NUM: Solved problems—Numerical analysis

All problems can be done without a calculator. But you can use it if you want.
Problems marked with the symbol (+) are a bit more theoretical, these can appear at the oral exam or in the theoretical section of the exam.

Approximation.

a0a. Find an approximating formula for the function $\ln(x)$ on a neighborhood of $a = 1$ with error $O(h^4)$.

a0b. Find a linear approximation for the function $\frac{1}{x}$ on a neighborhood of $a = 2$.

a0c. Find an approximating formula for the function $\arctan(x)$ on a neighborhood of $a = 0$ with error $O(h^3)$.

a1a. Use linear approximation to estimate $\sqrt{4.5}$.

a1b. Use quadratic approximation to estimate $e^{0.4}$.

a1c. Use quadratic approximation to estimate $\cos(\pi + 0.5)$.

Integrals.

i0a. Deduce the rectangle formula for approximating a definite integral and explain with a picture.

i0b. Deduce the trapezoid formula for approximating a definite integral and explain with a picture.

i0c. Explain the notion of order of method for numerical integration. What is the order for the rectangle, trapezoid and Simpson methods?

i0d. We used a method of order 3 to approximate a certain integral with partition size $n = 100$. We have a reason to believe that the error is bounded by the number $e_n = 0.01$. Estimate the error of the approximation obtained using the partition number $n = 200$.
What partition size should we use if we want to get an error at most $\varepsilon = 0.0001$?

i0e. We used the trapezoid method to approximate a certain integral with partition size $n = 100$. We have a reason to believe that the error is bounded by the number $e_n = 0.016$. Estimate the error of the approximation obtained using the partition number $n = 200$.
What partition size should we use if we want to get an error at most $\varepsilon = 0.001$?

i1a. Use the rectangle method to approximate the integral $\int\limits_0^2 \sqrt{x}\,dx$ s potem dlen $n = 2$.

i1b. Use the rectangle method to approximate the integral $\int\limits_2^6 \frac{1}{2}x - 1\,dx$ s krokem $h = 2$.

i2a. Use the trapezoid method to approximate the integral $\int\limits_0^2 \sqrt{x}\,dx$ s krokem $h = 1$.

i2b. Use the trapezoid method to approximate the integral $\int\limits_2^6 \frac{1}{2}x - 1\,dx$ s potem dlen $n = 2$.

Roots of functions.

r0a. Explain the notion of order of iterative method for finding roots of functions. What is the order for the bisection and Newton methods?

r0b. Using an iterating method of order 2 we obtained estimates x_6, x_7 of a number r. We have a reason to believe that their errors are approximately $E_6 = 0.01$, $E_7 = 0.0003$.
Estimate the error E_8 of the next iteration.

r1a. Write the algorithm of the bisection method for finding roots. Explain it with a picture.

r1b. Apply the bisection method for root finding to the following problem: We want to solve the equation $x^2 = x + 1$ on an interval $[0, 4]$.
Show the first three steps of the iteration.

r1c. Apply the bisection method for root finding to the following problem: We want to solve the equation $\frac{1}{x} = x - 2$ on an interval $[1, 9]$.
Show the first three steps of the iteration.

r2a. Write the algorithm of the Newton method for finding roots. Explain it with a picture.

r2b. Apply the Newton method for root finding to the following problem: We want to solve the equation $x^2 = x + 1$, the initial guess is $x_0 = 0$.
Show the first two steps of the iteration.

r2c. Apply the Newton method for root finding to the following problem: We want to solve the equation $\frac{1}{x} = x - 2$, the initial guess is $x_0 = 1$.
Show the first two steps of the iteration.

r2d. Use the Newton method to deduce an iterating scheme that should find the number x satisfying $x^3 = A$ (that is, we are trying to find $\sqrt[3]{A}$).

r2e. Use the Newton method to deduce an iterating scheme that should find the number x satisfying $e^{-x} = x$.

r2f(+). Derive the general formula for the Newton method for root finding.

r3a. Write the iterating algorithm for finding fixed points of functions. Explain how to estimate whether it will converge. Explain how and why we use relaxation.

r3b. Apply the fixed point approach to the following problem: We want to solve the equation $x^2 = x + 1$, the initial guess is $x_0 = 3$.
Show the first two steps of the iteration.
Inquire whether the behaviour around x_0 looks hopeful.

r3c. Apply the fixed point approach to the following problem: We want to solve the equation $\frac{1}{x} = x - 2$, the initial guess is $x_0 = 3$.
Show the first two steps of the iteration.
Inquire whether the behaviour around x_0 looks hopeful.

r3d(+). For the iterations in 3b and 3c, write the general relaxed iterative formula and then find the optimal λ for the given x_0.

Differential equations.

d0a. Explain the notion of order of method for solving initial value problems (differential equations). What is the order for the Euler method? What is the order of one very popular quality method of Runge-Kutta type?

d0b. Consider the initial value problem $y' = f(x, y)$, $y(x_0) = y_0$. We approximated its solution on an interval $[x_0, x_0 + T]$ using a certain method of order 2 with step size h and we have a reason to believe that the global error is about 0.0027.
Estimate the error of approximation obtained with step size $\frac{1}{3}h$?
What step size should we use if we want to get an error at most 0.00001?

d0c. Consider the initial value problem $y' = f(x, y)$, $y(x_0) = y_0$. We approximated its solution on an interval $[x_0, x_0 + T]$ using a certain method of order 4 with step size h and we have a reason to believe that the global error is about 0.004.
Estimate the error of approximation obtained with step size $\frac{1}{2}h$?

d1a. Consider the initial value problem $y' = x + y$, $y(1) = 13$. Set up iterative formulas for approximating the solution on the interval $[1, 5]$ with step size $h = 1$ using the Euler method.
Calculate the first three points.

d1b. Consider the initial value problem $y' = 2xy$, $y(1) = 3$. Set up iterative formulas for approximating the solution on the interval $[1, 3]$ with partition size $n = 4$ using the Euler method.
Calculate the first three points.

Systems of linear equations.

s0a. Explain the notion of computational complexity of a method.
State the computational complexity of the Gaussian elimination and the backward (or forward) substitution. Discuss computational complexity of iterative methods for solving systems of equations.

s0b. We have a method for some matrix manipulation with computational complexity n^3. If this program ran for 5 hours when applied to a problem with $n = 1000$, how long will it probably run when applied to a problem with $n = 2000$?

s0c. Explain how systems of linear equations can be solved using the Gaussian elimination and the backward substitution. Discuss computational complexity of this approach.
Explain how systems of linear equations can be solved using iterative methods.

s1a. Consider the system $\left[\begin{array}{l} x + y - z = -2, \\ -x \quad + 2z = 3, \\ x + y + z = 2. \end{array}\right.$
Use it to explain how systenms of equations can be solved using the Gaussian elimination and the backward substitution.

s1b. Consider the system $\left[\begin{array}{l} x_1 + 2x_2 + x_3 = 0, \\ x_1 + 3x_2 \quad = -2, \\ -2x_1 - 4x_2 \quad = 2. \end{array}\right.$
Use it to explain how systenms of equations can be solved using the Gaussian elimination and the backward substitution.

s2a. Consider the system $\left[\begin{array}{l} x \quad - z = 1, \\ 2x + y - z = 1, \\ x + 2y - z = -1. \end{array}\right.$ Set up an iterating scheme for solving this system using the Gauss-Seidel method.
Show the first two steps with the initial vector $\vec{x}_0 = (0, 0, 0)$.
(+) Set up an iterating scheme for solving this system using the Jacobi method.
Show the first two steps with the initial vector $\vec{x}_0 = (0, 0, 0)$.

s2b. Consider the system $\left[\begin{array}{l} x \quad + z = 2 \\ x - y + 2z = 1. \\ x + 2y + z = 1 \end{array}\right.$ Set up an iterating scheme for solving this system using the Gauss-Seidel method.
Show the first two steps with the initial vector $\vec{x}_0 = (0, 0, 0)$.
(+) Set up an iterating scheme for solving this system using the Jacobi method.
Show the first two steps with the initial vector $\vec{x}_0 = (0, 0, 0)$.

Solutions

a0a. $\ln(1+h) = h - \frac{1}{2}h^2 + \frac{1}{3}h^3 + O(h^4)$ nebo $\ln(x) = (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 + O((x-1)^4)$.

a0b. We actuall need a tangent line. $\frac{1}{x} = \frac{1}{2} - \frac{1}{4}(x-2) + O((x-2)^2)$ or $\frac{1}{2+h} = \frac{1}{2} - \frac{1}{4}h + O(h^2)$.

a0c. $\arctan(x) = (x-0) + O(x^3) = x + O(x^3)$ or $\arctan(h) = \arctan(0+h) = h + O(h^3)$.

a1a. Choice: $a=4$. $\sqrt{4+h} \approx 2 + \frac{1}{4}h$, hence $\sqrt{4.5} \approx 2 + \frac{1}{4}\cdot 0.5 = 2.125$.

a1b. Choice: $a=0$. $e^h \approx 1 + h + \frac{1}{2}h^2$, hence $e^{0.4} \approx 1 + 0.4 + \frac{1}{2}\cdot 0.16 = 1.48$.

a1c. $\cos(\pi+h) \approx -1 + \frac{1}{2}h^2$, hence $\cos(\pi+0.5) \approx -1 + \frac{1}{2}\cdot 0.0.25 = -0.875$.

i0c. For every integral there is some C so that $|E_n| \le C\frac{1}{n^q}$. Practical version: $|E_h| \le Ch^q$.
Method of left/right rectangles: order 1. Trapezoid method: order 2. Simpson method: order 4.

i0d. $E_{2n} \approx C\frac{1}{(2n)^3} = \frac{1}{8}C\frac{1}{n^3} \approx \frac{1}{8}E_n$. Thus $E_{200} \approx \frac{1}{8}E_{100} = 0.00125$.
We want $0.0001 = E_{100a} \approx \frac{1}{a^3}E_{100} = \frac{1}{a^3}\cdot 0.01$, hence $a = \sqrt[3]{100}$, so we want $n = 100\cdot\sqrt[3]{100}$.

i0e. Order 2. $E_{2n} \approx C\frac{1}{(2n)^2} = \frac{1}{4}C\frac{1}{n^3} \approx \frac{1}{4}E_n$. Thus $E_{200} \approx \frac{1}{4}E_{100} = 0.004$.
We want $0.001 = E_{100a} \approx \frac{1}{a^2}E_{100} = \frac{1}{a^2}\cdot 0.016$, hence $a = \sqrt{16} = 4$, so we want $n = 400$.

i1a. $h=1$, points $0,1,2$. Two possibilities.
Left rectangles: $I \approx 1\cdot[\sqrt{0}+\sqrt{1}] = 1$. Right rectangles: $I \approx 1\cdot[\sqrt{1}+\sqrt{2}] = 1+\sqrt{2}$.

i1b. $n=2$, points $2,4,6$. Two possibilities.
Left rectangles: $I \approx 2\cdot[0+1] = 2$. Right rectangles: $I \approx 2\cdot[1+2] = 6$.

i2a. $n=2$, points $0,1,2$. $I \approx \frac{1}{2}\cdot 1\cdot[\sqrt{0}+2\sqrt{1}+\sqrt{2}] = 1 + \frac{1}{2}\sqrt{2}$.

i2b. $h=2$, points $2,4,6$. $I \approx \frac{1}{2}\cdot 2\cdot[0+2\cdot 1+2] = 4$.

r0a. For a particular sequence $\{x_k\}$ generated by a method of order q the following should be (approximately, for large values of k, in case of convergence) true: $|E_{k+1}| \approx C|E_k|^q$, where C is a number specific for this sequence (not a general parameter of the method).
Bisection: $q=1$. Newton: $q=2$.

r0b. We expect that $|E_7| \approx C|E_6|^2$, that is, $0.0003 = c\cdot 0.0001$. Hence $c \approx 3$. Therefore
$|E_8| \approx c|E_7|^2 \approx 3\cdot 0.00000009 = 0.00000027$.

r1b. Rewrite: $x^2 - x - 1 = 0$, $f(x) = x^2 - x - 1$.
Check: $f(0) = -1 < 0$, $f(4) = 11 > 0$, it's OK.
(0) $a_0 = 0$, $f(a_0) < 0$; $b_0 = 4$, $f(b_0) > 0$.
Center $m_0 = \frac{1}{2}(0+4) = 2$, $f(2) = 1 > 0$, hence $a_1 = a_0$, $b_1 = m_0$.
(1) $a_1 = 0$, $f(a_1) < 0$; $b_1 = 2$, $f(b_1) > 0$.
Center $m_1 = 1$, $f(1) = -1 < 0$, hence $a_2 = m_1$, $b_2 = b_1$.
(2) $a_2 = 1$, $f(a_2) < 0$; $b_2 = 2$, $f(b_2) > 0$.
Center $m_2 = 1.5$, $f(1.5) = -0.25 < 0$, hence $a_3 = m_2$, $b_3 = b_2$.

r1c. Rewrite: $\frac{1}{x} - x + 2 = 0$, $f(x) = \frac{1}{x} - x + 2$.
Check: $f(1) = 2 > 0$, $f(9) = \frac{1}{9} - 7 < 0$, it's OK.
(0) $a_0 = 1$, $f(a_0) > 0$; $b_0 = 9$, $f(b_0) < 0$.
Center $m_0 = \frac{1}{2}(1+9) = 5$, $f(5) = \frac{1}{5} - 3 < 0$, hence $a_1 = a_0$, $b_1 = m_0$.
(1) $a_1 = 1$, $f(a_1) > 0$; $b_1 = 5$, $f(b_1) < 0$.
Center $m_1 = 3$, $f(3) = \frac{1}{3} - 1 < 0$, hence $a_2 = a_1$, $b_2 = m_1$.
(2) $a_2 = 1$, $f(a_2) > 0$; $b_2 = 3$, $f(b_2) < 0$.
Center $m_2 = 2$, $f(2) = \frac{1}{2} > 0$, hence $a_3 = m_2$, $b_3 = b_2$.

r2b. Rewrite: $x^2 - x - 1 = 0$, $f(x) = x^2 - x - 1$, then $f'(x) = 2x - 1$.
$x_{k+1} = x_k - \frac{x_k^2 - x_k - 1}{2x_k - 1} = \frac{x_k^2+1}{2x_k-1}$. $x_0 = 0$, $x_1 = -1$, $x_2 = -\frac{2}{3}$, ...

r2c. Rewrite: $\frac{1}{x} - x + 2 = 0$, $f(x) = \frac{1}{x} - x + 2$, then $f'(x) = -\frac{1}{x^2} - 1$.
$x_{k+1} = x_k - \frac{\frac{1}{x_k} - x_k + 2}{-\frac{1}{x_k^2} - 1} = \frac{2x_k + 2x_k^2}{1+x_k^2}$. $x_0 = 1$, $x_1 = 2$, $x_2 = \frac{12}{5}$, ...

r2d. $f(x)=x^3-A$, then $f'(x)=3x^2$ and so $x_{k+1}=x_k-\frac{x_k^3-A}{3x_k^2}=\frac{1}{3}\big(2x_k+\frac{A}{x_k^2}\big)$.

r2e. $f(x)=e^{-x}-x$, then $f'(x)=-e^{-x}-1$ and so
$x_{k+1}=x_k-\frac{e^{-x_k}-x_k}{-e^{-x_k}-1}=x_k+\frac{e^{-x_k}-x_k}{e^{-x_k}+1}=\frac{x_k+1}{1+e^{x_k}}$.

r3b. Rewrite to fixed point: for instance $x^2-1=x$, hence $\varphi=x^2-1$.
Iteration: $x_{k+1}=\varphi(x_k)=x_k^2-1$. $x_0=3$, $x_1=3^2-1=8$, $x_2=8^2-1=63$, ...
We look at $\varphi'(x)=2x$, for $x=3$ we get $\varphi'(3)=6\geq 1$, does not look good.
Alternative reformulation to fixed point: $x=\sqrt{x+1}$, hence $\varphi=\sqrt{x+1}$.
Iteration: $x_{k+1}=\varphi(x_k)=\sqrt{x_k+1}$. $x_0=3$, $x_1=\sqrt{3+1}=2$, $x_2=\sqrt{2+1}=\sqrt{3}$, ...
We look at $\varphi'(x)=\frac{1}{2\sqrt{x+1}}$, for $x=3$ we get $\varphi'(3)=\frac{1}{4}<0$, this looks hopeful. The second question is whether φ maps some interval I around $x=3$ onto itself; this could be investigated or we simply try this iteration.

r3c. Rewrite to fixed point: for instance $\frac{1}{x}+2=x$, hence $\varphi=\frac{1}{x}+2$.
Iteration: $x_{k+1}=\varphi(x_k)=\frac{1}{x_k}+2$. $x_0=3$, $x_1=\frac{1}{3}+2=\frac{7}{3}$, $x_2=\frac{3}{7}+2=\frac{17}{7}$, ...
We look at $\varphi'(x)=-\frac{1}{x^2}$, for $x=3$ we get $|\varphi'(3)|=\frac{1}{9}<1$, this looks hopeful. The second question is whether φ maps some interval I around $x=3$ onto itself; this could be investigated or we simply try this iteration.
Alternative reformulation to fixed point: $x=\frac{1}{x-2}$, hence $\varphi=\frac{1}{x-2}$.
Iteration: $x_{k+1}=\varphi(x_k)=\frac{1}{x_k-2}$. $x_0=3$, $x_1=\frac{1}{3-2}=1$, $x_2=\frac{1}{1-2}=-1$, ...
We look at $\varphi'(x)=\frac{-1}{(x-2)^2}$, for $x=3$ we get $|\varphi'(3)|=1$, this does not look too good but also not too bad). An experiment tells us more.

r3d(+).
Re: r3b. Iteration: $x_{k+1}=\lambda(x_k^2-1)+(1-\lambda)x_k$.
$\varphi'_\lambda(3)=0\implies\lambda_{\text{opt}}=-\frac{1}{5}$.
Alternative: Iteration: $x_{k+1}=\lambda\sqrt{x_k+1}+(1-\lambda)x_k$.
$\varphi'_\lambda(3)=0\implies\lambda_{\text{opt}}=\frac{4}{3}$.
Re: r3c. Iteration: $x_{k+1}=\lambda\big(\frac{1}{x_k}+2\big)$.
$\varphi'_\lambda(3)=0\implies\lambda_{\text{opt}}=\frac{9}{10}$.
Alternative: Iteration: $x_{k+1}=\lambda\frac{1}{x_k-2}+(1-\lambda)x_k$.
$\varphi'_\lambda(3)=0\implies\lambda_{\text{opt}}=\frac{1}{2}$.

d0b. The order of the method implies that errors should satisfy $E_h\approx ch^2$. Therefore
$E_{h/3}\approx c\big(\frac{1}{3}h\big)^2=\frac{1}{9}ch^2=\frac{1}{9}\cdot 0.0027=0.0003$.
We want $0.00001=E_{ah}=a^2E_h=a^2\cdot 0.0027$, hence $a=\frac{1}{\sqrt{270}}$, so we want the step $\frac{h}{\sqrt{270}}$.

d0c. The order of the method implies that errors should satisfy $E_h\approx ch^4$. Therefore
$E_{h/2}\approx c\big(\frac{1}{2}h\big)^4=\frac{1}{16}ch^2=\frac{1}{16}\cdot 0.004=0.00025$.

d1a. The main iterative formula is $y_{k+1}=y_k+h\cdot f(x_k,y_k)$. Step size is given, from that we get the partition size $n=4$. Scheme:
(0) $x_0=1$, $y_0=13$.
(1) $x_{k+1}=x_k+1$, $y_{k+1}=y_k+1\cdot(x_k+y_k)=x_k+2y_k$ for $i=0,\dots,3$.
Points: $(1,13)$, $(2,27)$, $(3,56)$.

d1b. The main iterative formula is $y_{k+1}=y_k+h\cdot f(x_k,y_k)$. The partition size is given, from that we get the step size $h=\frac{1}{2}$. Scheme:
(0) $x_0=1$, $y_0=3$.
(1) $x_{k+1}=x_k+\frac{1}{2}$, $y_{k+1}=y_k+\frac{1}{2}\cdot(2x_ky_k)=(x_k+1)y_k$ for $i=0,\dots,3$.
Points: $(1,3)$, $(1.5,6)$, $(2,15)$.

s0b. Cubic complexity means that the runtime of the program is proportional to n^3, that is, $T_n\approx cn^3$.
If we double n, we get $T_{2n}=c(2n)^3=8cn^3=8T_n$.
So for $n=2000=2\cdot 1000$ we can expect approximate runtime $8\cdot 5=40$ hours.

s1a. Step 1 (GEM): The extended matrix of the system $\begin{pmatrix} 1 & 1 & -1 & -2 \\ -1 & 0 & 2 & 3 \\ 1 & 1 & 1 & 2 \end{pmatrix}$ is reduced using Gaussian elimination to upper-triangular form: $\begin{pmatrix} 1 & 1 & -1 & -2 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 2 & 4 \end{pmatrix}$

Step 2 (BS): The new system of equations $\left[\begin{aligned} x+y-z&=-2 \\ y+z&=1 \\ 2z&=4 \end{aligned}\right.$ is solved last to first:
$z=2$, $y=1-z=-1$, $x=-2-y+z=1$.

s1b. Step 1 (GEM): The extended matrix of the system $\begin{pmatrix} 1 & 2 & 1 & 0 \\ 1 & 3 & 0 & -2 \\ -2 & -4 & 0 & 2 \end{pmatrix}$ is reduced using Gaussian elimination to upper-triangular form: $\begin{pmatrix} 1 & 2 & 1 & 0 \\ 0 & 1 & -1 & -2 \\ 0 & 0 & 2 & 2 \end{pmatrix}$

Step 2 (BS): The new system of equations $\left[\begin{aligned} x_1+2x_2+x_3&=0 \\ x_2-x_3&=-2 \\ 2x_3&=2 \end{aligned}\right.$ is solved last to first
$x_3=1$, $x_2=-2+x_3=-1$, $x_1=-2x_2-x_3=1$.

s2a. We rewrite the system as $\left[\begin{aligned} x&=1+z \\ y&=1-2x+z \\ z&=1+x+2y \end{aligned}\right.$.
These are the iteration formulas. If we use the updated values of variables, we obtain the Gauss-Seidel iteration. Formally:

$$\left[\begin{aligned} x_{k+1}&=1+z_k \\ y_{k+1}&=1-2x_{k+1}+z_k \\ z_{k+1}&=1+x_{k+1}+2y_{k+1} \end{aligned}\right. \qquad \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \implies \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \implies \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix} \implies \cdots$$

The Jacobi method updates only after the whole iterative step is finished, so

$$\left[\begin{aligned} x_{k+1}&=1+z_k \\ y_{k+1}&=1-2x_k+z_k \\ z_{k+1}&=1+x_k+2y_k \end{aligned}\right. \qquad \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \implies \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \implies \begin{pmatrix} 2 \\ 0 \\ 4 \end{pmatrix} \implies \cdots$$

s2b. We rewrite the system as $\left[\begin{aligned} x&=2-z \\ y&=-1+x+2z \\ z&=1-x-2y \end{aligned}\right.$
These are the iteration formulas. If we use the updated values of variables, we obtain the Gauss-Seidel iteration. Formally:

$$\left[\begin{aligned} x_{k+1}&=2-z_k \\ y_{k+1}&=-1+x_{k+1}+2z_k \\ z_{k+1}&=1-x_{k+1}-2y_{k+1} \end{aligned}\right. \qquad \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \implies \begin{pmatrix} 2 \\ 1 \\ -3 \end{pmatrix} \implies \begin{pmatrix} 5 \\ -2 \\ 0 \end{pmatrix} \implies \cdots$$

The Jacobi method updates only after the whole iterative step is finished, so

$$\left[\begin{aligned} x_{k+1}&=2-z_k \\ y_{k+1}&=-1+x_k+2z_k \\ z_{k+1}&=1-x_k-2y_k \end{aligned}\right. \qquad \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} \implies \begin{pmatrix} 2 \\ -1 \\ 1 \end{pmatrix} \implies \begin{pmatrix} 1 \\ 3 \\ 1 \end{pmatrix} \implies \cdots$$