DELFT UNIVERSITY OF TECHNOLOGY
FACULTY OF ELECTRICAL ENGINEERING, MATHEMATICS AND COMPUTER SCIENCE

ANSWERS OF THE TEST NUMERICAL METHODS FOR DIFFERENTIAL EQUATIONS
(CTB2400)
Tuesday July 18 2023, 13:30-16:30

1. (a) The local truncation error is given by

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

where z_{n+1} is obtained by one step of the method starting in t_n and y_n. We determine y_{n+1} by the use of Taylor expansions around t_n:

$$y_{n+1} = y_n + \Delta t y'(t_n) + \frac{\Delta t^2}{2} y''(t_n) + \mathcal{O}(\Delta t^3). \tag{2}$$

We bear in mind that

$$y'(t_n) = f(t_n, y_n)$$

$$\begin{aligned} y''(t_n) &= \frac{df(t_n, y_n)}{dt} = \frac{\partial f(t_n, y_n)}{\partial t} + \frac{\partial f(t_n, y_n)}{\partial y} y'(t_n) \\ &= \frac{\partial f(t_n, y_n)}{\partial t} + \frac{\partial f(t_n, y_n)}{\partial y} f(t_n, y_n). \end{aligned}$$

Hence

$$y_{n+1} = y_n + \Delta t y'(t_n) + \frac{\Delta t^2}{2} \left(\frac{\partial f(t_n, y_n)}{\partial t} + \frac{\partial f(t_n, y_n)}{\partial y} f(t_n, y_n) \right) + \mathcal{O}(\Delta t^3). \tag{3}$$

After substitution of $k_1 = f(t_n, y_n)$ and $k_2 = f(t_{n+1}, y_n + \Delta t k_1)$ into $w_{n+1} = y_n + \frac{\Delta t}{2}(k_1 + k_2)$, and after using a Taylor expansion around (t_n, y_n), we obtain for z_{n+1}:

$$\begin{aligned} z_{n+1} &= y_n + \frac{\Delta t}{2} \left(f(t_n, y_n) + f(t_n + \Delta t, y_n + \Delta t f(t_n, y_n)) \right) \\ &= y_n + \frac{\Delta t}{2} \left(2 f(t_n, y_n) + \Delta t \left(\frac{\partial f(t_n, y_n)}{\partial t} + f(t_n, y_n) \frac{\partial f(t_n, y_n)}{\partial y} \right) + \mathcal{O}(\Delta t^2) \right). \end{aligned}$$

Herewith, one obtains

$$y_{n+1} - z_{n+1} = \mathcal{O}(\Delta t^3), \text{ and hence } \tau_{n+1}(\Delta t) = \frac{\mathcal{O}(\Delta t^3)}{\Delta t} = \mathcal{O}(\Delta t^2). \tag{4}$$

(b) *Note: Every miscalculation in the calculation of $\underline{k}_1$ and $\underline{k}_2$ gives a subtraction of 1/4 point, with at most 1/2 point being subtracted.*

Note: The calculation of $\underline{w}_1$ must be consistent with the value for $\underline{k}_1$ and $\underline{k}_2$. If not, 1 point is subtracted.

Note: Every miscalculation in the calculation of $\underline{w}_1$ gives a subtraction of 1/4 point, with at most 1 point being subtracted.

Application of the integration method to the system $\underline{x}' = A\underline{x} + \underline{f}$, gives

$$\begin{aligned} \underline{k}_1 &= A\underline{w}_0 + \underline{f}_0, \\ \underline{k}_2 &= A(\underline{w}_0 + \Delta t \underline{k}_1) + \underline{f}_1 \\ \underline{w}_1 &= \underline{w}_0 + \tfrac{\Delta t}{2}\left(\underline{k}_1 + \underline{k}_2\right). \end{aligned} \tag{5}$$

With the initial condition $\underline{w}_0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ and $\Delta t = 0.5$, this gives the following result

$$\underline{k}_1 = \begin{pmatrix} -2 & 1 \\ 0 & -4 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} + \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ -7 \end{pmatrix}. \tag{6}$$

$$\underline{k}_2 = \begin{pmatrix} -2 & 1 \\ 0 & -4 \end{pmatrix} \left(\begin{pmatrix} 1 \\ 2 \end{pmatrix} + 0.5 * \begin{pmatrix} 0 \\ -7 \end{pmatrix} \right) + \begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -3.5 \\ 6 \end{pmatrix}. \tag{7}$$

The final result is calculated as follows

$$\begin{aligned} \underline{w}_1 &= \begin{pmatrix} 1 \\ 2 \end{pmatrix} + 0.25\left(\underline{k}_1 + \underline{k}_2\right) \\ &= \begin{pmatrix} 0.125 \\ 1.75 \end{pmatrix} \end{aligned}$$

(c) Consider the test equation $y' = \lambda y$, then one gets

$$\begin{aligned} k_1 &= \lambda w_n, \\ k_2 &= \lambda(w_n + \Delta t \lambda w_n), \\ w_{n+1} &= w_n + \frac{\Delta t}{2}(k_1 + k_2) \\ &= \left(1 + \Delta t \lambda + \frac{(\Delta t \lambda)^2}{2}\right) w_n. \end{aligned}$$

Hence the amplification factor is given by

$$Q(\lambda \Delta t) = 1 + \lambda \Delta t + \frac{(\lambda \Delta t)^2}{2}. \tag{8}$$

(d) *Note: Every miscalculation in the calculation of $|Q(\lambda_1 \Delta t)|^2$ gives a subtraction of 1/4 point, with at most 1/2 point being subtracted.*

Note: The calculation of $|Q(\lambda_1 \Delta t)|^2$ must be consistent with the eigenvalues found. If not, 1/2 point is subtracted.

First, we determine the eigenvalues of the matrix A. Subsequently, the eigenvalues are substituted into the amplification factor.

The eigenvalues of the matrix A are given by $\lambda_1 = -4$ and $\lambda_2 = -2$.

Since $\lambda_1 = -4$ is the smallest eigenvalue it is sufficient to check if $|Q(\lambda_1 \Delta t)| \leq 1$. Since $Q(\lambda_1 \Delta t) = 1 + \lambda_1 \Delta t + \frac{1}{2}(\lambda_1 \Delta t)^2$ we have to check that $|1 - 4\Delta t + 8(\Delta t)^2| \leq 1$. This leads to

$$-1 \leq 1 - 4\Delta t + 8(\Delta t)^2 \leq 1.$$

We start with the left inequality:

$$-1 \leq 1 - 4\Delta t + 8(\Delta t)^2$$

This can be written as

$$0 \leq 2 - 4\Delta t + 8(\Delta t)^2$$

This is a second order polynomial. Since the discriminant $(-4)^2 - 4 \times 2 \times 8$ is negative there are no real roots. The inequality holds for $\Delta t = 0$ so it holds for all Δt-values. For the right inequality we have:

$$1 - 4\Delta t + 8(\Delta t)^2 \leq 1.$$

This is equivalent to

$$-4\Delta t + 8(\Delta t)^2 \leq 0.$$

Dividing

$$8(\Delta t)^2 \leq 4\Delta t$$

by $8\Delta t$ leads to

$$\Delta t \leq \frac{1}{2}.$$

So the method is stable for all $\Delta t \leq \frac{1}{2}$.

(e) The largest advantage is that the stability condition for equation $x_1' = -2x_1 + x_2$ is $\Delta t \leq 1$. So for this equation the time step can be chosen two times a large than for the complete system. This means that less work is needed for this approach.

2. (a) Taylor polynomials are:

$$\begin{aligned} d(0) &= d(2h) - 2hd'(2h) + 2h^2 d''(2h) - \frac{(2h)^3}{6} d'''(\xi_0)\ , \\ d(h) &= d(2h) - hd'(2h) + \frac{h^2}{2} d''(2h) - \frac{h^3}{6} d'''(\xi_1)\ , \\ d(2h) &= d(2h). \end{aligned}$$

We know that $Q(h) = \frac{\alpha_0}{h^2} d(0) + \frac{\alpha_1}{h^2} d(h) + \frac{\alpha_2}{h^2} d(2h)$, which should be equal to $d''(2h)$ + remainder term. This leads to the following conditions:

$$\begin{array}{rcrcrcl} \frac{\alpha_0}{h^2} & + & \frac{\alpha_1}{h^2} & + & \frac{\alpha_2}{h^2} & = & 0\ , \\ -2\frac{\alpha_0}{h} & - & \frac{\alpha_1}{h} & & & = & 0\ , \\ 2\alpha_0 & + & \frac{1}{2}\alpha_1 & & & = & 1\ . \end{array}$$

(b) The truncation error follows from the Taylor polynomials:

$$d''(2h) - Q(h) = d''(2h) - \frac{d(0) - 2d(h) + d(2h)}{h^2} = \frac{\frac{8h^3}{6} d'''(\xi_0) - 2(\frac{h^3}{6} d'''(\xi_1))}{h^2} = hd'''(\xi).$$

(c) Using the formula with $h = 10$ we obtain the estimate:

$$\frac{d(0) - 2d(10) + d(20)}{100} = \frac{0 - 2 \times 40 + 100}{100} = 0.2\ (\mathrm{m/s^2}).$$

3. (a) A fixed point p satisfies the equation $p = g(p)$. Substitution gives: $p = \frac{p^3}{6} + \frac{23}{48}$. Rewriting this expression gives:

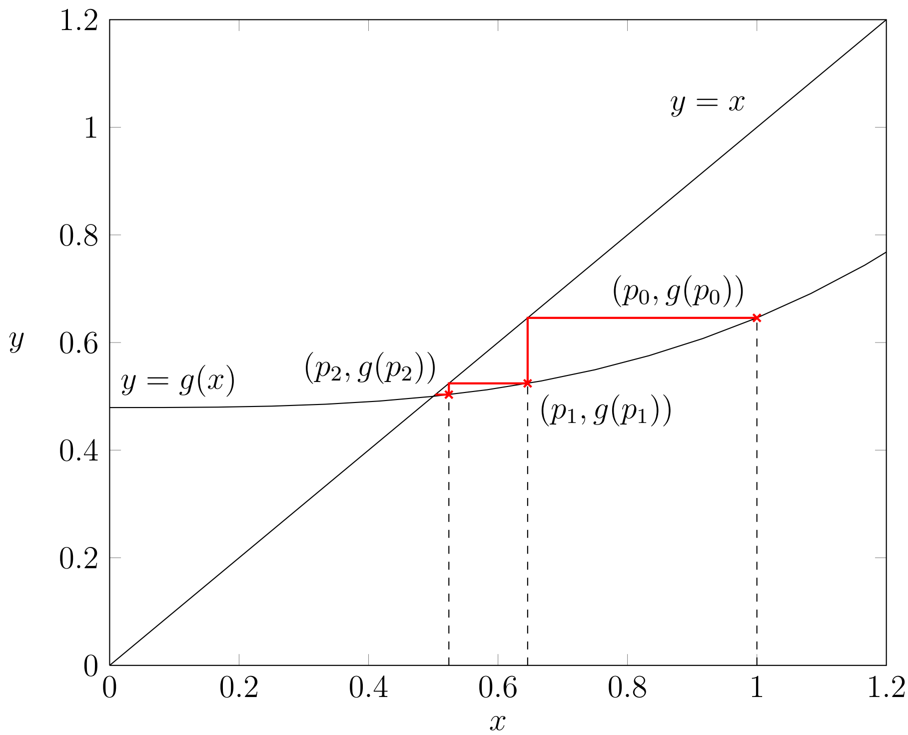
$$\begin{aligned} -\tfrac{p^3}{6} + p - \tfrac{23}{48} &= 0 \\ \Rightarrow \quad -p^3 + 6p - \tfrac{23}{8} &= 0 \\ \Rightarrow \quad f(p) &= 0, \end{aligned}$$

which shows that a fixed point of $g(x)$ also a root of $f(x)$ is.

(b) Starting with $p_0 = 1$ we obtain:

$$\begin{aligned} p_1 &= \approx 0.6458, \\ p_2 &= \approx 0.5241, \\ p_3 &= \approx 0.5032. \end{aligned}$$

A sketch of this fixed-point iteration is given by



(c) For the convergence three conditions should be satisfied:

- $g \in C[0, 1]$.
- $g(p) \in [0, 1]$ for all $p \in [0, 1]$.
- $|g'(p)| \leq k < 1$ for all $p \in [0, 1]$.

Since $g(p) = \frac{p^3}{6} + \frac{23}{48}$ it follows that g is continuous everywhere, so the first condition holds.

Furthermore, $g'(x) = \frac{x^2}{2}$. Note that $g'(p) \geq 0$ for all $p \in [0, 1]$. This implies that $g(x)$ is increasing on $[0, 1]$. A lower bound for $g(x)$ is given by

$$g(x) \geq g(0) = \frac{23}{48} \geq 0,$$

and an upper bound is given by

$$g(x) \leq g(1) = \frac{31}{48} \leq 1.$$

So $0 \leq g(x) \leq 1$ and the second conditions holds.
For the third condition we note that $|g'(x)| = \frac{x^2}{2} \leq \frac{1}{2} = k < 1$ for all $x \in [0,1]$, so the third condition is also satisfied.
As all conditions are satisfied, the fixed point iteration converges for all $p_0 \in [0,1]$ to the fixed point $p \in [0,1]$.