A-stable high-order implicit time schemes

Mamadou N'DIAYE PhD student Department of Applied Mathematics

University of Pau & INRIA-Bordeaux Sud-Ouest, Team Magique-3D Advisors: Hélène BARUCQ & Marc DURUFLÉ

Conference Waves 2017 May 15-19, Minneapolis (Minnessota, USA)







Code Montjoie

- Research code maintained by Marc Duruflé at INRIA (open-source).
- Frequency-domain and time-domain modeling for wave propagation.

Main goals

- Several wave-field propagators
 - Acoustic, electromagnetic, aeroacoustic and elastodynamic.
- Different discretization methods
 - Spatial: FEM, DG
 - Time schemes: explicit and implicit one step, explicit multi-step.
- ✓ Uses several libraries like : Blas, Mumps, Lapack, SuperLU, GSL, ...

In my thesis we are concerned with the development of high-order:

- explicit time schemes with optimal CFL,
- A-stable implicit time schemes,
- and locally implicit time schemes for 1D, 2D and 3D wave equations.

We consider the following ODE (Ordinary Differential equation):

$$M_h \frac{dX}{dt} + K_h X = F(t)$$

where M_h is the mass matrix and K_h is the stiffness matrix, obtained after space discretization.

The ODE is stiff because

- the eigenvalues of $M_h^{-1}K_h$ can be large,
- due to the presence of **small elements** in the computational mesh,
- or high-order space discretization.

Consequence

- ✓ Explicit methods become too expensive.
- ✓ Implicit methods are needed.

Why implicit time schemes?



Spectrum of $M_h^{-1}K_h$ for the second order (Q2) LDG method with upwind fluxes on the regular grid $[-4,4]^2$. Acoustic wave equation with Neumann boundary condition.

Why implicit time schemes?



Spectrum of $M_h^{-1}K_h$ for LDG method with upwind fluxes on the regular grid $[-4,4]^2$ with 10 points along the *x* axis. Acoustic wave equation with Neumann boundary condition.

Objective

When we consider

$$y'(t) = i\lambda y(t), \quad y(t_0) = y_0, \quad t \in [t_0, T].$$
 (1)

The exact solution after one step $(h = t_{n+1} - t_n)$ is

$$y_n = \mathrm{e}^{ih\lambda} y_{n-1}.$$

Numerical solution is obtained by approximating

 $e^{ih\lambda} \approx R(ih\lambda).$

- The scheme is **A-stable** if $|\mathbf{R}(z)| \le 1, \forall z \in \mathbb{C}^-$,
- non dissipative if $|\mathbf{R}(z)| = 1, \forall z \in \mathbb{C}^-$,
- non dispersive if $\Phi(z) = z arg(\mathbf{R}(z)) = 0, \forall z \in \mathbb{C}^-$.

 $\forall z \in \mathbb{C}, R(z)$ is called the stability function of the numerical scheme.

Main goals

Find R(z) such that the scheme is **A-stable**, **low dissipative** and **low dispersive**.

Barucq, Duruflé and N'Diaye

- 1) General idea in this framework
- 2 Padé schemes
- 3 "Linear" Singly Diagonal Runge-Kutta schemes (Linear-SDIRK)
- Comparison Padé vs "Linear"-SDIRK schemes



In this framework...

We recall the ODE:

$$\begin{cases} M_h \frac{dX}{dt} + K_h X = F(t) \quad t \in (0,T] \\ X(0) = X_0 \end{cases}$$

with X_0 the initial condition.

After each step $\Delta t = t_{n+1} + t_n$ the analytical solution to (2) reads

$$X(t_{n+1}) = \mathbf{e}^{\Delta t \mathbf{A}} \left(X(t_n) + \int_0^{\Delta t} \mathbf{e}^{-u \mathbf{A}} M_h^{-1} F(n \Delta t + u) \, \mathrm{d}u \right),$$

where $A = -M_h^{-1}K_h$. The numerical solution is constructed by finding *R* such that

 $e^{\Delta tA} \approx R(\Delta tA).$

- *R* polynomial function \Rightarrow explicit scheme
- *R* rational polynomial function \Rightarrow implicit scheme

Barucq, Duruflé and N'Diaye

(2)

Stability functions

Let us define the stability function as:

$$\mathbf{R}(z) = \frac{N(z)}{D(z)}, \, \forall z \in \mathbb{C}^-,$$

where N(z) and D(z) are polynomials of z.

The stability region is defined as

 $S = \{z \in \mathbb{C} \text{ such that } |R(z)| \le 1\}.$

And the method is A-stable if and only if $S \supset \mathbb{C}^-$.

We consider two cases:

- *D*(*z*) has distinct roots
- D(z) has only one real root

Remark For the scheme to be A-stable the degree of D must be greater or equal to that of N.

Barucq, Duruflé and N'Diaye

Final numerical scheme

From the analytical solution

$$X(t_{n+1}) = \mathbf{e}^{\Delta t A} \left(X(t_n) + \int_0^{\Delta t} \mathbf{e}^{-u A} M_h^{-1} F(n \Delta t + u) \, \mathrm{d}u \right),$$

we take $e^{\Delta tA} \approx R(\Delta tA) = [D(\Delta tA)]^{-1}N(\Delta tA)$, the solution becomes:

$$D(\Delta tA)X(t_{n+1}) = N(\Delta tA)X(t_n) + \phi, \qquad (3)$$

where

$$\phi = N(\Delta tA) \int_0^{\Delta t} e^{-uA} M_h^{-1} F(n\Delta t + u) \, \mathrm{d}u + O(\Delta t^{p+1}).$$

To compute ϕ we rather use

$$\phi = D(\Delta tA)X(t_{n+1}) - N(\Delta tA)X(t_n)$$
(4)

and the final numerical schemes of order p reads

$$D(\Delta tA)X_{n+1} = N(\Delta tA)X_n + \phi_n$$
(5)

$$D(\Delta tA)X(t_{n+1}) = N(\Delta tA)X(t_n) + \phi,$$
(3)

To compute ϕ we use

$$\phi = D(\Delta tA)X(t_{n+1}) - N(\Delta tA)X(t_n)$$
(4)

and the final numerical schemes of order *p* reads

$$D(\Delta tA)X_{n+1} = N(\Delta tA)X_n + \phi_n$$
(5)

 ϕ_n is an approximation of ϕ (up to a term in $O(\Delta t^{p+1})$) in the form:

$$\phi_n = \sum_{r=1}^m A^{r-1} \Delta t^r \sum_{i=0}^{n_{\omega}-1} \omega_i^r F(t_n + \Delta t c_i)$$
(6)

where *m* is the degree of the polynomial *N* or *D* and n_{ω} depends on the scheme.



- Stability function of diagonal Padé schemes
- Computation of ϕ

Stability function of diagonal Padé schemes

We introduce the following notations

$$\rho_i^m = \binom{m}{i} \frac{(2m-i)!}{(2m)!} \quad \text{and} \quad C_k = \frac{1}{k! \ 2^{k-1}}, \quad \forall m, k \in \mathbb{N}.$$

The stability function of a diagonal Padé scheme of order 2m [Butcher, 2008] is given by:

$$\mathbf{R}_m(z) = \frac{N_m(z)}{D_m(z)},$$

where

$$N_m(z) = \sum_{i=0}^m \rho_i^m z^i$$
 and $D_m(z) = N_m(-z).$

Remark $R_m(z)$ is also the stability function of the *m*-stage Gauss Runge-Kutta schemes (Theorem 5.2 [Hairer and Wanner, 2010]).

Proposition

The stability function of numerical schemes obtained using the diagonal Padé approximation verify: $\forall z \in \mathbb{C}^-$

 $|\mathbf{R}_m(z)| \leq 1, \quad \forall m \in \mathbb{N}.$

Furthermore, if $z = ib, b \in \mathbb{R}$,

$$|\mathbf{R}_m(z)|=1, \quad \forall m\in\mathbb{N}.$$

Consequences

- The Diagonal Padé schemes are "always" stable,
- the Diagonal Padé schemes when applied to the test equation $y' = i\lambda y$ are not dissipative.

Stability function of diagonal Padé schemes



Dispersion of diagonal Padé schemes of order 2 to 10. $z = h\lambda$ and m, degree of D, is equivalent to the number of linear systems to be solved.

Barucq, Duruflé and N'Diaye

Computation of source term I

We recall here:

$$\phi = D_m(\Delta tA)X(t_{n+1}) - N_m(\Delta tA)X(t_n)$$
(7)

The Taylor expansion of $X(t_{n+1})$ and $X(t_n)$ around $t_n + \frac{\Delta t}{2}$ at order 2m gives

$$\phi = \sum_{i=2p, p \in \mathbb{N}}^{m} \rho_i^m (\Delta t A)^i \sum_{k=2q+1, q \in \mathbb{N}}^{2m} C_k \Delta t^k \mathbf{X}^{(k)}$$

$$- \sum_{i=2p+1, p \in \mathbb{N}}^{m} \rho_i^m (\Delta t A)^i \sum_{k=2q, q \in \mathbb{N}}^{2m} C_k \Delta t^k \mathbf{X}^{(k)} + O(\Delta t^{2m+1}).$$
(8)

To evaluate $X^{(k)}$ we derive the following relation (k-1)-times

$$\frac{dX(t)}{dt} - AX(t) = F(t),$$

to obtain:

$$\mathbf{X}^{(k)} = \sum_{j=1}^{k} A^{k-j} F^{(j-1)} + A^k X^{(0)}.$$
 (9)

 $F^{(j)}$ is the *j*-th derivative of the function *F* at point $t_n + \frac{\Delta t}{2}$, $F^{(0)} = F(t_n + \frac{\Delta t}{2})$. Using formula (9) in the expression of ϕ allows us to write:

Proposition

The expression of ϕ is equivalent to

$$\phi = \sum_{r=1}^{m} \Delta t^{r} A^{r-1} \sum_{j=1}^{2m-r+1} \alpha_{j-1}^{r} \Delta t^{j-1} F^{(j-1)} + O(\Delta t^{2m+1})$$
(10)

with

$$\mathbf{t}_{j}^{r} = \begin{cases} \sum_{i=0}^{\min(m,r-1)} (-1)^{i} \rho_{i}^{m} C_{r+j-i}, & \text{if } j-r \equiv 0[2] \\ 0 & \text{otherwise} \end{cases}$$
(11)

We consider the following approximation:

$$\sum_{i=0}^{2m} \alpha_i^r \Delta t^i F^{(i)}\left(t_n + \frac{\Delta t}{2}\right) \approx \sum_{i=0}^{m-1} \omega_i^r F\left(t_n + \Delta t c_i + \frac{\Delta t}{2}\right),$$
(12)

where c_i are quadrature points chosen in [0, 1] and ω_i^r are the weights.

- We choose Gauss-Legendre points for c_i,
- ω_i^r are solution of linear system of type Vandermonde.

Finally we replace c_i and ω_i^r in (12) to approximate ϕ :

$$\phi_n = \sum_{r=1}^m A^{r-1} \Delta t^r \sum_{i=0}^{m-1} \omega_i^r F(t_n + \Delta t c_i)$$



- Stability function of Linear"-SDIRK schemes
- Linear-SDIRK methods *s* + *l*-stages of order *s* + 1

We define the stability function of "Linear"-SDIRK schemes as follows:

$$R_{s}^{l}(z) = rac{N_{s}^{l}(z)}{D_{s}^{l}(z)} = rac{N_{s}^{l}(z)}{(1 - \gamma z)^{s+l}},$$

where γ is a real positive number and s + l is the number of stages. We recall the numerical scheme is now:

$$D_s^l(\Delta tA)X_{n+1} = N_s^l(\Delta tA)X_n + \phi_n$$

We propose to find the numerator $N_s^l(z)$ and γ such that:

- The method is A-stable $(|\mathcal{R}_{s}^{l}(z)| \leq 1, \forall z \in \mathbb{C}^{-}),$
- The method is of order s + 1.

The stability function of the $(s+1)^{th}$ -order "Linear"-SDIRK schemes is given by

$$\mathsf{R}_{s}^{l}(z) = 1 + z + \frac{z}{2!} + \dots + \frac{z^{s}}{s!} + \frac{z^{s+1}}{(s+1)!} + \frac{\alpha_{1}z^{s+2} + \dots + \alpha_{s+l}z^{2s+l+1}}{(1 - \gamma z)^{s+l}}.$$

where s is a fixed number of stage and l represents the number of additional stages.

 $R_{s}^{l}(z)$ is found by minimizing the coefficient η_{1} such that

$$e^{z} - R_{s}^{l}(z) = \eta_{1} z^{s+2} + O(z^{s+3})$$

subject to $|R_s^l(z)| \le 1$ (A-stability condition).

For l = 0, we search for the stability function R_s^0 of the form:

$$R_s^0(z) = 1 + z + \frac{z}{2!} + \dots + \frac{z^{s+1}}{(s+1)!} + \frac{\alpha_1 z^{s+2} + \dots + \alpha_s z^{2s+1}}{(1-\gamma z)^s}.$$

 \Rightarrow the method is by construction of order s + 1

- Reduce $R_s^0(z)$ to the common denominator
- Constants $\alpha_1, \alpha_2, \cdots, \alpha_s$ are set to cancel polynomials of degree $\geq s + 2$.
- The stability function takes the form

$$R_s^0(z) = \frac{a_0(\gamma) + a_1(\gamma)z + \cdots + a_s(\gamma)z^s + a_{s+1}(\gamma)z^{s+1}}{(1-\gamma z)^s}.$$

Linear-SDIRK methods s-stages of order s + 1

The stability function $R_s^0(z)$ reads:

$$R_s^0(z) = \frac{a_0(\gamma) + a_1(\gamma)z + \dots + a_s(\gamma)z^s + a_{s+1}(\gamma)z^{s+1}}{(1 - \gamma z)^s}$$

- Compute the roots of the polynomial in $a_{s+1}(\gamma)$.
- Choose γ such that $|R_s^0(z)| \leq 1$.

We retrieved the same coefficients as that presented in [Burrage, 1977]:

Minimal stage linear-SDIRK of order s+1					
S	Value for γ	comment			
1	0.5	-			
2	0.788675134594813	-			
3	1.068579021301629	-			
4	x	No A-stable schemes			
5	0.473268391258295	-			
6	х	No A-stable schemes			
$r \ge 7$	x	No A-stable schemes			

For l = 1, we search for the stability function R_s^1 of the form:

$$R_s^1(z) = 1 + z + \frac{z}{2!} + \dots + \frac{z^{s+1}}{(s+1)!} + \frac{\alpha_1 z^{s+2} + \dots + \alpha_{s+1} z^{2s+2}}{(1-\gamma z)^{s+1}}.$$

 \Rightarrow the method is by construction of order s + 1.

- Reduce $R_s^1(z)$ to the common denominator.
- Constants $\alpha_1, \alpha_2, \cdots$, α_{s+1} are set to cancel polynomials of degree $\geq s+2$.
- The stability function takes the form

$$R_s^1(z) = \frac{a_0(\gamma) + a_1(\gamma)z + \dots + a_{s+1}(\gamma)z^{s+1}}{(1 - \gamma z)^{s+1}}$$

Linear-SDIRK methods (s + 1)-stages of order s + 1

For l = 1:

- One free parameter γ
- $R_s^1(z)$ is found by minimizing η_1 :

$$e^{z} - R_{s}^{1}(z) = \eta_{1} z^{s+2} + O(z^{s+3})$$

• Choose γ such that $|R_s^1(z)| \leq 1$.

We retrieved the same interval for γ as that presented in [Burrage, 1977]:

Linear-SDIRK of order s+1 with one additional stage				
S	interval for γ	γ_{opt}		
1	$[\frac{1}{4},\infty[$	leads to third order		
2	$\left[\frac{1}{3}, 1.06866\right]$	leads to fourth order		
3	[0.39434, 1.28057]	0.394337567297406		
4	$[.24651, .3618] \cup [.42079, .47326]$	leads to sixth order		
5	[0.28407, 0.5409]	0.284064638011798		
6	x	No A-stable schemes		
7	[0.21705, 0.26471]	0.217049743094303		
$r \ge 8$	X	No A-stable schemes		

Linear-SDIRK methods (s + 2)-stages of order s + 1

For l = 2:

- Two free parameters γ and α_1
- $R_s^2(z)$ is found by minimizing η_1 :

$$e^{z} - R_{s}^{2}(z) = \eta_{1} z^{s+2} + O(z^{s+3})$$

• Subject to $|R_s^2(z)| \le 1$

Linear-SDIRK of order s+1 with two additional stages					
S	γ_{opt}	$\alpha_{1_{opt}}$	comment		
$1 \le s \le 4$	-	-	No unicity		
5	0.20394	0.000198393849206349	-		
6	-	-	leads to eighth order		
7	0.16688	$2.92592592592593e^{-06}$	-		
8	х	х	No A-stable schemes		
9	0.1419402	$2.29828167588584e^{-08}$	-		
$r \ge 10$	х	Х	No A-stable schemes		

Linear-SDIRK methods (s + 3)-stages of order s + 1

For l = 3:

- Three free parameters γ , α_1 and α_2
- $R_s^3(z)$ is found by minimizing η_1 :

$$e^{z} - R_{s}^{3}(z) = \eta_{1} z^{s+2} + O(z^{s+3})$$

• Subject to $|R_s^3(z)| \le 1$

Linear-SDIRK of order s+1 with three additional stages					
S	γ_{opt}	$\alpha_{1_{opt}}$	$\alpha_{2_{opt}}$		
$1 \le s \le 6$	-	-	-		
7	0.13588	$2.76727843915344e^{-06}$	$-3.45238095238095e^{-06}$		
8	-	-	-		
9	0.15176	$2.4583633958634e^{-08}$	$-4.31522566939234e^{-08}$		
10	х	х	х		
11	0.134	$1.65054852554853e^{-10}$	$-2.93702705657071e^{-10}$		
$r \ge 12$	х	Х	Х		



Comparison Padé vs "Linear"-SDIRK schemes

- Numerical dissipation and dispersion
- Convergence curves and numerical results in 1-D
- Numerical results in 2-D



Figure: Dispersion and dissipation curves of diagonal Padé schemes of order 4 compared with that of the Linear-SDIRK, when applied to the test equation (1). LS2 - l and LS3 - l represents the s = 2 and s = 3 plus l additional stages Linear-SDIRK of order 3 and 4.



Figure: Dispersion and dissipation curves of diagonal Padé schemes of order 6 compared with that of the Linear-SDIRK, when applied to the test equation (1). LS5 – l represents the s = 5 plus l additional stages Linear-SDIRK of order 6.



Figure: Dispersion and dissipation curves of diagonal Padé schemes of order 8 compared with that of the Linear-SDIRK, when applied to the test equation (1). LS7 – l represents the s = 7 plus l additional stages Linear-SDIRK of order 8.



Figure: Dispersion and dissipation curves of diagonal Padé schemes of order 10 compared with that of the Linear-SDIRK, when applied to the test equation (1). LS9 - l represents the s = 9 plus l additional stages Linear-SDIRK of order 10.

We consider the first order system for the acoustic wave equation:

$$\begin{cases} \rho \,\partial_t u - \operatorname{div} v = 0, \quad \forall (x,t) \in \Omega \times \mathbb{R}^+ \\ \mu^{-1} \partial_t v - \nabla u = 0, \quad \forall (x,t) \in \Omega \times \mathbb{R}^+ \end{cases}$$
(13)

where

- Ω is the computational domain,
- *n* is the outgoing normal to the considered boundary,
- ρ and μ are physical indexes, which are piecewise constant,
- the scalar field *u* and vector field *v* depend on the space *x* and time *t*.

The space discretization is performed using the mixed spectral element method implemented in the code Montjoie [Duruflé, 2006].

Convergence curves and numerical results in 1-D

The wave equation (13) is solved in 1-D with homogeneous media

$$\rho = \mu = 1$$

in the computational domain $\Omega = [0, 500]$.

 \checkmark An inhomogeneous Dirichlet condition is set on the left extremity

$$u(x=0,t) = e^{-i\omega t} \exp\left(-\frac{1}{2}\left(\frac{t-T}{\tau}\right)^2\right)$$

where

$$\omega = 2\pi, \quad \tau = \frac{20}{2\sqrt{2\log 2}}, \quad T = 100$$

✓ An homogeneous Neumann condition is set on the right extremity

- ✓ Mixed spectral element of order 16 is used in space, 16 points per wavelength, to have around 10^{-12} as space discretization error,
- \checkmark we choose [0, 1000] for the time interval

Convergence curves and numerical results in 1-D

The wave equation (13) is solved in 1-D with homogeneous media

 $\rho = \mu = 1$

in the computational domain $\Omega = [0, 500]$.

- ✓ An inhomogeneous Dirichlet condition is set on the left extremity
- $\checkmark\,$ An homogeneous Neumann condition is set on the right extremity
- ✓ Mixed spectral element of order 16 is used in space, 16 points per wavelength, to have around 10^{-12} as space discretization error,
- \checkmark we choose [0, 1000] for the time interval

For this case, we can compare the numerical results with the following analytical solution (before reflection)

$$u^{\text{exact}}(x,t) = e^{i\omega(x-t)} \exp\left(-\frac{1}{2}\left(\frac{t-T-x}{\tau}\right)^2\right)$$

After the first reflection, the solution *u* will be conjugated.
Convergence curves and numerical results in 1-D



Figure: Relative L^2 error between numerical solution and exact solution for t = 200 versus the time step. Comparison of diagonal Padé and Linear-SDIRK of order 4 and 6. LSs – *l* represents the *s* plus *l* additional stages Linear-SDIRK of order s+1. The space discretization error is about 10^{-12} .

Convergence curves and numerical results in 1-D



Figure: Relative L^2 error between numerical solution and exact solution for t = 200 versus the time step. Comparison of diagonal Padé and Linear-SDIRK of order 8 and 10. LSs – *l* represents the *s* plus *l* additional stages Linear-SDIRK of order s+1. The space discretization error is about 10^{-12} .

The number of time iterations and computational time needed to reach 1% of relative error are given:

Table: Computational time after imposing 1% of relative errors (1-D case) t = 1000

	Pade4	LS3-0	LS3-1	Pade6	LS5-0	LS5-1	LS5-2
Nb	33333	110000	25960	8360	18835	11540	7355
steps							
CPU	1mn36	5mn23	1mn48 <i>s</i>	37s	1mn31	1mn09	53s
time							
	Pade8	LS7-1	LS7-2	LS7-3	Pade10	LS9-2	LS9-3
Nb	3875	5830	4600	3700	2326	3106	2845
steps							
CPU	24s	46s	43 s	38s	17 s	34 s	34s
time							

Mesh of a resonant cavity

- 927 elements
- 1989 edges



(a) Local refinement



(b) Disk resonant mesh

- An homogeneous Neumann boundary condition is set on all the boundaries except at the external circle
- An ABC is set at the external of the circle

$$\mu \partial_n u + \sqrt{\rho \mu} \ \partial_t u = \mu \partial_n u^{\text{inc}} + \sqrt{\rho \mu} \ \partial_t u^{\text{inc}}, \quad \text{on } \Gamma_A$$

where the incident field is given by

$$u^{\text{inc}} = h(t - 1.5 - x)$$
 where $h(t) = \sin(\omega t)e^{-b(t-T)^2}$

with $\omega = 16\pi, \ b = 4, \ T = \sqrt{\log(10^6)}$

The wave equation (13) is solved in the 2-D resonant cavity mesh

- ✓ An homogeneous Neumann boundary condition is set on all the boundaries except at the external circle
- ✓ An ABC is set at the external of the circle
- $\checkmark\,$ Mixed spectral element method is used in space, to have around 10^{-6} as space discretization error,
- \checkmark We choose [0, 10] as time interval
- ✓ The reference solution is computed on this mesh with $\Delta t = 0.01$ and eighth-order Padé scheme

The number of time iterations and computational time needed to reach 0.1% of relative error are given:

Table: Computational time after imposing 0.1% of L^2 relative error for the scattering of a resonant cavity.

	Pade4	LS3-0	LS3-1	Pade6	LS5-0	LS5-1	LS5-2	
Nb	2370	7745	1955	623	1393 856		823	
steps								
CPU	5mn25	21mn17	7mn27	2mn16	6mn27	4mn49	5mn26	
time								
	Pade8	LS7-1	LS7-2	LS7-3	Pade10	LS9-2	LS9-3	
Nb	297	449	370	911	181	478	481	
steps								
UPU	1mn33	3mn23	3mn13	8mn40	1mn19	5mn05	5mn37	

Numerical results in 2-D

Using an homogeneous ABC and the initial condition

$$u = \exp\left(-7\frac{(x^2 + y^2)}{0.15^2}\right)$$

instead of 0 we found

Table: Computational time after imposing 0.1% of L^2 relative error with an initial condition.

			Pade	4	LS3	- 1	Pac	de6	LS	5 - 0	LS	5 - 1	LS	35 - 2]
	Nb		2500		2059		738		164	42	100	07	67	'9	
	step	S													
	CPL	ן נ	5mn3	1	7mn	30	2m	n37	7m	n38	5m	in41	4n	nn28	
	time	;													
		Ρ	ade8	LS	7 - 1	LS7	7 - 2	LS7	-3	Pade	e10	LS9-	-2	LS9-	3
Ν	lb	- 38	81	560	0	442		365		248		316		298	
S	teps														
С	PU	1	mn53	4m	n18	3mi	n49	3mn	29	1 mn	36 <i>s</i>	3mn1	0	3mn1	7
ti	me														

Simulation results in 2-D



Simulation results of the wave equation (13) in 2D on the irregular mesh resonant cavity at t = 2, ..., 7.

Barucq, Duruflé and N'Diaye

Conclusion

✓ Padé schemes are more efficient and more robust than Linear-SDIRK schemes

Ongoing works

- Optimization of CFL for explicit schemes ✓.
- Development of locally implicit schemes in 1D and 2D ✓.
- Validation tests of the locally implicit implemented.

Perspectives

- Improve the treatment of the source for Linear-SDIRK.
- Test the **locally** implicit method for 3-D Maxwell's equations.
- Perform validation tests with realistic wave problems.

Conclusion

✓ Padé schemes are more efficient and more robust than Linear-SDIRK schemes

Ongoing works

- Optimization of CFL for explicit schemes ✓.
- Development of locally implicit schemes in 1D and 2D ✓.
- Validation tests of the locally implicit implemented.

Perspectives

- Improve the treatment of the source for Linear-SDIRK.
- Test the **locally** implicit method for 3-D Maxwell's equations.
- Perform validation tests with realistic wave problems.

Thank you for your attention. Questions?







Burrage, K. (1977).

A special family of runge-kutta methods for solving stiff differential equations.

BIT, pages 22–41.



Butcher, J. C. (2008).

Numerical Methods for Ordinary Differential Equations SECOND EDITION.

John Willey & Sons Ltd.



Duruflé, M. (2006).

Intégration numérique et éléments finis d'ordre élevé appliqués aux équations de Maxwell en regime harmonique. PhD thesis, Université Paris Dauphine.

Hairer, E. and Wanner, G. (2008). Solving ordinary differential equations I Nonstiff problem. Springer.

Hairer, E. and Wanner, G. (2010).

Solving ordinary differential equations II Stiff and differential-algebraic problem. Springer.



Skvortsov, L. M. (2006). Diagonally implicit runge-kutta methods for stiff problems. *Computational Mathematics and Computational Physics*, 46:2110–2123.

How to efficiently implement these schemes





Hints and tips for an efficient implementation of Padé schemes

Hints and tips for the implementation I

We take $C = \Delta t A$ and we recall the numerical scheme given

$$D_m(C)(X_{n+1} - X_n) = (N_m - D_m)(C)X_n + \phi_n$$

We note $G = (N_m - D_m)(C)X_n + \phi_n$ and $Y = X_{n+1} - X_n$. Let λ_k be zeroes of D_m :

- All λ_k , $1 \le k \le m$ are complex conjugate if *m* is even
- Unless there is one real root, the m − 1 other roots are complex conjugate
- We write *D_m* as a product of second degree polynomial factors

Assuming *m* is even, the scheme read then

$$Y_{0} = G,$$

$$\prod_{k=1}^{m/2} \underbrace{\left(I - \frac{C}{\lambda_{k}}\right) \left(I - \frac{C}{\overline{\lambda_{k}}}\right)}_{P_{2}(C)} Y_{k} = Y_{k-1}.$$
(14)

Hints and tips for the implementation II

Let
$$a_k = -rac{1}{\lambda_k}$$
 and $ar{a}_k = -rac{1}{\overline{\lambda}_k}$, then

$$P_2(C) = (I + a_k C) \left(I + \overline{a}_k C \right) = I + 2Re(a_k)C + a_k \overline{a}_k C^2.$$

A partial fraction decomposition of the polynomial $P_2^{-1}(x) = \frac{1}{(1 + a_k x)(1 + \bar{a}_k x)}$ allows to write

$$P_2^{-1}(C) = b_k (I + a_k C)^{-1} + \overline{b}_k (I + \overline{a}_k C)^{-1},$$
(15)

with $b_k = \frac{a_k}{a_k - \bar{a}_k}$. Then Y_k can be computed as follows:

$$(I + a_k C) u = Y_{k-1},$$

$$(I + \overline{a}_k C) v = Y_{k-1},$$

$$Y_k = b_k u + \overline{b}_k v.$$

Assuming Y_{k-1} is a real vectors we get

$$\mathbf{v} = (I + \bar{\mathbf{a}}_k C)^{-1} Y_{k-1} = \bar{\mathbf{u}}$$

Hints and tips for the implementation III

It suffices to solve only one system, giving the following algorithm:

$$(I + a_k C) u = Y_{k-1},$$

$$Y_k = b_k u + \overline{b}_k \overline{u} = 2 \operatorname{Re}(b_k u).$$
(16)

- For Y_{k-1} complex, we solve twice (16). One for the real part and one for the imaginary part.
- If *m* is odd and λ the only real root of D_m (14) becomes

$$\begin{pmatrix} I - \frac{C}{\lambda} \end{pmatrix} Y_0 = G,$$

$$\prod_{k=1}^{(m-1)/2} \left(I - \frac{C}{\lambda_k} \right) \left(I - \frac{C}{\overline{\lambda_k}} \right) Y_k = Y_{k-1}.$$
(17)

Remark For diagonal Padé schemes of order 2m we need to solve at most $\frac{m}{2}$ complex systems of size *N*, *N* being the number of unknowns.

Linear-SDIRK methods *s*-stages of order s + 1

Order 2: for the method of order 2 we take s = 1 and l = 0. The stability function is computed from:

$$R_1^0(z) = 1 + z + rac{z^2}{2} + rac{lpha_0 z^3}{(1 - \gamma z)}.$$

We reduce to the common denominator:

$$R_1^0(z) = \frac{1 + (1 - \gamma)z + \left(\frac{1}{2} - \gamma\right)z^2 + \left(\alpha_0 - \frac{\gamma}{2}\right)z^3}{(1 - \gamma z)}.$$

In order to have a A-stable scheme, we need to satisfy at least:

$$\gamma = \frac{1}{2}$$
 and $\alpha_0 = \frac{\gamma}{2}$.

As a result, we obtain

$$R_1^0(z) = \frac{1+\frac{z}{2}}{1-\frac{z}{2}},$$

which is the same as the stability function of the Crank-Nicolson scheme.

Linear-SDIRK methods s-stages of order s + 1

Order 3: to construct method of order 3 we take s = 2 and l = 0. We search α_1 , α_2 and γ such that:

$$R_2^0(z) = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{\alpha_1 z^4 + \alpha_2 z^5}{(1 - \gamma z)^2}$$

After reducing to the common denominator, α_1 , α_2 and γ are solution of:

$$\alpha_2 + \frac{\gamma^2}{6} = 0, \quad \alpha_1 - \frac{\gamma}{3} + \frac{\gamma^2}{2} = 0,$$
(18)
$$\frac{1}{6} - \gamma + \gamma^2 = 0.$$
(19)

The value that leads to *A*-stable scheme is for $\gamma = \frac{1}{2} + \frac{1}{2\sqrt{3}}$, since the modulus of the asymptote of $R_2^0(z)$ when *z* tends to $+\infty$ satisfies

$$\frac{\frac{1}{2}-2\gamma+\gamma^2}{\gamma^2}\bigg|<1,$$

which is a necessary condition to have an A-stable scheme.

Barucq, Duruflé and N'Diaye

Linear-SDIRK methods *s*-stages of order s + 1

Order 3: to construct method of order 3 we take s = 2 and l = 0. We search α_1 , α_2 and γ such that:

$$R_2^0(z) = 1 + z + \frac{z^2}{2!} + \frac{z^3}{3!} + \frac{\alpha_1 z^4 + \alpha_2 z^5}{(1 - \gamma z)^2}.$$

After reducing to the common denominator, α_1 , α_2 and γ are solution of:

$$\alpha_{2} + \frac{\gamma^{2}}{6} = 0, \quad \alpha_{1} - \frac{\gamma}{3} + \frac{\gamma^{2}}{2} = 0,$$
(18)
$$\frac{1}{6} - \gamma + \gamma^{2} = 0.$$
(19)

The obtained stability function is then given by

$$R_2^0(z) = \frac{1 + (1 - 2\gamma)z + \left(\frac{1}{2} - 2\gamma + \gamma^2\right)z^2}{(1 - \gamma z)^2}.$$

General case:

Let R_s^0 be the stability function of the scheme. We search for R_s^0 of the form:

$$\mathcal{R}_{s}^{0}(z) = 1 + z + \frac{z}{2!} + \dots + \frac{z^{s+1}}{(s+1)!} + \frac{\alpha_{1}z^{s+2} + \dots + \alpha_{s}z^{2s+1}}{(1-\gamma z)^{s}}.$$

We note the polynomial *P* that appears while reducing to the common denominator:

$$P(z) = (1 - \gamma z)^{s} \left(1 + z + \frac{z^{2}}{2!} + \dots + \frac{z^{s+1}}{(s+1)!} \right)$$

= $a_{0}(\gamma) + a_{1}(\gamma)z + \dots + a_{2s+1}(\gamma)z^{2s+1}.$

The constants α_i are chosen to balance higher-order terms of *P*, i.e.

$$\alpha_i = -a_{s+1+i}(\gamma), \quad i = 1 \dots s.$$

General case:

A necessary condition for A-stability property is that the term in z^{s+1} vanishes:

$$a_{s+1}(\gamma) = \sum_{i=0}^{s} \frac{(-\gamma)^{i} {s \choose i}}{(s+1-i)!} = 0.$$
 (18)

Finally, the numerator of $R_s^0(z)$ is given as

$$N_s^0(z) = a_0(\gamma) + a_1(\gamma)z + a_2(\gamma)z^2 + \cdots + a_s(\gamma)z^s.$$

We have

$$R_s^0(z) = \frac{N_s^0(z)}{(1-\gamma z)^s}.$$

General case:

A necessary condition for A-stability property is that the term in z^{s+1} vanishes:

$$a_{s+1}(\gamma) = \sum_{i=0}^{s} \frac{(-\gamma)^{i} {s \choose i}}{(s+1-i)!} = 0.$$
 (18)

To ensure the A-stability condition we choose γ as follows:

- for each γ root of $a_{s+1}(\gamma)$,
- if $\lim_{z\to\infty} \frac{R_s^0(z)}{z} \leq 1$:
 - we compute the roots of the following polynomial equation:

$$|N(i\sqrt{z})|^{2} = |D(i\sqrt{z})|^{2}.$$
(19)

- If this polynomial equation has no real roots except zero, then the scheme is A-stable,
- otherwise the scheme is not A-stable.
- Else no A-stable schemes.

Handling a source term

Again we recall

$$\phi = D_s^l(C)X(t_{n+1}) - N_s^l(C)X(t_n) + O(\Delta t^{s+2}).$$

After much work, like for Padé schemes, we end up with

$$\phi = \Delta t \sum_{r=1}^{s+1} (\Delta tA)^{r-1} \sum_{j=1}^{s+2-r} \alpha_{j-1}^{r,l} \Delta t^{j-1} F^{(j-1)} + O(\Delta t^{s+2})$$

where

$$\alpha_{j-1}^{r,l} = \sum_{i=0}^{\min(r-1,s+l)} \frac{1}{2^{r+j-i-1}(r+j-i-1)!} \left(D_i - (-1)^{r+j-i-1} N_i \right).$$
(20)

And its approximation is given by

$$\phi_n = \sum_{r=1}^{s+l} A^{r-1} \Delta t^r \sum_{i=0}^s \omega_i^{r,l} F\left(t_n + \Delta t \, c_i\right).$$

Maxwell's equations

$$\begin{cases} \nabla \cdot D = \rho_v & \text{Gauss's law} \\ \nabla \cdot B = 0 & \text{Gauss Magnetism law} \\ \nabla \times H = \frac{\varepsilon \partial E}{\partial t} + J & \text{Amper's law} \\ \nabla \times E = -\frac{\partial B}{\partial t} & \text{Faraday's law} \end{cases}$$
(21)

where

- E is the electric field,
- *H* is the magnetic field,
- $J = \sigma E$, from Ohm's law, is the electric current density,
- $D = \varepsilon E$ is the electric flux density with ε the permittivity of the medium,
- $B = \mu H$ is the magnetic flux density with μ the permeability of the medium,
- σ is the electric conductivity of the medium.

Derivation of the wave equations

We consider the vector identity

$$\nabla \times \nabla \times H = \nabla (\nabla \cdot H) - \Delta H.$$

We suppose we are in a free source region, then

$$\nabla(\nabla \cdot H) = 0.$$

From Faraday's law, we have

$$\nabla \times \nabla \times E = -\mu \frac{\partial}{\partial t} (\nabla \times H).$$

By substituting in Amper's law, we get

$$egin{aligned} \nabla imes
abla imes E &= -\Delta E = -\mu rac{\partial}{\partial t} (
abla imes H) \ &= -\mu arepsilon rac{\partial}{\partial t} \left(rac{\partial E}{\partial t} + J
ight). \end{aligned}$$

Since J = 0 because of source free region, the vector wave equation reads

$$\Delta E = \mu \varepsilon \frac{\partial^2 E}{\partial t^2}.$$
 (22)



Some classical Runge-Kutta schemes

The *s*-stage RK scheme applied to $y'(t) = i\lambda y(t)$ is given by

$$y_{n+1} = y_n + h \sum_{j=1}^s b_j K_j, \quad K_j = i\lambda \left(y_n + h \sum_{l=1}^s a_{j,l} K_l \right).$$
 (23)

Using the Butcher table [Butcher, 2008], it can be represented as follows

where **c** corresponds to the position of stage value within the time-step. *A* is the matrix $(a_{j,l})$ and vector **b** corresponds to the weight coefficients of the quadratic formula.

The *s*-stage RK scheme applied to $y'(t) = i\lambda y(t)$ is given by

$$y_{n+1} = y_n + h \sum_{j=1}^{s} b_j K_j, \quad K_j = i\lambda \left(y_n + h \sum_{l=1}^{s} a_{j,l} K_l \right).$$
 (23)

Which is equivalent to

$$y_{n+1} = R(ih\lambda)y_n$$

where *R* is the stability function defined by

$$\mathbf{R}(z) = \mathbf{1} + zb^{T}(I - zA)^{-1}\mathbf{1}, \quad z \in \mathbb{C}.$$
 (24)

A is the matrix $(a_{i,l})$ and b is the vector (b_i) .

For implicit RK methods R(z) is a rational polynomial function while for explicit RK methods it is a polynomial function [Hairer and Wanner, 2010] and [Hairer and Wanner, 2008].

Here are some ERK schemes (in Butcher table):



Stability region, black when $|R(z)| \leq 1$.

We considered the following IRK methods:

- SDIRK34 of Crouzeix [Hairer and Wanner, 2010].
- S54b of Skvortsov [Skvortsov, 2006].
- SDIRK55 of Cooper & Sayfy [Hairer and Wanner, 2010].



Stability region, black when $|R(z)| \le 1$.



Dissipation





Dispersion









Application to 1D and 2D Maxwell's equations

1D test case

- \checkmark 10-th order FEM is used in space.
- $\checkmark\,$ Time interval: [0,40], domain with 20 λ and 10,000 dofs.

Validation tests of the Padé schemes (1D test case)

Computational times after imposing 1% of errors							
Schemes Time steps Relative errors Times in s							
Padé2	0.004	0.009692	1.535				
Padé4	0.069	0.010023	0.379				
Padé6	0.196	0.009752	0.242				
Padé8	0.350	0.008919	0.195				
ERK44 0.006 0.009978 1.665							
SDIRK34	0.022	0.011980	0.845				
SDIRK55	0.08	0.009847	0.439				

Simulation results for Maxwell's equations in 2D

2D test case on regular square mesh

- \checkmark 10-th order FEM is used on regular mesh $[-4, 4]^2$.
- \checkmark Time interval: [0, 10].

Validation tests of the Padé schemes (2D regular square mesh)

Computational times after imposing 0.1% of errors.								
Schemes Time steps Relative errors Times in s								
Padé2	0.003	0.001241	1713.04					
Padé4	0.054	0.000988	181.09					
Padé6	0.153	0.000871	183.84					
Padé8	0.285	0.001051	149.50					
SDIRK34	0.0166	0.001030	985.66					
SDIRK55	0.066	0.001244	455.39					
Simulation results for Maxwell's equations in 2D

2D test case on irregular mesh

 $\checkmark~$ 6-th order FEM is used on the irregular mesh resonant disk.

 \checkmark Time interval: [0, 10].

Simulation results on the irregular mesh resonant disk

Comparison Padé vs ERK44			
Schemes	Time steps	Relative errors	Times in s
Padé4	0.0077	4.97×10^{-4}	8mn02s
Padé6	0.0286	4.62×10^{-4}	4mn42s
Padé8	0.05	4.99×10^{-4}	3mn51s
Padé4	5.00×10^{-4}	1.90×10^{-9}	296mn19s
Padé6	1.67×10^{-3}	1.69×10^{-9}	61mn01s
Padé8	6.67×10^{-3}	1.69×10^{-9}	18mn20s
ERK44	7.20×10^{-5}	$39.13 \times 10^{+3}$	30mn55s
ERK44	6.67×10^{-5}	1.69×10^{-9}	33mn21s