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

ANSWERS OF THE TEST NUMERICAL METHODS FOR DIFFERENTIAL EQUATIONS (WI3097 TU and CTB2400)
Thursday August 14 2014, 18:30-21:30

1. a) The local truncation error is defined as

$$\tau_{n+1} = \frac{y_{n+1} - z_{n+1}}{h}, \tag{1}$$

where y_{n+1} is the exact solution at t_{n+1} and z_{n+1} the value obtained by applying the given method at the exact solution point (t_n, y_n):

$$\begin{aligned} k_1 &= hf(t_n, y_n) \\ k_2 &= hf(t_n + h, y_n + k_1) \\ z_{n+1} &= y_n + \beta\ k_1 + (1-\beta)\ k_2. \end{aligned} \tag{2}$$

Both y_{n+1} and z_{n+1} have to be expanded into a Taylor series at the point (t_n, y_n). To start with z_{n+1}, k_1 and k_2 are substituted into the corrector part (2):

$$z_{n+1} = y_n + \beta\ hf(t_n, y_n) + (1-\beta)\ hf(t_n + h, y_n + hf(t_n, y_n)). \tag{3}$$

Next $f(t_n + h, y_n + hf(t_n, y_n))$ is expanded:

$$\begin{aligned} f(t_n + h, y_n + hf(t_n, y_n)) &= f(t_n, y_n) + h\frac{\partial f}{\partial t}(t_n, y_n) + hf(t_n, y_n)\frac{\partial f}{\partial y}(t_n, y_n) + \ldots \\ &= y'_n + h[\frac{\partial f}{\partial t} + f\frac{\partial f}{\partial y}](t_n, y_n) + O(h^2), \end{aligned} \tag{4}$$

using the differential equation $y' = f(t, y)$.
In this expression $[\frac{\partial f}{\partial t} + f\frac{\partial f}{\partial t}](t_n, y_n)$ can be replaced by $y''(t_n) = y''_n$, for

$$y'' = \frac{dy'}{dt} = \frac{df}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial y}\frac{dy}{dt} = \frac{\partial f}{\partial t} + f\frac{\partial f}{\partial y},$$

again using $y' = f(t, y)$ in the last step.
As a result, (4) becomes:

$$f(t_n + h, y_n + hf(t_n, y_n)) = y'_n + hy''_n + O(h^2).$$

Substitution of this expression into (3) gives:

$$\begin{aligned} z_{n+1} &= y_n + \beta\ hy'_n + (1-\beta)\ h\ (y'_n + hy''_n + O(h^2)) \\ &= y_n + hy'_n + (1-\beta)h^2 y''_n + O(h^3). \end{aligned}$$

Substitution of this expansion, together with the expansion for y_{n+1}:

$$y_{n+1} = y_n + hy'_n + \frac{1}{2}h^2 y''_n + O(h^3),$$

into (1) yields:

$$\begin{aligned}\tau_{n+1} &= \frac{y_n + hy'_n + \frac{1}{2}h^2 y''_n + O(h^3) - [y_n + hy'_n + (1-\beta)h^2 y''_n + O(h^3)]}{h} \\ &= (\beta - \frac{1}{2})\ h\ y''_n + O(h^2)\end{aligned}$$

It turns out that the truncation error is O(h), except for $\beta = \frac{1}{2}$. Note that the predictor-corrector method is just Modified Euler for $\beta = \frac{1}{2}$.

b) The amplification factor is found by applying the method to the homogeneous test equation $y' = \lambda y$:

$$\begin{aligned}k_1 &= h\lambda w_n \\ k_2 &= h\lambda(w_n + h\lambda w_n) = h\lambda(1 + h\lambda)w_n \\ w_{n+1} &= w_n + \beta\ h\lambda w_n + (1-\beta)\ h\lambda(1 + h\lambda)w_n \\ &= [1 + h\lambda + (1-\beta)(h\lambda)^2]w_n.\end{aligned}$$

The amplification factor $Q(h\lambda)$ is seen to be $1 + h\lambda + (1-\beta)(h\lambda)^2$.

c) To derive the stability condition we need the eigenvalues of the system

$$\mathbf{x}' = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \mathbf{x}.$$

These are purely imaginary: $\lambda_{1,2} = \pm i$, as can be seen easily.
For stability we require $|Q(\pm hi)| < 1$ or, more conveniently,
$|Q(\pm hi)|^2 < 1$.
From c):

$$\begin{aligned}|1 \pm hi + (1-\beta)(\pm hi)^2|^2 &< 1 \leftrightarrow \\ |1 - (1-\beta)h^2 \pm hi|^2 &< 1 \leftrightarrow \\ (1 - (1-\beta)h^2)^2 + h^2 &< 1 \leftrightarrow \\ 1 - 2(1-\beta)h^2 + (1-\beta)^2h^4 + h^2 &< 1 \leftrightarrow \\ (1-\beta)^2h^2 &< 2(1-\beta) - 1 = 1 - 2\beta.\end{aligned}$$

(Note: the squared modulus of a complex number equals the sum of the squares of it's real and imaginary part.)
It now follows that

$$h^2 < \frac{1 - 2\beta}{(1-\beta)^2}$$

is required for stability.
Clearly, stability is possible only for $\beta < \frac{1}{2}$.

d) We have optimal stability if the upper bound for h is as large as possible. So we have to investigate the behavior of the function $g(\beta) = \frac{1-2\beta}{(1-\beta)^2}$ for $\beta < \frac{1}{2}$. The derivative of $g(\beta)$ is given: $\frac{-2\beta}{(1-\beta)^2}$. This derivative is positive for $\beta < 0$ and negative for $0 < \beta < \frac{1}{2}$. So $g(\beta)$ assumes its maximum for $\beta = 0$, $g(0)$ being 1. The optimal stability condition for the considered system is therefore $h < 1$.

e) First we compute the vectorial counterpart of k_1:

$$\mathbf{k}_1 = hA\mathbf{y}_0 = 0.5A \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix}.$$

Then we get

$$\mathbf{y}_0 + \mathbf{k}_1 = \begin{pmatrix} \frac{3}{2} \\ \frac{1}{2} \end{pmatrix}.$$

Hence

$$\mathbf{k}_2 = 0.5A(\mathbf{y}_0 + \mathbf{k}_1) = 0.5A \left(\begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{pmatrix} \right) = 0.5A \begin{pmatrix} \frac{3}{2} \\ \frac{1}{2} \end{pmatrix} = \begin{pmatrix} \frac{1}{4} \\ -\frac{3}{4} \end{pmatrix}.$$

With $\beta = 0$, we get

$$\mathbf{w}_1 = \mathbf{y}_0 + \mathbf{k}_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \frac{1}{4} \\ -\frac{3}{4} \end{pmatrix} = \begin{pmatrix} \frac{5}{4} \\ \frac{1}{4} \end{pmatrix}.$$

2. (a) The linear Lagrangian interpolatory polynomial, with nodes x_0 and x_1, is given by

$$p_1(x) = \frac{x - x_1}{x_0 - x_1} f(x_0) + \frac{x - x_0}{x_1 - x_0} f(x_1). \tag{5}$$

This is evident from application of the given formula.

(b) The quadratic Lagrangian interpolatory polynomial with nodes x_0, x_1 and x_2 is given by

$$p_2(x) = \frac{(x - x_1)(x - x_2)}{(x_0 - x_1)(x_0 - x_2)} f(x_0) + \frac{(x - x_0)(x - x_2)}{(x_1 - x_0)(x_1 - x_2)} f(x_1) + \frac{(x - x_0)(x - x_1)}{(x_2 - x_0)(x_2 - x_1)} f(x_2). \tag{6}$$

This is also evident from application of the given formula.

(c) To this extent, we compute $p_1(0.5)$ and $p_2(0.5)$ for both linear and quadratic Lagrangian interpolation as approximations at $x = 0.5$. For linear interpolation, we have

$$p_1(0.5) = 0.5 + \frac{1}{2} \cdot 2 = \frac{3}{2}, \tag{7}$$

and for quadratic interpolation, one obtains

$$p_2(0.5) = \frac{(0.5-1)(0.5-2)}{1\cdot(-2)}\cdot 1 + \frac{(0.5-0)(0.5-2)}{1\cdot(-1)}\cdot 2 + \frac{(0.5-0)(0.5-1)}{2\cdot 1}\cdot 4 = \frac{11}{8} = 1.375. \tag{8}$$

(d) The difference between the exact polynomial p and the perturbed polynomial $\hat{p}$ is bounded by

$$|p(x)-\hat{p}(x)| \leq \frac{|x_1-x||f(x_0)-\hat{f}(x_0)| + |x-x_0||f(x_1)-\hat{f}(x_1)|}{x_1-x_0} \leq \frac{|x_1-x|+|x-x_0|}{x_1-x_0}\varepsilon\ .$$

For interpolation we know that $x_0 \leq x \leq x_1$ so the inequality simplifies to

$$|p(x)-\hat{p}(x)| \leq \frac{x_1-x+x-x_0}{x_1-x_0}\varepsilon = \frac{x_1-x_0}{x_1-x_0}\varepsilon = \varepsilon,$$

so the maximal error is bounded by ε.

(e) The iteration process is a fixed point method. If the process converges we have: $\lim_{n\to\infty} x_n = p$. Using this in the iteration process yields:

$$\lim_{n\to\infty} x_{n+1} = \lim_{n\to\infty}[x_n + h(x_n)(x_n^3-3)]$$

Since h is a continuous function one obtains:

$$p = p + h(p)(p^3-3)$$

so

$$h(p)(p^3-3) = 0.$$

Since $h(x) \neq 0$ for each $x \neq 0$ it follows that $p^3 - 3 = 0$ and thus $p = 3^{\frac{1}{3}}$.

(f) The convergence of a fixed point method $x_{n+1} = g(x_n)$ is determined by $g'(p)$. If $|g'(p)| < 1$ the method converges, whereas if $|g'(p)| > 1$ the method diverges. For all choices we compute the first derivative in p. For the first method we elaborate all steps. For the other methods we only give the final result. For h_1 we have $g_1(x) = x - \frac{x^3-3}{x^4}$. The first derivative is:

$$g_1'(x) = 1 - \frac{3x^2\cdot x^4 - (x^3-3)\cdot 4x^3}{(x^4)^2}$$

Substitution of p yields:

$$g_1'(p) = 1 - \frac{3p^6 - (p^3-3)\cdot 4p^3}{p^8}.$$

Since $p = 3^{\frac{1}{3}}$ the final term cancels:

$$g_1'(p) = 1 - \frac{3p^6}{p^8} = 1 - 3^{\frac{1}{3}} = -0.4422.$$

This implies that the method is convergent with convergence factor 0.4422.

For the second method we have:

$$g_2'(p) = 1 - \frac{3p^4 - (p^3 - 3) \cdot 2p}{p^4} = 1 - \frac{3p^4}{p^4} = -2$$

Thus the method diverges.

For the third method we have:

$$g_3'(p) = 1 - \frac{9p^4 - (p^3 - 3) \cdot 6p}{9p^4} = 1 - \frac{9p^4}{9p^4} = 0$$

Thus the method is convergent with convergence factor 0.

Concluding we note that the third method is the fastest.

(g) To estimate the error in p we first approximate the function f in the neighboorhood of p by the first order Taylor polynomial:

$$P_1(x) = f(p) + (x - p)f'(p) = (x - p)f'(p).$$

Due to the measurement errors we know that

$$(x - p)f'(p) - \epsilon_{max} \leq \hat{P}_1(x) \leq (x - p)f'(p) + \epsilon_{max}.$$

This implies that the perturbed root $\hat{p}$ is bounded by the roots of $(x-p)f'(p) - \epsilon_{max}$ and $(x - p)f'(p) + \epsilon_{max}$, which leads to

$$p - \frac{\epsilon_{max}}{|f'(p)|} \leq \hat{p} \leq p + \frac{\epsilon_{max}}{|f'(p)|}.$$