Numerical Integration of Differential Equations

Methods with uniform step size

We consider methods to numerically integrate the initial value problem

$$y' = f(t, y), \qquad\qquad y(a) = y_0 \tag{1}$$

on the interval $a \le t \le b$. We divide the interval $[a, b]$ into N subintervals of length $h = (b - a)/N$, and let $t_n = a + nh, n = 0, 1, \ldots, N$. We let $y(t)$ denote the exact solution of the IVP (1). The value of y at the point t_n is $y(t_n)$. The values of the numerical solution at the point t_n we denote by y_n. The error at the nth step is $e_n = y(t_n) - y_n$.

The Euler Method

The Euler method is the simplest example of an *explicit* method. In the Euler Method we calculate the values y_n by making tangent line approximations:

$$y_{n+1} = y_n + hf(t_n, y_n). \tag{2}$$

To estimate the error we expand the exact solution $y(t)$ at the point t_n:

$$\begin{aligned} y(t) &= y(t_n) + y'(t_n)(t - t_n) + O((t - t_n)^2) \\ &= y(t_n) + f(t_n, y(t_n))(t - t_n) + O((t - t_n)^2). \end{aligned}$$

Putting $t = t_{n+1} = t_n + h$, we find

$$y(t_n + h) = y(t_n) + f(t_n, y(t_n))h + O(h^2). \tag{3}$$

Now subtract (2) from (3) to find

$$\begin{aligned} e_{n+1} &= e_n + [f(t_n, y(t_n)) - f(t_n, y_n)]h + O(h^2) \\ &= e_n[1 + f_y(t_n, \eta_n)h] + O(h^2) \end{aligned} \tag{4}$$

where η_n is some point in the interval between y_n and $y(t_n)$. If $y(t_n) = y_n$, so that $e_n = 0$, then $e_{n+1} = O(h^2)$. This is the local error, or formula error. If we assume that there is a constant $L \ge 0$ such that $|f_y(t, y)| \le L$, and that $|y''(t)| \le M$, using (4) we can prove that Euler's method converges in the sense that the cumulative error

$$|y(b) - y_N| \le Ch.$$

This means that as $N \to \infty$, i.e., $h \to 0$, the value y_N converges to $y(b)$. We say that Euler's method has order $p = 1$ because the cumulative error tends to zero as the first power of h.

Stability We say that the ODE (1) is *stable* if $f_y \le 0$. Intuitively, this means that the solution curves are not spreading apart as t increases. Note that equation may stable for some values of (t, y) and not for other values.

From (4) we see that the factor $[1 + hf_y]$ multiplies the error e_n This factor is the amplification (magnifcation) factor of the Euler method. If $f_y > 0$, i.e., if the ODE is unstable, the error is amplified. We will say that a *method* (like Euler's method) is *stable* if the amplication factor is less than or equal to 1 in absolute value. In the case of Euler's method we require that

$$|1 + hf_y| \leq 1.$$

The interval of stability for Euler's method is

$$-2 \leq hf_y \leq 0.$$

Thus for Euler's method to be stable we must have

$$h \leq \frac{2}{|f_y|}.$$

Backward Euler Method

The backward Euler method is an *implicit* method. It is given by

$$y_{n+1} = y_n + hf(t_{n+1}, y_{n+1}). \tag{5}$$

The value y_{n+1} is defined implicitly by the equation (5). It must be determined at each step by some numerical method of solving equations. To estimate the error e_{n+1} in terms of e_n, we expand the exact solution of (1) at the point t_{n+1}:

$$\begin{aligned} y(t) &= y(t_{n+1}) + y'(t_{n+1})(t - t_{n+1}) + O((t - t_{n+1})^2) \\ &= y(t_{n+1}) + f(t_{n+1}, y(t_{n+1}))(t - t_{n+1}) + O((t - t_{n+1})^2). \end{aligned}$$

Now set $t = t_n = t_{n+1} - h$, to deduce

$$y(t_n) = y(t_{n+1}) - hf(t_{n+1}, y(t_{n+1})) + O(h^2),$$

whence

$$y(t_{n+1}) = y(t_n) + hf(t_{n+1}, y(t_{n+1})) + O(h^2). \tag{6}$$

Now subtract (5) from (6) to find

$$\begin{aligned} e_{n+1} &= e_n + h[f(t_{n+1}, y(t_{n+1})) - f(t_{n+1}, y_{n+1})] + O(h^2) \\ &= e_n + hf_y(t_{n+1}, \eta_n)e_{n+1} + O(h^2). \end{aligned}$$

When we solve for e_{n+1} we find

$$e_{n+1} = \left(\frac{1}{1 - hf_y}\right)e_n + O(h^2). \tag{7}$$

It follows that if the equation is stable, i.e., $f_y \leq 0$, the the amplification factor $0 \leq 1/(1 - hf_y) \leq 1$. Thus the backward Euler method is stable whenever the ODE is stable.

The explicit trapezoid method

This method is motivated by the trapezoid approximation to the integral

$$\begin{aligned} y(t+h) &= y(t) + \int_t^{t+h} f(s, y(s))ds \\ &\approx y(t) + (h/2)[f(t, y(t)) + f(t+h, y(t+h))]. \end{aligned}$$

The explicit trapezoid method is a two stage procedure: In the first stage we take the slope to be that of the Euler method,

$$s_1 = f(t_n, y_n)$$

and in the second stage we take the slope to be

$$s_2 = f(t_{n+1}, y_n + hs_1).$$

For our formula we use an average of these values:

$$y_{n+1} = y_n + \frac{h}{2}(s_1 + s_2). \tag{8}$$

To derive an estimate of the error e_{n+1} in terms of the error e_n, we expand the exact solution $y(t)$ at $t = t_n$ up to second order:

$$y(t) = y(t_n) + y'(t_n)(t - t_n) + \frac{1}{2}y''(t_n)(t - t_n)^2 + O((t - t_n)^3). \tag{9}$$

Now we use the fact that for the exact solution of (1), $y' = f(t, y(t))$, and

$$y''(t) = f_t(t, y(t)) + f_y(t, y(t))f(t, y(t)).$$

Substituting this into equation (9) with $t = t_{n+1} = t_n + h$, we obtain

$$\begin{aligned} y(t_{n+1}) &= y(t_n) + hf(t_n, y(t_n)) \\ &+ \frac{h^2}{2}[f_t(t_n, y(t_n)) + f_y(t_n, y(t_n))f(t_n, y(t_n))] + O(h^3). \end{aligned} \tag{10}$$

To compare this formula with (9), we must expand s_2 about t_n:

$$\begin{aligned} s_2 &= f(t_n, y_n) + h[f_t(t_n, y_n) + f_y(t_n, y_n)hs_1] + O(h^2) \\ &= f(t_n, y_n) + hf_t(t_n, y_n) + hf_y(t_n, y_n)f(t_n, y_n) + O(h^2). \end{aligned}$$

Putting this expression into (9) we arrive at

$$y_{n+1} = y_n + hf(t_n, y_n) + \frac{h^2}{2}[f_t(t_n, y_n) + f_y(t_n, y_n)f(t_n, y_n)] + O(h^3). \tag{11}$$

Let $g(t, y) = f_t(t, y(t)) + f_y(t, y(t))f(t, y(t))$. We subtract (11) from (10) to find

$$\begin{aligned} e_{n+1} &= e_n + h[f(t_n, y(t_n)) - f(t_n, y_n)] \\ &+ \frac{h^2}{2}[g(t_n, y_n) - g(t, y_n)] + O(h^3). \end{aligned}$$

Applying the mean value theorem in the y variable to both f and g, we find that

$$e_{n+1} = e_n[1 + hf_y(t_n, \xi_n) + \frac{h^2}{2} g_y(t_n, \eta_n)] + O(h^3). \tag{12}$$

It follows that the local error is on the order of h^3. The method converges and the cumulative error tends to zero as h^2. The explicit trapezoid method has order $p = 2$. The amplification factor of the error is

$$1 + hf_y + \frac{h^2}{2} g_y.$$

Since this expression is usually too difficult to use, we concentrate on the case that $f(t, y) = a$ where a is a constant. Then the amplification factor becomes

$$1 + ah + \frac{(ah)^2}{2}$$

from which it can be seen that the explicit trapezoid method is stable when

$$-2 \leq ah + \frac{(ah)^2}{2} \leq 0.$$

Since the ODE is stable only when $a \leq 0$, we see that in the case that $f(t, y) = ay$, the explicit trapezoid method is stable for $h \leq 2/|a|$.