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

**ANSWERS OF THE TEST NUMERICAL METHODS FOR DIFFERENTIAL EQUATIONS ( WI3097 TU / Minor AESB2210 )**
**Thursday January 28 2016, 18:30-21:30**

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

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

where

$$z_{n+1} = y_n + \Delta t f(t_n, y_n), \tag{2}$$

for the Forward Euler method. A Taylor expansion for $y_{n+1}$ around $t_n$ is given by

$$y_{n+1} = y_n + \Delta t y'(t_n) + \frac{\Delta t^2}{2} y''(\xi), \quad \exists\, \xi \in (t_n, t_{n+1}). \tag{3}$$

Since $y'(t_n) = f(t_n, y_n)$, we use equation (1), to get

$$\tau_h = \frac{\Delta t}{2} y''(\xi), \quad \exists\, \xi \in (t_n, t_{n+1}). \tag{4}$$

Hence, the truncation error is of first order.

(b) We define $y_1 := y$ and $y_2 := y'$, hence $y_1' = y_2$. Further, we use the differential equation to obtain

$$y'' + \varepsilon y' + y = y_1'' + \varepsilon y_1' + y_1 = y_2' + \varepsilon y_2 + y_1. \tag{5}$$

Hence, we obtain

$$y_2' = -y_1 - \varepsilon y_2 + \sin(t). \tag{6}$$

Hence the system is given by

$$\begin{aligned} y_1' &= y_2, \\ y_2' &= -y_1 - \varepsilon y_2 + \sin(t). \end{aligned} \tag{7}$$

The initial conditions are given by

$$\begin{aligned} 1 &= y(0) = y_1(0), \\ 0 &= y'(0) = y_1'(0) = y_2(0). \end{aligned} \tag{8}$$

(c) First, we use the test equation, $y' = \lambda y$, to analyse numerical stability. For the Forward Euler method, we obtain

$$w_{n+1} = w_n + \Delta t \lambda w_n = Q(\lambda \Delta t) w_n, \tag{9}$$

hence the amplification factor becomes

$$Q(\lambda \Delta t) = 1 + \lambda \Delta t. \tag{10}$$

The numerical solution is stable if and only if $|Q(\lambda \Delta t)| \leq 1$.

Next, we deal with the case $\varepsilon = 0$, to obtain the following system

$$\begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}. \tag{11}$$

This system gives the following eigenvalues $\lambda_{1,2} = \pm i$, where $i$ is the imaginary unit. Hence, the amplification factor is given by

$$Q(\lambda \Delta t) = 1 \pm i \Delta t. \tag{12}$$

Then, it is immediately clear that $|Q(\lambda \Delta t)| > 1$ for all $\Delta t > 0$ since

$$|1 \pm i| = \sqrt{1^2 + (\Delta t)^2}| > 1. \tag{13}$$

Hence, we conclude that the forward Euler method is never stable if $\varepsilon = 0$.

(d) From part (c) we know that if $\varepsilon = 0$, the eigenvalues of the system are purely imaginary. This implies that the system is analytically (zero) stable if $\varepsilon = 0$.

Non-zero values of $\varepsilon$ give the following system

$$\begin{pmatrix} y_1' \\ y_2' \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & \varepsilon \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}. \tag{14}$$

then we get the following eigenvalues $\lambda_{1,2} = \frac{\varepsilon}{2} \pm \frac{1}{2}\sqrt{\varepsilon^2 - 4}$ (real-valued), if $\varepsilon^2 - 4 \geq 0$ and $\lambda = \frac{\varepsilon}{2} \pm \frac{i}{2}\sqrt{4 - \varepsilon^2}$ (nonreal-valued) if $\varepsilon^2 - 4 < 0$. Hence, we consider two cases: real-valued and nonreal-valued eigenvalues.

*Real-valued eigenvalues*

In this case $|\varepsilon| \geq 2$, and $0 \leq \varepsilon^2 - 4 < \varepsilon^2$, and hence the real-valued eigenvalues have the same sign, which is determined by the sign of $\varepsilon$. Hence, if $\varepsilon \leq -2$, then, the system is stable. Furthermore, if $\varepsilon \geq 2$, then, the system is unstable.

*Nonreal-valued eigenvalues*

In this case $|\varepsilon| < 2$. The system is analytically unstable if and only if the real part of the eigenvalues is positive. Further, the real part of the eigenvalues is positive if and only if $\varepsilon > 0$. Hence, the system is analytically unstable if and only if $\varepsilon > 0$. Hence, the system is stable if and only if $(-2 <)\varepsilon \leq 0$.

From these arguments, it follows that the system is stable if and only if $\varepsilon \leq 0$.

(e) Since currently the discriminant, $\varepsilon^2 - 4$, is negative, the eigenvalues are nonreal. Substitution into the amplification factor yields

$$Q(\lambda\Delta t) = 1 + \frac{\varepsilon}{2}\Delta t \pm \frac{i\Delta t}{2}\sqrt{4-\varepsilon^2}. \tag{15}$$

Hence, numerical stability is warranted if

$$|Q(\lambda\Delta t)|^2 = (1 + \frac{\varepsilon}{2}\Delta t)^2 + \frac{\Delta t^2}{4}(4-\varepsilon^2) \leq 1. \tag{16}$$

Hence for stability, we have

$$1 + \varepsilon\Delta t + \frac{\varepsilon^2\Delta t^2}{4} + \Delta t^2 - \frac{\varepsilon^2\Delta t^2}{4} = 1 + \Delta t\varepsilon + \Delta t^2 \leq 1. \tag{17}$$

Since $\Delta t > 0$, we obtain the following stability criterion

$$\Delta t \leq -\varepsilon = |\varepsilon|. \tag{18}$$

If $\varepsilon = -2$, then both eigenvalues are real-valued and given by $\lambda_{1,2} = -1$. For this case, we obtain $Q(\lambda\Delta t) = 1 - \Delta t$, and stability is warranted if and only if $-1 \leq Q(\lambda\Delta t) \leq 1$, hence $\Delta t \leq 2(= |\varepsilon|$.

We conclude that for $-2 \leq \varepsilon < 0$, we have a numerically stable solution if and only if $\Delta t \leq |\varepsilon|$.

2. (a) Using central differences for the second order derivative at a node $x_j = j\Delta x$ gives

$$y''(x_j) \approx \frac{y_{j+1} - 2y_j + y_{j-1}}{\Delta x^2} =: Q(\Delta x). \tag{19}$$

Here, $y_j := y(x_j)$. Next, we will prove that this approximation is second order accurate, that is $|y''(x_j) - Q(\Delta x)| = \mathcal{O}(\Delta x^2)$.
Using Taylor's Theorem around $x = x_j$ gives

$$\begin{aligned} y_{j+1} &= y(x_j + \Delta x) = y(x_j) + \Delta x y'(x_j) + \tfrac{\Delta x^2}{2}y''(x_j) + \tfrac{\Delta x^3}{3!}y'''(x_j) + \tfrac{\Delta x^4}{4!}y''''(\eta_+), \\ y_{j-1} &= y(x_j - \Delta x) = y(x_j) - \Delta x y'(x_j) + \tfrac{\Delta x^2}{2}y''(x_j) - \tfrac{\Delta x^3}{3!}y'''(x_j) + \tfrac{\Delta x^4}{4!}y''''(\eta_-). \end{aligned} \tag{20}$$

Here, $\eta_+$ and $\eta_-$ are numbers within the intervals $(x_j, x_{j+1})$ and $(x_{j-1}, x_j)$, respectively. Substitution of these expressions into $Q(\Delta x)$ gives

$$|y''(x_j) - Q(\Delta x)| = \mathcal{O}(\Delta x^2).$$

This leads to the following discretisation formula for internal grid nodes:

$$\frac{-w_{j-1} + 2w_j - w_{j+1}}{\Delta x^2} + (x_j + 1)w_j = x_j^3 + x_j^2 - 2. \tag{21}$$

Here, $w_j$ represents the numerical approximation of the solution $y_j$. To deal with the boundary $x = 0$, we use a virtual node at $x = -\Delta x$, and we define $y_{-1} := y(-\Delta x)$. Then, using central differences at $x = 0$ gives

$$0 = y'(0) \approx \frac{y_1 - y_{-1}}{2\Delta x} =: Q_b(\Delta x). \tag{22}$$

Using Taylor's Theorem, gives

$$\begin{aligned} Q_b(\Delta x) &= \\ &= \frac{y(0) + \Delta x y'(0) + \frac{\Delta x^2}{2} y''(0) + \frac{\Delta x^3}{3!} y'''(\eta_+)}{2\Delta x} \\ &- \frac{y(0) - \Delta x y'(0) + \frac{\Delta x^2}{2} y''(0) - \frac{\Delta x^3}{3!} y'''(\eta_-)}{2\Delta x} \\ &= y'(0) + \mathcal{O}(\Delta x^2). \end{aligned}$$

Again, we get an error of $\mathcal{O}(\Delta x^2)$.

(b) With respect to the numerical approximation at the virtual node, we get

$$\frac{w_1 - w_{-1}}{2\Delta x} = 0 \quad \Leftrightarrow \quad w_{-1} = w_1. \tag{23}$$

The discretisation at $x = 0$ is given by

$$\frac{-w_{-1} + 2w_0 - w_1}{\Delta x^2} + w_0 = -2. \tag{24}$$

Substitution of equation (23) into the above equation, yields

$$\frac{2w_0 - 2w_1}{\Delta x^2} + w_0 = -2. \tag{25}$$

Subsequently, we consider the boundary $x = 1$. To this extent, we consider its neighbouring point $x_{n-1}$ and substitute the boundary condition $w_n = y(1) = y_n = 1$ into equation (21) to obtain

$$\begin{aligned} &\frac{-w_{n-2} + 2w_{n-1}}{\Delta x^2} + (x_{n-1} + 1)w_{n-1} && (26) \\ = \ & x_{n-1}^3 + x_{n-1}^2 - 2 + \frac{1}{\Delta x^2} && (27) \\ = \ & (1 - \Delta x)^3 + (1 - \Delta x)^2 - 2 + \frac{1}{\Delta x^2}. && (28) \end{aligned}$$

This concludes our discretisation of the boundary conditions. In order to get a symmetric discretisation matrix, one divides equation (25) by 2.

Next, we use $\Delta x = 1/3$. From equations (21, 25, 28) we obtain the following system

$$\begin{aligned} 9\frac{1}{2}w_0 - 9w_1 &= -1 \\ -9w_0 + 19\frac{1}{3}w_1 - 9w_2 &= -\frac{50}{27} \\ -9w_1 + 19\frac{2}{3}w_2 &= \frac{209}{27}. \end{aligned}$$

(c) The Gershgorin circle theorem states that the eigenvalues of a square matrix $\mathbf{A}$ are located in the complex plane in the union of circles

$$|z - a_{ii}| \le \sum_{\substack{j \neq i \\ j=1}}^{n} |a_{ij}| \quad \text{where} \quad z \in \mathbb{C} \tag{29}$$

For the $3 \times 3$ matrix derived in part (b) we have

- For $i = 1$:
$$\left| z - 9\frac{1}{2} \right| \le 9 \quad \Rightarrow \quad |\lambda_1|_{\min} \ge \frac{1}{2} \tag{30}$$
- For $i = 2$:
$$\left| z - 19\frac{1}{3} \right| \le 18 \quad \Rightarrow \quad |\lambda_2|_{\min} \ge 1\frac{1}{3} \tag{31}$$
- For $i = 3$:
$$\left| z - 19\frac{2}{3} \right| \le 9 \quad \Rightarrow \quad |\lambda_3|_{\min} \ge 10\frac{2}{3} \tag{32}$$

Hence, a lower bound for the smallest eigenvalue is $\frac{1}{2}$. For a symmetric matrix $\mathbf{A}$ we have

$$\|\mathbf{A}^{-1}\| = \frac{1}{|\lambda|_{\min}} \le 2 \tag{33}$$

This proves that the finite-difference scheme is stable, e.g., with constant $C = 2$.

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:

$$\begin{aligned} -\frac{p^3}{6} + p - \frac{23}{48} &= 0 \\ -p^3 + 6p - \frac{23}{8} &= 0 \\ -p^3 + 6p - 2\frac{7}{8} &= 0 \\ f(p) &= 0 \end{aligned}$$

The fixed point iteration is defined by: $p_{i+1} = g(p_i)$. Starting with $p_0 = 1$ one obtains:

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

(b) The fixed point iteration is illustrated in figure 1.

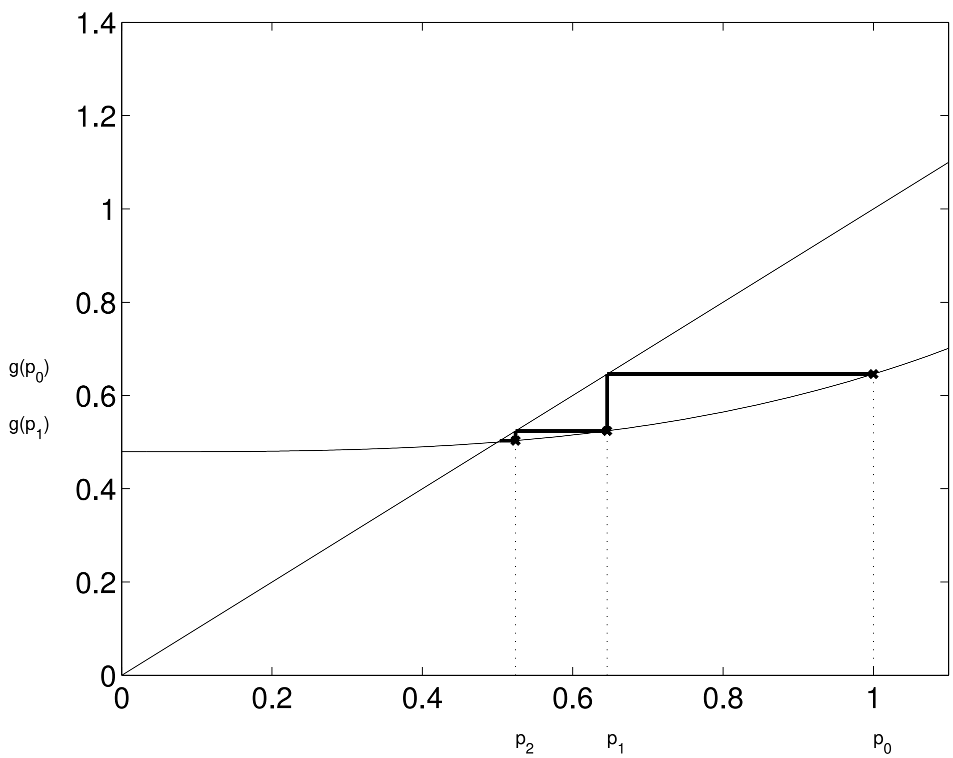


Figure 1: Graphical illustration of the fixed point iteration

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

- $g(p) \in [0, 1]$ 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 $g'(p) = \frac{p^2}{2}$. Note that $g'(p) \geq 0$ for all $p \in [0, 1]$. This implies that

$$0 < \frac{23}{48} = g(0) \leq g(p) \leq g(1) = \frac{31}{48} < 1 \quad \text{for all} \quad p \in [0, 1], \tag{34}$$

so the first condition holds.

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