

Runge-Kutta time-stepping

Hundsdorfer - Verwer, 2003 p. 140 - 214

for ODE system $w'(t) = F(t, w(t))$, $w(0) = w_0$

General form of RK: $w^{n+1} = w^n + \Delta t \sum_{k=1}^s b_k \cdot F(t_n + c_k \Delta t, w^{n,k})$, $c_k = \sum_{l=1}^s a_{kl}$

$$w^{n,k} = w^n + \Delta t \sum_{l=1}^s a_{kl} F(t_n + c_l \Delta t, w^{n,l}), k=1, \dots, s$$

often represented by the Butcher-arrays: $c_1 \mid a_{11} \cdots a_{1s}$ $s = \# \text{ of stages}$

$$\begin{array}{c|ccccc} & & a_{21} & \cdots & a_{2s} \\ c_s \mid & a_{s1} & \cdots & a_{ss} \\ \hline & b_1 & \cdots & b_s \end{array}$$

$p = \text{order}$ hard to figure out

$q = \text{stage-order}$

Such a method is explicit if $a_{lk} = 0$ for $l \geq k$ ($[a_{lk}]$ lower triangular)

For up to $s=4$, order $p=s$, but for $p \geq 5$, $p < s$.

e.g. $\begin{array}{c|c} 0 & 0 \\ \hline 1 & 1 \end{array}$ $w^{n+1} = w^n + \Delta t \cdot F(t_n, w^n)$ is the explicit Forward Euler; $s=1$

$\begin{array}{c|cc} 0 & 1 & 1 \\ \hline 1 & 1 & 1 \end{array}$ is backward Euler

$\begin{array}{c|cc} 0 & 1 & 1 \\ \hline 1 & 1/2 & 1/2 \end{array}$ $w^{n+1} = w^n + \frac{\Delta t}{2} F(t_n, w^n) + \frac{\Delta t}{2} F(t_n + \Delta t, w^n + \Delta t F(t_n, w^n))$

has $s=2$, $p=2$ is the explicit trapezoidal rule (modified Euler)

$$\begin{array}{c|cccc} 0 & & & & \\ \hline 1/2 & 1/2 & & & \\ 1/2 & 0 & 1/2 & & \\ \hline 1 & 0 & 0 & 1 & \\ \hline & 1/6 & 1/3 & 1/3 & 1/6 \end{array}$$

is the 4-order, 4-stage Classical RK (rk4 of matlab)

The CFL stability condition may depend on s , restricting Δt more for higher s .

Order reduction: Very often, RK used as time-integrator for PDEs suffer order reduction

e.g. a 4th order method may produce order 2 or 2.5!

Thus, the extra cost of using more stages (to get higher order)
may not pay off!

It has been studied to death, there are some tricks to diminish it. for some PDEs,
but problem dependent, no good way out...

Multistep methods

Hundsdorfer-Verner 2003 p. 170 -

for ODE system $\dot{w}(t) = F(t, w(t))$, $w(0) = w_0$ $w(t) \in \mathbb{R}^m$

Linear k-step method: use k previous steps w_n, \dots, w^{n+k-1} to compute w^{n+k}

$$\sum_{j=0}^k \alpha_j w^{n+j} = \tau \sum_{j=0}^k \beta_j F(t_{n+j}, w^{n+j}), \quad n=0, 1, \dots$$

$\alpha_k > 0$

explicit if $\beta_k = 0$, else implicit

For $k=1$ it is also RK, θ-scheme can be viewed as 1-step multistep

For $k \geq 2$ these are essentially different than single-step methods like RK.

Advantages: 1. only one F-evaluation per time step (vs s evaluations for s-stage RK)

for explicit methods, or only one system to be solved for implicit methods.

2. easier to develop higher order schemes

3. adaptive schemes are easier to develop and implement than for RK

4. great for predictor-corrector

5. for BDF methods for stiff problems

Disadvantage: not self-starting: needs k steps w^0, \dots, w^{k-1} to begin

but only have initial values w^0 . Some single-step method must be used to cook up w^1, \dots, w^{k-1} , either some RK
(or often, Euler with (very) small Δt)

Adams methods: $\alpha_k = 1, \alpha_{k-1} = -1, \alpha_j = 0$ for $j = k-2, \dots, 1, 0$.

β_j chosen for optimal order

Adams-Basforth are explicit Adams of order k :

$k=1$ is forward Euler

$$k=2: w^{n+2} = w^{n+1} + \frac{\tau}{2} [3F^{n+1} - F^n]$$

$$F^j := F(t_j, w^j)$$

$$k=3: w^{n+3} = w^{n+2} + \frac{\tau}{12} [23F^{n+2} - 16F^{n+1} + 5F^n]$$

Adams-Moulton are implicit Adams of order $k+1$ (system to be solved)

$k=1$ is Trapezoidal rule

$$k=2: w^{n+2} = w^{n+1} + \frac{\tau}{12} [5 \cdot F^{n+2} + 8F^{n+1} - F^n]$$

Predictor-corrector schemes: use an explicit k -step Adams to predict \tilde{W}^{n+k} , then plug into $F^{n+k} \approx F(t_{n+k}, \tilde{W}^{n+k})$ to use in an implicit Adams.

Overall scheme is explicit of order $k+1$
For $k=1$ it is the explicit trapezoidal rule

Backward Differentiation Formulas (BDF): implicit with $\beta_k = 1$, $\beta_j = 0$ for $j=0, \dots, k-1$
 α_j chosen to get order k .

$k=1$ is Backward Euler

$$k=2: \frac{3}{2}W^{n+2} - 2W^{n+1} + \frac{1}{2}W^n = \tau \cdot F^{n+2}$$

They have good stability for stiff

Were introduced by Curtiss-Hirschfelder (1952), popularized by Gear (1971)

Positivity requirement is severe on multistep methods,

Most of the standard multistep with $k \geq 2$ cannot be positive!
and BDF

Some may be under Δt restriction, often severe

For k -step of order p : $\gamma \leq \frac{k-p}{k-1}$, γ = positivity threshold: $\alpha \cdot \Delta t \leq \gamma$ [H-V p.193]

Adam-Basforth has $p=k$ so $\gamma \leq 0$! cannot be positive

Adams-Moulton has $p=k+1$ so $\gamma \leq \frac{1}{k-1}$ bad for high k

Stability for linear system: $\vec{y}' = A\vec{y}$, $A_{n \times n} \Rightarrow$ exact sol $\vec{y} = e^{tA} \vec{y}_0$

Euler scheme: $\vec{Y}^{n+1} = \vec{Y}^n + \Delta t \cdot A \vec{Y}^n$ [set $\tau = \Delta t$]

$$= [I + \tau A] \vec{Y}^n$$

$$= [I + \tau A]^2 \vec{Y}^{n-1} = \dots = [I + \tau A]^{n+1} \vec{Y}^0$$

propagator: $R(\tau A) = I + \tau A$ so $R(z) = I + z$ for Forward Euler ($\approx e^z$ to 1st order)

Any scheme can be viewed as

$$\begin{aligned}\vec{Y}^{n+1} &= R(\tau A) \vec{Y}^n \\ \vec{Y}^n &= R(\tau A)^n \vec{Y}^0\end{aligned}$$

$R(z)$ called the stability function of the scheme
 $z \in \mathbb{C}$ It is a polynomial or rational function

$R(\tau A)^n$ is the propagator from \vec{Y}^0 to \vec{Y}^n , clearly we need it to remain bounded

For θ -scheme: $R(z) = \frac{1 + (1-\theta)z}{1 - \theta z}$, $R(\tau A) = (I - \theta \tau A)^{-1} (I + (1-\theta)\tau A)$

existence of \uparrow means the implicit system has

stable scheme if $\|R(\tau A)^n\| \leq C_s \quad \forall n \tau \leq t_{\max}$ (Lax-Richtmyer stability) unique solution

Stability region of scheme is the set $S = \{z \in \mathbb{C} : |R(z)| \leq 1\}$



If " τA " $\in S$ the scheme will be stable. $R(z)$ is rational function, cont'd on S meaning?

Thm [Hundsdorfer-Verner p.39] If $A = P \Lambda P^{-1}$ with $\Lambda = \text{diag}(\lambda_j)$, $\text{cond}(P) \leq K$ eigenvalues

then $\tau \lambda_j \in S, j=1, \dots, M \Rightarrow \|R(\tau A)\| \leq K \quad \forall n = 1, 2, \dots$

i.e. $|R(\tau \lambda_j)| \leq 1$

(in any absolute vector norm)

but $\text{cond}(P)$ can be very large, $A \sim \frac{1}{\Delta x^2}$ for diffusion

Thm [H-V p.39]: If A is normal matrix then $\tau \lambda_j \in S \Rightarrow \|R(\tau A)\|_2 \leq 1 \Rightarrow \|R(\tau A)^n\|_2 \leq 1$

Good for diffusion where $A = \frac{\Delta x^2}{\Delta t} \Delta x^2$ symm, positive definite, not good for advection, A not normal

Thm [H-V p.40]: If $\operatorname{Re}(\vec{z}, A\vec{z}) \leq \omega \|\vec{z}\|^2 \quad \forall \vec{z} \in \mathbb{C}^M$ then in the induced norm $\| \cdot \| = \sqrt{\langle \cdot, \cdot \rangle}$

$$\|R(\tau A)\| \leq \max_{\operatorname{Re} z \leq \tau \omega} \|R(z)\| \quad \text{provided } 1 - \theta \tau \omega > 0 \quad (\text{energy norm})$$

$\operatorname{Re} z \leq \tau \omega$

(stab. condition)

Thm [Lekque 1990 p.107]: If $\|R(\tau A)\| \leq 1 + \alpha \cdot \tau$ for $\tau \leq \tau_0$ then $\|R(\tau A)^n\| \leq (1 + \alpha \tau)^n \leq e^{\alpha n \tau} \leq e^{\alpha t_{\max}}$ $\forall n \tau \leq t_{\max}$

Consistency -

Stability - convergence for linear system $\vec{y}' = A \vec{y}$, A $m \times m$
for linear schemes

$$\text{Scheme: } \vec{Y}^{n+1} = R(\tau A) \vec{Y}^n$$

$$\text{on exact } y(t): \quad \tilde{Y}^{n+1} = R(\tau A) y(t_n) \quad \text{but}$$

$$\text{discretization error } \varepsilon_{n+1} = Y^{n+1} - y(t_{n+1}) = Y^{n+1} - \tilde{Y}^{n+1} + \tilde{Y}^{n+1} - y(t_{n+1})$$

$$= R(\tau A) Y^n - R(\tau A) y(t_n) + R(\tau A) y(t_n) - y(t_{n+1})$$

$$= R(\tau A) [Y^n - y(t_n)]$$

$$\begin{aligned} & \text{but } y(t_{n+1}) = R(\tau A) y(t_n) + \tau p_n \text{ for Euler} \\ & (I - \theta \tau A) y(t_{n+1}) = \quad \quad \quad \text{for } \theta \text{ schemes} \\ & - \tau p_n \end{aligned}$$

$$\boxed{\varepsilon_{n+1} = R(\tau A) \varepsilon_n + \delta_n}$$

$$\delta_n = -\tau p_n \text{ for Euler}, \quad \delta_n = (I - \theta \tau A)^{-1} \tau p_n \text{ for } \theta > 0$$

global new error = propagated error + new error at this step

(result of applying the scheme on exact $y(t_n)$)

If the local truncation error $p_n = O(\Delta t^p)$ then $\|\delta_n\| \leq C \cdot \tau \|p_n\|$ if $\|(I - \theta \tau A)^{-1}\| \leq C$
 $= O(\Delta t^{p+1})$

$$\begin{aligned} \text{By recursion: } \varepsilon_n &= R(\tau A) \varepsilon_{n-1} + \delta_{n-1} \\ &= R(\tau A) [R(\tau A) \varepsilon_{n-2} + \delta_{n-2}] = R^2 \varepsilon_{n-2} + R \delta_{n-2} + \delta_{n-1} \\ &\vdots \\ &= R(\tau A)^n \varepsilon_0 + R^{n-1} \delta_0 + \dots + R \delta_{n-2} + \delta_{n-1} \end{aligned}$$

$$\Rightarrow \|\varepsilon_n\| \leq \|R(\tau A)^n\| \cdot \|\varepsilon_0\| + \sum_{j=0}^{n-1} \|R(\tau A)^{n-1-j}\| \|\delta_j\|$$

therefore we need stability: $\boxed{\text{If } \|R(\tau A)^n\| \leq K, \quad n = 0, 1, \dots, n \tau \leq t_{\max}}$

$$\text{then } \|\varepsilon_n\| \leq K \|\varepsilon_0\| + K \sum_{j=0}^{n-1} \|\delta_j\|, \quad \text{so if } \|\delta_j\| \leq C \cdot \tau^{p+1}$$

$$\leq K \|\varepsilon_0\| + K C (n \tau) \cdot \tau^p, \quad n \tau \leq t_{\max}$$

$$\leq K \|\varepsilon_0\| + K C t_{\max} \cdot \tau^p$$

If $\varepsilon_0 = Y^0 - y(0) = 0$ then $\varepsilon_n = O(\tau^p)$ same order as consistency error

\therefore stability \Rightarrow convergence, and by Lax Equivalence Thm also \Leftarrow

