ODE - order of a method, Collatz (midpoint) method

First-order initial value problem

$$\mathbf{Y}'(x) = \mathbf{F}(x, \mathbf{Y}(x)) \quad \text{with an initial condition } \mathbf{Y}(x^{(0)}) = \mathbf{Y}^{(0)}\,, \tag{1}$$

where

$$\mathbf{Y}(x) = \begin{bmatrix} y_1(x) \\ y_2(x) \\ \vdots \\ y_n(x) \end{bmatrix}, \ \mathbf{Y}'(x) = \begin{bmatrix} y_1'(x) \\ y_2'(x) \\ \vdots \\ y_n'(x) \end{bmatrix}, \ \mathbf{F}(x, \mathbf{Y}) = \begin{bmatrix} f_1(x, y_1, y_2, \dots, y_n) \\ f_2(x, y_1, y_2, \dots, y_n) \\ \vdots \\ f_n(x, y_1, y_2, \dots, y_n) \end{bmatrix}$$

The system of (1) is called *linear*, if functions f_i are linear with respect to all variables y_j , i.e., they have the form

$$f_i(x, y_1, y_2, \dots, y_n) = g_{i0}(x) + g_{i1}(x)\, y_1 + g_{i2}(x)\, y_2 + ... + g_{in}(x)\, y_n\,, \quad i = 1, \dots n\,. \tag{2}$$

Existence and uniqueness of the (exact) solution

The standard theorem, which formulates *sufficient* conditions for existence and uniqueness of the solution of the initial value problem (1):

Theorem 1

Suppose that functions f_i are continuous in some region D defined by $a \le x \le b,\ a_j < y_j < b_j$ for $j = 1, \dots n$ and that F satisfies *Lipschitz condition* with respect to $\mathbf{Y}$ in D, i. e., there exists a constant L such that

$$\|F(x, \mathbf{Y}) - F(x, \mathbf{Z})\| \le L\,\|\mathbf{Y} - \mathbf{Z}\| \quad \forall\ (x, \mathbf{Y}),\ (x, \mathbf{Z}) \in D\,. \tag{3}$$

Then for any $(x^{(0)}, \mathbf{Y}^{(0)}) \in D$ there exists a unique maximal solution $\mathbf{Y}(x) \subset D$ of the problem (1). Moreover, if $D = I \times R^n$, then the unique maximal solution is defined on the whole interval I.

For a **linear system** (2) the following holds:

Suppose that all functions $g_{ij}(x)$ are continuous on an interval $I =< a, b >$. Then the assumptions of Theorem 1 are satisfied in $D = I \times R^n$ with $L = \max_{x \in I} |g_{ij}(x)|$, so the unique maximal solution is defined on the whole interval I.

The Lipschitz condition (3) can be interpreted as requiring a little more than continuity but a little less than differentiability:

Theorem 2

Suppose that functions f_i are continuous in some region D defined by $a \le x \le b,\ a_j < y_j < b_j$ for $j = 1, \dots n$, and that they have also continuous partial derivatives with respect to y_j, $j = 1, \dots, n$ there. Then the inequality (3) is satisfied in D with $L = \sup_D \|\frac{\partial f_i}{\partial y_j}\|$ and if $L < \infty$, then the assumptions of Theorem 1 hold in D.

Throughout the text, we assume that the problem (1) satisfies the assumptions of Theorem 1 in some region D and $(x^{(0)}, \mathbf{Y}^{(0)}) \in D$.

It also implicates that the problem is *well-posed*, which means (roughly speaking) that it's solution depends continuously on the given data: that a small perturbation of $\mathbf{F}$ or $\mathbf{Y}^{(0)}$ leads to a small change of the solution.

Global truncation error (GTE) of the approximate solution $\mathbf{Y}^{(k)}$ at $x^{(k)}$ is defined as

$$e_k = \| \mathbf{Y}^{(k)} \ - \ \mathbf{Y}(x^{(k)})\| . \qquad (4)$$

Local truncation error (LTE) at $x^{(k+1)}$ is defined as

$$E_{k+1} = \| \mathbf{Y}^{(k+1)} \ - \ \widehat{\mathbf{Y}}(x^{(k+1)})\| , \qquad (5)$$

where $\widehat{\mathbf{Y}}(x^{(k+1)})$ is the value of the exact solution for initial condition $\widehat{\mathbf{Y}}(x^{(k)}) = \mathbf{Y}^{(k)}$.

Order of a method: The method is of p-th order of accuracy, if some norm of a vector $\|\mathbf{e}\| = \|(e_1, \dots e_n)^T\|$ of the global errors (4) is $\mathcal{O}(h^p)$.

Note: Sometimes the method is said to be of p-th order, if LTE is $\mathcal{O}(h^{p+1})$. Both definitions are compatible: the global error at the n-th time step can be reasonably guessed to be n-times the LTE; then for fixed $a = x^{(0)}$ and $b = x^{(n)}$, n is proportional to $1/h$ and GTE at b can be expected to be proportional to $1/h$ times LTE, that is $\mathcal{O}(h^p)$.

How to compute the order of LTE and GTE will be illustrated for the explicit Euler method on a scalar equation as the simplest case of the system (1):

$$y'(x) = f(x, y(x)) \ , \qquad y(^{(0)}) = y^{(0)} \ . \qquad (6)$$

The technique is allways the same: use expansion in **Taylor series**.

Explicit Euler method: $x^{(k+1)} = x^{(k)} + h, \quad y^{(k+1)} = y^{(k)} + \ h \ f(x^{(k)}, y^{(k)})$

Local error (5) at $x^{(k+1)}$ is

$E_{k+1} = \| y^{(k+1)} - \ \widehat{y}(x^{(k+1)})\| = \| y^{(k)} + \ h \ f(x^{(k)}, y^{(k)}) \ - \ \widehat{y}(x^{(k+1)})\| \ ,$

where $\widehat{y}(x^{(k+1)})$ is the value of the exact solution for initial condition $\widehat{y}(x^{(k)}) = y^{(k)}$.

Taylor expansion at $\widehat{y}(x^{(k)})$ and using $y'(x) = f(x, y(x))$ and $\widehat{y}(x^{(k)}) = y^{(k)}$ gives

$\widehat{y}(x^{(k+1)}) \equiv \widehat{y}(x^{(k)} + h) = \widehat{y}(x^{(k)}) + h\, \widehat{y}\,'(x^{(k)}) + \mathcal{O}(h^2) = y^{(k)} + h\, f(x^{(k)}, y^{(k)}) + \mathcal{O}(h^2)$

and after substitution we have

$E_{k+1} = \| y^{(k)} + \ h \ f(x^{(k)}, y^{(k)}) \ - (\ y^{(k)} + h\, f(x^{(k)}, y^{(k)}) + \mathcal{O}(h^2) \,)\| = \mathcal{O}(h^2) \ .$

Conclusion: **Explicit Euler method is of the first order**.

Global error (4) at x_{k+1} is $e_{k+1} = \| y^{(k+1)} \ - \ y(x_{k+1})\|$.

(a) $y(x_{k+1}) = y(x_k) + h\, y'(x_k) + \mathcal{O}(h^2) \ \dots$ Taylor expansion of exact solution

(b) $y^{(k+1)} = y^{(k)} + \ h \ f(x_k, y^{(k)}) \ \dots$ Euler method

(a) - (b):

$y(x_{k+1}) - y^{(k+1)} = y(x_k) - y^{(k)} + h\, (f(x_k, y(x_k)) - f(x_k, y^{(k)})) + \mathcal{O}(h^2)$

$\qquad e_{k+1} \qquad \leq \qquad e_k \qquad + h\, \|f(x_k, y(x_k)) - f(x_k, y^{(k)})\| + c\, h^2$ for some $c \in R$.

The second term at the right hand side can be bounded using Lipschitz condition (3) as $\|f(x_k, y(x_k)) - f(x_k, y^{(k)})\| \leq L\, \|y(x_k) - y^{(k)}\| = L\, e_k$, and so $e_{k+1} \leq e_k\, (1 + h\, L) + c\, h^2$.

By recursion and using notation $a = (1 + h\, L)$ it follows

$e_1 \leq c\, h^2$ (the local error at the first step)
$e_2 \leq e_1\, a + c\, h^2 \leq a\, c\, h^2 + c\, h^2 = (a + 1)\, c\, h^2$
$e_3 \leq e_2\, a + c\, h^2 \leq (a + 1)\, c\, h^2\, a + c\, h^2 = (a^2 + a + 1)\, c\, h^2$
$\dots$
$e_n \leq (a^{n-1} + \dots \ a^2 + a + 1)\, c\, h^2 = \frac{a^n - 1}{a - 1}\, c\, h^2 = \frac{(1 + h\, L)^n - 1}{L}\, c\, h \leq (\mathrm{e}^{Lhn} - 1)\, \frac{1}{L}\, c\, h$
from which it follows that the global error is $\mathcal{O}(h)$ if $hn = |x_0 - x_n|$ is constant, i. e., if we consider some bounded interval only.

Collatz (or midpoint) method

Motivation: Euler (explicit) method is deduced by substitution of the first forward difference instead of the derivative on the left hand side of the differential equation. The first forward difference approximates the derivative with $\mathcal{O}(h)$ error:

$$\begin{aligned} y'(x^{(k)}) &= f(x^{(k)}, y(x^{(k)})) \\ \frac{y(x^{(k+1)}) - y(x^{(k)})}{h} + \mathcal{O}(h) &= f(x^{(k)}, y(x^{(k)})) \\ y(x^{(k+1)}) &= y(x^{(k)}) + h\, f(x^{(k)}, y(x^{(k)})) + \mathcal{O}(h^2) \\ y^{(k+1)} &= y^{(k)} + h\, f(x^{(k)}, y^{(k)}) \end{aligned}$$

What if we used the first central difference with $\mathcal{O}(h^2)$ error instead? As we do not want to involve values at other nodes than at $x^{(k)}$ and $x^{(k+1)}$, we need to use half-step for the central difference. Then

$$\begin{aligned} y'\left(x^{(k)} + \frac{h}{2}\right) &= f\left(x^{(k)} + \frac{h}{2},\ y\left(x^{(k)} + \frac{h}{2}\right)\right) \\ \frac{y(x^{(k+1)}) - y(x^{(k)})}{h} + \mathcal{O}(h^2) &= f\left(x^{(k)} + \frac{h}{2},\ y\left(x^{(k)} + \frac{h}{2}\right)\right) \\ y(x^{(k+1)}) &= y(x^{(k)}) + h\ f\left(x^{(k)} + \frac{h}{2},\ y\left(x^{(k)} + \frac{h}{2}\right)\right) + \mathcal{O}(h^3) \end{aligned}$$

At the right hand side, there is an unknown value of $y\left(x^{(k)} + \frac{h}{2}\right)$ which we want to approximate by values of $x^{(k)}$ and $y^{(k)}$ in order to obtain an explicit method. Using forward Euler method with half step we have

$$y\left(x^{(k)} + \tfrac{h}{2}\right) \approx y^{(k)} + \tfrac{h}{2}\, f(x^{(k)}, y^{(k)})$$

and so the final formula for the **Collatz method** is

$$y^{(k+1)} = y^{(k)} + h\ f\left(x^{(k)} + \frac{h}{2},\ y^{(k)} + \frac{h}{2}\, f(x^{(k)}, y^{(k)})\right). \tag{7}$$

Local error (5) at x_{k+1} is $E_{k+1} = \|\, y^{(k)} + h\ f(x_k + \frac{1}{2}\, h, y^{(k)} + \frac{1}{2}\, h\, f(x_k, y^{(k)})) - \widehat{y}(x_{k+1})\|$,

where $\widehat{y}(x_{k+1})$ is the value of the exact solution for initial condition $\widehat{y}(x_k) = y^{(k)}$.

Taylor expansion for $\widehat{y}(x_{k+1})$ gives

$\widehat{y}(x_{k+1}) \equiv \widehat{y}(x_k + h) = \widehat{y}(x_k) + h\, \widehat{y}\,'(x_k) + \frac{1}{2}\, h^2\, \widehat{y}\,''(x_k) + \mathcal{O}(h^3)$,

then substitute $y'(x) = f(x, y(x))$ and $y''(x) = \dfrac{d}{dx} f = \dfrac{\partial f}{\partial x}\dfrac{dx}{dx} + \dfrac{\partial f}{\partial y}\dfrac{dy}{dx} = \dfrac{\partial f}{\partial x} + \dfrac{\partial f}{\partial y} f$ (chain rule), to obtain

$$\begin{aligned} \widehat{y}(x_{k+1}) &= \widehat{y}(x_k) + h\, f(x_k, \widehat{y}(x_k)) + \tfrac{1}{2}\, h^2 \left[\frac{\partial f}{\partial x}(x_k, \widehat{y}(x_k)) + \frac{\partial f}{\partial y}(x_k, \widehat{y}(x_k))\, f(x_k, \widehat{y}(x_k))\right] + \mathcal{O}(h^3) = \\ &= \widehat{y}(x_k) + h \left[f(x_k, \widehat{y}(x_k)) + \frac{\partial f}{\partial x}(x_k, \widehat{y}(x_k))\, \tfrac{1}{2}\, h + \frac{\partial f}{\partial y}(x_k, \widehat{y}(x))\, \tfrac{1}{2}\, h\, f(x_k, \widehat{y}(x_k))\right] + \mathcal{O}(h^3). \end{aligned}$$

Now compare the term in the brackets with the right hand side of the Taylor expansion

$$f(x_k + h_1, \widehat{y}(x_k) + h_2) = f(x_k, \widehat{y}(x_k)) + \frac{\partial f}{\partial x}(x_k, \widehat{y}(x_k))\, h_1 + \frac{\partial f}{\partial y}(x_k, \widehat{y}(x))\, h_2 + \mathcal{O}(h_1^2 + h_1 h_2 + h_2^2)$$

– choice $h_1 = \frac{1}{2}\,h$ and $h_2 = \frac{1}{2}\,h\,f(x_k, \widehat{y}(x_k))$ enables replacing that term by

$f(x_k + h_1, \widehat{y}(x_k) + h_2) + \mathcal{O}(h^2)$, which after substitution of $\widehat{y}(x_k) = y^{(k)}$ and rearranging leads to

$$\widehat{y}(x_{k+1}) = y^{(k)} + h\,\left[f(x_k + h_1, y^{(k)} + h_2) + \mathcal{O}(h^2)\right] + \mathcal{O}(h^3) =$$

$$= y^{(k)} + h\,f(x_k + \tfrac{1}{2}\,h, y^{(k)} + \tfrac{1}{2}\,h\,f(x_k, \widehat{y}(x_k)) + \mathcal{O}(h^3)\ ,$$

$$E_{k+1} = \|\, y^{(k)} + h\ f(x_k + \tfrac{1}{2}\,h, y^{(k)} + \tfrac{1}{2}\,h\,f(x_k, y^{(k)}))\ - (\ y^{(k)} + h\,f(x_k + h_1, y^{(k)} + h_2) + \mathcal{O}(h^3)\,)\| = \mathcal{O}(h^3)$$

.

The local error of Collatz method is $\mathcal{O}(h^3)$.

Conclusion: **Collatz (midpoint) method is of the second order**.

It can be shown that global error of Collatz method is $\mathcal{O}(h^2)$, using similar technique as for explicit Euler method.

Collatz (or midpoint) method for a system of equations

choose a step size h and for $i = 0, 1, 2, \ldots$

1. compute an auxiliary point $[x_p, \mathbf{Y}_p]$ using forward Euler method with half-step:
   $\mathbf{K}_1 = \mathbf{F}(x^{(i)},\, \mathbf{Y}^{(i)})$
   $x_p = x^{(i)} + \frac{1}{2}h$
   $\mathbf{Y}_p = \mathbf{Y}^{(i)} + \frac{1}{2}h\,\mathbf{K}_1$
2. compute the derivative $\mathbf{K}_2$ at the auxiliary point $[x_p, \mathbf{Y}_p]$ as
   $\mathbf{K}_2 = \mathbf{F}(x_p,\, \mathbf{Y}_p)$
3. set
   $x^{(i+1)} = x^{(i)} + h$
   $\mathbf{Y}^{(i+1)} = \mathbf{Y}^{(i)} + h\,\mathbf{K}_2$

Example 1 - from the previous tutorial, continued. Consider Cauchy problem

$$y' = \frac{y}{x^2}\,, \quad y(1) = 2\ .$$

Compute an approximate value of $y(1.4)$ using Collatz method with step size $h = 0.2$ and compare its performance with previous results of Euler method.

The solution

The results are summarized in Table 1. These results show that Collatz method gives more precise solution than Euler method (both explicit and implicit), even in the case when for Collatz method, step size twice as long as for Euler was used (which represents comparable work).

Computation:

$h = 0.2, \quad x^{(0)} = 1, \quad y^{(0)} = 2$

$k_1 \equiv f(x^{(0)}, y^{(0)}) = \dfrac{y^{(0)}}{(x^{(0)})^2} = \dfrac{2}{1^2} = 2,$

$x_p = x^{(0)} + \frac{1}{2}h = 1 + 0.1 = 1.1, \quad y_p = y^{(0)} + \frac{1}{2}h\,k_1 = 2 + 0.1 \cdot 2 = 2.2$

$k_2 \equiv f(x_p, y_p) = \dfrac{y_p}{x_p^2} = \dfrac{2.2}{1.1^2} = 1.8182$

$x^{(1)} = x^{(0)} + h = 1 + 0.2 = 1.2, \quad y^{(1)} = y^{(0)} + h\,k_2 = 2 + 0.2 \cdot 1.8182 = 2.3636$

$k_1 \equiv f(x^{(1)}, y^{(1)}) = \dfrac{y^{(1)}}{(x^{(1)})^2} = \dfrac{2.3636}{1.2^2} = 1.6414$

$x_p = x^{(1)} + \frac{1}{2}h = 1.2 + 0.1 = 1.3$

$y_p = y^{(1)} + \frac{1}{2}h\,k_1 = 2.3636 + 0.1 \cdot 1.6414 = 2.5278$

$k_2 \equiv f(x_p, y_p) = \dfrac{y_p}{x_p^2} = \dfrac{2.5278}{1.3^2} = 1.4957$

$x^{(2)} = x^{(1)} + h = 1.2 + 0.2 = 1.4$

$y^{(2)} = y^{(1)} + h\,k_2 = 2.3636 + 0.2 \cdot 1.4957 = 2.6628$

$y(1.4)$ is approximately equal to $y^{(2)} = 2.6628$.

$x^{(i)}$	exact $y(x^{(i)})$	Euler Explic.	$h = 0.1$ Implic.	Euler Explic.	$h = 0.2$ Implic.	Collatz $h = 0.2$
1	2.0000	2.0000	2.0000	2.0000	2.0000	2.0000
1.1	2.1903	2.2000	2.1802			(2.2000)
1.2	2.3627	2.3818	2.3429	2.4000	2.3226	2.3636
1.3	2.5191	2.5472	2.4902			(2.5278)
1.4	2.6614	2.6979	2.6241	2.7333	2.5865	2.6628
1.5	2.7912	2.8356	2.7462			(2.7986)
1.6	2.9100	2.9616	2.8578	3.0122	2.8057	2.9115
1.7	3.0190	3.0773	2.9602			(3.0253)
1.8	3.1192	3.1838	3.0545	3.2476	2.9903	3.1209
1.9	3.2118	3.2821	3.1415			(3.2172)
2.0	3.2974	3.3730	3.2221	3.4480	3.1477	3.2992

Table 1: **Example 1**. The first column represents values of x, where the approximate solution is computed. At the second column there is exact solution, the third and the fourth columns present solution obtained by Euler method with step size $h = 0.1$ and $h = 0.2$, respectively, and the last column presents approximate solution obtained by Collatz method with step size $h = 0.2$.

Example 2 - from the previous tutorial, solved using Collatz method

Consider Cauchy problem

$$\mathbf{Y}' = \begin{bmatrix} y_1 \sin(x) + y_3 \\ y_2 \ln(x+1) - 4 \\ 2y_1 - \dfrac{y_3}{x-2} \end{bmatrix}, \qquad \mathbf{Y}(1) = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix}$$

Choose a step size $h = 0.2$ and compute an approximate value of $\mathbf{Y}(1.2)$ using Collatz method.

The solution

$x^{(0)} = 1$, $\mathbf{Y}^{(0)} = (-1, 1, 2)^T$, $h = 0.2$:

$$\mathbf{K}_1 = \mathbf{F}(x^{(0)}, \mathbf{Y}^{(0)}) = \begin{bmatrix} -1 \cdot \sin(1) + 2 \\ 1 \cdot \ln(1+1) - 4 \\ 2 \cdot (-1) - \dfrac{2}{1-2} \end{bmatrix} = \begin{bmatrix} -0.84147 + 2 \\ 0.69315 - 4 \\ -2 + 2 \end{bmatrix} = \begin{bmatrix} 1.1585 \\ -3.3068 \\ 0 \end{bmatrix}$$

$$x_p = x^{(0)} + \tfrac{1}{2}\, h = 1 + 0.1 = 1.1$$

$$\mathbf{Y}_p = \mathbf{Y}^{(0)} + \frac{1}{2}\, h\, \mathbf{K}_1 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} + 0.1 \begin{bmatrix} 1.1585 \\ -3.3068 \\ 0 \end{bmatrix} = \begin{bmatrix} -0.8842 \\ 0.6693 \\ 2 \end{bmatrix}$$

$$\mathbf{K}_2 = \mathbf{F}(x_p, \mathbf{Y}_p) = \begin{bmatrix} -0.8842 \cdot \sin(1.1) + 2 \\ 0.6693 \cdot \ln(1.1+1) - 4 \\ 2 \cdot (-0.8842) - \dfrac{2}{1.1-2} \end{bmatrix} = \begin{bmatrix} -0.7880 + 2 \\ 0.6693 \cdot 0.74194 - 4 \\ -1.7684 + 2.2222 \end{bmatrix} = \begin{bmatrix} 1.2120 \\ -3.5034 \\ 0.45380 \end{bmatrix}$$

$$\mathbf{Y}^{(1)} = \mathbf{Y}^{(0)} + h\, \mathbf{K}_2 = \begin{bmatrix} -1 \\ 1 \\ 2 \end{bmatrix} + 0.2 \begin{bmatrix} 1.2120 \\ -3.5034 \\ 0.4538 \end{bmatrix} = \begin{bmatrix} -0.7576 \\ 0.2993 \\ 2.091 \end{bmatrix}$$

The value of $\mathbf{Y}(1.2)$ is approximately $\mathbf{Y}^{(1)} = (-0.7782,\ 0.2993,\ 2.091)^T$.