a)^{n+1}$$

(ξ between x and a)

- **Absolute and relative error**

$$\Delta_x = \tilde{x} - x, \qquad \frac{\Delta_x}{x} \approx \frac{\Delta_x}{\tilde{x}}, \qquad \Delta_{x+y} = \Delta_x + \Delta_y, \qquad \frac{\Delta_{xy}}{xy} \approx \frac{\Delta_x}{x} + \frac{\Delta_y}{y}$$

- **Error propagation formulas, condition number (1D)**

$$\Delta f \approx f'(x)\Delta x, \qquad \left|\frac{\Delta f/f}{\Delta x/x}\right| \approx \left|\frac{xf'(x)}{f(x)}\right|$$

$$\Delta f \approx f''(x)\frac{\Delta x^2}{2}$$

- **Correct decimals**
  $|\Delta x| \leq 0.5 \cdot 10^{-t}$

- **Numbers in base B**
  $x = x_m B^m + x_{m-1}B^{m-1} + \ldots + x_0 B_0 + x_{-1}B^{-1} + \ldots = (x_m x_{m-1} \ldots x_0 . x_{-1} \ldots)_B$

- **Iterative methods**

Bisection method:

```
c=(a+b)/2;
while (b-a)>2*tol
    if f(c)*f(a)>0
        a=c;
    else
        b=c;
    end
    c=(a+b)/2;
end
```

Newton-Raphson: $\quad x_{n+1} = x_n - \dfrac{f(x_n)}{f'(x_n)}, \qquad J(\mathbf{x}_n)(\mathbf{x}_{n+1} - \mathbf{x}_n) + \mathbf{f}(\mathbf{x}_n) = 0$

The secant method: $\quad x_{n+1} = x_n - f(x_n)\dfrac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})},$

$$e_n = x_n - x^*, \qquad |x_{n+1} - x^*| < \bar{c}|x_n - x^*|^p, \qquad \lim_{n\to\infty} \tfrac{|e_{n+1}|}{|e_n|^p} = c$$

- **Equation systems**

  $A\mathbf{x} = \mathbf{b}$, residual $\mathbf{r} = \mathbf{b} - A\widetilde{\mathbf{x}}$

LU-factorization:	$A = LU, \quad PA = LU$
QR-factorization:	$A = QR, \quad Q^T Q = I$
(Iterative methods)	$A = D + L + U$
Jacobi methods:	$\mathbf{x}^{(k)} = -D^{-1}(L+U)\mathbf{x}^{(k-1)} + D^{-1}\mathbf{b}$
Gauss-Seidel:	$\mathbf{x}^{(k)} = -(D+L)^{-1}U\mathbf{x}^{(k-1)} + (D+L)^{-1}\mathbf{b}$

Backward: $\|\mathbf{r}\|_\infty$, forward: $\|\mathbf{x} - \widetilde{\mathbf{x}}\|_\infty$

- **Norms and condition numbers**

Let $\mathbf{x} = (x_1, \dots, x_n)$ be a vector:

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p\right)^{1/p}, \quad \|\boldsymbol{x}\|_1 = \sum_{i=1}^n |x_i|, \quad \|\boldsymbol{x}\|_2 = \sqrt{x_1^2 + \dots + x_n^2}, \quad \|\mathbf{x}\|_\infty = \max_i |x_i|.$$

Let A be a $n \times n$ matrix:

$$\|A\| = \sup_{\mathbf{x}\neq 0} \frac{\|A\mathbf{x}\|}{\|\mathbf{x}\|}, \qquad \|A\mathbf{x}\| \le \|A\| \cdot \|\mathbf{x}\|, \qquad \kappa(A) = \|A\| \cdot \|A^{-1}\|$$

$A(\mathbf{x} + \delta\mathbf{x}) = \mathbf{b} + \delta\mathbf{b}$

$$\frac{\|\delta\mathbf{x}\|}{\|\mathbf{x}\|} \le \kappa(A)\frac{\|\delta\mathbf{b}\|}{\|\mathbf{b}\|}, \qquad econd(A) = \frac{\|\delta\mathbf{x}\|/\|\mathbf{x}\|}{\|\delta\mathbf{b}\|/\|\mathbf{b}\|} \le \kappa(A),$$

- **Interpolation**

Let $(x_0, y_0), \dots, (x_n, y_n)$ be $n+1$ points in the xy-plane.

Monomial:	$P(x) = a_0 + a_1 x + \ldots + a_n x^n$
Lagrangre:	$P(x) = \sum_{j=0}^n y_j \ell_j(x), \qquad \ell_j(x) = \prod_{\substack{0\le m\le n \\ m\neq j}} \frac{x - x_m}{x_j - x_m}$

Newton's divided differences:

$$P(x) = [y_0] + [y_0, y_1](x - x_0) + \dots + [y_0, \dots, y_k](x - x_0)(x - x_1)\cdots(x - x_{n-1}).$$

$$[y_0] = f(x_0), \quad [y_0, y_1] = \frac{f(x_1) - f(x_0)}{x_1 - x_0}, \quad [y_0, y_1, y_2] = \frac{[y_1, y_2] - [y_0, y_1]}{x_2 - x_0}, \dots$$

Interpolation errors:

$$R(x) = f(x) - P(x) = (x - x_0)(x - x_1)\cdots(x - x_n)\frac{f^{n+1}(\xi)}{(n+1)!}, \qquad x_0 < \xi < x_n,$$

- **Least squares, normal equations** $A^T A\mathbf{x} = A^T\mathbf{b}$, residual $\mathbf{r} = \mathbf{b} - A\mathbf{x}$

- **Finite differences**

$$\frac{f(x+h) - f(x)}{h} = f'(x) + f''(\xi)\frac{h}{2} \qquad \xi \in [x, x+h]$$

$$\frac{f(x) - f(x-h)}{h} = f'(x) - f''(\xi)\frac{h}{2} \qquad \xi \in [x-h, x]$$

$$\frac{f(x+h) - f(x-h)}{2h} = f'(x) + f^{(3)}(\xi)\frac{h^2}{6} \qquad \xi \in [x-h, x+h]$$

$$\frac{f(x+h) - 2f(x) + f(x-h)}{h^2} = f''(x) + f^{(4)}(\xi)\frac{h^2}{12} \qquad \xi \in [x-h, x+h]$$

- **Trapezoidal rule, Simpson's rule**

$$\int_a^b f(x)dx = \tfrac{h}{2}\left(f(x_0) + 2\textstyle\sum_{k=1}^{n-1} f(x_k) + f(x_n)\right) - \tfrac{(b-a)h^2}{12} f''(\xi), \qquad h = \frac{b-a}{n}$$

$$\int_a^b f(x)dx = \tfrac{h}{3}\left(f(x_0) + 4\textstyle\sum_{k=1}^{n} f(x_{2k-1}) + 2\textstyle\sum_{k=1}^{n-1} f(x_{2k}) + f(x_{2n})\right) - \tfrac{(b-a)h^4}{180} f^{(4)}(\xi), \qquad h = \frac{b-a}{2n}$$

$a < \xi < b$

- **Richardson extrapolation**
  $Q = F(h) + kh^n + \mathcal{O}(h^{n+1}), \qquad Q = \frac{2^n F(h/2) - F(h)}{2^n - 1} + \mathcal{O}(h^{n+1})$

- **Romberg** $R_{i,1} = T(h/2^{i-1})$, $R_{ij} = \frac{4^{j-1} R_{i,j-1} - R_{i-1,j-1}}{4^{j-1} - 1}$

- **Numerical solutions of differential equations**

Differential equation $y' = f(x,y)$ with initial condition $y(x_0) = y_0$

Euler forward ($g_i \sim \mathcal{O}(h)$): $\quad y_{n+1} = y_n + hf(x_n, y_n)$

Euler backward ($g_i \sim \mathcal{O}(h)$): $\quad y_{n+1} = y_n + hf(x_{n+1}, y_{n+1})$

Heun's method ($g_i \sim \mathcal{O}(h^2)$): $\quad \begin{cases} y_{n+1} = y_n + \frac{1}{2}(k_1 + k_2) \\ k_1 = hf(x_n, y_n) \\ k_2 = hf(x_n + h, y_n + k_1) \end{cases}$

RK4 ($g_i \sim \mathcal{O}(h^4)$): $\quad \begin{cases} y_{n+1} = y_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4) \\ k_1 = hf(x_n, y_n) \\ k_2 = hf(x_n + h/2, y_n + k_1/2) \\ k_3 = hf(x_n + h/2, y_n + k_2/2) \\ k_4 = hf(x_n + h, y_n + k_3) \end{cases}$

where $x_{n+1} = x_n + h$.

- **Boundary value problems**

Two-point boundary problem $y'' = f(x, y, y')$ with initial condition $y(a) = \alpha$ and $y(b) = \beta$.

Shooting method: The BVP is rewritten as a IVP. Vary the modified initial condition until the boundary conditions are satisfied with desired accuracy.

Finite difference method: Derivatives are approximated by finite difference quotients. Exact solution y is replaced by y_i such that $y_i \approx y(x_i)$

- **Eigenvalue problems**

The power method: $\mathbf{v}_{k+1} = A\mathbf{v}_k / \|A\mathbf{v}_k\|$ and $\lambda_1 \approx \mathbf{v}_k^T A \mathbf{v}_k$.

The QR-method. Let $A = Q_0 R_0$ be a QR-decomposition of a real matrix A. Set $A_1 = R_0 Q_0$ and inductively (if $A_{n-1} = Q_{n-1} R_{n-1}$ is a QR-decomposition) $A_n = R_{n-1} Q_{n-1}$.