Numerical ODEs

Supplementary notes.

Convergence

Given the ordinary differential equation

$$(1) \quad x' = f(t,x), \quad x_0 = x(t_0)$$

In the following, we assume that a unique solution $x(t)$ exist, that $f(t,x)$ is "sufficiently smooth", and we assume that f satisfy the Lipschitz condition

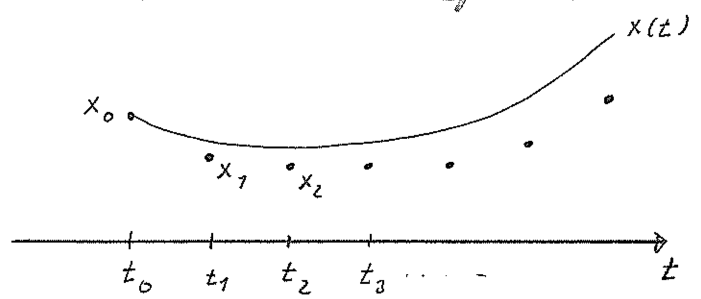
$$|f(t,x) - f(t,\tilde{x})| \leq L \cdot |x - \tilde{x}|$$

for all $x, \tilde{x}$, where $L \geq 0$ is some constant.

We are looking for approximations to $x(t)$ at some given points, that is

$$x_i \approx x(t_i), \quad t_i = t_0 + i \cdot h, \quad i = 1, 2, \cdots$$

where h is the stepsize.



Now, let T be some fixed point, and assume that we use n steps with our method to find an approximation to $x(T)$.
The <u>global error</u> is the error

$$E_n = x(T) - x_n$$

The stepsize used is $h = \dfrac{T - t_0}{n}$.

The method is <u>convergent</u> if

$$E_n \underset{n \to \infty}{\to} 0 \qquad (\text{or } h \to 0)$$

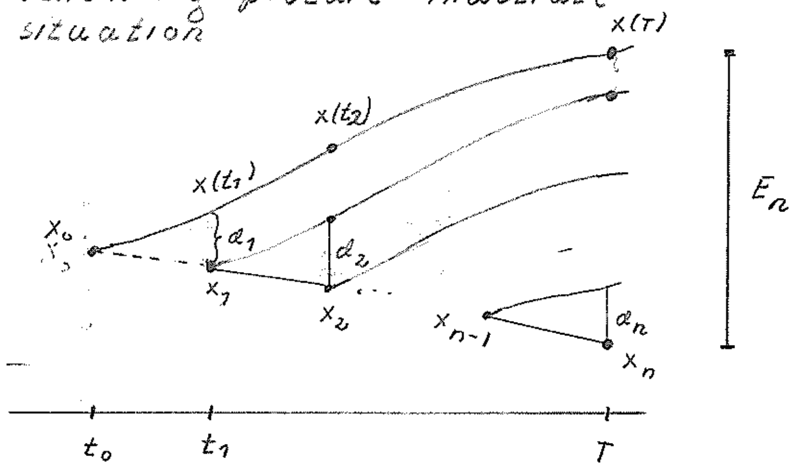
for all ODEs satisfying the assumptions, and the method is of <u>order p</u> if

$$E_n = \mathcal{O}(h^p).$$

The global error depends on two factors:

- The local error, d_i which is the error made in each step.
- The propagation of the errors.

The following picture illustrate the situation



Let us illustrate these concepts by the famous Eulers method, given by

$$x_{i+1} = x_i + hf(t_i, x_i), \quad i = 0,1,2,\dots$$

The local truncation error is found by comparing the numerical and the exact solution after one step, assuming $x_i = x(t_i)$. Thus

$$d_{i+1} = x(t_i + h) - x_i - hf(t_i, x_i)$$

$$= x(t_i) + hx'(t_i) + \tfrac{1}{2}h^2 x''(t_i + \xi h)$$

$$- x_i - hf(t_i, x_i)$$

Where $\xi \in (0,1)$. Using the ODE (1), we get

$$d_i = \tfrac{1}{2}h^2 x''(t_i + \xi h)$$

If x'' is bounded in the region of interest, then there is a $C > 0$ such that

(3) $\quad |d_i| \leq C \cdot h^2.$

Convergence of the Euler method.

Let $E_i = x(t_i) - x_i$ be the global error after i steps.

$$x(t_i + h) = x(t_i) + h f(t_i, x(t_i)) + d_i .$$

$$x_{i+1} = x_i + h f(t_i, x_i)$$

and

$$E_{i+1} = E_i + h\left(f(t_i, x(t_i)) - f(t_i, x_i)\right) + d_i$$

Using the Lipschitz condition (2) and the bound for the local truncation error (3) we get

$$|E_{i+1}| \leq (1 + h\cdot L)|E_i| + C\cdot h^2 , \quad i = 0,1,2,\ldots$$

such that

$$|E_1| \leq (1+hL)|E_0| + C\cdot h^2$$

$$|E_2| \leq (1+hL)^2 |E_0| + (1+hL+1)C\cdot h^2$$

$$\vdots$$

$$|E_n| \leq (1+hL)^n |E_0| + \sum_{i=0}^{n-1} (1+hL)^i \, Ch^2$$

$$= (1+hL)^n \cdot |E_0| + \frac{(1+hL)^n - 1}{h\cdot L} \cdot Ch^2$$

Remember that E_n is the error $x(T) - x_n$ for the case where n steps of stepsize $h = (T-t_0)/n$ has been used.
Also, use $1 + hL \leq e^{hL}$ since $hL > 0$.
Then

$$|E_n| \leq e^{hL\cdot n} |E_0| + \frac{e^{hLn} - 1}{L} C\cdot h$$

or

$$\boxed{|E_n| \leq e^{L(T-t_0)} |E_0| + \frac{e^{L(T-t_0)} - 1}{L} \cdot C\cdot h}$$

The first term gives an upper bound for the propagation of any initial error $E_0 = x(t_0) - x_0$. If this is zero, which we usually assume, we see that the method is of order 1, and thereby convergent.

In general, one step methods like Runge-Kutta methods can be written as

$$x_{i+1} = x_i + h\,\Phi(t_i, x_i; h) \quad , \; i = 0,1,2,$$

where Φ is some function depending on f and the method.
In this case, the local truncation error is

$$d_i = x(t_{i+1}) - x(t_i) - h\,\Phi(t_i; x(t_i); h).$$

If Φ satisfy a Lipschitz condition

$$|\Phi(t,x;h) - \Phi(t,\tilde{x};h)| \leq M \cdot |x - \tilde{x}|$$

and $|d_i| \leq C \cdot h^{p+1}$ then the argument on the previous page can be repeated, to prove that

$$|E_n| \leq \frac{e^{M(T-t_0)} - 1}{M} \cdot C \cdot h^p$$

(assuming $E_0 = 0$).

Numerical ODE

Supplementary notes.

Runge-Kutta methods.

An s-stage Runge-Kutta (RK) method is defined as

$$K_i = h f(t_0 + c_i \cdot h,\ x_0 + \sum_{j=1}^{s} a_{ij} K_j),\quad i = 1, \dots, s$$

$$x_1 = x_0 + \sum_{i=1}^{s} b_i K_i$$

The particular methods is given by the coefficients c_i, a_{ij}, b_i, which often is presented in a Butcher tableau

$$\begin{array}{c|cccc} c_1 & a_{11} & a_{12} & \cdots & a_{1s} \\ c_2 & a_{21} & a_{22} & & a_{2s} \\ \vdots & & & & \\ c_s & a_{s1} & a_{s2} & \cdot & a_{ss} \\ \hline & b_1 & b_2 & \cdots & b_3 \end{array}$$

A method is called explicit if $a_{ij} = 0$ $j \geq i$ otherwise it is called implicit.

Examples:

Eulers method:

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array}$$

Heuns method:

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & 1/2 & 1/2 \end{array}$$

Trapezoidal rule:
(Implicit)

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1/2 & 1/2 \\ \hline & 1/2 & 1/2 \end{array}$$

Runge-Kutta 4th order method:

$$\begin{array}{c|cccc} 0 & & & & \\ 1/2 & 1/2 & & & \\ 1/2 & 0 & 1/2 & & \\ 1 & 0 & 0 & 1 & \\ \hline & 1/6 & 1/3 & 1/3 & 1/6 \end{array}$$

An RK method is of order p, that is

$$|x(t_0+h) - x_1| \leq C \cdot h^{p+1}$$

if

$$c_i = \sum_{j=1}^{s} a_{ij}, \quad i = 1, \cdots, s$$

and:

$p = 1$: $\sum_i b_i = 1$

$p = 2$: $\sum_i b_i c_i = \frac{1}{2}$

$p = 3$: $\sum_i b_i c_i^2 = \frac{1}{3}$

$\sum_{i,j} b_i a_{ij} c_j = \frac{1}{6}$

$p = 4$: $\sum_i b_i c_i^3 = \frac{1}{4}$

$\sum_{i,j} b_i a_{ij} c_j^2 = \frac{1}{12}$

$\sum_{ij} b_i c_i a_{ij} c_j = \frac{1}{8}$

$\sum_{ijk} b_i a_{ij} a_{jk} c_k = \frac{1}{24}$

For $p = 5$ there are 9 additional cond.
$p = 6$ 20
$p = 7$ 48
$p = 8$ 115
$p = 9$ 286
$p = 10$ 719

So, for a method of order 10, a total number of 1205 conditions has to be satisfied.