Part II - Lent Term 2005
Numerical Analysis II – Model Solutions

Example (Error analysis for solving ODE)

Let $f : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ be a sufficiently smooth function with a Lipshitz constant $L > 0$,

$$\|f(t,x) - f(t,y)\| \le L\,\|x - y\|, \quad x, y \in \mathbb{R}^d, \quad t \ge 0.$$

Prove from the first principles that the implicit midpoint rule

$$y_{n+1} = y_n + hf\left(t_{n+1/2}, \tfrac{1}{2}(y_n + y_{n+1})\right), \quad n \ge 0, \tag{1}$$

converges to the exact solution of the ordinary differential equation

$$y' = f(t,y), \quad t \ge 0, \quad y(0) = y_0,$$

as $h \to 0$ uniformly for $0 \le t \le t^*$, where t^* is any finite constant.

Solution

1) Local truncation error η (which gives the order of the method). This is the error of the numerical scheme (1) applied to the exact solution $y'(t) = f(t, y(t))$:

$$y(t_{n+1}) = y(t_n) + hf\left(t_{n+1/2}, \tfrac{1}{2}(y(t_n) + y(t_{n+1})\right) + \eta. \tag{2}$$

As we cannot use $y'(t) = f(t, y(t))$ immediately for determining η, we slightly modify the right-hand side and write

$$y(t_{n+1}) = y(t_n) + hf(t_{n+1/2}, y(t_{n+1/2})) + \eta_1 + \eta. \tag{3}$$

Then, by Lipschitz condition,

$$\begin{aligned} |\eta_1| &= h\left|f\left(t_{n+1/2}, \tfrac{1}{2}(y(t_n) + y(t_{n+1}))\right) - f(t_{n+1/2}, y(t_{n+1/2}))\right| \\ &\le \tfrac{1}{2}hL\left|y(t_n) + y(t_{n+1}) - 2y(t_{n+1/2})\right|, \end{aligned}$$

and by Taylor expansion at t_n, with $y = y(t_n)$, we have

$$|\eta_1| \le \tfrac{1}{2}hL\left|y + (y + hy') - 2(y + \tfrac{1}{2}hy') + \mathcal{O}(h^2)\right| = \mathcal{O}(h^3).$$

Now, from (3), since $f(t, y(t)) = y'(t)$, we obtain

$$\eta_1 + \eta = y(t_{n+1}) - y(t_n) - hy'(t_{n+1/2}) = hy' + \tfrac{1}{2}h^2y'' - h\left[y' + \tfrac{1}{2}hy''\right] + \mathcal{O}(h^3) = \mathcal{O}(h^3),$$

hence

$$|\eta| \le |\eta_1| + \mathcal{O}(h^3) \le ch^3.$$

2) Approximation error $e_n = y(t_n) - y_n$. The recurrence relation for e_n is obtained by subtracting (1) from (2):

$$e_{n+1} = e_n + h\left[f\left(t_{n+1/2}, \tfrac{1}{2}(y(t_n) + y(t_{n+1}))\right) - f\left(t_{n+1/2}, \tfrac{1}{2}(y_n + y_{n+1})\right)\right] + \eta.$$

The Lipschitz condition implies

$$\begin{aligned} |e_{n+1}| &\le |e_n| + \tfrac{1}{2}hL\,|y(t_n) - y_n + y(t_{n+1}) - y_{n+1})| + |\eta| \\ &\le |e_n| + \tfrac{1}{2}hL\,(|e_n| + |e_{n+1}|) + ch^3. \end{aligned}$$

So, finally,

$$|e_{n+1}| \le \frac{1 + \frac{1}{2}hL}{1 - \frac{1}{2}hL}|e_n| + ch^3.$$

Now, the relations of the form

$$|e_{n+1}| \le A\,|e_n| + B, \quad n \ge 0, \quad e_0 = 0,$$

provide the estimate

$$|e_{n+1}| \le B(1 + A + A^2 + \cdots + A^n) = B\frac{A^{n+1} - 1}{A - 1}.$$

With our particular A and B, since $\frac{1}{2} < 1 - \frac{1}{2}hL < 1$ for small h, we have

$$A - 1 = \frac{hL}{1 - \frac{1}{2}hL} \ge hL,$$

$$A^{n+1} = \left(1 + \frac{hL}{1 - \frac{1}{2}hL}\right)^{n+1} \le (1 + 2Lh)^{n+1} \le e^{2Lh(n+1)} \le e^{2Lt^*},$$

whence

$$|e_{n+1}| \le \frac{1}{hL} e^{2Lt^*} ch^3 \le c_1 h^2,$$

with a constant c_1 that depends only on L, $\|y'''\|$ and t^*, i.e., we have uniform convergence in any finite interval $[0, t^*]$.

Example (QR algorithm)

Prove that, for each positive integer k the matrix A_{k+1} of the QR algorithm is related to the starting matrix $A_1 = A$ by the formula $A_{k+1} = \bar{Q}_k^T A \bar{Q}_k$, where $\bar{Q}_k \bar{R}_k$ is the QR factorization of the matrix A^k.

Let the QR algorithm be applied to the matrix

$$A = \begin{bmatrix} 1 & 0 \\ \lambda & 1 \end{bmatrix}$$

for some λ. Deduce from the above formula the elements of A_{k+1}. Also identify the limiting value of A_{k+1} as $k \to \infty$ and comment briefly on your result.

Solution

Since

$$A_i = Q_i R_i, \quad \text{and} \quad A_{i+1} = R_i Q_i = Q_{i+1} R_{i+1},$$

we have

$$\begin{aligned} A^k &= Q_1 R_1 Q_1 R_1 Q_1 \ \cdots \ R_1 Q_1 R_1 Q_1 R_1 \\ &= Q_1 Q_2 R_2 Q_2 R_2 \ \cdots \ Q_2 R_2 Q_2 R_2 R_1 \\ &= Q_1 Q_2 Q_3 \cdots \ Q_k R_k \ \cdots \ R_3 R_2 R_1 \\ &= \bar{Q}_k \bar{R}_k. \end{aligned}$$

Since also $R_i = Q_i^T A_i$, we have

$$\begin{aligned} A_{k+1} = R_k Q_k = Q_k^T A_k Q_k &= Q_k^T Q_{k-1}^T A_{k-1} Q_{k-1} Q_k = \cdots \\ &= Q_k^T Q_{k-1}^T \cdots Q_1^T A_1 Q_1 \cdots Q_{k-1} Q_k \\ &= \bar{Q}_k^T A \bar{Q}_k. \end{aligned}$$

(In particular, since for orthogonal matrices $Q^T = Q^{-1}$, the matrix A_k is similar to A, i.e., $A_k = SAS^{-1}$ for some matrix S, hence it has the same eigenvalues as the original matrix A).

2) Now, for the matrix A given, we compute

$$A^k = \begin{bmatrix} 1 & 0 \\ k\lambda & 1 \end{bmatrix}, \quad \bar{Q}_k^T = \frac{1}{\sqrt{1+k^2\lambda^2}} \begin{bmatrix} 1 & k\lambda \\ -k\lambda & 1 \end{bmatrix},$$

whence

$$\begin{aligned} A_{k+1} &= \frac{1}{1+k^2\lambda^2} \begin{bmatrix} 1 & k\lambda \\ -k\lambda & 1 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ \lambda & 1 \end{bmatrix} \times \begin{bmatrix} 1 & -k\lambda \\ k\lambda & 1 \end{bmatrix} \\ &= \frac{1}{1+k^2\lambda^2} \begin{bmatrix} 1+k\lambda^2+k^2\lambda^2 & -k^2\lambda^3 \\ \lambda & 1-k\lambda^2+k^2\lambda^2 \end{bmatrix} \end{aligned}$$

and

$$\lim_{k\to\infty} A_{k+1} = \begin{bmatrix} 1 & -\lambda \\ 0 & 1 \end{bmatrix},$$

the first column being $\lambda_{\max} e_1$, with $\lambda_{\max}$ the largest eigenvalue of A.

Example (Runge–Kutta method)

Consider the Runge–Kutta method

$$\begin{aligned} k_1 &= f(t_n, y_n), \\ k_2 &= f(t_n + a_{21}h, y_n + a_{21}k_1 h), \\ k_3 &= f\left(t_n + (a_{31} + a_{32})h, y_n + (a_{31}k_1 + a_{32}k_2)h\right), \\ y_{n+1} &= y_n + (b_1 k_1 + b_2 k_2 + b_3 k_3)h \end{aligned}$$

for solving $y'(t) = f(t, y)$, $t \in \mathbb{R}$. Show that its order is at least two if the equations

$$b_1 + b_2 + b_3 = 1 \quad \text{and} \quad b_2 a_{21} + b_3(a_{31} + a_{32}) = \tfrac{1}{2} \qquad (*)$$

are satisfied, and it is three if the parameters have the values

$$a_{21} = 1, \quad a_{31} = 4/9, \quad a_{32} = 2/9, \quad b_1 = 1/4, \quad b_2 = 0, \quad b_3 = 3/4. \qquad (**)$$

Solution

We need to show that, for $y_n = y(t_n)$, where y is the exact solution of $y'(t) = f(t, y)$, and with particular parameters given in $(*)$ and $(**)$, we obtain

$$\begin{aligned} (*) \qquad & y(t_{n+1}) = y(t_n) + (b_1 k_1 + b_2 k_2 + b_3 k_3)h + \mathcal{O}(h^3), \\ (**) \qquad & y(t_{n+1}) = y(t_n) + (b_1 k_1 + b_2 k_2 + b_3 k_3)h + \mathcal{O}(h^4), \end{aligned}$$

respectively. We will prove it by showing that the right-hand side coincide with the Taylor expansion of y at t_n up to a given order, i.e.,

$$(b_1 k_1 + b_2 k_2 + b_3 k_3)h = hy' + \frac{1}{2}h^2 y'' + \mathcal{O}(h^3) = hy' + \frac{1}{2}h^2 y'' + \frac{1}{6}h^3 y''' \mathcal{O}(h^4),$$

So, we make the Taylor expansion of k_i around the point $(t_n, y(t_n))$, and use the formulae for previous k_j-s to express k_i in terms of f and its partial derivatives:

$$\begin{aligned} k_1 &= f, \\ k_2 &= f + h\left[a_{21} f_t + a_{21} k_1 f_y\right] + \frac{1}{2}h^2\left[a_{21}^2 f_{tt} + 2a_{21}^2 k_1 f_{ty} + a_{21}^2 k_1^2 f_{yy}\right] + \mathcal{O}(h^3) \\ &= f + h\left[a_{21}(f_t + f f_y)\right] + \frac{1}{2}h^2\left[a_{21}^2 (f_{tt} + 2f f_{ty} + f^2 f_{yy})\right] + \mathcal{O}(h^3), \\ k_3 &= f + h\left[(a_{31} + a_{32}) f_t + (a_{31}k_1 + a_{32}k_2) f_y\right] \\ &\quad + \frac{1}{2}h^2\left[(a_{31} + a_{32})^2 f_{tt} + 2(a_{31} + a_{32})(a_{31}k_1 + a_{32}k_2) f_{ty} + (a_{31}k_1 + a_{32}k_2)^2 f_{yy}\right] + \mathcal{O}(h^3) \\ &= f + h\left[(a_{31} + a_{32})(f_t + f f_y)\right] + h^2\left[a_{32}a_{21}(f_t + f f_y) f_y\right] \\ &\quad + \frac{1}{2}h^2\left[(a_{31} + a_{32})^2 (f_{tt} + 2f f_{ty} + f^2 f_{yy})\right] + \mathcal{O}(h^3)\,. \end{aligned}$$

Now, summing up and combining terms at powers of h, we derive

$$\begin{aligned} h(b_1 k_1 + b_2 k_2 + b_3 k_3) &= h(b_1 + b_2 + b_3) f + h^2\left[b_2 a_{21} + b_3(a_{31} + a_{32})\right](f_t + f f_y) \\ &\quad + h^3\left\{b_3 a_{32} a_{21}(f_t + f f_y) f_y + \left[\tfrac{1}{2}b_2 a_{21}^2 + \tfrac{1}{2}b_3(a_{31} + a_{32})^2\right](f_{tt} + 2f f_{ty} + f^2 f_{yy})\right\} + \mathcal{O}(h^4). \end{aligned}$$

Differentiating the identity $y'(t) = f(t, y)$ with respect to t we find that, for the exact solution y,

$$\begin{aligned} y' &= f, \\ y'' &= f_t + f_y f, \\ y''' &= (f_{tt} + f_{ty} f + f_y f_t) + (f_{yt} + f_{yy} f + f_y^2) f = (f_{tt} + 2f_{ty} f + f_{yy} f^2) + (f_t + f_y f) f_y\,. \end{aligned}$$

So, in case $(*)$, we indeed have order two, and we get order three if

$$b_3 a_{32} a_{21} = \tfrac{1}{2}b_2 a_{21}^2 + \tfrac{1}{2}b_3(a_{31} + a_{32})^2 = \tfrac{1}{6}\,,$$

and that is the case for the parameteres in $(**)$.

Example (Multistep method)

Determine the order of the multistep method

$$y_{n+2} - (1+\alpha)y_{n+1} + \alpha y_n = \frac{1}{12}h\left[(5+\alpha)f_{n+2} + (8-8\alpha)f_{n+1} - (1+5\alpha)f_n\right]$$

for the solution of ODEs $y' = f(t,y)$ for different choices of α.

Solution

The order of a numerical method $y_{n+1} = \phi(t_0..t_n, y_0..y_n)$ for solving $y' = f(t,y)$ is the order of its error on the exact solution y reduced by one, i.e., it is the largest integer p such that

$$y(t_{n+1}) = \phi(t_0..t_n, y(t_0)..y(t_n)) + \mathcal{O}(h^{p+1}).$$

In our case, since $f_k = f(t_k, y_k) = y'(t_k)$ for the exact solution y, it is the error of the formula

$$\begin{aligned} &y(t_{n+2}) - (1+\alpha)y(t_{n+1}) + \alpha y(t_n) \\ &\quad = \tfrac{1}{12}h\left[(5+\alpha)y'(t_{n+2}) + (8-8\alpha)y'(t_{n+1}) - (1+5\alpha)y'(t_n)\right] + \mathcal{O}(h^{p+1}) \end{aligned}$$

for all sufficiently smooth functions y. We see at once that this p is the highest term until which the Taylor expansions of the right- and of the left-hand sides coincide.

Method 1. Equivalently, it is the largest p for which the above formula is identity for any polynomial of degree p. It is sufficient to verify this formula for monomials $y(t) = t^k$ and for any particular h and t_n, say $h = 1$ and $t_n = -1$ (so that $t_{n+1} = 0$ and $t_{n+2} = 1$). We have

$$\begin{aligned} y(t) = t^0, \quad & 0 &=& \ 0, \\ y(t) = t^1, \quad & 1-\alpha &=& \ \tfrac{1}{12}(12 - 12\alpha) = 1-\alpha, \\ y(t) = t^2, \quad & 1+\alpha &=& \ \tfrac{2}{12}((5+\alpha) + (1+5\alpha)) = 1+\alpha, \\ y(t) = t^3, \quad & 1-\alpha &=& \ \tfrac{3}{12}((5+\alpha) - (1+5\alpha)) = 1-\alpha, \\ y(t) = t^4, \quad & 1+\alpha &=& \ \tfrac{4}{12}((5+\alpha) + (1+5\alpha)) = 2(1+\alpha), \\ y(t) = t^5, \quad & 1-\alpha &=& \ \tfrac{5}{12}((5+\alpha) - (1+5\alpha)) = \tfrac{5}{3}(1-\alpha). \end{aligned}$$

So, the method is of order three for any α, and it is of order four if $\alpha = -1$.

Method 2. By Theorem 3.11 of Lecture Notes, we have

$$\rho(w) = w - (1+\alpha) + 1/w, \quad \sigma(w) = \tfrac{1}{12}[(5+\alpha)w + (8-8\alpha) - (1+5\alpha)/w]\,.$$

(Here, we made a shift $n+1 \to n$ to simplify expressions involved.) Then

$$\begin{aligned} &\rho(e^z) - z\sigma(e^z) \\ &\quad = [1 + z + \tfrac{1}{2!}z^2 + \tfrac{1}{3!}z^3 + \tfrac{1}{4!}z^4 + \tfrac{1}{5!}z^5] - (1+\alpha) + \alpha[1 - z + \tfrac{1}{2!}z^2 - \tfrac{1}{3!}z^3 + \tfrac{1}{4!}z^4 - \tfrac{1}{5!}z^5] \\ &\qquad - \tfrac{1}{12}z\left\{(5+\alpha)[1 + z + \tfrac{1}{2!}z^2 + \tfrac{1}{3!}z^3 + \tfrac{1}{4!}z^4] + (8-8\alpha)\right. \\ &\qquad\qquad \left. + (1+5\alpha)[1 - z + \tfrac{1}{2!}z^2 - \tfrac{1}{3!}z^3 + \tfrac{1}{4!}z^4]\right\} + \mathcal{O}(z^6) \\ &\quad = -\tfrac{1}{4!}(1+\alpha)z^4 - \tfrac{1}{5!}\tfrac{2}{3}(1-\alpha)z^5 + \mathcal{O}(z^6), \end{aligned}$$

where calculations of the coefficients at z^k is the same procedure as for $y(t) = t^k$ in the Method 1. So, again, the method is of order three for any α, and it is of order four if $\alpha = -1$.