Numerical Analysis – Lecture 12[1]

Methods 3.25 A general ν-stage *Runge–Kutta method* is

$$\boldsymbol{k}_l = \boldsymbol{f}\left(t_n + c_l h, \boldsymbol{y}_n + h\sum_{j=1}^{\nu} a_{l,j}\boldsymbol{k}_j\right) \quad \text{where} \quad \sum_{j=1}^{\nu} a_{l,j} = c_l, \qquad l = 1,2,\ldots,\nu,$$

$$\boldsymbol{y}_{n+1} = \boldsymbol{y}_n + h\sum_{l=1}^{\nu} b_l \boldsymbol{k}_l.$$

Obviously, $a_{l,j} = 0$ for all $l \le j$ yields the standard *explicit* RK. Otherwise, an RK method is said to be *implicit.*

Example 3.26 Consider the 2-stage method

$$\begin{aligned}\boldsymbol{k}_1 &= \boldsymbol{f}\left(t_n, \boldsymbol{y}_n + \tfrac{1}{4}h(\boldsymbol{k}_1 - \boldsymbol{k}_2)\right),\\ \boldsymbol{k}_2 &= \boldsymbol{f}\left(t_n + \tfrac{2}{3}h, \boldsymbol{y}_n + \tfrac{1}{12}h(3\boldsymbol{k}_1 + 5\boldsymbol{k}_2)\right),\\ \boldsymbol{y}_{n+1} &= \boldsymbol{y}_n + \tfrac{1}{4}h(\boldsymbol{k}_1 + 3\boldsymbol{k}_2).\end{aligned}$$

In order to analyse the order of this method, we restrict our attention to scalar, autonomuous equations of the form $y' = f(y)$.[2] For brevity, we use the convention that all functions are evaluated at $y = y_n$, e.g. $f_y = \mathrm{d}f(y_n)/\mathrm{d}y$. Thus,

$$\begin{aligned}k_1 &= f + \tfrac{1}{4}hf_y(k_1 - k_2) + \tfrac{1}{32}h^2 f_{yy}(k_1 - k_2)^2 + \mathcal{O}\big(h^3\big),\\ k_2 &= f + \tfrac{1}{12}hf_y(3k_1 + 5k_2) + \tfrac{1}{288}h^2 f_{yy}(3k_1 + 5k_2)^2 + \mathcal{O}\big(h^3\big).\end{aligned}$$

We have $k_1, k_2 = f + \mathcal{O}(h)$ and substitution in the above equations yields $k_1 = f + \mathcal{O}\big(h^2\big)$, $k_2 = f + \frac{2}{3}hf_y f + \mathcal{O}\big(h^2\big)$. Substituting again, we obtain

$$\begin{aligned}k_1 &= f - \tfrac{1}{6}h^2 f_y^2 f + \mathcal{O}\big(h^3\big),\\ k_2 &= f + \tfrac{2}{3}hf_y f + h^2\left(\tfrac{5}{18}f_y^2 f + \tfrac{2}{9}f_{yy}f^2\right) + \mathcal{O}\big(h^3\big)\\ \Rightarrow \quad y_{n+1} &= y + hf + \tfrac{1}{2}h^2 f_y f + \tfrac{1}{6}h^3(f_y^2 f + f_{yy}f^2) + \mathcal{O}\big(h^4\big).\end{aligned}$$

But $y' = f \Rightarrow y'' = f_y f \Rightarrow y''' = f_y^2 f + f_{yy}f^2$ and we deduce from Taylor's theorem that the method is at least of order 3. (It is easy to verify that it isn't of order 4, for example applying it to the equation $y' = \lambda y$.)

Phenomenon 3.27 Consider the linear system

$$\boldsymbol{y}' = A\boldsymbol{y} \qquad \text{where} \qquad A = \begin{bmatrix} -100 & 1 \\ 0 & -\frac{1}{10} \end{bmatrix}.$$

The exact solution is a linear combination of $\mathrm{e}^{-t/10}$ and e^{-100t}: the first decays gently, whereas the second becomes practically zero almost at once. Suppose that we solve the ODE with the *forward Euler* method. As will be shown soon, the requirement that $\lim_{n\to\infty}\boldsymbol{y}_n = \boldsymbol{0}$ (for fixed $h > 0$) leads to an unacceptable restriction on the size of h.

With greater generality, let us solve $\boldsymbol{y}' = A\boldsymbol{y}$, for general $N \times N$ constant matrix A, with Euler's method. Then $\boldsymbol{y}_{n+1} = (I + hA)\boldsymbol{y}_n$, therefore $\boldsymbol{y}_n = (I + hA)^n\boldsymbol{y}_0$. Let the eigenvalues of A be $\lambda_1, \ldots, \lambda_N$, with corresponding linearly-independent eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N$. Let $D = \mathrm{diag}\boldsymbol{\lambda}$ and $V = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_N]$,

[1] Please email all corrections and suggestions to these notes to `A.Iserles@damtp.cam.ac.uk`. All handouts are available on the WWW at the URL `http://www.damtp.cam.ac.uk/user/na/PartII/Handouts.html`.

[2] This procedure might lead to loss of generality for methods of order ≥ 5.

whence $A = VDV^{-1}$. We assume further that $\mathrm{Re}\,\lambda_l < 0$, $l = 1, \ldots, N$. In that case it is easy to prove that $\lim_{t\to\infty} \boldsymbol{y}(t) = \boldsymbol{0}$, e.g. by representing the exact solution of the ODE explicitly as

$$\boldsymbol{y}(t) = \mathrm{e}^{tA}\boldsymbol{y}_0, \qquad \text{where} \qquad \mathrm{e}^{tA} = \sum_{k=0}^{\infty} \frac{1}{k!} t^k A^k = V \mathrm{e}^{tD} V^{-1}.$$

However, $\boldsymbol{y}_n = V(I + hD)^n V^{-1}\boldsymbol{y}_0$, where $A = VDV^{-1}$ and the matrix D is diagonal, therefore $\lim_{n\to\infty} \boldsymbol{y}_n = \boldsymbol{0}$ for all initial values $\boldsymbol{y}_0$ iff $|1 + h\lambda_l| < 1$, $l = 1, \ldots, N$. In our example we thus require $|1 - \frac{1}{10}h|, |1 - 100h| < 1$, hence $h < \frac{1}{50}$.

It is important to realise that this restriction, necessary to recovery of correct asymptotic behaviour, has *nothing* to do with local accuracy, since, for large n, the genuine 'unstable' component is exceedingly small. Its purpose is solely to prevent this component from leading to an unbounded growth in the numerical solution.

Definition 3.28 We say that the ODE $\boldsymbol{y}' = \boldsymbol{f}(t, \boldsymbol{y})$ is *stiff* if (for some methods) we need to depress h to maintain *stability* well beyond requirements of accuracy. An important example of stiff systems occurs when an equation is linear, $\mathrm{Re}\,\lambda_l < 0$, $l = 1, 2, \ldots, N$, and the quotient $\max|\lambda_k| / \min|\lambda_k|$ is large: a ratio of 10^{20} is not unusual in real-life problems!

Stiff equations, mostly nonlinear, occur throughout applications, whenever we have two (or more) different timescales in the ODE. A typical example are equations of *chemical kinetics*, where each timescale is determined by the speed of reaction between two compounds: such speeds can differ by many orders of magnitude.

Definition 3.29 Suppose that a numerical method, applied to $y' = \lambda y$, $y(0) = 1$, with constant h, produces the solution sequence $\{y_n\}_{n\in\mathbb{Z}^+}$. We call the set

$$\mathcal{D} = \{h\lambda \in \mathbb{C} : \lim_{n\to\infty} y_n = 0\}$$

the *linear stability domain* of the method. Noting that the set of $\lambda \in \mathbb{C}$ for which $y(t) \overset{t\to\infty}{\longrightarrow} 0$ is the left half-plane $\mathbb{C}^- = \{z \in \mathbb{C} : \mathrm{Re}\, z < 0\}$, we say that the method is *A-stable* if $\mathbb{C}^- \subseteq \mathcal{D}$.

Example 3.30 We have already seen that for Euler's method $y_n \to 0$ iff $|1 + h\lambda| < 1$, therefore $\mathcal{D} = \{z \in \mathbb{C} : |1 + z| < 1\}$. Moreover, solving $y' = \lambda y$ with the *trapezoidal rule*, we obtain $y_{n+1} = [(1 + \frac{1}{2}h\lambda)/(1 - \frac{1}{2}h\lambda)]y_n$ thus, by induction, $y_n = [(1 + \frac{1}{2}h\lambda)/(1 - \frac{1}{2}h\lambda)]^n y_0$. Therefore

$$z \in \mathcal{D} \qquad \Leftrightarrow \qquad \left|\frac{1 + \frac{1}{2}z}{1 - \frac{1}{2}z}\right| < 1 \qquad \Leftrightarrow \qquad \mathrm{Re}\, z < 0$$

and we deduce that $\mathcal{D} = \mathbb{C}^-$. Hence, the method is A-stable.
It can be proved by similar means that for *backward Euler* it is true that $\mathcal{D} = \{z \in \mathbb{C} : |1 - z| > 1\}$, hence that the method is also A-stable.

Note that A-stability does not mean that *any* step size will do! We need to choose h small enough to ensure the right accuracy, but we don't want to depress it much further to prevent instability.

Discussion 3.31 A-stability analysis of multistep methods is considerably more complicated. However, according to the *second Dahlquist barrier*, no multistep method of order $p \geq 3$ may be A-stable. Note that the $p = 2$ barrier for A-stability is attained by the trapezoidal rule.

The Dahlquist barrier implies that, in our quest for higher-order methods with good stability properties, we need to pursue one of the following strategies:

- either relax the definition of A-stability
- or consider other methods in place of multistep.

The two courses of action will be considered in the next lecture.