MA4255 Numerical Methods in Differential Equations

Chapter 2: One-step methods

The problem and the standing assumption

We consider the IVP

$$y'(x) = f(x, y(x)), \qquad y(x_0) = y_0.$$

We suppose throughout that f satisfies the conditions of Picard's Theorem on the rectangle R and that the IVP has a unique solution defined on $[x_0, X_M]$, $-\infty < x_0 < X_M < \infty$.

2.1 Euler's method and its relatives: the θ-method

The simplest one-step method: Euler's method

Problem: approximate the soln $y : [x_0, X_M] \to \mathbb{R}$ to the IVP

$$y'(x) = f(x, y(x)) \quad \text{for} \quad x \in (x_0, X_M), \qquad y(x_0) = y_0.$$

For $N \in \mathbb{N}$, we divide the interval $[x_0, X_M]$ by the $N+1$ **mesh-points**

$$x_0, \quad x_1 = x_0 + h, \quad x_2 = x_0 + 2h, \quad \cdots \quad , \quad x_N = x_0 + Nh = X_M,$$

where $h := \frac{X_M - x_0}{N} > 0$ is the so-called **step size**. For $n \in \{0, \dots, N\}$, we want to find an approximation y_n to $y(x_n)$. $n = 0$ is already done by i.c.!

Idea: $y(x_{n+1}) = y(x_n) + \int_{x_n}^{x_{n+1}} y'(x)\,\mathrm{d}x = y(x_n) + \int_{x_n}^{x_{n+1}} f(x, y(x))\,\mathrm{d}x.$

$$\text{Rectangle rule:} \quad \int_a^b g(x)\,\mathrm{d}x \approx (b-a)g(a)$$

$\Longrightarrow y(x_{n+1}) \approx y(x_n) + hf(x_n, y(x_n))$. **(Explicit) Euler method**:

$$y_{n+1} = y_n + hf(x_n, y_n), \quad n \in \{0, \dots, N-1\}.$$

Generalization: the θ-method

Idea: As before,

$$y(x_{n+1}) = y(x_n) + \int_{x_n}^{x_{n+1}} f(x, y(x))\,\mathrm{d}x.$$

Now, instead of rectangle rule, let us use the integration rule

$$\int_a^b g(x)\,\mathrm{d}x \approx (b-a)\left[(1-\theta)g(a) + \theta g(b)\right], \qquad \theta \in [0,1].$$

We find that

$$y(x_{n+1}) \approx y(x_n) + h\left[(1-\theta)f(x_n, y(x_n)) + \theta f(x_{n+1}, y(x_{n+1}))\right].$$

The θ-method with $\theta \in [0,1]$:

$$y_{n+1} = y_n + h\left[(1-\theta)f(x_n, y_n) + \theta f(x_{n+1}, y_{n+1})\right], \quad n \in \{0, \ldots, N-1\}.$$

To compute y_{n+1}, only need the previous value $y_n \Longrightarrow$ **One-step method**

The choices $\theta = 0$, $\theta = 1$, and $\theta = \frac{1}{2}$

Recall the θ-method with $\theta \in [0, 1]$:

$$y_{n+1} = y_n + h\left[(1-\theta) f(x_n, y_n) + \theta f(x_{n+1}, y_{n+1})\right], \quad n \in \{0, \ldots, N-1\}.$$

- The choice $\theta = 0$: **Explicit Euler method**

$$y_{n+1} = y_n + hf(x_n, y_n), \qquad n \in \{0, \ldots, N-1\}.$$

- The choice $\theta = 1$: **Implicit Euler method**

$$y_{n+1} = y_n + hf(x_{n+1}, y_{n+1}), \qquad n \in \{0, \ldots, N-1\}.$$

- The choice $\theta = \frac{1}{2}$: **Trapezium rule method**

$$y_{n+1} = y_n + h\frac{f(x_n, y_n) + f(x_{n+1}, y_{n+1})}{2}, \qquad n \in \{0, \ldots, N-1\}.$$

Rk: If instead of averaging $f(x_n, y_n), f(x_{n+1}, y_{n+1})$, we evaluate f at the average of x_n, x_{n+1} and y_n, y_{n+1} we obtain the **implicit midpoint rule**:

$$y_{n+1} = y_n + hf\left(\frac{x_n + x_{n+1}}{2}, \frac{y_n + y_{n+1}}{2}\right), \qquad n \in \{0, \ldots, N-1\}.$$

So, which $\theta \in [0,1]$ is best? An experiment

Consider the IVP

$$y'(x) = x - [y(x)]^2 \quad \text{for} \quad x \in (0, 0.4), \qquad y(0) = 0.$$

(Note: here, $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,z) := x - z^2$.) We take the mesh points

$$x_0 := 0, \quad x_1 := 0.1, \quad x_2 := 0.2, \quad x_3 := 0.3, \quad x_4 := 0.4 \qquad (h := 0.1).$$

Choice $\theta = 0$ (explicit Euler): $y_{n+1} = y_n + hf(x_n, y_n)$

$$\begin{aligned}
y_0 &= y(x_0) = 0,\\
y_1 &= y_0 + hf(x_0, y_0) = 0 + 0.1 \cdot f(0,0) = 0,\\
y_2 &= y_1 + hf(x_1, y_1) = 0 + 0.1 \cdot f(0.1,0) = 0.01,\\
y_3 &= y_2 + hf(x_2, y_2) = 0.01 + 0.1 \cdot f(0.2,0.01) = 0.02999,\\
y_4 &= y_3 + hf(x_3, y_3) = 0.02999 + 0.1 \cdot f(0.3,0.02999) \approx 0.05990.
\end{aligned}$$

Values of the true solution y at $x_1, \ldots, x_4$ (to 5 digits after comma):

$$y(x_1) \approx 0.00500, \quad y(x_2) \approx 0.01998, \quad y(x_3) \approx 0.04488, \quad y(x_4) \approx 0.07949.$$

We see that the largest error $|e_n| := |y(x_n) - y_n|$ is $|e_4| \approx 1.96 \cdot 10^{-2}$.

So, which $\theta \in [0,1]$ is best? An experiment

Consider the IVP

$$y'(x) = x - [y(x)]^2 \quad \text{for} \quad x \in (0, 0.4), \qquad y(0) = 0.$$

(Note: here, $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,z) := x - z^2$.) We take the mesh points

$$x_0 := 0, \quad x_1 := 0.1, \quad x_2 := 0.2, \quad x_3 := 0.3, \quad x_4 := 0.4 \qquad (h := 0.1).$$

Choice $\theta = 1$ (implicit Euler): $y_{n+1} = y_n + hf(x_{n+1}, y_{n+1})$

$$\begin{aligned}
y_0 &= y(x_0) = 0,\\
y_1 &= y_0 + hf(x_1, y_1) = 0.1(0.1 - y_1^2) \quad \Longrightarrow \quad y_1 \approx 0.00999,\\
y_2 &= y_1 + hf(x_2, y_2) \quad \Longrightarrow \quad y_2 \approx 0.02990,\\
y_3 &= y_2 + hf(x_3, y_3) \quad \Longrightarrow \quad y_3 \approx 0.05955,\\
y_4 &= y_3 + hf(x_4, y_4) \quad \Longrightarrow \quad y_4 \approx 0.09857.
\end{aligned}$$

Values of the true solution y at $x_1, \ldots, x_4$ (to 5 digits after comma):

$$y(x_1) \approx 0.00500, \quad y(x_2) \approx 0.01998, \quad y(x_3) \approx 0.04488, \quad y(x_4) \approx 0.07949.$$

We see that the largest error $|e_n| := |y(x_n) - y_n|$ is $|e_4| \approx 1.91 \cdot 10^{-2}$.

So, which $\theta \in [0, 1]$ is best? An experiment

Consider the IVP

$$y'(x) = x - [y(x)]^2 \quad \text{for} \quad x \in (0, 0.4), \qquad y(0) = 0.$$

(Note: here, $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, z) := x - z^2$.) We take the mesh points

$$x_0 := 0, \quad x_1 := 0.1, \quad x_2 := 0.2, \quad x_3 := 0.3, \quad x_4 := 0.4 \qquad (h := 0.1).$$

Choice $\theta = \frac{1}{2}$ (trap. rule method): $y_{n+1} = y_n + \frac{h}{2}(f(x_n, y_n) + f(x_{n+1}, y_{n+1}))$.

$y_0 = y(x_0) = 0,$

$y_1 = y_0 + \frac{h}{2}(f(x_0, y_0) + f(x_1, y_1)) = 0.05(0.1 - y_1^2) \quad \Longrightarrow \quad y_1 \approx 0.00500,$

$y_2 = y_1 + \frac{h}{2}(f(x_1, y_1) + f(x_2, y_2)) \quad \Longrightarrow \quad y_2 \approx 0.01998,$

$y_3 = y_2 + \frac{h}{2}(f(x_2, y_2) + f(x_3, y_3)) \quad \Longrightarrow \quad y_3 \approx 0.04486,$

$y_4 = y_3 + \frac{h}{2}(f(x_3, y_3) + f(x_4, y_4)) \quad \Longrightarrow \quad y_4 \approx 0.07944.$

Values of the true solution y at $x_1, \ldots, x_4$ (to 5 digits after comma):

$y(x_1) \approx 0.00500, \quad y(x_2) \approx 0.01998, \quad y(x_3) \approx 0.04488, \quad y(x_4) \approx 0.07949.$

We see that the largest error $|e_n| := |y(x_n) - y_n|$ is $|e_4| \approx 5 \cdot 10^{-5}$.

So, which $\theta \in [0, 1]$ is best? An experiment

$\Longrightarrow$ $\theta = \frac{1}{2}$ seems to be much better than $\theta = 0$ or $\theta = 1$. Why?

$\Longrightarrow$ Will be clear after the next section.

Remark: The true solution in this example is not available in closed form. We have used Picard iteration

$$y_0(x) \equiv 0, \qquad y_k(x) = \int_0^x \left(t - [y_{k-1}(t)]^2\right) \, \mathrm{d}t, \quad k \in \mathbb{N}.$$

to get a very fine approximation of the soln, which acts as true soln.

$$y_0(x) \equiv 0 \qquad y_1(x) = \frac{1}{2}x^2, \qquad y_2(x) = \frac{1}{2}x^2 - \frac{1}{20}x^5, \qquad \cdots$$

By induction, one shows that

$$y(x) = \frac{1}{2}x^2 - \frac{1}{20}x^5 + \frac{1}{160}x^8 - \frac{1}{4400}x^{11} + O\left(x^{14}\right).$$

2.2 Error analysis of the θ-method

The goal of this section

Consider the θ-method for $\theta \in [0, 1]$:

$$y_{n+1} = y_n + h\left[(1-\theta)f(x_n, y_n) + \theta f(x_{n+1}, y_{n+1})\right], \quad n \in \{0, \dots, N-1\}.$$

We define the **global error** e_n by

$$e_n := y(x_n) - y_n \quad \text{for} \quad n \in \{0, \dots, N\}.$$

We investigate the decay of the global error for the θ-method with respect to the reduction of the mesh size h.

For simplicity, we consider $\theta = 0$ (i.e., explicit Euler) and then only state the result for general $\theta \in [0, 1]$ in the end (exercise).

Error analysis for explicit Euler. 1) Consistency error

Explicit Euler method:

$$y_{n+1} = y_n + hf(x_n, y_n), \qquad n \in \{0, 1, \ldots, N-1\}, \qquad y_0 = y(x_0).$$

Introduce the **consistency error** (or **truncation error**):

$$T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - f(x_n, y(x_n)), \qquad n \in \{0, 1, \ldots, N-1\},$$

By $f(x_n, y(x_n)) = y'(x_n)$ and Taylor's Theorem, $\exists\ \xi_n \in (x_n, x_{n+1})$ s.t.

$$|T_n| = \frac{|y(x_{n+1}) - y(x_n) - hy'(x_n)|}{h} = \frac{\frac{1}{2}h^2|y''(\xi_n)|}{h} \leq \frac{h}{2}M_2$$

with $M_2 := \max_{x\in[x_0,X_M]} |y''(x)|$, where we have assumed f is sufficiently smooth so that $y''(x) = \frac{\mathrm{d}}{\mathrm{d}x}[f(x, y(x))]$ exists and is bounded on $[x_0, X_M]$.

Error analysis for explicit Euler. 2) Compare e_{n+1} to e_n

By definition of explicit Euler:

$$0 = \frac{y_{n+1} - y_n}{h} - f(x_n, y_n). \tag{1}$$

By definition of consistency error:

$$T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - f(x_n, y(x_n)). \tag{2}$$

Recall $e_n = y(x_n) - y_n$. Subtract (1) from (2):

$$T_n = \frac{e_{n+1} - e_n}{h} - \left[f(x_n, y(x_n)) - f(x_n, y_n)\right],$$

or equivalently,

$$e_{n+1} = e_n + h[f(x_n, y(x_n)) - f(x_n, y_n)] + hT_n.$$

Assuming that $|y_n - y_0| \leq Y_M$, from the Lipschitz condition we get

$$|e_{n+1}| \leq |e_n| + h|f(x_n, y(x_n)) - f(x_n, y_n)| + h|T_n| \leq (1 + hL)|e_n| + h|T_n|.$$

Now, let $T := \max\{|T_0|, |T_1|, \ldots, |T_{N-1}|\}$; then,

$$|e_{n+1}| \leq (1 + hL)|e_n| + hT \qquad \forall n \in \{0, \ldots, N-1\}.$$

Error analysis for explicit Euler. 3) Compare e_n to e_0

Using $|e_{n+1}| \le (1+hL)|e_n| + hT$ from Step 2, we find

$$\begin{aligned}|e_n| &\le (1+hL)|e_{n-1}| + hT \\ &\le (1+hL)\left((1+hL)|e_{n-2}| + hT\right) + hT \\ &\vdots \\ &\le (1+hL)^n|e_0| + \frac{T}{L}\left[(1+hL)^n - 1\right].\end{aligned}$$

Noting that $1 + x \le e^x \ \forall x \in \mathbb{R}$, and $nh = x_n - x_0$, we have

$$|e_n| \le e^{nhL}|e_0| + \frac{T}{L}\left[e^{nhL} - 1\right] = e^{L(x_n - x_0)}|e_0| + \frac{T}{L}\left[\mathrm{e}^{L(x_n - x_0)} - 1\right].$$

Using $T \le \frac{h}{2}M_2$ from Step 1 gives

$$|e_n| \le \mathrm{e}^{L(x_n - x_0)}|e_0| + h\frac{M_2}{2L}\left[\mathrm{e}^{L(x_n - x_0)} - 1\right] \qquad \forall n \in \{0, \ldots, N\}.$$

Error analysis for explicit Euler. 4) Conclude

We have found in Step 3 that

$$|e_n| \leq \mathrm{e}^{L(x_n - x_0)}|e_0| + h\frac{M_2}{2L}\left[\mathrm{e}^{L(x_n-x_0)} - 1\right] \qquad \forall n \in \{0, \ldots, N\}.$$

Observation: $e_0 = y(x_0) - y_0 = 0.$ (at least in theory; in practice, y_0 is the nearest floating point number to $y(x_0)$). Thus,

$$|e_n| \leq h\frac{M_2}{2L}\left[\mathrm{e}^{L(x_n-x_0)} - 1\right] \leq h\frac{M_2}{2L}\left[\mathrm{e}^{L(X_M-x_0)} - 1\right] \quad \forall n \in \{0, \ldots, N\}.$$

We conclude that

$$\max_{n\in\{0,\ldots,N\}} |e_n| = \mathcal{O}(h)$$

as $h \searrow 0$.

Error analysis for the θ-method with $\theta \in [0,1]$

Exercise.

For $\theta \neq \frac{1}{2}$, one finds that

$$\max_{n \in \{0,\dots,N\}} |e_n| = \mathcal{O}(h),$$

and for $\theta = \frac{1}{2}$, one find that

$$\max_{n \in \{0,\dots,N\}} |e_n| = \mathcal{O}(h^2).$$

$\Longrightarrow \theta = \frac{1}{2}$ "best" value of θ regarding speed of convergence to true soln.

So, why not always use $\theta = \frac{1}{2}$?

While the trapezium rule method leads to more accurate approximations than the explicit Euler method, it is less convenient from the computational point of view because it requires the solution of implicit equations at each mesh point x_{n+1} to compute y_{n+1}.

Attractive compromise: use explicit Euler to compute an initial crude approximation to $y(x_{n+1})$ and then use this value within the trapezium rule to obtain a more accurate approximation for $y(x_{n+1})$:

$$y_{n+1} = y_n + \frac{h}{2}[f(x_n, y_n) + f(x_{n+1}, y_n + hf(x_n, y_n))], \quad n \in \{0, \dots, N-1\}$$

This is called the **improved Euler method**.

2.3 General one-step methods

Formal definition of a one-step method

Definition (One-step method)

A **one-step method** is a fct Ψ that takes $(\xi, \eta; h) \in \mathbb{R} \times \mathbb{R} \times (0, \infty)$ and a fct $f(\cdot, \cdot)$, and computes an approximation $\Psi(\xi, \eta; h, f) \in \mathbb{R}$ of $y(\xi + h)$, which is the soln at $x = \xi + h$ of the IVP

$$y'(x) = f(x, y(x)), \qquad y(\xi) = \eta. \tag{3}$$

We assume that (3) has a unique soln, and therefore $y(\xi + h)$ exists. Here, h may need to be sufficiently small for Ψ to be well-defined.

- explicit Euler: $\Psi(\xi, \eta; h, f) = \eta + hf(\xi, \eta)$.
- implicit Euler: Ψ is defined implicitly, by

$$\Psi(\xi, \eta; h, f) = \eta + hf(\xi + h, \Psi(\xi, \eta; h, f)).$$

If f satisfies a **global Lipschitz condition** (in 2nd arg.), i.e., $\exists L > 0$:

$$|f(x, z) - f(x, \tilde{z})| \le L|z - \tilde{z}| \qquad \forall (x, z), (x, \tilde{z}) \in [x_0, X_M] \times \mathbb{R},$$

then for $(\xi, \eta) \in \mathbb{R}^2$ and $h \in (0, \frac{1}{L})$, $\exists!$ soln $\Psi(\xi, \eta; h, f) \in \mathbb{R}$.

Explicit one-step methods

A general **explicit one-step method** can be written as

$$y_{n+1} = y_n + h\Phi(x_n, y_n; h, f), \quad n \in \{0, \dots, N-1\}, \quad y_0 = y(x_0),$$

where $\Phi = \Phi(\xi, \eta; h, f)$ is a continuous fct in the arguments $\xi, \eta; h$.
In this case,

$$\Psi(\xi, \eta; h, f) = \eta + h\Phi(\xi, \eta; h, f).$$

Example: for explicit Euler, we have $\Phi(\xi, \eta; h, f) = f(\xi, \eta)$ and $\Psi(\xi, \eta; h, f) = \eta + hf(\xi, \eta)$.

Rk: $\Phi(\xi, \eta; h, f)$ and $\Psi(\xi, \eta; h, f)$ can be explicitly computed.

From now on: we do not indicate the dependence of $\Phi(\xi, \eta; h, f)$ on f, and will write $\Phi(\xi, \eta; h)$ instead. E.g., for explicit Euler: $\Phi(\xi, \eta; h) = f(\xi, \eta)$.

2.4 General explicit one-step methods

Global error and consistency error

Recall: a general **explicit one-step method** can be written as

$$y_{n+1} = y_n + h\Phi(x_n, y_n; h), \quad n \in \{0, \dots, N-1\}, \quad y_0 = y(x_0),$$

where $\Phi(\cdot, \cdot; \cdot)$ is a continuous fct.

We define the **global error**, e_n, by

$$e_n := y(x_n) - y_n, \qquad n \in \{0, \dots, N\}.$$

We define the **consistency error**, T_n, by

$$T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - \Phi(x_n, y(x_n); h), \qquad n \in \{0, \dots, N-1\}.$$

Rk: For an implicit one-step method $y_{n+1} = y_n + h\Phi(x_n, y_n, y_{n+1}; h)$, the consistency error is defined by

$$T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - \Phi(x_n, y(x_n), y(x_{n+1}); h), \quad n \in \{0, \dots, N-1\}.$$

An error bound

Theorem (Error bound for general explicit one-step methods)

Consider the general one-step method

$$y_{n+1} = y_n + h\Phi(x_n, y_n; h), \quad n \in \{0, \ldots, N-1\}, \quad y_0 = y(x_0),$$

where, in addition to being a continuous fct, Φ *is assumed to satisfy a Lipschitz condition (w.r.t. its 2nd arg.); namely,* $\exists L_\Phi, h_0 > 0$ *s.t., for* $h \in [0, h_0]$ *and for the same region* R *as in Picard's Theorem,*

$$|\Phi(x, z; h) - \Phi(x, \tilde{z}; h)| \leq L_\Phi |z - \tilde{z}|, \qquad \forall (x, z), (x, \tilde{z}) \in \mathsf{R}.$$

Assume that $|y_n - y_0| \leq Y_M \; \forall n \in \{0, \ldots, N\}$. *Then,*

$$|e_n| \leq e^{L_\Phi (x_n - x_0)} |e_0| + \frac{e^{L_\Phi (x_n - x_0)} - 1}{L_\Phi} T, \qquad n \in \{0, \ldots, N\},$$

where $T := \max_{n \in \{0, \ldots, N-1\}} |T_n|$.

Proof of the error bound

From the defn of the method and the defn of T_n, we have

$$y_{n+1} = y_n + h\Phi(x_n, y_n; h), \quad y(x_{n+1}) = y(x_n) + h\Phi(x_n, y(x_n); h) + hT_n.$$

Subtracting the first equality from the second, we find that

$$e_{n+1} = e_n + h[\Phi(x_n, y(x_n); h) - \Phi(x_n, y_n; h)] + hT_n.$$

Since $(x_n, y(x_n)), (x_n, y_n) \in \mathrm{R}$, the Lipschitz condition implies

$$|e_{n+1}| \leq |e_n| + hL_\Phi|e_n| + h|T_n| \leq (1 + hL_\Phi)|e_n| + hT.$$

By induction, we then can show

$$|e_n| \leq (1 + hL_\Phi)^n|e_0| + \frac{(1 + hL_\Phi)^n - 1}{L_\Phi}T.$$

Finally, using $1 + x \leq e^x \ \forall x \in \mathbb{R}$, and $nh = x_n - x_0$, we have

$$|e_n| \leq e^{nhL_\Phi}|e_0| + \frac{e^{nhL_\Phi} - 1}{L_\Phi}T = e^{L_\Phi(x_n - x_0)}|e_0| + \frac{e^{L_\Phi(x_n - x_0)} - 1}{L_\Phi}T.$$

□

Applying the thm to a specific IVP solved by expl. Euler

Consider the IVP

$$y'(x) = \arctan(y(x)) \quad \text{for} \quad x \in (0,1), \qquad y(0) = 1.$$

Here, $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,z) := \arctan(z)$. (Exercise: Show $\exists!$ soln on $[0,1]$.)
Use explicit Euler to approximate the soln $\implies \Phi(x,z;h) := f(x,z)$.

Compute a Lip. const. L_Φ: $|\partial_z f(x,z)| = \frac{1}{1+z^2} \leq 1 \quad \forall (x,z) \in \mathbb{R}^2$.
$\implies \Phi$ satisfies a global Lip. cond. with $L_\Phi := 1$.

Rk: As Φ satisfies a global Lip.cond., we see from the proof of the general error bound that the assumption $|y_n - y_0| \leq Y_M \; \forall n$ is not necessary.

The general error bound gives for our case (using $e_0 = 0$):

$$|e_n| \leq \frac{e^{L_\Phi(x_n - x_0)} - 1}{L_\Phi} T \leq (e^{x_n} - 1)\frac{M_2}{2} h, \qquad M_2 := \max_{x \in [0,1]} |y''(x)|,$$

where we have used that $T = \max_{n \in \{0,\dots,N-1\}} |T_n| \leq \frac{M_2}{2} h$ for expl. Euler.

Applying the thm to a specific IVP solved by expl. Euler

Consider the IVP

$$y'(x) = \arctan(y(x)) \quad \text{for} \quad x \in (0,1), \qquad y(0) = 1.$$

Here, $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x,z) := \arctan(z)$. Use explicit Euler to approximate the soln. We have shown

$$|e_n| \leq (e^{x_n} - 1)\frac{M_2}{2}h, \qquad M_2 := \max_{x \in [0,1]} |y''(x)|.$$

Let us bound M_2: For any $x \in [0,1]$, we have

$$|y''(x)| = \left|\frac{\mathrm{d}}{\mathrm{d}x}(\arctan(y(x)))\right| = \frac{|y'(x)|}{1 + [y(x)]^2} = \frac{|\arctan(y(x))|}{1 + [y(x)]^2} \leq \frac{\pi}{2}.$$

So, $M_2 \leq \frac{\pi}{2}$ and thus,

$$|e_n| \leq (e^{x_n} - 1)\frac{\pi}{4}h \leq \frac{\pi(e-1)}{4}h \qquad \forall n \in \{0, \ldots, N\}.$$

$\Longrightarrow$ For given $\mathtt{TOL} > 0$, we have $|e_n| \leq \mathtt{TOL}$ $\forall n$ when $h \leq \frac{4}{\pi(e-1)}\mathtt{TOL}$.

Consistency

Error bd for general explicit one-step method (assuming $e_0 = 0$):

$$|e_n| \le \frac{e^{L_\Phi(x_n - x_0)} - 1}{L_\Phi} T, \qquad n \in \{0, \ldots, N\}.$$

$\Longrightarrow$ Consistency error "decides" whether global error converges to 0.

Definition (Consistent method)

An explicit one-step scheme is **consistent** with the DE if the consistency error is such that $\forall \varepsilon > 0 \; \exists h_\varepsilon > 0$ s.t. $|T_n| < \varepsilon$ for all $h \in (0, h_\varepsilon)$ and any points $(x_n, y(x_n))$, $(x_{n+1}, y(x_{n+1}))$ on the graph of y.

Recall the defn

$$T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - \Phi(x_n, y(x_n); h).$$

We see that for any $x \in [x_0, X_M]$, we have

$$\lim_{h\to 0, n\to\infty, x_n \to x \in [x_0, X_M]} T_n = y'(x) - \Phi(x, y(x); 0) = f(x, y(x)) - \Phi(x, y(x); 0)$$

$\Longrightarrow$ the one-step method is consistent iff $\Phi(x, y; 0) \equiv f(x, y)$.

Convergence theorem for general explicit one-step methods

Theorem

Suppose the true soln of the IVP and its approximation lie in R *when* $h \le h_0$. *Suppose also that* $\Phi(\cdot, \cdot; \cdot)$ *is uniformly continuous on* $\mathsf{R} \times [0, h_0]$ *and satisfies the consistency condition* $\Phi(x, y; 0) \equiv f(x, y)$ *and the Lipschitz condition*

$$|\Phi(x, z; h) - \Phi(x, \tilde{z}; h)| \le L_\Phi |z - \tilde{z}| \qquad \forall (x, z, h), (x, \tilde{z}, h) \in \mathsf{R} \times [0, h_0].$$

Then, if successive approximation sequences (y_n), *generated for* $x_n = x_0 + nh$, $n \in \{1, \ldots, N\}$, *are obtained from the method with successively smaller values of* h, *each less than* h_0, *we have convergence of the numerical solution to the solution of the IVP in the sense that*

$$|y(x) - y_n| \longrightarrow 0 \qquad \text{as} \quad h \to 0,\ n \to \infty,\ x_n \to x \in [x_0, X_M].$$

Order of accuracy of explicit one-step methods

We saw that for explicit Euler we have $|T_n| = \mathcal{O}(h)$, i.e., $\exists h_0, K > 0$ s.t.

$$|T_n| \leq Kh \qquad \forall h \in (0, h_0].$$

However, there are other one-step methods (which we will discuss later) for which we can do better $(\mathcal{O}(h^2), \mathcal{O}(h^3), \dots)$.

Definition (Order of accuracy)

The method is said to have **order of accuracy** p (or **order of consistency** p), if $p \in \mathbb{N}$ is the largest natural number s.t., for any sufficiently smooth solution curve $(x, y(x))$ in R of the IVP we have

$$|T_n| = \mathcal{O}(h^p),$$

i.e., there exist constants $h_0, K > 0$ such that $|T_n| \leq Kh^p$ for all $h \in (0, h_0]$, for any pair of points $(x_n, y(x_n))$, $(x_{n+1}, y(x_{n+1}))$ on the solution curve.

2.5 Explicit Runge–Kutta methods

Explicit Runge–Kutta methods

Motivation: explicit Euler is only first-order accurate (but cheap to implement because we only need one evaluation of f in each step).

Runge–Kutta (RK) methods: higher accuracy by sacrificing the efficiency of explicit Euler through re-evaluating f:

General R-**stage explicit RK family**: $y_{n+1} = y_n + h\Phi(x_n, y_n; h)$, where

$$\begin{aligned}
\Phi(x,z;h) &= \sum_{r=1}^{R} c_r k_r(x,z;h), \\
k_1(x,z;h) &= f(x,z), \\
k_r(x,z;h) &= f\left(x + ha_r, z + h\sum_{s=1}^{r-1} b_{rs} k_s(x,z;h)\right), \qquad r \in \{2,\dots,R\}.
\end{aligned}$$

General Runge–Kutta methods

General version of a R-stage RK method:

$$y_{n+1} = y_n + h \sum_{r=1}^{R} c_r k_r,$$

$$k_r = f\left(x_n + ha_r, y_n + h \sum_{s=1}^{R} b_{rs} k_s\right) \quad \text{for} \quad r \in \{1, \ldots, R\}.$$

If the method is not a R-stage explicit RK method, then it is called a R-**stage implicit RK method**.

A RK method is usually displayed in the so-called **Butcher tableau**

$$\begin{array}{c|c} a & B \\ \hline & c^{\mathsf{T}} \end{array}$$

where

$$a = (a_1, \ldots, a_R)^{\mathrm{T}}, \qquad B = (b_{ij})_{1 \le i,j \le R}, \qquad c = (c_1, \ldots, c_R)^{\mathrm{T}}.$$

For explicit RK methods, the matrix B is strictly lower-triangular.

1-stage explicit RK methods

Suppose that $R = 1$, i.e., $y_{n+1} = y_n + h\Phi(x_n, y_n; h)$, where

$$\begin{aligned}\Phi(x, z; h) &= c_1 k_1(x, z; h), \\ k_1(x, z; h) &= f(x, z).\end{aligned}$$

$\Longrightarrow y_{n+1} = y_n + hc_1 f(x_n, y_n)$.

The method is consistent iff $\Phi(x, y; 0) \equiv f(x, y)$, i.e., iff $c_1 = 1$.

The resulting one-stage explicit RK method is the explicit Euler method:

$$y_{n+1} = y_n + hf(x_n, y_n).$$

This is first-order accurate. Butcher tableau:

$$\begin{array}{c|c} 0 & 0 \\ \hline & 1 \end{array}$$

2-stage explicit RK methods

Suppose that $R = 2$, i.e., $y_{n+1} = y_n + h\Phi(x_n, y_n; h)$, where

$$\begin{aligned} \Phi(x, z; h) &= c_1 k_1(x, z; h) + c_2 k_2(x, z; h), \\ k_1(x, z; h) &= f(x, z), \\ k_2(x, z; h) &= f(x + a_2 h, z + b_{21} h k_1(x, z; h)). \end{aligned}$$

Q: Can we choose a_2, b_{21}, c_1, c_2 s.t. the method is second-order accurate (or even better)?

Consistency $\iff \Phi(x, y; 0) \equiv f(x, y) \iff c_1 f(x, y) + c_2 f(x, y) \equiv f(x, y)$, i.e., the method is consistent iff $c_1 + c_2 = 1$.

Consistency error: $T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - \Phi(x_n, y(x_n); h)$, i.e.,

$$T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - c_1 f(x_n, y(x_n)) - c_2 f(x_n + a_2 h, y(x_n) + b_{21} h f(x_n, y(x_n)))$$

2-stage explicit RK methods: Expansion of T_n

We want to expand T_n in powers of h. Recall:

$$T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - c_1 f(x_n, y(x_n)) - c_2 f(x_n + a_2 h, y(x_n) + b_{21} h f(x_n, y(x_n)))$$

Taylor expansion for the first term:

$$\frac{y(x_{n+1}) - y(x_n)}{h} = y'(x_n) + \frac{1}{2} h y''(x_n) + \frac{1}{6} h^2 y'''(x_n) + \mathcal{O}(h^3).$$

Taylor expansion for the third term:

$$\begin{aligned}
&f(x_n + a_2 h, y(x_n) + b_{21} h f(x_n, y(x_n))) \\
&= \left[f + a_2 h f_x + b_{21} h f f_z + \frac{a_2^2}{2} h^2 f_{xx} + a_2 b_{21} h^2 f f_{xz} + \frac{b_{21}^2}{2} h^2 f^2 f_{zz} \right] (x_n, y(x_n)) \\
&\quad + \mathcal{O}(h^3).
\end{aligned}$$

Noting that $y'(x_n) = f(x_n, y(x_n))$ and recalling $c_1 + c_2 = 1$, we find

$$\begin{aligned}
T_n =& \frac{1}{2} h y''(x_n) + \frac{1}{6} h^2 y'''(x_n) - c_2 h [a_2 f_x + b_{21} f f_z](x_n, y(x_n)) \\
&- c_2 h^2 \left[\frac{1}{2} a_2^2 f_{xx} + a_2 b_{21} f f_{xz} + \frac{1}{2} b_{21}^2 f^2 f_{zz} \right] (x_n, y(x_n)) + \mathcal{O}(h^3).
\end{aligned}$$

We have obtained that

$$T_n = \frac{1}{2} h y''(x_n) + \frac{1}{6} h^2 y'''(x_n) - c_2 h [a_2 f_x + b_{21} f f_z](x_n, y(x_n))$$
$$- c_2 h^2 \left[\frac{1}{2} a_2^2 f_{xx} + a_2 b_{21} f f_{xz} + \frac{1}{2} b_{21}^2 f^2 f_{zz} \right] (x_n, y(x_n)) + \mathcal{O}(h^3).$$

To continue, note that from $y'(x) = f(x, y(x))$, we find

$$y''(x) = f_x(x, y(x)) + y'(x) f_z(x, y(x)) = [f_x + f f_z](x, y(x)) = F_1(x, y(x)),$$
$$y'''(x) = [f_{xx} + f_x f_z + f f_{xz}](x, y(x)) + y'(x)[f_{xz} + f_z^2 + f f_{zz}](x, y(x))$$
$$= [f_x f_z + f f_z^2 + f_{xx} + 2 f f_{xz} + f^2 f_{zz}](x, y(x)) = [f_z F_1 + F_2](x, y(x))$$

where $F_1 := f_x + f f_z$ and $F_2 := f_{xx} + 2 f f_{xz} + f^2 f_{zz}$. We obtain

$$T_n = h \left[\frac{1}{2} F_1 - a_2 c_2 f_x - b_{21} c_2 f f_z \right] (x_n, y(x_n))$$
$$+ h^2 \left[\frac{1}{6} f_z F_1 + \frac{1}{6} F_2 - \frac{1}{2} a_2^2 c_2 f_{xx} - a_2 b_{21} c_2 f f_{xz} - \frac{1}{2} b_{21}^2 c_2 f^2 f_{zz} \right] (x_n, y(x_n))$$
$$+ \mathcal{O}(h^3).$$

2-stage explicit RK methods: Conditions for $T_n = \mathcal{O}(h^2)$

We have obtained that

$$\begin{aligned} T_n = h &\left[\frac{1}{2}F_1 - a_2c_2f_x - b_{21}c_2ff_z\right](x_n, y(x_n)) \\ &+ h^2\left[\frac{1}{6}f_zF_1 + \frac{1}{6}F_2 - \frac{1}{2}a_2^2c_2f_{xx} - a_2b_{21}c_2ff_{xz} - \frac{1}{2}b_{21}^2c_2f^2f_{zz}\right](x_n, y(x_n)) \\ &+ \mathcal{O}(h^3), \end{aligned}$$

where $F_1 := f_x + ff_z$ and $F_2 := f_{xx} + 2ff_{xz} + f^2f_{zz}$.

Q: Can we get $T_n = \mathcal{O}(h^2)$ for any f?

$\Longrightarrow$ Yes! Choose the parameters s.t. $a_2c_2 = \frac{1}{2}$, $b_{21}c_2 = \frac{1}{2}$, $c_1 + c_2 = 1$, i.e.,

$$b_{21} = a_2, \qquad c_2 = \frac{1}{2a_2}, \qquad c_1 = 1 - \frac{1}{2a_2}.$$

This still leaves one free parameter, a_2, but no choice of the parameters will yield $T_n = \mathcal{O}(h^3)$ for any f.

Examples of second-order explicit RK methods

Recall general 2-stage explicit RK method:

$$y_{n+1} = y_n + h\left[c_1 f(x_n, y_n) + c_2 f(x_n + a_2 h, y_n + b_{21} h f(x_n, y_n))\right].$$

Recall: if $b_{21} = a_2$, $c_2 = \frac{1}{2a_2}$, $c_1 = 1 - \frac{1}{2a_2}$, then second-order accurate.

- **The modified Euler method:** $a_2 := \frac{1}{2}$, $b_{21} := \frac{1}{2}$, $c_1 := 0$, $c_2 := 1$.

$$y_{n+1} = y_n + h\, f\left(x_n + \frac{h}{2}, y_n + \frac{h}{2} f(x_n, y_n)\right).$$

$$\begin{array}{c|cc} 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ \hline & 0 & 1 \end{array}$$

- **The improved Euler method:** $a_2 := 1$, $b_{21} := 1$, $c_1 := \frac{1}{2}$, $c_2 := \frac{1}{2}$.

$$y_{n+1} = y_n + \frac{h}{2}\left[f(x_n, y_n) + f(x_n + h, y_n + h f(x_n, y_n))\right].$$

$$\begin{array}{c|cc} 0 & 0 & 0 \\ 1 & 1 & 0 \\ \hline & \frac{1}{2} & \frac{1}{2} \end{array}$$

3-stage explicit RK methods: the autonomous case

Suppose that $R = 3$ and $f(x,z) = \tilde{f}(z)$ is independent of x (the ODE $y'(x) = \tilde{f}(y(x))$ is called **autonomous**). Then,

$$y_{n+1} = y_n + h\Phi(x_n, y_n; h),$$

where $\Phi(x,z;h) = c_1\tilde{k}_1(z;h) + c_2\tilde{k}_2(z;h) + c_3\tilde{k}_3(z;h)$ and

$$\begin{aligned}
\tilde{k}_1(z;h) &= \tilde{f}(z), \\
\tilde{k}_2(z;h) &= \tilde{f}(z + hb_{21}\tilde{k}_1(z;h)), \\
\tilde{k}_3(z;h) &= \tilde{f}(z + hb_{31}\tilde{k}_1(z;h) + hb_{32}\tilde{k}_2(z;h)).
\end{aligned}$$

Q: Can we choose the parameters s.t. the method is third-order accurate (or even better)?

Consistency $\iff \Phi(x,y;0) \equiv \tilde{f}(y) \iff c_1\tilde{f}(y) + c_2\tilde{f}(y) + c_3\tilde{f}(y) \equiv \tilde{f}(y)$ $\iff c_1 + c_2 + c_3 = 1$.

Consistency error: $T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - \Phi(x_n, y(x_n); h)$, i.e.,

$$T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - c_1\tilde{f}(y(x_n)) - c_2\tilde{k}_2(y(x_n);h) - c_3\tilde{k}_3(y(x_n);h).$$

3-stage exp. RK methods, autonomous case: expanding T_n

We want to expand T_n in powers of h. Recall:

$$T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - c_1 \tilde{f}(y(x_n)) - c_2 \tilde{k}_2(y(x_n); h) - c_3 \tilde{k}_3(y(x_n); h)$$

Taylor expansion for $\tilde{k}_2(y(x_n); h)$:

$$\tilde{k}_2(y(x_n); h) := \tilde{f}(y(x_n) + hb_{21}\tilde{f}(y(x_n))) = \left[\tilde{f} + hb_{21}\tilde{f}\tilde{f}' + h^2 \frac{b_{21}^2}{2}\tilde{f}^2\tilde{f}''\right](y(x_n)) + \mathcal{O}(h^3).$$

Taylor expansion for $\tilde{k}_3(y(x_n); h)$:

$$k_3(y(x_n); h) := \tilde{f}(y(x_n) + hb_{31}\tilde{f}(y(x_n)) + hb_{32}\tilde{k}_2(y(x_n); h))$$

$$= \left[\tilde{f} + h\left(b_{31}\tilde{f}\tilde{f}' + b_{32}(\tilde{f} + hb_{21}\tilde{f}\tilde{f}' + \mathcal{O}(h^2))\tilde{f}'\right) + h^2\left(\frac{1}{2}(b_{31}\tilde{f} + b_{32}(\tilde{f} + \mathcal{O}(h)))^2\tilde{f}''\right)\right](y(x_n))$$

$$+ \mathcal{O}(h^3)$$

$$= \left[\tilde{f} + h\left((b_{31} + b_{32})\tilde{f}\tilde{f}'\right) + h^2\left(b_{21}b_{32}\tilde{f}\tilde{f}'^2 + \frac{1}{2}(b_{31} + b_{32})^2\tilde{f}^2\tilde{f}''\right)\right](y(x_n)) + \mathcal{O}(h^3).$$

We find that

$$\begin{aligned}\Phi(x_n, y(x_n); h) &= c_1 \tilde{f}(y(x_n)) + c_2 \tilde{k}_2(y(x_n); h) + c_3 \tilde{k}_3(y(x_n); h) \\ &= (c_1 + c_2 + c_3)\tilde{f}(y(x_n)) + h\left[(b_{21}c_2 + (b_{31} + b_{32})c_3)\tilde{f}\tilde{f}'\right](y(x_n)) \\ &\quad + h^2\left[b_{21}b_{32}c_3\tilde{f}\tilde{f}'^2 + \frac{b_{21}^2 c_2 + (b_{31} + b_{32})^2 c_3}{2}\tilde{f}^2\tilde{f}''\right](y(x_n)) + \mathcal{O}(h^3)\end{aligned}$$

and we also have that

$$\frac{y(x_{n+1}) - y(x_n)}{h} = y'(x_n) + \frac{1}{2}hy''(x_n) + \frac{1}{6}h^2 y'''(x_n) + \mathcal{O}(h^3)$$

$$= 1 \cdot \tilde{f}(y(x_n)) + h\left[\frac{1}{2}\tilde{f}\tilde{f}'\right](y(x_n)) + h^2\left[\frac{1}{6}\tilde{f}\tilde{f}'^2 + \frac{1}{6}\tilde{f}^2\tilde{f}''\right](y(x_n)) + \mathcal{O}(h^3).$$

Noting that $T_n = \frac{y(x_{n+1}) - y(x_n)}{h} - \Phi(x_n, y(x_n); h)$, we conclude that we can achieve $T_n = \mathcal{O}(h^3)$ if we choose the parameters s.t.

$$c_1 + c_2 + c_3 = 1, \qquad b_{21}c_2 + (b_{31} + b_{32})c_3 = \frac{1}{2},$$

$$b_{21}^2 c_2 + (b_{31} + b_{32})^2 c_3 = \frac{1}{3}, \qquad b_{21}b_{32}c_3 = \frac{1}{6}.$$

Solving for the six unknowns, we obtain a two-parameter family of third-order accurate 3-stage explicit RK methods.

Examples of third-order explicit RK methods

Recall third-order accuracy conditions (for the autonomous case):

$$c_1 + c_2 + c_3 = 1, \qquad b_{21}c_2 + (b_{31} + b_{32})c_3 = \frac{1}{2},$$
$$b_{21}^2 c_2 + (b_{31} + b_{32})^2 c_3 = \frac{1}{3}, \qquad b_{21}b_{32}c_3 = \frac{1}{6},$$

- **Heun's method**: $c_1 = \frac{1}{4}$, $c_2 = 0$, $c_3 = \frac{3}{4}$, $b_{21} = \frac{1}{3}$, $b_{31} = 0$, $b_{32} = \frac{2}{3}$.

$$\begin{aligned} y_{n+1} &= y_n + h\left(\frac{1}{4}\tilde{k}_1 + \frac{3}{4}\tilde{k}_3\right), \\ \tilde{k}_1 &= \tilde{f}(y_n), \quad \tilde{k}_2 = \tilde{f}\left(y_n + \frac{1}{3}h\tilde{k}_1\right), \quad \tilde{k}_3 = \tilde{f}\left(y_n + \frac{2}{3}h\tilde{k}_2\right). \end{aligned}$$

- **Standard third-order explicit RK method**: $c_1 = \frac{1}{6}$, $c_2 = \frac{2}{3}$, $c_3 = \frac{1}{6}$, $b_{21} = \frac{1}{2}$, $b_{31} = -1$, $b_{32} = 2$.

$$\begin{aligned} y_{n+1} &= y_n + h\left(\frac{1}{6}\tilde{k}_1 + \frac{2}{3}\tilde{k}_2 + \frac{1}{6}\tilde{k}_3\right), \\ \tilde{k}_1 &= \tilde{f}(y_n), \quad \tilde{k}_2 = \tilde{f}\left(y_n + \frac{1}{2}h\tilde{k}_1\right), \quad \tilde{k}_3 = \tilde{f}\left(y_n - h\tilde{k}_1 + 2h\tilde{k}_2\right) \end{aligned}$$

R-stage explicit RK methods with $R \geq 4$

For $R = 4$, an analogous argument leads to a two-parameter family of four-stage RK methods of order four. Popular example:

$$y_{n+1} = y_n + \frac{1}{6}h\left(k_1 + 2k_2 + 2k_3 + k_4\right),$$

where

$$k_1 = f(x_n, y_n), \qquad k_2 = f\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_1\right),$$

$$k_3 = f\left(x_n + \frac{1}{2}h, y_n + \frac{1}{2}hk_2\right), \qquad k_4 = f(x_n + h, y_n + hk_3).$$

We have constructed R-stage explicit RK methods of order of accuracy $\mathcal{O}(h^R)$, $R = 1, 2, 3, 4$.

Q: Is there an R-stage method of order R for $R \geq 5$?

$\Longrightarrow$ No. For $R = 5, 6, 7, 8, 9$, the highest order that can be attained by a R-stage RKm is $4, 5, 6, 6, 7$. For $R \geq 10$, highest order is $\leq R - 2$.

2.6 Absolute stability of explicit Runge–Kutta methods

The model problem

We consider the **model problem**

$$y'(x) = \lambda y(x), \quad y(0) = y_0,$$

with $\lambda < 0$ and $y_0 \neq 0$. The true solution to this IVP is

$$y(x) = y_0 \, e^{\lambda x}.$$

Observe that $y(x) \to 0$ as $x \to \infty$.

Q: Under what conditions on the step size h does a RK method reproduce this behavior?

$\Longrightarrow$ the answer gives information on how to choose h in the approximation of an IVP by an explicit RK method over $[x_0, X_M]$ with $X_M \gg x_0$.

We answer this question for R-stage explicit RK methods of order of accuracy R, for $R = 1, 2, 3, 4$.

Absolute stability of 1st-order accurate 1-stage explicit RK

The only first-order accurate 1-stage explicit RK method is explicit Euler:

$$y_{n+1} = y_n + hf(x_n, y_n).$$

We apply this to the model problem

$$y'(x) = \lambda y(x), \qquad y(0) = y_0.$$

Here, $f : \mathbb{R}^2 \to \mathbb{R}$, $f(x, z) := \lambda z$. We find

$$y_{n+1} = y_n + hf(x_n, y_n) = y_n + \lambda h y_n = (1 + \bar{h}) y_n,$$

where $\bar{h} := \lambda h$. Thus, for any $n \in \mathbb{N}_0$ we have

$$y_n = (1 + \bar{h})^n y_0.$$

$\Longrightarrow$ $(y_n)_{n \in \mathbb{N}_0}$ will converge to 0 iff $|1 + \bar{h}| < 1$, i.e., iff $\bar{h} \in (-2, 0)$.

For h chosen s.t. $\bar{h} \in (-2, 0)$, explicit Euler is said to be **absolutely stable**, and $(-2, 0)$ is called **interval of absolute stability** of the method.

Absolute stability of 2nd-order accurate 2-stage explicit RK

Consider a second-order accurate 2-stage explicit RK method:

$$y_{n+1} = y_n + h(c_1 k_1 + c_2 k_2), \quad k_1 = f(x_n, y_n), \quad k_2 = f(x_n + a_2 h, y_n + b_{21} h k_1)$$

with parameters satisfying $c_1 + c_2 = 1$ and $a_2 c_2 = b_{21} c_2 = \frac{1}{2}$ (to have second-order accuracy). Apply this to model problem ($f(x, z) = \lambda z$):

$$\begin{aligned} y_{n+1} &= y_n + h(c_1 \lambda y_n + c_2 \lambda (y_n + b_{21} h \lambda y_n)) \\ &= \left(1 + (c_1 + c_2)\lambda h + b_{21} c_2 \lambda^2 h^2\right) y_n = \left(1 + \bar{h} + \frac{1}{2}\bar{h}^2\right) y_n. \end{aligned}$$

(Recall $\bar{h} := \lambda h$.) Thus, for any $n \in \mathbb{N}_0$ we have

$$y_n = \left(1 + \bar{h} + \frac{1}{2}\bar{h}^2\right)^n y_0.$$

$\Longrightarrow$ Method absolutely stable ($y_n \to 0$ as $n \to \infty$) iff $\left|1 + \bar{h} + \frac{1}{2}\bar{h}^2\right| < 1$ iff $\frac{1}{2}(\bar{h} + 1)^2 + \frac{1}{2} < 1$ iff $(\bar{h} + 1)^2 < 1$ iff $\bar{h} + 1 \in (-1, 1)$ iff $\bar{h} \in (-2, 0)$.

Absolute stability of 3rd-order accurate 3-stage explicit RK

For third-order accurate 3-stage explicit RK methods, an analogous argument yields

$$y_{n+1} = \left(1 + \bar{h} + \frac{1}{2}\bar{h}^2 + \frac{1}{6}\bar{h}^3\right) y_n$$

and hence, for any $n \in \mathbb{N}_0$,

$$y_n = \left(1 + \bar{h} + \frac{1}{2}\bar{h}^2 + \frac{1}{6}\bar{h}^3\right)^n y_0.$$

$\Longrightarrow$ Method absolutely stable ($y_n \to 0$ as $n \to \infty$) iff

$$\left|1 + \bar{h} + \frac{1}{2}\bar{h}^2 + \frac{1}{6}\bar{h}^3\right| < 1,$$

which yields the interval of absolute stability: $\bar{h} \in (-a, 0)$ with $a \approx 2.51$.

Absolute stability of 4th-order accurate 4-stage explicit RK

For fourth-order accurate 4-stage explicit RK methods, an analogous argument yields

$$y_{n+1} = \left(1 + \bar{h} + \frac{1}{2}\bar{h}^2 + \frac{1}{6}\bar{h}^3 + \frac{1}{24}\bar{h}^4\right) y_n,$$

and the interval of absolute stability is $\bar{h} \in (-a, 0)$ with $a \approx 2.78$.

Rk: For $R \geq 5$, applying the explicit RK method to model problem gives

$$y_{n+1} = A_R(\bar{h}) y_n,$$

however, unlike the cases $R = 1, 2, 3, 4$, in addition to $\bar{h}$ the expression $A_R(\bar{h})$ also depends on the coefficients of the explicit RK method.

End of “Chapter 2: One-step methods”.