Lecture 2

I. The error estimate for the Euler explicit method applied to the initial value problem

Exact problem

$$(1) \qquad \begin{cases} y' = f(t, y) \\ y(t_0) = y_0 \end{cases}$$

Assumptions:

1. the function $f(t, y)$ is regular (i.e. it has continuous derivatives $f_t(t, y)$, $f_y(t, y)$, $f_{tt}(t, y)$, $f_{ty}(t, y)$, $f_{yy}(t, y)$), in particular it satisfies the Lipschitz condition

$$|f(t, y_1) - f(t, y_2)| \leq L|y_1 - y_2|$$

   for all t, y_1, y_2, where $L > 0$ is some constant.

2. let
   $M = \max_{(t,y)}\{|f(t, y)|, |f_t(t, y)|, |f_x(t, y)|, |f_{tt}(t, y)|, |f_{xt}(t, y)|, |f_{xx}(t, y)|\}$

We define the approximate problem as follows:

Find a series of approximate values using the explicit Euler algorithm

$$y_0, \quad y_{i+1} = y_i + hf(t_i, y_i), \; i = 0, 1, 2, \ldots, n - 1$$

The global error of approximation is defined as follows

$$e(h) = \max_{1 \leq i \leq n} |y(t_i) - y_i|.$$

To analyze $e(h)$ we consider the sequence of the errors at the individual grid points

$$e_i = y(t_i) - y_i, \quad i = 1, 2, \ldots, n,$$

Then by Taylor's formula* (up to the first term) we get

$$(2) \qquad y(t_{i+1}) = y(t_i) + hy'(t_i) + R(h),$$

where $|R(h)| \leq Ch^2$,

$$C = \frac{1}{2} \max_{t_i \le t \le t_{i+1}} |y''(t)| = \frac{1}{2} \max_{t_i \le t \le t_{i+1}} |f_t + f \cdot f_y| \le \frac{1}{2}(M + M^2) \le M^2$$

$$y'' = (y')' = \frac{d}{dt} f(t, y) = f_t(t, y) + y' f_y(t, y) = f_t + f \cdot f_y$$

It follows from the differential equations that $y'(t_i) = f(t_i, y_i)$, therefore

(3) $\quad y(t_{i+1}) = y(t_i) + hf(t_i, y_i) + R(h)$

* - Taylor's formula up to $n - th$ term has a form:

$y(t+h) = y(t) + \frac{h}{1!}y'(t) + \frac{h^2}{2!}y''(t) + \cdots + \frac{h^n}{n!}y^{(n)}(t) + R_n(h)$,

where $R_n(h)$ is the remainder, for which we have an estimate
$|R_n(h) \le Ch^{n+1}$.
Since

(4) $\quad y_{i+1} = y_i + hf(t_i, y_i)$,

it follows from (**??**) and (**??**) that

(5) $\quad e_{i+1} = y(t_{i+1}) - y_{i+1} = (y(t_i) - y_i) + h(f(t_i, y(t_i)) - f(t_i, y_i)) + R(h)$.

Since f satisfies the Lipschitz condition

$$|f(t_i, y(t_i)) - f(t_i, y_i)| \le L|y(t_i) - y_i|,$$

from (**??**) we get inequality

(6) $\quad |e_{i+1}| \le (1 + Lh)|e_i| + Ch^2, \quad i = 0, 1, 2, \ldots, n,$

where $C > 0$ is some constant independent of h. We discuss this inequality beginning with the following auxilliary lemma

Lemma 1.

Let the sequence $\{e_i\}$ satisfy

(7) $\quad |e_{i+1}| \le A|e_i| + B, \quad i = 0, 1, 2, \ldots$

where $A, B > 0$ are some constants. Then for each e_i we have the following explicit estimate

(8) $\quad |e_i| \le A^i|e_0| + B\dfrac{A^i - 1}{A - 1}, \quad i = 0, 1, 2, \ldots$

Proof of lemma. The proof is by induction. First we observe that

$$|e_1| \leq A|e_0| + B.$$

Assuming (**??**) for i, for $i+1$ we get inequality

$$|e_{i+1}| \leq A|e_i| + B \leq A\left(A^i|e_0| + B\frac{A^i - 1}{A - 1}\right) + B = A^{i+1}|e_0| + B\frac{A^i - 1}{A - 1}.$$

An inductive argument ends the proof.

Now we apply lemma 1 to our sequence, where $e_0 = 0$, $A = 1 + Lh$ and $B = Ch^2$. Then we get

$$e_i \leq \frac{(1 + hL)^i - 1}{hL} Ch^2 \leq \frac{C}{L}(1 + hL)^n h, \quad 0 \leq i \leq n.$$

Since

$$t_0 < t_1 < \cdots < t_n = T, \quad nh = T - t_0$$

we have

$$(1 + hL)^n = \left[(1 + hL)^{\frac{1}{hL}}\right]^{nhL} \leq \exp((T - t_0)L),$$

The last inequality follows from the known inequality $(1 + x)^{\frac{1}{x}} \leq e$ for $x > 0$. Finally we get the estimate of the global error for the explicit Euler method

$$e(h) = \max_{0 \leq i \leq n} |e_i(h)| \leq \frac{C}{L} e^{(T - t_0)L}\ h = C_1 h. \tag{9}$$

Let us compare the global error (**??**) for the explicit Euler method with the local error. For the latter error we have an estimate

$$e_l^i(h) = |y(t_{i+1}) - y(t_i) - hf(t_i, y(t_i))| = |y(t_{i+1}) - y(t_i) - hy'(t_i)| = |r(h)| \tag{10}$$

where $r(h)$ is the remainder in Taylor's formula, which satisfies the inequality $|r(h)| \leq Ch^2$. Hence it follows that

$$e_l(h) = \max_{0 \leq i \leq n} (e_l^i(h)) \leq Ch^2.$$

Generally, for difference methods between local and global errors we usually have the following relation

(11) $e_l(h) \le ch^{p+1}$ and $e(h) \le ch^p$,

where $p > 0$ is the order of the method.

III Practical methods of higher order - the Runge-Kutta's methods

1. The method of the second order

$$y_{i+1}: \quad k_1 = f(t_i, y_i)h$$
$$k_2 = f(t_i + \frac{1}{2}h, y_i + \frac{1}{2}k_1)h$$
$$y_{i+1} = y_i + k_2$$

2. The method of the fourth order

$$y_{i+1}: \quad k_1 = f(t_i, y_i)h$$
$$k_2 = f(t_i + \frac{1}{2}h, y_i + \frac{1}{2}k_1)h$$

$$k_3 = f(t_i + \frac{1}{2}h, y_i + \frac{1}{2}k_2)h$$
$$k_4 = f(t_i + h, y_i + k_3)h$$
$$y_{i+1} = y_i + \frac{1}{6}k_1 + \frac{1}{3}k_2 + \frac{1}{3}k_3 + \frac{1}{6}k_4$$

Example 1. The second order Runge-Kutta method for the system of equations

$$\begin{cases} x' = f(t, x, y) \\ y' = g(t, x, y) \end{cases}$$

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} : \quad k_1 = \begin{bmatrix} k_{11} \\ k_{12} \end{bmatrix} = \begin{bmatrix} f(t_i, x_i, y_i)h \\ g(t_i, x_i, y_i)h \end{bmatrix}$$

$$k_2 = \begin{bmatrix} k_{21} \\ k_{22} \end{bmatrix} = \begin{bmatrix} f(t_i + \frac{1}{2}h, x_i + \frac{1}{2}k_{11}, y_i + \frac{1}{2}k_{12})h \\ \\ g(t_i + \frac{1}{2}h, x_i + \frac{1}{2}k_{11}, y_i + \frac{1}{2}k_{12})h \end{bmatrix}$$

$$\begin{bmatrix} x_{i+1} \\ y_{i+1} \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} + \begin{bmatrix} k_{21} \\ k_{22} \end{bmatrix}$$

A practical guide for verification of the order $p > 0$ in (**??**) of the convergence rate of the algorithm.

1. Asume that $e(h) \approx Ch^p$.
2. Let choose a step h_0 for example $h_0 = 0.05$ and perform numerical tests for a number of step sizes of h, for example $h = h_0, h_0/2, h_0/4, \ldots, h_0/2^5$.
3. For each value of h, determine errors: $e(h_0)$, $e(h_0/2)$, $e(h_0/4)$,$\ldots e(h_0/2^5)$ of the approximation of the exact solution $y(t)$ by an approximate solution y_h.
4. If in point 3. the exact solution $y(t)$ is not available, you can take instead the approximate solution $y_h$obtained for the last value of step h.
5. It follows from our assumption that

$$e(h_0) \approx Ch_0^p, \quad e(h_0/2) \approx Ch_0^p 2^{-p}, \quad e(h_0/4) \approx C2^{-2p}h_0^p, \quad \ldots$$
$$e(h_0/2^k) \approx C2^{-kp}h_0^p,$$

So we may expect that

$$e(h_0/2^i)/e(h_0/2^{i+1}) \approx 2^p, \; i = 0, 1, 2, \ldots,$$

which is verifiable by computation. Hence we retrieve the value of p.
6. To illustrate results you can draw the points

$$P_k = (-\ln(h_0/2^k), -\ln(e(h_0/2^k)).$$

Since

$$-\ln(e(h)) \approx p\ln(-h) + \ln(-C),$$

we expect that points P_k are arranged along the line $y = p \cdot x + b$. The coefficients p and b can be retrieved by using a Matlab command 'polyfit'.