MA2501 Numerical Methods
Exam preparation

April 16-17, 2024

Consider the function

$$f(x) := 2x - \sin(x) + 2.$$

In order to solve the equation $f(x) = 0$, it is possible to apply a fixed point iteration of the form

$$x_{k+1} = x_k - \frac{1}{2} f(x_k).$$

Question:
Show that the equation $f(x) = 0$ has a unique solution $\hat{x}$, and that the iteration converges for every starting value $x_0 \in \mathbb{R}$ to $\hat{x}$.

Answer:
We first note that $\hat{x}$ is a solution of the equation $f(x) = 0$ iff $\hat{x}$ is a fixed point of the mapping $x \mapsto \Phi(x)$.
Since $f(x) = 2x - \sin(x) + 2$, the expression for $\Phi(x)$ becomes

$$\Phi(x) := x - \frac{1}{2} f(x) = \frac{1}{2} \sin(x) - 1$$

Next, we must show that our iteration is a contraction.
This requires Lipschitz continuity, hence a bound on the derivative:

$$\sup_{x \in \mathbb{R}} |\Phi'(x)| = \sup_{x \in \mathbb{R}} \left| \frac{1}{2} \cos(x) \right| = \frac{1}{2}$$

The final derivation becomes

$$\begin{aligned} |\Phi(x) - \Phi(y)| &= \left| \int_x^y \Phi'(s)\, ds \right| \\ &\leq \int_x^y |\Phi'(s)|\, ds \\ &\leq \frac{1}{2}|x - y| \end{aligned}$$

Hence, Φ is a contraction on $\mathbb{R}$ with contraction factor $1/2 < 1$. We can use Banachs fixed point theorem to conclude that Φ has a unique fixed point $\hat{x}$, and our fixed point iteration $x_{k+1} = \Phi(x_k)$ converges for all starting values $x_0 \in \mathbb{R}$ to $\hat{x}$.

Question:
Compute one step of the fixed point iteration with a starting value $x_0 = 0$. Use your result to estimate, after how many steps we have $|x_k - \hat{x}| \leq 2^{-20}$.

Answer:
Direct insertion of $x_0 = 0$ yields

$$x_{k+1} = \frac{1}{2}\sin(0) - 1 = -1$$

We apply formula (1.10) in the proof of the Theorem 1.4:

$$|x_k - x_1| \leq \frac{L^k}{1-L}|x_0 - x_1|$$

Direct insertion of relevant values yields

$$|x_k - x_1| \leq \frac{1}{2^{k-1}}$$

For $k \geq 21$, the right-hand side is smaller than or equal to 2^{-20}. Therefore, the required accuracy is reached after at most 21 steps.

Given the following nonlinear system of equations:

$$x_1^2 + x_2^2 = 1$$
$$x_1^3 - x_2 = 2$$

This system has two sets of solutions, one in the domain $-1 \le x_1, x_2 \le 0$ and one in the domain $0 \le x_1, x_2 \le 1$.

Question:
Set up Newton's method for the nonlinear equation system.

Answer:
We rewrite the system of equations as

$$F(X) = \begin{bmatrix} x_1^2 + x_2^2 - 1 \\ x_1^3 - x_2 - 2 \end{bmatrix}$$

The Jacobi matrix is defined as

$$J(X) = \begin{bmatrix} 2x_1 & 2x_2 \\ 3x_1^2 & -1 \end{bmatrix}$$

Newton's method in 2D becomes

$$X^{(n+1)} = X^{(n)} - J(X^{(n)})^{-1} F(X^{(n)})$$

Question:
Select a set of appropriate initial values for x_1 and x_2 and make two iterations of Newtons method.

Answer:
We must ensure that the Jacobian is not zero:

$$\det(J(X)) = -2x_1(1 + 3x_1x_2) \neq 0$$

The initial values must be kept away from two curves:

$$x_1 = 0 \quad , \quad 3x_1x_2 = -1$$

First, we recall the formula for inversion of a 2×2-matrix:

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix} \implies M^{-1} = \frac{1}{ad - bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

The inverse Jacobi-matrix is given by

$$J(X)^{-1} = -\frac{1}{2x_1(1 + 3x_1x_2)} \begin{bmatrix} -1 & -2x_2 \\ -3x_1^2 & 2x_1 \end{bmatrix}$$

Thus, we can find an explicit expression for our scheme:

$$
\begin{aligned}
& X - J(X)^{-1}F(X) \\
=& \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \frac{1}{2x_1(1+3x_1x_2)} \begin{bmatrix} -1 & -2x_2 \\ -3x_1^2 & 2x_1 \end{bmatrix} \begin{bmatrix} x_1^2 + x_2^2 - 1 \\ x_1^3 - x_2 - 2 \end{bmatrix} \\
=& \frac{1}{2x_1(1+3x_1x_2)} \begin{bmatrix} 4x_1^3x_2 + x_1^2 + x_2^2 + 4x_2 + 1 \\ 3x_1^2(x_2^2 - x_1^2 + 3) - 4x_1 \end{bmatrix}
\end{aligned}
$$

The final expression for each step is

$$
\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \mapsto \frac{1}{2x_1(1+3x_1x_2)} \begin{bmatrix} 4x_1^3x_2 + x_1^2 + x_2^2 + 4x_2 + 1 \\ 3x_1^2(x_2^2 - x_1^2 + 3) - 4x_1 \end{bmatrix}
$$

Consider the matrix

$$\mathbf{A} = \begin{bmatrix} 9 & -3 & -3 \\ -3 & 10 & 1 \\ -3 & 1 & 5 \end{bmatrix}$$

Question:
Show that $\mathbf{A}$ has a unique Cholesky factorization, without computing it.

Answer:
$\mathbf{A}$ is both real and symmetric, so the eigenvalues are real.

It is also strictly diagonally dominant with positive diagonal elements, as shown below:

$$\begin{aligned} |-3| + |-3| &= 6 < 9 \\ |-3| + |1| &= 4 < 10 \\ |-3| + |1| &= 4 < 5 \end{aligned}$$

It follows from Gerschgorin's Theorem that all the eigenvalues of are positive. This in turn implies that $\mathbf{A}$ is positive definite. Hence, $\mathbf{A}$ is symmetric positive definite (SPD) and consequently has a unique Cholesky factorization.

Question:
Compute the Cholesky factorization of $\mathbf{A}$, and use it to solve the linear system $\mathbf{A}\mathbf{x} = \mathbf{b}$ with $\mathbf{b} = [-9, 1.5, 5]^T$.

Answer:
The algorithm for Cholesky factorization yields

$$
\begin{aligned}
l_{11} &= \sqrt{a_{11}} &&= \sqrt{9} &&= 3 \\
l_{21} &= a_{21}/l_{11} &&= -3/3 &&= -1 \\
l_{31} &= a_{31}/l_{11} &&= -3/3 &&= -1 \\
l_{22} &= \sqrt{a_{22} - l_{21}^2} &&= \sqrt{10 - (-1)^2} &&= 3 \\
l_{32} &= (a_{32} - l_{21}l_{31})/l_{22} &&= (1 - (-1)(-1))/3 &&= 0 \\
l_{33} &= \sqrt{a_{33} - l_{31}^2 - l_{32}^2} &&= \sqrt{5 - (-1)^2 - (0)^2} &&= 2
\end{aligned}
$$

The factorization is $\mathbf{A} = \mathbf{L}\mathbf{L}^T$, where

$$\mathbf{L} = \begin{bmatrix} 3 & 0 & 0 \\ -1 & 3 & 0 \\ -1 & 0 & 2 \end{bmatrix}$$

We must solve two separate equation systems:

$$\mathbf{L}\mathbf{y} = \mathbf{b} \qquad , \qquad \mathbf{L}^T\mathbf{x} = \mathbf{y}$$

The first system is solved with forward substitution:

$$
\begin{aligned}
y_1 &= \frac{b_1}{l_{11}} && = \frac{-9}{3} && = -3 \\
y_2 &= \frac{b_2 - l_{21} y_1}{l_{22}} && = \frac{-1.5 - (-1)(-3)}{3} && = -1.5 \\
y_3 &= \frac{b_3 - l_{31} y_1 - l_{32} y_2}{l_{33}} && = \frac{5 - (-1)(-3) - 0(-1.5)}{2} && = 1
\end{aligned}
$$

Spring 2015, Problem 2b)

The second system is solved with backward substitution:

$$
\begin{aligned}
x_1 &= \frac{y_3}{l_{11}} && = \frac{-9}{3} && = 0.5 \\
x_2 &= \frac{y_2 - l_{32}x_3}{l_{22}} && = \frac{-1.5 - 0(0.5)}{3} && = -0.5 \\
x_3 &= \frac{y_1 - l_{31}x_3 - l_{21}x_2}{l_{11}} && = \frac{-3 - (-1)(0.5) - (-1)(-0.5)}{3} && = -1
\end{aligned}
$$

The solution of $\mathbf{Ax} = \mathbf{b}$ is therefore

$$\mathbf{x} = [-1, -0.5, 0.5]$$

Question:

Perform 1 iteration of the SOR method with relaxation parameter $\omega = 1.1$ for the linear system $\mathbf{Ax} = \mathbf{b}$ from b). Use the starting point $x^{(0)} = [0, 0, 0]^T$.

Does it look like the iterations will converge towards the solution?

Will the iterations converge for an arbitrary starting point?

Answer:
The first iteration componentwise is

$$
\begin{aligned}
x_1^{(1)} &= \omega \frac{b_1 - a_{12}x_2^{(0)} - a_{13}x_3^{(0)}}{a_{11}} + (1-\omega)x_1^{(0)} &&= -1.1 \\
x_1^{(1)} &= \omega \frac{b_1 - a_{21}x_1^{(0)} - a_{23}x_3^{(0)}}{a_{22}} + (1-\omega)x_2^{(0)} &&= -0.528 \\
x_1^{(1)} &= \omega \frac{b_1 - a_{31}x_1^{(0)} - a_{32}x_2^{(0)}}{a_{33}} + (1-\omega)x_3^{(0)} &&= 0.49016
\end{aligned}
$$

Spring 2015, Problem 2c)

The relative error in every component is at most 10% after just a single iteration, and the error decreases in the later components, where we use more updated values.
The iterations will in fact converge regardless of the starting point because $\mathbf{A}$ is SPD.

Question:

Verify that the matrices

$$M = \begin{bmatrix} 3 & 0 & 0 \\ 1 & 6 & 0 \\ 2 & 3 & 1 \end{bmatrix} \quad , \quad N = \begin{bmatrix} 2 & 6 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

give a LU decomposition of the matrix

$$A = \begin{bmatrix} 6 & 18 & 3 \\ 2 & 12 & 1 \\ 4 & 15 & 3 \end{bmatrix}$$

Use the LU decomposition to solve the linear system

$$Ax = (3, 19, 0)^T$$

Answer:

Direct computation yields

$$\begin{bmatrix} 3 & 0 & 0 \\ 1 & 6 & 0 \\ 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} 2 & 6 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 6 & 18 & 3 \\ 2 & 12 & 1 \\ 4 & 15 & 3 \end{bmatrix}$$

First, we set $y = Nx$ and solve $My = (3, 19, 0)^T$, which yields the temporary solution

$$y = (1, 3, -11)^T$$

Then, we return to $y = Nx$ and solve $Nx = (1, 3, -11)^T$, which yields the final solution

$$x = (-3, 3, -11)^T$$

Question:
Given the data set

x	1	3/2	2
y	-1	3	3

Find the lowest-degree polynomial $p(x)$ that interpolates the set.

Answer:
The Lagrange interpolation formula is

$$p(x) = \sum_{i=0}^{2} y_i l_i(x)$$

We find the component functions directly:

$$
\begin{aligned}
l_1(x) &= \frac{(x-3/2)(x-2)}{(1-3/2)(1-2)} && = (2x-3)(x-2) \\
l_2(x) &= \frac{(x-1)(x-2)}{(3/2-1)(3/2-2)} && = -4(x-1)(x-2) \\
l_3(x) &= \frac{(x-1)(x-3/2)}{(2-1)(2-3/2)} && = (x-1)(2x-3)
\end{aligned}
$$

The final polynomial becomes

$$p(x) = -1l_0(x) + 3l_1(x) + 3l_2(x) = -8x^2 + 28x - 21$$

Question:

Determine the constants a, b and c such that $p(x)$ interpolates the function

$$f(x) = a\cos(\pi x) + b\sin(\pi x) + c$$

in the three points $(1, -1)$, $(3/2, 3)$ and $(2, 3)$.

Answer:

The system of equations is

$$\begin{aligned} f(1) &= -a + c = -1 \\ f(3/2) &= -b + c = 3 \\ f(2) &= a + c = 3 \end{aligned}$$

The solution is $a = 2$, $b = -2$ and $c = 1$.

Question:
Find an upper limit for the error $|f(x) - p(x)|$ when $x \in [1, 2]$.

Answer:
Since the interpolation point are uniformly distributed, we can invoke the general formula

$$|f(x) - p(x)| \leq \left(\frac{h^{n+1}}{4(n+1)}\right) \max_{1 \leq x \leq 2} |f^{(n+1)}(x)|$$

We need the 3rd and 4th derivatives:

$$\begin{aligned} f(x) &= 2[\cos(\pi x) - \sin(\pi x)] + 1 \\ f^{(3)}(x) &= 2\pi^3[\sin(\pi x) + \cos(\pi x)] \\ f^{(4)}(x) &= 2\pi^4[\cos(\pi x) - \sin(\pi x)] \end{aligned}$$

Setting $f^{(4)}(x) = 0$ yields

$$\begin{aligned} \tan(\pi x) &= 1 \\ \implies \pi x &= \frac{\pi}{4} + k\pi \\ \implies x &= \frac{4k+1}{4} \end{aligned}$$

Since $x \in [1, 2]$, we choose $x = 5/4$ and test this point, in addition to the endpoints:

$$f^{(3)}(1) = -2\pi^3$$
$$f^{(3)}(5/4) = -2\sqrt{2}\pi^3$$
$$f^{(3)}(2) = 2\pi^3$$

We insert relevant values and get

$$|f(x) - p(x)| \leq \frac{1}{12}(2\sqrt{2}\pi^3)\left(\frac{1}{2}\right)^3 = \frac{\pi^3\sqrt{2}}{48} \approx 0.9135$$

Denote by f_n, $n \in \mathbb{N}$, the polynomial of degree n that interpolates the function $f(x) = e^x + e^{-x}$ in equidistant interpolation points in the interval $[0, 1]$.

Question:
Show that $f_n(x) \to f(x)$ for every $x \in \mathbb{R}$.

Answer:
For every $x \in \mathbb{R}$ and $n \in \mathbb{N}$ there exists ξ (depending on both x and n) lying either in the interval $[0, 1]$ or between x and the interval $[0, 1]$ such that

$$f(x) - f_n(x) = \frac{1}{(n+1)!} f^{(n+1)}(\xi) \prod_{i=1}^{n} \left(x - \frac{i}{n} \right)$$

The n-th order derivative has an analytical expression:

$$f(\xi) = e^{\xi} + e^{-\xi}$$
$$\implies f^{(n)}(\xi) = e^{\xi} + (-1)^n e^{-\xi}$$

Since ξ lies either in $[0,1]$ or between x and this interval, we have the following bounds for e^{ξ} and $e^{-\xi}$:

$$\begin{aligned} e^{\xi} &\leq \max\{e^{x}, e^{1}\} \\ e^{-\xi} &\leq \max\{e^{-x}, e^{0}\} \end{aligned}$$

Thus, we have an upper bound for the derivative:

$$\begin{aligned} |f^{(n+1)}| &= |e^{\xi} + (-1)^{n} e^{-\xi}| \\ &\leq e^{\xi} + e^{-\xi} \\ &\leq \max\{e^{x}, e\} + \max\{e^{-x}, 1\} \\ &:= C \end{aligned}$$

We have three different inequalities for the latter product term:

$$0 \le x \le 1 \quad \Longrightarrow \quad \left| \prod_{i=1}^{n} \left(x - \frac{i}{n} \right) \right| \le 1$$

$$x > 1 \quad \Longrightarrow \quad \left| \prod_{i=1}^{n} \left(x - \frac{i}{n} \right) \right| \le x^{n+1}$$

$$x < 0 \quad \Longrightarrow \quad \left| \prod_{i=1}^{n} \left(x - \frac{i}{n} \right) \right| \le (-x+1)^{n+1}$$

Since $x^{n+1} \leq (x+1)^{n+1}$ and $1 \leq (|x|+1)^{n+1}$, we can sum up all the three inequalities into one single:

$$\left|\prod_{i=1}^{n}\left(x-\frac{i}{n}\right)\right| \leq (|x|+1)^{n+1}$$

The final inequality for everything becomes

$$|f(x)-f_n(x)| = \frac{C}{(n+1)!}(|x|+1)^{n+1}$$

We have the universal limit

$$\lim_{n\to\infty}\frac{(|x|+1)^{n+1}}{(n+1)!} = 0 \qquad , \qquad x \in \mathbb{R}$$

Hence, we have shown that $f_n(x) \to f(x)$.

Question:

Provide an estimate for

$$\sup_{0 \le x \le 1} |f_5(x) - f(x)|$$

Answer:

For equidistant interpolation points on the interval $[0, 1]$ we have the universal estimate

$$\sup_{0 \le x \le 1} |f(x) - f_n(x)| \le \frac{h^{n+1}}{4(n+1)} \sup_{0 \le x \le 1} |f^{(n+1)}(x)|$$

Since $h = 1/n$ and $n = 5$, we obtain

$$\sup_{0\le x\le 1} |f(x) - f_5(x)| \le \frac{1}{5^6 \cdot 4 \cdot 6} \sup_{0\le x\le 1} |e^x + e^{-x}|$$

Since $e^x + e^{-x}$ is convex, it attains its maximum on the interval's boundary. Thus

$$\sup_{0\le x\le 1} |e^x + e^{-x}| = e + e^{-1}$$

Combining all values yields

$$\sup_{0\le x\le 1} |f(x) - f_5(x)| \le \frac{e + e^{-1}}{375000} \approx 8.23 \cdot 10^{-6}$$

Question:

Find coefficients a and b such that the expression

$$\int_{-1}^{1} [ax^2 + b\sin(x) - e^x]^2 \, dx$$

is as small as possible.

Answer:

This is a least squares problem:

$$\min_{a,b\in\mathbb{R}} \int_{-1}^{1} I(a,b;x) \, dx$$

The optimum is found by setting the gradient equal to zero:

$$\nabla_{a,b} \left(\int_{-1}^{1} I(a,b;x) \, dx \right) = \mathbf{0}$$

First, find the gradient of the integrand I with respect to a and b:

$$\begin{aligned} I &= [ax^2 + b\sin(x) - e^x]^2 \\ \frac{\partial I}{\partial a} &= 2[ax^2 + b\sin(x) - e^x] \cdot x^2 \\ \frac{\partial I}{\partial b} &= 2[ax^2 + b\sin(x) - e^x] \cdot \sin(x) \end{aligned}$$

By setting the gradient of the integrand equal to 0 and writing the system on matrix form, we get

$$\begin{bmatrix} \int_{-1}^{1} x^4\,dx & \int_{-1}^{1} x^2 \sin(x)\,dx \\ \int_{-1}^{1} x^2 \sin(x)\,dx & \int_{-1}^{1} \sin^2(x)\,dx \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \int_{-1}^{1} x^2 e^x\,dx \\ \int_{-1}^{1} e^x \sin(x)\,dx \end{bmatrix}$$

The off-diagonal elements are zero due to odd integrands:

$$\int_{-1}^{1} x^2 \sin(x)\, dx = 0$$

The diagonal elements are

$$\int_{-1}^{1} x^4\, dx = \frac{2}{5} \qquad , \qquad \int_{-1}^{1} \sin^2(x)\, dx = 1 - \frac{\sin(2)}{2}$$

The right-hand side elements are

$$\int_{-1}^{1} x^2 e^x\, dx = e - 5e^{-1}$$
$$\int_{-1}^{1} e^x \sin(x)\, dx = \sin(1)\cosh(1) - \cos(1)\sinh(1)$$

Since the equation system has a diagonal matrix, we can solve it trivially and obtain the final values directly:

$$a = \frac{\int_{-1}^{1} x^2 e^x \, dx}{\int_{-1}^{1} x^4 \, dx} \qquad , \qquad b = \frac{\int_{-1}^{1} e^x \sin(x) \, dx}{\int_{-1}^{1} \sin^2(x) \, dx}$$

Explicitly, we get the expressions

$$\begin{aligned} a &= \frac{5[e - 5e^{-1}]}{2} \\ b &= \frac{2[\sin(1)\cosh(1) - \cos(1)\sinh(1)]}{2 - \sin(2)} \end{aligned}$$

Question:
Estimate the value of the integral

$$\int_1^3 x \ln(x)\, dx$$

using the composite Simpson's rule. Choose the number of subintervals n such that the absolute integration error is guaranteed to not exceed 10^{-4}.

Answer:
The error term for composite Simpson with f on $[a,b]$ and n subintervals is given by

$$e = -\frac{(b-a)^5}{180n^4} f^{(4)}(\xi)$$

For our function $f = x\ln(x)$, we have

$$f^{(4)}(x) = \frac{2}{x^3} \qquad , \qquad \max_{1\le x\le 3} \left|\frac{2}{x^3}\right| = 2$$

Continuation 2013, Problem 5)

By inserting relevant values and using the tolerance $\tau = 10^{-4}$ instead of e, we get:

$$\frac{(3-1)^5}{180n^4} \cdot 2 \leq \tau \quad \Longrightarrow \quad n \geq \frac{2}{(45\tau)^{1/4}}$$

This yields $n \geq 7.72$, and since n must be even, we get $n \geq 8$.

Using the composite rule yields $S = 2.9437737349$, which is very close to the exact value $4.5\ln(3) - 2 \approx 2.943755299$.

Consider the nodes

$$c_1 = \frac{1}{6} \qquad , \qquad c_2 = \frac{1}{2} \qquad , \qquad c_3 = \frac{5}{6}$$

and the corresponding quadrature formula

$$Q(f) = w_1 f(c_1) + w_2 f(c_2) + w_3 f(c_3)$$

which approximates the integral $\int_0^1 f(x)\,dx$.

Question:
Determine the weights w_1, w_2, w_3 so that the formula is exact for polynomials of degree up to 2, that is

$$Q(P) = \int_0^1 P(x)\,dx \qquad , \qquad P \in \mathbb{P}^2$$

Answer:
Since the interval is $[0,1]$ instead of $[-1,1]$, we must transfer the monomials $1, x, x^2, \ldots$ to the new domain, and they become

$$1,\ (x-1/2),\ (x-1/2)^2$$

The quadrature formula is

$$Q(f) = w_1 f(1/6) + w_2 f(1/2) + w_3 f(5/6)$$

From this, we get three equations:

$$\begin{aligned} w_1 + w_2 + w_3 &= 1 \\ -\frac{1}{3}w_1 + \frac{1}{3}w_3 &= 0 \\ \frac{1}{9}w_1 + \frac{1}{9}w_3 &= \frac{1}{12} \end{aligned}$$

The solution of the equation system is

$$w_1 = \frac{3}{8} \qquad , \qquad w_2 = \frac{1}{4} \qquad , \qquad w_3 = \frac{3}{8}$$

The final expression is

$$Q(f) = \frac{3}{8} f\left(\frac{1}{6}\right) + \frac{1}{4} f\left(\frac{1}{2}\right) + \frac{3}{8} f\left(\frac{1}{6}\right)$$

Question:

Compute the error $E_k = \left| Q(x^k) - \int_0^1 x^k \, dx \right|$ for the lowest integer k such that E_k is not zero.

Answer:

We test for $k = 3$ first:

$$\begin{aligned} Q((x-1/2)^3) = \frac{3}{8}\frac{1}{6^3} + \frac{1}{4}\frac{1}{2^4} + \frac{3}{8}\frac{5^3}{6^3} &= \frac{1}{4} \\ \int_0^1 x^3 \, dx &= \frac{1}{4} \end{aligned}$$

Since $E_3 = 0$, we can proceed with $k = 4$:

$$\begin{aligned} Q((x-1/2)^4) &= \frac{3}{8}\frac{1}{6^4} + \frac{1}{4}\frac{1}{2^4} + \frac{3}{8}\frac{5^4}{6^4} &&= \frac{439}{2160} \\ \int_0^1 x^4\,dx & &&= \frac{1}{5} \end{aligned}$$

We get $E_4 = 7/2160 \approx 3.24 \times 10^{-3}$, so $k = 4$.

Consider the second order differential equation for $y(t)$

$$y'' + y' \sin(y) = 0$$

with initial conditions

$$y(0) = 1 \quad , \quad y'(0) = 2$$

Question:
We introduce the new variables $x_1 = y$ and $x_2 = y'$. Rewrite the initial value problem into a system of first-order differential equations in the variables $X = [x_1, x_2]^T$.
Denote by $X_i = [x_{1i}, x_{2i}]^T$ the approximation from a numerical method to $X(t_i)$ with $t_i = t_0 + ih$ for $i = 0, 1, 2, \cdots$. Approximate $X(0.2)$ for this initial value problem, by taking two steps with Euler's method and step size $h = 0.1$.

Continuation 2013, Problem 6a)

Answer:
The system is given by

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_2 \\ -x_2 \sin(x_1) \end{bmatrix}$$

Hence, Euler's method becomes

$$\begin{bmatrix} x_{1,i+1} \\ x_{2,i+1} \end{bmatrix} = \begin{bmatrix} x_{1i} \\ x_{2i} \end{bmatrix} + h \begin{bmatrix} x_{2i} \\ -x_{2i} \sin(x_{1i}) \end{bmatrix}$$

If we compute directly with $h = 0.1$, the first two steps become

$$\begin{aligned} X_1 &= \begin{bmatrix} 1 \\ 2 \end{bmatrix} + 0.1 \begin{bmatrix} 2 \\ -2\sin(1) \end{bmatrix} \\ &= \begin{bmatrix} 1.2 \\ 1.831705803 \end{bmatrix} \\ X_2 &= \begin{bmatrix} 1.2 \\ 1.8317058030 \end{bmatrix} + 0.1 \begin{bmatrix} 1.8317058030 \\ -1.8317058030\sin(1.2) \end{bmatrix} \\ &= \begin{bmatrix} 1.8317058030 \\ 1.5509836627 \end{bmatrix} \end{aligned}$$

Question:

Consider here a general autonomous system of first-order differential equations

$$X' = F(X)$$

Euler's explicit and implicit methods read

$$X_{n+1} = X_n + hF(X_n)$$
$$X_{n+1} = X_n + hF(X_{n+1})$$

We generate a higher order method by combining these two methods in the following way:

1. One step of size $h/2$ with explicit Euler from X_n to $X_{n+1/2}$.
2. One step of size $h/2$ with implicit Euler from $X_{n+1/2}$ to X_{n+1}.

Show that we get a Runge-Kutta method. Write down its Butcher tableau. Determine the order.

Answer:
The two steps of our method are

$$X_{n+1/2} = X_n + \frac{h}{2}F(X_n)$$
$$X_{n+1} = X_{n+1/2} + \frac{h}{2}F(X_{n+1})$$

If we merge these steps, the resulting scheme becomes

$$X_{n+1} = X_n + h\frac{F(X_n) + F(X_{n+1})}{2}$$

By setting $K_1 = F(X_n)$ and, $K_2 = F(X_{n+1})$, we get

$$X_{n+1} = X_n + \frac{h(K_1 + K_2)}{2}$$

The Butcher tableau for the trapezoidal method is

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array}$$

The order conditions read

$$\sum_{i=1}^{2} b_i = \frac{1}{2} + \frac{1}{2} = 1$$

$$\sum_{i=1}^{2} b_i c_i = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 1 = \frac{1}{2}$$

$$\sum_{i=1}^{2} b_i c_i^2 = \frac{1}{2} \cdot 0^2 + \frac{1}{2} \cdot 1^2 = \frac{1}{2} \neq \frac{1}{3}$$

Our method is second order.

Consider the following boundary value problem for the unknown function $u(x)$:

$$u_{xx} - 2u_x = f(x) \quad , \quad 0 < x < 1 \quad , \quad u(0) = 2 \, , \; u(1) = 1$$

where $f(x) = \sin(\pi x)$.

Question:
Construct a finite difference method using central differences to approximate both u_{xx} and u_x with equidistant grid points on $[0, 1]$. In other words, obtain the discretized linear system $A_h U = F$, and specify what A_h, U and F are.

Answer:
If we use central difference approximation with $h = 1/(M+2)$ for $M+1$ intervals, then the 1st and 2nd order derivatives become

$$u_x(x_j) \approx \frac{u(x_{j+1}) - u(x_{j-1})}{2h}$$
$$u_{xx}(x_j) \approx \frac{u(x_{j+1}) - 2u(x_j) + u(x_{j-1})}{h^2}$$

Since $u_{xx} - 2u_x = f(x)$, A_h becomes a tridiagonal matrix, given by

$$A_h = \frac{1}{h^2} tridiag(1+h, -2, 1-h)$$

By invoking boundary conditions, we obtain

$$U = \begin{bmatrix} U_1 \\ \vdots \\ U_M \end{bmatrix} \quad , \quad F = \begin{bmatrix} f_1 - \frac{2}{h^2} - \frac{2}{h} \\ f_2 \\ \vdots \\ f_{M-1} \\ f_M - \frac{1}{h^2} + \frac{1}{h} \end{bmatrix}$$

where U_j approximates $u(x_j)$ and $f_j = f(x_j)$.

Question:
Compute the quantity $\lim_{h\to 0^+}\rho(A_h^{-1})$, where ρ is the spectral radius. You can use the following fact without proof:
For a tridiagonal matrix $tridiag(c, a, b)$ with $bc > 0$, the eigenvalues are given by

$$\lambda_s = a + 2\sqrt{bc}\cos\left(\frac{s\pi}{M+1}\right) \qquad , \qquad s = 1, \dots, M$$

Answer:

In our case $A_h = h^{-2}tridiag(1+h, -2, 1-h)$, and that yields

$$\lambda_s = \frac{-2 + 2\sqrt{1-h^2}\cos(\pi s h)}{h^2} \qquad , \qquad s = 1, \dots, M$$

The spectral radius of A_h^{-1} is found as follows:

$$\rho(A_h) = \max|\lambda_s|$$
$$\rho(A_h^{-1}) = \frac{1}{\min|\lambda_s|} = \frac{1}{|\lambda_1|}$$

The following expansions around $h = 0$ are valid:

$$\sqrt{1-h^2} = 1 - \frac{h^2}{2} + \mathcal{O}(h^4)$$
$$\cos(\pi h) = 1 - \frac{(\pi h)^2}{2} + \mathcal{O}(h^4)$$

We approximate λ_1 as follows:

$$\begin{aligned}
|\lambda_1| &= 2\frac{1 - \left(1 - \frac{h^2}{2} + \mathcal{O}(h^4)\right)\left(1 - \frac{(\pi h)^2}{2} + \mathcal{O}(h^4)\right)}{h^2} \\
&= \pi^2 + 1 + \mathcal{O}(h^2)
\end{aligned}$$

Thus, we have shown that

$$\lim_{h \to 0^+} \rho(A_h^{-1}) = \frac{1}{\pi^2 + 1}$$