Numerical Solution of Differential Equations

Mirko Navara
http://cmp.felk.cvut.cz/~navara/
Center for Machine Perception, Department of Cybernetics, FEE, CTU
Karlovo náměstí, building G, office 104a

http://math.feld.cvut.cz/nemecek/nummet.html

January 14, 2016

Restriction: ordinary (not partial) differential equations, Cauchy initial value problem, only one differential equation of the first order
Task: On interval $[x_0, x_n]$, solve differential equation

$$y'(x) = f(x, y(x))$$

with initial condition

$$y(x_0) = y_0\,,$$

where f is a function of two variables and $y_0 \in \mathbb{R}$.

Comment: If f does not depend on y, i.e., $f(x, y) = g(x)$, we get numerical integration as a special case—differential equation

$$y'(x) = g(x)$$

Existence and uniqueness of the solution

It is not guaranteed in general:
Example: Consider differential equation with initial condition:

$$y'(x) = \sqrt[3]{y(x)}, \qquad y(0) = 0\,,$$

where the third root is a real function defined also for negative arguments. It has solutions, e.g., $y(x) = 0$ and $y(x) = \pm\left(\frac{2}{3}\,x\right)^{\frac{3}{2}}$.
Theorem: Let function f be defined and continuous at $[x_0, x_n] \times \mathbb{R}$ (e.g., for all $x \in [x_0, x_n]$, $y \in \mathbb{R}$).
Let the **Lipschitz condition**

$$\exists L \in \mathbb{R}\ \forall x \in [x_0, x_n]\ \forall y_1, y_2 \in \mathbb{R} : |f(x, y_1) - f(x, y_2)| \le L\,|y_1 - y_2|$$

be satisfied. Then the solution on $[x_0, x_n]$ exists and it is unique.
Sufficient condition: $\frac{\partial f}{\partial y}$ continuous and bounded on $[x_0, x_n] \times \mathbb{R}$.

Interpretation of the problem and principle of solution

Comment: Equivalent formulation of the problem: Solution

$$y(x) = y_0 + \int_{x_0}^{x} f(t, y(t))\,dt$$

may be understood as an integral of an (unknown) function $g(t) = f(t, y(t))$ of one variable or as a curve integral of a known function f along an (unknown) curve with parametrization $(t, y(t))$, $t \in [x_0, x_n]$.
We split interval $[x_0, x_n]$ to n subintervals of length $h = (x_n - x_0)/n$. We get **nodes** $x_i = x_0 + i\,h$, $i = 0, \ldots, n$.
Correct values at nodes, $y(x_i)$, are replaced by their estimates y_i.
Values of the derivative: $f_i = f(x_i, y_i)$.

General principle of solution

We generate a sequence y_i, $i = 0, \ldots, n$. In step $i+1$, we use estimates $y_0, \ldots, y_i$ for estimation of y_{i+1}. Exact solution

$$y(x_{i+1}) - y(x_i) = \int_{x_i}^{x_{i+1}} f(t, y(t))\, dt$$

is estimated by

$$\Delta y_i = y_{i+1} - y_i \approx \int_{x_i}^{x_{i+1}} f(t, y(t))\, dt\,.$$

$$y_{i+1} = y_i + \Delta y_i\,.$$

Particular methods differ only by the estimate Δy_i.

Runge–Kutta methods 1: Euler's method

It is a generalization of the left sum method of integration; function $f(t, y(t))$ is replaced by its value $f(x_i, y_i)$ at x_i

$$\Delta y_i = \int_{x_i}^{x_{i+1}} f(x_i, y_i)\, dt = h\, f(x_i, y_i)\,,$$

$$y_{i+1} = y_i + h\, f(x_i, y_i) = y_i + h\, f_i\,.$$

Geometrical interpretation: $f_i = f(x_i, y_i)$ is the slope of the line segment with endpoints (x_i, y_i), (x_{i+1}, y_{i+1}).

Estimate of the error

Evaluate the Taylor expansion of function y with center x_0 at x_1:

$$y(x_1) = y(x_0) + h\, y'(x_0) + \frac{h^2}{2}\, y''(\xi)\,,$$

where $\xi \in [x_0, x_1]$.

$$y(x_1) = \underbrace{y(x_0) + h\, f(x_0, y_0)}_{y_1} + \frac{h^2}{2}\, y''(\xi)\,,$$

$$y(x_1) - y_1 = \frac{h^2}{2}\, y''(\xi)\,.$$

The error at the end of the first step is proportional to h^2.
In subsequent steps, we use an initial condition which is not exact. Nevertheless, usually the error is proportional to h^2 and the number of steps $n = \frac{x_n - x_0}{h}$.
The error at the end of the interval is proportional to $\frac{1}{h}\, h^2 = h$ $\quad \Rightarrow$ method of the 1st order.

Runge–Kutta methods 2: First modification of Euler's method

Generalization of rectangular (midpoint) integration method; we approximate function $f(t, y(t))$ by its value at $\frac{x_i + x_{i+1}}{2} = x_i + \frac{h}{2}$. The second argument of f is the result of an auxiliary step of length $h/2$, made by Euler's method:

$$\eta_i = y_i + \frac{h}{2}\, f_i\,.$$

$$f(t, y(t)) \approx f(x_i + \frac{h}{2}, \eta_i)$$

$$\Delta y_i = \int_{x_i}^{x_{i+1}} f\left(x_i + \frac{h}{2}, \eta_i\right)\, dt = h\, f\left(x_i + \frac{h}{2}, \eta_i\right)\,.$$

Method of 2nd order.

Second modification of Euler's method (Heun's method)

Generalization of trapezoidal integration method; we approximate function $f(t, y(t))$ by a linear function going through the endpoints of the interval:
at x_i: $f_i = f(x_i, y_i)$,
at x_{i+1}: the lack of knowledge of the yth coordinate is compensated by an auxiliary step of length $h/2$, made by Euler's method:

$$\theta_i = y_i + h\, f_i\,.$$

Function $f(t, y(t))$ is approximated by a linear function whose graph goes through points $\big(x_i, f(x_i, y_i)\big)$, $\big(x_{i+1}, f(x_{i+1}, \theta_i)\big)$.

$$\Delta y_i = \frac{h}{2}\left(f(x_i, y_i) + f(x_{i+1}, \theta_i)\right).$$

Method of 2nd order.

Runge–Kutta methods 4: Runge–Kutta method of 4th order

Generalization of Simpson's method; we first compute auxiliary points and derivatives in them,

$$\begin{aligned}
k_{i,1} &= f(x_i, y_i)\,,\\
k_{i,2} &= f\left(x_i + \frac{h}{2}, y_i + \frac{h}{2}\,k_{i,1}\right),\\
k_{i,3} &= f\left(x_i + \frac{h}{2}, y_i + \frac{h}{2}\,k_{i,2}\right),\\
k_{i,4} &= f(x_i + h, y_i + h\,k_{i,3})\,.
\end{aligned}$$

The integral is approximated by a linear combination of these values:

$$\Delta y_i = \frac{h}{6}\left(k_{i,1} + 2k_{i,2} + 2k_{i,3} + k_{i,4}\right).$$

Runge–Kutta methods 5: General Runge–Kutta methods

They estimate the integral $\int_{x_i}^{x_{i+1}} f(t, y(t))\,dt$ from several values of function f at points, obtained from the initial values x_i, y_i and auxiliary steps. These values are combined so that the errors of the lowest orders are compensated.

Mulitistep methods

Methods

- one-step: they use only x_i, y_i and $f_i = f(x_i, y_i)$
  (e.g. Runge–Kutta),
- mulitistep: they use also the results of previous steps, i.e., x_j, y_j a $f_j = f(x_j, y_j)$, $j = i, i-1, \ldots, i-s+1$
  (for an s-step method).

Mulitistep methods admit to increase the order without auxiliary steps.
However, the initialization of an s-step method requires s values $y_0, y_1, \ldots, y_{s-1}$. These are obtained by a **starting method** (one-step).

Adams–Bashforth methods (explicit)

We approximate s values of the derivative, $f_i, f_{i-1}, \ldots, f_{i-s+1}$
at nodes $x_i, x_{i-1}, \ldots, x_{i-s+1}$
by the interpolating polynomial φ_i, which is integrated instead of $f(t, y(t))$:

$$\Delta y_i = \int_{x_i}^{x_{i+1}} \varphi_i(t)\,dt\,.$$

We do not need to compute φ_i because

$$\Delta y_i = h \sum_{j=0}^{s-1} w_j\, f_{i-j}\,,$$

where w_j are known coefficients.
We use polynomial approximation of the derivative $y'(t) = f(t, y(t))$, not of the solution, $y(t)$!

For $s = 1$:
$\varphi_i = f_i$ is constant $\Rightarrow$ Euler's method.
For $s = 2$:
φ_i is a linear polynomial going through points (x_i, f_i), (x_{i-1}, f_{i-1}),

$$\varphi_i(t) = f_i + \frac{f_i - f_{i-1}}{h}\,(t - x_i)$$

$$\Delta y_i = \int_{x_i}^{x_{i+1}} \varphi_i(t)\,dt = h\,f_i + \frac{h}{2}\,(f_i - f_{i-1}) = \frac{h}{2}\,(3f_i - f_{i-1})\,.$$

For $s = 3$:

$$\Delta y_i = \frac{h}{12}\,(23f_i - 16f_{i-1} + 5f_{i-2})\,,$$

For $s = 4$:

$$\Delta y_i = \frac{h}{24}\,(55f_i - 59f_{i-1} + 37f_{i-2} - 9f_{i-3})\,.$$

The order of these methods is s=number of points used in the approximation.
Advantage:

- simplicity

Disadvantages:

- different signs of coefficients ($\Rightarrow$ round-off errors)
- systematic error caused by the polynomial extrapolation

$\Rightarrow$ effort to avoid extrapolation

Adams–Moulton methods (implicit)

Function $f(t, y(t))$ is approximated by an interpolating polynomial φ_i through values $f_i, f_{i-1}, \ldots, f_{i-s+1}$ and value at x_{i+1}, i.e., $f_{i+1} = f(x_{i+1}, y_{i+1})$.
Again it reduces to a linear combination

$$y_{i+1} - y_i = \Delta y_i = h \sum_{j=-1}^{s-1} w_j\,f_{i-j}\,,$$

where w_j are known coefficients (different from the above).
We obtain an **equation**

$$y_{i+1} = y_i + h\,w_{-1}\,f(x_{i+1}, y_{i+1}) + h \sum_{j=0}^{s-1} w_j\,f_{i-j}$$

for an unknown value y_{i+1}, which is determined only **implicitly**.

For $s = 1$: φ_i is a linear polynomial going through points (x_i, f_i), (x_{i+1}, f_{i+1}), e.g.,

$$\varphi_i(t) = f_i + \frac{f_{i+1} - f_i}{h}\,(t - x_i)\,.$$

$$\Delta y_i = \int_{x_i}^{x_{i+1}} \varphi_i(t)\,dt = \frac{h}{2}\,(f_{i+1} + f_i)\,,$$

with substitution $f_{i+1} = f(x_{i+1}, y_{i+1})$

$$y_{i+1} - y_i = \frac{h}{2}\,\big(f(x_{i+1}, y_{i+1}) + f_i\big)\,.$$

For $s = 2$:

$$\Delta y_i = \frac{h}{12}\,\big(5f(x_{i+1}, y_{i+1}) + 8f_i - f_{i-1}\big)\,,$$

For $s = 3$:

$$\Delta y_i = \frac{h}{24}\left(9f(x_{i+1}, y_{i+1}) + 19f_i - 5f_{i-1} + f_{i-2}\right).$$

The order of these methods is $s + 1$=number of points used in the approximation.
Advantage:

- higher precision

Disadvantages:

- difficult solution of the implicit equation (an analytical solution is usually impossible, numerical solution increases the computational complexity)
- even the polynomial **interpolation** can cause large systematic errors

Predictor–corrector methods

Based on a **corrector**, which might be some of implicit methods, in which the corresponding equation is solved numerically.
In the mth iteration, we compute an estimate $y_{i+1,m}$ of y_{i+1}, where we use the results of the preceding iteration, $y_{i+1,m-1}$, on the right-hand side:

$$y_{i+1,m} = y_i + h\sum_{j=0}^{s-1} w_j\, f_{i-j} + h\, w_{-1}\, f(x_{i+1}, y_{i+1,m-1}).$$

The initial estimate $y_{i+1,0}$ is computed from the results of previous steps using another method, **predictor**, e.g., some of explicit methods.

Control mechanism

P = Predictor
C = Corrector
E = Evaluation
Most common choices:

- The cycle of the corrector is repeated until the difference $y_{i+1,m} - y_{i+1,m-1}$ is sufficiently small.
- Constant number k of repetitions of the corrector, P(EC)^{k}E.
- Single use of the corrector, PECE.

Adams methods

Predictor: Adams–Bashforth method
Corrector: Adams–Moulton method
Example: The simplest variant of an Adams method, $s = 1$:
Predictor: Euler's method (1st order)

$$y_{i+1,0} = y_i + h\, f_i.$$

Corrector: Adams–Moulton method of 2nd order

$$y_{i+1,m} = y_i + \frac{h}{2}\left(f_i + f(x_{i+1}, y_{i+1,m-1})\right).$$

Choice of starting methods (of their order)
Step size control

Richardson's extrapolation in solution of differential equations

$\tilde{y}(x,h)$ ... numerical solution at x with step h
$\tilde{y}(x,2h)$ ... numerical solution at x with step $2h$
(here $q=2$)
The error of estimate $\tilde{y}(x,h)$ is approximately $2^p\times$ smaller than the error of estimate $\tilde{y}(x,2h)$
$\Rightarrow$ **estimate of the error of** $\tilde{y}(x,h)$:

$$\tilde{y}(x,h) - y(x) \approx \frac{1}{2^p - 1}(\tilde{y}(x,2h) - \tilde{y}(x,h))\,.$$

Estimate improved by Richardson's extrapolation:

$$y(x) \approx \tilde{y}(x,h) + \frac{1}{2^p - 1}(\tilde{y}(x,h) - \tilde{y}(x,2h))\,.$$

Richardson's extrapolation

- passive
- active