euler's method for ODEs

Fall semester 2024

https://www.phys.uconn.edu/~rozman/Courses/P2200_24F/

Last modified: September 30, 2024

Introduction

We are interested in the numerical solution of the following initial value problem (IVP) for an ordinary differential equation:

$$\frac{\mathrm{d}y}{\mathrm{d}t} = f(t,y), \quad a \le t \le b, \quad y(a) = y_1. \tag{1}$$

The idea is to start from $t = a$ (since we know $y(a)$), increment t by sufficiently small integration step h, and use Eq. (1) to determine $y(t+h)$. The process is then repeated until we reach $t = b$.

We denote the value of independent variable at the ith integration step by t_{i+1}, $i = 1,2,\ldots$, $t_1 = a$; the computed solution at the ith step by y_{i+1},

$$y_{i+1} \equiv y(t_{i+1}), \quad i = 1,\ldots,n-1; \tag{2}$$

the value of the right hand side of Eq. (1) at the ith integration step by f_{i+1},

$$f_{i+1} \equiv f(t_{i+1}, y_{i+1}). \tag{3}$$

The step size h (assumed to be a constant for the sake of simplicity) is:

$$h = t_i - t_{i-1} = \frac{b-a}{n-1}. \tag{4}$$

Here $n-1$ is the total number of integration steps (corresponding to n function evaluations of the right hand side of Eq. (1)).

The error that is induced at every time-step, ϵ_i, is referred to as the *local truncation error* (LTE) of the method. The local truncation error is different from the *global error* g_n, which is defined as the absolute value of the difference between the true solution and the computed solution,

$$g_n = \left| y_{\text{exact}}(t_n) - y_n \right|. \tag{5}$$

In most cases, we do not know the exact solution and hence cannot evaluate the global error. However, it is reasonable to assume that the global error at the nth time step is n times the LTE. Since h is proportional to $\frac{1}{n}$ (i.e. $n \sim \frac{1}{h}$ for $n \gg 1$, g_n should be proportional to $\frac{\epsilon}{h}$. A method with $\epsilon \sim h^{k+1}$ is said to be of kth order. This implies that for a kth order method, the global error scales as h^k.

Euler's method

The Taylor series expansion of $y(t_{j+1})$ about t_j correct up to the h^2 term is as following,

$$y(t_{j+1}) = y(t_j + h) = y(t_j) + h\,\frac{\mathrm{d}y}{\mathrm{d}t}\bigg|_{t_j} + \frac{h^2}{2}\,\frac{\mathrm{d}^2y}{\mathrm{d}t^2}\bigg|_{t_j} + O(h^3). \tag{6}$$

Using Eq. (1) for $\dfrac{\mathrm{dy}}{\mathrm{d}t}$,

$$y(t_{j+1}) = y(t_j) + h\,f(t_j, y_j) + \alpha\frac{h^2}{2} + O(h^3), \tag{7}$$

or

$$y_{j+1} = y_j + hf_j + \alpha\frac{h^2}{2} + O(h^3), \tag{8}$$

where α is an unknown constant.

Ignoring the quadratic in h and higher order terms, we obtain the expression for Euler's integration step:

$$y_{j+1} = y_j + hf_j. \tag{9}$$

In addition to deriving Eq. (9), we learned that the leading in h error term dropped in Eq. (9) is quadratic in h, therefore Euler's method is a first order method.

Richardson extrapolation

Based on out knowledge that the local truncation error for the Euler's method is $O(h^2)$, let's use Richardson extrapolation to construct a an integrator with a smaller truncation error than $O(h^2)$.

The local error of Euler's method of the step h is

$$y_{exact}(t+h) - \mathrm{Euler}_h(t+h) = \alpha\frac{h^2}{2}. \tag{10}$$

The local error of Euler's method of two steps of $h/2$ is twice as small:

$$y_{exact}(t+h) - \mathrm{Euler}_{h/2}(t+h) = 2\alpha \frac{(h/2)^2}{2} = \alpha \frac{h^2}{4}. \tag{11}$$

Combining Eq. (10) and Eq. (11), we can eliminate the leading error term, obtaining

$$y_{exact}(t+h) - 2\mathrm{Euler}_{h/2}(t+h) + \mathrm{Euler}_h(t+h) = O(h^3). \tag{12}$$

Therefore the integration method

$$y(t+h) = 2\mathrm{Euler}_{h/2}(t+h) - \mathrm{Euler}_h(t+h) \tag{13}$$

has the local truncation error $O(h^3)$.

Explicitly,

$$\begin{aligned}
y_{j+1} &= y_j + h f(t_j, y_j), && (14)\\
y_{j+1/2} &= y_j + \frac{h}{2} f(t_j, y_j), && (15)\\
y_{j+1/2+1/2} &= y_{j+1/2} + \frac{h}{2} f(t_{j+1/2}, y_{j+1/2}) \\
&= y_j + \frac{h}{2} f(t_j, y_j) + \frac{h}{2} f\left(t_j + h/2, y_j + \frac{h}{2} f(t_j, y_j)\right), && (16)
\end{aligned}$$

The method that we obtained is called *midpoint method*:

$$y_{j+1} = 2y_{j+1/2+1/2} - y_{j+1} = y_j + hf(t_j + h/2, y_j + h/2\, f_j). \tag{17}$$