Linnaeus University
Department of Mathematics
Sofia Eriksson

Written Exam on Numerical Methods, 1MA930/1MA931, 3 hp (7.5 hp/5 hp)
Thursday 1st of June 2023, 08.00–13.00.

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) Which ones of the following numbers (i-iv) can be represented exactly in floating point arithmetics (IEEE double precision)? If a number can not be represented exactly, state to what value it is rounded by the computer:

   (i) $(1 + 2^{-58}) - 1$
   (ii) $(1 + 2^{-17}) - 1$
   (iii) 2^{-58}
   (iv) 2^{-17}

   (b) In (v-vi), identify for which values of x there is subtraction of nearly equal numbers, and find an alternate form that avoids the problem:

   (v) $\frac{\sqrt{1+x}-\sqrt{1-x}}{x}$
   (vi) $\frac{1-\cos(x)}{1+\cos(x)}$ (2p+3p)

2. (a) Use the Newton-Raphson method to find approximations of all solutions to the equation $x^3 + 18x^2 - 39x + 11 = 0$. Answer with 6 correct decimals.

   (b) If you instead would use the Bisection method to find one of the roots, and would start with an initial interval $[a, b]$ of width 1 (that is having $b = a + 1$): How many iterations would you have to do to obtain an answer with 6 correct decimals? (4p+1p)

3. Consider an n-by-n system of linear equations where the coefficient matrix is in upper triangular form, as visualised below for a 4-by-4 system.

$$\left(\begin{array}{cccc|c} \times & \times & \times & \times & \times \\ 0 & \times & \times & \times & \times \\ 0 & 0 & \times & \times & \times \\ 0 & 0 & 0 & \times & \times \end{array}\right)$$

   (a) Derive the number of operations needed to solve this n-by-n system (for an arbitrary positive integer n) using back-substitution. (*By operations we mean addition, subtraction, multiplication and division*)

   (b) Explain why and in what situations it is advantageous to use LU-factorisation. (3p+2p)

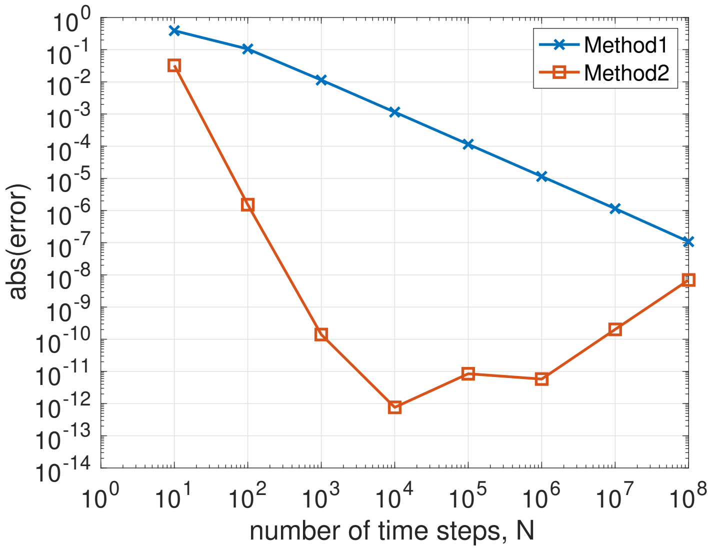
Please turn, the questions continue on next page!

4. (a) Use the (composite) Simpson's method to compute the integral $I = \int_0^1 e^x \, dx$ for two different step sizes h ($h = 1/2$ and $h = 1/4$). Then use Richardson extrapolation on the two results to further improve the approximation of the integral.

(b) Show that the finite difference formula $D_h = \frac{f(x+h)-f(x-h)}{2h}$ is second order accurate.

(3p+2p)

5. The differential equation $y' = -2y(1 + 4\cos(4t))$ with initial condition $y(0) = 1$ is solved using some of the numerical solvers mentioned in this course.



(a) The figure above shows the resulting errors at time $t = 1.3$ (obtained using two different explicit methods). Given that information, answer the following:

- What methods do "Method1" and "Method2" refer to? Motivate.
- Explain the errors of "Method2" in terms of why the curve looks as it does for *fewer* time steps ($N \lessapprox 10^4$) and *more* time steps ($N \gtrapprox 10^4$), respectively.
- Predict the behaviour of the curves as N increases even more, either by drawing a picture or by giving a crude guess for how large you think the errors produced by the two methods will be when the number of time steps are around $N \approx 10^{11}$.

(b) Find the numerical solution to the above-mentioned initial value problem, at time $t = 0.1$. Use the Euler Backward method with time step 0.05. (3p+2p)

6. Consider the boundary value problem (BVP)

$$\frac{d^2y}{dx^2} = \frac{x+y}{25}, \qquad x \in [0,\ 20]$$
$$y(x=0) = 1$$
$$y(x=20) = -7$$

(a) Approximate the boundary value problem described above as a finite difference problem with step size $\Delta x = h = 5$, and present the resulting system of equations in matrix form. *You don't need to solve the system!*

(b) Rewrite the BVP such that it could be solved using the shooting method. (3p+2p)

Good luck!

List of formulas for the exam in Numerical Methods, 2023

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)

- **Operation count**

$$1+2+3+4+\cdots+n = \frac{n(n+1)}{2}, \qquad 1^2+2^2+3^2+4^2+\cdots+n^2 = \frac{n(n+1)(2n+1)}{6}$$

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

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

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

• **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:	$A = QR, \quad Q^TQ = I$
(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 $\mathbf{x} = (x_1, \ldots, x_n)$ be a vector:

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

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 $(x_0, y_0), \ldots, (x_n, y_n)$ 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:

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

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

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** $A^TA\mathbf{x} = A^T\mathbf{b}$, 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**

$$\int_a^b f(x)dx = \tfrac{h}{2}\Big(f(x_0) + 2\textstyle\sum_{k=1}^{n-1} f(x_k) + f(x_n)\Big) - \tfrac{(b-a)h^2}{12} f''(\xi), \qquad h = \frac{b-a}{n}$$
$$\int_a^b f(x)dx = \tfrac{h}{3}\Big(f(x_0) + 4\textstyle\sum_{k=1}^{n} f(x_{2k-1}) + 2\textstyle\sum_{k=1}^{n-1} f(x_{2k}) + f(x_{2n})\Big) - \tfrac{(b-a)h^4}{180} f^{(4)}(\xi), \qquad h = \frac{b-a}{2n}$$

$a < \xi < b$

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

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

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 $A = Q_0R_0$ be a QR-decomposition of a real matrix A. Set $A_1 = R_0Q_0$ and inductively (if $A_{n-1} = Q_{n-1}R_{n-1}$ is a QR-decomposition) $A_n = R_{n-1}Q_{n-1}$.