Faculty of Electrical Engineering, Mathematics and Computer Science
Numerieke Methoden I, TW2060, BSc Technische Wiskunde
Answers to the exam of August 15, 2018 test

1 a We search α_{-1}, α_0 and α_1 such

$$f''(x)+f'(x) \approx Q(h) = \frac{\alpha_{-1}f(x-h)+\alpha_0 f(x)+\alpha_1 f(x+h)}{h^2}.$$

To this extent, we use Taylor series around x, this is where we aim to approximate the derivative, to obtain

$$\begin{aligned} f''(x)+f'(x) &= \frac{\alpha_{-1}(f(x)-hf'(x)+\frac{h^2}{2}f''(x)-\frac{h^3}{6}f'''(x)+\frac{h^4}{24}f''''(x))}{h^2}+\frac{\alpha_0 f(x)}{h^2}+ \\ &\frac{\alpha_1(f(x)+hf'(x)+\frac{h^2}{2}f''(x)+\frac{h^3}{6}f'''(x)+\frac{h^4}{24}f''''(x))}{h^2}+\mathcal{O}(h^3)= \\ &\frac{\alpha_{-1}+\alpha_0+\alpha_1}{h^2}f(x)+\frac{-\alpha_{-1}+\alpha_1}{h}f'(x)+\frac{1}{2}(\alpha_{-1}+\alpha_1)f''(x)+\frac{h}{6}(-\alpha_{-1}+\alpha_1)f'''(x)+ \\ &\frac{h^2}{24}(\alpha_{-1}+\alpha_1)f''''(x)+\mathcal{O}(h^3). \end{aligned} \tag{1}$$

In the last step, we factorised according to the derivatives of the function $f(x)$. Since we want to determine the three α–values, we also need a 3×3-system of equations. We choose these equations such that the error is as small as possible, which boils down to using the first three (lowest order, which are orders h^{-2}, h^{-1} and 1) terms corresponding to $f(x)$, $f'(x)$ and $f''(x)$ only. Other combinations can give an error of negative order. Therefore, we write the above equation as

$$f''(x)+f'(x) = Q(h)+\frac{h}{6}(-\alpha_{-1}+\alpha_1)f'''(x)+\frac{h^2}{24}(\alpha_{-1}+\alpha_1)f''''(x)+\mathcal{O}(h^3). \tag{2}$$

Here, we have

$$Q(h) = \frac{\alpha_{-1}+\alpha_0+\alpha_1}{h^2}f(x)+\frac{-\alpha_{-1}+\alpha_1}{h}f'(x)+\frac{1}{2}(\alpha_{-1}+\alpha_1)f''(x). \tag{3}$$

Hence, equating the terms corresponding to $f(x)$, $f'(x)$ and $f''(x)$, gives

$$\begin{aligned} &\alpha_{-1}+\alpha_0+\alpha_1 = 0, && \text{terms in front of } f(x), \\ &\frac{-\alpha_{-1}+\alpha_1}{h} = 1, && \text{terms in front of } f'(x), \\ &\frac{1}{2}(\alpha_{-1}+\alpha_1) = 1, && \text{terms in front of } f''(x). \end{aligned} \tag{4}$$

It is easily verified that the solution of this system of equations is given by

$$\alpha_{-1} = 1-\frac{h}{2}, \quad \alpha_0 = -2, \quad \alpha_1 = 1+\frac{h}{2}.$$

b Here we can answer the question in at least two ways. One way, which is more straightforward, is by setting the obtained values for the α-values in the formula for $Q(h)$, see equation (5) and directly use Taylor Series around x.

It is, however, more efficient to combine the results in equations (1) and (4). If we do so, then we see from equation (4) that $\alpha_{-1}+\alpha_1 = h$ and $\alpha_{-1}+\alpha_1 = 2$, and these two expressions are substituted into the equation(1), to get

$$\begin{aligned} f''(x)+f'(x) &= Q(h)+\frac{h}{6}(-\alpha_{-1}+\alpha_1)f'''(x)+\frac{h^2}{24}(\alpha_{-1}+\alpha_1)f''''(x)+\mathscr{O}(h^3) = \\ & Q(h)+\frac{h^2}{6}f'''(x)+\frac{1}{12}h^2f''''(x)+\mathscr{O}(h^3). \end{aligned} \tag{5}$$

Hence

$$\frac{f''(x)+f'(x)-Q(h)}{h^2} = \frac{1}{12}(2f'''(x)+f''''(x))+\mathscr{O}(h),$$

and herewith, it follows that

$$\lim_{h\to 0}\frac{f''(x)+f'(x)-Q(h)}{h^2} = \frac{1}{12}\,(2f'''(x)+f''''(x)) = K.$$

Hence $K = \frac{1}{12}\,(2f'''(x)+f''''(x))$.

c Using the exact (non-available) values of $f(x)$ at x, $x-h$ and $x+h$, we have

$$Q(h) = \frac{f(x-h)-2f(x)+f(x+h)}{h^2}+\frac{f(x+h)-f(x-h)}{2h},$$

and using the available data gives

$$\tilde{Q}(h) = \frac{\tilde{f}(x-h)-2\tilde{f}(x)+\tilde{f}(x+h)}{h^2}+\frac{\tilde{f}(x+h)-\tilde{f}(x-h)}{2h}.$$

We are interested in the magnitude of the total error, that is we want

$$|f'(x)+f''(x)-\tilde{Q}(h)|,$$

This can be processed as follows

$$\begin{aligned} |f'(x)+f''(x)-\tilde{Q}(h)| &= |f'(x)+f''(x)-Q(h)+Q(h)-\tilde{Q}(h)| \\ &\leq |f'(x)+f''(x)-Q(h)|+|Q(h)-\tilde{Q}(h)|. \end{aligned} \tag{6}$$

The inequality is based on the triangular inequality. The first term on the right-hand side represents the truncation error which was given to have the upper bound $|f'(x)+f''(x)-Q(h)| \leq \tilde{K}h^2$. Next, we work on the second term. From the expressions for $Q(h)$ and $\tilde{Q}(h)$, we obtain

$$|Q(h)-\tilde{Q}(h)| =$$

$$|\frac{f(x-h)-2f(x)+f(x+h)}{h^2}+\frac{f(x+h)-f(x-h)}{2h}-(\frac{\tilde{f}(x-h)-2\tilde{f}(x)+\tilde{f}(x+h)}{h^2}+\frac{\tilde{f}(x+h)-\tilde{f}(x-h)}{2h})|$$

$$= |\frac{(f(x-h)-\tilde{f}(x-h))-2(f(x)-\tilde{f}(x))+(f(x+h)-\tilde{f}(x+h))}{h^2}+\frac{(f(x+h)-\tilde{f}(x+h))-(f(x-h)-\tilde{f}(x-h))}{2h}|$$

$$\leq \frac{|f(x-h)-\tilde{f}(x-h)|+2|f(x)-\tilde{f}(x)|+|f(x+h)-\tilde{f}(x+h)|}{h^2}+\frac{|f(x+h)-\tilde{f}(x+h)|+|f(x-h)-\tilde{f}(x-h)|}{2h}$$

$$\leq \frac{\varepsilon+2\varepsilon+\varepsilon}{h^2}+\frac{\varepsilon+\varepsilon}{2h} = \frac{4\varepsilon}{h^2}+\frac{\varepsilon}{h}.$$

Substituting this contribution into inequality (6), gives the following final result

$$|f'(x)+f''(x)-\tilde{Q}(h)| \leq \tilde{K}h^2+\frac{4\varepsilon}{h^2}+\frac{\varepsilon}{h}.$$

d It can be seen that the error contains a h^{-2} (and h^{-1}) term. If h is approaches zero, then h^{-2} gets arbitrarily large, whereas the truncation error, which is of order h^2 approaches zero, and hence the error due to rounding will dominate if h is too small. This is undesirable.

2 We consider interpolation and numerical integration

a We are looking for a linear relation such that the line contains the points $(x_i, f(x_i))$ and $(x_{i+1}, f(x_{i+1}))$. A linear relation reads as

$$L(x) = c_0 + c_1 x,$$

and we determine c_0 and c_1 such that $L(x_i) = f(x_i)$ and $L(x_{i+1}) = f(x_{i+1})$, this gives

$$c_0 + c_1 x_i = f(x_i),$$

$$c_0 + c_1 x_{i+1} = f(x_{i+1}).$$

This gives $c_0 = f(x_i) - \frac{x_i\ (f(x_{i+1}) - f(x_i))}{x_{i+1} - x_i}$ and $c_1 = \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}$, hence

$$L(x) = f(x_i) + \frac{f(x_{i+1}) - f(x_i)}{x_{i+1} - x_i}(x - x_i) = f(x_i)\frac{x - x_{i+1}}{x_i - x_{i+1}} + f(x_{i+1})\frac{x - x_i}{x_{i+1} - x_i}.$$

b We derive the Trapezoidal Rule by integration of the interpolation polynomial, this gives

$$\int_{x_i}^{x_{i+1}} f(x)dx \approx \int_{x_i}^{x_{i+1}} L(x)dx = \int_{x_i}^{x_{i+1}} f(x_i)\frac{x - x_{i+1}}{x_i - x_{i+1}} + f(x_{i+1})\frac{x - x_i}{x_{i+1} - x_i}dx =$$

$$\frac{f(x_i)}{x_i - x_{i+1}}\int_{x_i}^{x_{i+1}} (x - x_{i+1})dx + \frac{f(x_{i+1})}{x_{i+1} - x_i}\int_{x_i}^{x_{i+1}} (x - x_i)dx = \frac{1}{2}(x_{i+1} - x_i)(f(x_i) + f(x_{i+1})).$$

c The magnitude of the error is given by

$$|\int_{x_i}^{x_{i+1}} f(x) - L(x)dx| \leq \int_{x_i}^{x_{i+1}} |f(x) - L(x)|dx \leq \int_{x_i}^{x_{i+1}} M(x - x_i)(x_{i+1} - x)dx.$$

The first upper bound follows from the triangular inequality, and the second upper bound follows from the assertion that $|f''(x)| \leq M$ on (x_i, x_{i+1}). The above integral in the upper bound can be evaluated by integration by parts or using the substitution rule. This gives

$$\int_{x_i}^{x_{i+1}} (x - x_i)(x_{i+1} - x)dx = \frac{1}{12}(x_{i+1} - x_i)^3.$$

Combination of these expressions gives

$$|\int_{x_i}^{x_{i+1}} f(x) - L(x)dx| \leq \int_{x_i}^{x_{i+1}} M(x - x_i)(x_{i+1} - x)dx = \frac{M}{12}(x_{i+1} - x_i)^3.$$

Hence $\alpha = \frac{1}{12}$.

d i Let I_T be the approximation of $I = \int_a^b f(x)dx$, then

$$I = \int_a^b f(x)dx = \sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(x)dx \approx \sum_{i=0}^{n-1} \frac{h}{2}(f(x_i) + f(x_{i+1})) =$$

$$\frac{h}{2}(f(a) + f(x_1) + f(x_1) + f(x_2) + \ldots + f(x_{n+1}) + f(b)) = h(\frac{f(a)}{2} + f(x_1) + \ldots + f(x_{n-1}) + \frac{f(b)}{2}) = I_T.$$

ii The magnitude of the global error is given by

$$|\int_a^b f(x)dx - I_T| = |\sum_{i=0}^{n-1} \int_{x_i}^{x_{i+1}} f(x)dx - \frac{h}{2}(f(x_i)+f(x_{i+1}))| \leq$$

$$\sum_{i=0}^{n-1} |\int_{x_i}^{x_{i+1}} f(x)dx - \frac{h}{2}(f(x_i)+f(x_{i+1}))| \leq \sum_{i=0}^{n-1} \alpha M_i h^3 = \sum_{i=0}^{n-1} \frac{M_i}{12} h^3.$$

The last inequality follows from the result in assignment 2c and $M_i \geq |f''(x)|$ on (x_i, x_{i+1}). Let $|f''(x)| \leq M$ on (a,b), then $M_i \leq M$ for $i = 0, \ldots, n-1$, then

$$|\int_a^b f(x)dx - I_T| \leq \sum_{i=0}^{n-1} \alpha M_i h^3 \leq \sum_{i=0}^{n-1} \alpha M h^3 = n\alpha M h^3 = nh\alpha M h^2 = (b-a)\alpha M h^2 = (b-a)\frac{M}{12}h^2.$$

3 We solve the one-dimensional heat equation for $y = y(x,t)$

$$\begin{cases} \dfrac{\partial y}{\partial t} = \dfrac{\partial^2 y}{\partial x^2}, & x \in (0,1),\ t > 0, \\ y(0,t) = 0 = y(1,t), & t > 0, \\ y(x,0) = y_0(x), & x \in (0,1). \end{cases} \tag{7}$$

a For the second spatial derivative, we get

$$\frac{\partial^2 y(x_i,t)}{\partial x^2} = \frac{y_{i-1}(t) - 2y_i(t) + y_{i+1}(t)}{\Delta x^2} + \mathscr{O}(\Delta x^2), \tag{8}$$

where $y_j(t) = y(x_i,t)$. Using the heat equation, neglecting the error and replacing y_i with w_i gives

$$\frac{dw_i}{dt} = \frac{w_{i-1}(t) - 2w_i(t) + w_{i+1}(t)}{\Delta x^2}, \text{ for } i = 1, \ldots, n. \tag{9}$$

Since we have $y_0(t) = y_{n+1}(t) = 0$, we obtain for $i = 1$

$$\frac{dw_1}{dt} = \frac{-2w_1(t) + w_2(t)}{\Delta x^2}, \tag{10}$$

and for $i = n$

$$\frac{dw_n}{dt} = \frac{w_{n-1}(t) - 2w_n(t)}{\Delta x^2}. \tag{11}$$

Herewith, we have $\mathbf{w}' = A\mathbf{w}$, where

$$\begin{aligned} a_{ii-1} &= \tfrac{1}{\Delta x^2}, \text{ for } i = 2, \ldots, n, \\ a_{ii} &= -\tfrac{2}{\Delta x^2}, \text{ for } i = 1, \ldots, n, \\ a_{ii+1} &= \tfrac{1}{\Delta x^2}, \text{ for } i = 1, \ldots, n-1. \end{aligned} \tag{12}$$

b Gershgorin's Theorem says: *Given an $n \times n$-matrix $A \in mat_{n\times n}(\mathbb{C})$, and let $\lambda(A) = \{\lambda_1, \ldots, \lambda_n\}$, where λ_i is an eigenvalue of A, then*

$$\lambda(A) \subset \bigcup_{i=1}^{n} \{\lambda \in \mathbb{C} : |\lambda - a_{ii}| \leq \sum_{j \in \{1,\ldots,n\}\setminus\{i\}} |a_{ij}|\}. \tag{13}$$

(Note that this theorem is also valid for unsymmetric/non-hermitian matrices).

First of all, symmetry of A implies that the eigenvalues of A are real-valued. Application of Gershgorin's Theorem to the first and last row of the matrix A gives:

$$|\lambda + \frac{2}{\Delta x^2}| \leq |-\frac{1}{\Delta x^2}| = \frac{1}{\Delta x^2}. \tag{14}$$

Since A has real eigenvalues only, this implies

$$-\frac{1}{\Delta x^2} \leq \lambda + \frac{2}{\Delta x^2} \leq \frac{1}{\Delta x^2}. \tag{15}$$

This implies

$$-\frac{3}{\Delta x^2} \leq \lambda \leq -\frac{1}{\Delta x^2}. \tag{16}$$

For the remaining rows, we get

$$|\lambda + \frac{2}{\Delta x^2}| \leq |-\frac{1}{\Delta x^2}| + |-\frac{1}{\Delta x^2}| = \frac{2}{\Delta x^2}. \tag{17}$$

Since A has real eigenvalues only, this implies

$$-\frac{2}{\Delta x^2} \leq \lambda + \frac{2}{\Delta x^2} \leq \frac{2}{\Delta x^2}. \tag{18}$$

This implies

$$-\frac{4}{\Delta x^2} \leq \lambda \leq 0. \tag{19}$$

Since Gershgorin gives the union of the discs in the complex plane for the location of the eigenvalues, and since the bounds from the first and last rows of the matrix are contained within the other rows, it follows directly that the eigenvalues of A are between the following bounds:

$$-\frac{4}{\Delta x^2} \leq \lambda \leq 0. \tag{20}$$

The highest lower bound of the eigenvalues of A is given by $-\frac{4}{\Delta x^2}$ and the lowest upper bound of the eigenvalues of A is given by 0.

c We apply Euler's forward method to

$$\mathbf{w}' = A\mathbf{w}. \tag{21}$$

It was given that the eigenvalues of the matrix A, range in $[-\frac{4}{\Delta x^2}, 0]$, that is

$$\lambda(A) \subset [-\frac{4}{\Delta x^2}, 0].$$

For the sake of stability, we need the amplification factor, which is derived by the use of the test equation, which reads as $y' = \lambda y$. For Euler's forward method, we get

$$w_{n+1} = w_n + \Delta t \lambda w_n = (1 + \Delta t \lambda) w_n, \tag{22}$$

and hence the amplification factor is given by

$$Q(\Delta t \lambda) = 1 + \Delta t \lambda. \tag{23}$$

Stability of the numerical method requires that

$$|Q(\Delta t \lambda)| \leq 1, \text{ for all eigenvalues in } \lambda(A). \tag{24}$$

Hence, for the current system, in which the eigenvalues are real-valued, we get

$$Q(\Delta t \lambda) \in [1 - \Delta t \frac{4}{\Delta x^2}, 1], \tag{25}$$

which gives the following condition for numerical stability

$$1 - \frac{4\Delta t}{\Delta x^2} \geq -1. \tag{26}$$

Processing the above inequality further, gives

$$\Delta t \leq \frac{\Delta x^2}{2}. \tag{27}$$

Note that since $\lambda \in \mathbb{R}$, it is also allowed to use $\Delta t \leq \frac{2}{-\lambda}$ directly.

d The local truncation error is defined by

$$\tau_{n+1}(\Delta t) := \frac{y_{n+1} - z_{n+1}}{\Delta t}, \tag{28}$$

where z_{n+1} is obtained from using the exact solution $y_n = y(t_n)$ at time t_n. For the exact solution, we have

$$y_{n+1} = y(t_n + \Delta t) = y_n + \Delta t y'(t_n) + \frac{\Delta t^2}{2} y''(\zeta), \text{ with } \zeta \in (t_n, t_{n+1}). \tag{29}$$

Using $y'(t) = f(t,y)$ gives $y'(t_n) = f(t_n, y_n)$ and hence

$$y_{n+1} = y(t_n + \Delta t) = y_n + \Delta t f(t_n, y_n) + \frac{\Delta t^2}{2} y''(\zeta), \text{ with } \zeta \in (t_n, t_{n+1}). \tag{30}$$

Since Euler's forward method applied using the exact solution at the previous time, gives the following expression for z_{n+1}

$$z_{n+1} = y_n + \Delta t f(t_n, y_n), \tag{31}$$

we get, using equation (1),

$$\tau_{n+1}(\Delta t) = \frac{y_n + \Delta t f(t_n, y_n) + \frac{\Delta t^2}{2} y''(\zeta) - (y_n + \Delta t f(t_n, y_n))}{\Delta t} = \frac{\Delta t}{2} y''(\zeta) = \mathscr{O}(\Delta t). \tag{32}$$

One may relate $y''(t)$ to f by $y''(t) = \frac{dy'(t)}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial y} y'(t) = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial y} f(t,y)$ (but this is not necessary).

e Lax Equivalence Theorem states: *A stable, consistent scheme gives a converging numerical solution.* Consistency of the scheme means that $\lim_{\Delta t \to 0} \tau_{n+1}(\Delta t) = 0$. In assignment 3d, we proved that $\tau_{n+1}(\Delta t) = \mathscr{O}(\Delta t)$ and hence $\lim_{\Delta t \to 0} \tau_{n+1}(\Delta t) = 0$ and herewith the scheme is consistent (the local truncation error tends to zero as the time-step tends to zero).

We also demonstrated that the scheme is stable if $\Delta t \leq \frac{\Delta x^2}{2}$. Hence the numerical solution converges if $\Delta t \leq \frac{\Delta x^2}{2}$.