DELFT UNIVERSITY OF TECHNOLOGY
Faculty of Electrical Engineering, Mathematics and Computer Science

ANSWERS OF THE TEST NUMERICAL METHODS FOR DIFFERENTIAL EQUATIONS (CTB2400)

Thursday June 29 2023, 13:30-16:30

1. (a) Replace $f(t, y)$ by λy in the RK_4 formulas:

$$\begin{aligned} k_1 &= \lambda \Delta t w_n \\ k_2 &= \lambda \Delta t (w_n + \tfrac{1}{2} k_1) = \lambda \Delta t (1 + \tfrac{1}{2}\lambda \Delta t) w_n \\ k_3 &= \lambda \Delta t (w_n + \tfrac{1}{2} k_2) = \lambda \Delta t (1 + \tfrac{1}{2}\lambda \Delta t (1 + \tfrac{1}{2}\lambda \Delta t)) w_n \\ k_4 &= \lambda \Delta t (w_n + k_3) = \lambda \Delta t (1 + \lambda \Delta t (1 + \tfrac{1}{2}\lambda \Delta t (1 + \tfrac{1}{2}\lambda \Delta t)) w_n \end{aligned}$$

Substitution of these expressions into:

$$w_{n+1} = w_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4),$$

and collecting like powers of $\lambda \Delta t$ yields:

$$w_{n+1} = [1 + \lambda \Delta t + \frac{1}{2}(\lambda \Delta t)^2 + \frac{1}{6}(\lambda \Delta t)^3 + \frac{1}{24}(\lambda \Delta t)^4] w_n.$$

The amplification factor is therefore:

$$Q(\lambda \Delta t) = 1 + \lambda \Delta t + \frac{1}{2}(\lambda \Delta t)^2 + \frac{1}{6}(\lambda \Delta t)^3 + \frac{1}{24}(\lambda \Delta t)^4.$$

(b) The local truncation error is defined as

$$\tau_{n+1} = \frac{y(t_{n+1}) - z_{n+1}}{\Delta t}, \quad (1)$$

where z_{n+1} is the numerical solution at t_{n+1}, obtained by starting from the exact value $y(t_n)$ in stead of w_n. Repeating the derivation under (a), with w_n replaced by $y(t_n)$, gives:

$$z_{n+1} = Q(\lambda \Delta t) y(t_n).$$

Using furthermore $y(t_{n+1}) = e^{\lambda \Delta t} y(t_n)$ in (1) it follows that

$$\tau_{n+1} = \frac{e^{\lambda \Delta t} - Q(\lambda \Delta t)}{\Delta t} y(t_n).$$

Canceling the first five terms of the expansion of $e^{\lambda \Delta t}$ against $Q(\lambda \Delta t)$, the required order of magnitude of τ_{n+1} follows.

(c) Use the transformation:

$$\begin{aligned} y_1 &= y, \\ y_2 &= y', \end{aligned}$$

This implies that

$$\begin{aligned} y_1' &= y' = y_2 \, , \\ y_2' &= y'' = -qy_1 - py_2 + \sin t \, , \end{aligned}$$

So the matrix $\mathbf{A}$ and vector $\mathbf{g}$ are:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ -q & -p \end{pmatrix}; \quad \mathbf{g}(t) = \begin{pmatrix} 0 \\ \sin t \end{pmatrix}.$$

Characteristic equation: $\lambda^2 + p\lambda + q = 0$. $\lambda_{1,2} = \frac{-p \pm \sqrt{p^2 - 4q}}{2}$.

(d) Substitution of the values of p and q into the matrix $\mathbf{A}$ yields the eigenvalues $\lambda_{1,2} = -500 \pm 1$. From the hint it follows that $\Delta t \leq 2.8/501 = 0.00559$ as the stability condition.

(e)

$$y'' + py' + qy = \sin t, \quad y(0) = y_0, \; y'(0) = y_0'. \tag{2}$$

After a short time the solution is close to a linear combination of $\sin t$ and $\cos t$, which is called a smooth solution.

The smooth solution can be integrated accurately by RK_4 with a 'large' step size: a step size of 0.1, let us say, would give an error of order 10^{-4} which is sufficient for most engineering purposes. However stability, governed by the eigenvalues, requires that the step size be restricted (see part (d)) to 0.00559. So the stability requirement forces us to choose a step size yielding an unnecessarily accurate solution, which is inefficient.

The Trapezoidal rule, on the other hand, is stable for all step sizes. So the step size is restricted by accuracy requirements only. The Trapezoidal rule has a global error of order Δt^2 such that a good accuracy may be expected for step sizes of about 0.01, which is much larger than the step size for RK4: 0.00559. An efficiency gain may be obtained in spite of the extra work connected with the implicitness of the method.

2. (a) The given formula for $s(x)$ consists of two polynomials of degree 3 on disjunct intervals dividing $[-1, 1]$, so $s(x)$ is a piecewise function consisting of polynomials of degree 3 or lower.

(b) The nodes of the spline are $x = -1$, $x = 0$ and $x = 1$.

We will evaluate $s(x)$ in these three nodes and show that it is equal to $f(x)$ in these nodes:

$$\begin{aligned}
s(-1) &= -\frac{3}{4}x^3 - \frac{9}{4}x^2 + \frac{1}{2}x + 2\Big|_{x=-1} \\
&= -\frac{3}{4}(-1)^3 - \frac{9}{4}(-1)^2 + \frac{1}{2}(-1) + 2 \\
&= \frac{3}{4} - \frac{9}{4} - \frac{1}{2} + 2 \\
&= 0 \\
&= f(-1), \\
s(0) &= \frac{3}{4}x^3 - \frac{9}{4}x^2 + \frac{1}{2}x + 2\Big|_{x=0} \\
&= \frac{3}{4}(0)^3 - \frac{9}{4}(0)^2 + \frac{1}{2}(0) + 2 \\
&= 0 - 0 + 0 + 2 \\
&= 2 \\
&= f(0), \\
s(1) &= \frac{3}{4}x^3 - \frac{9}{4}x^2 + \frac{1}{2}x + 2\Big|_{x=1} \\
&= \frac{3}{4}(1)^3 - \frac{9}{4}(1)^2 + \frac{1}{2}(1) + 2 \\
&= \frac{3}{4} - \frac{9}{4} + \frac{1}{2} + 2 \\
&= 1 \\
&= f(1).
\end{aligned}$$

(c) Because $s(x)$ consists of polynomials, the only possible point of discontinuity is the node $x = 0$, so $s(x)$ is continuous if it is continuous in $x = 0$.

Therefore we have to show

$$\lim_{x \to 0^-} s(x) = \lim_{x \to 0^+} s(x).$$

The left limit equals:

$$\begin{aligned}
\lim_{x \to 0^-} s(x) &= \lim_{x \to 0^-} -\frac{3}{4}x^3 - \frac{9}{4}x^2 + \frac{1}{2}x + 2 \\
&= 2.
\end{aligned}$$

The right limit equals:

$$\begin{aligned}
\lim_{x \to 0^+} s(x) &= \lim_{x \to 0^+} \frac{3}{4}x^3 - \frac{9}{4}x^2 + \frac{1}{2}x + 2 \\
&= 2.
\end{aligned}$$

So $s(x)$ is continuous.

The derivative $s'(x)$ is given by

$$s'(x) = \begin{cases} -\frac{9}{4}x^2 - \frac{9}{2}x + \frac{1}{2} & \text{if} \quad x \in [-1, 0), \\ \frac{9}{4}x^2 - \frac{9}{2}x + \frac{1}{2} & \text{if} \quad x \in [0, 1]. \end{cases}$$

$s'(x)$ is continuous if it is continuous in $x = 0$, so we have to show

$$\lim_{x \to 0^-} s'(x) = \lim_{x \to 0^+} s'(x).$$

The left limit equals:

$$\begin{aligned} \lim_{x \to 0^-} s'(x) &= \lim_{x \to 0^-} -\frac{9}{4}x^2 - \frac{9}{2}x + \frac{1}{2} \\ &= \frac{1}{2}. \end{aligned}$$

The right limit equals:

$$\begin{aligned} \lim_{x \to 0^+} s'(x) &= \lim_{x \to 0^+} \frac{9}{4}x^2 - \frac{9}{2}x + \frac{1}{2} \\ &= \frac{1}{2}. \end{aligned}$$

So $s'(x)$ is continuous.
The second derivative $s''(x)$ is given by

$$s''(x) = \begin{cases} -\frac{9}{2}x - \frac{9}{2} & \text{if} \quad x \in [-1, 0), \\ \frac{9}{2}x - \frac{9}{2} & \text{if} \quad x \in [0, 1]. \end{cases}$$

$s''(x)$ is continuous if it is continuous in $x = 0$, so we have to show

$$\lim_{x \to 0^-} s''(x) = \lim_{x \to 0^+} s''(x).$$

The left limit equals:

$$\begin{aligned} \lim_{x \to 0^-} s''(x) &= \lim_{x \to 0^-} -\frac{9}{2}x - \frac{9}{2} \\ &= -\frac{9}{2}. \end{aligned}$$

The right limit equals:

$$\begin{aligned} \lim_{x \to 0^+} s''(x) &= \lim_{x \to 0^+} \frac{9}{2}x - \frac{9}{2} \\ &= -\frac{9}{2}. \end{aligned}$$

So $s''(x)$ is continuous.

(d) Evualuating $s''(x)$ in $x = -1$ gives:

$$s''(-1) = -\frac{9}{2}x - \frac{9}{2}\bigg|_{x=-1} = \frac{9}{2} - \frac{9}{2} = 0,$$

and evaluation at $x = 1$ gives

$$s''(1) = \frac{9}{2}x - \frac{9}{2}\bigg|_{x=1} = \frac{9}{2} - \frac{9}{2} = 0,$$

so indeed $s''(x) = 0$ in the end points.

(e) $x = \frac{1}{2}$ lies in the right interval, so we need to perform the next calculation:

$$\begin{aligned} f\left(\frac{1}{2}\right) &\approx s\left(\frac{1}{2}\right) \\ &= \frac{3}{4}x^3 - \frac{9}{4}x^2 + \frac{1}{2}x + 2\bigg|_{x=\frac{1}{2}} \\ &= \frac{3}{4}\left(\frac{1}{2}\right)^3 - \frac{9}{4}\left(\frac{1}{2}\right)^2 + \frac{1}{2}\left(\frac{1}{2}\right) + 2 \\ &= 1.7812 \end{aligned}$$

3. (a) The right composite Rectangle rule is given by

$$\int_a^b y(x)dx \approx h \sum_{j=1}^{n} y(x_j),$$

with $hn = b - a$ and $x_j = a + jh$ for $j = 0, \ldots, n$.

From $h = \pi/2$, $a = 0$ and $b = 2\pi$, it follows that $n = 4$ and the following table also follows:

j	0	1	2	3	4
x_j	0	$\pi/2$	π	$3\pi/2$	2π
$y(x_j)$	2	1	0	1	2

Note: Every miscalculation in the calculation below gives a subtraction of 1/2 point, with at most 1 point being subtracted.

Applying the right composite Rectangle rule with $h = \pi/2$ gives

$$\begin{aligned}\int_0^{2\pi} y(x)dx &\approx \frac{\pi}{2}\left(y\left(\frac{\pi}{2}\right) + y(\pi) + y\left(\frac{3\pi}{2}\right) + y(2\pi)\right), \\ &= \frac{\pi}{2}(1 + 0 + 1 + 2), \\ &= 2\pi.\end{aligned}$$

(b) The composite Trapezoidal rule is given by

$$\int_a^b y(x)dx \approx h \sum_{j=1}^{n} \frac{1}{2}\left(y(x_{j-1}) + y(x_j)\right),$$

with $hn = b - a$ and $x_j = a + jh$ for $j = 0, \ldots, n$.

Note: Every miscalculation in the calculation below gives a subtraction of 1/2 point, with at most 1/2 point being subtracted.

Applying the composite Trapezoidal rule with $h = \pi/2$ gives

$$\begin{aligned}\int_0^{2\pi} y(x)dx &\approx \frac{\pi}{2}\left(\frac{1}{2}y(0) + y\left(\frac{\pi}{2}\right) + y(\pi) + y\left(\frac{3\pi}{2}\right) + \frac{1}{2}y(2\pi)\right), \\ &= \frac{\pi}{2}\left(\frac{2}{2} + 1 + 0 + 1 + \frac{2}{2}\right), \\ &= 2\pi.\end{aligned}$$

(c) *Note: Your answers should be consistent with each other. For each inconsistency 1/4 point will be subtracted, with at most 1 1/2 points being subtracted.*

The derivatives of the function y are given by

$$\begin{aligned}y'(x) &= -\sin(x), \\ y''(x) &= -\cos(x).\end{aligned}$$

From this it follows

$$\begin{aligned}\max_{x\in[0,2\pi]} |y'(x)| &= 1, \\ \max_{x\in[0,2\pi]} |y''(x)| &= 1.\end{aligned}$$

Therefore the explicit upper bounds for ε_R and ε_T are given by

$$\varepsilon_R \leq \pi h,$$
$$\varepsilon_T \leq \frac{\pi}{6} h^2.$$

(d) *Note: No points are given if one of the following holds:*

- *no arguments are presented;*
- *the selected method is inconsistent with the arguments.*

Note: Incorrect arguments on topics other than the amount of work and accuracy give per such argument a subtraction of $^1/_4$ *point, with at most* $^1/_2$ *points being subtracted.*

From the above upper bounds one can conclude that

$$\varepsilon_T < \varepsilon_R$$

if $h < 6$. Hence, the error for the composite Trapezoidal method is much smaller for small h then the error for the right composite Rectangle rule.

Furthermore, with $n = {}^{b-a}/_h$, the number of function evaluations of the right composite Rectangle rule is n, and $n+1$ for the composite Trapezoidal rule. It also holds that

$$\frac{n+1}{n} \approx 1,$$

for large n. Hence, for small h the amount of work within both methods is similar.

Therefore the composite Trapezoidal method should be preferred for small h.