Linnaeus University
Department of Mathematics
Sofia Eriksson

Written Exam on Numerical Methods, 1MA930, 3 hp (7.5 hp)
Thursday 2nd of June 2022, 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. Do the following sums by hand in IEEE double precision computer arithmetic, using the Rounding to Nearest Rule:

   (a) $(1 + 2^{-54}) - 1$,

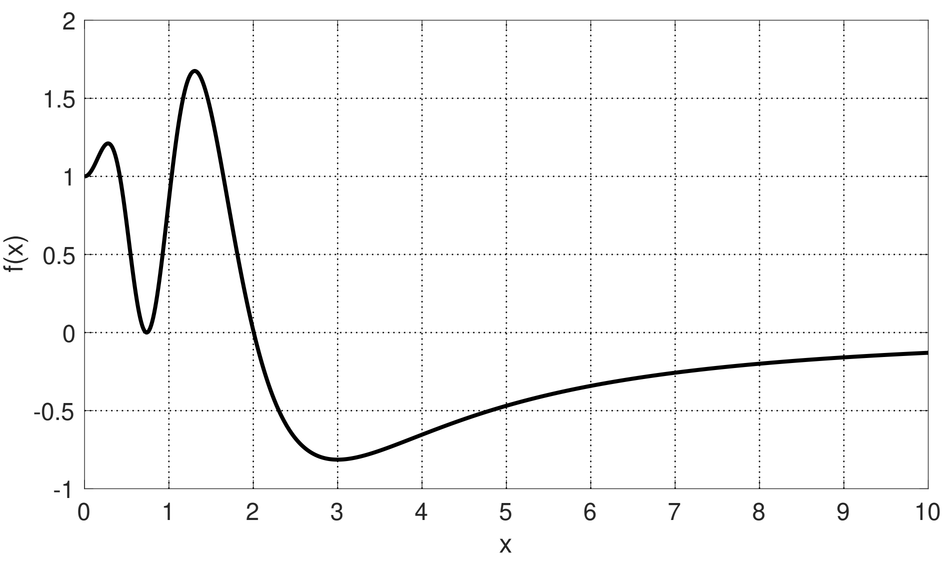
   (b) $2^{-60} + 2^{-75}$.

   (c) Identify for which values of x there is subtraction of nearly equal numbers, and find an alternate form that avoids the problem:
$$\frac{1}{1+x} - \frac{1}{1-x}.$$

   (d) A derivative $f'(x)$ is approximated by the central finite difference $D_{x,h} = \frac{f(x+h)-f(x-h)}{2h}$ in a computer.

   Sketch the resulting error $|D_{x,h} - f'(x)|$, as a function of h. It should be a log-log plot (logarithmic scaling on both x-axis and y-axis) for values of h ranging from 10^{-20} to 1. *You may assume that $x \sim \mathcal{O}(1)$, that $f(x)$ and its derivatives are well-defined in a neighbourhood around x and that $f''(x) \neq 0$.* (1p+1p+1p+2p)

2. Assume that you seek the roots to a non-linear equation $f(x) = 0$. To get an understanding for the problem, you first plot the function $f(x)$. The result is depicted in the figure below (the roots visible in the figure are the only ones that exist for $x \geq 0$):



Please turn, the questions continue on next page!

(a) Without doing any actual calculations, describe what will happen if you use the Newton-Raphson method with (i) $x_0 = 1$, (ii) $x_0 = 2$, (iii) $x_0 = 3$ or (iv) $x_0 = 4$ as initial guess. *Thus, for each initial guess (i-iv): Does the method converge? If it converges – to which root and with what convergence rate? If it does not converge – why not?*

(b) Find the real solution of $x^3+6x = 3x^2+11$. Answer with 4 correct decimals. (3p)+(2p)

3. Given the following set of points (x, y): $(1, 2)$, $(3, 6)$, $(4, 5)$:

(a) Determine the corresponding interpolating polynomial (of lowest possible degree),

(b) Evaluate the function value of the interpolating polynomial for $x = 2$.

(c) Assume that the points in (a) were produced by the function $f(x) = x^3-9x^2+25x-15$. Find an upper bound for the interpolation error at $x = 2$.

(2p+1p+2p)

4. The function values y are given for a few points according to

x	1	1.25	1.5	1.75	2
$y(x)$	0.1250	0.8350	1.0280	0.5740	−1.1630

a) Compute an approximation to the integral $\int_1^2 y(x)dx$ using Simpson's method. All available function values must be used.

b) Compute an approximation of $\int_1^2 y(x)dx$ using Simpson's method again, but now with the step length h doubled. Approximately how large is the error obtained in (b) compared to the error obtained in (a)? Motivate your answer.

c) Use Richardson extrapolation on the values obtained in (a) and (b) in order to find an improved approximation of the integral $\int_1^2 y(x)dx$. (2p+2p+1p)

5. Let $y(x)$ be the solution of $y'(x) = x - xy$ for which $y(0) = 2$.

(a) Find an approximation of $y(2)$ using Euler forward with step length $h = 1$ and another approximation using $h = 0.5$. Answer using 4 correctly rounded decimals.

(b) Using Richardson extrapolation, calculate an improved approximation of y(2) using the results obtained in (a).

(c) What are the advantages and drawbacks when comparing Euler forward with Euler backward? (3p+1p+1p)

6. Consider the boundary value problem

$$\frac{d^2y}{dx^2} = \frac{-15}{x+1}, \qquad x \in [0,\ 6]$$
$$y(x=0) = -8$$
$$y(x=6) = 3$$

(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.

(b) Solve the system and plot the solution in an appropriate coordinate system. (3p+2p)

Good luck!

List of formulas for the exam in Numerical Methods, 2022

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**

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

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

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 $\mathbf{x} = (x_1, \dots, x_n)$ 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:	$P(x) = a_0 + a_1x + \ldots + a_nx^n$
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** $A^TA\mathbf{x} = A^T\mathbf{b}$, 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**

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

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

$a < \xi < b$

- **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) $A_n = R_{n-1} Q_{n-1}$.